Production Planning with Load Dependent Lead Times and Sustainability Aspects

Institute of Information Systems
Department of Business Sciences
University of Hamburg

In Partial Fulfillment of the Requirements for the Degree of

Doktor der Wirtschaftswissenschaften
(Dr. rer. pol.)

Cumulative Dissertation

submitted by

Julia Pahl

Head of board of examiners: Prof. Dr. Knut Haase
First examiner: Prof. Dr. Stefan Voß
Second examiner: Prof. Dr. Hartmut Stadtler
Date of thesis discussion: 18. May 2012

Contents

Table of Contents **1**

I Framework of the Thesis **2**

1 Production Planning with Load-Dependent Lead Times and Sustainability Aspects **3**
1.1 List of Related Research Articles and Reports 4
1.2 Course of Research 5
1.3 Conclusions and Research Directions 7
References 9

2 Cumulative Doctoral Thesis **10**
2.1 Three Thematically Related Research Articles and Reports 10
2.2 Co-Authors and Substantial Contribution of Candidate 11
2.3 Publication of Research Articles and Reports 12

3 Curriculum Vitae **13**

II Literature **21**

Part I

Framework of the Thesis

Chapter 1

Production Planning with Load-Dependent Lead Times and Sustainability Aspects

The research contained in this thesis was undertaken partly as an external doctoral candidate and partly as a research and teaching assistant at the Institute of Information Systems, University of Hamburg. It contains eight articles and a technical report in the field of aggregate production planning and supply chain management. The research question immanent to this work is how lead times that are load dependent are taken into account in mathematical models for the tactical planning level, how they influence other planning levels and the resulting production plan. There is empirical evidence that lead times exponentially increase with the increase of capacity utilization measured in workload or work-in-process in the production system long before the capacity limit is reached. This can lead to significant differences in planned and realized lead times. Abstraction from such nonlinearities frequently takes place, mainly in favor of complexity reduction of mathematical models for use in practice and implementation into standard software such as advanced planning systems or enterprise resource planning. Variabilities of lead times may become significant when targeting 100% resource utilization. This leads to longer waiting times of production parts and products and lead to quality losses of such items. In the worst case, they cannot be used for their original purpose, anymore, and have to be discarded or reworked if technically possible. Therefore, another research question is how quality losses and lifetime restrictions are taken into account in the literature up to date. The latter mentioned aspects pertain to the highly dynamic research field of green supply chain management that contains, among others, questions of rendering production processes more environmentally friendly. This ranges from product design that influences all fields of production to optimal recycling, remanufacturing, and rework processes. It further comprises actions from wastage or disposal reductions also in terms of energy usage translated in related costs that not only imply CO_2 emission reductions, but also long-term recovery actions of exploited lands and environments. Related literature surveys show that there do not exist mathematical formulations taking into account all mentioned aspects. We closed this gap by developing discrete dynamic models formulated as *mixed integer programming* (MIPs) including production smoothing that accounts for *load-dependent lead times* (LDLT) thus aiming at avoiding peaks of resource utilization together with lifetime constraints of items as well as rework of items that pass their useful lifetime. Moreover, remanufacturing of externally returned items is integrated together with waiting-dependent rework processing times.

1.1 List of Related Research Articles and Reports

This cumulative dissertation contains the following selection of research articles and reports.

Peer-Reviewed Articles in Journals

1. Production Planning with Load Dependent Lead Times (with S. Voß and D.L. Woodruff), *4OR: A Quarterly Journal for Operations Research*, 3(4): 257-302, 2005, Abbreviation in Figure 1.1: 4OR (05).

2. Production Planning with Load Dependent Lead Times: An Update of Research (with S. Voß and D.L. Woodruff), *Annals of Operations Research* 153(1): 297-345, 2007, Abbreviation in Figure 1.1: Annals of OR (07).

Peer-Reviewed Articles in Journals Under Review

3. Depreciation Effects in Production Planning and Supply Chain Management: A Survey, invited review of *European Journal of Operations Research*, under review, 2011, Abbreviation in Figure 1.1: EJOR (11).

Peer-Reviewed Conference Proceedings

4. Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling (with S. Voß and D.L. Woodruff), In: Kotzab H., S. Seuring S., Müller M., and Reiner G. (Eds.), Research Methodologies in Supply Chain Management, Physica, Heidelberg, 540–554, 2005, Abbreviation in Figure 1.1: RMSCM (05).

5. Supply Chain Integration: Improvements of Global Lead Times with SCEM (with A. Mies and S. Voß) in: Geldermann J., Treitz M., Schollenberger H., Rentz O. (Eds.) Challenges for Industrial Production, Workshop of the PepOn Project: Integrated Process Design for Inter-Enterprise Plant Layout Planning of Dynamic Mass Flow Networks, Karlsruhe, Nov. 7-8., 79–89, 2005, Abbreviation in Figure 1.1: SCEM (05).

6. Production Planning and Deterioration Constraints: A Survey (with S. Voß and D.L. Woodruff), In: Ceroni J.A. (Ed.) The Development of Collaborative Production and Service Systems in Emergent Economies, *Proceedings of the 19th International Conference on Production Research* (IFPR), Valparaiso, Chile, 1–6, 2007, Abbreviation in Figure 1.1: IFPR (07).

7. Discrete Lot-Sizing and Scheduling with Sequence-Dependent Setup Times and Costs including Deterioration and Perishability Constraints (with S. Voß and D.L. Woodruff), *IEEE*, HICSS-44, 1–10, 2011, Abbreviation in Figure 1.1: HICSS-44 (11).

Lecture Notes

8. Discrete Lot-Sizing and Scheduling Including Deterioration and Perishability Constraints (with S. Voß), In: Danglmainer W., Blecken A., Delius R., and Klöpfer S. (ed.), In: Advanced Manufacturing and Sustainable Logistics, Proceedings of the 8th International Heinz Nixdorf Symposium, IHNS 2010, *Lecture Notes in Business Information Processing*, Springer, Berlin, Heidelberg, 345–357, 2010, Abbreviation in Figure 1.1: LNBIP (10).

Working Papers and Technical Reports

9. Production Planning with Load Dependent Lead Times and Sustainability Aspects, Technical Report, Institute of Information Systems, University of Hamburg, 2012, Abbreviation in Figure 1.1: Tech Rep (12).

Further research not included, but undertaken in this time period that complements this thesis is given subsequently (see also the Curriculum Vitae, Chapter 3 on Page 3):

Peer-Reviewed Conference Proceedings

10. Integrated Aircraft Scheduling Problem: An Auto-Adapting Algorithm to Find Robust Aircraft Assignments for Large Flight Plans (with T. Reiners, M. Maroszek, C. Rettig), *IEEE*, HICSS-45, 2012.

Books and Manuscripts

11. Production Planning with Deterioration and Perishability, Draft Manuscript, 2011.
12. Logistik Management - Systeme, Methoden, Integration (with S. Voß and S. Schwarze (Ed.)), Logistics Management - partly in German and partly in English, Physica, Heidelberg, 2009.
13. Network Optimization - International Network Optimization Conference, INOC 2011 (with T. Reiners and S. Voß (Eds.)), Hamburg, Germany, 2011, *Lecture Notes in Computer Sciences (LNCS)*, Springer, Heidelberg, 2011.

Working Papers

14. WIP Management in Practice: Survey and Analysis (with F. Schulte), Institute of Information Systems, University of Hamburg, 2011.
15. Efficient Formulations for Lot-Sizing Models with Perishability (with C. Seipl), Institute of Information Systems, University of Hamburg, 2011.

1.2 Course of Research

The course of research is depicted in Figure 1.1. Abbreviations of related articles and reports are provided in Section 1.1. Investigations of LDLT and their influences on planning and production outputs present an introduction into the research field; see Pahl et al. (2005b) [RMSCM (05)], Pahl et al. (2005c) [4OR (05)], Pahl et al. (2007a) [Annals of OR (07)]. Special emphasis is on aggregate or mid term planning for production and supply chain planning. The operational and control level is addressed by analyzing methods and tools to enhance communication of supply chain partners; see Pahl et al. (2005a) [SCEM (05)]. A related tool that emerged approximately a decade ago is *supply chain event management* (SCEM). It identifies deviations between original plans and their execution across the whole supply network and propose corrective actions according to predefined rules in order to improve decision making, shorten lead times, and decrease variability/disruptions of processes. Research revealed that methods and software tools for SCEM are still in its infancy and far from being of practical employment, but they can provide a database on process information

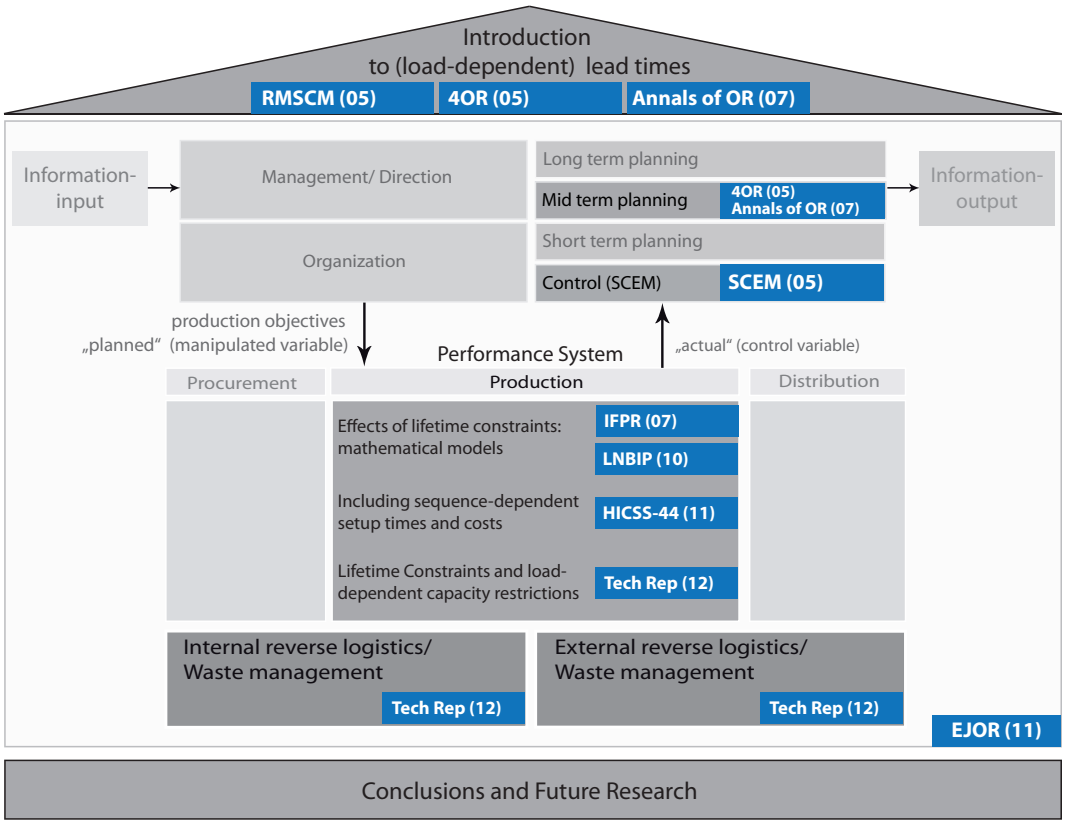

Figure 1.1: Framework of this thesis.

that can be analyzed for performance improvement especially regarding lead times and its variabilities.

The subsequent body of research investigates the question of how lifetime constraints of items and, in general, value loss, are taken into account in mathematical models for aggregate production and supply chain planning; see Pahl et al. (2007b) [IFPR (07)] and Pahl (2011) [EJOR (11)]. The book presented in Pahl and Voß (2011) discusses and analyzes mathematical formulations for depreciation effects. It is still in preparation and, therefore, not included in this thesis. The mentioned reviews show that depreciation and quality loss of products are taken into account in various fields of production planning ranging from procurement including ordering and replenishment policies to inventory management as well as lot sizing and rework/recycling management. A classification scheme of approaches is provided in Pahl (2011). These approaches are gouped regarding their assumed quality of information and with respect to their regarded time horizon. The survey reveals that little research is done regarding multiple products and thus setup times and costs. The major number of approaches integrates *shrinkage* of stored items due to deterioration or perishability into their models. Lifetime constraints that limit storage times are rarely incorporate. It is further revealed that deterministic model formulations prevail over stochastic as well as static over dynamic approaches. A small number of discrete dynamic models including lifetime constraints exist, but they are not extensively explored although their importance for planning is significant due to their aim to avoid quality losses and related disposals in the first place. Furthermore, such approaches may be integrated in frequently employed planning software such as ERP and solved using exact methods or heuristic algorithms.

Lifetime constraints for integration in classical inventory and lot size models are developed and their effects on lot sizing, production planning, and inventory management studied

in Pahl and Voß (2010) [LNBIP (10)]. Decisions on the production mix are important regarding LDLT, because their variations significantly influence resource utilization. Therefore, special emphasis is on setup times and costs that are assumed to be sequence-dependent in Pahl et al. (2011) [HICSS-45 (11)]. Hybrid lot sizing models including some scheduling information are regarded. The inclusion of lifetimes is also interesting from a mathematical point of view, because the implication of a maximal lifetime might reduce setup variables for periods exceeding these lifetimes. Numerical studies show that short lifetimes permit fast solution finding due to the application of lot-for-lot policy, i.e., no aggregation of lot sizes and thus no inventory holding takes place and demand is satisfied in the related period. This is not the case for longer lifetimes that permit different production-inventory variations.

The results of the aforementioned research state highlight regarding discrete dynamic deterministic approaches for aggregate production planning and a lack of models that include LDLT while accounting for sustainability aspects such as the avoidance of quality losses/disposals or wastages due to deterioration and perishability, lifetime constraints, as well as rework of passed items. This gap is closed in Pahl (2012) [Tech Rep (12)].

1.3 Conclusions and Research Directions

In this thesis, extensive research on models regarding LDLT and depreciation constraints is conducted. They highlight research gaps and promising directions for further research. In the technical report, we developed optimization models for aggregate production planning based on classical lot sizing models that take into account the problem of lead times that are dependent on the workload of production systems. This phenomenon is prevailing in practice and should be regarded in planning. Classical methods like material requirements planning, material resource planning that are implemented in enterprise resource planning systems do not provide related tools or planning methods. An increase of workload in the system leads to increases of lead times, so that items have to wait at various steps of the production system. There is empirical evidence that lead times increase exponentially with the workload long before 100% resource utilization is reached, so that planned and realized lead times vary significantly. Additionally empirical evidence shows that flustering due to the workload of a production resource takes place, so that outputs are not produced at a specific constant production rate, but at a fraction of the original rate, so that planned outputs are not realized. There are few model approaches accounting for the link between order releases, planning and capacity decisions to lead times taking into account the system workload as well as lot sizing and sequencing decisions. We studied different model formulations that directly approach the problem with lead times that are load dependent using diverse methods. These have their advantages and disadvantages. Therefore, the choice was made to pursuit with the proposed approach that account for lot sizing decisions in order to directly capture the planning circularity regarding lot sizes, setups, and capacity utilization. We developed different formulations for production smoothing for the tactical planning level. Up to date, production smoothing is rather accounted for the operational planning level regarding scheduling decisions mainly in the environment of synchronized assembly line systems. No model approaches have been proposed for the tactical planning level regarding lot sizing, so far. Our production smoothing approaches distribute production jobs evenly or in a way that a certain target utilization sector is achieved, so that LDLT are avoided.

Besides, a major objective of this work is to regard effects of LDLT on the quality of items. If items have a certain lifetime, i.e., time interval in which they are useful for further employment in production or consumption, lead times that are load-dependent might lead to enhanced quality losses of waiting items. In the worst case, they cannot be used for demand

satisfaction and need to be discarded or reworked if technically possible. The consideration of limited lifetimes is a hard constraint on inventory management and its integration into mathematical models for lot sizing, production, and inventory management an important issue. One may argue that longer storage times are avoided by mathematical models by accounting for inventory holding costs in the objective function. This is true, but there is a difference between being able to store products for long time periods and not being able due to lifetime constraints. In this regard, we further regarded aspects of sustainability such as avoidance of disposals caused by limited lifetimes of items, rework and remanufacturing issues. We proposed related models further including features, e.g., overtime, minimal lot size constraints, and setup carry overs in order to minimize their effects on resource utilization thus already taking into account some scheduling at this stage. Two models are developed based on the capacitated lot sizing problem with perishability that include possibilities of remanufacturing of externally returned items and rework of internally passed items that is waiting dependent. An extensive numerical study on this model is executed analyzing functionality and characteristics regarding complexity and its implications on the solution finding process. The influence of constraints, e.g., minimal lot size constraints, lifetimes, workloads, etc. are studied. Two different returns rates of items were assumed that significantly influenced results, so that return rates turn out to be a very important input factor to the model which confirms conclusions from other authors that the return or recycling rates significantly influences the planning process. Besides, minimal lot size constraints are a complicating factor of planning, not only regarding lifetime constraints, but the overall planning process. Therefore, more research is needed on this influencing factor.

Further research directions that are pursuited are the development of network flow formulations for proposed models and their comparison regarding complexity and efficiency of solution finding. Besides, we are interested in integrating finished goods inventory distribution decisions to the customer accounting for aspects of packaging and decision making on transportation modes in order to maintain quality while minimizing negative influences on the environment. This includes the reformulation of different objectives in the objective function of mathematical models that are dominated by profit maximization or cost minimization up to date. Additionally, stochastic influences are prevalent regarding LDLT, so that another research direction is to study the influence of stochastic events on the model behavior and the resulting plan using szenario techniques and deterministic/stochastic decomposition approaches. With the introduction of the GSCM and related research, a change is perceptible in the research literature. We pursuit this important topic by enhancing consciousness and sensitivity regarding sustainability in the production (planning) process.

Bibliography

J. Pahl. Integrating deterioration and lifetime constraints in production and supply chain planning: A survey. Invited survey; under review at the European Journal of Operations Research, 2011.

J. Pahl. Production planning with load dependent lead times and sustainability aspects. Technical report, Institute of Information Systems, University of Hamburg, 2012.

J. Pahl and S. Voß. Production planning with deterioration and perishability. Draft Manuscript, 2011.

J. Pahl and S. Voß. Discrete lot-sizing and scheduling including deterioration and perishability constraints. In *Advanced Manufacturing and Sustainable Logistics*, number 46 in Lecture Notes in Business Information Processing, pages 345–357. Springer, Berlin Heidelberg, 2010.

J. Pahl, S. Voß, and A. Mies. Supply chain integration: Improvements of global lead times with scem. In *Challenges in Industrial Production*, pages 79–88. University of Karlsruhe, 2005a.

J. Pahl, S. Voß, and D. L. Woodruff. Load dependent lead times - from empirical evidence to mathematical modeling. In H. Kotzab, S. Seuring, M. Müller, and G. Reiner, editors, *Research Methodologies in Supply Chain Management*, pages 539–554. Physica, Heidelberg, 2005b.

J. Pahl, S. Voß, and D. L. Woodruff. Production planning with load dependent lead times. *4OR: A Quarterly Journal of Operations Research*, 3(4):257–302, 2005c.

J. Pahl, S. Voß, and D. L. Woodruff. Production planning with load dependent lead times: An update of research. *Annals of Operations Research*, 153(1):297–345, 2007a.

J. Pahl, S. Voß, and D. L. Woodruff. Production planning with deterioration constraints: A survey. In J.A. Ceroni, editor, *The Development of Collaborative Production and Service Systems in Emergent Economies, Proceedings of the 19th International Conference on Production Research, IFPR*, pages 1–6, Valparaiso, Chile, 2007b.

J. Pahl, S. Voß, and D. L. Woodruff. Discrete lot-sizing and scheduling with sequence-dependent setup times and costs including deterioration and perishability constraints. In *Hawaiian International Conference on Systems Sciences (HICSS)*, pages 1–10, 2011.

Chapter 2

Cumulative Doctoral Thesis

2.1 Three Thematically Related Research Articles and Reports

The focus of this thesis is on tactical production and supply chain planning with LDLT as well as sustainability aspects. Related mathematical models are studied and research gaps regarding formulations and characterizations highlighted. These gaps are closed by developing discrete dynamic deterministic approaches including the aforementioned aspects. Related surveys are given as follows:

- Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling (with S. Voß and D.L. Woodruff), In: Kotzab H., S. Seuring S., Müller M., and Reiner G. (Eds.), Research Methodologies in Supply Chain Management, Physica, Heidelberg, 540–554, 2005.
- Production Planning with Load Dependent Lead Times (with S. Voß and D.L. Woodruff), *4OR: A Quarterly Journal for Operations Research*, 3(4): 257-302, 2005.
- Production Planning with Load Dependent Lead Times: An Update of Research (with S. Voß and D.L. Woodruff), *Annals of Operations Research* 153(1): 297-345, 2007.
- Production Planning and Deterioration Constraints: A Survey (with S. Voß and D.L. Woodruff), In: Ceroni J.A. (Ed.) The Development of Collaborative Production and Service Systems in Emergent Economies, *Proceedings of the 19th International Conference on Production Research* (IFPR), Valparaiso, Chile, 1–6, 2007.
- Depreciation Effects in Production Planning and Supply Chain Management: A Survey, invited review of *European Journal of Operations Research*, under review, 2011.

Related mathematical formulations are developed and analyzed in the following publications:

- Discrete Lot-Sizing and Scheduling Including Deterioration and Perishability Constraints (with S. Voß), In: Danglmainer W., Blecken A., Delius R., and Klöpfer S. (ed.), In: Advanced Manufacturing and Sustainable Logistics, Proceedings of the 8th International Heinz Nixdorf Symposium, IHNS 2010, *Lecture Notes in Business Information Processing*, Springer, Berlin, Heidelberg, 345–357, 2010.
- Discrete Lot-Sizing and Scheduling with Sequence-Dependent Setup Times and Costs including Deterioration and Perishability Constraints (with S. Voß and D.L. Woodruff), *IEEE*, HICSS-44, 1–10, 2011.
- Production Planning with Load Dependent Lead Times and Sustainability Aspects, Technical report, Institute of Information Systems, University of Hamburg, 2012.

where the latter mentioned technical report unifies LDLT and aforementioned sustainability aspects providing mathematical models further including rework and remanufacturing. This implies a study on closed-loop supply chain management and reverse logistics that pertain to the field of green supply chain management. Moreover, an in-depth discussion on lead time/capacity management, and related planning circularities is provided. Developed and extended models are evaluated. While the first two publications concentrate on lifetime constraints, the third contribution comprises all aspects and responds to the initial research objective that includes the investigation of models accounting for LDLT while items have limited lifetimes and need to rework them if they passed their useful lifetime.

2.2 Co-Authors and Substantial Contribution of Candidate

The doctorate regulations of the University of Hamburg require a determination of the score regarding the doctorate performance according to the following equation:

$$\left(\frac{2}{n+1}\right) \tag{2.1}$$

where n denotes the number of authors per publication. The score of this doctorate thesis is 5.67. Details are given in Table 2.1.

Table 2.1: Determination of score regarding the doctoral thesis

Nr.	Title	Nr. of authors	Score
1.	Production Planning with Load Dependent Lead Times	3	0.5
2.	Production Planning with Load Dependent Lead Times: An Update of Research	3	0.5
3.	Depreciation Effects in Production Planning and Supply Chain Management: A Survey	1	1
4.	Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling	3	0.5
5.	Supply Chain Integration: Improvements of Global Lead Times with SCEM	3	0.5
6.	Production Planning and Deterioration Constraints: A Survey	3	0.5
7.	Discrete Lot-Sizing and Scheduling with Sequence-Dependent Setup Times and Costs including Deterioration and Perishability Constraints	3	0.5
8.	Discrete Lot-Sizing and Scheduling Including Deterioration and Perishability Constraints	2	0.67
9.	Production Planning with Load Dependent Lead Times and Sustainability Aspects	1	1
		Total:	**5.67**

The contribution of the doctoral candidate regarding the presented research is substantial which is also reflected by the order of authors on related publications. The form of the contribution given in Pahl (2012) is selected due to constraints regarding time and space. In general, journal publications should not exceed 15 – 20 pages unless they are surveys

that are allowed to extend to 31 pages. Additionally, the investigative character of the study should be preserved, so that a book form is chosen. The final publication form is decided at a later stage. No contribution included in this thesis constitutes an element of a current or concluded dissertation project.

2.3 Publication of Research Articles and Reports

The publication and provision of research results is a fundamental requirement to take part in the advancement of research. The interdisciplinary character of information systems (management) and operations research/management science is reflected in the selection of publication character where business management may prefers publications in professional journals whereas informatics favor publications in conference proceedings.

The evaluation of related publications may also be done by rankings that vary depending on different criteria. Table 2.2 provides the ranking of A.-W. Harzing, JQL 2000-2010[1] for Annals of Operations research (Annals), European Journal of Operations Research (EJOR), and 4OR: A Quarterly Journal for Operations Research (4OR).

Table 2.2: Ranking of selected publications

	Rankings				
Journals	VHB2003	Aeres08	Wie08	VHB08	VHB JQ 2.1
Annals	A	A	A	B	B
EJOR	A	A	A	A	A
4OR	-	-	-	-	C

[1] http://www.harzing.com/download/jql.zip.

Chapter 3

Curriculum Vitae

Personal Data

Name	Julia Pahl
Title	Diploma in Business Administration
Residence	Blankeneser Hauptstrasse 82, 22587 Hamburg, Germany
Date of Birth	27. June 1975 in Hamburg, Germany
Homepage	http://iwi.econ.uni-hamburg.de/IWIWeb/Default.aspx?tabid=174

Professional Career

Since Jul 2008	Researcher and Teaching Assistant, University of Hamburg, Germany
Jan 2006 – Jul 2008	Business Analyst, Quality Authorized Agent for ISO 9001, Lead Auditor EFQM, Deutsches Zentrum für Luft- und Raumfahrt (DLR e.V.), Cologne, Germany
May 2005 – Jan 2006	Researcher, University of Hamburg, Germany
May 2004 – May 2005	Young Graduate Trainee, Project Controlling, VEGA Small Launcher Department, Frascati, Italy
Mar 2004 – May 2004	Assistance and Support in Sales and Marketing, IBM Deutschland GmbH, Hamburg, Germany

Education and Studies

Oct 1997 – May 2003	Diploma in Business Administration on 22. May 2003, University of Hamburg with the grade *befriedigend (B)*
Oct 1999 – Oct 2000	Study abroad at the Università degli Studi Tor Vergata, Rome, Italy with the grade *excellent (A)*
Jun 1983 – Jun 1995	Secondary School at Ernst-Schlee Gymnasium, with the grade *befriedigend (B)*

Stays Abroad

May 2004	Young Graduate Traineeship, Frascati, Italy
Oct 2008	Research Visit in Davis, California, USA

Honors and Awards

May 2004	Young Graduate Traineeship at VEGA Small Launcher
Oct 2010	Research Grant for HICSS-44
Jan 2012	DAAD-PPP Vigoni Research Project, Cooperation with Universitá degli Studi di Salerno, Dipartimento di Informatica, Fiscano, Italy, Total budget 13.000,– Euro

Languages

Fluent	German (Native Speaker), English, French, Italian
Advanced Knowledge	Spanish

Abilities

Aviation:	Private Pilot License PPL-A, JAR-FCL

Description of Academic Employment and Experiences (Selection)

1. **Research Interests**

Research:	Doctoral Thesis (PhD)
Title:	Title “Production Planning with Load Dependent Lead Times and Sustainability Aspects.”
Research Fields:	Green Production and Supply Chain Planning, Lead Time Management, Sustainability, Network Planning, Airline and Airport Management.

2. **Assistance to Lectures, Exercises**

Course Title:	Production and Logistics
Level:	Bachelor, Diploma, Master
Details:	Introduction to Production (Planning) and Supply Chain Management with interest in organizational aspects and quantitative models for lot-sizing and mrp/MRP II/ERP. The course further introduces into complexity theory, decision science, and heuristic methods. Additional discussion is on scheduling, assembly line balancing, flow shops, postponement strategies, lead time management, Computer Integrated Manufacturing (CIM), Computer Aided Design (CAD), disruption management, stochastic optimization models and methods.

Course Title:	Information Management
Level:	Bachelor, Master
Details:	Introduction to Information Management especially focussing on social aspects of organization, e.g., Principal-Agent Theory and discussing various Decision Support Systems including Data Management, Information Requirement Analysis, Knowledge Management, Logistics of Information, Data Warehouse, Online Analytical Processing, Data Mining.

Course Title:	Statistics
Level:	Bachelor, Diploma, Master
Details:	Introduction to Conclusive Statistics, Probability Mass Functions, Theoretical Distributions, Distribution Model for Discrete Random Variable, Parameter Estimation, Maximum-Likelihood, Hypothesis Testing (Tests Regarding Single Control Samples, Comparison of Parameters Regarding Independent Samples, Welch Test), Analysis of Variances, Chi-Quadrat-Adjustment, Wilcoxon Rang Sum Test.

3. **Conference Organizations**

Conference Title:	LM09, INOC 2011
Level:	ca. 150 participants
Details:	Planning, preparation, and execution of several conferences including the preparation of LNCS proceedings, communication with publishing houses, financial planning and control, cost control, program planning and execution.

Description of Professional Employment and Experiences

1. **Deutsches Zentrum für Luft- und Raumfahrt e.V., Cologne, Germany**

   Job Title: Business Analyst, Quality Manager
   Dep.: Finance and Controlling
   Years: 2006–2008
   Details: **Corporate Controlling** of the administrative and technical infrastructure, controlling of Project Execution Organizations, coordination of the overhead cost planning, reengineering of standard planning and (ad hoc) reporting,
   Quality management according to ISO 9001 standards and EFQM, Project coordination and support of the implementation of Quality Management in the overall DLR administration,
   Risk management incl. system administration and coordinator corresponding to KonTraG Standards, preparation of risk reports and continuous monitoring of high-risk classes, representation of the RMS vis-a-vis financial auditors or international delegations such as EASA.

2. **ESA, ESRIN, Frascati, Italy**

   Job Title: Young Graduate Trainee
   Dep.: VEGA Small Launcher
   Years: 2004–2005
   Details: **Process Optimization Analysis**: Mapping and reengineering of the budget and financial flows and processes of the VEGA Small Launcher Project, preparation of standard reports of the Contract Status Monitoring, Quarterly Report to Council, Backdating, Geopgraphical Return, monitoring of the status of milestones and re-scheduling, Cost Controlling (Cost-at-Completion, Geographical Return, Mission Expenditure Reporting), development of online reports, overall documentation.

3. **IBM Deutschland GmbH, Hamburg, Germany**

   Job Title: Project Assistant
   Dep.: Marketing and Sales
   Years: 2004–2004
   Details: **Optimization of Supply Chain Processes** between IBM and subcontractors, due date monitoring, critical status reporting of orders, initializing escalation process for overdue orders, standard monitoring reporting.

4. **Hapag-Lloyd AG**

   Job Title: Internship
   Dep.: Corporate Controlling
   Years: 2002–2003
   Details: **Development and Consulting**: Development of a planning reporting system using MIS Alea (OLAP technology), analysis and optimization of personnel controlling reporting, development of the user interface and standard reports, development of user documentations, end user training regarding the developed tools.

Additional Academic Activities

1. **Editorial Board**

Journal:	International Journal of Operations Research and Information Systems (IJORIS)

2. **Programme Commitee, Reviewer (selection)**

Conference Organization:	LM09, INOC2011
Programme Commitee:	MIC (2009), LM (2009), ICCL (2011)
Reviewer:	IEEE Transactions, IJPR, IJPE, IJORIS, ITOR, Information Sciences.

3. **Memberships and Affiliations**

INFORMS	Institute for Operations Research and the Management Science,
GOR	Society of Operations Research,
VORMS	Virtual Operations Research and Management Science,
INFORMS	Aviation Application Section,
INFORMS	Transportation and Logistics Section,

4. **International Research Cooperations**

1. Prof. Dr. David L. Woodruff, Graduate School of Management, Davis, CA, USA
2. Assistant Prof. Dr. Monica Gentili, University of Salerno, Fiscano, Italy
3. Senior Lecturer Dr. Torsten Reiners, Curtin University, Perth, Australia

5. **Expertise in Modern Technologies (Selection)**

SAP R/3	Modules FI, CO, MM, SD,
MS Office	including Access, Project, One Note, Publisher,
OLAP	MIS Alea,
Optimization Software	ILOG OPL, Xpress-IVE, AMPL/CPLEX,
Programming	Java, VB.Net, PHP, MySQL

6. **Assessment and Professional Training**

Quality Management	EFQM Assessor, Lead Auditor LRQA (complete formation), Internal Auditor, TÜV Rheinland Groupe and DGQ
Risk Management	Assigned Risk Management Coordinator
Project Management	Intensiv Traning
Controller Academy	Level I and II,

Statistic of Publications

Peer-reviewed Journal Articles:	3†
Peer-reviewed Proceedings Articles:	5
Books:	3⋆
Lecture Notes:	1
Technical Report:	1
Working Papers:	2
Sum:	**15**

† of this 1 invited review, currently under review

⋆ additional 1 manuscript to be finalized

Publications

Peer-reviewed Articles

1. Production Planning with Load Dependent Lead Times (with S. Voß and D.L. Woodruff), *4OR: A Quarterly Journal for Operations Research*, 3(4): 257-302, 2005.

2. Production Planning with Load Dependent Lead Times: An Update of Research (with S. Voß and D.L. Woodruff), *Annals of Operations Research* 153(1), 297-345, 2007.

Peer-Reviewed Submitted Article in Review

3. Depreciation Effects in Production Planning and Supply Chain Management: A Survey, invited review of *European Journal of Operations Research*, under review, 2011.

Peer-Reviewed Conference Proceedings

4. Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling (with S. Voß and D.L. Woodruff), In: Kotzab H., S. Seuring S., Müller M., and Reiner G. (Eds.), Research Methodologies in Supply Chain Management, Physica, Heidelberg, 540–554, 2005.

5. Supply Chain Integration: Improvements of Global Lead Times with SCEM (with A. Mies and S. Voß) in: Geldermann J., Treitz M., Schollenberger H., Rentz O. (Eds.) Challenges for Industrial Production, Workshop of the PepOn Project: Integrated Process Design for Inter-Enterprise Plant Layout Planning of Dynamic Mass Flow Networks, Karlsruhe, Nov. 7-8., 79–89, 2005.

6. Production Planning and Deterioration Constraints: A Survey (with S. Voß and D.L. Woodruff), In: Ceroni J.A. (Ed.) The Development of Collaborative Production and Service Systems in Emergent Economies, *Proceedings of the 19th International Conference on Production Research* (IFPR), Valparaiso, Chile, 1–6, 2007.

7. Discrete Lot-Sizing and Scheduling with Sequence-Dependent Setup Times and Costs including Deterioration and Perishability Constraints (with S. Voß and D.L. Woodruff), *IEEE*, HICSS-44, 1–10, 2011.

8. Integrated Aircraft Scheduling Problem: An Auto-Adapting Algorithm to Find Robust Aircraft Assignments for Large Flight Plans (with T. Reiners, M. Maroszek, C. Rettig), *IEEE*, HICSS-45, 2012.

Books

9. Production Planning with Deterioration and Perishability, Draft Manuscript, 2011.

10. Logistik Management - Systeme, Methoden, Integration (with S. Voß and S. Schwarze (Ed.)), Logistics Management - partly in German and partly in English, Physica, Heidelberg, 2009.

11. Network Optimization - International Network Optimization Conference, INOC 2011 (with T. Reiners and S. Voß (Eds.)), Hamburg, Germany, 2011, *Lecture Notes in Computer Sciences (LNCS)*, Springer, Heidelberg, 2011.

Lecture Notes

12. Discrete Lot-Sizing and Scheduling Including Deterioration and Perishability Constraints (with S. Voß), In: Danglmainer W., Blecken A., Delius R., and Klöpfer S. (ed.), In: Advanced Manufacturing and Sustainable Logistics, Proceedings of the 8th International Heinz Nixdorf Symposium, IHNS 2010, *Lecture Notes in Business Information Processing*, Springer, Berlin, Heidelberg, 345–357, 2010.

Working Papers and Technical Reports

13. WIP Management in Practice: Survey and Analysis (with F. Schulte), Institute of Information Systems, University of Hamburg, 2011.

14. Efficient Formulations for Lot-Sizing Models with Perishability (with C. Seipl), Institute of Information Systems, University of Hamburg, 2011.

15. Production Planning with Load Dependent Lead Times and Sustainability Aspects, Technical Report, Institute of Information Systems, University of Hamburg, 2012.

Part II

Literature

Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling

Julia Pahl, Stefan Voß, David L. Woodruff

in: Kotzab, H., Seuring, S., Müller, M., Reiner, G. (eds.) (2005): Research Methodologies in Supply Chain Management, Physica, Heidelberg, p. 540-554.

1 Introduction 540

2 Load Dependent Lead Times – Empirical Evidence 541

3 Models Including Load Dependent Lead Times 544

4 Conclusions 552

5 References 553

Summary:
As organizations move from creating plans for individual production lines to entire supply chains it is increasingly important to recognize that decisions concerning utilization of production resources impact the lead times that will be experienced. In this paper we give some insights into why this is the case by looking at the queuing that results in delays. In this respect, special mention should be made that it is difficult to experience related empirical data, especially for tactical planning issues. We use these insights to survey and suggest optimization models that take into account load dependent lead times and related "complications."

Keywords:
Supply Chain Management, Load Dependent Lead Times, Lead Times, Tactical Planning, Aggregate Planning

1 Introduction

Let us define the lead time as the time between the release of an order to the shop floor or to a supplier and the receipt of the items. Lead time considerations are essential with respect to the global competitiveness of firms, because long lead times impose high costs due to rising WIP (work in process) inventory levels as well as larger safety stocks caused by increased uncertainty about production prerequisites and constraints. Despite this, considerations about load dependent lead times are rare in the literature. The same is valid for models linking order releases, planning and capacity decisions to lead times, and take into account factors influencing lead times such as the system workload, batching and sequencing decisions or WIP levels.

Present practice for manufacturing supply chains is dominated by the use of material requirements planning (mrp) with its inherent problems. Many companies do not use adequate planning tools at all. Accordingly, problems arise when fixed, constant or "worst case" lead times are assumed at an aggregate planning level, e.g., to have enough "buffer time" to securely meet demands. In order to meet due dates there is also a tendency to release jobs into the system much earlier than necessary, leading to very high WIP levels and, therefore, longer queuing (waiting times) causing even longer lead times. This overreactional behavior becomes a self-fulfilling prophecy and is addressed in the literature as the *lead time syndrome* which results from the fact that the relationship between WIP, output, workload and average flow times is ignored (Zäpfel & Missbauer, 1993; Tatsiopoulos & Kingsman, 1983). Moreover, most mrp and enterprise resource planning (ERP) models implement sequential planning algorithms which neither consider uncertainties nor resource and production flow constraints of raw material, WIP and finished goods inventory (FGI), leading to suboptimal or infeasible production plans (Caramanis & Ahn, 1999).

Another fundamental problem of manufacturing and production planning models is the omission of modeling nonlinear dependencies, e.g., between lead times and the workload of a production system or a production resource. This happens even though there is empirical evidence that lead times increase nonlinearly long before resource utilization reaches 100% (Asmundsson et al., 2003; Karmarkar, 1987); see Figure 1. This may lead to significant differences in planned and realized lead times. There is a lack of models allowing the analysis of behavior of lead times and WIP levels considering the facility workload under variable demand patterns like seasonal demand. In addition, it seems likely that queuing tends to be correlated so that a machine failure at one point of the system will cause queuing at other stations which leads to the presumption that lead time distributions tend to be fat-tailed and skewed. However, to the best of our knowledge there is no comprehensive (empirical) work on this topic currently available. Furthermore, it seems that there is no model which analyzes load dependent lead times in the context of stochastic demand evidently prevailing in practice.

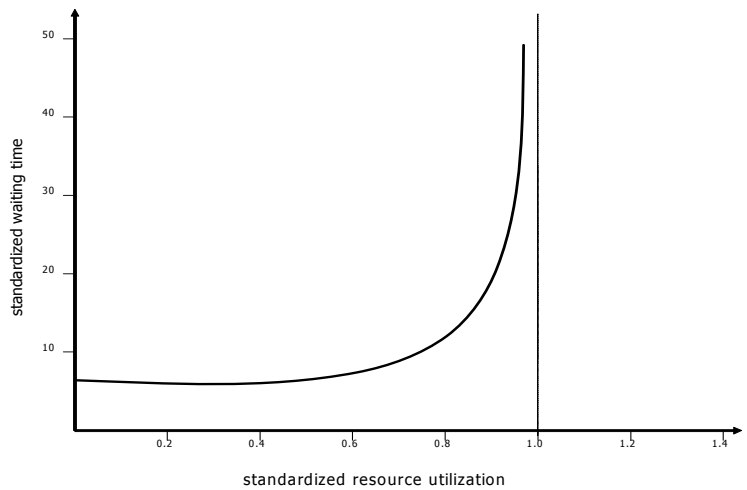

Figure 1: Nonlinear Relationship between Waiting Time and Resource Utilization (Voß & Woodruff, 2003: 162)

It is necessary to examine the problem of lead time dynamics at individual links to better understand the effects and the modelling requirements and complexities at the aggregate planning level for the entire supply network. The aim of this paper is to demonstrate, based on empirical evidence obtained by a survey and interviews recently executed and briefly sketched in the next section, the need for aggregate planning models with the following features: being able to take into account the nonlinear relation between lead times and workload, while remaining tractable to be adapted to complex production systems and supply chains. The remainder of this paper is organized as follows. In Section 2 we point out the empirical evidence of load dependent lead times by means of the results obtained from interviews and a survey recently executed. Then we survey methods and models dealing with load dependent lead times and examine indirect approaches, aspects of queuing theory, and introduce so-called clearing functions in Section 3. The paper concludes with some remarks and suggestions for future research directions.

2 Load Dependent Lead Times – Empirical Evidence

Production planning is a complex issue especially in the context of variable demand patterns or stochastic demand. In numerous production environments demand quantities are not known at the beginning of the production planning

process. As a result it is difficult to create forecasts of (highly) variable demand patterns. Production uncertainties and unforeseen events such as machine breakdowns, unavailability of production resources, illness of workers, etc. raise the instability of the production process with queues building up in front of machines resulting in increased WIP and FGI levels and consequently in raised lead times. Therefore, production processes in such environments tend to become somewhat hectic with overtime in peak situations leading to unbalanced utilization of production resources. This is especially true for the food (or semiconductor) industry which also has to account for various deteriorating rates of their production material, which is another complication issue of (load dependent) lead times. As a result, the forecast quality is very important in tactical (and operational) production planning since it prevents precipitated releases of jobs (orders) in the production process and, therefore, should be linked with aggregate production planning and order release control. Nevertheless, there is a lack of practical and useful tools for tactical production planning which permits companies to account for variable demand and unforeseen events causing load dependent lead times. This is one of the principle outcomes of our survey and interviews with companies from different industrial sectors such as transportation, logistics and inventory, aerospace, industry automation and mineral oil, and the chemical industry.

The study includes enterprises of various sizes producing different types of products with very different product life cycles including base polyols, load cels, indicators/ transmitters, software, IT- and logistic services and satellite launchers. These companies face diverse demand patterns and environmental challenges they have to take into consideration in the overall production planning process, especially with regard to the planning of resources and their utilization levels. Many companies face variable (seasonal) and not easily predictable demand for their foremost products. This seems to be the main uncertainty in the production process, because further potential precarious factors such as, e.g., the cooperations with supply chain partners and delieveries from supply chain partners are not validated as highly impacting the production process. This is due to the fact that, e.g., the launcher business has very long production cycle times (the production of one launcher like Ariane 5 or Vega takes on average 2.5 years). Here, problems concerning the cooperation of supply chain partners are not very critical in terms of time compared to other branches like the automotive industry where JIT production is mainly implemented and late deliveries of components cause the whole production process to stop. Nevertheless, late deliveries of important components for, e.g., a launcher also cause the whole assembly process to stop which leads to great financial losses not only because of idle times of production resources and very expensive WIP waiting in the queue, but also because of the costs for the client associated with the delay (lost profits of satellite services). Other companies do not experience supply chain cooperation problems due to long endurance with few supply chain partners, leading to a stabilized and well defined work flow.

Usually medium or large organisations are characterized by a large number of supply chain partners and more or less complex production processes. The range of surveyed companies and their production systems cover synchronized facilities, job shops and make-to-order-systems with different core objectives in tactical production planning as, e.g., maximizing resource utilization in order to avoid idle times (mostly in synchronized production facilities), minimizing lead times or cycle times (this holds true for make-to-order situations) and minimizing WIP and FGI levels. Only a few companies use specific tools for tactical production planning such as SAP R/3, SAP APO (APO SNP for tactical production planning and APO PP/DS for operational production planning). The survey confirms the prevailing use of mrp-based systems together with estimated lead times (or planned lead times) leading to the problems outlined above. Most of the companies experience rising lead times due to machine breakdowns as well as rising WIP levels and consequent queuing in front of machines. Nevertheless, because of the unavailability of data sets (surveys) which are necessary to execute a detailed empirical analysis, it is not clear whether this occurs before reaching 100% utilization, but despite the lack of information, queuing theory emphasizes the impact of resource utilization on load dependent lead times.

The main goals in tactical production planning of the surveyed companies consist in minimizing lead (or cycle) times, as well as WIP and FGI inventory levels, and maximizing resource utilization in order to avoid idle times. For this purpose some of them use, e.g., some "worst case lead times" in order to have enough buffer time at certain (critical) points in the production system and to secure that demands can be met. Others implement estimates of lead times derived from historical data of the production system, which gets problematic when production processes change. Consequently, the underlying data for estimated or planned lead times is neither reliable nor useful in order to achieve the mentioned objectives. However, companies are aware of the fact that decisions on the workload in the production system (and of single resources), on scheduling and sequencing, and on lot sizing and setup times are key factors influencing (load dependent) lead times. To summarize, they lack models (included in comprehensive, usable and useful software tools) providing them with, e.g., "if-then"-analysis in order to better understand the impact of decisions of resource utilization levels, and not only for one single machine or production resource, but even for the whole supply chain network, and furthermore, in order to permit them to use better estimates of lead times. Until now useful models did not exist which provide production planners with necessary information about the lead times which will be experienced in case of diverse resource utilization levels.

As mentioned above, load dependent lead times are the result of production planning processes and should not be an input factor for production planning and scheduling. Moreover, the surveyed companies state the interest in models which take into account load dependent lead times and their impact on the performance of production. Thus it is necessary to analyse the nonlinear relationship of

resource utilization and lead times, as well as influencing factors in more detail in order to integrate them in aggregate production planning. Finally, the integration of supply chain partners in an overall supply chain network tends to precede in the right direction by connecting the participants through information system tools, e.g., the same production planning software or linking them together by add-ons in order to guarantee real time information of their production process and those of the supply chain partners. Nevertheless, this is still an ongoing process.

3 Models Including Load Dependent Lead Times

Load dependent lead times are primarily considered in the framework of capacity planning models and order release control mechanisms. Traditional models aim at "filling time buckets" which represent the available capacity of a production system in discrete time periods, while linear programming models are typically employed using hard capacity constraints which ignore the phenomenon that in asynchronous systems, prevailing in practice, queues build up long before 100% resource utilization is reached. Furthermore, they do not impose costs until the capacity constraint is violated, i.e., the constraint only tightens in case of 100% utilization (Karmarkar, 1989). Moreover, these models neither account for WIP and other lead time related cost factors that increase with queues and delays and accordingly with longer lead times (Karmarkar, 1993; Zipkin, 1986), nor do they include WIP costs and lead time consequences of capacity loading which can have significant effects on the performance of the production system.

3.1 Indirect Approaches

There are several ways to address problems associated with load dependent lead times. Some authors do not directly consider the difficulty of modeling nonlinear dependencies of lead times and workload, but try to solve the problem indirectly by influencing parameters that have an effect on lead times such as decisions on job release policies, influences of the demand side, changes in production plans or by smoothing demand variability, e.g., by implementing a make to stock policy, or shifts (away) from bottlenecks in order to increase capacity. Other approaches concentrate on lot sizing as an influencing factor or on production system characteristics as well as employing queuing theory as an analytical method.

3.2 Aspects of Queuing Theory

Analysis of production system performance and important key factors like throughput, WIP levels and load dependent lead times are frequently executed in

the context of queuing theory due to the fact that a large percentage of lead times are waiting times. It has been shown that 90% of the total flow time is due to transit times, where 85% consists of waiting (queuing) time, 3% of quality control, and 2% of transportation time; only 10% is due to value added processing operations (Tatsiopoulos & Kingsman, 1983). Queuing network models highlight the relationship between the capacity, loading and production mix as well as the resulting WIP levels and effect on lead times (Karmarkar, 1987) and provide important information on the causes of congestion phenomena. Furthermore, they show that delays predominantly depend on the service variability, i.e., the processing time of a resource, the variability of the arrival rate of work at a resource and the current workload as well as scale effects with major delays near the maximum capacity usage (Srinivasan et al., 1988).

Congestion phenomena are inherent problems of production systems complicating the planning process. They emerge at different and frequently changing times and places which are hardly predictable. Therefore, it is crucial to better understand the reasons for congestion phenomena like the limited capacity of a machine (resource) to respond to demand variation over time (Lautenschläger, 1999) and to account for them in aggregate planning models. The literature on queuing and congestion phenomena is multitudinous; see, e.g. (Chen et al., 1988; Karmarkar, 1987, 1989; Spearman, 1991; Suri & Sanders, 1993; Zipkin, 1986). Spearman (1991) develops a cyclic closed queuing network model with three parameters, viz. the bottleneck capacity, the "raw processing time" (i.e., increasing failure rate processing time, IFR) and a congestion coefficient which specifies a unique throughput/WIP curve in order to analyze the dependency between mean cycle time (synonymously used for "flow time" in many references) and WIP for the whole production system, i.e., single resources and their processing times are not considered. The model indicates a relationship between mean cycle time and WIP level and can be used to predict the average cycle time in exponential as well as in IFR closed queuing networks. Chen et al. (1988) provide a network queuing model for semiconductor wafer facilities which points out that congestion and delays are due to variability in the operating environment. So this variability has to be smoothed in order to obtain shorter production cycle times.

It is useful to start with a queuing model in order to obtain some approximations for the key parameters or objective functions to be implemented in an aggregate planning model (Buzacott & Shantikumar, 1993).

3.3 Indirect Integration of Load Dependent Lead Times

There are only a few approaches which try to integrate load dependent lead times directly into mathematical programming models. For instance, Zijm & Buitenhek (1996) developed a manufacturing planning and control framework for a machine shop which includes workload oriented lead time estimates. For this purpose they

suggest a method that determines the earliest possible completion time of arriving jobs with the restriction that the delivery performance of any other job in the system will not be adversly affected, i.e., that every job can be completed and delivered on time. Their aim is to determine reliable planned lead times based on workload which guarantee that due dates are met and can be implemented at a capacity planning level, serving as input for a final detailed capacity scheduling procedure that also takes into account additional resources, job batching decisions as well as machine setup characteristics. Their framework is partly based on the work of (Karmarkar, 1987; Karmarkar et al., 1985). Missbauer (1998) focuses on the hierarchical production planning concept in which all partial problems can be included (e.g., aggregate production planning, capacity planning, lot sizing, scheduling etc.), but which avoids the problems of a comprehensive model, e.g., problems of data procurement, limited computational storage space, CPU times that are too long for calculation etc.

Graves (1986) studied the dependencies between production capability, variability (uncertainty) of the production requirements, and level of WIP inventory in a tactical planning model and analyzes to which extent the job flow time (or WIP inventory) depends on the utilization of each resource of a job shop or production stage. He further concentrates on analyzing the interrelationship of flow time and production mix. For this purpose he employs a network model where multiple routings of jobs are possible so that the lack of a dominant work flow renders production control, which aims at reducing the variance of planned lead times. In addition, he uses a queuing model that includes flexible production rates of resources which can be set by a tactical planning model in order to smooth the work flow and to avoid the overload of resources. Moreover, he implements a control rule at each resource that determines the amount of work performed during a time period which is a fixed portion α_j of the queue of work at j remaining at the start of the period at a specific resource j: $P_{jt} = \alpha_j Q_{jt} \quad \forall j,t$ with P_{jt} denoting the production of resource j in time period t, α_j a smoothing parameter with $0 < \alpha_j \leq 1$ and Q_{jt} the queue of work at j at the beginning of time period t. This parameter α_j is called "proportional factor" by Missbauer (1998) and "clearing factor" by Graves (1986), because it indicates the quantity of jobs (orders) which can be cleared or finished (and passed to another station) in one time period. Here, the clearing factor implies infinite capacity since the resource is able to complete the fixed portion α_j even when the workload (WIP) is infinitely high. The major drawback of this model is the employment of a linear function and consequently the omission of the nonlinear relationship of WIP and lead times. Nevertheless, Graves (1986) seems to be the first reference accounting for the dependency between lead times and workload and giving a practical aid on how to set planned lead times in, e.g., mrp models considering the workload of the production system.

3.4 Clearing Functions

Taking up the idea of Graves and integrating it in a model with so-called "clearing factors" *α(WIP)* which are nonlinear functions of the WIP yields a clearing function of the following form (Karmarkar, 1989; Srinivasan et al., 1988):

$$Capacity = \alpha(WIP) * WIP = f(WIP)$$

where *f* represents the clearing function which models capacity as a function of the workload. The clearing factor specifies the fraction of the actual WIP which can be completed, i.e., "cleared," by a resource in a given period of time (Asmundsson et al., 2003). Missbauer (1998) calls this function "utilization function."

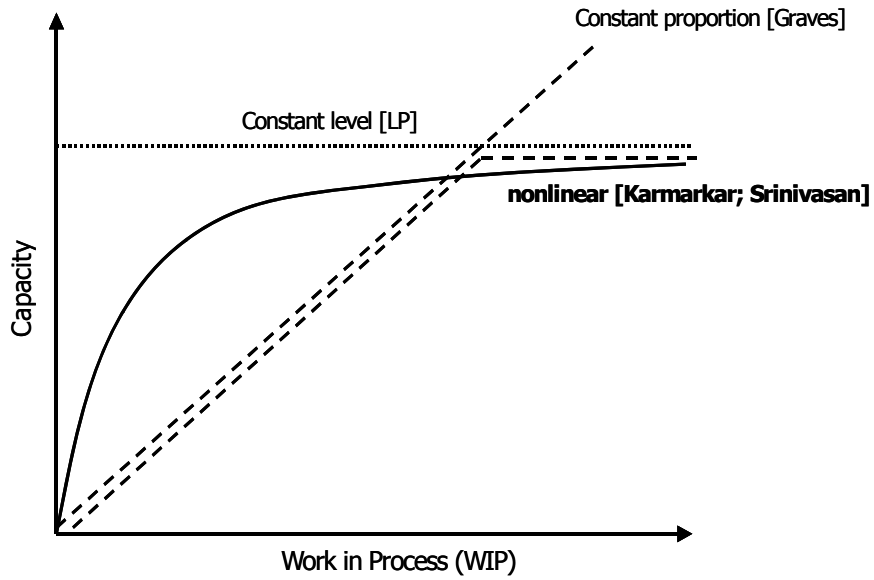

Figure 2: Different Clearing Functions (Karmarkar, 1993)

Figure 2 depicts some possible clearing functions where the constant level clearing function corresponds to an upper bound for capacity as mainly employed by linear programming models. This implies instantaneous production without lead time constraints since production takes place independently of WIP in the production system. The constant proportion clearing function represents the control rule given by Graves (1986) which implies infinite capacity and hence allows for unlimited output. In contrast to the nonlinear clearing function of Karmarkar and Srinivasan et al., the combined clearing function in some region underestimates and in others overestimates capacity. Moreover, the nonlinear clearing function relates WIP levels to output and lead times to WIP levels which are influenced by the work-

load of the production system (Karmarkar, 1993) and, therefore, is able to capture the behavior of load dependent lead times. By applying Little's Law, the clearing function can be reinterpreted in terms of lead time or WIP turn. Additionally, the slope of the clearing function represents the inventory turn with lead times given by the inverse of the slope (Karmarkar, 1989). Asmundsson et al. (2003) combine aspects of queuing theory with the clearing function concept by employing a clearing function of the above given form and by defining the performance of a resource (work center) as dependent on the workload using a G/G/1 queuing model.

In order to develop the clearing function, two approaches can be found in the literature to date. The first is the analytical derivation from queuing network models and the second an empirical approximation using a functional form which can be fitted to empirical data. Because of the large amount of detail in practical systems the complete identification of the clearing function will not be possible, so we have to work with approximations. Asmundsson et al. (2002) integrate the estimated clearing function in a mathematical programming model where the framework is based on the production model of (Hackman & Leachman, 1989) with an objective function that minimizes the overall costs. It is assumed that backorders do not occur and that all demand must be met on time. We concentrate on this model as an example for the direct integration of load dependent lead times in aggregate production planning models. The model is then stated as follows:

$$Min \sum_t \left[\sum_n \left[\sum_i (\phi_t^n X_{it}^n + \varpi_{it}^n W_{it}^n + \pi_{it}^n I_{it}^n + \rho_{it}^n R_{it}^n) \right] + \sum_j \theta_{jt} Y_{jt} \right]$$

subject to :

$$W_{it}^n = W_{i,t-1}^n - \frac{1}{2}\left(X_{it}^n + X_{i,t-1}^n\right) + R_{it}^n + \sum_{j \in A(n,i)} Y_{jt} \qquad \forall n,t,i$$

$$I_{it}^n = I_{i,t-1}^n + \frac{1}{2}\left(X_{it}^n + X_{i,t-1}^n\right) - D_{it}^n - \sum_{j \in B(n,i)} Y_{jt} \qquad \forall n,t,i$$

$$X_{\bullet t}^n \le f_{nt}\left(W_{\bullet t}^n\right) \qquad \forall n,t$$

$$X_{it}^n, W_{it}^n, Y_{jt}, I_{it}^n, R_{it}^n \ge 0 \qquad \forall n,t,i,j$$

where $\phi_t^n X_{it}^n$ denotes the costs of the total amount of production over the latter half of period t and the first half of period t+1, represented by X_{it}^n, with ϕ_t^n referring to the corresponding unit costs at node n in period t. The WIP costs and the FGI costs of item i at node n at the end of period t are denoted by $\omega_{it}^n W_{it}^n$ and $\pi_{it}^n I_{it}^n$, respectively, with ω_{it}^n and π_{it}^n being the corresponding unit costs and W_{it}^n and I_{it}^n representing the WIP and the inventory, respectively. Likewise, the costs of releases of raw material of item i at node n during period t are represented

by $\rho_{it}^{n} R_{it}^{n}$ with unit costs ρ_{it}^{n} and finally the transfer costs on arc *j* during period *t* are given by $\theta_{jt} Y_{jt}$ with corresponding unit costs θ_{jt}. $X_{\bullet t}^{n}$ is the production quantity and $W_{\bullet t}^{n}$ the WIP level summarized over all items *i*.

The first two constraints denote the flow conservation for WIP and FGI, which is different from classical models since inventory levels at each node in the network are connected with the throughput rate. (*A(n,i)* / *B(n,i)* represent a set of transportation arcs contributing to inflow / outflow of item *i* at node *n*.) In contrast to Ettl et al. (2000), the nonlinear dynamic is incorporated in the clearing function and thus not included in the objective function, but modeled as a constraint of the model. Furthermore, the planning circularity which is one of the most significant shortcomings of mrp systems is overcome by not modeling the lead time explicitly in the mathematical program. Consequently, there is no need to employ fixed and / or estimated lead times ignoring the nonlinear relationship between lead times and WIP. Instead, they are calculated using Little's Law:

$$L_{it}^{n} = \frac{W_{it}^{n}}{X_{it}^{n}}$$

where L_{it}^{n} denotes the expected lead time for the last job of item *i* which arrived before the end of period *t*. We are also interested in deriving the lead times for single items *i* in order to consider multiple product types with different resource consumption patterns. For that purpose we assume the standard case of FIFO processing for which the following relationship holds:

$$\frac{X_{it}^{n}}{X_{\bullet t}^{n}} = \frac{W_{it}^{n}}{W_{\bullet t}^{n}}$$

Taking this relation and multiplying the production quantities X_{it}^{n} with their so-called resource consumption factor ξ_{it}^{n} which defines the capacity consumption per unit produced for item *i* at node *n*, we derive a new variable Z_{it}^{n} of the following form:

$$Z_{it}^{n} = \frac{\xi_{it}^{n} X_{it}^{n}}{X_{\bullet t}^{n}} = \frac{\xi_{it}^{n} W_{it}^{n}}{W_{\bullet t}^{n}} \qquad \forall n, t, i$$

By implementing this variable Z_{it}^{n} we obtain the clearing function for each item *i* which is called the partitioned clearing function:

$$\xi_{it}^{n} X_{it}^{n} \le Z_{it}^{n} f\left(\frac{\xi_{it}^{n} W_{it}^{n}}{Z_{it}^{n}}\right) \qquad \forall n, t, i$$

with the following properties of Z_{it}^{n}:

$$\sum_i Z_{it}^n = 1 \qquad \forall i,n,t$$

$$Z_{it}^n \geq 0 \qquad \forall i,n,t$$

The partitioned clearing function is depicted in Figure 3.

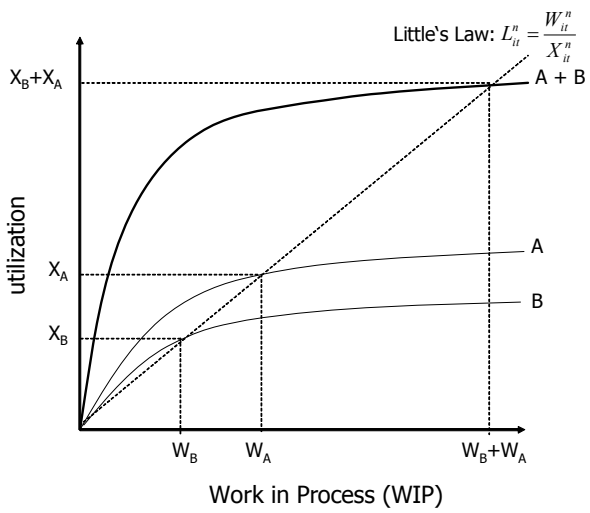

Figure 3: Clearing Function for Products A and B (Asmundsson et al., 2002)

In order to relax the assumed priority rule (FIFO) we only suppose that Z_{it}^n satisfies the properties stated above, but has an arbitrary functional form. With this formulation Asmundsson et al. (2002) succeed in integrating the nonlinear relationship between WIP and lead times in a mathematical model. The second goal is to transform this model in a tractable form which allows even the relatively large planning problems to be dealt with. For this reason we use a linear programming formulation by representing the partitioned clearing function through a set of linear constraints. To be more precise, the clearing function is approximated by the convex hull of straight lines which is possible because of its concavity:

$$f_{nt}\left(W_{\bullet t}^n\right) = \min_c \left\{\alpha_{nt}^c W_{\bullet t}^n + \beta_{nt}^c\right\} \qquad \forall n,t$$

The individual lines of the items are denoted by the index c. The β coefficients represent the intersection with the y-axis and indicate the capacity splitting (sharing) across the items while the α coefficients represent the slope of the clearing function. Applying this formulation to the partitioned clearing function leads to the following form:

$$Z_{it}^{n}\left(\frac{\xi_{it}^{n}W_{it}^{n}}{Z_{it}^{n}}\right)=Z_{it}^{n}\cdot\min_{c}\left\{\alpha_{nt}^{c}\frac{\xi_{it}^{n}W_{it}^{n}}{Z_{it}^{n}}+\beta_{nt}^{c}\right\}=\min_{c}\left\{\alpha_{nt}^{c}\xi_{it}^{n}W_{it}^{n}+\beta_{nt}^{c}Z_{it}^{n}\right\}$$

Replacing the former capacity constraint of the original nonlinear mathematical programming model with nonlinear lead time and capacity dynamics gives the complete linearized formulation:

$$Min\sum_{t}\left[\sum_{n}\left[\sum_{i}(\phi_{it}^{n}X_{it}^{n}+\varpi_{it}^{n}W_{it}^{n}+\pi_{it}^{n}I_{it}^{n}+\rho_{it}^{n}R_{it}^{n})\right]+\sum_{j}\theta_{jt}Y_{jt}\right]$$

subject to :

$$W_{it}^{n}=W_{i,t-1}^{n}-\frac{1}{2}\left(X_{it}^{n}+X_{i,t-1}^{n}\right)+R_{it}^{n}+\sum_{j\in A(n,i)}Y_{jt}\qquad\forall n,t,i$$

$$I_{it}^{n}=I_{i,t-1}^{n}+\frac{1}{2}\left(X_{it}^{n}+X_{i,t-1}^{n}\right)-D_{it}^{n}-\sum_{j\in B(n,i)}Y_{jt}\qquad\forall n,t,i$$

$$\xi_{it}^{n}X_{it}^{n}\le\alpha_{nt}^{c}\xi_{it}^{n}W_{it}^{n}+Z_{it}^{n}\beta_{nt}^{c}\qquad\forall n,t,i,c$$

$$\sum_{i}Z_{it}^{n}=1\qquad\forall n,t$$

$$X_{it}^{n},W_{it}^{n},Y_{jt},I_{it}^{n},R_{it}^{n},Z_{it}^{n}\ge 0\qquad\forall n,t,i,j$$

The approximation of the partitioned clearing function is depicted in Figure 4.

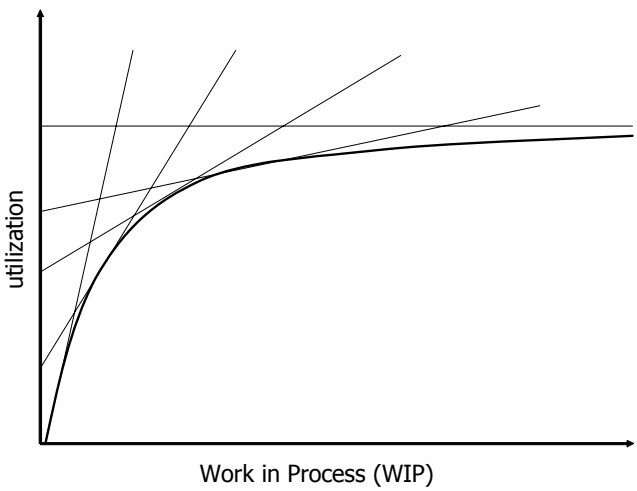

Figure 4: Linearization of the Partitioned Clearing Function

Advantages of this approach lie in the fact that the marginal cost of capacity and the marginal benefit of adding WIP are strictly positive, because of the fact that the constraints are always active as opposed to classical models where, e.g., the capacity constraint is only active at 100% utilization. However, this is only likely to be the fact at the bottleneck of the production system (Asmundsson et al., 2002). In order to examine the relevance and performance of this approach the authors consider an example of a simple single stage system with three products, which gives very good results. Furthermore, the sensitivity of the estimated clearing function to diverse shop floor scheduling algorithms, different demand patterns and techniques of production planning using a simulation model is analyzed. To summarize, the clearing function model reflects the characteristics and capabilities of the production system better than models using fixed planned lead times (like mrp) and derives realistic and robust plans with better on time delivery performance, lower WIP and system inventory (Asmundsson et al., 2003). In addition, the model captures the effects of congestion phenomena on lead times and WIP and, therefore, the fundamental trade-off between anticipatory production to account for possible demand peaks and just in time production to avoid higher costs due to preventable FGI. Finally, the releases generated by the partial clearing function model are smoother and lead to enhanced lead time performance. Moreover, interactions between clearing functions and shop floor execution systems such as the dependency of load dependent lead times on the various priority rules have to be analyzed more closely. This is a circularity, because clearing functions are dependent on the employed scheduling policy and, therefore, on the result of the scheduling algorithm. Moreover, the schedules are dependent on the release schedule and consequently on the planning algorithm.

4 Conclusions

We have seen that considerations on load dependent lead times are rare in the literature to date which is also true for aggregate planning and control models. This is particularly noteworthy, because reflections on lead times are essential with respect to the global competitiveness of firms. Furthermore, we have seen the importance to account for the nonlinear relationship between lead times and workload of production systems and further influencing factors such as product mix, scheduling policies, batching or lot sizing, variable demand patterns, deterioration etc. Analytical (queuing) models emphasize the nonlinear relationship between lead times and workload which is included only in a few mathematical planning models. Additionally, there is a lack of models which analyze load dependent lead times in the context of stochastic demand and uncertainties evidently prevailing in practice. The approach of modeling clearing functions in order to account for load dependent lead times as outlined in this paper is considered very promising and will be implemented in a stochastic

framework by using queuing models with the purpose of integrating the problem of variable demand patterns, and in order to analyze the behavior of load dependent lead times. This will be used as a starting point for more sophisticated modelling of production systems where we try to model single production units (resources, workstations, etc.) as queuing models in order to derive their specific clearing functions. Furthermore, it has been stated that load dependent lead times mainly arise due to congestion phenomena which are pervasive problems of production systems, complicating the planning process by emerging at different and frequently changing times and places which are hardly predictable due to various factors like machine breakdowns, variable demand patterns or deteriorating items. For future work we shall develop approaches for aggregate production planning which take empirical values of the probability of machine breakdowns into account as well as the other mentioned causes of congestion phenomena. This can be achieved by applying, e.g., a learning algorithm which allows for learning the behavior of production units (resources or workstations) as well as the overall system behavior, and including this information into aggregate production planning. Information on downtimes is rarely considered or integrated in mathematical models. It is not even considered in the latest and sophisticated supply chain management software like SAP APO. This will be a subject for further research.

5 References

Asmundsson, J., Rardin, R. L., Uzsoy, R. (2002): Tractable Nonlinear Capacity Models for Aggregate Production Planning, Working Paper, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., Rardin, R. L., Uzsoy, R. (2003): An Experimental Comparison of Linear Programming Models for Production Planning Utilizing Fixed Lead Time & Clearing Functions, Working Paper, School of Industrial Engineering, Purdue University, West Lafayette.

Buzacott, J. A., Shantikumar, J. G. (1993): Stochastic Models of Manufacturing Systems, Englewood Cliffs, New York.

Caramanis, M. C., Ahn, O. M. (1999): Dynamic Lead Time Modeling for JIT Production Planning, Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, May 10-15, v2, 1450-1455.

Chen, H., Harrison, M. J., Mandelbaum, A., van Ackere, A., Wein, L. M. (1988): Empirical Evaluation of a Queuing Network Model for Semiconductor Wafer Fabrication, Operations Research, 36(2): 202-215.

Ettl, M., Feigin, G. E., Lin, G. Y., Yao, D. D. (2000): A Supply Network Model with Base-Stock Control and Service Requirements, Operations Research, 48(2): 216-232.

Graves, S. (1986): A Tactical Planning Model for Job Shops, Operations Research, 34(4): 522 – 533.

Hackman, S. T., Leachman, R. C. (1989): A General Framework for Modeling Production, Management Science, 35(4): 478-495.

Karmarkar, U. S. (1987): Lot Sizes, Lead Times and In-Process Inventories, Management Science, 33(3): 409-418.

Karmarkar, U. S. (1989): Capacity Loading and Release Planning with Work-In-Process (WIP) and Lead Times, Journal of Manufacturing and Operations Management, 2(2): 105-123.

Karmarkar, U. S. (1993): Manufacturing Lead Times, Order Release and Capacity Loading, in: Graves, S., Rinnooy Kan, A., Zipkin, P. (eds.): Logistics of Production and Inventory, Handbooks in Operations Research and Management Science, Vol. 4, Amsterdam: p. 287-329.

Karmarkar, U. S., Kekre, S., Kekre, S. (1985): Lotsizing in Multi-Item Multi Machine Job Shops, IIE Transactions, 17(3): 290-297.

Lautenschläger, M. (1999): Mittelfristige Produktionsprogrammplanung mit auslastungsabhängigen Vorlaufzeiten ("Tactical Production Planning with Workload Dependent Forward Production Times"), PhD Thesis, TU Darmstadt.

Missbauer, H. (1998): Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen ("Inventory Control as a Basis for a New Concept for PPS-Systems"), Physica, Heidelberg.

Spearman, M.L. (1991): An Analytic Congestion Model for Closed Production Systems with IFR Processing Times, Management Science, 37(8): 1015-1029.

Srinivasan, A., Carey, M., Morton, T. E. (1990): Resource Pricing and Aggregate Scheduling in Manufacturing Systems, Working Paper, GSIA, 1988 (Revised December 1990).

Suri, R., Sanders, J.L. (1993): Performance Evaluation of Production Networks, in: Graves, S., Rinnooy Kan, A., Zipkin, P. (eds.): Logistics of Production and Inventory, Handbooks in Operations Research and Management Science, Vol. 4, Amsterdam: p. 199-286.

Tatsiopoulos, I.P., Kingsman, B.G. (1983): Lead Time Management, European Journal of Operational Research, 14(4): 351-358.

Voß, S., Woodruff D.L. (2003): An Introduction to Computational Optimization Models for Production Planning in a Supply Chain, Springer, Berlin.

Zäpfel, G., Missbauer, H. (1993): New Concepts for Production Planning and Control, European Journal of Operational Research, 67(3): 297-320.

Zijm, W. H. M., Buitenhek, R. (1996): Capacity Planning and Lead Time Management, International Journal of Production Economics, 46-47: 165-179.

Zipkin, P. H. (1986): Models for Design and Control of Stochastic, Multi-Item Batch Production Systems, Operations Research, 34(1): 91-104.

4OR 3: 257–302 (2005) **Invited survey**

DOI: 10.1007/s10288-005-0083-9

Production planning with load dependent lead times

Julia Pahl[1], **Stefan Voß**[1], **and David L. Woodruff**[2]

[1] Institut für Wirtschaftsinformatik, Universität Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany (e-mail: pahl@econ.uni-hamburg.de; stefan.voss@uni-hamburg.de)
[2] Graduate School of Management, University of California, Davis CA 95616, USA (e-mail: dlwoodruff@ucdavis.edu)

Received: September 2005 / Revised version: November 2005

Abstract. Lead times impact the performance of the supply chain significantly. Although there is a large literature concerning queuing models for the analysis of the relationship between capacity utilization and lead times, and there is a substantial literature concerning control and order release policies that take lead times into consideration, there have been only few papers describing models at the aggregate planning level that recognize the relationship between the planned utilization of capacity and lead times. In this paper we provide an in-depth discussion of the state-of-the art in this literature, with particular attention to those models that are appropriate at the aggregate planning level.

Key words: Production planning, load dependent lead times, mathematical programming, supply chain management

MSC classification: 46N10

1 Introduction

Let us define the *lead time* as the time between the release of an order to the shop floor or to a supplier and the receipt of the items. Lead time considerations are essential with respect to the global competitiveness of firms, because long lead times impose high costs due to rising *work in process* (WIP) inventory levels as well as larger safety stocks caused by increased uncertainty about demand. Large planning models typically treat lead times as static input data, but in most situations the output of a planning model implies capacity utilizations which, in turn, imply lead

All correspondence to: Stefan Voß

4OR A Quarterly Journal of Operations Research
© Springer-Verlag 2005

times. Lead times are among the most important properties of items in production planning and *supply chain management* (SCM).

Despite this, considerations about lead times dependent upon resource utilization (*load dependent lead times*: LDLT) are rare in the literature. The same is valid for models linking order releases, planning and capacity decisions to lead times, and taking into account factors influencing lead times such as the system workload, batching and sequencing decisions or WIP levels. That is, in SCM and production planning models nonlinear dependencies, e.g., between lead times and the workload of a production system or a production resource, are usually omitted. This happens even though there is empirical evidence that lead times increase nonlinearly long before resource utilization reaches 100%, which may lead to significant differences in planned and realized lead times.

In this paper we present different approaches on how to account for LDLT. These are indirect approaches primarily aiming at avoiding LDLT and their causes before they arise, e.g., demand variability smoothing methods, job release policies and production planning design, as well as models addressing the problem by implementing LDLT in aggregate production planning using clearing functions.

The literature we review comes from well-known journals as well as less well-known dissertations, which provides the reader a broad set of sources for further studies. We start with a few introductory remarks in the next section. That is, before describing the specific models related to LDLT in the literature, we devote Sect. 2 to the provision of some context in the larger supply chain and production planning literatures. In Sect. 3 we provide detailed descriptions of recent research aimed at the development of aggregate production planning models that are able to capture the nonlinear dependency of lead times and workload directly and nevertheless remain computationally tractable. Finally, some conclusions are provided.

2 Supply chain management and lead time discussion

SCM rose to prominence as a major management issue in the last decades. While the focus of managing supply chains has undergone a drastic change as a result of improved information technology (IT), production planning remains a critical issue. For some references on SCM see Simchi-Levi et al. (2002), Stadtler and Kilger (2005); Voß and Woodruff (2006), just to mention some.

In this section we provide context for the models that will be discussed in Sect. 3 below. Besides a few remarks on information sharing (Sect. 2.1) we provide some links to literature concerning common planning concepts (Sect. 2.2) as well as release mechanisms (Sect. 2.3).

Before going into detail, let us reconsider our lead time definition. Besides our definition given above, Yücesan and de Groote (2000) define lead times as the time between the authorization of production to the completion of processing, at which point the material is ready to fill a customer order. So, lead time consists of queuing time, processing time, batching time, and transportation/handling time.

Even if it is not related to the literature on LDLT it should be mentioned that various forms of lead times may be discussed, such as effectuation lead times (the time between a decision is made and the time the consequences of this decision can be observed throughout the supply chain), minimal lead times, and actual lead times; see, e.g., de Kok and Fransoo (2003).

2.1 Information sharing

Many companies use modern information technology to help them gain competitive advantages in the marketplace. The rapid information technology advancements have provided tools to enable supply chain partners to share information with each other. Yet, questions concerning the benefits that can be gained through the sharing of information are frequently raised. Several researchers have examined the impact of information sharing on business performance; see, e.g., Thonemann (2002) and Zhao et al. (2002). A survey on the impacts of sharing information including some lead time discussion is provided by Huang et al. (2003).

One of the most common information sharing issues is the so-called bullwhip effect. It relates to the observed behavior when forecasts for intermediate products within a supply chain are based only on the direct demand experienced for those products, i.e., variability in the demand pattern is magnified. For references on the bullwhip effect see, e.g., Lee et al. (1997), Chen et al. (2000), Simchi-Levi et al. (2002).

The dynamics and the globalization of markets imply competition in time-based factors like flexibility or responsiveness to consumers, especially in case of variable demand patterns as pointed out by Spearman and Zhang (1999). This results in the need to share business information, especially information related to the production process and the inventory levels for use in planning models. Without proper communication and planning models, decreasing lead times at individual links of the supply chain do not automatically lead to improvements in global supply chain lead times. If, e.g., an individual link of the supply chain is able to respond flexibly to production requirement changes by modulating its own lead time accordingly, this does not result in a benefit for the overall supply chain as long as the information about this flexibility in lead time dynamics is not communicated to the other partners (see Graves and Willems 2000 or Hall and Potts 2003). For instance, Caramanis and Anli (1999b) propose an iterative learning algorithm that shows how efficiency can be increased by sharing information at each individual production and decision node of a production cell. These nodes can also be interpreted as companies related to each other in a supply chain.

2.2 Common planning concepts

Present practice for manufacturing supply chains is dominated by the utilization of models based on *materials requirements planning* (mrp) which are associated

with a variety of problems. Homem-de-Mello et al. (1999) and Caramanis and Anli (1999b) highlight that one of these problems lies in the assumption of fixed and constant lead times at the aggregate level of the planning process and, what is even worse, in the utilization of the "worst case lead time," i.e., the maximal lead time for an individual link. This is done in order to have enough "buffer time" to ensure that the demand requirements can be met in any case (see also Enns and Suwanruji 2004). Additionally, in order to meet due dates, there is a tendency to release jobs into the production system much earlier than necessary, leading to very high WIP levels and therefore longer waiting times. Consequently, lead times get even longer. This over-reaction becomes a self-fulfilling prophecy and is addressed in the literature as the *lead time syndrome*. It results from the fact that the relationship between WIP, output, workload and average flow times is ignored (see Zäpfel and Missbauer 1993a,b or Tatsiopoulos and Kingsman (1983)). Furthermore, these planning models hide the variability and the dependency of lead times on planning factors by defining planned lead times which are much longer than necessary. Additionally, they use planning techniques that ignore consistency constraints in determining production targets for single production cells (stations or resources) or, as Caramanis and Anli (1999b) point out, tighten them so as to linearize and make them time invariant.

Most mrp and *enterprise resource planning* (ERP) models implement sequential planning algorithms which omit many functional interdependencies or uncertainties. In Asmundsson et al. (2003) and Karmarkar (1987) we see that one of the fundamental problems of various manufacturing and production planning models like mrp or *manufacturing resources planning* (MRP II) is the omission of modeling nonlinear dependencies. These include, e.g., the relationship between lead times and the workload of a production system or a production resource. However, there is empirical evidence that lead times increase nonlinearly long before resource utilization reaches 100%. This is also highlighted in Pahl et al. (2005). As a consequence, production plans tend to be inaccurate because of the fact that lead times are extremely dependent on decisions about the product mix, workload and scheduling policies which is also emphasized in Asmundsson et al. (2002, 2003), Caramanis (2000), Caramanis and Anli (1999a,b), Yano (1987), Zijm and Buitenhek (1996). This is particularly true for unsynchronized manufacturing shops where products can follow diverse routings and resources have limited capacity. Nevertheless, mrp systems treat the production lead time of a product, part, or *stock keeping unit* (SKU) as a constant (external) parameter independent of other production factors and, therefore, as an attribute of the part (see Dauzère-Pérès and Lasserre 2002).

Karmarkar (1993b) points out that the lead time of a part is rather a property of the state of the production facility and, consequently, depends on the overall available capacity and workload which can vary significantly over time. Hackman and Leachman (1989) propose a framework for modeling production processes that generalizes and improves the accuracy of simple production models integrated in *linear programming* (LP), mrp and CPM (*critical path method*) based on modeling

and planning techniques by reformulating the constraints of components of production lead times. Billington et al. (1983) address the shortcoming of mrp models to treat the capacity as infinite. They present a mathematical programming approach dealing with capacity constraints which are scheduled considering other production stages and unpredictable production lead times in order to account for the fact that lead times are related to capacity planning. Clark and Armentano (1995) develop a heuristic model for a resource capacitated multi-stage lot sizing problem including general production structures, setup costs, WIP inventory costs and production lead times. The heuristic first calculates a production plan for the uncapacitated case and then shifts production backward in time so as to generate capacity feasible production plans.

Missbauer (2002) integrates the functional relationships between, e.g., the mean level of WIP and important core variables of production planning like flow time or capacity utilization in lot sizing models to analyze the impact of lot sizing on these specific values and dependencies. Moreover, Enns and Choi (2002) and Enns and Suwanruji (2004) develop methods for setting dynamically planned lead times based on the current system workload and lot sizes in MRP II systems. In this respect, they study the use of exponentially smoothed order flow time feedback for the dynamic planned lead time setting in a $GI|G|1$ queuing model. Results reveal that the assumptions considering the *lot inter-arrival time variability* have a great influence on lot sizes and the overall system performance since lead times are significantly affected by the system workload which, in turn, is dependent on capacity and, therefore, setup times.

Voß and Woodruff (2004) concentrate on the fact that decisions about the utilization of production resources in the aggregate planning process affect realized lead times. This is even more important to be captured when widening the aggregate planning process to whole supply chains. They further emphasize the need to simultaneously consider lead time consequences and alternative routings and/or suppliers – which they model as additional routing alternatives – arising during the planning process. This is due to the fact that bottlenecks shift as a result of the production and planning process. The important insight, which is also highlighted in Goldratt and Fox (1986), is that the consideration of interactions of bottlenecks in the production system is of great importance also because of the fact that a resource becoming a bottleneck is dependent on the workload. It therefore depends on decisions about loading and alternative routings. These, in turn, are determined much earlier in the aggregate production planning process and, thus, from the point of view of Voß and Woodruff (2004), should not be distributed to the control level. Their aim is to provide a useful planning model with realistic assumptions from which feasible (realizable) schedules can be derived and which accounts for consequences of LDLT on planning decisions for supply chain networks.

Recent versions of *advanced planning systems* (APS) incorporate finite capacity schedule techniques based on diverse models targeted toward determining accurate start times and schedules for jobs without ignoring capacity constraints

(see, e.g., Fleischmann and Meyr 2003 and the contributions in Stadtler and Kilger 2005). Together with *manufacturing execution systems* (MES) which use network technology in order to provide real time tracking of important production resource information (machines, labor, WIP levels), the APS receive the current production status of the shop floor and the planned order releases from an ERP system with the objective to optimize the production schedule (see, e.g., Homem-de-Mello et al. 1999).

2.3 Release mechanisms

The use of pull-based control mechanisms can increase the predictability of lead times and thereby improve the performance of planning systems. There is vast literature on this subject; see, e.g., Ben-Daya and Raouf (1994), Buzacott and Shantikumar (1993), Hopp et al. (1990), Ouyang et al. (1997), Srinivasan et al. (1988), Wu (2001). In addition, pull-based control systems like KANBAN or CONWIP aim at controlling workflow in order to reduce overall inventories and the related costs; see, e.g., Spearman et al. (1989, 1990). The KANBAN concept is an early suggestion from Japan to implement the pull principle in manufacturing systems. The specific information system includes the KANBAN card, which can be regarded as an authorization for production. The main characteristic of the KANBAN systems is reflected by the closed loops of every production unit. For more details on KANBAN systems see, e.g., Krieg (2003). In CONWIP systems, the manufacturing system is comprised of one closed loop card system, so that one card is attributable to one product container and, consequently, inventory is equal to the number of cards in the system. The name "*CONWIP*" derives from "*CONstant Work In Process*." The CONWIP system can be regarded as a closed queuing network where the orders correspond to the constant number of cards in the system (see, e.g., Hopp and Spearman 2001).

There exist many studies in the literature to date that examine pull release mechanisms based on the workload of production facilities as, e.g., Bertrand and Wortmann (1981), Bertrand (1983) and methods of workload oriented production control as, e.g., Bechte (1982), Bertrand and Wortmann (1981), Bertrand (1983), Wiendahl and Wedemeyer (1990). According to Homem-de-Mello et al. (1999), neither KANBAN nor CONWIP integrate due dates in order to smooth releases and therefore workload, which leads to inaccurate order releases.

Adoption of KANBAN and *just in time* (JIT) techniques in complex production systems that do not have the proper characteristics such as flow balance between production stages, highly similar products etc., is problematic as pointed out by Zäpfel and Missbauer (1993a,b). Moreover, the problem of workload oriented production controls lies in the fact that often, from a supply chain perspective, they do not really solve any problem regarding lead times, but rather increase their mean and variance by shifting the workload from inside to outside the manufacturing shop. That is, jobs that would wait within the shop floor are forced to wait in front

of the shop. This is because workload controls only release jobs when enough capacity is available which is also stated by Zijm and Buitenhek (1996) or Philipoom and Fry (1992).

Other authors argue that holding (delaying) jobs in a so-called "job pool" and thereby controlling a hierarchy of backlog lengths provides a good mechanism to control the shop floor throughput times. As has been argued by Hendry and Kingsman (1991) the latter are an important part of lead times and consequently reduce congestion situations in the production system, WIP levels, LDLT and related costs. Moreover, Philipoom and Fry (1992) investigate the benefits and drawbacks of not only delaying jobs in a job pool, but rejecting customer orders when the system workload or utilization is very high in order to avoid long queues in front of the entry point to the production facility and to improve delivery reliability. Rejecting orders will surely lead to the loss of goodwill, but accepting customer orders including due dates and failing to deliver does lead to even higher loss of goodwill. For further reading see Philipoom and Fry (1992). They also point out that unconditionally accepting customer orders leads to the increase of average flow times resulting in raised uncertainty whether agreed upon due dates are met for all other jobs in the production system. It is surely dependent on business culture in the specific industry whether rejecting orders is practical, but they show that rejecting a small percentage of orders can lead to considerable improvements in system performance. Job release policies and release controls aim at avoiding overload of the production system and, consequently, restrict WIP to reduce queues and therefore the dependency of the lead time on the workload which then allows for the usage of simpler planning models. Witt and Voß (2005) proposed controlling the WIP in a hybrid flow shop by modifying release quantities based on limited intermediate storage.

Agnew (1976) proposes a dynamic model of a general congestion-prone system which illustrates the relationship between throughput and workload of the system. He analyzes the system behavior and the form of optimal policies for most accurately controlling congestions. Land and Gaalman (2003) propose a comprehensive production *workload control model* (WLC) which buffers the shop floor against external dynamics like rush orders or demand peaks and tries to equilibrate the queuing of jobs in the shop floor in a stationary process using workload norms. Various authors have extended these WLC models to hierarchical capacity-oriented production control concepts for job shops as, e.g., Bechte (1988), Bertrand and Wortmann (1981). Missbauer (2002) presents a decision model for the optimization of aggregate order releases that takes into account the dependency of the work center output on the WIP and thus can be considered as a workload control method. Homem-de-Mello et al. (1999) develop a job release method using simulation-based optimization for production scheduling that aims at setting release times for jobs with fixed due dates and determined sequences in stochastic production lines including machine down times, where job release criteria are WIP levels as well as customer service (rates). One objective is to keep flow or cycle

time (sum of processing and queuing time) at a minimum level in order to increase responsiveness and flexibility in manufacturing and to reduce WIP. They consider a transient situation, as opposed to other papers which deal with setting release times using steady state models.

Glassey and Resende (1988) study optimal control policies including information about the type of machine failure in order to minimize the average inventory in two stage flow lines. They describe the optimal control policy by defining a specific threshold level for each type of failure of downstream resources (stations). They show that dynamic control policies reacting to machine failures are most valuable in the case that the second stage resource is the bottleneck, the downtime is not too small and the ratio of the repair rate to the processing rate of the second resource is small.

Further work includes studies on closed-loop job release mechanisms for job shops with machine failures and repair as a main source of randomness. The so-called "starvation avoidance" control mechanism aims at maximizing utilization of critical bottlenecks while controlling the development of the WIP. By monitoring actual WIP levels this job control procedure only releases jobs into the production system when inventories are below a predetermined safety stock level. Besides this, Kim et al. (1996) propose a release control mechanism designed for semiconductor wafer fabrication that uses a tracking system which provides the model with actual information about the shop floor situation. A heuristic is proposed in order to minimize the mean flow time and mean tardiness while observing given output rates. Simulation results are provided for a simplified semiconductor wafer fabrication environment. A more detailed overview of order release algorithms can be found in Hendry and Kingsman (1991), Philipoom and Fry (1992), and Tatsiopoulos and Kingsman (1983).

Wiendahl and Wedemeyer (1990) describe the general dilemma of production planning, scheduling and release control which has its roots in the conflicting planning objectives and resulting trade-offs as, e.g., minimization of lead times and WIP inventory levels while maintaining high utilization levels. They argue that, despite of these problems, it is possible to find an optimal solution to each situation by employing an order release control method. In addition, Zäpfel and Missbauer (1993a) give an overview of production planning and control concepts that aim at overcoming the deficiencies of traditional models such as MRP II. These include concepts like *Optimized Production Technology* (OPT), *Just in Time Production Planning and Control* (JIT-PPC) as well as production planning and control including progressive figures which provide an instrument for monitoring material flows employing a graphic or diagram, but do not include methods for capacity planning.

The OPT concept considers the performance of the production system as dependent on the bottleneck in the system and takes it as a basic constraint (starting point) to calculate the detailed production plan. The OPT scheduling algorithm generates the detailed schedule for the bottleneck and then performs the forward

scheduling of the orders also including the determination of transfer batches and process batches. The resulting flow times and delivery times are the outcome of the planning process which is the most important result of this concept, namely that it does not use estimated planned lead times. The order release process is determined on the so-called "drum beat" of the bottleneck resource, i.e., its production rate which is calculated by the drum-buffer-rope approach (DBR) and determines the schedule for the bottleneck resource. The concept is discussed extensively in Goldratt (1988) and Zäpfel and Missbauer (1993a).

3 Models incorporating load dependent lead times

In this section we will concentrate on the problem of how to develop aggregate production planning models which are able to directly capture the nonlinear dependency of lead times and workload and nevertheless remain tractable computationally. The first reference that we review is Zijm and Buitenhek (1996). They consider the relationship of lead times and workload without using the concept of clearing functions. This is incorporated in the model of Graves (1986) which opens the discussion on clearing functions. Models from Srinivasan et al. (1988), Karmarkar (1987, 1989a), Asmundsson et al. (2002, 2003), and Hwang and Uzsoy (2004) follow in this regard. Moreover, the model of Missbauer (1998) emphasizes lot sizing decisions and integrates them into a clearing function model. Besides, the work of Lautenschläger (1999) concentrates on forward shifting functions for production planning. However, Caramanis and Anli (1999a) are the first who extend the view to the overall supply chain. Mendoza (2003) emphasizes stochastic issues and develops a *transient Little's Law*.

3.1 The planning and control model with workload oriented lead times of Zijm and Buitenhek

Zijm and Buitenhek (1996) develop a manufacturing planning and control framework for a machine shop that includes workload oriented lead time estimates in order to account for the need to consider both lead time and capacity management in one management planning tool. For that purpose they suggest a method that determines the earliest possible completion times of arriving jobs with the restriction that the delivery performance of any other job in the system will not be adversely affected, i.e., that every job can be completed and delivered in time.

The goal is to determine reliable planned lead times based on the workload that results in due dates for jobs that can be met. The model can be implemented at a capacity planning level, serving there as an input for a final detailed capacity scheduling procedure that also takes into account additional resources, job batching decisions as well as machine setup characteristics. They use queuing network techniques to determine the mean and variance of lead times dependent on lot sizes,

production mix and expected production volume. Their framework is partly based on the work by Karmarkar (1993a) and Karmarkar et al. (1985a) as they employ for each network service station a $M|G|1$ queuing model with multiple part types.

Employing different approximations for queuing networks can lead to improvements in the result, e.g., when modeling production system nodes as $GI|G|c$ networks in case of parallel machines. To find an approximation for the average work in the production system that has to be processed at a specific machine, Little's Law is used:

$$\overline{Y}_{ijk} = \lambda_{ijk}\overline{L}_{ijk}$$

where $\overline{Y}_{ijk}$ denotes the mean number of batches of part type i in queue or in process at resource (work station) j for their k-th operation, $\overline{L}_{ijk}$ is the mean lead time corresponding to the k-th operation of part type i at resource j and λ_{ijk} the arrival rate of batches of part type i at resource j for their k-th operation (defined by $\lambda_{ijk} = \frac{D_i}{Q_i}\delta_{ijk}$, where D_i is the demand rate of part type i, Q_i the deterministic lot size of i and δ_{ijk} an indicator function which fixes the part type dependent routing of i). Furthermore, the batch processing time p_{ijk} of part type i at resource j for the k-th operation includes the setup time S_{ijk} and is given by $p_{ijk} = S_{ijk} + Q_i\xi_{ijk}$, where ξ_{ijk} denotes the deterministic processing time for part type i at resource j for the k-th operation. The n-th moment of the part type- and operation-independent batch processing time p_j is then represented by the following formula:

$$E(p_j)^n = \sum_{i=1}^{M}\sum_{k=1}^{K}\frac{\lambda_{ijk}}{\lambda_j}(p_{ijk})^1$$

Here, λ_j denotes the overall arrival rate at resource j. Consequently, the average workload ρ_j at workstation j is given by:

$$\rho_j = \lambda_j E(p_j)^n$$

Assuming Poisson arrival processes at each resource, the Pollaczek-Khintchine formula yields the following expression for the mean waiting time in queue $\overline{Wt}_j$ at workstation j:

$$\overline{Wt}_j = \frac{\lambda_j E(p_j)^2}{2(1-\rho_j)}$$

Moreover, the mean lead time $\overline{L}_{ijk}$ is determined by the mean waiting time in queue at resource j ($\overline{Wt}_j$) and the batch processing time p_{ijk}: $\overline{L}_{ijk} = \overline{Wt}_j + p_{ijk}$. We clearly note the incorporation of the relationship between lead times and workload in this formula. The approximation for the average work in the production system $\overline{W}_j$ is then given as follows:

$$\overline{W}_j = \sum_{i=1}^{M}\sum_{k=1}^{K}\overline{Y}_{ijk}\left(p_{ijk} + \sum_{\hat{k}=k+1}^{K} p_{ij\hat{k}}\right) + \sum_{i=1}^{M}\sum_{\substack{\hat{j}=1\\ \hat{j}\neq j}}^{N}\sum_{k=1}^{K}\overline{Y}_{i\hat{j}k}\left(\sum_{\hat{k}=k+1}^{K} p_{i\hat{j}\hat{k}}\right)$$

The first term refers to the jobs currently present at resource j and includes both current (k) and future operations ($\hat{k}$). The second term represents the jobs currently present at other resources ($\hat{j} \neq j$) which could still have to visit resource j. Furthermore, $p_{ij\hat{k}}$ denotes the batch processing time of the future operation $\hat{k}$ on part type i at resource j. The term can be interpreted as the number of jobs at resource j on which the jobs have to undergo present and future operations plus the number of jobs which have not already been processed on resource j. Subsequent to the determination of both mean and variance of the lead times and the average work in the production system is a decomposition-based job shop scheduling algorithm (the *shifting bottleneck procedure*). The capacity planning procedure is then used to calculate the individual planned lead times for each batch, based on the actual workload of the production system.

To be more precise, first a scheduling algorithm is employed to minimize the maximum lateness for each job under the special characteristic that a job is allowed to visit a workstation, once, more than once, or not at all. Despite the fact that this is an NP-hard problem the shifting bottleneck heuristic gives satisfactory results. The decomposition algorithm then distributes the overall problem to a series of single machine scheduling problems for determining the virtual release and due dates for the operations with the objective of minimizing the maximum lateness. Once assigned, these due dates are fixed, i.e., modifications are not possible.

The capacity planning method based on aggregate scheduling finally assigns the due dates to each job depending on the current workload of the production system under the aforementioned condition that all jobs in the system can be finished without delay. These due dates further give the input for the detailed shop floor scheduling. What is obviously different here is the explicit consideration of different routings and machines at their capacity limit which is not included in classical "bucket filling" capacity planning models as the latter tend to smooth the needed capacity over the scheduling time period and consequently ignore precedence relations. The model represented here is designed for a highly simplified machine shop, but can be modified for several shop configurations. Although it does not consider the nonlinear relationship between workload and LDLT, it provides a first approach for considering the workload of a production system at an aggregate planning level.

3.2 The tactical planning model of Graves

Graves (1986) develops a tactical planning model for a job shop for the purpose of studying the dependencies between production capability, variability (uncertainty) of the production requirements, and level of WIP inventory. He tries to analyze to which extent the job flow time (or WIP inventory) depends on the utilization of each resource of a job shop or production stage, respectively. Furthermore, the interrelationship of flow time and production mix is analyzed. The underlying production system is a flexible job shop, modeled as a network of queues where multiple routings of jobs are possible so that the lack of a dominant work flow

complicates production control which aims at considerably reducing the variance of planned lead times. Nonetheless, input/output control systems try to manage the work flow through the shop by stabilizing queues at a predetermined level.

Planned lead times are used as the key decision parameters for implementing an input/output control in order to describe production rates by resource and by time period. For that purpose a queuing model is employed that includes *flexible production rates* of resources which can be set by a tactical planning model as to smooth the work flow in order to avoid overload of resources. According to Karmarkar (1989a) this imposes difficulties, because it implies that, applying Little's Law, lead times are fixed even when workload varies. Similar to Bertrand and Wortmann (1981), Graves (1986) develops a time discrete, aggregate planning model which attempts to control the flow time of jobs and hence the workload in the production system. This discrete-time, continuous flow model considers workload rather than single jobs. It is assumed that the movement of jobs from one resource to another and the arrival of new jobs only occur at the beginning or at the end of a time period, respectively. A control rule is implemented at each resource that determines the amount of work performed during a time period, which is a fixed portion α_j of the queue of work remaining at the start of the period at resource j:

$$X_{jt} = \alpha_j W_{jt} \quad \forall j, t$$

where X_{jt} denotes the production at resource j in time period t, α_j a smoothing parameter with $0 < \alpha_j < 1$ and W_{jt} the queue of work at the beginning of time period t. This smoothing parameter is called "proportional factor" by Missbauer (1998). Infinite capacity is implied since the resource is able to complete the fixed portion α_j even when the workload (WIP) is infinitely high. The major drawback of the model is that it employs fixed planned lead times, i.e., it is assumed that production processing rates can be regulated in order to account for workload in the production system. This implies, as mentioned above, by applying Little's Law, that lead times are fixed even when the workload varies. So lead time consequences, due to varying workload in the production system, are not really captured in the model. Additionally, the model assumes static average loading with stochastic variations. Nevertheless, Karmarkar (1989a) argues that these restrictions can be overcome and extensions to generate a dynamic model can be created.

3.3 Clearing functions

Karmarkar (1989a) and Srinivasan et al. (1988) independently take up the idea of Graves (1986) and integrate it in a mathematical planning model with so-called clearing factors $\alpha(\text{WIP})$ which are nonlinear functions of the WIP, thus yielding a clearing function of the following form:

$$Capacity = \alpha(\text{WIP}) \cdot \text{WIP} = f(\text{WIP})$$

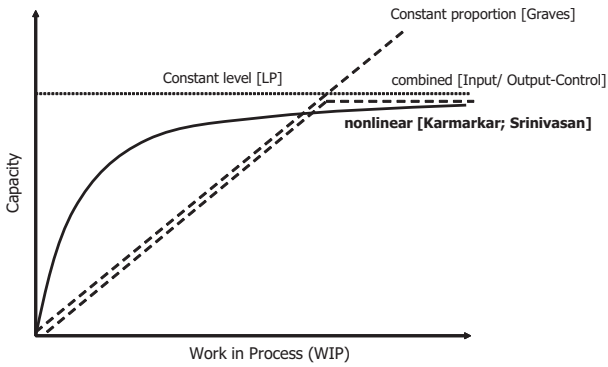

Fig. 1. Diverse forms of clearing functions

where f represents the clearing function that models capacity as a function of the workload. In Asmundsson et al. (2003) the clearing factor specifies the fraction of the actual WIP which can be completed, i.e., "cleared," by a resource in a given period of time. (Missbauer 1998 refers to this function as a "utilization function".)

Figure 1 depicts some possible forms of clearing functions where the constant level clearing function corresponds to an upper bound for capacity as mainly employed by LP models. This implies instantaneous production without lead time constraints since production takes place independently of WIP in the production system. The constant proportion clearing function represents the control rule given by Graves (1986) which implies infinite capacity and hence allows for unlimited output. Moreover, in contrast to the nonlinear clearing function of Karmarkar (1989a,b) and Srinivasan et al. (1988), the combined clearing function in some region underestimates and in others overestimates capacity. The nonlinear clearing function relates WIP levels to output and lead times to WIP levels which are influenced by the workload of the production system (see also Karmarkar 1993b) and therefore is able to capture the behavior of LDLT. By applying Little's Law the clearing function can be reinterpreted in terms of lead time or WIP turn, respectively. Additionally, the slope of the clearing function represents the inventory turn with lead time given by the inverse of the slope, which is stated by Karmarkar (1989a).

Next we provide various clearing function models.

3.3.1 Clearing function model by Srinivasan, Carey and Morton

Srinivasan et al. (1988) construct a nonlinear aggregate discrete-time continuous flow model with variable job arrival times and finite horizon settings which aims at determining aggregate schedules and optimal WIP inventory levels for each workstation taking into account the effects of WIP levels on (throughput) capacity and on lead time as well as the marginal value of capacity and therefore determining

prices for each workload situation of a workstation. A workstation is defined as an aggregation of individual machines, e.g., different machines, flexible machine centers, flow lines, parallel machines or a department. A more detailed lower level model will derive plans for each single machine. Furthermore, the model captures the continuously changing lead times at each workstation, thus allowing for better lead time estimates for due date setting as well as the changing utilization of workstations due to varying demand patterns reflected in the alternating value of capacity. Finally, an optimal routing scheme for each job through the production network is determined.

There exist two different approaches of production planning models in the literature to date, namely the *monolithic* and the *hierarchical* planning approach. The former includes both tactical production planning and operational scheduling in one broad (overall) mathematical programming model whereas the latter captures the need for planning control at each organizational level of the production process. Hierarchical systems employ an aggregate model that serves as a framework for the more detailed lower level models providing primal solution information serving as further modeling constraints. For further reading see Bitran et al. (1981, 1982), and Dempster et al. (1981).

The optimization model presented by Srinivasan et al. (1988) resides between tactical planning and detailed scheduling. The production system is viewed as a network (congestion) queuing model with the workstations connected by transportation links where jobs compete for machine capacity (contention system). Capacity is modeled as a concave function of WIP at the workstation or the length of the dynamic line. Additionally, it is assumed that capacity increases with a declining rate thereby allowing for the existence of implicit resource prices at each utilization level. This is not possible when capacity is modeled as an upper bound such that implicit resource prices are only non-zero at 100% utilization.

The objective of the model is to minimize the overall costs of production, namely the holding costs, the processing costs and the tardiness/earliness costs, stated in the following objective function (with T indicating the number of periods, M the number of jobs, and N denoting the number of workstations without considering the final workstation d):

$$\min \sum_{i=1}^{M} \sum_{t=1}^{T} \left[\sum_{j=1}^{N} \left(\omega_{ijt} W_{ijt} + \phi_{ijt} X_{ijt} \right) + w_i e_i(t) X_{idt} \right]$$

with $\phi_{ijt} X_{ijt}$ denoting the processing costs (ϕ_{ijt} is the processing cost factor which only considers variable and labor costs directly attributable to units of each processed job and X_{ijt} the number of jobs i at workstation j at the beginning of the t-th time period in external or internal queues at individual machines or in process (the production costs do not need to vary over time) and $w_{ijt} W_{ijt}$ representing the holding costs per job i at workstation j in period t (w_{ijt} is the holding cost factor and W_{ijt} denotes the WIP level of job i at workstation j in the t-th time period). The

last term of the objective function indicates the tardiness or earliness costs with w_i being a priority weight for job i, $e_i\ (t)$ representing the tardiness/earliness function for job i completed in period t which includes penalty costs for late delivery, lost sales and lost goodwill and X_{idt} denoting the actual throughput of job i at the final workstation d in the t-th time period. The tardiness/earliness function $e_i\ (t)$ can be specified in the following form:

$$e_i(t) = \left\{ \begin{array}{rll} e_E(\delta_i - t) & if & t < \delta_i \\ 0 & if & t = \delta_i \\ e_L(t - \delta_i) & if & t > \delta_i \end{array} \right\}$$

where $e_E(\cdot)$ and $e_L(\cdot)$ are convex functions of the tardiness and earliness, respectively. The δ_i's represent the due dates for job i. Moreover, quadratic or linear tardiness/earliness functions can be specified. This early/tardy-objective is a nonregular performance measure. Furthermore, the objective function is subject to several constraints. For the first three constraints, the associated Lagrangian multipliers are shown in set braces. They are subsequently used to derive the implicit prices for capacity.

$$\{\psi_{ijt}\}\ W_{ij,t+1} = W_{ijt} - X_{ijt} + F_{ijt} + \sum_{a \in A_I(i,j)} Y_{iat} \quad \forall i, j, t = 1, ..., T-1$$

$$\{\mu_{ijt}\}\ X_{ijt} = \sum_{a \in A_O(i,j)} Y_{iat} \quad \forall i, j, t = 1, ..., T-1$$

$$\{\eta_{jt}\} \sum_{i=1}^{M} X_{ijt} \leq f_{jt}\left(W_{1jt}, ..., W_{ijt}, ..., W_{Mjt}\right) \quad \forall j, t = 1, ..., T-1$$

$$W_{ij1} = \dot{W}_{ij1} \quad \forall i, j$$

$$W_{ijt}, Y_{iat}, F_{ijt} \geq 0 \quad \forall i, j, a, t$$

The first constraint represents a workload oriented conservation equation for each workstation. It gives the number of jobs i at workstation j in period $t+1$ ($W_{ij,t+1}$) which equals the difference of the number of jobs i of the previous period and the actual throughput of workstation j in period t plus the external inflow F_{ijt} of jobs i at workstation j in period t. The last term sums over inflows $A_I(i, j)$ denoting the transportation arcs pointing into workstation j for job i and of inflows from preceding workstations ($\sum_{a \in A_I(i,j)} Y_{iat}$). Here, Y_{iat} denotes the number of units of jobs i on arc a in period t. The F_{ijt}-values may be seen as "demand" for resource utilization. Based on that the authors include them in the nonnegativity constraints.

The second constraint states that the throughput of job i at workstation j in period t, the "clearing factor," equals the number of jobs transported to all its subsequent workstations ($\sum_{a \in A_O(i,j)} Y_{iat}$) where $A_O(i, j)$ denotes the set of transportation arcs pointing out of workstation j for job i. This "clearing factor" is

in turn restricted by the capacity function $f_{jt}\left(W_{1jt}, ..., W_{ijt}, ..., W_{Mjt}\right)$, represented in the third constraint, where any slack, i.e., $f_{jt}\left(W_{1jt}, ..., W_{ijt}, ..., W_{Mjt}\right) - \sum_{i \in M} X_{ijt} > 0$, represents the WIP ("number of units of jobs that are held over at workstation j from period t to $t+1$"). In order to prevent congestion downstream in the production chain, holding WIP at a workstation can be efficient even if this causes additional holding costs. The fourth restriction forces the initial loading of each workstation j to be non-zero with a given initial parameter $\dot{W}_{ij1} > 0$. For more details see Srinivasan et al. (1988). If the objective function and the constraints are strictly convex there is a unique optimal solution. The fifth restriction summarizes the usual non-negativity constraints.

The Karush-Kuhn-Tucker conditions for the optimization problem are, for all $i \in M$ (by f'_{jt} we denote the first derivative of the capacity function):

$$
\begin{aligned}
&(K1)\ \left\{W_{ijt} \geq 0\right\} \omega_{ijt} - \psi_{ij,t-1} - \eta_{jt} f'_{jt} + \psi_{ijt} \geq 0 \qquad \forall i, j, t \\
&(K2)\ \{Y_{iat} \geq 0\} \mu_{ijt} - \psi_{iat} \geq 0 \qquad \forall i, jt, a \in A_O(i, j) \\
&(K3)\ \left\{X_{ijt}\right\} \eta_{jt} - \psi_{ijt} + \mu_{ijt} = \phi_{ijt} \qquad \forall i, j, t \\
&(K4)\ \{X_{iat}\} w_i e_i(t) - \psi_{iat} = 0 \qquad \forall i, a, t \\
&(K5)\ \eta_{jt} \geq 0 \qquad \forall j, t
\end{aligned}
$$

together with complementary slackness conditions for all pairs of inequalities in $(K1)$ and $(K2)$. Considering that the function f_{jt} is bounded by the WIP and 0, the variables X_{ijt} and the second constraint can be omitted from the formulation of the problem which leads to the following compact presentation of constraints of the optimization model:

$$
\begin{aligned}
&W_{ij,t+1} \geq W_{ijt} - f_{jt}\left(..., W_{ijt}, ...\right) + F_{ijt} + \sum_{a \in A_I(i,j)} Y_{iat} \ \forall i, j, t = 1, ..., T-1 \\
&W_{ij1} = \dot{W}_{ij1} \quad \forall i, j \\
&W_{ijt}, F_{ijt}, Y_{iat} \geq 0 \quad \forall i, j, t, a
\end{aligned}
$$

The capacity function f_{jt} is assumed to be a concave non-decreasing function. This is due to the fact that throughput increases as the number of jobs at a workstation increases, but at a declining rate because of congestion.

Srinivasan et al. (1988) discuss different forms of capacity functions and ways to operationalize them. In order to derive implicit capacity prices at all levels of utilization, the Lagrangian multipliers of the mathematical model are employed, where ψ_{ijt} represents the *marginal remaining processing cost* (MRPC) for all jobs i at workstation j in period t, μ_{ijt} denotes the MRPC incurring for job i after its processing by workstation j, and η_{jt} denotes the value of an additional unit of workstation j or the prices of the capacity of workstation j in period t, respectively. When $X_{ijt} > 0$, then the price of capacity of workstation j (η_{jt}) is equal to the sum of marginal production costs and marginal waiting costs caused by job i at

workstation j, implied by $(K3)$ of the Karush-Kuhn-Tucker conditions:

$$\eta_{jt} = \phi_{ij} + \psi_{ijt} - \mu_{ijt}$$

where ϕ_{ijt} is the marginal production cost incurred at workstation j and $\psi_{ijt} - \eta_{ijt}$ the marginal cost entailed by job i waiting at the workstation j in period t. However, when $W_{ijt} > 0$, the price of the capacity of workstation j in period t is stated by the following equation implied by the complementary slackness of $(K1)$:

$$\eta_{jt} = \frac{\omega_{ijt} + \left(\psi_{ijt} - \psi_{ij,t-1}\right)}{f'_{jt}}$$

where η_{jt} denotes the capacity price of the workstation, ω_{ijt} represents the holding costs, $\psi_{ijt} - \psi_{ij,t-1}$ states the changes in the MRPC and f'_{jt} denotes the slope of the capacity function. When $\psi_{ijt} - \psi_{ij,t-1} > 0$, the MRPC is rising over time and so does congestion. The price of the workstation capacity is bounded from below by the holding costs ω_{ijt}. Furthermore, when the slope of the capacity function approaches zero, $\psi_{ijt} - \psi_{ij,t-1}$ approaches ω_{ijt}. Besides this, the marginal value of WIP (β_{ijt}) is given by the change of the slope of the capacity function and the marginal value of additional capacity given by η_{jt}:

$$\beta_{ijt} = f'_{jt}\eta_{jt} = \omega_{ijt} + \left(\psi_{ijt} - \psi_{ij,t-1}\right)$$

This again follows from the complementary slackness $(K1)$ for $W_{ijt} > 0$. It states that the marginal value of WIP is the marginal saving due to marginally increased capacity utilization and therefore lower WIP levels. In periods of rising congestion $\left(\psi_{ijt} - \psi_{ij,t-1} > 0\right)$ the holding costs bound the marginal value of WIP from below, i.e., $\beta_{ijt} > \omega_{ijt}$ and vice versa for the case of declining congestion.

Srinivasan et al. (1988) examine the existence of a set of equilibrium flows for a given long time horizon which is interesting for input/output-control in order to arrive at steady state production rates that allow for low WIP levels and consequently low production costs. Assuming constant external inflow over a sufficiently long time horizon F_{ijt} (which will be subject to relaxation, thus, the t-index is kept for F_{ijt}) and an objective function only depending on the WIP, there exists a solution to the following problem:

$$\min \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{t=1}^{T} \omega_{ijt} f_{jt}^{-1} \left(\sum_{a \in A_O(i,j)} Y_{iat} \right)$$

subject to:

$$\left\{\mu_{ijt}\right\} \sum_{a \in A_O(i,j)} Y_{iat} = F_{ijt} + \sum_{a \in A_I(i,j)} Y_{iat} \quad \forall i, j, t = 1, ..., T-1$$

$$X_{iat} \geq 0 \qquad \forall i, j, t, a \in A_O(i, j)$$

If $f_{jt}(\cdot)$ is strictly concave for all j, there exists a unique optimal solution, because then the inverse of $f_{jt}(\cdot)$ is a strictly convex function and the above stated problem would be a minimum convex cost problem. In steady state there is no time variation in the MRPC and consequently the price of workstation j is only a function of the holding costs and the utilization, and the marginal value of WIP is constant and equal to the holding costs. Thus the price of the workstation capacity is given by

$$\eta_{jt} = \frac{\omega_{ijt}}{f'_{jt}}$$

which results from substituting $\left(\psi_{ijt} = \psi_{ij,t-1}\right)$ in the complementary slackness condition associated with $(K1)$, because in equilibrium the MRPC remain constant over time.

3.3.2 Clearing function model of Karmarkar

Karmarkar (1989a) develops a capacity and release planning model which explicitly takes into account WIP costs and lead time consequences caused by the production system workload. For that purpose it is based on order release and batching models and applies the traditional capacity planning methodology that combines release planning, master scheduling issues and seasonal planning. Additionally, it aims at surmounting the shortcomings of aggregate production planning models like those of Graves (1986), Kekre and Kekre (1985), Karmarkar et al. (1987), and the limitations of input/output-control models. Like Srinivasan et al. (1988), Karmarkar (1989a) uses the nonlinear "clearing function" to represent the output as a function of the average WIP in the production system. The form of the curve is also valid for synchronous deterministic flow lines with batched flows.

The clearing function is now discussed for different types of production systems. For example, examining a synchronous production process with no queuing as well as batching finished and leaving at the end of each production cycle gives the following function for the actual production rate X:

$$X = \frac{Q}{C} = \frac{PQ}{(PS + Q)}$$

Q denotes the batch size for all products, C stands for the cycle time at a workstation, also denoted as $C = S + Q/P$, P represents the nominal production rate at all workstations, and S denotes the setup time between the batches. Substituting Q in terms of WIP with $Q = W/N$ (where N denotes all workstations) gives:

$$X = \frac{PW}{NPS + W}$$

This clearing function has the same nonlinear form as can be seen in Fig. 1. Standard results for an $M|M|1$ queuing system give the utilization ρ (with arrival rate Λ) as follows:

$$\rho = \lambda X = \frac{X}{P} + \frac{XS}{Q}$$

This is equal to the formula of the average workload of Zijm and Buitenhek (1996). Furthermore, the average manufacturing lead time $\overline{L}$ is given as follows:

$$\overline{L} = \frac{S + \frac{Q}{P}}{1 - \frac{X}{P} - \frac{XS}{Q}}$$

In order to keep things simple we ignore the setup times and write the total WIP as $W = \overline{L}X$, substituting $\overline{L} = W/X$:

$$\frac{W}{X} = \frac{\frac{Q}{P}}{1 - \frac{X}{P}}$$

Solving this relation for X results in the following equation:

$$X = \frac{PW}{Q + W}$$

For more detail see Karmarkar et al. (1985b) or Karmarkar (1987). This function is clearly similar to the clearing function, although derived in a different context. Furthermore, applying an $M|G|1$ model for an asynchronous multi-product production facility with lead times and WIP as a function of the batching policy, ignoring setup times and assuming that the average queuing time $\overline{Wt}$ is given by the Pollaczek-Khintchine formula leads to the following form:

$$\overline{Wt} = \frac{\sum_{i=1}^{M} \frac{X_i}{Q_i} \left(\frac{Q_i}{P_i}\right)^2}{2\left(1 - \sum_{i=1}^{M} \frac{X_i}{P_i}\right)}$$

with i denoting the index for item or production part. This is equal to the formula of Zijm and Buitenhek (1996). The total average lead time in the system for batches of items i is represented by:

$$\overline{L}_i = X_i + \overline{Wt}$$

while the WIP for this production system is derived as follows:

$$W_i = X_i(\overline{Wt} + \frac{Q_i}{P_i})$$

In the case of a multi-product system there exist i equations for i products. Furthermore, the output of each product is affected by the WIP level of all other products. In order to keep things simple, Karmarkar (1989a) considers a discrete period model for a single product production system to proceed to the dynamic reformulation of the model. Assuming that demand is satisfied, the mathematical formulation is stated as follows for all t:

$$W_{t-1} + R_t = W_t + X_t$$

$$I_{t-1} + X_t = I_t + D_t$$
$$X_t = f\,(W_{t-1}, R_t, P_t)$$

Here, R_t represents the job releases in period t, W_t denotes the WIP at the end of period t carried over to period $t+1$ and W_{t-1} denotes WIP at the start of period t. The same is valid for I_{t-1} and I_t representing the inventory. Moreover, D_t denotes the demand in period t. As mentioned above, the output or actual production in period t is a function of WIP, job releases and the maximum production possible in period t denoted by P_t. The first equation states that when WIP levels are small, output cannot exceed the amount in the system. In order to formulate the mathematical model as an optimization model we need to determine the costs of the production system like holding costs for WIP and *finished goods inventory* (FGI), processing costs and penalty costs etc.

The mathematical programming model is then stated as follows:

$$\min \sum_{t=1}^{T} [\omega_t W_t + \pi_t I_t + \phi_t X_t + \vartheta_t B_t + r_t R_t]$$

subject to:

$$W_{t-1} + R_t - W_t - X_t = 0$$
$$I_{t-1} + X_t - I_t + B_t = D_t$$
$$X_t - f\,(W_{t-1}, R_t, P_t) \leq 0$$
$$X_t, R_t, W_t, B_t, I_t \geq 0$$

with $\omega_t, \pi_t, \phi_t, \vartheta_t$ and r_t for all periods denoting the WIP holding costs, the FGI holding costs, the variable production costs, the backorder costs due to unsatisfied demand, and the release costs due to, e.g., raw material expenditures. Therefore, B_t denotes the backorders in period t. The first equation represents the mass balance restriction of WIP and releases. The second equation gives information about how demand is met. The third inequality gives the clearing function and the fourth one states the usual non-negativity assumptions.

In contrast to common LP formulations, the effects of capacity loading are captured by the costs of WIP, which in turn is necessitated by the clearing function. Moreover, the resulting release plan will differ from the production plan X, because of lead times which are affected by capacity loading. This interdependence is explicitly accounted for in the model.

The clearing function seems to be the key to explicitly taking into account lead time effects and related dependencies. Its functional form and parameterization has to be studied in more detail as will subsequently be the case. Furthermore, in the literature up to date there only exists one paper that analyzes the fitting of the clearing function to empirical data, namely the work of Asmundsson et al. (2003). Besides this, the case of multi-product manufacturing systems has to be examined more closely especially concentrating on the form of the clearing function, because

output levels may also depend on WIP of each product with WIP being a mix of all products. We will now discuss an approach that pursues this objective.

3.3.3 Clearing function model by Asmundsson, Rardin and Uzsoy

The mathematical programming approach of Asmundsson et al. (2002, 2003) models the nonlinear dependency between lead times and WIP (workload) by employing a clearing function, too. Special properties of the clearing function allow for formulating a linear program version in order to develop a model which remains numerically tractable. In accordance with the procedure of Karmarkar (1989a), Asmundsson et al. (2002, 2003) define the performance of a resource (work center) as dependent on the workload. For this reason they use a $G|G|1$ queuing model:

$$W = \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho^2}{1-\rho} + \rho$$

$$\rho = \frac{\sqrt{(W+1)^2 + 4W\left(c^2 - 1\right)} - (W+1)}{2\left(c^2 - 1\right)}$$

The first equation gives the WIP for a resource which is presented with the coefficient of variation for the service time c_a and arrival time c_s. The second equation represents the utilization ρ of a resource as a function of the WIP with c denoting a parameter to be defined. Basically, the authors simplify the parameters from the first equation according to: $c^2 = \frac{c_a^2+c_s^2}{2}$. As mentioned above, batching and lot sizing have an effect on lead times especially when small batches give rise to frequent setup changes leading to time losses for production, lower throughput and eventual starvation of resources on further production stages. Since the formulation given above only includes starvation we can reformulate the resource utilization also accounting for setup time as follows:

$$\frac{Q\overline{\sigma}}{Q\overline{\sigma} + S} \frac{W + 1 - \sqrt{W^2 + 2c^2W + 1}}{1 - c^2}$$

where Q denotes the lot size, $\overline{\sigma}$ the mean service time and S the setup time. The lot size is assumed to be set independently of the system utilization. Here $Q\overline{\sigma} + S$ is the average total time to service a lot. It can be noticed that when setup time S increases the utilization decreases and so does the lot size Q. So we have another nonlinear relationship as lot sizes tend to be increased for high WIP levels in order to clear queuing in front of a resource and vice versa (see Karmarkar 1989a). When setup time increases the clearing function shifts downwards.

The clearing function is assumed to have the characteristics depicted in Fig. 2 where the first inequality states that the clearing function is bounded by zero and f^{max}. The second and the third inequality reflect the property that the clearing function is monotonically increasing in a defined measure of WIP with a declining

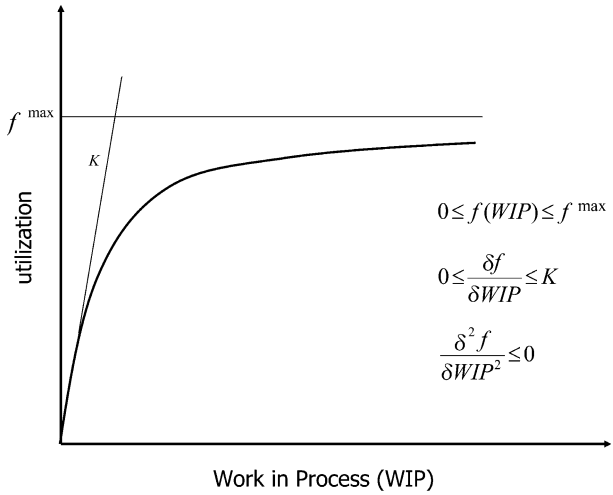

Fig. 2. Characteristics of the clearing function

rate, i.e., f is concave. Besides this, the second inequality imposes the restriction that the slope of the function is bounded from above by K. Furthermore, the function should be well-behaved in the presence of variability and lot sizing as mentioned above and also in Asmundsson et al. (2002).

In order to develop the clearing function, there are two methods available in the literature to date, where the first is the analytical derivation from queuing network models and the second an empirical approximation using a functional form which can be fitted to empirical data. Because of the large complexity of practical systems the complete identification of the clearing function will not be possible, so we have to work with approximations. Asmundsson et al. (2002) integrate the estimated clearing function in a mathematical programming model where the framework is based on the production model of Hackman and Leachman (1989) with an objective function that minimizes the overall costs. It is assumed that backorders do not occur and that all demand must be met on time. One may assume a representation as a graph with N nodes refering to the workstations and a set A of transportation arcs between these nodes. Predecessors of workstations are indicated by A_I and successors by A_O, respectively. The model is then stated as follows:

$$\min \sum_{i=1}^{M} \sum_{t=1}^{T} \left[\sum_{j=1}^{N} (\phi_{ijt} X_{ijt} + \omega_{ijt} W_{ijt} + \pi_{ijt} I_{ijt} + r_{ijt} R_{ijt}) + \sum_{\gamma \in A} \theta_{i\gamma t} \Gamma_{i\gamma t} \right]$$

subject to:

$$W_{ijt} = W_{ij,t-1} - \frac{1}{2}\left(X_{ijt} + X_{ij,t-1}\right) + R_{ijt} + \sum_{\gamma \in A_I(i,j)} \Gamma_{i\gamma t} \qquad \forall i, j, t$$

$$I_{ijt} = I_{ij,t-1} + \frac{1}{2}\left(X_{ijt} + X_{ij,t-1}\right) - D_{ijt} - \sum_{\gamma \in A_O(i,j)} \Gamma_{i\gamma t} \qquad \forall i, j, t$$

$$X_{\cdot jt} \leq f_{jt}\left(W_{\cdot jt}\right) \qquad \forall j, t$$

$$X_{ijt}, W_{ijt}, \Gamma_{i\gamma t}, I_{ijt}, R_{ijt} \geq 0 \qquad \forall i, j, t, \gamma$$

where $\phi_{ijt} X_{ijt}$ denotes the costs of the total amount of production over the latter half of period t and the first half of period $t+1$, represented by X_{ijt}, with ϕ_{ijt} refering to the corresponding unit costs of item i at node j in period t. The WIP costs and the FGI costs of item i at workstation j at the end of period t are denoted by $\omega_{ijt} W_{ijt}$ and $\pi_{ijt} I_{ijt}$, respectively, with ω_{ijt} and π_{ijt} being the corresponding unit costs and W_{ijt} and I_{ijt} representing the WIP and the inventory, respectively. Likewise, the costs of releases of raw material of item i at workstation j during period t are represented by $r_{ijt} R_{ijt}$ with unit costs r_{ijt}. The transfer costs of item i on arc γ during period t are given by $\theta_{i\gamma t} \Gamma_{i\gamma t}$ with corresponding unit costs $\theta_{i\gamma t}$. Asmundsson et al. (2002) assume these costs as independent from the item.

The first two constraints denote the flow conservation for WIP and FGI which is different from classical models since inventory levels at each node in the network are connected with the throughput rate. Note that because X_{ijt} denotes production that spans the period boundary, the balance equations average the adjacent times.

In contrast to Ettl et al. (2000) the nonlinear dynamic is incorporated in the clearing function in the constraints and thus not included in the objective function. Lead times are calculated using Little's Law:

$$\widehat{L}_{ijt} = \frac{W_{ijt}}{X_{ijt}}$$

where $\widehat{L}_{ijt}$ denotes the expected lead time for the last job of item i which arrived before the end of period t. In order to consider multiple product types with different resource consumption patterns, the lead times for single items i must be derived. For that purpose, FIFO processing is assumed, which results in the following relationship:

$$\frac{X_{ijt}}{X_{\cdot jt}} = \frac{W_{ijt}}{W_{\cdot jt}} \qquad \forall i, j, t$$

Here, $X_{\cdot jt}$ is the production quantity and $W_{\cdot jt}$ the WIP level summarized over all items i. Taking this relation and multiplying the production quantities X_{ijt} by their resource consumption factor ξ_{ijt} which defines the capacity consumption per unit produced for item i at workstation j, we derive a new variable Z_{ijt} of the following form:

$$Z_{ijt} = \frac{\xi_{ijt} X_{ijt}}{X_{\cdot jt}} = \frac{\xi_{ijt} W_{ijt}}{W_{\cdot jt}} \qquad \forall i, j, t$$

By implementing this variable Z_{ijt}, we obtain the clearing function for each item i which is called the partitioned clearing function:

$$\xi_{ijt} X_{ijt} \leq Z_{ijt} f_{jt}\left(\frac{\xi_{ijt} W_{ijt}}{Z_{ijt}}\right) \qquad \forall i, j, t$$

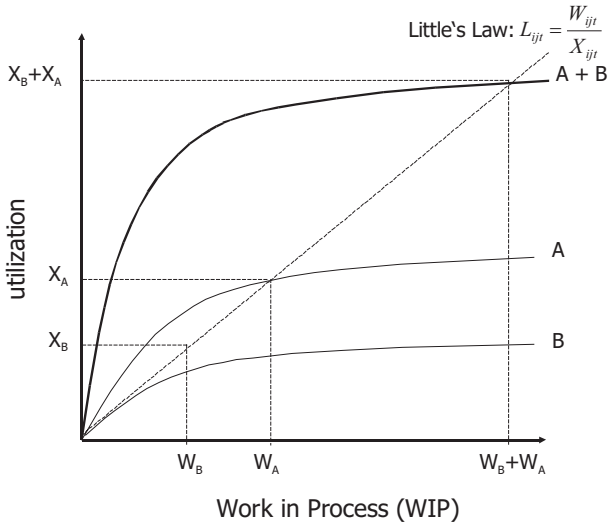

Fig. 3. Partitioned Clearing function for products A and B

with the following properties:

$$\sum_{i=1}^{M} Z_{ijt} = 1 \qquad \forall j, t$$

$$Z_{ijt} \geq 0 \qquad \forall i, j, t$$

The partitioned clearing function is depicted in Fig. 3 (note that the functions are symboling rather than detailed to scale).

In order to relax the assumed priority rule (FIFO) we only suppose that Z_{ijt} satisfies the properties stated above, but has an arbitrary functional form. With this formulation Asmundsson et al. (2002) succeed in integrating the nonlinear relationship between WIP and lead times in a mathematical programming model. The second goal is to transform this model in a tractable form which allows for dealing even with relatively large planning problems. For that purpose, an LP formulation is used that represents the partitioned clearing function through a set of linear constraints. To be more precise, the clearing function is approximated by the minimum of straight lines which is possible because of its concavity as shown in Fig. 4:

$$f_{jt}\left(W_{\cdot jt}\right) = \min_{c} \left\{\alpha_{jt}^{c} W_{\cdot jt} + \beta_{jt}^{c}\right\} \qquad \forall j, t, c$$

The individual lines of the items are denoted by the index c. The α_{jt}^{c} and β_{jt}^{c} coefficients represent the slope and the intersection of line c with the y-axis. The capacity splitting (sharing) across the items appears in sharing the β_{jt}^{c} coefficients.

Applying this formulation to the partitioned clearing function leads to the following form:

$$Z_{ijt} f_{jt}\left(\frac{\xi_{ijt} W_{ijt}}{Z_{ijt}}\right) = Z_{ijt} \cdot \min_c \left\{\alpha_{jt}^c \frac{\xi_{ijt} W_{ijt}}{Z_{ijt}} + \beta_{jt}^c\right\}$$
$$= \min_c \left\{\alpha_{jt}^c \xi_{ijt} W_{ijt} + \beta_{jt}^c Z_{ijt}\right\}$$

Replacing the former capacity constraint of the original nonlinear mathematical programming model with nonlinear lead time and capacity dynamics gives the complete linearized formulation:

$$\min \sum_{i=1}^{M} \sum_{t=1}^{T} \left[\sum_{j=1}^{N} (\phi_{ijt} X_{ijt} + \omega_{ijt} W_{ijt} + \pi_{ijt} I_{ijt} + r_{ijt} R_{ijt}) + \sum_{\gamma \in A} \theta_{i\gamma t} \Gamma_{i\gamma t} \right]$$

subject to:

$$W_{ijt} = W_{ij,t-1} - \frac{1}{2}\left(X_{ijt} + X_{ij,t-1}\right) + R_{ijt} + \sum_{\gamma \in A_I(i,j)} \Gamma_{i\gamma t} \qquad \forall i, j, t$$

$$I_{ijt} = I_{ij,t-1} + \frac{1}{2}\left(X_{ijt} + X_{ij,t-1}\right) - D_{ijt} - \sum_{\gamma \in A_O(i,j)} \Gamma_{i\gamma t} \qquad \forall i, j, t$$

$$\xi_{ijt} X_{ijt} \leq \alpha_{jt}^c \xi_{ijt} W_{ijt} + Z_{ijt} \beta_{jt}^c \qquad \forall i, j, t, c$$

$$\sum_{i=1}^{M} Z_{ijt} = 1 \qquad \forall j, t$$

$$X_{ijt}, W_{ijt}, \Gamma_{i\gamma t}, I_{ijt}, R_{ijt}, Z_{ijt} \geq 0 \qquad \forall i, j, t, \gamma$$

An advantage of this approach is that the marginal cost of capacity and the marginal benefit of adding WIP are strictly positive, since the constraints are always active as opposed to classical models where the capacity constraint is only active at bottleneck resources that are 100% utilized. For the purpose of examining the relevance and the performance of the mathematical approach, Asmundsson et al. (2002) consider an example of a simple single stage system with three products which gives very good results. Additionally, they analyze the sensitivity of the estimated clearing function to diverse shop floor scheduling algorithms, different demand patterns and techniques of production planning using a simulation model of the respective production system. They find that the clearing function model reflects the characteristics and capabilities of the production system better than models using fixed planned lead times (like mrp) and derives realistic and robust plans with better on time delivery performance, lower WIP and system inventory. In addition, the model captures the effects of congestion on lead times and WIP and therefore the fundamental trade-off between anticipated production to account for possible peaks in demand and just in time production to avoid higher costs due

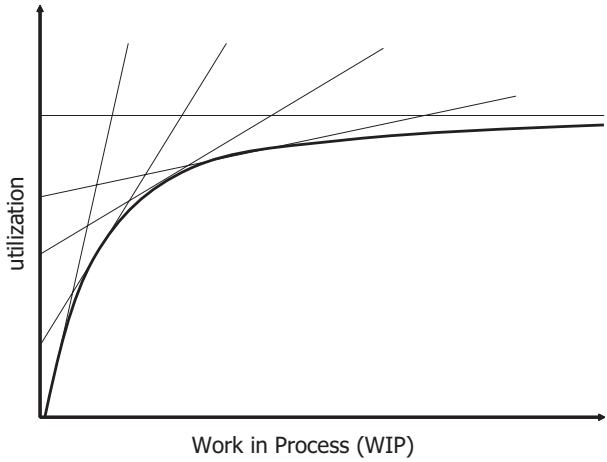

Fig. 4. Linearization of the clearing function

to preventable FGI. Releases generated by this partial clearing function model are smoother and lead to enhanced lead time performance. Further analysis of different aspects of this mathematical programming model is needed, e.g., the extension of the approach to systems of multiple production stages.

Besides this, interactions between clearing functions and shop floor execution systems such as the dependency of LDLT on the various priority rules need further analysis. This is another circularity, because clearing functions are dependent on the employed scheduling policy and therefore on the result of the scheduling algorithm. Moreover, as pointed out by Asmundsson et al. (2002), the schedules are dependent on the release schedule and consequently on the planning algorithm.

3.3.4 Clearing function model of Hwang and Uzsoy

Hwang and Uzsoy (2004) combine the work of Karmarkar (1987) and Asmundsson et al. (2002) and add lot sizing in order to show how small or large lot sizes influence the resulting production plans. For that purpose they present a single-product dynamic lot sizing model which takes into account WIP and congestion using queuing models such as those presented by Karmarkar (1987, 1993b). They develop a clearing function which captures the dependency of the expected throughput of a single-stage production system closely related to the classical Wagner-Whitin model (see Wagner and Whitin 1958) including setups, expected WIP levels and lot sizes which is then integrated into a dynamic lot sizing model. Results demonstrate that their proposed model provides significantly more realistic performance and, therefore, production plans, than models ignoring the relationships between lead times, workload, production mix (setups), throughput and lot sizes. The importance of lot sizing is emphasized especially in the context of capacity planning and pro-

duction coordination, because different product types compete for scarce resources and require significant setup times (see Hwang and Uzsoy 2004 and the references therein).

In general, large lot sizes allow for high resource utilization due to the reduced number of setups, but result in increased FGI and may lead to delays of important jobs waiting to be processed. However, small batch sizes minimize WIP levels, but necessitate frequent setups impeding the usage of resources for production, creating congestion and high WIP levels in front of the resource resulting in long lead times. So, lead times depend not only on the workload of the system, but also on the work assignments to resources which, in turn, are determined by the planning models and, therefore, representing a further planning circularity which must be accounted for. Given a WIP level, their model determines the lot size which maximizes system throughput.

The clearing function aims at minimizing the sum of release costs, production costs, WIP and FGI holding costs over the planning horizon. It is derived with the use of a $M|G|1$ queuing model (see Buzacott and Shantikumar 1993), assuming a Poisson process as arrival rate of the lots and a general distribution of the processing times at the resource following a *first come first serve* (FCFS) process order for the products (jobs). The so-called *lot sizing clearing function* (LCF) which is derived using Little's Law $X = \frac{W}{L}$ (expected throughput rate equals the expected WIP over the expected lead time) is stated as follows:

$$f(Q, W) = \frac{\left(Q + W - \sqrt{Q^2 + W^2 + 2QWc^2}\right)}{\left(1 - c^2\right)\widehat{p}}$$

where Q denotes the lot size and $\widehat{p}$ the expected processing time of a batch. The coefficient of variation of the system processing time is given by c (for more details on the derivation of the LCF see Hwang and Uzsoy 2004). The LCF is integrated in a single-stage manufacturing system where the decision variables are the release quantity denoted by R_t, the lot size Q_t, the production quantity Y_t ("number of lots produced in period t") and the level of WIP denoted by W_t and FGI denoted by I_t. No backorders are allowed and the variables are assumed to be continuous. The objective function does not contain setup costs, because these are captured in the LCF by showing directly their influencing effects on capacity. The complete dynamic lot sizing model is stated as follows:

$$\min \sum_{t=1}^{T} [\phi_t Q_t Y_t + \omega_t W_t + \pi_t I_t + r_t R_t]$$

subject to:

$$W_t = W_{t-1} + R_t - Q_t Y_t \quad \forall t$$
$$I_t = I_{t-1} + Q_t Y_t - D_t \quad \forall t$$

$$
\begin{aligned}
&Q_t Y_t \leq f(Q_t, W_t) \quad \forall t \\
&Q_t > \frac{S_t D_t}{(1 - \frac{D_t}{P})} \quad \forall t \\
&R_t, Q_t, Y_t, W_t, I_t \geq 0
\end{aligned}
$$

The first constraint represents the WIP balance, the second the FGI balance (with the demand in period t, D_t), and the third reveals the LCF. The fourth constraint represents a lower bound on the lot size in order to ensure system stability by bounding utilization below 1 where S_t denotes the setup time per lot and P gives the production rate of the manufacturing system. The model is reformulated by setting $Q_t Y_t = W_{t-1} - W_t + R_t$:

$$
\min \sum_{t=1}^{T} \left[\phi_{t-1} W_{t-1} + (\omega_t - \phi_t) W_t + \pi_t I_t + (r_t + \phi_t) R_t \right]
$$

subject to:

$$
\begin{aligned}
&I_t + W_t = I_{t-1} + W_{t-1} + R_t - D_t \quad \forall t \\
&W_{t-1} - W_t + R_t \leq f(Q_t, W_t) \quad \forall t \\
&Q_t > \frac{S_t D_t}{(1 - \frac{D_t}{P})} \quad \forall t \\
&R_t, Q_t, Y_t, W_t, I_t \geq 0 \quad \forall t
\end{aligned}
$$

This has the advantage that the Y_t are eliminated from the objective function. Furthermore, the second constraint now combines the WIP and the FGI constraint which reduces the number of constraints.

In computing the model they examine several properties of the clearing function and find that, for a given WIP level, there exists a unique lot size which balances the system throughput and LDLT. Future research concentrates on the extension of the model to multiple products and to investigate how capacity is distributed among them when they are subject to different demand, manufacturing requirements and variability.

3.4 The production planning and order release concept of Missbauer

Missbauer (1998) emphasizes the need to create a structure for production planning and scheduling in which all partial problems are included, e.g., aggregate production planning, capacity planning, lot sizing, scheduling etc., but which does not imply the problems and complexities which arise from comprehensive models such as, e.g., problems of data procurement, limited computational storage space, too long CPU times for calculation etc. Therefore, he concentrates on the *hierarchical production planning concept* where all stages or areas of the production planning and

scheduling process are included and can be integrated in traditional production planning (mrp) systems, but nevertheless leaves room for decisions on implementations of different planning approaches like KANBAN, OPT or priorities on the various planning stages or areas. The focus lies on a multi-stage, multi-product production system which is the underlying concept in most production planning and scheduling systems prevailing in practice. The planning process is executed in a decentralized manner with two planning levels: one central, over arching planning unit, called "production unit control" as well as decentralized production units, called "goods flow control" by Bertrand and Wijngaard (1986).

The production control unit defines the sequence of jobs and their due dates and the moment at which the material flow control is passed on the decentralized goods flow units (order release process). In order to supervise and regulate the actual production situation at the operating level of the production process, measures like due date fulfillment, lead times, WIP, utilization levels etc. are reported from the goods flow control to the central production unit control. The organizational type of the goods flow control, i.e., if it is an internal department or a supplier, is of secondary importance so that the concept is also applicable and extendable to supply chain networks. Therefore, in order to control the production system status of the goods flow units it is necessary for the production unit control to rely on information about the characteristics and system dynamics inside and between the decentralized goods flow units and also on information about the functional dynamics of the control measures like, e.g., the relationship between lead times, WIP and capacity. These relationships and dynamics specify the system states that can be reached. Furthermore, they are easily depicted graphically and used as a basis for models trying to anticipate future system states from current states. Nevertheless, it is even more important to understand the structure and order of decision events at the central planning stage, the production unit control.

Missbauer (1998) distinguishes four different hierarchical planning levels, namely the aggregate production planning level, the master production schedule, the planning of the material flow between the decentralized goods flow units (determination of lot sizes, due dates, safety inventory and buffer times) and the order release control level. In order to achieve coordination, the models of the upper planning stages take into account the dynamics of the subordinate planning models. Moreover, a capacity oriented planning approach is used which, as opposed to the product oriented approach, does not plan on the basis of the specific product parts, but considers aggregate product units as capacity units. This is because in the long run of the planning horizon, i.e., after a completed production cycle, the stored capacity is relevant for the planning process whereas it is less important which and how many specific parts are at stock. This becomes important in the short horizon, detailed production planning stage. The capacity oriented planning approach divides the order release process into two decision problems: assuming a certain value of the aggregate demand on end products (expressed in capacity units), the first decision problem concerns the time to release jobs (expressed in quantity

of working hours per period or capacity units) in order to achieve and maintain a certain system state including further decisions about the system performance as well as aggregate FGI inventories. This determines the budget of capacity (expressed in working hours). The second decision problem is about choosing specific orders and their concrete release times based on product oriented criteria. In doing so the capacity oriented approach is very useful, because it eliminates the difficulties arising from uncertain demand estimates for specific products underlying the product oriented approach. It only needs information about demand estimates aggregated across all products and their capacity utilization in order to plan capacity utilization in the long run. The detailed scheduling for specific products takes place on subsequent production planning stages which not only reduces the information requirements but also the complexity of the decision models.

Missbauer (1998) further distinguishes between models of aggregate production (and release) planning with constant or variable planned workload (non-stationary system behavior). Because we are interested in considerations of LDLT in the aggregate production planning process and the effects on them caused by demand peaks or machine breakdowns, we will concentrate on models with variable (planned) workload. The model of Missbauer (1998) presented here is based on a job shop production system which depicts the material flow as a continuous or fluid flow. The planning problem is further divided into two partial problems where first the material flow between the production unit and the goods flow units is modeled and secondly the flow of jobs (work) between the single goods flow units.

The partial formulation of the problem of the material flow between the decentralized production units mainly consists of the inventory equation which defines the WIP inventory level at the production units as well as the FGI inventory level. Furthermore, the dependencies of these inventory levels on the demand quantity of products, on the performance of the system in one period and on the order release quantities are modeled. The basic assumptions are that the production system consists of N production units ($j = 1, \ldots, N$; index v denotes the preceding production units) with M products to be manufactured ($i = 1, \ldots, M$). Moreover, each product i contains $K(i)$ operations (production processes with $k = 1, \ldots, K(i)$) where the technical machine sequence as well as the jobs and completion due dates for each product are given. In addition, the demand for product i in period t is measured in work hours and denoted by D_{it}. Jobs are arbitrarily divisible and setup times proportional so that any job of any product can be processed at any time. This is a critical assumption, because it does not reflect reality and leads to impractical anticipated system states. We will refer to that later on.

In front of each production unit there is a stock of WIP which is measured in work hours or capacity units. The WIP inventory equation has the following form:

$$W_{ijt} = W_{ij,t-1} + \sum_{v=1}^{N} \sum_{k=2}^{K(i)} \delta_{ijk} \delta_{iv,k-1} X_{ivt} \frac{p_{ij}}{p_{iv}} + R_{it} \delta_{ij1} \frac{p_{ij}}{p_i} - X_{ijt} \qquad \forall i, j, t$$

where W_{ijt} denotes the WIP of product i at production unit j at the end of period t and δ_{ijk} is a Boolean variable which is 1 if operation k of product i is executed at production unit j and 0 otherwise. $\delta_{iv,k-1}$ is a Boolean variable denoting if a preceding operation $k-1$ has been executed at production unit v. By X_{ijt} we can denote the throughput of item i in workstation j during period t. With this X_{ivt} may represent the performance of the preceding production unit v in period t attributable to product i. p_{ij} denotes the processing time of a job of product i at production unit j and p_i the aggregation across all production units j. Additionally, R_{it} is the order release quantity of product i in period t. The second term of the equation reflects an increase of WIP from all predecessors v. The third term represents an increase of WIP at the production system starting point or after order release and the fourth one is the work performed on product i which can be considered as the cleared WIP of product i at production unit j. The FGI inventory equation has the following form:

$$I_{it} = I_{i,t-1} + \sum_{j=1}^{N} \delta_{ij,K(i)} X_{ijt} \frac{p_{ij}}{p_i} - D_{it} \qquad \forall i, t$$

where I_{it} (FGI inventory) consists of the FGI inventory of the previous period, the increase added in period t (measured at the end of period t where $\delta_{ij,K(i)}$ denotes the last operation of product i if $K(i)$ is executed on production unit j) reduced by the demand for product i in period t. This formulation specifies that the increase from the previous production units in period t only depends on the performance of these production units in this period and consequently does not assume a production rhythm. Instead, it only supposes a movement at the transition times from one period to another, which implies that release jobs could pass an infinite number of production units and which is similar to overlapping production with infinite small transfer lot sizes. It means that the delay caused by processing times of jobs is not yet modeled. Missbauer (1998) states that this feature is easily added.

It is important to note that the model overestimates the system performance due to its underlying assumptions, especially when direct inventory levels are small. This is due to the assumption that jobs are infinitely divisible and setups are independent. So system performance quantities X_{ijt} depend on the progress of production and thus on the course of inventory levels or system workload on time, assuming given order release, and consequently on the utilization function or clearing function. The determination of the clearing function is part of the formulation of the second partial problem. The second partial problem highlights the relationship between the performance quantities X_{ijt} and the clearing function. Moreover, in order to achieve good production system state predictions and therefore realistic system performance, the influence of the discontinuities of material workflow due to non-divisible jobs needs to be reflected, assuming the estimated arriving work or WIP as the maximal performance in one period which can be distributed to specific products. The clearing function is given as $f(\widehat{W}_{ijt})$ where the term $\widehat{W}_{ijt}$ denotes the estimated inflow of products i at production unit j in period t and is comprised

of the estimated WIP level of products i at the beginning of period t ($\widehat{W}_{ij,t-1}$) and the estimated increase ($\widehat{X}_{ijt}$), so that $\widehat{W}_{ijt}$ is given by

$$\widehat{W}_{ijt} = \widehat{W}_{ij,t-1} + \widehat{X}_{ijt}$$

An $M|M|1$ model is used to derive the form of the clearing function. In the steady state the average performance in one period t equals the average increase of work ρ which is the degree of utilization ("traffic intensity") of one production unit, so that $\widehat{W}_{ijt} = \rho_j \tau_t$ where τ_t denotes the length of period t. The average utilization or workload $\overline{W}_{ijt}$ is the sum of the average inventory represented by $\overline{I}_{ijt} = \overline{I} = \frac{\rho}{1-\rho}$ and of the average performance $\overline{X}_{ijt}$ with $\overline{I}$ denoting the average number of jobs in the queue:

$$\overline{W}_{ijt} = \frac{\rho}{1-\rho} + \overline{X}_{ijt}$$

The clearing function (in disaggregated form) then gets the following form:

$$f(\widehat{W}_{ijt}) = \frac{1}{2}\left(1 + \widehat{W}_{ijt} + \tau_t - \sqrt{1 + 2\widehat{W}_{ijt} + \left(\widehat{W}_{ijt}\right)^2 + 2\tau_t - 2\widehat{W}_{ijt}\tau_t + \tau_t^2}\right)$$

The performance in one period t can be represented as a function of the utilization or workload in one period by generalizing the above stated formulation for which the statement $X_{ijt} = \beta_j \overline{W}_{ijt} \;/\; \left(\overline{W}_{ijt} + \alpha_j\right)$ is valid with α_j giving the slope of the utilization function and β_j being a parameter to be estimated. We define the utilization or workload for the stationary case as $W_{ijt} = \overline{W}_{ijt} + X_{ijt}$:

$$\begin{aligned} &f(W_{ijt}) \\ &= \frac{1}{2}\left[\beta_j + \alpha_j + W_{ijt} - \sqrt{\beta_j^2 + 2\beta_j\alpha_j + \alpha_j^2 - 2\beta_j W_{ijt} + 2\alpha_j W_{ijt} + \left(W_{ijt}\right)^2}\right] \end{aligned}$$

where the parameters β_j and α_j need to be estimated. It should be mentioned that the functional relationship between the performance in one period and the utilization of one period further depends on the composition of the expected workload, i.e., new work arrivals and workload that remain in the production system for a longer time. The clearing function makes implicit assumptions on this relationship that may not be valid in case of non-stationary systems. Therefore, this procedure is only an approximation. The entire disaggregated model now has the following form:

$$\min \sum_{i=1}^{M}\sum_{t=1}^{T}\left(\sum_{j=1}^{N} \omega_{ijt} W_{ijt} + \pi_{it} I_{it}\right)$$

subject to:

$$W_{ijt} = W_{ij,t-1} + \sum_{v=1}^{N}\sum_{k=2}^{K(i)} \delta_{ijk}\delta_{iv,k-1} X_{ivt}\frac{p_{ij}}{p_{iv}} + R_{it}\delta_{ij1}\frac{p_{ij}}{p_i} - X_{ijt} \qquad \forall i, j, t$$

$$I_{it} = I_{i,t-1} + \sum_{j=1}^{N} \delta_{ij,K(i)} X_{ijt} \frac{p_{ij}}{p_i} - D_{it} \qquad \forall i, t$$

$$\sum_{i=1}^{M} X_{ijt} \leq \alpha_j^c + \beta_j^c \sum_{i=1}^{M} (W_{ijt} + X_{ijt}) \qquad \forall t, j, c$$

$$X_{ijt}, I_{it}, R_{it}, W_{ijt} \geq 0$$

The objective function specifies minimization of the costs for inventory of WIP and FGI. The clearing function is given by the third restriction with α_j^c and β_j^c representing the position of the clearing function and the c's denoting the degree of linearization. As a result, the model determines the flow of jobs (materials) through the production system and, consequently, the order release R_{it}. Furthermore, it decides which jobs of the waiting W_{ijt} are to be processed next and therefore determines the production schedule for each workstation j. The order release priority rule is not explicitly taken into account, but would be determined implicitly by the objective function where adherence to due dates (delivery dates) would be the restriction.

In order to reduce the size of the optimization problem it is possible to aggregate over products and capacity which means that products with a similar machine (processing) sequence and similar processing times p_{ij} can be grouped in product families and production units can be aggregated with the requirement that the dynamic behavior is not distorted. For more details see Missbauer (1998).

It is important to note that the question how far to aggregate the products and which products to group is very difficult and widely discussed in the literature without an unambiguous answer. On the one hand aggregation leads to reduction of complexity and model size, but on the other, it can impede the achievement of optimal results. Besides this, the time delays caused by the transition from production units j to N are defined for product i, resulting in a time delay function.

Additionally, the problem of the delay caused by the fact that jobs are not arbitrarily divisible is integrated in the model in order to arrive at the final formulation which is a network model of bottlenecks. The advantages of the final model are that the clearing function and the delay function between the production units give a good description of the characteristics of real production using parameters which are directly observable and therefore empirically determinable. Nevertheless, the clearing function used here is derived from characteristics of the stationary system case. Missbauer (1998) reveals that an analysis of non-stationary behavior of queuing systems and their integration in the models of the production units can be an improvement and, consequently, represents an objective for further research. Moreover, it would be interesting to know if the use of nonlinear programming could be a better representation of the system behavior, especially for considering the relationship between lead times and workload. Besides this, the representation of the production system structure as one bottleneck with the other production units being modeled as delaying functions is interesting, because it would lead to a

great reduction of complexity concentrating on only a small number of production units and aggregated products. This is especially interesting in production planning where multiple production units and stages have to be considered simultaneously.

3.5 The planning model with forward shifting preparation times by Lautenschläger

Lautenschläger (1999) integrates workload dependent forward shifting production times in a multi-level capacity oriented production planning model with and without lot sizing. In contrast to approaches in the literature which use constant minimum forward shifting production times he considers these times to be dependent on the utilization of resources (workstations or production units) in a capacity oriented production planning model. Therefore, he uses a functional relationship that specifies the portion of the production volume that is to be shifted forward in time due to (and dependent on) resource utilization. Extensions of the basic model contain linear as well as nonlinear forward shifting functions, the integration of overtime or complete preliminary production and effects of shifting production before the beginning of the planning horizon or planning period. Finally, procedures for smoothing resource utilization are proposed that use ideas from digital control engineering and a filter accounting for utilization levels of resources.

Simulation experiments demonstrate that part of the production volume is shifted forward in time due to the utilization level assumed and varied in the planning model. He also shows that employing the same forward shifting production function and varying only the utilization levels of the resources leads to increases in parts of the production volume that is shifted forward. Moreover, integrating forward shifting times reduces the occurrence of backorders independent of production system design which is also due to smoothing utilization levels in the production planning process as well as to the coordination of production segments and finally leads to a reduction of lead times and inventory without an increase of backorders. The basic model is a *multi-item, multi-level capacitated lot sizing problem* (MLCLSP) which includes detailed lot sizing and resource deployment planning assuming the data to be deterministic within a finite planning horizon which is divided into T sub-periods of equal length. Furthermore, demand may vary from period to period, but backorders are not allowed. Inventory costs are proportional to the inventory stock at the end of a period and, because production and material costs per unit are assumed to be constant, they are not relevant for the planning decision. There are K operations represented in the model with multi-level product structure where every operation k is assigned to a resource and is only carried out once in a period causing setup costs and times. There is no restriction concerning the number of operations executed in one period, but the production system consists of N different resources ($j = 1, \ldots, N$) subject to capacity restrictions which can be extended by overtime.

The objective is the minimization of setup, inventory and overtime costs with the complete model as follows:

$$\min \sum_{k=1}^{K} \sum_{t=1}^{T} [\pi_k \cdot I_{kt} + \zeta_k \cdot S_{kt}] + \sum_{j=1}^{N} \sum_{t=1}^{T} o_j \cdot O_{jt}$$

subject to:

$$\begin{aligned}
& I_{k,t-1} + X_{k,t-l_k} = PD_{kt} + \sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t} + I_{kt} \qquad \forall k, j, t \\
& \sum_{k=1}^{K} (\xi_{jk} \cdot X_{kt} + \zeta_{jk} \cdot S_{kt}) \leq \kappa_{jt} + O_{jt} \qquad \forall j, t \\
& X_{kt} \leq TOP \cdot S_{kt} \qquad \forall k, t \\
& I_{kt}, O_{jt}, X_{kt} \geq 0 \qquad \forall k, j, t \\
& S_{kt} \in \{0, 1\} \qquad \forall k, t
\end{aligned}$$

where the objective function has three terms representing the inventory costs, the setup costs and the overtime costs and the index k standing for one operation, j denoting a resource (or workstation or production unit, respectively) and t representing the period. The first constraint is the inventory balance equation stating that the sum of inventory stock of the previous period ($I_{k,t-1}$) and the production volume in period t, reduced by the constant minimal preliminary shift of operation k ($X_{k,t-l_k}$) must equal the sum of primary part demand (PD_{kt}), the secondary part demand ($\sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t}$ with $d_{j\hat{k}}^{(k)}$ denoting the direct requirement coefficient from the preceding to the subsequent operation) and inventory holding I_{kt} at the end of period t. This means that the processing of operation k must be initialized not later than l_k and finished before use of resource j. The constant minimal preliminary shift of operation k contains times for transfer, drying, etc., but does not take into account waiting times in front of resources due to occupation by other operations. The second restriction represents the capacity constraint stating that the sum of resource utilization for production ($\sum_{k=1}^{K} \xi_{jk} \cdot X_{kt}$) and the resource utilization for setups ($\sum_{k=1}^{K} \zeta_{jk} \cdot S_{kt}$) must be smaller or equal to the sum of capacity and overtime of resource j in period t ($\kappa_{jt} + O_{jt}$). The third constraint denotes the linkage of the production volume to the setup variables in order to take account of the fact that production only takes place after a setup (with TOP denoting a big number). The remaining constraints are the usual non-negativity restrictions and the definition of S_{kt} as a Boolean variable where $S_{kt} = 1$ indicates a setup for operation k in period t.

In order to balance the preliminary shifts with the smallest possible lead times for tactical production planning it is advisable to determine preliminary shifts dynamically and dependent on the production program. As a result, production process dependent waiting times are included in preliminary shifting times. Moreover, they need not be set for production planning, but result from the plan, so that one main source of inaccuracy (planning circularity) of production planning, i.e., the use of predetermined and estimated planned lead times in mrp systems, is eliminated. Consequently, the model is modified accordingly by redefining the variable for the production volume: Two modes for the production volume are defined which distinguish between production and availability of the product. That is, mode 1 represents an operation k which is produced as well as available in period t and mode 2 denotes an operation k produced in period $t-1$ and available in period t. So the selection of the production mode determines whether there has to be one period in between two operations or not. The production volume produced in mode m is then denoted as X_{kt}^{m} and the total production volume is defined as

$$X_{kt}^{\text{total}} = X_{kt}^{1} + X_{k,t+1}^{2} \qquad \forall k, t$$

which is the sum of the production volume produced in period t for period t (mode 1) and the production volume produced in period t for period $t+1$ (mode 2). The position of the forward shifting production function is defined by two parameters b_j and e_j which are related to the part of production which had to be shifted forward in the previous period. Here b_j states the utilization level from which shifting production forward becomes necessary and e_j denotes the part of production volume which needs to be shifted forward at a utilization level of 100%. The production volume that can be produced in mode 1 depends on the capacity not used for production, namely $\kappa_{jt} = \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}}$. This constraint is adapted to the forward shifting production function by means of the parameters α_j and β_j which depend on b_j and e_j and modeled for each single resource:

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{1} \leq \alpha_j \kappa_{jt} - \beta_j \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \qquad \forall j, t$$

The parameters α_j and β_j have to be determined in order to derive the forward shifting production function for each resource. They depend on the before mentioned parameters b_j and e_j as follows (for more details see Lautenschläger 1999): $\alpha_j = \frac{e_j b_j}{1-b_j}$ and $\beta_j = \frac{e_j + b_j - 1}{1-b_j}$. Subsequently, these extensions are integrated in the MLCLSP by inserting the new function and by modifying the inventory balance equation:

$$\min \sum_{k=1}^{K} \left[T \pi_k I_{k0} + \sum_{t=1}^{T} \left((T-t+1) \cdot (l_k X_{kt}^{\text{total}} - \pi_k P D_{kt}) + \zeta_k S_{kt} \right) \right]$$

subject to:

$$X_{kt}^{\text{total}} = \sum_{m=1}^{Mod} X_{kt}^{m} \qquad \forall k, t$$

$$I_{k,t-1} + \sum_{m=1}^{Mod} X_{kt}^{m} = PD_{kt} + \sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t}^{\text{total}} + I_{kt} \qquad \forall k, t$$

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \leq \kappa_{jt} \qquad \forall j, t$$

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{1} \leq \alpha_j \kappa_{jt} - \beta_j \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \qquad \forall j, t$$

$$X_{kt}^{\text{total}} \leq TOP \cdot S_{kt} \qquad \forall k, t$$

$$I_{kt}, X_{kt}^{m}, X_{kt}^{\text{total}} \geq 0 \qquad \forall k, t, m$$

$$S_{kt} \in \{0, 1\} \qquad \forall k, t$$

The objective function now aims at minimizing the sum of inventory and setup costs including l_k which is the echelon inventory cost coefficient for the increase of inventory due to operation k. Overtime is not modeled here. Furthermore, in this model the forward shifting production function is supposed to be a linear constraint for each resource.

The nonlinear relationship between the forward shifting volume and capacity utilization can perhaps be modeled by a linear approximation similar to the method described by Asmundsson et al. (2002) or else by the description of a nonlinear constraint. Finally, in order to smooth variation of utilization rates of resources to account for the relationship between utilization and lead times, a controller from digital engineering is combined with the forward shifting production function. For further reading see Lautenschläger and Stadtler (1998); Lautenschläger (1999).

3.6 The supply chain coordination-model by Caramanis and Anli

Caramanis and Anli (1999b) examine the potential value of integrating variable lead time relationship information and information sharing between plant level and cell decision nodes in the supply chain coordination process of a multiple part production facility which is expected to consist of decreased inventory and backlog costs. For that purpose they propose a master problem algorithm for vertical supply chain coordination of planning and operational decision making for manufacturing cells with concave as well as convex (nonlinear) lead time relationships. Horizontal coordination constraints like safety stocks and KANBAN levels are not considered. Furthermore, they evaluate the problem of generating the required information through calculation of the sub problems at each production cell modeled

as small networks of queues which are communicated to a master problem solving coordinator and they examine the numerical problems associated with system complexity. The objective of Caramanis and Anli (1999b) is to determine optimal and consistent (achievable) production targets (plans) for each production cell during a time period aiming at reducing WIP, FGI and backlog costs across the supply chain for a given demand forecast and to compare the benefits of employing higher complexity issues to classical methods of supply chain coordination. Plan consistency refers to adherence to horizontal operational production constraints such as capacity, lead time and safety stocks.

A production planning model for an assembly production system is defined as consisting of three production cells with the objective to minimize the WIP and FGI (surplus or backlog) costs over the planning horizon. The relevant optimization problem takes the following form:

$$\min \sum_{i=1}^{M} \sum_{c=1}^{C} \sum_{t=1}^{T} \left[\omega_{it} W_{ict} + \pi_{it} \hat{I}_{ict} + \vartheta_{it} B_{ict} \right]$$

where $\omega_{it} W_{ict}$ are the costs of WIP of part type i in cell c at the end of period t with ω_{it} representing the cost factor and with $W_{ict} = \sum_{j=1}^{N} q_{icjt}$, i.e., the WIP of part type i in cell c at the end of period t is equal to the sum of all buffers q_{ijct} across all workstations $j = 1, 2, ..., N$. Based on the inventory positions I_{ict}, $\pi_{it} \hat{I}_{ict}$ denotes the costs of the positive FGI at the end of period t with $\hat{I}_{ict} = \max\{0, I_{ict}\}$ and $\vartheta_{it} B_{ict}$ represents the costs of the backlogged demand at the end of period t with $B_{ict} = \max\{0, -I_{ict}\}$. The objective function is restricted by the following constraints enforced for $t \in \{1, ..., T\}$:

$$\begin{aligned}
&W_{ict} = W_{ic,t-1} + R_{ict} - X_{ict} \qquad \forall i, c, t \\
&I_{ict} = I_{ic,t-1} + X_{ict} - R_{i3t} \qquad \forall i, t, \qquad c = 1, 2 \\
&I_{ict} = I_{ic,t-1} + X_{ict} - D_{it} \qquad \forall i, t, \qquad c = 3 \\
&\sum_{i=1}^{M} \xi_{icjt} X_{ict} \leq \rho_{cjt} \qquad \forall c, j, t \\
&\sum_{t=1}^{t^*} R_{ict} + \dot{W}_{ic,t+1} \geq \sum_{t=1}^{t^*} X_{ict} + \sum_{t=t^*+1}^{T} X_{ict} 1\left\{t - L_{ict} \leq t^*\right\} \qquad \forall i, c \\
&\overline{W}_{ict} = \alpha W_{ic,t-1} + (1-\alpha) W_{ict} \\
&W_{ict}, X_{ict}, \dot{W}_{ict}, \overline{W}_{ict}, R_{ict} \geq 0 \qquad \forall i, c, t \\
&I_{ict} \geq 0 \qquad \forall i, t, \qquad c = 1, 2 \\
&I_{ict} \quad \text{unrestricted in sign } c = 3
\end{aligned}$$

where the first three constraints denote the material flow relations with R_{ict} being the release rate of part type i during period t into cell c and X_{ict} stating the production

target of part i by cell c during period t. The fourth restriction is the capacity constraint where the sum of the expected time needed by machine j at cell c to produce one part of type i during time period t multiplied by the production target of part i by cell c during period t ($\sum_{i=1}^{M} \xi_{icjt} X_{ict}$) is forced to be smaller or equal to the maximum allowed utilization of machine j in cell c during time period t. The fifth restriction represents the lead time constraint where $\dot{W}_{ict}$ is the WIP of part type i in cell c at the end of period t_0, hence available at the beginning of period t_1. An indicator function is represented by $1\{\cdot\}$ which equals 1 if $\{\cdot\}$ is true and 0 otherwise. Furthermore, an equality constraint is given by the sixth restriction which denotes the weighted average WIP of part i in cell c during period t, which is given by $0 \leq \alpha \leq 1$. The last two constraints are the usual non-negativity restrictions. The system is modeled as a Markovian closed queuing network with exponential processing times, variable probabilistic routing, multiple part types and multiple machines (resources). Lead time estimates for each production cell are determined by *Mean Value Analysis* (MVA; see, e.g., Suri and Hildenbrant (1984)) and, assuming that the production period is long enough to lead to a steady state, rewritten by employing Little's Law:

$$X_{ict} \leq \frac{\overline{W}_{ict}}{L_{ict}}$$

Since L_{ict} is a function of all lead time requirements the constraint is rewritten as follows:

$$X_{ict} \leq g_{ic}(\overline{W}_{ict}, X_{ict}, X_{i^*ct}, \forall i^* \neq i)$$

where g_{ic} is a nonlinear function which represents the lead time in the production system for each production cell modeled as a function of the specific workload of a cell in a considered time period. In order to analyze algorithms which are able to coordinate effectively the supply chain under complicating factors, Caramanis and Anli (1999b) concentrate on concave as well as convex lead time constraints.

In the case of concave lead time restrictions linearization is achieved by decentralized calculations at appropriately selected points of the production system generated by the iterative generalized Benders' master problem algorithm which handles the information exchange transactions between the sub problem decision making including information feedback determination and the master problem coordinator. In the case of convex lead time restrictions a modified generalized Benders' master algorithm is used. The iterative algorithm generates tentative sets of production targets for each production cell and then examines their feasibility assuming the feasibility of the targets of all other cells. Subsequently, the production targets are converted in the planning horizon to a point on the lead time constraint surface by adjusting W_{ict} or X_{ict} as implied by the master problem iteration and returning the linearized constraint with the following form:

$$X_{ict} \leq \widetilde{g} + \frac{\delta g}{\delta W_{ict}}(W_{ict} - \widetilde{W}^n_{ict}) + \sum_{i^* \neq i} \frac{\delta g}{\delta X_{i^*ct}}(X_{i^*ct} - \widetilde{X}^n_{i^*ct})$$

with $\widetilde{(\cdot)}$ indicating initialization values. A feasible starting point on the surface of g_{ict} is obtained by using the tentative solution of iteration n:

$$\left[\widetilde{W}^n_{ict}, \widetilde{X}^n_{ict}, \widetilde{X}^n_{i^*ct}, \forall i^* \neq i\right]$$

Since the cell sub problems use stochastic dynamic models with finer time scales than the master problem, a sensitivity analysis is performed and time averaged in order to be consistent with the master problem. Learning effects about the nonlinear lead time constraints are obtained after each iteration possibly leading to an optimal and consistent production plan. As a result the complex issue of nonlinear lead time constraints is solved using linear programming which is similar to the solution of Asmundsson et al. (2003).

Caramanis and Anli (1999b) analyze three different types of lead time constraints, namely an unambiguous concave g_{ict}-function at all cells c and for all part types i, a medium non-concave function, and a severely non-concave function due to FCFS queue policy and significant different part type processing times with the result that the implemented algorithm leads to an effective coordination even with the existence of complicating factors like severely non-concave lead time constraints. Furthermore, in order to deal with combinatorial problems arising with production system size and therefore complexity issues of sub problems at the cell level, ongoing advances with controlled queuing networks in combination with concurrent simulation and neurodynamic programming can be further research directions.

3.7 The production planning model with transient Little's Law by Mendoza

In his thesis Mendoza (2003) develops a transient Little's Law which is the time-dependent analogue of Little's Law and denotes the expected queue length in a general input/output model in terms of *transient Palm probabilities* (see, e.g., Baccelli and Bremaud 1987) of the waiting time and of the expected value of the input process. He uses the transient Little's Law to create an *integro-differential equation* for the expected WIP and develops an approximation method to solve it numerically. An optimization based production planning model is presented which establishes the optimal release of the material into the system in order to closely match the defined requirements, e.g., capacity constraints for a finite planning horizon. In doing so he can approximate the transient behavior of general multi-class queuing networks such as supply chains including random lead times and time-dependent WIP. Randomness can be due to material entering the production system and coming out in several periods according to some weights $w_i(s, t)$ which indicate the proportion of input of product i during period s that leaves the production system by period t. These weights implicitly contain information about the lead time distributions of the production system, i.e., they are a function of the system load with two properties: First, they represent how the lead times depend on the load of the system and, second, how the lead times are affected by the randomness inherent

in the production system. The numerical method presented for transient analysis gives an approximation solution to the expected value of the WIP in the system as a function of time, calculating the weights for the lead time distribution. In case of known weights the nonlinear optimization problem is reduced to an LP solvable with standard LP methods. Here we describe the basic production planning model of Mendoza (2003) in order to show how he integrates LDLT in production planning.

The model is defined as a production network for producing several products. The model has the objective of determining an optimal release point for raw material into the system in order to meet some predefined requirements which are assumed to be deterministic (this is relaxed later). The products move through the system according to specific routes which may also be random, and feedback (rerouting) is allowed. The input of the production system $\widehat{X}_{qt}$ (the cumulative expected quantity of items that enter the buffer q at time t ($\widehat{X} = \left\{\widehat{X}_{qt} : t > 0, q \in B\right\}$) is transformed into the output ($X = \left\{X_{qt} : t > 0, q \in B\right\}$) through a function f stated as follows:

$$X = f_{qt}(\widehat{X}, W)$$

with W denoting the load on all buffers in the system and f denoting a function describing how the items are delayed due to queuing and processing times. The complete production model is stated as follows:

$$\min \sum_{t=1}^{T} \sum_{q \in O(q)} \left[\phi_{qt} X_{qt} + \vartheta_{qt} B_{qt}\right] + \sum_{t=1}^{T} \sum_{q \in B} \omega_{qt} W_{qt}$$

subject to:

$$\begin{aligned}
&X_{qt} = f_{qt}(\widehat{X}, W) && q \in B, \forall t \\
&Z_{qut} = \alpha_{qu} \widehat{X}_{ut} && q \in \Im(u), u \in B \setminus \Im, \forall t \\
&Z_{qut} = \beta_{qu} \widehat{X}_{qt} && u \in O(q), q \in B \setminus O, \forall t \\
&\widehat{X}_{qt} = W_{qt} + X_{qt} && q \in B, \forall t \\
&\widehat{X}, X, Z \geq 0
\end{aligned}$$

with ϕ_{qt} denoting the holding costs for final products in the output buffer $q \in O$ during period t, X_{qt} representing the cumulative output from buffer $q \in B$ (with $B(s)$ being buffers whose items are processed at station $s \in S$) being the cumulative requirements of final product in buffer $q \in O$ up to time t. The variable Z_{qut} is the cumulative flow from buffer $q \in B \setminus O$ to buffer $u \in B \setminus \Im$. $\widehat{X}_{qt}$ represents the cumulative input (or output) to buffer $q \in B$ and, finally, α_{qu} and β_{qu} denote the quantity of flow Z_{qut} needed to make one unit of input to buffer q or one unit of output, respectively.

The objective function minimizes all relevant costs. The first constraint describes the load dependent transformation of input into output. The following two equations describe the transfer of material between the buffers, and the fourth

equation gives the conservation of material flow. The load-dependent function $\hat{X}_{qt} = f_{qt}(X, W)$ is nonlinear, includes the above mentioned weights which embody stochastic queuing and service and describes the expected cumulative output X_{qt} in terms of input $\widehat{X}$ as a function of the load in the system. After a transformation (see Mendoza 2003 for more detail) we obtain for the cumulative output from buffer $q \in B$:

$$X_{qt} = \sum_{s=1}^{u} x_{qt} w_q(s, t)$$

where $x_{qt} = \widehat{X}_{qt} - \widehat{X}_{q,t-1}$ is the marginal amount of input to buffer q during period t.

Mendoza (2003) includes in his model further stochastic issues like random demand and random output and formulates an optimization algorithm in order to find a feasible solution based on the computed weights which are consistent with the load of the system which means that planned production schedules will also be feasible. So, the model is able to schedule later releases in certain situations, e.g., if requirements of the production changes during the production horizon, and consequently reducing FGI while still satisfying production requirements.

4 Conclusions

We have seen that although considerations about load dependent lead times are rare in the literature to date, there are some very interesting aggregate planning models that take them into account. This is particularly noteworthy, because lead times are of critical importance with respect to the global competitiveness of firms. Furthermore, we have discussed the importance of accounting for the nonlinear relationship between lead times and workload of production systems as well as other factors such as, e.g., product mix, scheduling policies, batching or lot sizing, variable demand patterns etc. Queuing models highlight the evidence that lead times increase nonlinearly long before resource utilization reaches 100% which is only included in few mathematical planning models. Additionally, there is a lack of models that analyze LDLT in the context of stochastic demand and other uncertainties evidently prevailing in practice.

The approach of modeling clearing functions in order to account for LDLT is considered very promising and should be implemented in a stochastic framework by using queuing models with the purpose to integrate the problem of variable demand patterns and to analyze the behavior of LDLT as proposed by Buzacott and Shantikumar (1993). This could be used as a starting point for additional modeling of production systems. It is evident that not all concepts and methods are well suited for each type of production system (job shop, flexible manufacturing system, etc.) with one or more production stages or products (one level – multi level production systems, one item – multi items or products), respectively. In case of LDLT and their integration into mathematical programming models for the aggregate production

planning process it is wise to begin with the simplest production system and then proceed to more complex systems and situations in order to arrive at overall supply networks with the aim of finding general principles of modeling and accounting for LDLT.

For future research we also need some careful consideration of the interaction between planning and scheduling. For a better understanding of the factors that determine the performance of a supply chain or a supply network, it is necessary to create formal and structured approaches and then incorporate the resulting knowledge and findings in software programs which can help to support planning decisions about the production system and help to find optimal decisions with respect to the complexities, mutual dependencies inherent in the system. The increasing level of automation, especially at the operational level, not only aims at improving the physical flow of materials through the supply network, but also tries to combine the understanding of system behavior with information and computational support in order to explore different courses of action and find optimal production plans.

References

Agnew C (1976) Dynamic modeling and control of congestion-prone systems. *Operations Research* 24(3):400–419

Asmundsson J, Rardin RL, Uzsoy R (2002) Tractable nonlinear capacity models for aggregate production planning. Working paper, School of Industrial Engineering, Purdue University, West Lafayette

Asmundsson J, Rardin RL, Uzsoy R (2003) An experimental comparison of linear programming models for production planning utilizing fixed lead time and clearing functions. Working paper, School of Industrial Engineering, Purdue University, West Lafayette

Baccelli F, Bremaud P (1987) *Palm Probabilities and Stationary Queues*. Springer, Berlin Heidelberg New York

Bechte W (1982) Controlling manufacturing lead time and work-in-process inventory by means of load-oriented order release. In: *APICS*, pp. 67–72. APICS, 25th Ann. International Conference of Production of the American Production Inventory Control Society

Bechte W (1988) Theory and practice of load-oriented manufacturing control. *International Journal of Production Research* 26(3):375–395

Ben-Daya M, Raouf A (1994) Inventory lead time as a decision variable. *Journal of the Operational Research Society* 45(5):579–582

Bertrand JWM (1983) The effect of workload dependent due dates on job shop performance. *Management Science* 29(7):799–816

Bertrand JWM, Wijngaard J (1986) The structuring of production control systems. *International Journal of Operations and Production Management* 6(2):5–20

Bertrand JWM, Wortmann J (1981) *Production Control and Information Systems for Component Manufacturing*. Elsevier, Amsterdam

Billington PJ, McClain JO, Thomas LJ (1983) Mathematical programming approaches to capacity-constrained mrp systems: Review, formulation and problem reduction. *Management Science* 29(10):1126–1137

Bitran GR, Haas EA, Hax AC (1981) Hierarchical production planning: A single stage system. *Operations Research* 29:717–743

Bitran GR, Haas EA, Hax AC (1982) Hierarchical production planning: A two stage system. *Operations Research* 30:232–251

Buzacott JA, Shantikumar JG (1993) *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, New York

Caramanis MC (2000) Multiple time scale decomposition in the production planning of stochastic manufacturing supply chain systems. http://www.bu.edu/pcms/docs/mcc/mtsdppsc.pdf, last call 18 Nov 2005

Caramanis MC, Anli OM (1999a) Dynamic lead time modeling for JIT production planning. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1450–1455

Caramanis MC, Anli OM (1999b) Modeling load-dependent lead time constraints for efficient supply chain coordination: Convexity issues. In: *Proceedings of the 38th IEEE Conference on Decision and Control*, vol. 1, pp. 900–906

Chen F, Drezner Z, Ryan JK, Simchi-Levi D (2000) Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science* 46:436–443

Clark AR, Armentano VA (1995) A heuristic for a resource-capacitated multi-stage lot-sizing problem with lead times. *Journal of the Operational Research Society* 46:1208–1222

Dauzère-Pérès S, Lasserre J-B (2002) On the importance of sequencing decisions in production planning and scheduling. *International Transactions in Operational Research* 9(6):779–793

de Kok TG, Fransoo JC (2003) Planning supply chain operations: Definition and comparison of planning concepts. In: de Kok AG, Graves SC (eds) *Supply chain management: design, coordination and operation*. Handbooks in operations research and management science, pp. 597–675. North-Holland, Amsterdam

Dempster MAH, Fisher ML, Jansen L, Lageweg BJ, Lenstra JK, Rinnooy Kan AHG (1981) Analytical evaluation of hierarchical planning systems. *Operations Research* 29(4):707–716

Enns ST, Choi S (2002) Use of GI/G/1 queuing approximations to set tactical parameters for the simulation of mrp systems. In: *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, pp. 1123–1129

Enns ST, Suwanruji P (2004) Work load responsive adjustment of planned lead times. *Journal of Manufacturing Technology Management* 15(1):90–100

Ettl M, Feigin GE, Lin GY, Yao DD (2000) A supply network model with base-stock control and service requirements. *Operations Research* 48:216–232

Fleischmann B, Meyr H (2003) Planning hierarchy, modeling and advanced planning systems. In: de Kok AG, Graves SC (eds) *Supply chain management: design, coordination and operation*. Handbooks in operations research and management science, pp. 455–523. North-Holland, Amsterdam

Glassey CR, Resende MGC (1988) Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 1:36–46

Goldratt EM (1988) Computerized shop floor scheduling. *International Journal of Production Research* 26(3):443–455

Goldratt EM, Fox RE (1986) *The race*. North River, Croton-on-Hudson

Graves SC (1986) A tactical planning model for job shops. *Operations Research* 34:522–533

Graves SC, Willems SP (2000) Optimizing strategic safety stock placement in supply chains. Manufacturing & Service Operations Management 2: 68–83

Hackman ST, Leachman RC (1989) A general framework for modeling production. *Management Science*, 35:478–495

Hall NG, Potts CN (2003) Supply chain scheduling: Batching and delivery. *Operations Research*, 51:566–584

Hendry LC, Kingsman BC (1991) Job release: Part of a hierarchical system to manage manufacturing lead times in make-to-order companies. *Journal of the Operational Research Society* 42:871–883

Homem-de-Mello T, Shapiro A, Spearman ML (1999) Finding optimal material release times using simulation-based optimization. *Management Science* 45:86–102

Hopp W, Spearman M (2001) *Factory Physics – Foundations of Manufacturing Management*. McGraw-Hill, Boston, 2nd ed

Hopp WJ, Spearman ML, Woodruff DL (1990) Practical strategies for lead time reduction. *Manufacturing Review* 3(2):78–84

Huang GQ, Lau JSK, Mak KL (2003) The impacts of sharing production information on supply chain dynamics: A review of the literature. *International Journal of Production Research* 41:1483–1517

Hwang S, Uzsoy R (2004) A single-product dynamic lot sizing model with work in process and congestion. Technical report, School of Industrial Engineering, Purdue University, West Lafayette

Karmarkar US (1987) Lot sizes, lead times and in-process inventories. *Management Science* 33:409–418

Karmarkar US (1989a) Batching to minimize flow times on parallel heterogeneous machines. *Management Science* 35:607–613

Karmarkar US (1989b) Capacity loading and release planning with work-in-process (WIP) and lead times. *Journal of Manufacturing and Operations Management* 2:105–123

Karmarkar US (1993a) Lot size, manufacturing lead times and utilization. Working paper no. QM8321, Graduate School of Management, University of Rochester

Karmarkar US (1993b) Manufacturing lead times, order release and capacity loading. In: Graves SC, Rinnooy Kan A, Zipkin P (eds) *Logistics of Production and Inventory*, vol. 4 of *Handbooks in Operations Research and Management Science*, pp. 287–329. North-Holland, Amsterdam

Karmarkar US, Kekre S, Kekre S (1985a) Lot sizing in multi-item multi-machine job shops. *IIE Transactions* 17:290–298

Karmarkar US, Kekre S, Kekre S (1987) Capacity analysis of a manufacturing cell. *Journal of Manufacturing Systems* 6(3):165–175

Karmarkar US, Kekre S, Kekre S, Freeman S (1985b) Lot-sizing and lead-time performance in a manufacturing cell. *Interfaces* 15(2):1–9

Kekre S, Kekre S (1985) Work-in-process considerations in job shop capacity planning. Working paper, GSIA, Carnegie Mellon University

Kim J, Leachman RC, Suh B (1996) Dynamic release control policy for the semiconductor wafer fabrication lines. *Journal of the Operational Research Society* 47:1516–1525

Krieg G (2003) *KANBAN-Controlled Manufacturing Systems*. Phd thesis, University of Ingolstadt, Ingolstadt

Land MJ, Gaalman GJ (2003) Workload control concept in job shops: A critical assessment. *International Journal of Production Economics* 46/47(11):535–548

Lautenschläger M (1999) *Mittelfristige Produktionsprogrammplanung mit auslastungsabhängingen Vorlaufzeiten*. PhD thesis, University of Darmstadt, Darmstadt

Lautenschläger M, Stadtler H (1998) Modelling lead times depending on capacity utilization. Technical report, Technische Hochschule Darmstadt, Darmstadt

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: The bullwhip effect. *Management Science* 43:546–558

Mendoza GR (2003) *Transient Behavior of Stochastic Networks: Application to Production Planning with Load-Dependent Lead Times*. PhD thesis, Georgia Institute of Technology

Missbauer H (1998) *Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen*. PhD thesis, University of Heidelberg, Physica Verlag, Heidelberg

Missbauer H (2002) Lot sizing in workload control systems. *Production Planning and Control* 13(7):649–664

Ouyang L, Yeh N, Wu K (1997) Mixture inventory models with backorders and lost sales for variable lead times. *Journal of the Operational Research Society* 47:829–832

Pahl J, Voß S, Woodruff DL (2005) Load dependent lead times – from empirical evidence to mathematical modeling. In: Kotzab H, Seuring S, Müller M, Reiner G (eds) *Research methodologies in supply chain management*, pp. 539–554. Physica, Heidelberg

Philipoom PR, Fry TD (1992) Capacity-based order review/release strategies to improve manufacturing performance. *International Journal of Production Research* 30(11):2559–2572

Simchi-Levi D, Kaminsky P, Simchi-Levi E (2002) *Designing and Managing the Supply Chain*. Irwin McGraw-Hill, Boston, 2 ed

Spearman ML, Hopp WJ, Woodruff DL (1989) A hierarchical control architecture for CONWIP production systems. *Journal of Manufacturing and Operations Management* 2:147–171

Spearman ML, Woodruff DL, Hopp WJ (1990) CONWIP: A pull alternative to Kanban. *International Journal of Production Research* 28:879–894

Spearman ML, Zhang RQ (1999) Optimal lead time policies. *Management Science* 45(2):290–295

Srinivasan A, Carey M, Morton TE (1988) Resource pricing and aggregate scheduling in manufacturing systems. Working paper, GSIA

Stadtler H, Kilger C (eds) (2005) *Supply chain management and advanced planning*. Springer, Berlin, 3rd ed

Suri R, Hildenbrant RR (1894) Modeling flexible manufacturing systems using mean value analysis. *Journal of Manufacturing Systems* 3:27–38

Tatsiopoulos IP, Kingsman BG (1983) Lead time management. *European Journal of Operational Research* 14:351–358

Thonemann UW (2002) Improving supply-chain performance by sharing advance demand information. *European Journal of Operational Research* 142:81–107

Voß S, Woodruff DL (2004) A model for multi-stage production planning with load dependent lead times. In: Sprague RH (ed) *Proceedings of the 37th Annual Hawaii International Conference on System Science*, pages DTVEA03 1–9, Piscataway, 2004

Voß S, Woodruff DL (2006) An introduction to computational optimization models for production plannung in a supply chain. Springer, Berlin, 2nd ed (to appear)

Wagner HM, Whitin TM (1958) Dynamic version of the economic lot size model. *Management Science* 5:89–96

Wiendahl H, Wedemeyer H (1990) Das Dilemma der Fertigungssteuerung. *Zeitschrift für Betriebswirtschaft* 60:407–422

Witt A, Voß S (2005) Simple heuristics for scheduling with limited intermediate storage. *Computers & Operations Research*, to appear

Wu K (2001) A mixed inventory model with variable lead time and random supplier capacity. *Production Planning and Control* 33:95–106

Yano CA (1987) Setting planned leadtimes in serial production systems with tardiness costs. *Management Science* 33:95–106

Yücesan E, de Groote X (2000) Lead times, order release mechanisms, and customer service. *European Journal of Operational Research* 120:118–130

Zäpfel G, Missbauer H (1993a) New concepts for production planning and control. *European Journal of Operational Research* 67:297–320

Zäpfel G, Missbauer H (1993b) Production planning and control (PPC) systems including load-oriented order release – problems and research perspectives. *International Journal of Production Economics* 30:107–122

Zhao X, Xie J, Leung J (2002) The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research* 142:321–344

Zijm WHM, Buitenhek R (1996) Capacity planning and lead time management. *International Journal of Production Economics*, 46/47:165–179, 1996.

Ann Oper Res (2007) 153: 297–345
DOI 10.1007/s10479-007-0173-5

Production planning with load dependent lead times: an update of research

Julia Pahl · Stefan Voß · David L. Woodruff

Published online: 3 May 2007
© Springer Science+Business Media, LLC 2007

Abstract Lead times impact the performance of the supply chain significantly. Although there is a large body of literature concerning queuing models for the analysis of the relationship between capacity utilization and lead times, and another body of work on control and order release policies that take lead times into consideration, there have been relatively few aggregate planning models that recognize the (nonlinear) relationship between the planned utilization of capacity and lead times. In this paper we provide an in-depth discussion of the state-of-the art in this area, with particular attention to those models that are appropriate at the aggregate planning level.

Keywords Production planning · Load dependent lead times · Mathematical programming · Supply chain management

1 Introduction

Let us define the *lead time* as the time between the release of an order to the shop floor or to a supplier and the receipt of the items. Lead time considerations are essential with respect to the global competitiveness of firms, because long lead times impose high costs due to rising *work in process* (WIP) inventory levels as well as larger safety stocks caused by increased uncertainty about demand. Large planning models typically treat lead times

An earlier version of this paper appeared in *4OR* 3, 257–302, 2005.

J. Pahl · S. Voß (✉)
Institut für Wirtschaftsinformatik, Universität Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
e-mail: stefan.voss@uni-hamburg.de

J. Pahl
e-mail: pahl@econ.uni-hamburg.de

D.L. Woodruff
Graduate School of Management, UC Davis, Davis, CA 95616, USA
e-mail: dlwoodruff@ucdavis.edu

as static, exogenous input data, but in most situations the output of a planning model implies capacity utilizations which, in turn, imply lead times. Lead times are among the most important properties of items in production planning and *supply chain management* (SCM).

Despite this, consideration of lead times that are dependent upon resource utilization (*load dependent lead times*: LDLT) is still rare in the literature. The same is valid for models linking order releases, planning and capacity decisions to lead times, and taking into account factors influencing lead times such as the system workload, batching and sequencing decisions or WIP levels. In short, in most SCM and production planning models nonlinear dependencies, e.g., between lead times and the workload of a production system or a production resource, are usually omitted. This happens even though there is empirical evidence that lead times begin to increase nonlinearly long before resource utilization reaches 100%, which may lead to significant differences in planned and realized lead times.

In this paper we present different approaches to modeling LDLT. There are indirect approaches, primarily aiming at avoiding LDLT and their causes before they arise, e.g., demand variability smoothing methods, job release policies and production planning design, as well as models addressing the problem by implementing LDLT in aggregate production planning using clearing functions. An earlier version of this paper has appeared as Pahl et al. (2005b).

The literature we review comes from both well-known journals and less well-known dissertations, which provides the reader a broad set of sources for further studies. We devote Sect. 2 to providing some context for LDLT in the larger supply chain and production planning literatures. In Sect. 3 we provide detailed descriptions of recent research aimed at the development of aggregate production planning models that are able to capture the nonlinear dependency of lead times and workload directly and nevertheless remain computationally tractable. Finally, some conclusions are provided.

2 Supply chain management and lead time discussion

SCM rose to prominence as a major management issue in the last decades. While the focus of managing supply chains has undergone a drastic change as a result of improved information technology (IT), production planning remains a critical issue. For some references on SCM see Simchi-Levi et al. (2002), Stadtler and Kilger (2005), and Voß and Woodruff (2006).

In this section we provide context for the models that will be discussed in Sect. 3 below. Besides a few remarks on information sharing (Sect. 2.1) we provide some links to literature concerning common planning concepts (Sect. 2.2) as well as release mechanisms (Sect. 2.3).

Before going into detail, let us reconsider our lead time definition. Besides our definition given above, Yücesan and de Groote (2000) define lead times as the time between the authorization of production to the completion of processing, at which point the material is ready to fill a customer order. So, lead time consists of queuing time, processing time, batching time, and transportation/handling time. For instance, Hung and Hou (2001) refer to this as the flow time.

Even if it is not related to the literature on LDLT it should be mentioned that various forms of lead times may be discussed, such as effectuation lead times (the time between a decision is made and the time the consequences of this decision can be observed throughout the supply chain), minimal lead times, and actual lead times; see, e.g., de Kok et al. (2003). Also, in semiconductor manufacturing and assembly line balancing the term cycle time is used quite frequently; see, e.g., Hopp and Spearman (2001) and Domschke et al. (1997).

2.1 Information sharing

Many companies use modern IT to help them gain competitive advantage in the marketplace. The rapid advances in IT have provided tools that enable supply chain partners to share information with each other. However, questions concerning the benefits that can be gained through the sharing of information are frequently raised. Several researchers have examined the impact of information sharing on business performance; see, e.g., Thonemann (2002) and Zhao et al. (2002). A survey of the impacts of sharing information, including some lead time discussion, is provided by Huang et al. (2003).

One of the most common information sharing issues is the so-called bullwhip effect. It relates to the observed behavior when forecasts for intermediate products within a supply chain are based only on the direct demand experienced for those products, i.e., variability in the demand pattern is magnified. For references on the bullwhip effect see, e.g., Lee et al. (1997), Chen et al. (2000), Simchi-Levi et al. (2002).

The dynamics and the globalization of markets imply competition on time-based factors like flexibility or responsiveness to consumers, especially in case of variable demand patterns as pointed out by Spearman and Zhang (1999). This results in the need to share business information, especially information related to the production process and the inventory levels for use in planning models. Without proper communication and planning models, reducing lead times at individual nodes of the supply chain does not automatically lead to improvements in global supply chain lead times. If, e.g., an individual node of the supply chain is able to respond flexibly to production requirement changes by modulating its own lead time, this does not result in a benefit for the overall supply chain if information about this flexibility in lead time dynamics is not communicated to the other partners (see Graves and Willems 2000 or Hall and Potts 2003). For instance, Caramanis and Anli (1999a) propose an iterative learning algorithm that shows how efficiency can be increased by sharing information at each individual production and decision node of a production cell. These nodes can also be interpreted as companies related to each other in a supply chain.

Companies are aware of the need not only to share business information, but also to actively and effectively manage their supply chains. However, the trend towards collaborations results in increases in supply chain size and system complexity. Consequently, decisions that were adequate on a small scale can have unexpected consequences on a large scale. Therefore, it is mandatory to understand the operational dynamics of the various aspects of supply chains. Orcun et al. (2006a) propose modeling tools for collaborative companies that allow them to design, develop and change aggregate models for supply chains including material and information flows in order to gain insights and assess alternative configurations. Effects of different types of uncertainty, e.g., lead times, demand patterns, etc., can be examined as well as the effects of competing goals regarding long or short-term contracts, volume commitments or price agreements on the entire system.

2.2 Common planning concepts

Present practice for manufacturing supply chains is dominated by the utilization of models based on *materials requirements planning* (mrp) which are associated with a variety of problems. Homem-de-Mello et al. (1999), Caramanis and Anli (1999a) and Hopp and Spearman (2001), among others, highlight that one of these problems lies in the assumption of fixed and constant lead times at the aggregate level of the planning process and, what is even worse, in the utilization of the "worst case lead time", i.e., the maximal lead time for an individual node. This is done in order to have enough "buffer time" to ensure that the

demand requirements can be met in any case (see also Enns and Suwanruji 2004 or Enns 2001). Additionally, in order to meet due dates, there is a tendency to release jobs into the production system much earlier than necessary, leading to very high WIP levels and therefore longer waiting times; see also Enns (2001). Consequently, lead times get even longer. This over-reaction becomes a self-fulfilling prophecy and is addressed in the literature as the *lead time syndrome*. It results from the fact that the relationship between WIP, output, workload and average flow times is ignored (see Zäpfel and Missbauer 1993a, 1993b or Tatsiopoulos and Kingsman 1983). For an extensive study see Selcuk et al. (2006). Furthermore, these planning models hide the inherent variability of lead times and their dependency on planning decisions by defining planned lead times which are much longer than necessary. Additionally, they use planning techniques that ignore consistency constraints in determining production targets for single production cells (stations or resources) or, as Caramanis and Anli (1999a) point out, tighten them so as to linearize and make them time invariant.

Hackman and Leachman (1989) propose a framework for modeling production processes that generalizes and improves the accuracy of simple production models integrated in *linear programming* (LP), mrp and the *critical path method* (CPM) based on modeling and planning techniques by reformulating the constraints of components of production lead times. Billington et al. (1983) address the shortcoming of mrp models to treat the capacity as infinite. They present a mathematical programming approach dealing with capacity constraints which are scheduled considering other production stages and unpredictable production lead times in order to account for the fact that lead times are related to capacity planning. Clark and Armentano (1995) develop a heuristic model for a resource capacitated multi-stage lot sizing problem including general production structures, setup costs, WIP inventory costs and production lead times. The heuristic first calculates a production plan for the incapacitated case and then shifts production backward in time so as to generate capacity feasible production plans. There is also some scheduling literature where lead time estimates are developed by simulation and used in an iterative framework; see, e.g., Vepsalainen and Morton (1988) and Lu et al. (1994).

Most mrp and *enterprise resource planning* (ERP) models implement sequential planning algorithms which omit many functional interdependencies or uncertainties. For instance, Karmarkar et al. (1993b) points out that the lead time of a part is rather a property of the state of the production facility and, consequently, depends on the overall available capacity and workload which can vary significantly over time. In Asmundsson et al. (2003, 2006a) and Karmarkar (1987) we see that one of the fundamental problems of various manufacturing and production planning models like mrp or *manufacturing resources planning* (MRP II) is the omission of modeling nonlinear dependencies. These include, e.g., the relationship between lead times and the workload of a production system or a production resource. However, there is empirical evidence that lead times increase nonlinearly long before resource utilization reaches 100%. This is also highlighted in Pahl et al. (2005a). As a consequence, production plans tend to be inaccurate because of the fact that lead times are extremely dependent on decisions about the product mix, workload and scheduling policies which is also emphasized in Asmundsson et al. (2002, 2003, 2006a), Caramanis (2000), Caramanis and Anli (1999a, 1999b), Yano (1987), Zijm and Buitenhek (1996). This is particularly true for unsynchronized manufacturing shops where products can follow diverse routings and resources have limited capacity. Nevertheless, mrp systems treat the production lead time of a product, part, or *stock keeping unit* (SKU) as a constant (exogenous) parameter independent of other production factors and, therefore, as an attribute of the part (see Dauzère-Pérès and Lasserre 2002).

Missbauer (2002) integrates the functional relationships between, e.g., the mean level of WIP and important core variables of production planning like flow time or capacity utilization in lot sizing models to analyze the impact of lot sizing on these specific values and dependencies. Enns and Choi (2002) and Enns and Suwanruji (2004) develop methods for setting dynamically planned lead times based on the current system workload and lot sizes in MRP II systems. They study the use of exponentially smoothed order flow time feedback for the dynamic planned lead time setting in a $GI|G|1$ queuing model. Results reveal that the assumptions considering the *lot inter-arrival time variability* have significant influence on lot sizes and overall system performance since lead times are significantly affected by the system workload which, in turn, is dependent on capacity and, therefore, setup times.

Selcuk et al. (2006) analyze the effects of updating lead times on the performance of a stochastic two level hierarchical planning system. Lead times are updated using periodic feedback and actual information on the production environment. Using discrete event simulation they reveal that updating lead times causes increases in lead times and production costs and that the main driver of this lead time inflation is material starvation. Consequently, updating lead times in a hierarchical setting may be questionable.

Hung and Hou (2001) suggest an iterative scheme for a flow time prediction model where the flow time encompasses transfer times, waiting times, and processing times. Their LP approach is an efficient alternative to simulation. It is based on empirical data describing the relationship between the loading rate and waiting time by a piecewise linear function. An approach to estimate lead times in a make-to-order system employing data mining is suggested by Öztürk et al. (2006). Furthermore, Jodlbauer (2004) studies the behavior of the input-weighted average standard processing time for a multi-item machine and shows that it is not influenced by scheduling, sequencing or lot sizing, but by the total input and the individual standard processing times of items. If the standard processing time is very short, its reduction results in an increase of the input-weighted average processing time which is caused by the convexity of the latter. Consequently, in order to minimize the average processing time the focus should be on products with large standard processing times. For a mathematical definition and a derivation of the properties of input-weighted average lead times see Jodlbauer (2005).

Voß and Woodruff (2004) concentrate on the fact that decisions about the utilization of production resources in the aggregate planning process affect realized lead times. This is even more important to capture when extending the aggregate planning process to whole supply chains. They further emphasize the need to simultaneously consider lead time consequences and alternative routings and/or suppliers—which they model as additional routing alternatives—arising during the planning process. This is due to the fact that bottlenecks shift as a result of the production planning process. The important insight, which is also highlighted in Goldratt and Fox (1986), is that the consideration of interactions of bottlenecks in the production system is of great importance also because of the fact that a resource becoming a bottleneck is dependent on the workload. It therefore depends on decisions about loading and alternative routings. These, in turn, are determined much earlier in the aggregate production planning process and, thus, from the point of view of Voß and Woodruff (2004), should not be distributed to the control level. Their aim is to provide a useful planning model with realistic assumptions from which feasible (realizable) schedules can be derived and which accounts for consequences of LDLT on planning decisions for supply chain networks. Instead of focusing on a single bottleneck one may consider several near-bottlenecks together in planning decisions; see, e.g., Horiguchi et al. (2001).

Recent versions of *advanced planning systems* (APS) incorporate finite capacity schedule techniques based on diverse models targeted toward determining accurate start times and

schedules for jobs without ignoring capacity constraints (see, e.g., Fleischmann et al. 2003 and the contributions in Stadtler and Kilger 2005). Together with *manufacturing execution systems* (MES) which use network technology in order to provide real time tracking of important production resource information (machines, labor, WIP levels), the APS receive the current production status of the shop floor and the planned order releases from an ERP system with the objective to optimize the production schedule (see, e.g., Homem-de-Mello et al. 1999).

2.3 Release mechanisms

The use of pull-based control mechanisms can increase the predictability of lead times and thereby improve the performance of planning systems. There is vast literature on this subject; see, e.g., Ben-Daya and Raouf (1994), Buzacott and Shantikumar (1993), Hopp et al. (1990), Ouyang et al. (1997), Srinivasan et al. (1988), Wu (2001). In addition, pull-based control systems like KANBAN or CONWIP aim at controlling workflow in order to reduce overall inventories and related costs; see, e.g., Spearman et al. (1989, 1990). The KANBAN concept is an early suggestion from Japan to implement the pull principle in manufacturing systems. The specific information system includes the KANBAN card, which can be regarded as an authorization for production. The main characteristic of KANBAN systems is reflected by the closed loops of every production unit. For more details on KANBAN systems see, e.g., Krieg (2003). In CONWIP systems, the manufacturing system is comprised of one closed loop card system, so that one card is attributable to one product container and, consequently, inventory is equal to the number of cards in the system. The name "*CONWIP*" is derived from "*CONstant Work In Process*". The CONWIP system can be regarded as a closed queuing network where the orders correspond to the constant number of cards in the system; see, e.g., Hopp and Spearman (2001).

There are many studies in the literature that examine pull release mechanisms based on the workload of production facilities as, e.g., Bertrand and Wortmann (1981), Bertrand (1983) and methods of workload oriented production control as, e.g., Bechte (1982), Bertrand and Wortmann (1981), Bertrand (1983), Wiendahl and Wedemeyer (1990). According to Homem-de-Mello et al. (1999), neither KANBAN nor CONWIP integrate due dates in order to smooth releases and therefore workload, which leads to inaccurate order releases.

Adoption of KANBAN and *just in time* (JIT) techniques in complex production systems that do not have the proper characteristics such as flow balance between production stages, highly similar products etc., is problematic as pointed out by Zäpfel and Missbauer (1993a, 1993b). Moreover, the problem of workload oriented production controls lies in the fact that often, from a supply chain perspective, they do not really solve any problem regarding lead times, but rather increase their mean and variance by shifting the workload from inside to outside the manufacturing shop. That is, jobs that would wait within the shop floor are forced to wait in front of the shop because workload controls only release jobs when enough capacity is available; see also Zijm and Buitenhek (1996) or Philipoom and Fry (1992).

Other authors argue that holding (delaying) jobs in a so-called "job pool" and thereby controlling a hierarchy of backlog lengths provides a good mechanism to control the shop floor throughput times. As has been argued by Hendry and Kingsman (1991) the latter are an important part of lead times and consequently reduce congestion situations in the production system, WIP levels, LDLT and related costs. Philipoom and Fry (1992) investigate the benefits and drawbacks of not only delaying jobs in a job pool, but rejecting customer orders when the system workload or utilization is very high in order to avoid long queues in

front of the entry point to the production facility and to improve delivery reliability. Rejecting orders will surely lead to the loss of goodwill, but accepting customer orders including due dates and failing to deliver leads to even higher loss of goodwill. For further reading see Philipoom and Fry (1992). They also point out that unconditionally accepting customer orders leads to increased average flow times, increasing uncertainty whether agreed upon due dates are met for all other jobs in the production system. They show that rejecting a small percentage of orders can lead to considerable improvements in system performance. Job release policies and release controls aim at avoiding overload of the production system and, consequently, restrict WIP to reduce queues and therefore the dependency of the lead time on the workload which then allows for the usage of simpler planning models. Witt and Voß (2005) propose controlling the WIP in a hybrid flow shop by modifying release quantities based on limited intermediate storage.

Agnew (1976) proposes a dynamic model of a general congestion-prone system which illustrates the relationship between throughput and workload of the system. He analyzes the system behavior and the form of optimal policies for most accurately controlling congestions. Land and Gaalman (1996) propose a comprehensive production *workload control model* (WLC) which buffers the shop floor against external dynamics like rush orders or demand peaks and tries to equilibrate the queuing of jobs in the shop floor in a stationary process using workload norms. Various authors have extended these WLC models to hierarchical capacity-oriented production control concepts for job shops as, e.g., Bechte (1988), Bertrand and Wortmann (1981). Missbauer (2002) presents a decision model for the optimization of aggregate order releases that takes into account the dependency of the work center output on the WIP and thus can be considered as a workload control method. A model that combines aggregate order release planning and workload control aiming at optimizing the aggregate flow of materials through a production unit is given by Missbauer (2006a). Homem-de-Mello et al. (1999) develop a job release method using simulation-based optimization for production scheduling that aims at setting release times for jobs with fixed due dates and determined sequences in stochastic production lines including machine down times, where job release criteria are WIP levels as well as customer service (rates). One objective is to keep flow or cycle time (sum of processing and queuing time) at a minimum level in order to increase responsiveness and flexibility in manufacturing and to reduce WIP. They consider a transient situation, as opposed to other papers which deal with setting release times using steady state models.

Hackman et al. (2002) propose a multi-product multi-period stochastic dynamic production planning model that determines the release times of products in regard of forecast requirements and various (time-varying) constraints on resources, production smoothing, machine breakdowns, rework, scrapping, delays in supplies, personnel availability, etc. Lead times are assumed to be random variables with a distribution that can be estimated, e.g., from historical data. The aim of the model is to determine the release times for the products that ensure meeting the predetermined requirements in each planning period with the desired probability at minimum costs. They suggest an LP model where release times are dynamically determined using a rolling planning horizon. Simulation results show that their model is able to derive more realistic plans than classical MRP II models with reduced production costs.

Glassey and Resende (1988) study control policies including information about the type of machine failure in order to minimize the average inventory in two stage flow lines. They define a specific threshold level for each type of failure of downstream resources (stations). They show that dynamic control policies reacting to machine failures are most valuable in the case that the second stage resource is the bottleneck, the downtime is not too small and

the ratio of the repair rate to the processing rate of the second resource is small. For an extensive description of this body of work in the semiconductor industry see Uzsoy et al. (1994).

Further work includes studies on closed-loop job release mechanisms for job shops with machine failures and repair as a main source of randomness. The so-called "starvation avoidance" control mechanism aims at maximizing utilization of critical bottlenecks while controlling the development of the WIP. By monitoring actual WIP levels this job control procedure only releases jobs into the production system when inventories are below a predetermined safety stock level. Kim et al. (1996) propose a release control mechanism designed for semiconductor wafer fabrication that uses a tracking system which provides the model with actual information about the shop floor situation. A heuristic is proposed in order to minimize the mean flow time and mean tardiness while observing given output rates. Simulation results are provided for a simplified semiconductor wafer fabrication environment. A more detailed overview of order release algorithms can be found in Hendry and Kingsman (1991), Philipoom and Fry (1992), and Tatsiopoulos and Kingsman (1983).

Wiendahl and Wedemeyer (1990) describe the general dilemma of production planning, scheduling and release control which has its roots in the conflicting planning objectives and resulting trade-offs as, e.g., minimization of lead times and WIP inventory levels while maintaining high utilization levels. They argue that, despite of these problems, it is possible to find an optimal solution to each situation by employing an order release control method. In addition, Zäpfel and Missbauer (1993a) give an overview of production planning and control concepts that aim at overcoming the deficiencies of traditional models such as MRP II. These include concepts like *Optimized Production Technology* (OPT), *Just in Time Production Planning and Control* (JIT-PPC) as well as production planning and control including progressive figures which provide an instrument for monitoring material flows employing a graphic or diagram, but do not include methods for capacity planning.

The OPT concept considers the performance of the production system as dependent on the bottleneck in the system and takes it as a basic constraint (starting point) to calculate the detailed production plan. The OPT scheduling algorithm generates the detailed schedule for the bottleneck and then performs the forward scheduling of the orders also including the determination of transfer batches and process batches. The resulting flow times and delivery times are the outcome of the planning process which is the most important result of this concept, namely that it does not use estimated planned lead times. The order release process is determined on the so-called "drum beat" of the bottleneck resource, i.e., its production rate which is calculated by the drum-buffer-rope approach (DBR) and determines the schedule for the bottleneck resource. The concept is discussed extensively in Goldratt (1988) and Zäpfel and Missbauer (1993a).

3 Models incorporating load dependent lead times

In this section we shall concentrate on the problem of how to develop aggregate production planning models which are able to directly capture the nonlinear dependency of lead times and workload and nevertheless remain tractable computationally. The first reference that we review is Zijm and Buitenhek (1996). They consider the relationship of lead times and workload without using the concept of clearing functions. This is incorporated in the model of Graves (1986) which opens the discussion of clearing functions. Models from Srinivasan et al. (1988), Karmarkar (1987, 1989a), Asmundsson et al. (2002, 2003, 2006a), and Hwang

and Uzsoy (2004) follow in this regard. Kim and Uzsoy (2006a) represent a capacity expansion problem with clearing functions and Upasani and Uzsoy (2005) integrate pricing decisions in their clearing function model.

The model of Missbauer (1998) emphasizes lot sizing decisions and integrates them into a clearing function model. The work of Lautenschläger (1999) concentrates on forward shifting functions for production planning. However, Caramanis and Anli (1999b) are the first who extend the view to the overall supply chain. Riaño (2002) emphasizes stochastic issues and develops a *transient Little's Law*. Finally, Selcuk et al. (2005) employ a short term clearing function in their hierarchical supply chain operations planning model aiming at linking the effects of order releases, throughput and scheduling decisions.

3.1 The planning and control model with workload oriented lead times of Zijm and Buitenhek

Zijm and Buitenhek (1996) develop a manufacturing planning and control framework for a machine shop that includes workload oriented lead time estimates in order to consider both lead time and capacity management in one management planning tool. For this purpose they suggest a method that determines the earliest possible completion times of arriving jobs with the restriction that the delivery performance of any other job in the system will not be adversely affected, i.e., that every job can be completed and delivered in time.

The goal is to determine reliable planned lead times based on the workload that results in due dates for jobs that can be met. The model can be implemented at a capacity planning level, serving there as an input for a final detailed capacity scheduling procedure that also takes into account additional resources, job batching decisions as well as machine setup characteristics. They use queuing network techniques to determine the mean and variance of lead times dependent on lot sizes, production mix and expected production volume. Their framework is partly based on the work by Karmarkar (1993a) and Karmarkar et al. (1985a) as they represent each network service station as a $M|G|1$ queuing model with multiple part types.

Employing different approximations for queuing networks can lead to improvements in the result, e.g., when modeling production system nodes as $GI|G|c$ networks in case of parallel machines. To find an approximation for the average work in the production system that has to be processed at a specific machine, Little's Law is used:

$$\overline{Y}_{ijk} = \lambda_{ijk}\overline{L}_{ijk}$$

where $\overline{Y}_{ijk}$ denotes the mean number of batches of part type i in queue or in process at resource (work station) j for their k-th operation, $\overline{L}_{ijk}$ is the mean lead time corresponding to the k-th operation of part type i at resource j, and λ_{ijk} the arrival rate of batches of part type i at resource j for their k-th operation (defined by $\lambda_{ijk} = \frac{D_i}{Q_i}\delta_{ijk}$, where D_i is the demand rate of part type i, Q_i the deterministic lot size of part type i and δ_{ijk} an indicator function which fixes the part type dependent routing of i). The batch processing time p^b_{ijk} of part type i at resource j for the k-th operation includes the setup time S_{ijk} and is given by $p^b_{ijk} = S_{ijk} + Q_i p_{ijk}$, where p_{ijk} denotes the deterministic processing time for part type i at resource j for the k-th operation. The n-th moment of the part type- and operation-independent batch processing time p^b_j is then represented as:

$$E(p^b_j)^n = \sum_{i=1}^{M}\sum_{k=1}^{K}\frac{\lambda_{ijk}}{\lambda_j}(p^b_{ijk})^1$$

where λ_j denotes the overall arrival rate at resource j. Consequently, the average workload or utilization ρ_j at workstation j is given by:

$$\rho_j = \lambda_j E(p_j^b)^n.$$

Assuming Poisson arrival processes at each resource, the Pollaczek–Khintchine formula yields the following expression for the mean waiting time in queue $\overline{Wt}_j$ at workstation j:

$$\overline{Wt}_j = \frac{\lambda_j E(p_j^b)^2}{2(1-\rho_j)}.$$

Moreover, the mean lead time $\overline{L}_{ijk}$ is determined by the mean waiting time in queue at resource j ($\overline{Wt}_j$) and the batch processing time p_{ijk}^b: $\overline{L}_{ijk} = \overline{Wt}_j + p_{ijk}^b$. We note the incorporation of the relationship between lead times and workload in this formula. The approximation for the average work in the production system $\overline{W}_j$ is then given as follows:

$$\overline{W}_j = \sum_{i=1}^{M}\sum_{k=1}^{K}\overline{Y}_{ijk}\left(p_{ijk}^b + \sum_{\hat{k}=k+1}^{K} p_{ij\hat{k}}^b\right) + \sum_{i=1}^{M}\sum_{\substack{\hat{j}=1\\ \hat{j}\neq j}}^{N}\sum_{k=1}^{K}\overline{Y}_{i\hat{j}k}\left(\sum_{\hat{k}=k+1}^{K} p_{i\hat{j}\hat{k}}^b\right).$$

The first term refers to the jobs currently present at resource j and includes both current (k) and future operations ($\hat{k}$). The second term represents the jobs currently present at other resources ($\hat{j} \neq j$) which could still have to visit resource j. $p_{ij\hat{k}}^b$ denotes the batch processing time of the future operation $\hat{k}$ on part type i at resource j. The term can be interpreted as the number of jobs at resource j on which the jobs have to undergo present and future operations plus the number of jobs which have not already been processed on resource j. Subsequent to the determination of both mean and variance of the lead times and the average work in the production system is a decomposition-based job shop scheduling algorithm (the *shifting bottleneck procedure*). The capacity planning procedure is then used to calculate the individual planned lead times for each batch, based on the actual workload of the production system.

To be more precise, first a scheduling algorithm is employed to minimize the maximum lateness for each job under the special characteristic that a job is allowed to visit a workstation, once, more than once, or not at all. Despite the fact that this is an NP-hard problem the shifting bottleneck heuristic gives satisfactory results. The decomposition algorithm then distributes the overall problem to a series of single machine scheduling problems for determining the virtual release and due dates for the operations with the objective of minimizing the maximum lateness. Once assigned, these due dates are fixed, i.e., modifications are not possible.

The capacity planning method based on aggregate scheduling finally assigns the due dates to each job depending on the current workload of the production system under the aforementioned condition that all jobs in the system can be finished without delay. These due dates further give the input for the detailed shop floor scheduling. What is obviously different here is the explicit consideration of different routings and machines at their capacity limit which is not included in classical "bucket filling" capacity planning models which tend to smooth the needed capacity over the scheduling time period and consequently ignore precedence relations. The model represented here is designed for a highly simplified machine shop, but can be modified for several shop configurations. Although it does not consider the nonlinear relationship between workload and LDLT, it provides a first approach for considering the workload of a production system at an aggregate planning level.

3.2 The tactical planning model of Graves

Graves (1986) develops a tactical planning model for a job shop for the purpose of studying the dependencies between production capability, variability (uncertainty) of the production requirements, and level of WIP inventory. He tries to analyze to which extent the job flow time (or WIP inventory) depends on the utilization of each resource of a job shop or production stage, respectively. Furthermore, the interrelationship of flow time and production mix is analyzed. The underlying production system is a flexible job shop, modeled as a network of queues where multiple routings of jobs are possible so that the lack of a dominant work flow complicates production control which aims at considerably reducing the variance of planned lead times. Nonetheless, input/output control systems try to manage the work flow through the shop by stabilizing queues at a predetermined level.

Planned lead times are used as the key decision parameters for implementing an input/output control in order to describe production rates by resource and by time period. For that purpose a queuing model is employed that includes *flexible production rates* of resources which can be set by a tactical planning model as to smooth the work flow in order to avoid overload of resources. According to Karmarkar (1989a) this imposes difficulties, because it implies that, applying Little's Law, lead times are fixed even when workload varies. Similar to Bertrand and Wortmann (1981), Graves (1986) develops a time discrete, aggregate planning model which attempts to control the flow time of jobs and hence the workload in the production system. This discrete-time, continuous flow model considers workload rather than single jobs. It is assumed that the movement of jobs from one resource to another and the arrival of new jobs only occur at the beginning or at the end of a time period, respectively. A control rule is implemented at each resource that determines the amount of work performed during a time period, which is a fixed portion α_j of the queue of work remaining at the start of the period at resource j:

$$X_{jt} = \alpha_j W_{jt} \quad \forall j, t$$

where X_{jt} denotes the production at resource j in time period t, α_j a smoothing parameter with $0 < \alpha_j < 1$ and W_{jt} the queue of work at the beginning of time period t. This smoothing parameter is called "proportional factor" by Missbauer (1998). Infinite capacity is implied since the resource is able to complete the fixed portion α_j even when the workload (WIP) is infinitely high. The major drawback of the model is that it employs fixed planned lead times, i.e., it is assumed that production processing rates can be regulated in order to account for workload in the production system. This implies, as mentioned above, by applying Little's Law, that lead times are fixed even when the workload varies. So lead time consequences, due to varying workload in the production system, are not really captured in the model. Additionally, the model assumes static average loading with stochastic variations. Nevertheless, Karmarkar (1989a) argues that these restrictions can be overcome and extensions to generate a dynamic model can be created.

3.3 Clearing functions

Karmarkar (1989a) and Srinivasan et al. (1988) independently take up the idea of Graves (1986) and integrate it in a mathematical planning model with so-called clearing factors $\alpha(\text{WIP})$ which are nonlinear functions of the WIP, thus yielding a clearing function of the following form:

$$\textit{Capacity} = \alpha(\text{WIP}) \cdot \text{WIP} = f(\text{WIP})$$

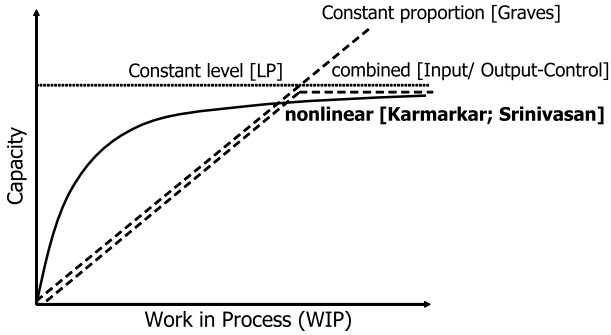

Fig. 1 Diverse forms of clearing functions

where f represents the clearing function that models capacity as a function of the workload. In Asmundsson et al. (2003, 2006a) the clearing factor specifies the fraction of the actual WIP which can be completed, i.e., "cleared", by a resource in a given period of time. (Missbauer (1998) refers to this function as a "utilization function" and Selcuk et al. (2005) as a "saturation function".)

Figure 1 depicts some possible forms of clearing functions where the constant level clearing function corresponds to an upper bound for capacity as mainly employed by LP models. This implies instantaneous production without lead time constraints since production takes place independently of WIP in the production system. The constant proportion clearing function represents the control rule given by Graves (1986) which implies infinite capacity and hence allows for unlimited output. Moreover, in contrast to the nonlinear clearing function of Karmarkar (1989a, 1989b) and Srinivasan et al. (1988), the combined clearing function in some region underestimates and in others overestimates capacity. The nonlinear clearing function relates WIP levels to output and lead times to WIP levels which are influenced by the workload of the production system (see also Karmarkar et al. 1993b) and therefore is able to capture the behavior of LDLT. By applying Little's Law the clearing function can be reinterpreted in terms of lead time or WIP turn, respectively. Additionally, the slope of the clearing function represents the inventory turn with lead time given by the inverse of the slope, which is stated by Karmarkar (1989a).

Next we discuss various clearing function models.

3.3.1 Clearing function model by Srinivasan, Carey and Morton

Srinivasan et al. (1988) construct a nonlinear aggregate discrete-time continuous flow model with variable job arrival times and finite horizon settings. It aims at determining aggregate schedules and optimal WIP inventory levels for each workstation taking into account the effects of WIP levels on (throughput) capacity and on lead time as well as the marginal value of capacity and therefore determining prices for each workload situation of a workstation. A workstation is defined as an aggregation of individual machines, e.g., different machines, flexible machine centers, flow lines, parallel machines or a department. A more detailed lower level model will derive plans for each single machine. The model captures the continuously changing lead times at each workstation, thus providing better lead time estimates for due date setting as well as the changing utilization of workstations due to varying demand patterns reflected in the alternating value of capacity. Finally, an optimal routing scheme for each job through the production network is determined.

There exist two different approaches of production planning models in the literature to date, namely the *monolithic* and the *hierarchical* planning approach. The former includes

both tactical production planning and operational scheduling in one broad (overall) mathematical programming model whereas the latter captures the need for planning control at each organizational level of the production process. Hierarchical systems employ an aggregate model that serves as a framework for the more detailed lower level models providing primal solution information serving as further modeling constraints. For further reading see Bitran et al. (1981, 1982), and Dempster et al. (1981).

The optimization model presented by Srinivasan et al. (1988) resides between tactical planning and detailed scheduling. The production system is viewed as a network (congestion) queuing model with the workstations connected by transportation links where jobs compete for machine capacity (contention system). Capacity is modeled as a concave function of WIP at the workstation or the length of the dynamic line. Additionally, it is assumed that capacity increases with a declining rate, allowing the existence of implicit resource prices at each utilization level. This is not possible when capacity is modeled as an upper bound such that implicit resource prices are only non-zero at 100% utilization.

The objective of the model is to minimize the overall costs of production, namely the holding costs, the processing costs and the tardiness/earliness costs, stated in the following objective function (with T indicating the number of periods, M the number of jobs, and N denoting the number of workstations without considering the final workstation d):

$$\min \quad \sum_{i=1}^{M}\sum_{t=1}^{T}\left[\sum_{j=1}^{N}(\omega_{ijt}W_{ijt}+\phi_{ijt}X_{ijt})+w_i e_i(t)X_{idt}\right]$$

where ϕ_{ijt} is the processing cost factor which only considers variable and labor costs directly attributable to units of each processed job and X_{ijt} the number of jobs i at workstation j at the beginning of the t-th time period in external or internal queues at individual machines or in process (the production costs do not need to vary over time) and $\omega_{ijt}W_{ijt}$ representing the holding costs per job i at workstation j in period t (ω_{ijt} is the holding cost factor and W_{ijt} denotes the WIP level of job i at workstation j in the t-th time period). The last term of the objective function indicates the tardiness or earliness costs, with w_i being a priority weight for job i, $e_i(t)$ representing the tardiness/earliness function for job i completed in period t which includes penalty costs for late delivery, lost sales and lost goodwill and X_{idt} denoting the actual throughput of job i at the final workstation d in the t-th time period. The tardiness/earliness function $e_i(t)$ can be specified in the following form:

$$e_i(t)=\left\{\begin{array}{ll} e_E(\delta_i-t) & \text{if } t<\delta_i \\ 0 & \text{if } t=\delta_i \\ e_L(t-\delta_i) & \text{if } t>\delta_i \end{array}\right\}$$

where $e_E(\cdot)$ and $e_L(\cdot)$ are convex functions of the tardiness and earliness, respectively. The δ_i's represent the due dates for job i. Moreover, quadratic or linear tardiness/earliness functions can be specified. This early/tardy-objective is a non-regular performance measure. Furthermore, the objective function is subject to several constraints. For the first three constraints, the associated Lagrangian multipliers are shown in set braces. They are subsequently used to derive the implicit prices for capacity.

$$\{\psi_{ijt}\}W_{ij,t+1}=W_{ijt}-X_{ijt}+F_{ijt}+\sum_{a\in A_I(i,j)}Y_{iat} \quad \forall i,j,t=1,\ldots,T-1,$$

$$\{\mu_{ijt}\}X_{ijt}=\sum_{a\in A_O(i,j)}Y_{iat} \quad \forall i,j,t=1,\ldots,T-1,$$

$$\{\eta_{jt}\}\sum_{i=1}^{M} X_{ijt} \le f_{jt}(W_{1jt},\ldots,W_{ijt},\ldots,W_{Mjt}) \quad \forall j,t=1,\ldots,T-1,$$

$$W_{ij1} = \dot{W}_{ij1} \quad \forall i,j,$$

$$W_{ijt}, Y_{iat}, F_{ijt} \ge 0 \quad \forall i,j,a,t.$$

The first constraint represents a workload conservation equation for each workstation. It gives the number of jobs i at workstation j in period $t+1$ ($W_{ij,t+1}$) which equals the difference of the number of jobs i of the previous period and the actual throughput of workstation j in period t plus the external inflow F_{ijt} of jobs i at workstation j in period t. The last term sums over inflows $A_I(i,j)$ denoting the transportation arcs pointing into workstation j for job i and of inflows from preceding workstations ($\sum_{a\in A_I(i,j)} Y_{iat}$). Here, Y_{iat} denotes the number of units of jobs i on arc a in period t. The F_{ijt}-values may be seen as "demand" for resource utilization. Hence the authors include them in the non-negativity constraints.

The second constraint states that the throughput of job i at workstation j in period t, the "clearing factor", equals the number of jobs transported to all its subsequent workstations ($\sum_{a\in A_O(i,j)} Y_{iat}$) where $A_O(i,j)$ denotes the set of transportation arcs pointing out of workstation j for job i. This "clearing factor" is in turn restricted by the capacity function $f_{jt}(W_{1jt},\ldots,W_{ijt},\ldots,W_{Mjt})$, represented in the third constraint, where any slack, i.e., $f_{jt}(W_{1jt},\ldots,W_{ijt},\ldots,W_{Mjt}) - \sum_{i\in M} X_{ijt} > 0$, represents the WIP ("number of units of jobs that are held over at workstation j from period t to $t+1$"). In order to prevent congestion downstream in the production chain, holding WIP at a workstation can be efficient even if this causes additional holding costs. The fourth restriction forces the initial loading of each workstation j to be non-zero with a given initial parameter $\dot{W}_{ij1} > 0$. For more details see Srinivasan et al. (1988). If the objective function and the constraints are strictly convex there is a unique optimal solution.

The Karush–Kuhn–Tucker conditions for the optimization problem are, for all $i \in M$ (by f'_{jt} we denote the first derivative of the capacity function):

$$\begin{aligned}
&\text{(K1)} \quad \{W_{ijt} \ge 0\}\omega_{ijt} - \psi_{ij,t-1} - \eta_{jt} f'_{jt} + \psi_{ijt} \ge 0 \quad \forall i,j,t,\\
&\text{(K2)} \quad \{Y_{iat} \ge 0\}\mu_{ijt} - \psi_{iat} \ge 0 \quad \forall i,j,t,a \in A_O(i,j),\\
&\text{(K3)} \quad \{X_{ijt}\}\eta_{jt} - \psi_{ijt} + \mu_{ijt} = \phi_{ijt} \quad \forall i,j,t,\\
&\text{(K4)} \quad \{X_{iat}\}w_i e_i(t) - \psi_{iat} = 0 \quad \forall i,a,t,\\
&\text{(K5)} \quad \eta_{jt} \ge 0 \quad \forall j,t
\end{aligned}$$

together with complementary slackness conditions for all pairs of inequalities in (K1) and (K2). Considering that the function f_{jt} is bounded by the WIP and 0, the variables X_{ijt} and the second constraint can be omitted from the formulation of the problem which leads to the following compact presentation of constraints of the optimization model:

$$W_{ij,t+1} \ge W_{ijt} - f_{jt}(\ldots,W_{ijt},\ldots) + F_{ijt} + \sum_{a\in A_I(i,j)} Y_{iat} \quad \forall i,j,t=1,\ldots,T-1,$$

$$W_{ij1} = \dot{W}_{ij1} \quad \forall i,j,$$

$$W_{ijt}, F_{ijt}, Y_{iat} \ge 0 \quad \forall i,j,t,a.$$

The capacity function f_{jt} is assumed to be a concave non-decreasing function. This is due to the fact that throughput increases as the number of jobs at a workstation increases, but at a declining rate because of congestion.

Srinivasan et al. (1988) discuss different forms of capacity functions and ways to operationalize them. In order to derive implicit capacity prices at all levels of utilization, the Lagrangian multipliers of the mathematical model are employed, where ψ_{ijt} represents the *marginal remaining processing cost* (MRPC) for all jobs i at workstation j in period t, μ_{ijt} denotes the MRPC incurring for job i after its processing by workstation j in period t, and η_{jt} denotes the value of an additional unit of workstation j or the prices of the capacity of workstation j, respectively. When $X_{ijt} > 0$, then the price of capacity of workstation j (η_{jt}) is equal to the sum of marginal production costs and marginal waiting costs caused by job i at workstation j, implied by (K3) of the Karush–Kuhn–Tucker conditions:

$$\eta_{jt} = \phi_{ijt} + \psi_{ijt} - \mu_{ijt}$$

where ϕ_{ijt} is the marginal production cost incurred at workstation j and $\psi_{ijt} - \eta_{ijt}$ the marginal cost entailed by job i waiting at the workstation j in period t. However, when $W_{ijt} > 0$, the price of the capacity of workstation j in period t is stated by the following equation implied by the complementary slackness of (K1):

$$\eta_{jt} = \frac{\omega_{ijt} + (\psi_{ijt} - \psi_{ij,t-1})}{f'_{jt}}$$

where η_{jt} denotes the capacity price of the workstation, ω_{ijt} represents the holding costs, $\psi_{ijt} - \psi_{ij,t-1}$ states the changes in the MRPC and f'_{jt} denotes the slope of the capacity function. When $\psi_{ijt} - \psi_{ij,t-1} > 0$, the MRPC is rising over time and so does congestion. The price of the workstation capacity is bounded from below by the holding costs ω_{ijt}. Furthermore, when the slope of the capacity function approaches zero, $\psi_{ijt} - \psi_{ij,t-1}$ approaches ω_{ijt}. Besides this, the marginal value of WIP (β_{ijt}) is given by the change of the slope of the capacity function and the marginal value of additional capacity given by η_{jt}:

$$\beta_{ijt} = f'_{jt}\eta_{jt} = \omega_{ijt} + (\psi_{ijt} - \psi_{ij,t-1}).$$

This again follows from the complementary slackness (K1) for $W_{ijt} > 0$. It states that the marginal value of WIP is the marginal saving due to marginally increased capacity utilization and therefore lower WIP levels. In periods of rising congestion ($\psi_{ijt} - \psi_{ij,t-1} > 0$) the holding costs bound the marginal value of WIP from below, i.e., $\beta_{ijt} > \omega_{ijt}$ and vice versa for the case of declining congestion.

Srinivasan et al. (1988) examine the existence of a set of equilibrium flows for a given long time horizon which is interesting for input/output-control in order to arrive at steady state production rates that allow for low WIP levels and consequently low production costs. Assuming constant external inflow over a sufficiently long time horizon F_{ijt} (which will be subject to relaxation, thus, the t-index is kept for F_{ijt}) and an objective function only depending on the WIP, there exists a solution to the following problem:

$$\min \quad \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{t=1}^{T}\omega_{ijt} f_{jt}^{-1}\left(\sum_{a\in A_O(i,j)} Y_{iat}\right)$$

subject to:

$$\{\mu_{ijt}\} \sum_{a \in A_O(i,j)} Y_{iat} = F_{ijt} + \sum_{a \in A_I(i,j)} Y_{iat} \quad \forall i, j, t = 1, \ldots, T-1,$$

$$X_{iat} \geq 0 \quad \forall i, j, t, a \in A_O(i, j).$$

If $f_{jt}(\cdot)$ is strictly concave for all j, there exists a unique optimal solution, because then the inverse of $f_{jt}(\cdot)$ is a strictly convex function and the above stated problem would be a minimum convex cost problem. In steady state there is no time variation in the MRPC and consequently the price of workstation j is only a function of the holding costs and the utilization, and the marginal value of WIP is constant and equal to the holding costs. Thus the price of the workstation capacity is given by

$$\eta_{jt} = \frac{\omega_{ijt}}{f'_{jt}}$$

which results from substituting ($\psi_{ijt} = \psi_{ij,t-1}$) in the complementary slackness condition associated with (K1), because in equilibrium the MRPC remain constant over time.

3.3.2 Clearing function model of Karmarkar

Karmarkar (1989a) develops a capacity and release planning model which explicitly takes into account WIP costs and lead time consequences caused by the production system workload. For that purpose it is based on order release and batching models and applies the traditional capacity planning methodology that combines release planning, master scheduling issues and seasonal planning. Additionally, it aims at surmounting the shortcomings of aggregate production planning models like those of Graves (1986), Kekre and Kekre (1985), Karmarkar et al. (1987), and the limitations of input/output-control models. Like Srinivasan et al. (1988), Karmarkar (1989a) uses the nonlinear "clearing function" to represent the output as a function of the average WIP in the production system. The form of the curve is also valid for synchronous deterministic flow lines with batched flows.

The clearing function is now discussed for different types of production systems. For example, examining a synchronous production process with no queuing as well as batching finished and leaving at the end of each production cycle gives the following function for the actual production rate X:

$$X = \frac{Q}{C} = \frac{PQ}{(PS + Q)}.$$

Q denotes the batch size for all products, C stands for the cycle time at a workstation, also denoted as $C = S + Q/P$, P represents the nominal production rate at all workstations, and S denotes the setup time between the batches. Substituting Q in terms of WIP with $Q = W/N$ (where N denotes all workstations) gives:

$$X = \frac{PW}{NPS + W}.$$

This clearing function has the same nonlinear form as can be seen in Fig. 1. Standard results for an $M|M|1$ queuing system give the utilization ρ (with arrival rate λ) as follows:

$$\rho = \lambda X = \frac{X}{P} + \frac{XS}{Q}.$$

This is equal to the formula of the average workload of Zijm and Buitenhek (1996). Furthermore, the average manufacturing lead time $\overline{L}$ is given as follows:

$$\overline{L} = \frac{S + \frac{Q}{P}}{1 - \frac{X}{P} - \frac{XS}{Q}}.$$

In order to keep things simple we ignore the setup times and write the total WIP as $W = \overline{L}X$, substituting $\overline{L} = W/X$:

$$\frac{W}{X} = \frac{\frac{Q}{P}}{1 - \frac{X}{P}}.$$

Solving this relation for X results in the following equation:

$$X = \frac{PW}{Q + W}.$$

For more detail see Karmarkar et al. (1985b) or Karmarkar (1987). This function is clearly similar to the clearing function, although derived in a different context. Furthermore, applying an $M|G|1$ model for an asynchronous multi-product production facility with lead times and WIP as a function of the batching policy, ignoring setup times and assuming that the average queuing time $\overline{Wt}$ is given by the Pollaczek–Khintchine formula leads to the following form:

$$\overline{Wt} = \frac{\sum_{i=1}^{M} \frac{X_i}{Q_i} (\frac{Q_i}{P_i})^2}{2(1 - \sum_{i=1}^{M} \frac{X_i}{P_i})}$$

with i denoting the index for an item or production part. This is equal to the formula of Zijm and Buitenhek (1996). The total average lead time in the system for batches of items i is represented by:

$$\overline{L}_i = X_i + \overline{Wt}$$

while the WIP for this production system is derived as follows:

$$W_i = X_i \left(\overline{Wt} + \frac{Q_i}{P_i} \right).$$

In the case of a multi-product system there exist i equations for i products. Furthermore, the output of each product is affected by the WIP level of all other products. In order to keep things simple, Karmarkar (1989a) considers a discrete period model for a single-product production system to proceed to the dynamic reformulation of the model. Assuming that demand is satisfied, the mathematical formulation is stated as follows for all t:

$$W_{t-1} + R_t = W_t + X_t,$$
$$I_{t-1} + X_t = I_t + D_t,$$
$$X_t = f(W_{t-1}, R_t, P_t).$$

Here, R_t represents the job releases in period t, W_t denotes the WIP at the end of period t carried over to period $t + 1$ and W_{t-1} denotes WIP at the start of period t. The same is valid for I_{t-1} and I_t representing the inventory. Moreover, D_t denotes the demand in period t.

As mentioned above, the output or actual production in period t is a function of WIP, job releases and the maximum production possible in period t denoted by P_t. The first equation states that when WIP levels are small, output cannot exceed the amount in the system. In order to formulate the mathematical model as an optimization model we need to determine the costs of the production system like holding costs for WIP and *finished goods inventory* (FGI), processing costs and penalty costs, etc.

The mathematical programming model is then stated as follows:

$$\min \quad \sum_{t=1}^{T} [\omega_t W_t + \pi_t I_t + \phi_t X_t + \vartheta_t B_t + r_t R_t]$$

subject to:

$$\begin{aligned}
&W_{t-1} + R_t - W_t - X_t = 0,\\
&I_{t-1} + X_t - I_t + B_t = D_t,\\
&X_t - f(W_{t-1}, R_t, P_t) \leq 0,\\
&X_t, R_t, W_t, B_t, I_t \geq 0
\end{aligned}$$

with $\omega_t, \pi_t, \phi_t, \vartheta_t$ and r_t for all periods denoting the WIP holding costs, the FGI holding costs, the variable production costs, the backorder costs due to unsatisfied demand, and the release costs due to, e.g., raw material expenditures. Therefore, B_t denotes the backorders in period t. The first equation represents the mass balance restriction of WIP and releases. The second equation gives information about how demand is met. The third inequality gives the clearing function and the fourth one states the usual non-negativity assumptions.

In contrast to common LP formulations, the effects of capacity loading are captured by the costs of WIP, whose explicit representation is necessitated by the clearing function. Moreover, the resulting release plan will differ from the production plan X, because of lead times which are affected by capacity loading. This interdependence is explicitly accounted for in the model.

The clearing function seems to be the key to explicitly taking into account lead time effects and related dependencies. Its functional form and parameterization has to be studied in more detail. Furthermore, in the literature up to date there only exists one work that analyzes the fitting of the clearing function to empirical data, namely the work of Asmundsson et al. (2003, 2006a). Besides this, the case of multi-product manufacturing systems has to be examined more closely especially concentrating on the form of the clearing function, because output levels may also depend on WIP of each product with WIP being a mix of all products. We will now discuss an approach that pursues this objective.

3.3.3 Clearing function model by Asmundsson, Rardin and Uzsoy

The mathematical programming approach of Asmundsson et al. (2002, 2003, 2006a) models the nonlinear dependency between leadtimes and WIP (workload) by employing a clearing function, too. Special properties of the clearing function allow for formulating an LP version in order to develop a model which remains numerically tractable. In accordance with the procedure of Karmarkar (1989a), Asmundsson et al. (2002, 2003, 2006a) define the performance of a resource (work center) as dependent on the workload. For this reason they use

a $G|G|1$ queuing model:

$$W = \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho^2}{1-\rho} + \rho,$$

$$\rho = \frac{\sqrt{(W+1)^2 + 4W(c^2-1)} - (W+1)}{2(c^2-1)}.$$

The first equation gives the WIP for a resource which is presented with the coefficient of variation for the service time c_a and arrival time c_s. The second equation represents the utilization ρ of a resource as a function of the WIP with c denoting a parameter to be defined. Basically, the authors simplify the parameters from the first equation according to: $c^2 = \frac{c_a^2+c_s^2}{2}$. As mentioned above, batching and lot sizing have an effect on lead times especially when small batches give rise to frequent setup changes leading to time losses for production, lower throughput and eventual starvation of resources on further production stages. Since the formulation given above only includes starvation we can reformulate the resource utilization also accounting for setup time as follows:

$$\frac{Q\bar{\sigma}}{Q\bar{\sigma}+S} \frac{W+1-\sqrt{W^2+2c^2W+1}}{1-c^2}$$

where Q denotes the lot size, $\bar{\sigma}$ the mean service time and S the setup time. The lot size is assumed to be set independently of the system utilization. Here $Q\bar{\sigma}+S$ is the average total time to service a lot. It can be noticed that when setup time S increases the utilization decreases and so does the lot size Q. So we have another nonlinear relationship as lot sizes tend to be increased for high WIP levels in order to clear queuing in front of a resource and vice versa (see Karmarkar 1989a). When setup time increases the clearing function shifts downwards.

The clearing function is assumed to have the characteristics depicted in Fig. 2 where the first inequality states that the clearing function is bounded by zero and $f^{\max}$. The second and the third inequality reflect the property that the clearing function is monotonically increasing in a defined measure of WIP with a declining rate, i.e., f is concave. Besides this, the

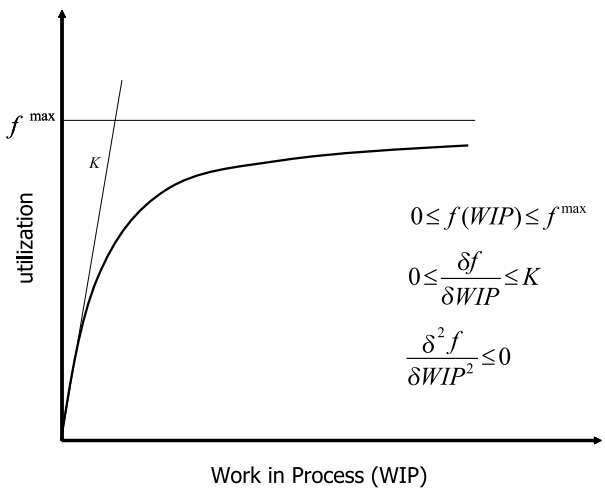

Fig. 2 Characteristics of the clearing function

second inequality imposes the restriction that the slope of the function is bounded from above by K. Furthermore, the function should be well-behaved in the presence of variability and lot sizing as mentioned above and also in Asmundsson et al. (2002, 2006a).

In order to develop the clearing function, there are two methods available in the literature to date, where the first is the analytical derivation from queuing network models and the second an empirical approximation using a functional form which can be fitted to empirical data. Because of the large complexity of practical systems the complete identification of the clearing function will not be possible, so we have to work with approximations. Asmundsson et al. (2002, 2006a) integrate the estimated clearing function in a mathematical programming model where the framework is based on the production model of Hackman and Leachman (1989) with an objective function that minimizes the overall costs. It is assumed that backorders do not occur and that all demand must be met on time. One may assume a representation as a graph with N nodes referring to the workstations and a set A of transportation arcs between these nodes. Predecessors of workstations are indicated by A_I and successors by A_O, respectively. The model is then stated as follows:

$$\min \quad \sum_{i=1}^{M}\sum_{t=1}^{T}\left[\sum_{j=1}^{N}(\phi_{ijt}X_{ijt}+\omega_{ijt}W_{ijt}+\pi_{ijt}I_{ijt}+r_{ijt}R_{ijt})+\sum_{\gamma\in A}\theta_{i\gamma t}\Gamma_{i\gamma t}\right]$$

subject to:

$$W_{ijt}=W_{ij,t-1}-\frac{1}{2}(X_{ijt}+X_{ij,t-1})+R_{ijt}+\sum_{\gamma\in A_I(i,j)}\Gamma_{i\gamma t}\quad \forall i,j,t,$$

$$I_{ijt}=I_{ij,t-1}+\frac{1}{2}(X_{ijt}+X_{ij,t-1})-D_{ijt}-\sum_{\gamma\in A_O(i,j)}\Gamma_{i\gamma t}\quad \forall i,j,t,$$

$$X_{\cdot jt}\le f_{jt}(W_{\cdot jt})\quad \forall j,t,$$

$$X_{ijt},W_{ijt},\Gamma_{i\gamma t},I_{ijt},R_{ijt}\ge 0\quad \forall i,j,t,\gamma$$

where $\phi_{ijt}X_{ijt}$ denotes the costs of the total amount of production over the latter half of period t and the first half of period $t+1$, represented by X_{ijt}, with ϕ_{ijt} referring to the corresponding unit costs of item i at node j in period t. The WIP costs and the FGI costs of item i at workstation j at the end of period t are denoted by $\omega_{ijt}W_{ijt}$ and $\pi_{ijt}I_{ijt}$, respectively, with ω_{ijt} and π_{ijt} being the corresponding unit costs and W_{ijt} and I_{ijt} representing the WIP and the inventory, respectively. Likewise, the costs of releases of raw material of item i at workstation j during period t are represented by $r_{ijt}R_{ijt}$ with unit costs r_{ijt}. The transfer costs of item i on arc γ during period t are given by $\theta_{i\gamma t}\Gamma_{i\gamma t}$ with corresponding unit costs $\theta_{i\gamma t}$. Asmundsson et al. (2002, 2006a) assume these costs as independent from the item.

We should note in passing that holding WIP of one product can create capacity for another, or, equivalently, an average system lead time is maintained by moving one product through very fast and simply holding WIP of the other to create capacity that allows the other product to be processed instantaneously. This is the entire motivation for the partitioned formulation, and is an important limitation that needs to be handled to obtain effective models.

The first two constraints denote the flow conservation for WIP and FGI which is different from classical models since inventory levels at each node in the network are connected

with the throughput rate. Note that because X_{ijt} denotes production that spans the period boundary, the balance equations average the adjacent times.

In contrast to Ettl et al. (2000) the nonlinear dynamic is incorporated in the clearing function in the constraints and thus not included in the objective function. Lead times are calculated using Little's Law:

$$\widehat{L}_{ijt} = \frac{W_{ijt}}{X_{ijt}}$$

where $\widehat{L}_{ijt}$ denotes the expected lead time for the last job of item i which arrived before the end of period t. In order to consider multiple product types with different resource consumption patterns, the lead times for single items i must be derived. For that purpose, FIFO processing is assumed, which results in the following relationship:

$$\frac{X_{ijt}}{X_{\cdot jt}} = \frac{W_{ijt}}{W_{\cdot jt}} \quad \forall i, j, t.$$

Here, $X_{\cdot jt}$ is the production quantity and $W_{\cdot jt}$ the WIP level summarized over all items i. Taking this relation and multiplying the production quantities X_{ijt} by their resource consumption factor ξ_{ijt} which defines the capacity consumption per unit produced for item i at workstation j, we derive a new variable Z_{ijt} of the following form:

$$Z_{ijt} = \frac{\xi_{ijt} X_{ijt}}{X_{\cdot jt}} = \frac{\xi_{ijt} W_{ijt}}{W_{\cdot jt}} \quad \forall i, j, t.$$

By implementing this variable Z_{ijt}, we obtain the clearing function for each item i which is called the partitioned clearing function:

$$\xi_{ijt} X_{ijt} \leq Z_{ijt} f_{jt}\left(\frac{\xi_{ijt} W_{ijt}}{Z_{ijt}}\right) \quad \forall i, j, t$$

with the following properties:

$$\sum_{i=1}^{M} Z_{ijt} = 1 \quad \forall j, t,$$
$$Z_{ijt} \geq 0 \quad \forall i, j, t.$$

The partitioned clearing function is depicted in Fig. 3 (note that the functions are symbolic rather than detailed to scale).

In order to relax the assumed priority rule (FIFO) we only suppose that Z_{ijt} satisfies the properties stated above, but has an arbitrary functional form. With this formulation Asmundsson et al. (2002, 2006a) succeed in integrating the nonlinear relationship between WIP and lead times in a mathematical programming model. The second goal is to transform this model in a tractable form which allows for dealing even with relatively large planning problems. For that purpose, an LP formulation is used that represents the partitioned clearing function through a set of linear constraints. To be more precise, the clearing function is approximated by the minimum of straight lines which is possible because of its concavity as shown in Fig. 4:

$$f_{jt}(W_{\cdot jt}) = \min_{c}\{\alpha_{jt}^{c} W_{\cdot jt} + \beta_{jt}^{c}\} \quad \forall j, t, c.$$

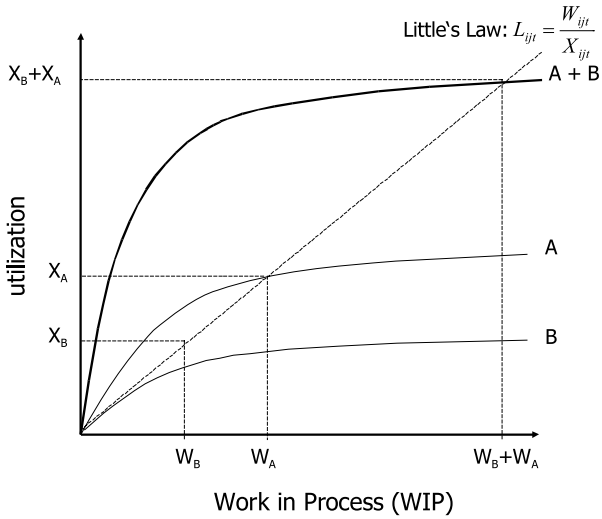

Fig. 3 Partitioned clearing function for products A and B

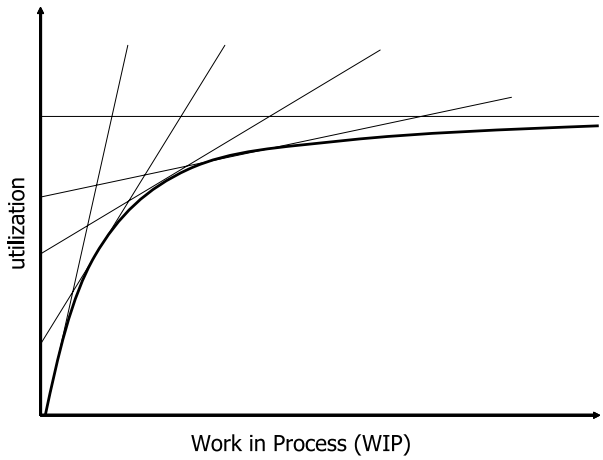

Fig. 4 Linearization of the clearing function

The individual lines of the items are denoted by the index c. The α_{jt}^c and β_{jt}^c coefficients represent the slope and the intersection of line c with the y-axis. The capacity splitting (sharing) across the items appears in sharing the β_{jt}^c coefficients. Applying this formulation to the partitioned clearing function leads to the following form:

$$\begin{aligned} Z_{ijt} f_{jt}\left(\frac{\xi_{ijt} W_{ijt}}{Z_{ijt}}\right) &= Z_{ijt} \cdot \min_c \left\{\alpha_{jt}^c \frac{\xi_{ijt} W_{ijt}}{Z_{ijt}} + \beta_{jt}^c\right\} \\ &= \min_c \{\alpha_{jt}^c \xi_{ijt} W_{ijt} + \beta_{jt}^c Z_{ijt}\}. \end{aligned}$$

Replacing the former capacity constraint of the original nonlinear mathematical programming model with nonlinear lead time and capacity dynamics gives the complete linearized

formulation:

$$\min \quad \sum_{i=1}^{M}\sum_{t=1}^{T}\left[\sum_{j=1}^{N}(\phi_{ijt}X_{ijt}+\omega_{ijt}W_{ijt}+\pi_{ijt}I_{ijt}+r_{ijt}R_{ijt})+\sum_{\gamma\in A}\theta_{i\gamma t}\Gamma_{i\gamma t}\right]$$

subject to:

$$W_{ijt}=W_{ij,t-1}-\frac{1}{2}(X_{ijt}+X_{ij,t-1})+R_{ijt}+\sum_{\gamma\in A_I(i,j)}\Gamma_{i\gamma t}\quad \forall i,j,t,$$

$$I_{ijt}=I_{ij,t-1}+\frac{1}{2}(X_{ijt}+X_{ij,t-1})-D_{ijt}-\sum_{\gamma\in A_O(i,j)}\Gamma_{i\gamma t}\quad \forall i,j,t,$$

$$\xi_{ijt}X_{ijt}\le\alpha^c_{jt}\xi_{ijt}W_{ijt}+Z_{ijt}\beta^c_{jt}\quad \forall i,j,t,c,$$

$$\sum_{i=1}^{M}Z_{ijt}=1\quad \forall j,t,$$

$$X_{ijt},W_{ijt},\Gamma_{i\gamma t},I_{ijt},R_{ijt},Z_{ijt}\ge 0\quad \forall i,j,t,\gamma.$$

For the purpose of examining the relevance and the performance of the mathematical approach, Asmundsson et al. (2002, 2006a) consider an example of a simple single stage system with three products which gives very good results. Additionally, they analyze the sensitivity of the estimated clearing function to diverse shop floor scheduling algorithms, different demand patterns and techniques of production planning using a simulation model of the respective production system. They find that the clearing function model reflects the characteristics and capabilities of the production system better than models using fixed planned lead times (like mrp) and derives realistic and robust plans with better on time delivery performance, lower WIP and system inventory. In addition, the model captures the effects of congestion on lead times and WIP and therefore the fundamental trade-off between anticipated production to account for possible peaks in demand and just in time production to avoid higher costs due to preventable FGI. Releases generated by this partial clearing function model are smoother and lead to enhanced lead time performance. This is also experienced by Asmundsson et al. (2006b) who apply the partitioned clearing function model to a semiconductor manufacturing environment which consists of a wafer fabrication with eleven machines performing related processes, three product families and different routings. Further analysis of different aspects of this mathematical programming model is needed, e.g., the extension of the approach to systems of multiple production stages. Orcun et al. (2006b) integrate the clearing function model in a simple system dynamics simulation and show that it captures the nonlinear changes in system performance, especially at high utilization, in consistence with results from queuing theory.

Nevertheless, interactions between clearing functions and shop floor execution systems such as the dependency of LDLT on the various priority rules need further analysis. This is another circularity, because clearing functions are dependent on the employed scheduling policy and therefore on the result of the scheduling algorithm. Moreover, as pointed out by Asmundsson et al. (2002, 2006a), the schedules are dependent on the release schedule and consequently on the planning algorithm.

3.3.4 Clearing function model of Hwang and Uzsoy

Hwang and Uzsoy (2004) combine the work of Karmarkar (1987) and Asmundsson et al. (2002, 2006a) and add lot sizing in order to show how small or large lot sizes influence the resulting production plans. They present a single-product dynamic lot sizing model which takes into account WIP and congestion using queuing models such as those presented by Karmarkar (1987, 1993b). They develop a clearing function which captures the dependency of the expected throughput of a single-stage production system closely related to the classical Wagner–Whitin model (see Wagner and Whitin 1958) including setups, expected WIP levels and lot sizes which is then integrated into a dynamic lot sizing model. Results demonstrate that their proposed model provides significantly more realistic performance and, therefore, production plans, than models ignoring the relationships between lead times, workload, production mix (setups), throughput and lot sizes. The importance of lot sizing is emphasized especially in the context of capacity planning and production coordination, because different product types compete for scarce resources and require significant setup times (see Hwang and Uzsoy 2004 and the references therein).

In general, large lot sizes allow for high resource utilization due to the reduced number of setups, but result in increased FGI and may lead to delays of important jobs waiting to be processed. However, small batch sizes minimize WIP levels, but necessitate frequent setups impeding the usage of resources for production, creating congestion and high WIP levels in front of the resource resulting in long lead times. So, lead times depend not only on the workload of the system, but also on the work assignments to resources which, in turn, are determined by the planning models and, therefore, representing a further planning circularity which must be accounted for. Given a WIP level, their model determines the lot size which maximizes system throughput.

The clearing function aims at minimizing the sum of release costs, production costs, WIP and FGI holding costs over the planning horizon. It is derived with the use of a $M|G|1$ queuing model (see Buzacott and Shantikumar 1993), assuming a Poisson process as arrival rate of the lots and a general distribution of the processing times at the resource following a *first come first serve* (FCFS) process order for the products (jobs). The so-called *lot sizing clearing function* (LCF) which is derived using Little's Law $X = W/L$ (expected throughput rate equals the expected WIP over the expected lead time) is stated as follows:

$$f(Q, W) = \frac{(Q + W - \sqrt{Q^2 + W^2 + 2QWc^2})}{(1 - c^2)\hat{p}^b}$$

where Q denotes the lot size and $\hat{p}^b$ the expected processing time of a batch. The coefficient of variation of the system processing time is given by c (for more details on the derivation of the LCF see Hwang and Uzsoy 2004). The LCF is integrated in a single-stage manufacturing system where the decision variables are the release quantity denoted by R_t, the lot size Q_t, the production quantity Y_t ("number of lots produced in period t") and the level of WIP denoted by W_t and FGI denoted by I_t. No backorders are allowed and the variables are assumed to be continuous. The objective function does not contain setup costs, because these are captured in the LCF by showing directly their influencing effects on capacity. The complete dynamic lot sizing model is stated as follows:

$$\min \quad \sum_{t=1}^{T} [\phi_t Q_t Y_t + \omega_t W_t + \pi_t I_t + r_t R_t]$$

subject to:

$$
\begin{aligned}
&W_t = W_{t-1} + R_t - Q_t Y_t \quad \forall t,\\
&I_t = I_{t-1} + Q_t Y_t - D_t \quad \forall t,\\
&Q_t Y_t \le f(Q_t, W_t) \quad \forall t,\\
&Q_t > \frac{S_t D_t}{(1 - \frac{D_t}{P})} \quad \forall t,\\
&R_t, Q_t, Y_t, W_t, I_t \ge 0.
\end{aligned}
$$

The first constraint represents the WIP balance, the second the FGI balance (with the demand in period t, D_t), and the third reveals the LCF. The fourth constraint represents a lower bound on the lot size in order to ensure system stability by bounding utilization below 1 where S_t denotes the setup time per lot and P gives the production rate of the manufacturing system. The model is reformulated by setting $Q_t Y_t = W_{t-1} - W_t + R_t$:

$$
\min \quad \sum_{t=1}^{T} [\phi_{t-1} W_{t-1} + (\omega_t - \phi_t) W_t + \pi_t I_t + (r_t + \phi_t) R_t]
$$

subject to:

$$
\begin{aligned}
&I_t + W_t = I_{t-1} + W_{t-1} + R_t - D_t \quad \forall t,\\
&W_{t-1} - W_t + R_t \le f(Q_t, W_t) \quad \forall t,\\
&Q_t > \frac{S_t D_t}{(1 - \frac{D_t}{P})} \quad \forall t,\\
&R_t, Q_t, W_t, I_t \ge 0 \quad \forall t.
\end{aligned}
$$

This has the advantage that the Y_t are eliminated from the objective function. Furthermore, the second constraint now combines the WIP and the FGI constraint which reduces the number of constraints.

Future research concentrates on the extension of the model to multiple products and to investigate how capacity is distributed among them when they are subject to different demand, manufacturing requirements and variability.

3.4 The capacity expansion problem with clearing functions of Kim and Uzsoy

Kim and Uzsoy (2006a) develop a capacity expansion approach considering congestion-prone production resources. Capacity expansion problems regard decisions on size and timing of capacity additions and deletions in order to achieve minimum total costs. Their approach is motivated by the fact that current work on this kind of problem does not include the relationship between demand, capacity, and operational performance measures, e.g., lead times and/or WIP levels. In order to fill this gap they employ a concave clearing function in a multi-stage multi-server capacity expansion approach and solve it with a column generation heuristic.

A heuristic is developed to find an initial feasible solution and to build a restricted master problem. An algorithm for the single stage case is used to solve sub-problems which generate new columns. The upper bound derived by the column generation approach is better than the one derived by the Lagrangian procedures.

The capacity expansion problem including a target WIP level is given as follows:

$$\min \quad \sum_{t=1}^{T}\sum_{l=1}^{\Lambda} H_{lt}(x_{lt}) + \sum_{t=1}^{T}\sum_{l=1}^{\Lambda} M_{lt}(j_{lt}) - \sum_{t=1}^{T}\sum_{l=1}^{\Lambda} G_{lt}(y_{lt})$$

subject to:

$$\begin{aligned}
&j_{lt} = j_{l,t-1} + x_{lt} - y_{lt} \quad \forall l, t,\\
&\sum_{l=1}^{\Lambda} W_{lt} \le \mathrm{WIP}_t \quad \forall t,\\
&j_{lt}, x_{lt}, y_{lt} \text{ integer} \quad \forall l, t,\\
&f_{lt}(W_{lt}, j_{lt}) \ge D_{lt} \quad \forall l, t,
\end{aligned}$$

with

$$f_{lt}(W_{lt}, j_{lt}) = \frac{[W_{lt} + j_{lt} - \sqrt{(W_{lt} + j_{lt})^2 + (2j_{lt} - c_{slt})(c_{alt} - 2W_{lt})}]j_{lt}X_{lt}}{(2j_{lt} - c_{alt})}.$$

The objective function minimizes the sum of investment, maintenance, and salvage costs across all stages $l = 1, \ldots, \Lambda$ over the planning horizon T where H_{lt} is the investment cost function at stage l in time period t and x_{lt} the number of newly introduced machines. The cost function for maintenance at stage l in time period t is denoted by M_{lt} and j_{lt} is the number of available machines, respectively. Moreover, y_{lt} is the number of salvaged machines on stage l at time period t and G_{lt} is the cost function of salvages at stage l in time period t. The first constraint represents the capacity balance equation for all stages and all time periods. Furthermore, the second constraint ensures that the total WIP at stage l in time period t does not exceed the target WIP level denoted by WIP_t. The fourth constraint is the clearing function which relates the WIP (W_{lt}) to the number of available machines j_{lt} at stage l in time period t. Parameter X_{lt} denotes the throughput rate. The clearing function is derived for a $GI/G/m$ system with the WIP equation expressed by the expected queue length:

$$W = \frac{c_a + \rho^2 c_s}{2(1-\rho)} + m\rho$$

where m is the standard notation for the number of servers in queuing theory. Thus, for the sake of clarity for those familiar with queuing theory, we stick to this notation. The variable ρ is the utilization or traffic intensity defined as: $\rho = \lambda/(m\mu)$ and c_a and c_s denote the squared coefficient of variation of the interarrival times and service times. It is important to note that the system is assumed to be in steady state.

Let j_l be a T component vector of the number of available machines at stage l in periods $t = 1, \ldots, T$ and S_l is a set of j_t such that

$$S_l = \{j_l \mid f_{lt}(W_{lt},\ j_{lt}) \ge D_{lt},\ j_{lt} = j_{l,t-1} + x_{lt} - y_{lt},\ j_{lt}, x_{lt}, y_{lt} \text{ integer},\ \forall t\}.$$

Furthermore, denoting $\Omega_l(j_l)$ as the total maintenance cost of the capacity profile defined by j_l over all planning periods T, i.e.,

$$\Omega_l(j_l) = \sum_{t=1}^{T}(H_{lt}(x_{lt}) + M_{lt}(j_{lt}) - G_{lt}(y_{lt}))$$

where the number of added and salvaged machines can be calculated from $j_{lt} = j_{l,t-1} + x_{lt} - y_{lt}$ and $x_{lt} \cdot y_{lt} = 0$, the original capacity expansion problem is reformulated as follows:

$$\min \quad \sum_{l=1}^{\Lambda} \Omega_l(j_l)$$

subject to:

$$\sum_{l=1}^{\Lambda} W_l(j_l) \leq \text{WIP},$$
$$j_l \in S_l \quad \forall l$$

where $W_l(j_l)$ is a T component vector that gives the minimum WIP corresponding to each j_{lt} that can be calculated using the WIP equation for a $GI/G/m$ queuing system. Introducing an index p for a capacity configuration plan and assuming that S_l contains a large but finite set of j_l vectors, the set of feasible capacity configurations is given as follows:

$$S_l = \left\{ j_l \,\middle|\, j_l = \sum_{p=1}^{p_l} j_{lp}\delta_{lp},\ \sum_{p=1}^{p_l} \delta_{lp} = 1,\ \delta_{lp} \in \{0, 1\},\ p = 1, \ldots, P_l \right\}.$$

The variable δ_{lp} is a boolean variable which becomes 1 if plan p is selected for stage l and 0 otherwise. Due to the fact that enumerating all columns in an LP model is often impossible, a restricted LP model is derived:

$$\min \quad \sum_{l=1}^{\Lambda} \sum_{p=1}^{P_l} \widetilde{\Omega}_l(j_{lp})\tilde{\delta}_{lp}$$

subject to:

$$\{\psi\} \sum_{l=1}^{\Lambda} \sum_{p=1}^{P_l} \widetilde{W}_l(j_{lp})\tilde{\delta}_{lp} \leq \text{WIP},$$
$$\{\eta\} \sum_{p=1}^{P_l} \tilde{\delta}_{lp} = 1 \quad \forall l,$$
$$\tilde{\delta}_{lp} \geq 0 \quad \forall l, p.$$

In order to obtain an initial feasible solution the dual variables ψ and η of the last two equations are derived. They are used to see if an improvement of the solution is possible by seeking a stage j such that

$$\min_p(\Omega_l(j_{lp}) - \psi W_l(j_{lp}) - \eta_l) < 0, \quad j_{lp} \in S_l$$

which is the sub-problem mentioned above. If there does not exist such a stage l then the derived solution is optimal. Nevertheless, there may be a problem if the solution contains fractions of machines. This can be overcome by executing the single stage algorithm given in Kim and Uzsoy (2006b) which is the basic model to the multi-stage model given here.

3.5 The integrated production planning and pricing model of Upasani and Uzsoy

Upasani and Uzsoy (2005) study integrated marketing-production models and give an overview on recent research. In this regard, they emphasize the lack of combined marketing/price-production models taking into account market sensitivities to price and lead times. They propose an integrated production planning and pricing model in a congestion-prone and capacity constrained environment including clearing functions. Although the model has some drawbacks, it is the first combined price-production model that accounts for the relationship between WIP levels, lead times, and price decisions. For instance, some drawbacks are that the model does not account for the transition behavior from one equilibrium state to another and that the marketing sub-model includes very simple price-demand relations. Furthermore, a price monopolist is assumed who can decide on price and sales quantities without external influencing factors which, obviously, is not realistic. Moreover, the model assumes perfect information and orders having the same size and processing time.

They incorporate the clearing function as given in Karmarkar (1989a):

$$f(W_t) = \frac{K_1 W_t}{p(K_2 + W_t)}$$

where K_1 is the theoretical maximum capacity and K_2 the slope or curvation of the clearing function. The parameter p denotes the raw processing time of one item (order, job). The demand function includes the maximum possible demand $\check{D}$ reduced by a term for price sensitivity $a_t P_t$ where a_t is the price sensitivity factor in demand and P_t the price in period t. Moreover, the demand function is reduced by a term stating lead time sensitivity $b_t L_t$ where b_t denotes the lead time sensitivity factor and L_t the lead time. Employing Little's Law the lead time L can be expressed in terms of WIP and throughput: $L_t = W_t / X_t$. Thus the demand function is:

$$D_t = \check{D} - a_t P_t - b_t \left(\frac{W_t}{X_t} \right).$$

As done in Asmundsson et al. (2002, 2003, 2006a) the production quantity is approximated by the average of the production quantities in two consecutive periods, viz. by $1/2(X_t + X_{t+1})$. The complete pricing-production model is stated as follows:

$$\max \quad \sum_{t=1}^{T} \left(P_t \left[\check{D} - a_t P_t - b_t \left(\frac{2W_t}{X_t + X_{t+1}} \right) \right] - r_t R_t - \pi_t I_t - w_t W_t \right)$$

subject to:

$$\begin{aligned}
&W_t = W_{t-1} - \frac{1}{2}(X_t + X_{t+1}) + R_t \quad \forall t, \\
&I_t = I_{t-1} + X_t - \left(\check{D} - a_t P_t - b_t \left(\frac{2W_t}{X_t + X_{t+1}} \right) \right) \quad \forall t, \\
&X_t \le \frac{K_1 W_t}{p(K_2 + W_t)} \quad \forall t, \\
&\check{D} - a_t P_t - b_t \left(\frac{2W_t}{X_t + X_{t+1}} \right) \ge 0 \quad \forall t, \\
&P_t, X_t, I_t, W_t, R_t \ge 0 \quad \forall t.
\end{aligned}$$

Different to all other models this approach maximizes profits which is stated by the difference of total revenue ($P_t D$) in each time period t and the operating costs including raw material costs, inventory holding costs, and WIP inventory costs ($r_t R_t + \pi_t I_t + w_t W_t$).

Upasani and Uzsoy (2005) compare this model to a fixed lead times model and present numerical examples. They reveal that price manipulations are easier in the short term for sales control than for lead time control due to the fact that lead times depend on WIP levels and the throughput rate that cannot be easily changed in a short time period.

3.6 The production planning and order release concept of Missbauer

Missbauer (1998) emphasizes the need to create a structure for production planning and scheduling in which all related subproblems are integrated, e.g., aggregate production planning, capacity planning, lot sizing, scheduling, etc., but which does not imply the complexities which arise from comprehensive models such as, e.g., problems of data procurement, limited computational storage space, too long CPU times for calculation, etc. He concentrates on the *hierarchical production planning concept* where all stages or areas of the production planning and scheduling process are included and can be integrated in traditional production planning (mrp) systems, but nevertheless leaves room for decisions on implementations of different planning approaches like KANBAN, OPT or priorities on the various planning stages or areas. The focus is on a multi-stage, multi-product production system which is the underlying concept in most production planning and scheduling systems prevailing in practice. The planning process is executed in a decentralized manner with two planning levels: one central, over arching planning unit, called "production unit control" as well as decentralized production units, called "goods flow control" by Bertrand and Wijngaard (1986).

The production control unit defines the sequence of jobs and their due dates and the moment at which the material flow control is passed on the decentralized goods flow units (order release process). In order to supervise and regulate the actual production situation at the operating level of the production process, measures like due date fulfillment, lead times, WIP, utilization levels etc. are reported from the goods flow control to the central production unit control. The organizational type of the goods flow control, i.e., if it is an internal department or a supplier, is of secondary importance so that the concept is also applicable and extendable to supply chain networks. Therefore, in order to control the production system status of the goods flow units it is necessary for the production unit control to rely on information about the characteristics and system dynamics inside and between the decentralized goods flow units and also on information about the functional dynamics of the control measures like, e.g., the relationship between lead times, WIP and capacity. These relationships and dynamics specify the system states that can be reached. Furthermore, they are easily depicted graphically and used as a basis for models trying to anticipate future system states from current states. Nevertheless, it is even more important to understand the structure and order of decision events at the central planning stage, the production unit control.

Missbauer (1998) distinguishes four different hierarchical planning levels, namely the aggregate production planning level, the master production schedule, the planning of the material flow between the decentralized goods flow units (determination of lot sizes, due dates, safety inventory and buffer times) and the order release control level. In order to achieve coordination, the models of the upper planning stages take into account the dynamics of the subordinate planning models. Moreover, a capacity oriented planning approach is used which, in contrast to the product oriented approach, does not plan on the basis of the specific product parts, but considers aggregate product units as capacity units. This is

because in the long run of the planning horizon, i.e., after a completed production cycle, the stored capacity is relevant for the planning process whereas it is less important which and how many specific parts are at stock. This becomes important in the short horizon, detailed production planning stage. The capacity oriented planning approach divides the order release process into two decision problems: assuming a certain value of the aggregate demand on end products (expressed in capacity units), the first decision problem concerns the time to release jobs (expressed in quantity of working hours per period or capacity units) in order to achieve and maintain a certain system state including further decisions about the system performance as well as aggregate FGI inventories. This determines the budget of capacity (expressed in working hours). The second decision problem is about choosing specific orders and their release times based on product oriented criteria. In doing so the capacity oriented approach is very useful, because it eliminates the difficulties arising from uncertain demand estimates for specific products underlying the product oriented approach. It only needs information about demand estimates aggregated across all products and their capacity utilization in order to plan capacity utilization in the long run. The detailed scheduling for specific products takes place on subsequent production planning stages which not only reduces the information requirements but also the complexity of the decision models.

Missbauer (1998) further distinguishes between models of aggregate production (and release) planning with constant or variable planned workload (non-stationary system behavior). Because we are interested in considerations of LDLT in the aggregate production planning process and the effects on them caused by demand peaks or machine breakdowns, we will concentrate on models with variable (planned) workload. The model of Missbauer (1998) presented here is based on a job shop production system which depicts the material flow as a continuous or fluid flow. The planning problem is further divided into two partial problems where first the material flow between the production unit and the goods flow units is modeled and secondly the flow of jobs (work) between the single goods flow units.

The partial formulation of the problem of the material flow between the decentralized production units mainly consists of the inventory equation which defines the WIP inventory level at the production units as well as the FGI inventory level. Furthermore, the dependencies of these inventory levels on the demand quantity of products, on the performance of the system in one period and on the order release quantities are modeled. The basic assumptions are that the production system consists of N production units ($j = 1, \ldots, N$; index v denotes the preceding production units) with M products to be manufactured ($i = 1, \ldots, M$). Moreover, each product i contains $K(i)$ operations (production processes with $k = 1, \ldots, K(i)$) where the technical machine sequence as well as the jobs and completion due dates for each product are given. In addition, the demand for product i in period t is measured in work hours and denoted by D_{it}. Jobs are arbitrarily divisible and setup times proportional so that any job of any product can be processed at any time. This is a critical assumption, because it does not reflect reality and leads to impractical anticipated system states. We will refer to that later on.

In front of each production unit there is a stock of WIP which is measured in work hours or capacity units. The WIP inventory equation has the following form:

$$W_{ijt} = W_{ij,t-1} + \sum_{v=1}^{N} \sum_{k=2}^{K(i)} \delta_{ijk} \delta_{iv,k-1} X_{ivt} \frac{p_{ij}}{p_{iv}} + R_{it} \delta_{ij1} \frac{p_{ij}}{p_i} - X_{ijt} \quad \forall i, j, t$$

where W_{ijt} denotes the WIP of product i at production unit j at the end of period t and δ_{ijk} is a Boolean variable which is 1 if operation k of product i is executed at production unit j and 0 otherwise. $\delta_{iv,k-1}$ is a Boolean variable denoting if a preceding operation $k-1$ has

been executed at production unit v. By X_{ijt} we can denote the throughput of item i in workstation j during period t. With this X_{ivt} may represent the performance of the preceding production unit v in period t attributable to product i. p_{ij} denotes the processing time of a job of product i at production unit j and p_i the aggregation across all production units j. Additionally, R_{it} is the order release quantity of product i in period t. The second term of the equation reflects an increase of WIP from all predecessors v. The third term represents an increase of WIP at the production system starting point or after order release and the fourth one is the work performed on product i which can be considered as the cleared WIP of product i at production unit j. The FGI inventory equation has the following form:

$$I_{it} = I_{i,t-1} + \sum_{j=1}^{N} \delta_{ij,K(i)} X_{ijt} \frac{p_{ij}}{p_i} - D_{it} \quad \forall i, t$$

where I_{it} (FGI inventory) consists of the FGI inventory of the previous period, the increase added in period t (measured at the end of period t where $\delta_{ij,K(i)}$ denotes the last operation of product i if $K(i)$ is executed on production unit j) reduced by the demand for product i in period t. This formulation specifies that the increase from the previous production units in period t only depends on the performance of these production units in this period and consequently does not assume a production rhythm. Instead, it only supposes a movement at the transition times from one period to another, which implies that release jobs could pass an infinite number of production units and which is similar to overlapping production with infinite small transfer lot sizes. It means that the delay caused by processing times of jobs is not yet modeled. Missbauer (1998) states that this feature is easily added.

It is important to note that the model overestimates the system performance due to its underlying assumptions, especially when direct inventory levels are small. This is due to the assumption that jobs are infinitely divisible and setups are independent. So system performance quantities X_{ijt} depend on the progress of production and thus on the course of inventory levels or system workload on time, assuming given order release, and consequently on the utilization function or clearing function. The determination of the clearing function is part of the formulation of the second partial problem. The second partial problem highlights the relationship between the performance quantities X_{ijt} and the clearing function. Moreover, in order to achieve good production system state predictions and therefore realistic system performance, the influence of the discontinuities of material workflow due to non-divisible jobs needs to be reflected, assuming the estimated arriving work or WIP as the maximal performance in one period which can be distributed to specific products. The clearing function is given as $f(\widehat{W}_{ijt})$ where the term $\widehat{W}_{ijt}$ denotes the estimated inflow of products i at production unit j in period t and is comprised of the estimated WIP level of products i at the beginning of period t ($\widehat{W}_{ij,t-1}$) and the estimated increase ($\widehat{X}_{ijt}$), so that $\widehat{W}_{ijt}$ is given by

$$\widehat{W}_{ijt} = \widehat{W}_{ij,t-1} + \widehat{X}_{ijt}.$$

An $M|M|1$ model is used to derive the form of the clearing function. In the steady state the average performance in one period t equals the average increase of work ρ which is the degree of utilization ("traffic intensity") of one production unit, so that $\widehat{W}_{ijt} = \rho_j \tau_t$ where τ_t denotes the length of period t. The average utilization or workload $\overline{W}_{ijt}$ is the sum of the average inventory represented by $\overline{I}_{ijt} = \overline{I} = \frac{\rho}{1-\rho}$ and of the average performance $\overline{X}_{ijt}$ with $\overline{I}$ denoting the average number of jobs in the queue:

$$\overline{W}_{ijt} = \frac{\rho}{1-\rho} + \overline{X}_{ijt}.$$

The clearing function (in disaggregated form) then has the following form:

$$f(\widehat{W}_{ijt}) = \frac{1}{2}\Big(1 + \widehat{W}_{ijt} + \tau_t - \sqrt{1 + 2\widehat{W}_{ijt} + (\widehat{W}_{ijt})^2 + 2\tau_t - 2\widehat{W}_{ijt}\tau_t + \tau_t^2}\Big).$$

The performance in one period t can be represented as a function of the utilization or workload in one period by generalizing the above stated formulation for which the statement $X_{ijt} = \beta_j \overline{W}_{ijt}/(\overline{W}_{ijt} + \alpha_j)$ is valid with α_j giving the slope of the utilization function and β_j being a parameter to be estimated. We define the utilization or workload for the stationary case as $W_{ijt} = \overline{W}_{ijt} + X_{ijt}$:

$$f(W_{ijt}) = \frac{1}{2}\Big[\beta_j + \alpha_j + W_{ijt} - \sqrt{\beta_j^2 + 2\beta_j\alpha_j + \alpha_j^2 - 2\beta_j W_{ijt} + 2\alpha_j W_{ijt} + (W_{ijt})^2}\Big]$$

where the parameters β_j and α_j need to be estimated. It should be mentioned that the functional relationship between the performance in one period and the utilization of one period further depends on the composition of the expected workload, i.e., new work arrivals and workload that remain in the production system for a longer time. The clearing function makes implicit assumptions on this relationship that may not be valid in case of non-stationary systems. Therefore, this procedure is only an approximation. The entire disaggregated model now has the following form:

$$\min \quad \sum_{i=1}^{M}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}\omega_{ijt}W_{ijt} + \pi_{it}I_{it}\right)$$

subject to:

$$W_{ijt} = W_{ij,t-1} + \sum_{v=1}^{N}\sum_{k=2}^{K(i)}\delta_{ijk}\delta_{iv,k-1}X_{ivt}\frac{p_{ij}}{p_{iv}} + R_{it}\delta_{ij1}\frac{p_{ij}}{p_i} - X_{ijt} \quad \forall i, j, t,$$

$$I_{it} = I_{i,t-1} + \sum_{j=1}^{N}\delta_{ij,K(i)}X_{ijt}\frac{p_{ij}}{p_i} - D_{it} \quad \forall i, t,$$

$$\sum_{i=1}^{M}X_{ijt} \le \alpha_j^c + \beta_j^c\sum_{i=1}^{M}(W_{ijt} + X_{ijt}) \quad \forall t, j, c,$$

$$X_{ijt}, I_{it}, R_{it}, W_{ijt} \ge 0.$$

The objective function specifies minimization of the costs for inventory of WIP and FGI. The clearing function is given by the third restriction with α_j^c and β_j^c representing the position of the clearing function and the c's denoting the degree of linearization. As a result, the model determines the flow of jobs (materials) through the production system and, consequently, the order release R_{it}. Furthermore, it decides which jobs of the waiting W_{ijt} are to be processed next and therefore determines the production schedule for each workstation j. The order release priority rule is not explicitly taken into account, but would be determined implicitly by the objective function where adherence to due dates (delivery dates) would be the restriction.

In order to reduce the size of the optimization problem it is possible to aggregate over products and capacity which means that products with a similar machine (processing) sequence and similar processing times p_{ij} can be grouped in product families and production

units can be aggregated with the requirement that the dynamic behavior is not distorted. For more details see Missbauer (1998).

It is important to note that the question how far to aggregate the products and which products to group is very difficult and widely discussed in the literature without an unambiguous answer. On the one hand aggregation leads to reduction of complexity and model size, but on the other, it can impede the achievement of optimal results. Besides this, the time delays caused by the transition from production units j to N are defined for product i, resulting in a time delay function.

Additionally, the problem of the delay caused by the fact that jobs are not arbitrarily divisible is integrated in the model in order to arrive at the final formulation which is a network model of bottlenecks. The advantages of the final model are that the clearing function and the delay function between the production units give a good description of the characteristics of real production using parameters which are directly observable and therefore empirically determinable. Nevertheless, the clearing function used here is derived from characteristics of the stationary system case. Missbauer (1998) reveals that an analysis of non-stationary behavior of queuing systems and their integration in the models of the production units can be an improvement and, consequently, represents an objective for further research. This is pursued by Missbauer (2006a) and Missbauer (2006b) which is discussed in the following.

Moreover, Missbauer (1998) states that it would be interesting to know if the use of nonlinear programming could be a better representation of the system behavior, especially for considering the relationship between lead times and workload. Besides this, the representation of the production system structure as one bottleneck with the other production units being modeled as delaying functions is interesting, because it would lead to a great reduction of complexity concentrating on only a small number of production units and aggregated products. This is especially interesting in production planning where multiple production units and stages have to be considered simultaneously.

Missbauer (2006b) examines these two research topics suggesting a model structure as described above. He presents the production system structure as a bottleneck with the other workstations being modeled as delaying functions. In this regard, the WIP inventory equation and the FGI inventory equation are reformulated as follows:

$$W_{ijt} = W_{i,j,t-1} + \sum_{v=1}^{N} \sum_{\tau=0}^{\infty} X_{iv,t-\tau} \overline{X}_{ijv} \xi_{ijv\tau} + \sum_{\tau=0}^{\infty} R_{i,t-\tau} \overline{X}_{i0j} \xi_{i0j\tau} - X_{ijt} \quad \forall i, j, t,$$

$$I_{it} = I_{i,t-1} + \sum_{j=1}^{N} \sum_{\tau=0}^{\infty} X_{ij,t-\tau} \overline{X}_{ij0} \xi_{ij0\tau} - D_{it} \quad \forall i, t.$$

The WIP and FGI level equations are expressed through the bottleneck in relation to the precedent workstations v where τ represents the time which is needed by the output of workstation v to arrive at workstation j. The parameter $\overline{X}_{ijv}$ denotes the average amount of work of product group i arriving at workstation j from workstation v. $\xi_{ijv\tau}$ denotes the proportion of output of product group i from workstation v to workstation j that arrives at j τ periods after completion at v. The input sources of the bottleneck station are now the output of the other workstations and the releases of orders. An index $v = 0$ indicates order releases. On the other hand, an index $j = 0$ indicates final products as given in the FGI inventory equation.

The amount of available work is used as the WIP measure. It represents the load of the workstation v in period t denoted by

$$\widehat{W}_{vt} = \widehat{W}_{v,t-1} + \widehat{X}_{vt}$$

where the load is defined as the sum of WIP at the beginning of period t and the planned input in period t.

The formulation of the clearing function is nearly the same as in the approach of Missbauer (1998) for an $M/G/1$ model in equilibrium and denotes as follows:

$$\widehat{X}_t = \frac{1}{2}(\beta + \alpha + \widehat{W}_t) - \sqrt{\beta^2 + 2\beta\alpha - \alpha^2 - 2\beta\widehat{W}_t + 2\alpha\widehat{W}_t + \widehat{W}_t^2}.$$

However, Missbauer (2006b) shows that the functional relationship between the expected load and the expected output of the system exists only if the system is in steady state. Changes in the relationship happen if the system is in a phase of transient state which gives rise to a further planning circularity: clearing functions are based on assumptions about the dynamic behavior of the system. Order releases are based on clearing functions, but directly determine system behavior. Thus they are based on assumptions which they directly influence. Furthermore, clearing function approaches assume that the shape of the clearing function does not change over time which need not be true in transient or stationary phases during the planning horizon. As stated by Missbauer (2006b), Lautenschläger (1999) and Karmarkar et al. (1993b) optimization models including clearing functions "get nervous" if steady state properties are assumed for short planning periods. This nervousness can be shown from optimization results for one workstation (see, e.g., Missbauer 2006b): demand varies between some values below the maximum capacity. The variations of planned output are lower than those of demand due to the nonlinear increase in WIP. Nevertheless, order releases exaggerate the demand variations. Consequently, clearing function models cannot be optimal at first sight, because they need to incorporate the characteristics of transient states in order to be optimal. Therefore, Missbauer (2006b) presents an extension of the clearing function employing a theory that describes the material flow through a workstation at an abstract level which is the theory of transient queuing networks. Thus he reformulates the clearing function as follows:

$$X_{ijt} = f_{ij}(\widehat{X}_{ij\tau}, X_{ij\tau}[\tau = 1, \ldots, t-1]; \widehat{X}_{ijt})$$

where $\widehat{X}_{ij\tau}$ is the estimated increase (or planned) work input of product group i at workstation j in period τ. As stated above, the expected output of product group i of workstation j in period t is a function of the (expected) inputs and (expected) outputs of the previous periods. As this function is difficult to estimate, Missbauer (2006b) searches for the factors which mainly influence the X_{ijt}-variables and, hence, should be included in the model by analyzing the transient behavior of single-stage systems.

He bases the extended clearing function model on two insights on transient behavior of single-stage queuing systems, namely, that overloaded systems can be approximated by fluid models and that the transition of workstations from one stationary state to another seems to follow an exponential function. The first insight suggests that stochastic effects become less important if the input per period is significantly higher than capacity. This is supported by Bertsimas and Sethuraman (2002) who state that the optimality loss for fluid relaxations decreases as the number of jobs in the system increases. This is also experienced by Stahlman and Cochran (1998) and Riaño et al. (2006) where the latter get slightly better approximations when the utilization values are high. Furthermore, their approximations perform less well for decreasing input functions when the system starts to empty due to the problem of transition stated by Missbauer (2006b). However, Missbauer (2006b) argues that the WIP level in the clearing function model is a decision variable that cannot be assumed to be generally high.

The second insight is supported by the analytical work of Odoni and Roth (1983). They reveal for a $M/M/1$ system that if a stationary state exists, requiring that the arrival rate is smaller than the departure rate, then the expected number of customers (or product groups) in the queue at time t, denoted by $\widehat{W}_{it}$, converges to the stationary value W_i. This process follows approximately an exponential function if t is sufficiently large:

$$\widehat{W}_t = \widehat{W}(1 - e^{-t/\beta}) \quad \forall t \geq 0$$

where $\widehat{W}$ and β can be determined by analytical methods or empirically. This is also experienced by Lin and Cochran (1990) and Stahlman and Cochran (1998). The latter describe the transition of a workstation from one state to another and back due to a provisional decrease in capacity. Following this, Missbauer (2006b) formulates the equation for the expected (or planned) input to workstation j assuming only one product group and a single workstation as follows:

$$\widehat{X}_{jt} = \sum_{j=1}^{N} \sum_{\tau=0}^{\infty} X_{v,t-\tau} \overline{X}_{vj} \xi_{vj\tau} + \sum_{\tau=0}^{\infty} R_{t-\tau} \overline{X}_{0j} \xi_{0j\tau}.$$

This equation denotes the output of the previous workstations and releases, thus the expected input for workstation j which are delayed by the non-bottleneck workstations. Assuming to know the functional relationship between average input/output and average WIP at workstation j for the stationary state (denoted by $\widehat{W}^{\widehat{X}_{jt}}$), the average convergence level of W_{jt} is also known if an input rate of $\widehat{X}_{jt}$ is maintained in the long run. Further assuming that the input rate $\widehat{X}_{jt}$ leads to an increase in the WIP level, Missbauer (2006b) derives the following equation for the WIP level of workstation j in time t:

$$W_{jt} = W_{j,t-1} + (\widehat{W}^{\widehat{X}_{jt}} - W_{j,t-1})(1 - e^{-t/\beta_d})$$

where β_d denotes the time constant for the transient behavior from the first state to the second state. Substituting this for W_{jt} in the balance equation and solving for X_{jt} yields:

$$X_{jt} = \widehat{X}_{jt} - (\widehat{W}^{\widehat{X}_{jt}} - W_{j,t-1})(1 - e^{-t/\beta_d}).$$

Numerical experiments show that for some values of β_d the transient behavior of $\widehat{X}_{jt}$ is not realistic immediately after the input change. Moreover, β_d depends on the workstation utilization and approaches infinity as the utilization reaches 100%. With the above equation the clearing function model becomes nonlinear.

This is the first attempt to account for the transient behavior of clearing function approaches. Nevertheless, this model is complex and solving it remains difficult. However, Missbauer (2006b) notes that the other clearing function models have to be regarded as approximations based on steady state assumptions.

3.7 The planning model with forward shifting preparation times by Lautenschläger

Lautenschläger (1999) integrates workload dependent forward shifting production times in a multi-level capacity oriented production planning model with and without lot sizing. In contrast to approaches in the literature which use constant minimum forward shifting production times he considers these times to be dependent on the utilization of resources (workstations or production units) in a capacity oriented production planning model. Therefore, he uses a functional relationship that specifies the portion of the production volume that is to be

shifted forward in time due to (and dependent on) resource utilization. Extensions of the basic model contain linear as well as nonlinear forward shifting functions, the integration of overtime or complete preliminary production and effects of shifting production before the beginning of the planning horizon or planning period. Finally, procedures for smoothing resource utilization are proposed that use ideas from digital control engineering and a filter accounting for utilization levels of resources.

Simulation experiments demonstrate that part of the production volume is shifted forward in time due to the utilization level assumed and varied in the planning model. He also shows that employing the same forward shifting production function and varying only the utilization levels of the resources leads to increases in parts of the production volume that is shifted forward. Moreover, integrating forward shifting times reduces the occurrence of backorders independent of production system design which is also due to smoothing utilization levels in the production planning process as well as to the coordination of production segments and finally leads to a reduction of lead times and inventory without an increase of backorders. The basic model is a *multi-item, multi-level capacitated lot sizing problem* (ML-CLSP) which includes detailed lot sizing and resource deployment planning assuming the data to be deterministic within a finite planning horizon which is divided into T sub-periods of equal length. Furthermore, demand may vary from period to period, but backorders are not allowed. Inventory costs are proportional to the inventory stock at the end of a period and, because production and material costs per unit are assumed to be constant, they are not relevant for the planning decision. There are K operations represented in the model with multi-level product structure where every operation k is assigned to a resource and is only carried out once in a period causing setup costs and times. There is no restriction concerning the number of operations executed in one period, but the production system consists of N different resources ($j = 1, \ldots, N$) subject to capacity restrictions which can be extended by overtime.

The objective is the minimization of setup, inventory and overtime costs with the complete model as follows:

$$\min \quad \sum_{k=1}^{K}\sum_{t=1}^{T}[\pi_k \cdot I_{kt} + \zeta_k \cdot S_{kt}] + \sum_{j=1}^{N}\sum_{t=1}^{T} o_j \cdot O_{jt}$$

subject to:

$$I_{k,t-1} + X_{k,t-l_k} = PD_{kt} + \sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t} + I_{kt} \quad \forall k, j, t,$$

$$\sum_{k=1}^{K}(\xi_{jk} \cdot X_{kt} + \zeta_{jk} \cdot S_{kt}) \le \kappa_{jt} + O_{jt} \quad \forall j, t,$$

$$X_{kt} \le TOP \cdot S_{kt} \quad \forall k, t,$$

$$I_{kt}, O_{jt}, X_{kt} \ge 0 \quad \forall k, j, t,$$

$$S_{kt} \in \{0, 1\} \quad \forall k, t$$

where the objective function has three terms representing the inventory costs, the setup costs and the overtime costs and the index k standing for one operation, j denoting a resource (or

workstation or production unit, respectively) and t representing the period. The first constraint is the inventory balance equation stating that the sum of inventory stock of the previous period ($I_{k,t-1}$) and the production volume in period t, reduced by the constant minimal preliminary shift of operation k ($X_{k,t-l_k}$) must equal the sum of primary part demand (PD_{kt}), the secondary part demand ($\sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t}$ with $d_{j\hat{k}}^{(k)}$ denoting the direct requirement coefficient from the preceding to the subsequent operation) and inventory holding I_{kt} at the end of period t. This means that the processing of operation k must be initialized not later than l_k and finished before use of resource j. The constant minimal preliminary shift of operation k contains times for transfer, drying, etc., but does not take into account waiting times in front of resources due to occupation by other operations. The second restriction represents the capacity constraint stating that the sum of resource utilization for production ($\sum_{k=1}^{K} \xi_{jk} \cdot X_{kt}$) and the resource utilization for setups ($\sum_{k=1}^{K} \zeta_{jk} \cdot S_{kt}$) must be smaller or equal to the sum of capacity and overtime of resource j in period t ($\kappa_{jt} + O_{jt}$). The third constraint denotes the linkage of the production volume to the setup variables in order to take account of the fact that production only takes place after a setup (with TOP denoting a big number). The remaining constraints are the usual non-negativity restrictions and the definition of S_{kt} as a Boolean variable where $S_{kt} = 1$ indicates a setup for operation k in period t.

In order to balance the preliminary shifts with the smallest possible lead times for tactical production planning it is advisable to determine preliminary shifts dynamically and dependent on the production program. As a result, production process dependent waiting times are included in preliminary shifting times. Moreover, they need not be set for production planning, but result from the plan, so that one main source of inaccuracy (planning circularity) of production planning, i.e., the use of predetermined and estimated planned lead times in mrp systems, is eliminated. Consequently, the model is modified accordingly by redefining the variable for the production volume: Two modes for the production volume are defined which distinguish between production and availability of the product. That is, mode 1 represents an operation k which is produced as well as available in period t and mode 2 denotes an operation k produced in period $t-1$ and available in period t. So the selection of the production mode determines whether there has to be one period in between two operations or not. The production volume produced in mode m is then denoted as X_{kt}^{m} and the total production volume is defined as

$$X_{kt}^{\text{total}} = X_{kt}^{1} + X_{k,t+1}^{2} \quad \forall k, t$$

which is the sum of the production volume produced in period t for period t (mode 1) and the production volume produced in period t for period $t+1$ (mode 2). The position of the forward shifting production function is defined by two parameters b_j and e_j which are related to the part of production which had to be shifted forward in the previous period. Here b_j states the utilization level from which shifting production forward becomes necessary and e_j denotes the part of production volume which needs to be shifted forward at a utilization level of 100%. The production volume that can be produced in mode 1 depends on the capacity not used for production, namely $\kappa_{jt} = \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}}$. This constraint is adapted to the forward shifting production function by means of the parameters α_j and β_j which depend on b_j and e_j and modeled for each single resource:

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{1} \leq \alpha_j \kappa_{jt} - \beta_j \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \quad \forall j, t.$$

The parameters α_j and β_j have to be determined in order to derive the forward shifting production function for each resource. They depend on the before mentioned parameters b_j and e_j as follows (for more details see Lautenschläger 1999): $\alpha_j = \frac{e_j b_j}{1-b_j}$ and $\beta_j = \frac{e_j+b_j-1}{1-b_j}$. Subsequently, these extensions are integrated in the MLCLSP by inserting the new function and by modifying the inventory balance equation:

$$\min \quad \sum_{k=1}^{K}\left[T\pi_k I_{k0} + \sum_{t=1}^{T}((T-t+1)\cdot(l_k X_{kt}^{\text{total}} - \pi_k PD_{kt}) + \zeta_k S_{kt})\right]$$

subject to:

$$X_{kt}^{\text{total}} = \sum_{m=1}^{\text{Mod}} X_{kt}^{m} \quad \forall k, t,$$

$$I_{k,t-1} + \sum_{m=1}^{\text{Mod}} X_{kt}^{m} = PD_{kt} + \sum_{\hat{k}=1}^{K} d_{j\hat{k}}^{(k)} \cdot X_{\hat{k}t}^{\text{total}} + I_{kt} \quad \forall k, t,$$

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \le \kappa_{jt} \quad \forall j, t,$$

$$\sum_{k=1}^{K} \xi_{jk} X_{kt}^{1} \le \alpha_j \kappa_{jt} - \beta_j \sum_{k=1}^{K} \xi_{jk} X_{kt}^{\text{total}} \quad \forall j, t,$$

$$X_{kt}^{\text{total}} \le TOP \cdot S_{kt} \quad \forall k, t,$$

$$I_{kt}, X_{kt}^{m}, X_{kt}^{\text{total}} \ge 0 \quad \forall k, t, m,$$

$$S_{kt} \in \{0, 1\} \quad \forall k, t.$$

The objective function now aims at minimizing the sum of inventory and setup costs including l_k which is the echelon inventory cost coefficient for the increase of inventory due to operation k. Overtime is not modeled here. Furthermore, in this model the forward shifting production function is supposed to be a linear constraint for each resource.

The nonlinear relationship between the forward shifting volume and capacity utilization can perhaps be modeled by a linear approximation similar to the method described by Asmundsson et al. (2002, 2006a) or else by the description of a nonlinear constraint. Finally, in order to smooth variation of utilization rates of resources to account for the relationship between utilization and lead times, a controller from digital engineering is combined with the forward shifting production function. For further reading see Lautenschläger and Stadtler (1998), Lautenschläger (1999).

3.8 The supply chain coordination-model by Caramanis and Anli

Caramanis and Anli (1999a) examine the potential value of integrating variable lead time relationship information and information sharing between plant level and cell decision nodes in the supply chain coordination process of a multiple part production facility which is expected to consist of decreased inventory and backlog costs. For that purpose they propose a master problem algorithm for vertical supply chain coordination of planning and operational decision making for manufacturing cells with concave as well as convex (nonlinear)

lead time relationships. Horizontal coordination constraints like safety stocks and KANBAN levels are not considered. Furthermore, they evaluate the problem of generating the required information through calculation of the sub-problems at each production cell modeled as small networks of queues which are communicated to a master problem solving coordinator and they examine the numerical problems associated with system complexity. The objective of Caramanis and Anli (1999a) is to determine optimal and consistent (achievable) production targets (plans) for each production cell during a time period aiming at reducing WIP, FGI and backlog costs across the supply chain for a given demand forecast and to compare the benefits of employing higher complexity issues to classical methods of supply chain coordination. Plan consistency refers to adherence to horizontal operational production constraints such as capacity, lead time and safety stocks.

A production planning model for an assembly production system is defined as consisting of three production cells with the objective to minimize the WIP and FGI (surplus or backlog) costs over the planning horizon. The relevant optimization problem takes the following form:

$$\min \quad \sum_{i=1}^{M}\sum_{c=1}^{C}\sum_{t=1}^{T}[\omega_{it}W_{ict}+\pi_{it}\hat{I}_{ict}+\vartheta_{it}B_{ict}]$$

where $\omega_{it}W_{ict}$ are the costs of WIP of part type i in cell c at the end of period t with ω_{it} representing the cost factor and with $W_{ict}=\sum_{j=1}^{N}q_{ijct}$, i.e., the WIP of part type i in cell c at the end of period t is equal to the sum of all buffers q_{ict} across all workstations $j=1,2,\ldots,N$. Based on the inventory positions I_{ict}, $\pi_{it}\hat{I}_{ict}$ denotes the costs of the positive FGI at the end of period t with $\hat{I}_{ict}=\max\{0,I_{ict}\}$ and $\vartheta_{it}B_{ict}$ represents the costs of the backlogged demand at the end of period t with $B_{ict}=\max\{0,-I_{ict}\}$. The objective function is restricted by the following constraints enforced for $t\in\{1,\ldots,T\}$:

$$\begin{aligned}
&W_{ict}=W_{ic,t-1}+R_{ict}-X_{ict} \quad \forall i,c,t,\\
&I_{ict}=I_{ic,t-1}+X_{ict}-R_{i3t} \quad \forall i,t,\ c=1,2,\\
&I_{ict}=I_{ic,t-1}+X_{ict}-D_{it} \quad \forall i,t,\ c=3,\\
&\sum_{i=1}^{M}\xi_{icjt}X_{ict}\le\rho_{cjt} \quad \forall c,j,t,\\
&\sum_{t=1}^{t^*}R_{ict}+\dot{W}_{ic,t+1}\ge\sum_{t=1}^{t^*}X_{ict}+\sum_{t=t^*+1}^{T}X_{ict}1\{t-L_{ict}\le t^*\} \quad \forall i,c,\\
&\overline{W}_{ict}=\alpha W_{ic,t-1}+(1-\alpha)W_{ict},\\
&W_{ict},X_{ict},\dot{W}_{ict},\overline{W}_{ict},R_{ict}\ge 0 \quad \forall i,c,t,\\
&I_{ict}\ge 0 \quad \forall i,t,\ c=1,2,\\
&I_{ict} \quad \text{unrestricted in sign } c=3
\end{aligned}$$

where the first three constraints denote the material flow relations with R_{ict} being the release rate of part type i during period t into cell c and X_{ict} stating the production target of part i by cell c during period t. The fourth restriction is the capacity constraint where the sum of the expected time needed by machine j at cell c to produce one part of type i during time period t multiplied by the production target of part i by cell c during period t

$(\sum_{i=1}^{M} \xi_{icjt} X_{ict})$ is forced to be smaller or equal to the maximum allowed utilization of machine j in cell c during time period t. The fifth restriction represents the lead time constraint where $\dot{W}_{ict}$ is the WIP of part type i in cell c at the end of period t_0, hence available at the beginning of period t_1. An indicator function is represented by $1\{\cdot\}$ which equals 1 if $\{\cdot\}$ is true and 0 otherwise. Furthermore, an equality constraint is given by the sixth restriction which denotes the weighted average WIP of part i in cell c during period t, which is given by $0 \leq \alpha \leq 1$. The last two constraints are the usual non-negativity restrictions. The system is modeled as a Markovian closed queuing network with exponential processing times, variable probabilistic routing, multiple part types and multiple machines (resources). Lead time estimates for each production cell are determined by *Mean Value Analysis* (MVA; see, e.g., Suri and Hildenbrant 1984) and, assuming that the production period is long enough to lead to a steady state, rewritten by employing Little's Law:

$$X_{ict} \leq \frac{\overline{W}_{ict}}{L_{ict}}.$$

Since L_{ict} is a function of all lead time requirements the constraint is rewritten as follows:

$$X_{ict} \leq g_{ic}(\overline{W}_{ict}, X_{ict}, X_{i^*ct}, \ \forall i^* \neq i)$$

where g_{ic} is a nonlinear function which represents the lead time in the production system for each production cell modeled as a function of the specific workload of a cell in a considered time period. In order to analyze algorithms which are able to coordinate effectively the supply chain under complicating factors, Caramanis and Anli (1999a) concentrate on concave as well as convex lead time constraints.

In the case of concave lead time restrictions linearization is achieved by decentralized calculations at appropriately selected points of the production system generated by the iterative generalized Benders' master problem algorithm which handles the information exchange transactions between the sub-problem decision making including information feedback determination and the master problem coordinator. In the case of convex lead time restrictions a modified generalized Benders' master algorithm is used. The iterative algorithm generates tentative sets of production targets for each production cell and then examines their feasibility assuming the feasibility of the targets of all other cells. Subsequently, the production targets are converted in the planning horizon to a point on the lead time constraint surface by adjusting W_{ict} or X_{ict} as implied by the master problem iteration and returning the linearized constraint with the following form:

$$X_{ict} \leq \widetilde{g} + \frac{\delta g}{\delta W_{ict}}(W_{ict} - \widetilde{W}^n_{ict}) + \sum_{i^* \neq i} \frac{\delta g}{\delta X_{i^*ct}}(X_{i^*ct} - \widetilde{X}^n_{i^*ct})$$

with $\widetilde{(\cdot)}$ indicating initialization values. A feasible starting point on the surface of g_{ict} is obtained by using the tentative solution of iteration n:

$$[\widetilde{W}^n_{ict}, \widetilde{X}^n_{ict}, \widetilde{X}^n_{i^*ct}, \forall i^* \neq i].$$

Since the cell sub-problems use stochastic dynamic models with finer time scales than the master problem, a sensitivity analysis is performed and time averaged in order to be consistent with the master problem. Learning effects about the nonlinear lead time constraints are obtained after each iteration possibly leading to an optimal and consistent production plan.

As a result the complex issue of nonlinear lead time constraints is solved using LP which is similar to the solution of Asmundsson et al. (2003).

Caramanis and Anli (1999a) analyze three different types of lead time constraints, namely an unambiguous concave g_{ict}-function at all cells c and for all part types i, a medium non-concave function, and a severely non-concave function due to FCFS queue policy and significant different part type processing times with the result that the implemented algorithm leads to an effective coordination even with the existence of complicating factors like severely non-concave lead time constraints. Furthermore, in order to deal with combinatorial problems arising with production system size and therefore complexity issues of sub-problems at the cell level, ongoing advances with controlled queuing networks in combination with concurrent simulation and neurodynamic programming can be further research directions.

3.9 The production planning model with transient Little's Law by Riaño

In his thesis Riaño (2002) develops a transient Little's Law which is the time-dependent analogue of Little's Law and denotes the expected queue length in a general input/output model in terms of *transient Palm probabilities* (see, e.g., Baccelli and Bremaud 1987, Riaño et al. 2003b) of the waiting time and of the expected value of the input process. He uses the transient Little's Law to create an *integro-differential equation* for the expected WIP and develops an approximation method to solve it numerically. An optimization based production planning model is presented which establishes the optimal release of the material into the system in order to closely match the defined requirements, e.g., capacity constraints for a finite planning horizon. In doing so he can approximate the transient behavior of general multi-class queuing networks such as supply chains including random lead times and time-dependent WIP. Randomness can be due to material entering the production system and coming out in several periods according to some weights $w_i(s,t)$ which indicate the proportion of input of product i during period s that leaves the production system by period t. These weights implicitly contain information about the lead time distributions of the production system, i.e., they are a function of the system load with two properties: First, they represent how the lead times depend on the load of the system and, second, how the lead times are affected by the randomness inherent in the production system. This is also done by Riaño et al. (2003a). The numerical method presented for transient analysis gives an approximation solution to the expected value of the WIP in the system as a function of time, calculating the weights for the lead time distribution. In case of known weights the nonlinear optimization problem is reduced to an LP solvable with standard LP methods. Here we describe the basic production planning model of Riaño (2002) in order to show how he integrates LDLT in production planning.

The model is defined as a production network for producing several products. The model has the objective of determining an optimal release point for raw material into the system in order to meet some predefined requirements which are assumed to be deterministic (this is relaxed later). The products move through the system according to specific routes which may also be random, and feedback (rerouting) is allowed. The input of the production system $\widehat{X}_{qt}$ (the cumulative expected quantity of items that enter the buffer q at time t ($\widehat{X} = \{\widehat{X}_{qt} : t > 0, q \in B\}$) is transformed into the output ($X = \{X_{qt} : t > 0, q \in B\}$) through a function f stated as follows:

$$X = f_{qt}(\widehat{X}, W)$$

with W denoting the load on all buffers in the system and f denoting a function describing how the items are delayed due to queuing and processing times. The complete production model is stated as follows:

$$\min \quad \sum_{t=1}^{T} \sum_{q \in O(q)} [\phi_{qt}(X_{qt} - R_{qt})^{+} + \vartheta_{qt}(X_{qt} - R_{qt})^{-}] + \sum_{t=1}^{T} \sum_{q \in B} \omega_{qt} W_{qt}$$

subject to:

$$\begin{aligned}
&X_{qt} = f_{qt}(\widehat{X}, W) \quad q \in B,\ \forall t,\\
&Z_{qut} = \alpha_{qu} \widehat{X}_{ut} \quad q \in \Im(u),\ u \in B \setminus \Im,\ \forall t,\\
&Z_{qut} = \beta_{qu} X_{qt} \quad u \in O(q),\ q \in B \setminus O,\ \forall t,\\
&\widehat{X}_{qt} = W_{qt} + X_{qt} \quad q \in B,\ \forall t,\\
&\widehat{X}, X, W, Z \geq 0
\end{aligned}$$

with ϕ_{qt} denoting the holding costs for final products in the output buffer $q \in O$ during period t, X_{qt} representing the cumulative output from buffer $q \in B$ (with $B(s)$ being buffers whose items are processed at station $s \in S$) being the cumulative requirements of final product in buffer $q \in O$ up to time t. The parameter R_{qt} denotes the cumulative requirements (or releases) of the final product $i \in O$ up to time t. Thus terms $(\cdot)^{+}$ and $(\cdot)^{-}$ in the objective function state that the objective function minimizes production costs ($\phi_{qt}(X_{qt} - R_{qt})^{+}$) if $(X_{qt} - R_{qt})^{+}$ is positive or backorder costs ($\vartheta_{qt}(X_{qt} - R_{qt})^{-}$) if $(X_{qt} - R_{qt})^{+}$ is negative. The variable Z_{qut} is the cumulative flow from buffer $q \in B \setminus O$ to buffer $u \in B \setminus \Im$. $\widehat{X}_{qt}$ represents the cumulative input to buffer $q \in B$ and, finally, α_{qu} and β_{qu} denote the quantity of flow Z_{qut} needed to make one unit of input to buffer q.

The objective function minimizes all relevant costs. The first constraint describes the load dependent transformation of input into output. The following two equations describe the transfer of material between the buffers, and the fourth equation gives the conservation of material flow. The load-dependent function $\widehat{X}_{qt} = f_{qt}(X, W)$ is nonlinear, includes the above mentioned weights which embody stochastic queuing and service and describes the expected cumulative output X_{qt} in terms of input $\widehat{X}$ as a function of the load in the system. After a transformation (see Riaño 2002 for more details) we obtain for the cumulative output from buffer $q \in B$:

$$X_{qt} = \sum_{s=1}^{t} x_{qs} w_q(s, t)$$

where $x_{qt} = \widehat{X}_{qt} - \widehat{X}_{q,t-1}$ is the marginal amount of input to buffer q during period t and $w_q(s, t)$ are the weights that tell what fraction of items that entered buffer q during period s (between $s - 1$ and s) has come out of the system by time t.

Riaño (2002) includes in his model further stochastic issues like random demand and random output and formulates an optimization algorithm in order to find a feasible solution based on the computed weights which are consistent with the load of the system which means that planned production schedules will also be feasible. So, the model is able to schedule later releases in certain situations, e.g., if requirements of the production changes during the production horizon, and consequently reducing FGI while still satisfying production requirements.

Algorithms for approximating time-varying performance parameters for $G_t/G/c$ queuing systems with transient Little Law and integrals of non-stationary Palm distributions of Sojourn times are developed by Riaño et al. (2006).

3.10 Supply chain operations planning of Selcuk, Fransoo and De Kok

Selcuk et al. (2005) represent a hierarchical framework for supply chain operations planning including decisions on capacity loading which are decomposed from order releases so that throughput during planned lead times can be determined in order to meet the planned delivery schedule. The throughput is defined as a random variable with its probability distribution being based on the available WIP during a time period and the short-term probabilistic behavior of the shop floor. In order to plan throughput a clearing function is used that anticipates the short term (operational) dynamic performance of the shop floor. It is a mathematical representation of the relation between WIP and throughput of the production process. This is different from other clearing function models that are based on long-term relationships and stationary conditions as, e.g., Karmarkar (1987) or Hopp and Spearman (2001). The clearing function of Selcuk et al. (2005) is based on short term operational system behavior, because they argue that such a formulation explicitly accounts for the interaction between timing and size of release decisions and operational execution. Consequently, it includes the consequences on system performance. The periodic performance is anticipated at the aggregate level of order release decisions in terms of capacity loading, throughput and flow times, thus through the clearing function.

They consider a single item multi-period order release and scheduling problem in a two-level decision hierarchy. The delivery schedule is derived using planned lead times. The system is periodically updated by information of the system state, available capacity and safety stocks. Backorders are allowed and the method of rolling schedules is employed. The supply chain structure contains a manufacturer and a warehouse. Nevertheless, the planning and control of material flows is executed in a centralized manner. Planned lead times are defined as an integer number of time periods after which a (at the start of a time period) released order is expected to arrive at the warehouse. Furthermore, it is assumed that the system does not run short of raw material and transportation times are negligible. Consequently, the aim of the model can be described by determining the throughput levels in order to satisfy planned order releases such that forecasted demand is met by holding the minimum amount of material at the manufacturer as well as at the warehouse.

The hierarchical planning model is derived as an LP formulation. The i-index indicating the part types is omitted for reason of clarity. All constants and variables are assumed to be non-negative.

$$\min \quad \sum_{s=1}^{T} \pi I_{t+s} + \sum_{s=1}^{T-1} (w W_{t+s} + w^f W_{t+s}^f) + \sum_{s=0}^{T-1} \vartheta I_{t+s}^{N-}$$

subject to:

$$\begin{aligned}
&I_{t+s+1} - B_{t+s+1} = I_{t+s} - B_{t+s} + Q_{t+s-L} + Q_{t+s}^{\text{open}} - D_{t+s}, \quad s = 0, 1, \ldots, T-1,\\
&W_{t+s+1} = W_{t+s} + R_{t+s} - X_{t+s}, \quad s = 0, 1, \ldots, T-2,\\
&X_{t+s} \le f^c(W_{t+s} + R_{t+s}), \quad s = 0, 1, \ldots, T-2,\ c = 1, \ldots, C,\\
&W_{t+s+1}^f = W_{t+s}^f + X_{t+s} - Q_{t+s+1-L} - Q_{t+s+1}^{\text{open}}, \quad s = 0, 1, \ldots, T-2,\\
&I_{t+s}^{N+} - I_{t+s}^{N-} = I_{t+s+1} - B_{t+s+1} - I^s, \quad s = 0, 1, \ldots, T-1.
\end{aligned}$$

The objective function minimizes the costs of holding inventory, WIP, and backorders where $w^f W^f_{t+s}$ denotes the costs of finished WIP at the manufacturer at the start of period $t+s$ just after the order scheduled for period $t+s$ is sent to the warehouse. ϑ is a cost factor for backorders, and π denotes the inventory cost factor.

The variable B denotes the backorders and I^{N+} or I^{N-} the net inventory level over and under the safety stock I^s at the warehouse at the end of period $t+s$, respectively.

The first constraint is the inventory balance at the warehouse between the consecutive planning periods using information on the actual schedule. The variable Q_{t+s-L} denotes the order size to be released at the start of period $t+s$ as decided at the start of period t with $s = 0, 1, \ldots, T-1$ where the L-index denotes the planned order time. Furthermore, Q^{open}_{t+s} is the total quantity among the actually open orders that are scheduled for receipt at the start of period $t+s$ as given at the start of period t with $s = 1, \ldots, T$.

The second constraint denotes the WIP balance where WIP increases with the new releases and decreases with throughput. The third constraint gives the relationship between throughput and capacity loading according to a piecewise-linear and concave clearing function where $c = 1, \ldots, C$ denote the individual lines of the linearization of the clearing function, thus referring to the c-th linear part. As the decision variable for raw materials R_{t+s} determines the range in which X_{t+s} can be anticipated, it can be noted that decisions on capacity loading are restricted by the desired level of throughput. The desired level of throughput is decided such that:

$$X_{t+s} = \min_{c=1,\ldots,C} (f^c(W_{t+s} + R_{t+s})), \quad t = 0, 1, \ldots, T-2.$$

Moreover, the fourth constraint implies that each order must be finished and delivered to the warehouse within the planned lead time and the fifth constraint determines the variables in relation to the net inventory and the safety stock I^s. Selcuk et al. (2005) study different types of clearing functions as shown in Fig. 1. Here we concentrate on the short term nonlinear clearing function.

In order to derive the probability that all items are cleared in a single time period, Selcuk et al. (2005) assume a given available WIP level W and exponential processing times. The probability is then formulated as follows:

$$Pr\{X^W = W\} = 1 - \sum_{k=0}^{W-1} e^{-p} \frac{p^k}{k!}$$

where X^W is the random variable for the throughput level in the regarded time period and the index k denotes the WIP items (with $k = 1, \ldots, W$). Consequently, the expected throughput in a time period with a WIP level W is

$$\widehat{X} = \sum_{k=1}^{W} k Pr\{X^W = k\}.$$

If the available WIP level approaches infinity, the throughput becomes a Poisson with a mean p, thus the short term nonlinear clearing function approaches the (nominal) production rate. After some reformulation the short term clearing function can be shown as follows:

$$f(W) = \sum_{k=1}^{W} Pr\{X^k = k\}.$$

Employing this clearing function and using simulation Selcuk et al. (2005) reveal that it performs best in terms of average periodic costs and robustness of the system. A coordinated material flow in the supply chain is realized for both manufacturer and warehouse.

4 Conclusions

We have seen that although considerations about load dependent lead times are rare in the literature to date, there are some very interesting aggregate planning models that take them into account. This is particularly noteworthy, because lead times are of critical importance with respect to the global competitiveness of firms. Furthermore, we have discussed the importance of accounting for the nonlinear relationship between lead times and workload of production systems as well as other factors such as, e.g., product mix, scheduling policies, batching or lot sizing, variable demand patterns, etc. Queuing models highlight the evidence that lead times increase nonlinearly long before resource utilization reaches 100% which is only included in few mathematical planning models. Additionally, there is a lack of models that analyze LDLT in the context of stochastic demand and other uncertainties evidently prevailing in practice.

The approach of modeling clearing functions in order to account for LDLT is considered very promising and should be implemented in a stochastic framework by using queuing models with the purpose to integrate the problem of variable demand patterns and to analyze the behavior of LDLT as proposed by Buzacott and Shantikumar (1993). This could be used as a starting point for additional modeling of production systems. It is evident that not all concepts and methods are well suited for each type of production system (job shop, flexible manufacturing system, etc.) with one or more production stages or products (one level–multi level production systems, one item–multi items or products), respectively. In case of LDLT and their integration into mathematical programming models for the aggregate production planning process it is wise to begin with the simplest production system and then proceed to more complex systems and situations in order to arrive at overall supply networks with the aim of finding general principles of modeling and accounting for LDLT.

For future research we also need some careful consideration of the interaction between planning and scheduling. For a better understanding of the factors that determine the performance of a supply chain or a supply network, it is necessary to create formal and structured approaches and then incorporate the resulting knowledge and findings in software programs which can help to support planning decisions about the production system and help to find optimal decisions with respect to the complexities, mutual dependencies inherent in the system. The increasing level of automation, especially at the operational level, not only aims at improving the physical flow of materials through the supply network, but also tries to combine the understanding of system behavior with information and computational support in order to explore different courses of action and find optimal production plans.

Acknowledgements The helpful comments of German Riano and Reha Uzsoy are greatly appreciated.

References

Agnew, C. (1976). Dynamic modeling and control of congestion-prone systems. *Operations Research*, *24*(3), 400–419.

Asmundsson, J., Rardin, R. L., & Uzsoy, R. (2002). *Tractable nonlinear capacity models for aggregate production planning*. Working paper, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., Rardin, R. L., & Uzsoy, R. (2003). *An experimental comparison of linear programming models for production planning utilizing fixed lead time and clearing functions*. Working paper, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., Rardin, R. L., Turkseven, C. H., & Uzsoy, R. (2006a). *Tractable nonlinear capacity models for production planning*. Working paper, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., Rardin, R. L., & Uzsoy, R. (2006b). Tractable nonlinear production planning models for semiconductor wafer fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing*, *19*(1), 95–111.

Baccelli, F., & Bremaud, P. (1987). *Palm probabilities and stationary queues*. Berlin: Springer.

Bechte, W. (1982). Controlling manufacturing lead time and work-in-process inventory by means of load-oriented order release. In *APICS, 25th annual international conference of production of the American production inventory control society* (pp. 67–72).

Bechte, W. (1988). Theory and practice of load-oriented manufacturing control. *International Journal of Production Research*, *26*(3), 375–395.

Ben-Daya, M., & Raouf, A. (1994). Inventory lead time as a decision variable. *Journal of the Operational Research Society*, *45*(5), 579–582.

Bertrand, J. W. M. (1983). The effect of workload dependent due dates on job shop performance. *Management Science*, *29*(7), 799–816.

Bertrand, J. W. M., & Wijngaard, J. (1986). The structuring of production control systems. *International Journal of Operations and Production Management*, *6*(2), 5–20.

Bertrand, J. W. M., & Wortmann, J. (1981). *Production control and information systems for component manufacturing*. Amsterdam: Elsevier.

Bertsimas, D., & Sethuraman, J. (2002). From fluid relaxations to practical algorithms for job shop scheduling: The makespan objective. *Mathematical Programming*, *92*(1), 61–102.

Billington, P. J., McClain, J. O., & Thomas, L. J. (1983). Mathematical programming approaches to capacity-constrained mrp systems: Review, formulation and problem reduction. *Management Science*, *29*(10), 1126–1137.

Bitran, G. R., Haas, E. A., & Hax, A. C. (1981). Hierarchical production planning: A single stage system. *Operations Research*, *29*, 717–743.

Bitran, G. R., Haas, E. A., & Hax, A. C. (1982). Hierarchical production planning: A two stage system. *Operations Research*, *30*, 232–251.

Buzacott, J. A., & Shantikumar, J. G. (1993). *Stochastic models of manufacturing systems*. New York: Englewood Cliffs.

Caramanis, M. C. (2000). *Multiple time scale decomposition in the production planning of stochastic manufacturing supply chain systems*. Internet source: http://www.bu.edu/pcms/docs/mcc/mtsdppsc.pdf, last call 29 Dec. 2006.

Caramanis, M. C., & Anli, O. M. (1999a). Modeling load-dependent lead time constraints for efficient supply chain coordination: Convexity issues. In *Proceedings of the 38th IEEE conference on decision and control* (Vol. 1, pp. 900–906).

Caramanis, M. C., & Anli, O. M. (1999b). Dynamic lead time modeling for JIT production planning. In *Proceedings of the IEEE international conference on robotics and automation* (Vol. 2, pp. 1450–1455).

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, *46*, 436–443.

Clark, A. R., & Armentano, V. A. (1995). A heuristic for a resource-capacitated multi-stage lot-sizing problem with lead times. *Journal of the Operational Research Society*, *46*, 1208–1222.

Dauzère-Pérès, S., & Lasserre, J.-B. (2002). On the importance of sequencing decisions in production planning and scheduling. *International Transactions in Operational Research*, *9*(6), 779–793.

de Kok, T. G., & Fransoo, J. C. (2003). Planning supply chain operations: Definition and comparison of planning concepts. In A. G. de Kok & S. C. Graves (Eds.), *Supply chain management: Design, coordination and operation*. *Handbooks in operations research and management science* (pp. 597–675). Amsterdam: North-Holland.

Dempster, M. A. H., Fisher, M. L., Jansen, L., Lageweg, B. J., Lenstra, J. K., & Rinnooy Kan, A. H. G. (1981). Analytical evaluation of hierarchical planning systems. *Operations Research*, *29*(4), 707–716.

Domschke, W., Scholl, A., & Voß, S. (1997). *Produktionsplanung: Ablauforganisatorische Aspekte* (2nd ed.). Berlin: Springer.

Enns, S. T. (2001). MRP performance effects due to lot size and planned lead time settings. *International Journal of Production Research*, *39*(1), 461–480.

Enns, S. T., & Choi, S. (2002). Use of GI/G/1 queuing approximations to set tactical parameters for the simulation of mrp systems. In *Proceedings of the 34th conference on winter simulation: Exploring new frontiers* (pp. 1123–1129).

Enns, S. T., & Suwanruji, P. (2004). Work load responsive adjustment of planned lead times. *Journal of Manufacturing Technology Management*, *15*(1), 90–100.

Ettl, M., Feigin, G. E., Lin, G. Y., & Yao, D. D. (2000). A supply network model with base-stock control and service requirements. *Operations Research*, *48*, 216–232.

Fleischmann, B., & Meyr, H. (2003). Planning hierarchy, modeling and advanced planning systems. In A. G. de Kok & S. C. Graves (Eds.), *Supply chain management: Design, coordination and operation. Handbooks in operations research and management science* (pp. 455–523). Amsterdam: North-Holland.

Glassey, C. R., & Resende, M. G. C. (1988). Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, *1*, 36–46.

Goldratt, E. M. (1988). Computerized shop floor scheduling. *International Journal of Production Research*, *26*(3), 443–455.

Goldratt, E. M., & Fox, R. E. (1986). *The race*. Croton-on-Hudson: North River.

Graves, S. C. (1986). A tactical planning model for job shops. *Operations Research*, *34*, 522–533.

Graves, S. C., & Willems, S. P. (2000). Optimizing strategic safety stock placement in supply chains. *Manufacturing & Service Operations Management*, *2*, 68–83.

Hackman, S., Rianõ, G., Serfozo, R., Ng, S. H., Lendermann, P., & Chan, L. P. (2002). *A stochastic production planning model*. Working paper, The Logistics Institute, Georgia Tech, and The Logistics Institute—Asia Pacific, National University of Singapore.

Hackman, S. T., & Leachman, R. C. (1989). A general framework for modeling production. *Management Science*, *35*, 478–495.

Hall, N. G., & Potts, C. N. (2003). Supply chain scheduling: Batching and delivery. *Operations Research*, *51*, 566–584.

Hendry, L. C., & Kingsman, B. C. (1991). Job release: Part of a hierarchical system to manage manufacturing lead times in make-to-order companies. *Journal of the Operational Research Society*, *42*, 871–883.

Homem-de-Mello, T., Shapiro, A., & Spearman, M. L. (1999). Finding optimal material release times using simulation-based optimization. *Management Science*, *45*, 86–102.

Hopp, W., & Spearman, M. (2001). *Factory physics-foundations of manufacturing management* (2nd ed.). Boston: McGraw-Hill.

Hopp, W. J., Spearman, M. L., & Woodruff, D. L. (1990). Practical strategies for lead time reduction. *Manufacturing Review*, *3*(2), 78–84.

Horiguchi, K., Raghavan, N., Uzsoy, R., & Venkateswaran, S. (2001). Finite capacity production planning algorithms for a wafer fabrication facility. *International Journal of Production Research*, *39*, 825–842.

Huang, G. Q., Lau, J. S. K., & Mak, K. L. (2003). The impacts of sharing production information on supply chain dynamics: A review of the literature. *International Journal of Production Research*, *41*, 1483–1517.

Hung, Y.-F., & Hou, M.-C. (2001). A production planning approach based on iterations of linear programming optimization and flow time prediction. *Journal of the Chinese Institute of Industrial Engineers*, *18*(3), 55–67.

Hwang, S., & Uzsoy, R. (2004). *A single-product dynamic lot sizing model with work in process and congestion*. Technical report, School of Industrial Engineering, Purdue University, West Lafayette.

Jodlbauer, H. (2004). Discussion of the standard processing time. *International Journal of Production Research*, *42*(7), 1471–1479.

Jodlbauer, H. (2005). Definition and properties of the input-weighted average lead-time. *European Journal of Operational Research*, *164*, 354–357.

Karmarkar, U. S. (1987). Lot sizes, lead times and in-process inventories. *Management Science*, *33*, 409–418.

Karmarkar, U. S. (1989a). Batching to minimize flow times on parallel heterogeneous machines. *Management Science*, *35*, 607–613.

Karmarkar, U. S. (1989b). Capacity loading and release planning with work-in-process (WIP) and lead times. *Journal of Manufacturing and Operations Management*, *2*, 105–123.

Karmarkar, U. S. (1993a). *Lot size, manufacturing lead times and utilization*. Working paper No. QM8321, Graduate School of Management, University of Rochester.

Karmarkar, U. S. (1993b). Manufacturing lead times, order release and capacity loading. In S. C. Graves, A. Rinnooy Kan & P. Zipkin (Eds.), *Logistics of production and inventory. Handbooks in operations research and management science* (Vol. *4*, pp. 287–329). Amsterdam: North-Holland.

Karmarkar, U. S., Kekre, S., & Kekre, S. (1985a). Lot sizing in multi-item multi-machine job shops. *IIE Transactions*, *17*, 290–298.

Karmarkar, U. S., Kekre, S., Kekre, S., & Freeman, S. (1985b). Lot-sizing and lead-time performance in a manufacturing cell. *Interfaces*, *15*(2), 1–9.

Karmarkar, U. S., Kekre, S., & Kekre, S. (1987). Capacity analysis of a manufacturing cell. *Journal of Manufacturing Systems*, *6*(3), 165–175.

Kekre, S., & Kekre, S. (1985). *Work-in-process considerations in job shop capacity planning*. Working paper, GSIA, Carnegie Mellon University.

Kim, J., Leachman, R. C., & Suh, B. (1996). Dynamic release control policy for the semiconductor wafer fabrication lines. *Journal of the Operational Research Society*, *47*, 1516–1525.

Kim, S., & Uzsoy, R. (2006a). A column generation heuristic for capacity expansion problems. *IIE Transactions*, to appear.

Kim, S., & Uzsoy, R. (2006b). *Exact and heuristic algorithms for capacity expansion problems*. Technical report, Laboratory for Extended Enterprises at Purdue, School of Industrial Engineering, Purdue University, West Lafayette.

Krieg, G. (2003). *KANBAN—controlled manufacturing systems*. PhD thesis, University of Ingolstadt, Ingolstadt.

Land, M. J., & Gaalman, G. J. C. (1996). Workload control concept in job shops: A critical assessment. *International Journal of Production Economics*, *46/47*(11), 535–548.

Lautenschläger, M. (1999). *Mittelfristige Produktionsprogrammplanung mit auslastungsabhängigen Vorlaufzeiten*. PhD thesis, University of Darmstadt, Darmstadt.

Lautenschläger, M., & Stadtler, H. (1998). *Modelling lead times depending on capacity utilization*. Technical report, Technische Hochschule Darmstadt, Darmstadt, Germany.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*, 546–558.

Lin, L., & Cochran, J. K. (1990). Metamodels of production line transient behaviour for sudden machine breakdowns. *International Journal of Production Research*, *28*(10), 1791–1806.

Lu, S. C. H., Ramaswamy, D., & Kumar, P. R. (1994). Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants. *IEEE Transactions on Semiconductor Manufacturing*, *7*, 374–388.

Missbauer, H. (1998). *Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen*. PhD thesis, University of Heidelberg, Heidelberg.

Missbauer, H. (2002). Lot sizing in workload control systems. *Production Planning & Control*, *13*(7), 649–664.

Missbauer, H. (2006a). Models of the transient behaviour of production units to optimize the aggregate material flow. Presented at 21st European conference on operational research, EURO XXI, Iceland, 2–5 July 2006.

Missbauer, H. (2006b). *Lead time management and order release in manufacturing planning and control systems*. Working Paper, University of Innsbruck.

Odoni, A. R., & Roth, E. (1983). An empirical investigation of the transient behavior of stationary queuing systems. *Operations Research*, *31*(3), 432–455.

Orcun, S., Asmundsson, J. M., Uzsoy, R., Clement, J. P., Pekny, J. F., & Rardin, R. L. (2006a, to appear). Supply chain optimization and protocol environment (SCOPE) for rapid prototyping and analysis of complex supply chains. *Production Planning & Control*.

Orcun, S., Uzsoy, R., & Kempf, K. (2006b). Using system dynamics simulations to compare capacity models for production planning. In *Proceedings of the 2006 winter simulation conference* (pp. 1–8).

Ouyang, L., Yeh, N., & Wu, K. (1997). Mixture inventory models with backorders and lost sales for variable lead times. *Journal of the Operational Research Society*, *47*, 829–832.

Öztürk, A., Kayaligil, S., & Özdemirel, N. E. (2006). Manufacturing lead time estimation using data mining. *European Journal of Operational Research*, *173*(2), 683–700.

Pahl, J., Voß, S., & Woodruff, D. L. (2005a). Load dependent lead times—from empirical evidence to mathematical modeling. In H. Kotzab, S. Seuring, M. Müller & G. Reiner (Eds.), *Research methodologies in supply chain management* (pp. 539–554). Heidelberg: Physica.

Pahl, J., Voß, S., & Woodruff, D. L. (2005b). Production planning with load dependent lead times. *4OR: A Quarterly Journal of Operations Research*, *3*(4), 257–302.

Philipoom, P. R., & Fry, T. D. (1992). Capacity-based order review/release strategies to improve manufacturing performance. *International Journal of Production Research*, *30*(11), 2559–2572.

Riaño, G. (2002). *Transient behavior of stochastic networks: Application to production planning with load-dependent lead times*. PhD thesis, Georgia Institute of Technology.

Riaño, G., Serfozo, R., & Hackman, S. (2003a). *Modeling random lead times in linear programs: A tutorial*. Technical report, COPA Centro de Optimización y Probabilidad Aplicada, Universidad de los Andes and Georgia Institute of Technology.

Riaño, G., Serfozo, R., & Hackman, S. (2003b). *A transient little law*. Technical report, COPA Centro de Optimización y Probabilidad Aplicada, Universidad de los Andes and Georgia Institute of Technology.

Riaño, G., Serfozo, R., & Hackman, S. (2006). *Transient behavior of queuing networks*. Technical report, COPA Centro de Optimización y Probabilidad Aplicada, Universidad de los Andes and Georgia Institute of Technology.

Selcuk, B., Fransoo, J. C., & De Kok, A. G. (2005). *Work-in-process clearing in supply chain operations planning*. Working paper, Department of Technology Management, Technische Universiteit Eindhoven, The Netherlands. Internet source: http://w3.tm.tue.nl/fileadmin/tm/Capaciteitsgroepen/OPAC/ClearingFunction.pdf, last call 25 December 2006.

Selcuk, B., Fransoo, J. C., & De Kok, A. G. (2006). The effect of updating lead times on the performance of hierarchical planning systems. *International Journal of Production Economics*, *104*(2), 427–440.

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2002). *Designing and managing the supply chain* (2nd ed.). Boston: McGraw-Hill.

Spearman, M. L., & Zhang, R. Q. (1999). Optimal lead time policies. *Management Science*, *45*(2), 290–295.

Spearman, M. L., Hopp, W. J., & Woodruff, D. L. (1989). A hierarchical control architecture for CONWIP production systems. *Journal of Manufacturing and Operations Management*, *2*, 147–171.

Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP: A pull alternative to Kanban. *International Journal of Production Research*, *28*, 879–894.

Srinivasan, A., Carey, M., & Morton, T. E. (1988). *Resource pricing and aggregate scheduling in manufacturing systems*. Working paper, GSIA.

Stadtler, H., & Kilger, C. (Eds.). (2005). *Supply chain management and advanced planning* (3rd ed.). Berlin: Springer.

Stahlman, E. J., & Cochran, J. K. (1998). Dynamic metamodelling in capacity planning. *International Journal of Production Research*, *36*(1), 197–209.

Suri, R., & Hildenbrant, R. R. (1984). Modeling flexible manufacturing systems using mean value analysis. *Journal of Manufacturing Systems*, *3*, 27–38.

Tatsiopoulos, I. P., & Kingsman, B. G. (1983). Lead time management. *European Journal of Operational Research*, *14*, 351–358.

Thonemann, U. W. (2002). Improving supply-chain performance by sharing advance demand information. *European Journal of Operational Research*, *142*, 81–107.

Upasani, A., & Uzsoy, R. (2005). *Integrated production planning and pricing decisions in congestion—prone capacitated production systems*. Research report, Laboratory for Extended Enterprises at Purdue, School of Industrial Engineering, Purdue University, West Lafayette.

Uzsoy, R., Lee, C. Y., & Martin-Vega, L. A. (1994). A review of production planning and scheduling models in the semiconductor industry, part II: Shop-floor control. *IIE Transactions*, *26*, 44–55.

Vepsalainen, A. P., & Morton, T. E. (1988). Improving local priority rules with global lead-time estimates: A simulation study. *Journal of Manufacturing and Operations Management*, *1*, 102–118.

Voß, S., & Woodruff, D. L. (2006). *Introduction to computational optimization models for production planning in a supply chain* (2nd ed.). Berlin: Springer.

Voß, S., & Woodruff, D. L. (2004). A model for multi-stage production planning with load dependent lead times. In R. H. Sprague (Ed.), *Proceedings of the 37th annual Hawaii international conference on system science* (pp. DTVE03 1–9). Piscataway: IEEE.

Wagner, H. M., & Whitin, T. M. (1958). Dynamic version of the economic lot size model. *Management Science*, *5*, 89–96.

Wiendahl, H., & Wedemeyer, H. (1990). Das Dilemma der Fertigungssteuerung. *Zeitschrift für Betriebswirtschaft*, *60*, 407–422.

Witt, A., & Voß, S. (2005, to appear). Simple heuristics for scheduling with limited intermediate storage. *Computers & Operations Research*.

Wu, K. (2001). A mixed inventory model with variable lead time and random supplier capacity. *Production Planning & Control*, *33*, 95–106.

Yano, C. A. (1987). Setting planned leadtimes in serial production systems with tardiness costs. *Management Science*, *33*, 95–106.

Yücesan, E., & de Groote, X. (2000). Lead times order release mechanisms, and customer service. *European Journal of Operational Research*, *120*, 118–130.

Zäpfel, G., & Missbauer, H. (1993a). Production planning and control (PPC) systems including load-oriented order release—problems and research perspectives. *International Journal of Production Economics*, *30*, 107–122.

Zäpfel, G., & Missbauer, H. (1993b). New concepts for production planning and control. *European Journal of Operational Research*, *67*, 297–320.

Zhao, X., Xie, J., & Leung, J. (2002). The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research*, *142*, 321–344.

Zijm, W. H. M., & Buitenhek, R. (1996). Capacity planning and lead time management. *International Journal of Production Economics*, *46/47*, 165–179.

SUPPLY CHAIN INTEGRATION: IMPROVEMENTS OF GLOBAL LEAD TIMES WITH SCEM

Julia Pahl [a], Stefan Voß [a], Andreas Mies [b]
[a] University of Hamburg, Hamburg, Germany
[b] Mummert Consulting AG, Hamburg, Germany

Supply Chain Integration is about integrating supply chain partners in a network and ensuring the uninterrupted flow of goods and information in order to increase global efficiency and responsiveness. As organizations move from creating plans for individual production lines to entire supply chains it is increasingly important to recognize that decisions concerning utilization of production resources impact the lead times that will be experienced (see, e.g., the bullwhip effect). Long lead times may impose high costs due to raised work in process (WIP) levels and larger safety stocks caused by increased uncertainty about production prerequisites and constraints. Without appropriate communication and planning models, decreased lead times at one link of the supply chain do not automatically lead to improvements in global supply chain lead times. Supply Chain Event Management (SCEM) addresses the problem of identifying deviations between an original plan and its execution across the whole network and generates corrective actions according to predefined rules and, thus, improves decision making. We examine the concept of SCEM and validate its effectiveness for production planning with load dependent lead times.

INTRODUCTION

The dynamics and the globalization of markets imply competition in time-based factors like flexibility or responsiveness to consumers, especially in case of variable demand patterns and various supply chain partners dispersed around the world (see (Huddleston, 2002), (Lee, 2000)). We define lead times as the time between the release of an order to the shop floor or to a supplier and the receipt of an item. Considerations about lead times are crucial to global competitiveness, because long lead times impose high costs due to rising WIP inventory levels and larger safety stocks attributable to increased uncertainty about production prerequisites and constraints (Karmarkar, 1989). Leadership of firms is achieved not only through a competitive price, but especially through factors like quality, direct guaranteed (reliable) deliverability, and short lead times (Andries & Gelders, 1995).

In recent years and with the literal explosion of new technologies considering computer power, speed and memory storage, a vast range of supply chain management (SCM) software systems and methods has flooded the market with the aim at overcoming, e.g., the time sensitive information provision problem offering tools and methods for better planning and control of the overall supply chain. Among these are, e.g., Advanced Planning Systems (APS), Enterprise Resource Planning (ERP), Quick Response Manufacturing (QRM), and SCEM (Supply Chain Event Management) just to mention a few. Problems arise when the management concepts behind this software are not well developed or these tools are not applicable or modifiable in order to reflect reality of shop floors. Furthermore, present tools and methods do not provide support to "intelligent" planning and decision making which is partly because most software systems have primarily evolved from advanced bookkeeping systems. This is, e.g., the case of Material Requirements Planning (mrp) and Manufacturing Resource Planning (MRP II) systems with its inherent problems discussed at a great length in the literature. For further reading see, e.g., (Zijm, 2000). A severe problem is the omission of capacity restrictions and nonlinear dependencies as, e.g., the relationship between lead times and

WIP (see (Voß & Woodruff, 2003) and (Voß & Woodruff, 2004). Further assuming "worst case" lead times for production planning in order to have enough buffer time leads to increased WIP and finished goods inventory (FGI) in the production system and, consequently, to waiting times which is addressed as the *lead time syndrome* in the literature (see (Zijm, 2000)). Intensifications arise when uncertainties and unexpected events like machine breakdowns and rush orders occur. In order to account for load dependent lead times it is mandatory to consider both events at the tactical and operational stage, because these two levels are linked very tightly and influence each other significantly (see (Dauzère-Pérès & Lasserre, 2002)).

The concept of SCEM can help to make the production process more transparent ("visible") for the decision maker within the supply chain helping to identify unexpected events as soon as they happen. This is especially important in times of outsourcing rendering supply chain processes more and more collaborative and complex through additional transportation modes, different handling methods etc. and further increasing time sensitivity with the application of Just In Time (JIT). This results in an increased vulnerability to irregularities and disruptions in supply processes (Zimmermann & Paschke, 2003). SCEM detects events defined as deviations between plan and its execution across the multitude of processes and actors in the supply chain in time, quantity or quality (Mies & Voß, 2004) and triggers the necessary, time sensitive information in order to launch corrective actions according to predefined rules (Otto, 2003). The main focus of SCEM is proactive behavior, viz. to act before a problem occurs (Zimmermann et al., 2006). Building on the concept of visibility in the supply chain considering internal as well as external resources of a company it allows systems, such as APS, to improve the quality of production plans and, therefore, to avoid unexpected events like capacity bottlenecks, inventory gaps or production changes. Companies' experience, e.g., 20% of inventory cost reductions and 30% reduction of supply chain cycle times (Vernon, 2004) through SCEM applications. It is a recent concept which has attracted only little attention in the literature so far. In this paper we examine the concept of SCEM and validate its effectiveness for production planning with load dependent lead times. We show that the improvements of transparency in the production planning can lead to significant lead time reductions of supply chains.

This paper is organized as follows. The following Section 2 highlights the theoretical concept of SCEM and the usage in practice with a critical evaluation of current approaches and software applications. Section 3 examines load dependent lead times giving a short overview on the problematic nature. Section 4 introduces formal modeling approaches of SCEM. Section 5 presents a modeling approach including SCEM as well as load dependent lead times in a supply chain planning model. The paper concludes with a critical evaluation and future research directions.

SUPPLY CHAIN EVENT MANAGEMENT

As plans and forecasts tend to deteriorate in accuracy due to continuously changing circumstances, SCEM provides an alerting mechanism and enables companies to respond quickly – or automatically – to unplanned events so that deviations from the plan can be considered at an early stage providing information based action opportunities (see (Mies & Voß, 2004), (Vernon, 2004)).

Theoretical Concept of SCEM

SCEM has evolved as an extension of process control (see (Marabotti, 2005)). Since SCEM applications are conceived as *add on's*, they need to be integrated into the overall SCM systems world (APS, ERP, Data Warehousing solutions, etc.) and linked with real-time applications like Electronic Data Interchange (EDI), Radio Frequency

Identification Device (RFID), mobile devices etc. (see (Zimmermann & Paschke, 2003)) which speeds up the supply chain by delivering status notes of parts, products and deliveries.

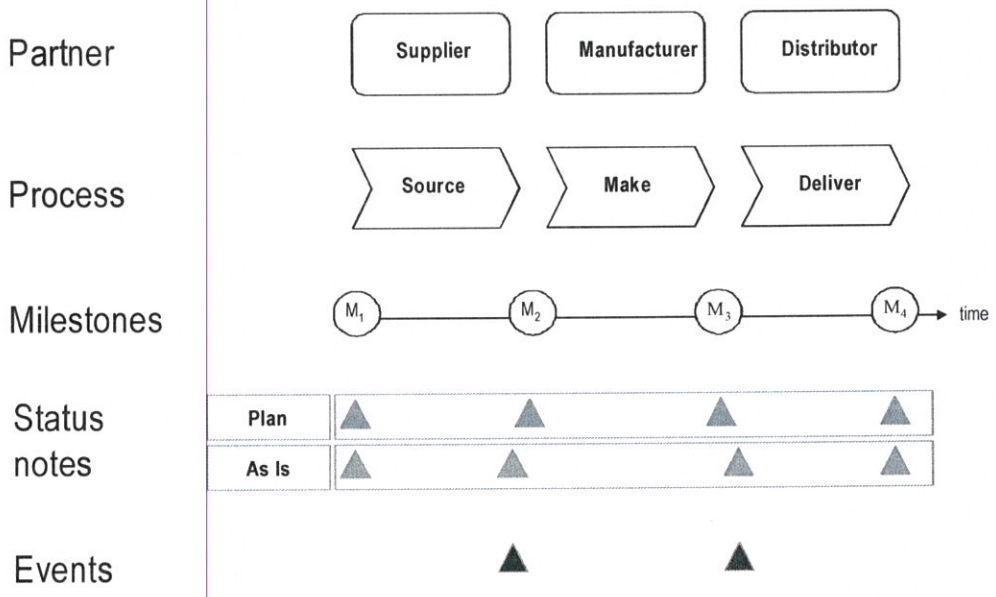

FIGURE 1: SCEM - The process oriented view

SCEM is an *event oriented* control concept for logistic processes where a process is defined as a continuous sequence of *milestones* with milestones being sensors deliberately positioned in the supply chain that report, processes, forwards and compares state notices to the plan, e.g., notices generated through tracking and tracing (T&T) systems. Thus, SCEM acts like a filter so as to avoid the time consuming search of deviations for the manager which can concentrate on value added activities, e.g., project control and decision making (Kilger et al., 2005). Only specifically classified notices (events) give impulses for action; see FIGURE 1 (Bretzke, 2003). Event information – categorized in high-, medium- or low-urgency alerts – are communicated to the responsible manager providing decision alternatives generated by norms and rules (Lane, 2002). By simulating the effects of occurred events, other events can be released in order to anticipate problems in the supply chain. Accordingly, SCEM acts like a mediator between supply chain planning (SCP) and -execution (SCE) by *monitoring, notifying, simulating, controlling* and *measuring* activities in the supply chain (see (Mies & Voß, 2004), (Otto, 2003)).

The success of every supply chain depends on the ability to prevent or resolve events. SCEM can use the following four modes to manage an event: *repair*, *re-schedule*, *re-plan* and *learn* (Otto, 2003). The *repair* mode describes an immediate reaction to an occurred deviation. The *re-schedule* mode re-calculates the production plan of the remaining processes and communicates the updated milestones to the relevant partners whereas the *re-plan* mode includes the re-planning of the complete process in case of incisive events. The *learn* mode includes the modification of planning policies derived from continuous analysis of occurred events.

Analysis of the Status Quo of SCEM

Regarding software, SCEM is a feature-functionality subset of SCM suites with its underlying problems and weaknesses influencing also the SCEM tool (see (Huang, 2002)). A study which analyses SCEM gives reference for the need of SCEM function-

ality in practice, but companies see more imperative call for action regarding the underlying systems themselves, e.g., T&T and SCM, emphasizing that the required environment for SCEM is not sufficiently matured so far. A general problem of SCEM applications lies in the definition of *deviations* (predefined thresholds) in order to avoid a flood of alerts when slight deviations occur. Since SCEM aggregates and processes data and information from other systems it heavily depends on the accuracy of this underlying data. If the data is outdated, because state notices and changes are registered with delay (or not at all), the system will send alerts and change proposal messages based on incorrect assumptions about reality (see (Kilger et al., 2005)). This is also subject to a study of (Kauremaa et al., 2004) that examines the general use of information technology (IT) in logistics and SCM which shows that, until now, only few supply chain partners are integrated with their SCM systems. Consequently, flexible supply networks using and sharing contemporary IT solutions are still a vision, but, nevertheless, on the way. For more detail see (Kauremaa et al., 2004). Besides, the decision support function of SCEM systems is still in its infancy according to (Lane, 2002). Nevertheless, the transparency within the supply chain provided by the SCEM concept and quick response generating preventive action obviously results into reduced lead times (see (Songini, 2001)). The higher the information visibility, the better the coordination of ordering, production and distribution with reduced costs and lead times avoiding effects highlighted with the Bullwhip effect.

LOAD DEPENDENT LEAD TIMES

Lead times are amongst the most important properties of items in the production planning and SCM, because short lead times are the key competition factor influencing flexibility and responsiveness to consumer, and service level, especially in case of variable demand patterns. They depend significantly on decisions about the production mix, scheduling policies, and system workload presented by WIP, FGI and capacity utilization (see (Asmundsson et al., 2002), (Asmundsson et al., 2003), (Caramanis & Anli, 1999)). This is ignored by traditional production planning models such as, e.g., mrp or MRP II, ERP or APS. In fact, many problems are inherent in these systems (see (Pahl et al. 2004), (Pahl et al., 2005)). One underlying assumption is a fixed or "*worst case*" lead time so that lead times are regarded as an attribute of an item (see (Voß & Woodruff, 2004)) although it is rather a consequence of system dynamics and dependencies. The *worst case* assumption serves to build buffer time rashly releasing orders in the production system, thus increasing system workload and, consequently, lead times which was attempted to be avoided. Therefore, lead times should not be fixed input parameters to production planning, but rather a result of the planning process. Another problem is that the afore mentioned systems lack of capacity constraints or model capacity as an upper bound with costs imposed only when the capacity constraint is violated, i.e. it tightens only in case of 100% capacity utilization. Nevertheless, empirical evidence demonstrates that lead times increase nonlinearly long before full resource utilization is reached. Thus, production plans tend to be inaccurate, because planned and realized lead times differ. Information inclusion about workload and unforeseen events which alter production plans could avoid or decrease load dependent lead times.

There exist several ways in the literature to account for load dependent lead times which range from indirect modelling approaches and queuing analysis to direct modelling approaches. The indirect modelling approaches try to influence parameters that have an effect on lead times such as decisions on release policies, influences on the demand side or shifts away from bottlenecks in order to increase capacity, lot sizing and production system characteristics (see e.g. (Missbauer, 1998)). Queuing analysis and

network models highlight the relationship between the capacity, loading and production mix as well as the resulting WIP levels and effects on lead times providing important information on the causes of congestion phenomena (e.g. (Chen et al., 1988)). Finally, direct approaches of modelling load dependent lead times integrate the nonlinear dependency between lead times and workload using clearing functions which model capacity as a function of system workload with a clearing factor that specifies the fraction of actual WIP "cleared" by a resource in a given time period. Work in this direction is provided by (Asmundsson et al., 2002), (Asmundsson et al., 2003), (Graves, 1986), (Karmarkar, 1989), (Missbauer, 1998). However, there does not exist work relating SCEM to load dependent lead times. In order to analyse how SCEM can be integrated in production planning systems (models) accounting for load dependent lead times we investigate the state of the art modelling approaches in the next section.

FORMAL MODELING APPROACHES

Approaches to integrate the SCEM concept into supply chains include Petri nets (Liu et al., 2004) or agent based technology (see (Zimmermann & Paschke, 2003), (Zimmermann et al., 2006). (Liu et al., 2004) present a Petri net approach which formulates supply chain event rules for analysing the cause-effect relationships between events. They categorize possible events in *task state related events*, *events produced by a task* and *external events* and present seven basic event patterns in order to capture cause-effect relationships which can be combined as building blocks so as to create more complex Petri-nets reflecting supply chains and their interdependencies or interactions. Agent-based technology is primarily used in supply chains to optimise schedules through decentralized coordination mechanisms or to support planning and execution of processes (see (Fox et al., 2000)). Nevertheless, an agent-based SCEM concept for order management is presented by (Zimmermann & Paschke, 2003). They argue that most SCEM systems whose capabilities and functionalities are added to ERP systems do not allow for adaptive behaviour regarding environmental changes resulting in extensive and time-consuming customisation of such systems and increased related costs which can be avoided using an agent-based approach. The authors provide SCEM functionalities concentrating on information gathering from different data sources across the supply chain with diverse types of agents (see Figure 2) analysing and comparing the data and sending reports and alerts (exception messages) if critical values are reached in order to monitor orders and suborders.

The SCEM algorithm is mapped on the agent architecture composed of three layers of agents: discourse agents, coordination agents, surveillance agents, and wrapper agents where the *discourse agents* on the first layer communicate the monitored information across the supply chain interacting with the other supply chain partner's agent systems. The tracking function is located at the second and third layer where the *coordination agent* executes the order profiling function triggering *surveillance agents* for every single monitored order that collects status information from, e.g., ERP systems, and communicates it to the discourse agents. *Wrapper agents* are located at the various information systems allowing for query functionalities. The monitored event data and state notes are stored in a data warehouse for analysis purposes through, e.g., data mining methods.

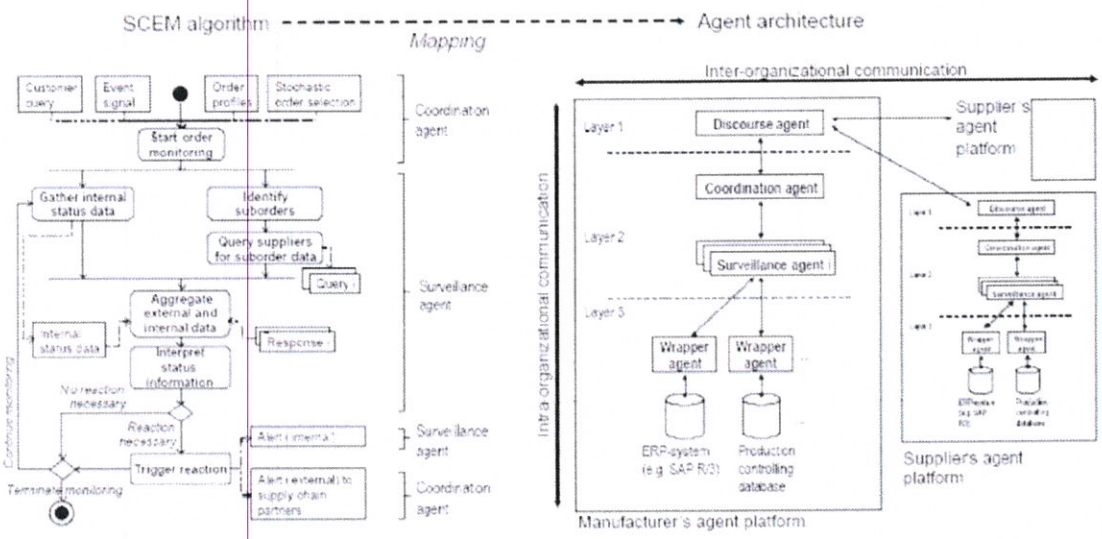

FIGURE 2: Agent-based software architecture (Zimmermann & Paschke, 2003)

In order to demonstrate the influence of SCEM on lead times, we provide a conceptual approach using key performance indicators (KPIs). We focus on the SCEM learn-function and on how it can be effectively integrated in mathematical programming models considering load dependent lead times for better production plans.

MODELING APPROACH OF SCEM INCLUDING LOAD DEPENDENT LEAD TIMES

Taking advantage of the functionality of the SCEM concept for production planning we focus on the learning and knowledge management component of SCEM. Therefore, we emphasize on *measures* derived from data analysis obtained by, e.g., surveillance agents as exposed above, e.g., protocols of delayed shipments, machine breakdowns leading to further delays or the need for re-routing etc., or by data mining methods in combination with data warehouse applications. These measures are then used to update (enhance) production plans in the short/ midterm run (e.g., "if machine X breaks down twice in a time period, then use machine Y") and to detect and repair inherent process problems in the long run (e.g. "buy a new machine of type X") in order to built a reliable supply chain with on-time arrivals of orders or manufacturing parts, thus low WIP inventories, short lead times and increased customer service. For a visualization of the concept see FIGURE 3. The Supply Chain Process Management (SCPM) comprises the SCEM functionality of data collection and event evaluation and the Supply Chain Performance Management (SCpM) with the functionality of documentation analysis, learning and derivation of KPIs through aggregation and evaluation of events.

With the help of the well known Supply Chain Operations Reference – model (SCOR) which allows for representing, analyzing and configuring complex supply chains and provides measures and indicators from highly aggregation to very detailed operational level, we can derive KPIs relevant for lead times. Some selected KPIs for evaluating supply chains are *Delivery Performance* (a), *Order Fulfilment Performance* (b), *Supply Chain Responsiveness* (c), *Production Flexibility* (d), *Total Logistics Management Cost* (e), *Value-Added Employees Productivity* (f), *Warranty Cost* (g), *Cycle Time* (h), *Inventory Days of Supply* (i) and *Asset Turns* (j) (Sürie & Wagner, 2002).

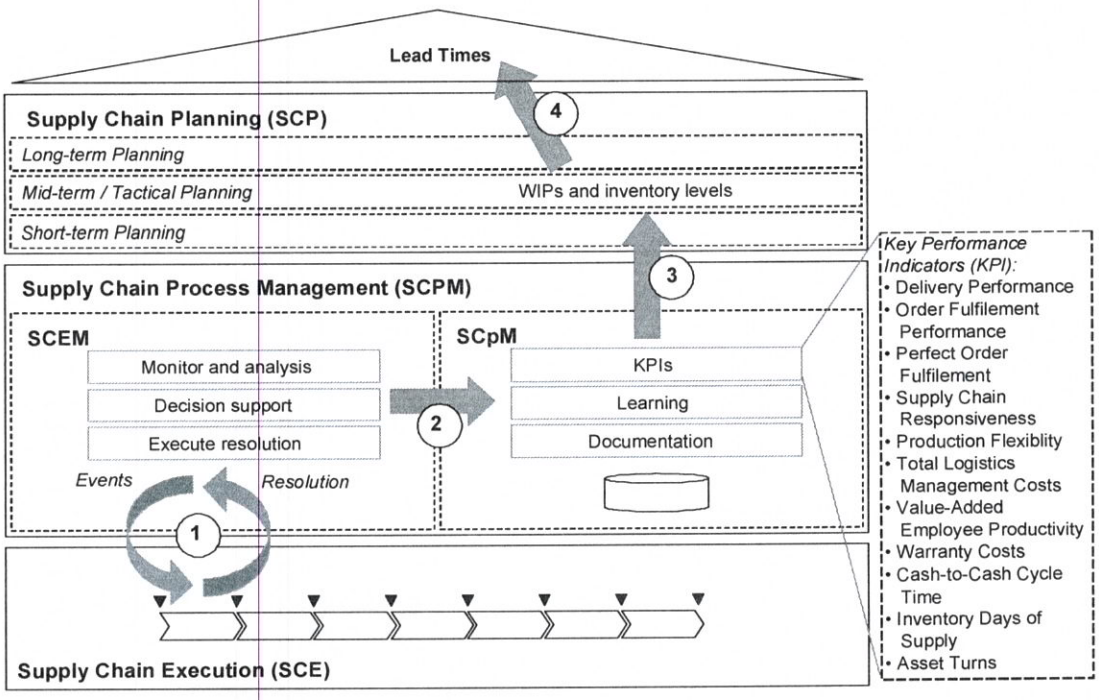

FIGURE 3: Lead time planning cycle

A further measure influencing lead times is the reliability of a workstation (Mirabedini, et al., 2001) or a supply chain partner which should be considered in production planning regarding lead time planning and integrated in mathematical programming models. Such a reliability measure highlights problems of a workstation, e.g., machine breakdowns, maintenance work, problems with setups etc., and accounts for the selection of a workstation for production or whether to prefer a more reliable one. Consequently, it contributes to more consistent and robust production plans at a tactical level, especially in regard of load dependent lead times and its effects also considered in selecting alternative production facilities (workstations or supply chain partners).

As outlined above, the dependency of lead times and workload can be captured by clearing functions whose functional form is reliant on diverse parameters. The basic idea of a clearing function is the expression of resource production in a time period depending on a function of diverse measures of workload as, e.g., WIP and FGI levels and order releases. They can be derived analytically using queuing theory or by fitting them into empirical data which is suggested by (Karmarkar, 1989) or (Asmundsson et al., 2003) who propose the following functional form:

$$(1) \qquad f(W) = \frac{K_1 W}{K_2 + W} \quad \text{or } f(W) = K_1\left(1 - e^{-K_2 W}\right)$$

Here, K_1 and K_2 are parameters that have to be determined by fitting the functions to data where K_1 represents the maximum capacity of a workstation and K_2 the slope of the clearing function.

To demonstrate the functionality of clearing functions in mathematical programming for production planning, we use a basic structure of a production model. The system is modelled as a network, where the set of nodes $N=\{1, \ldots, j, \ldots\}$ represents different workstations or supply chain partners that are connected by a set of directed transportation arcs $A_t = \{1, \ldots, k, \ldots\}$ arcs. The final workstation is denoted as $d \in N$. Let $N' = N - d$. The transportation arcs represent the flow of material between the different partners within the supply chain.

The model is stated as follows:

$$(2)\quad \min z = \sum_{t=1}^{T}\left[\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\omega_{jt}W_{ijt} + \pi_{jt}X_{ijt} + \phi_{jt}I_{ijt} + r_{jt}R_{ijt} + \sigma_{jt}S_{ijt} + \delta_{ijt}\gamma_{ijt}\Gamma_{ijt}\right) + \sum_{a=1}^{N'}\theta_{at}Y_{at}\right]$$

Conservation equations at each workstation:

$$(3)\quad W_{ijt} = W_{ij,t-1} - \frac{1}{2}\left(X_{ijt} - X_{ij,t-1}\right) + R_{ijt} + \sum_{\forall a \in A(j,i)} Y_{at} \qquad \forall i,j,t$$

$$(4)\quad I_{ijt} = I_{ij,t-1} - \frac{1}{2}\left(X_{ijt} - X_{ij,t-1}\right) - D_{ijt} - S_{ijt} - \sum_{\forall a \in B(j,i)} Y_{at} \qquad \forall i,j,t$$

State dependent capacity constraint (Clearing Function):

$$(5)\quad X_{jt} \le f_{jt}\left(W_{ijt}, W_{ij,t-1}, R_{ijt}\right) \qquad \forall j,t$$

Non-zero initial loading:

$$(6)\quad W_{ijt} = \overline{W}_{ij1} \text{ with } \overline{W}_{ij1} \ge 0$$

Reliability constraint:

$$(7)\quad \sum_{j=1}^{N'} \delta_{ijt} \log(\Gamma_{ijt}) \ge \log(\alpha_{ij}) \qquad \text{with } 0 \le \alpha \le 1, \forall i,t$$

Non-negativity constraint:

$$(8)\quad X_{ijt}, W_{ijt}, Y_{ijt}, I_{ijt}, R_{ijt}, S_{ijt}, \Gamma_{ijt} \ge 0$$

The objective function minimizes the costs of WIP ($\varpi_{ij}W_{ijt}$), production ($\pi_{jt}X_{ijt}$), FGI inventory ($\phi_{jt}I_{ijt}$), releases ($r_{jt}R_{ijt}$), shortages ($\sigma_{jt}S_{ijt}$) and transfer to other workstations a ($\theta_{jt}Y_{ijt}$). The conservation equations denote that the WIP of item i in workstation j in period t must equal the WIP from the preceding period minus the average production with X_{ijt} denoting production over the latter half of period t and the first half of period $t+1$, deduced the raw material releases and the transfer to subsequent workstations. Furthermore, FGI inventory of period t must equal the FGI inventory of the previous period $t-1$ minus average production, demand, shortages and transportation to subsequent workstations. The clearing function is given by (5). Equation (6) gives non-zero initial loading; the reliability constraint is presented by equation (7) and (8) gives the usual non-negativity constraint.

The reliability constraint deserves further consideration: the reliability measure Γ_{ijt}, derived by data mining methods from SCEM data based on KPIs, e.g., downtime, setup time (which could vary for specific products), throughout rate etc., must equal or be greater than a predefined threshold α_{ij} with $0 \le \alpha_{ij} \le 1$ which is a global reliability objective set as the product of all workstation reliabilities: $\prod_{a \in A(i,j)} \Gamma_{ijt} \ge \alpha_i$.

Information about the reliability measure can be derived by data-mining methods searching in data warehouse applications of SCEM systems. Furthermore, agent based approaches of SCEM can pass trend information about WIP and FGI inventory levels and, consequently, provide a better data basis for the derivation of clearing functions used in mathematical models.

CONCLUSIONS

In order to increase global efficiency and responsiveness within supply chains it is increasingly important to ensure the uninterruptible flow of goods and information. A critical success factor is the creation of reliable plans in tactical production planning.

Especially lead times are crucial for global competitiveness, because they impose high costs due to rising WIP inventory levels and larger safety stocks caused by increased uncertainty about production prerequisites and constraints. The concept of SCEM can help to generate the necessary visibility within the supply chain on an operational level. Information derived from SCEM can be used in the short-term to avoid deviations between the planning and the execution. In a long-term view we can collect and analyze this data in a data warehouse for performance improvement of supply chain processes using KPIs. The knowledge about these performance indicators can help us to continuously adjust the clearing functions in mathematical programs for tactical planning. The new type of clearing function can help to improve the reliability of production plans. Our thoughts are described in a mathematical model with an evolved type of clearing function. For the future, the research direction is to analyse the influence of different key performance indicators on clearing functions and the consequences for load dependent lead times.

REFERENCES

Andries, B. and L. Gelders (1995). Time-based manufacturing logistics. Logistics Information Management 8 (3), 30–36.

Asmundsson, J., R. Rardin, and R. Uzsoy (2002). Tractable nonlinear capacity models for aggregate production planning. *Working paper*, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., R. Rardin, and R. Uzsoy (2003). An experimental comparison of linear programming models for production planning utilizing fixed lead time and clearing functions. *Working paper*, School of Industrial Engineering, Purdue University, West Lafayette.

Bretzke, W.-R. (2003). Supply Chain Event Management – Bridging the Gap Between Planning and Realization. Internet Source: http://www.bbriefings.com/pdf/24/ifpm031_t_BEAR2.PDF.

Caramanis, M. and O. Anli (1999). Modeling load-dependent lead time constraints for efficient supply chain coordination: Convexity issues. *Technical report*, Proceedings of the IEEE Conference on Decision and Control, Phoenix, AZ.

Chen, H., M. Harrison, A. Mandelbaum, A. van Ackere, and L. Wein (1988). Empirical evaluation of a queuing network model for semiconductor wafer fabrication. Operations Research 36, 202–215.

Dauzère-Pérès, S., Lasserre, J.-B. (2002) On the Importance of Sequencing Decisions in Production Planning and Scheduling, International Transactions in Operational Research, 9 (6): 779 – 793

Fox, M., M. Barbuceanu, and R. Teigen (2000). Agent-oriented supply chain management. International Journal of Flexible Manufacturing Systems 12, 165–188.

Graves, S.C. (1986). A tactical planning model for job shops. Operations Reserach 34, 522–533.

Gustafsson, J. and A. Norrman (2001). Network managed supply – execution of real time replenishment in supply networks. Technical report, The 6th International Symposium on Logistics.

Huang, K. (2002). Supply chain visibility and event management takes on logistics. Internet Source: http://transentric.com/article/articlerelease.asp?intArticleID=29.

Huddleston, J. (2002) Supply Chain Visibility and Valuation, Internet: *http://www.21stGlobal.net*, Query: 05.04.2005 at 16:12.

Karmarkar, U.S. (1989) Capacity loading and release planning with work-in-process (WIP) and lead times, *Journal of Manufacturing and Operations Management*, 2, 105-123

Kauremaa, J., J. Auramo, K. Tanskanen, and M. K¨arkk¨a¨ainen (2004). The use of information technology in supply chains: Transactions and information sharing perspective. Technical report, LRN Conference.

Kilger, C., K. Ötschmann, and A. Schulz (2005). Anwendungen von Methoden des Supply Chain Managements auf das industrielle Projektgeschäft. Technical report, j&m Management Consulting AG. Internet Source: http://www.jnm.de/fileadmin/user upload/Dokumente/Artikel/jnm 03 Paderborner Tagung Methoden des SCM.pdf- 3. Sept. 2005.

Lane, R. (2002). From planning to control: Improving the high-tech supply chain. Technical report, Kleiner Perkins Caufield & Byers. Internet Source: http://whitepapers.zdnet.co.uk/0,39025945,60028796p-39000532q,00.htm.

Lee, H. (2000) Creating Value through Supply Chain Integration, *Supply Chain Management Review*, 30 – 36.

Liu, E., A. Kumar, and W. van der Aalst (2004). A formal modeling approach for supply chain event management. Technical report, Proceedings of the 10th ACM Symposium on Access Control Models and Technologies InternetSource: http: wwwpersonal.psu.edu/faculty/a/x/ axk41/WITS04–camera–SCEN.pdf.
Marabotti, D. (2005). Supply chain event management; category, function or just another buzzword? Internet Source: http://www.silvon.com/resourcectr/wtPaper.php.
Mies, A., Voß, S. (2004). The Value of Contemporary Information in the Context of Supply Chain Event Management, in: Li L, Yen KK (eds.), *Proceedings of the Third International Conference on Information, International Information Institute*, Tokyo, 482 – 485.
Mirabedini, A., S.D'Amours, and D. Ait-Kadi (2001). Tactical planning of reliable supply chain. Internet Source: http://www.umoncton.ca/cie/conferences/29thconf/29thlCCIE/ Papers%5C paper 039.pdf.
Missbauer, H. (1998). Bestandsregelung als Basis f'ur eine Neugestaltung von PPS Systemen. *Ph.D.-Thesis*, University of Heidelberg, Heidelberg. Physica Verlag.
Otto, A. (2003). Supply Chain Event Management: Three Perspectives, *The International Journal of Logistics Management*, 14, Nr. 2, 1 – 13.
Pahl, J., Voß, S., Woodruff D. L. (2004). Production Planning with Load Dependent Lead Times. *Working Paper*, University of Hamburg.
Pahl, J., Voß, S., Woodruff D. L. (2005). Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling in Kotzab H, Seuring S, Reiner G, (eds.) (2005): *Research Methodologies in Supply Chain Management, Physica*, Heidelberg, 540 – 554.
Sürie, C., Wagner, M. (2002). Supply Chain Analysis. In: Stadtler, H., Kilger, C. (2002). Supply Chain Management an Advanced Planning – Concepts, Models, Software and Case Studies. Springer Verlag, 29 – 44.
Vernon, M. (2004). Ready for anything? Internet Source: http://www.computerweekly.com/Articles/Article.aspx?liArticleID=200578.html.
Voß, S., Woodruff D (2003). Introduction to Computational Optimization Models for Production Planning in a Supply Chain, Springer, Berlin.
Voß, S., Woodruff D (2004) A model for multi-stage production planning with load dependent lead times. In: R.H. Sprague (ed.) *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, IEEE, Piscataway, DTVEA03: 1-9
Zijm, W. (2000). Towards Intelligent Manufacturing Planning and Control Systems, *OR Spektrum*, 22: 313 – 345
Zimmermann, R. and A. Paschke (2003). Pamas - an agentbased supply chain event management system. Technical report, Americas Conferene on Information Sstems (AMCIS03). Internet Source: http://ibis.in.tum.de/staff/paschke/docs/PAMAS AMCIS 03.pdf.
Zimmermann, R., Winkler, S., Bodendorf, F. (2006). Agent-based Supply Chain Event Management – Concept and Assessment. To appear at Hawai'I International Conference on System Sciences (HICSS).

PRODUCTION PLANNING WITH DETERIORATION CONSTRAINTS: A SURVEY

J. Pahl[1], S. Voß[2], D. L. Woodruff[3]

[1,2]Institute of Information Systems, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
[3]Graduate School of Management, University of California, Davis CA 95616, USA

Abstract

As organizations move from creating plans for individual production lines to entire supply chains it is increasingly important to recognize that planning decisions impact the quality of parts and products and, consequently, the production output. The consideration of product characteristics, e.g., fixed or uncertain lifetimes, is vital to production systems dealing with products that deteriorate over time. Products that pass their useful lifetime can impose high costs due to inventory loss or rework of items. This paper surveys the recent trends in modeling deterioration in the various fields of production planning and gives an extensive overview of the subject.

Keywords:
Deterioration, Ordering and Inventory Management, Production Planning, Rework

1 INTRODUCTION

Constraints on the lifetime of items force organizations to carefully plan their production in cooperation with their supply chain partners up- and downstream along the chain. This is important, because waiting times due to suboptimal planning give rise to increasing lead times and, consequently, to decreasing quality of items so that, in the worst case, they cannot be used. This has important consequences with increased costs as one concern. Another more troubling aspect is unsatisfied customers waiting for their products.

Deterioration has a great influence not only on inventory management, but on every area of production where items are stocked or forced to wait due to technical matters, variabilities or disruptions / stochastic influences in the production process. Not surprisingly, scrap is one of the major causes of inventory loss regardless of the types of industries or products.

The aspect of deterioration and perishability of items came to prominence with the modeling of inventory in blood bank management and the disruption of blood from transfusion centers to hospitals. Later, the interest shifted to other products as well, e.g., chemicals, food, various drugs, fashion clothes, technical components, newspapers etc. In accordance, the attention changed to a more general view on how to include deterioration and perishability constraints into mathematical models for the various perspectives of production planning, e.g., order / replenishment and inventory management, lot sizing, aggregate production planning, and rework. Rework encompasses all actions required to transform products that do no meet a pre-specified quality standard (anymore) into such that fulfill the quality requirements.

The deterioration of goods is regarded as the process of decay, damage or spoilage of items in such a way that they can no longer be used for their original purpose, i.e., they go through a change in storage and lose their utility partially or completely. This is a continuous process so that such items have a stochastic lifetime in contrast to perishable goods that are considered as items with a fixed, maximum lifetime. Such products become obsolete at some fixed point of time because of various reasons, e.g., external regulations for pharmaceuticals that predetermine their shelf-life. In order to make a clear distinction, the term “perishability” is used for items that cannot be employed anymore and lose all their utility at once after a certain point of time whereas “deterioration” is employed for items that lose their utility gradually. This is somewhat a fuzzy definition, because it depends on the physical status (fitness) and behavior of these items over time and on the production planner or quality controller who has to decide whether to still employ items or components having passed a certain age.

Older excellent and comprehensive surveys in this area seem to be outdated (see, e.g., [1], [2], [3], [4], [5], [6]) because this field has seen increasing interest especially in the last decade. This paper reviews the problem areas regarding production planning and supply chain planning with deterioration constraints and gives an overview of the work already done on this subject. The approaches in the literature are classified along the supply chain, viz. starting with order and replenishment management, inventory control, aggregate production planning and lot sizing, inventory management and delivery as well as external / internal reverse logistics including rework.

The remainder of this paper is organized as follows. In Section 2 we give a classification of deterioration including its characterization as well as a classification of approaches that include deterioration constraints. Section 3 provides a comprehensive overview of the problem of deterioration in the supply chain with emphasize on the production process. The various manufacturing segments (functions) are considered from order management through aggregate production planning and lot sizing to inventory management and delivery. Furthermore, rework is regarded as part of the internal and external reverse logistics system. The paper concludes with some future research directions.

2 CLASSIFICATION OF DETERIORATION

Deteriorating items came into prominence in the 1970’s where the prevailing concern of modeling deterioration, and more specifically perishability, was in the field of blood bank management where a large amount of work exists; see, e.g., [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], and [19]. Comprehensive overviews on the early approaches are given by [5] and [6]. A more recent summary on blood bank management is provided by [20]. More general reviews on deterioration and its effects especially in regard of inventory management are provided by [3], [1] and [2]. Additionally, an overview of models considering rework of items is presented by [21] which is a related subject.

2.1 Characterization of deterioration

Deterioration and perishability of an item can be characterized regarding different aspects, e.g., its lifetime,

its physical depletion or decay, its value or utility loss. The lifetime of an item can be fixed (time-independent) or random (time-dependent) where the first case indicates that the product's lifetime or period of usage is pre-specified and independent of, e.g., environmental factors. The utility of these products is given during a specific time. Thereafter their value perishes completely. In the second case, there exists no specific lifetime of a product. For instance, food items are, among others, dependent on the season of the year and on the environmental conditions, e.g., hot or cold, long or short summer / winter periods. In the literature, approaches can be found including time-dependent deterioration or lifetime probabilities as, e.g., gamma distributions, Weibull, two-parameter Weibull, or exponential distributions. The value or utility loss of items is emphasized by [2], [3], [22], and [4] who further distinguish between the deterioration of the actual functionality of items (e.g. fruits, vegetables, milk) and the customers' perceived utility loss where the actual functionality remains intact (e.g., fashion clothes, high technology products, newspapers).

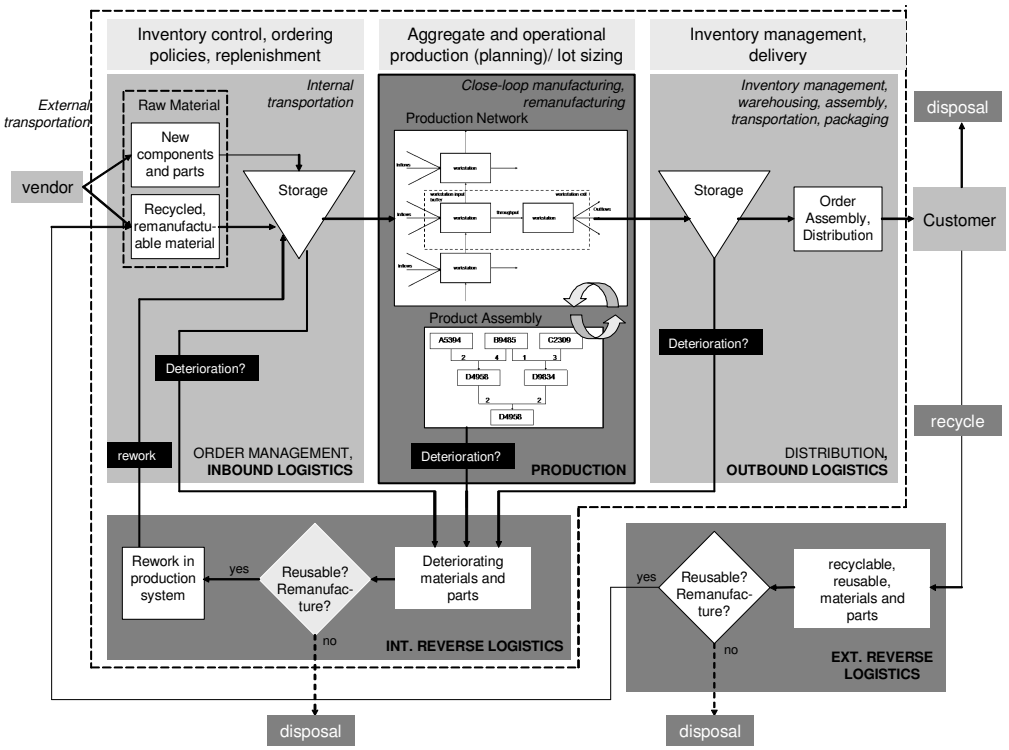

Figure 1: Material flow along the supply network, modified from [28]

2.2 Classification of approaches with deterioration

Frequently constraints are integrated into the mathematical models with the aim of rendering them more realistic. Consequently, models can be classified using the following list of several criteria (see [25]):

- Fixed or random lifetimes,
- Constant / varying deterioration rate,
- Deterministic / probabilistic and constant / time-varying demand,
- Single / multiple items,
- Single / multiple time periods,
- Shortages / no shortages with complete / partially backlogging,
- Economic models including price discounts, permissible delay in payments, inflation and the like,
- Queuing models,
- Purchase / manufacturing models
- Supply chain models (integrated setting, multi-stage models).

Additionally, demand plays a very important role in modeling deterioration. Thus different types of demand patterns can be found in the literature, i.e., constant or time-varying demand where the latter is frequently used to describe demand or sales in different product life cycle phases in the market (see [26] and [27]), e.g., deterministic demand together with time-dependent, stock-dependent, or price-dependent demand, or stochastic demand with known probability distributions or arbitrary probability distributions.

3 MODELING DETERIORATION

In order to describe the problem of deterioration and its impact on the various segments of manufacturing, Figure 1 shows the material flows along the supply network highlighting the interdependencies of the different manufacturing segments.

3.1 Ordering, replenishment and inventory control

The production process starts with the need for raw materials and, consequently, with the supplier (vendor) who interfaces with the order management of a manufacturer. The order management ("inbound logistics") has the assignment to provide the sufficient amount of raw materials of the desired quality for production at the right time at minimal ordering costs. This includes the avoidance of waste or the need to rework deteriorating items, if possible. Here, lot sizing already plays an important role, because raw materials are generally ordered in batches in order to take advantage of price and volume discounts and to minimize transportation costs. This might not hold for order contracts on a just in time basis or if the manufacturer has agreed on a vendor

managed inventory concept where the supplier is responsible for the inventory level and, consequently, for the replenishment batches of the manufacturer. This stream of literature is dominated by economic order models emphasizing the financial aspect that integrate different demand, pricing, payment and inflation scenarios and constraints, e.g., announced price increases / decreases or (volume) discounts, permissible delay in payments (see [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45]), inflation (see [46],) or, more generally, time value of money (see [47], [48], [49], [50], [51], [52], [53], [54], [55], [56]). Further constraints are included that aim at rendering the models more realistic, e.g., non-zero order lead times; see [56]. Other approaches with non-zero lead times model perishability; see [57], [58], [59], or [60] just to mention a few.

Deterioration complicates the replenishment process, especially if demand of *finished good inventories* (FGI) and raw material is uncertain or random. Items that have to wait too long for processing must be discarded or, if technically possible, reworked. Hence, deteriorated items are passed to the internal reverse logistic system which decides to rework or dispose them. Items can be reworked either on machines of the regular production system or on other machines parallel to the production system. These different approaches are discussed in Section 3.5.

Ordering decisions are highly dependent on the demand for FGI of the manufacturer and consequently on the decisions taken in aggregate production planning and forecasting. As a result, models exist that integrate the perspective of a vendor together with the perspective of a buyer show that significant cost reductions can be achieved by cooperative policies that determine, e.g., the optimal number of deliveries; see the approaches of [61], [62], [63], and [64].

3.2 Aggregate and operational production planning

Aggregate and operational production planning spans strategic, tactic and operational decisions of production planning whereas strategic questions deal with long-term decisions, e.g., in which markets to compete with which products and with which resources. This includes questions about the vertical integration (lean production, outsourcing, etc.). Tactical questions concern the implementation of the decisions from the strategic planning level and regard mid-range decisions, e.g., production network design and layout, arrangement and number of machines, and production mix. Finally, operational questions concern implementation of tactical decisions. Here we include scheduling and sequencing of orders on the machines and adjustments in case of disturbances and disruptive events, e.g., machine breakdowns, idleness due to missing raw material or *work in process* (WIP) due to moving bottlenecks.

The optimization of layouts and operations of an inspection system used for detecting malfunctioning processors in a multi-stage production system for deteriorating items is studied by [65]. They define a deteriorating item as one that has acquired a defect due to a malfunctioning machine. However, emphasis is on the planning of the overall inspection capacity, the assignment of operational tasks to the inspectors, and the scheduling of the tasks in order to avoid deterioration of items. Correspondingly, the deteriorating of capacity, viz. the physical deterioration of resources is studied by [66] including frequent machine breakdowns and downtimes due to the age of the resource. Deterioration affects capacity which is lost due to increased repair and, thus, downtimes. This further increases production (operating) costs due to maintenance actions and shortfalls of available production time to produce items that are incorporated in the objective function of the model. The combination of deteriorating items and processes is analyzed by [67]. They develop an integrated model for the joint determination of production quantity, inspection schedules, and quality control for an imperfect production process with deteriorating items. Similarly, optimal inspection policies for a (serial) production system including combinations of rework, repair, replacement, and disposal are proposed by [68].

Deterioration is especially important regarding the tactical and operational level of production. Scheduling and sequencing become more complex due to deterioration constraints that set limits on the time items can wait in front of a machine in order to be processed. A model that integrates production planning and inventory management in regard of deterioration with special emphasize on forecasting is proposed by [69]. He considers a production environment for cottage cheese subject to hardly predictable and highly influenced variable demand with peaks and lows as a result of, e.g., seasonal or customer fluctuations. Production planning is complex due to machine breakdowns or personnel illness and an increased deterioration rate. Therefore, production must closely follow variable demand leading to hectic production processes in peak situations and idle times in low demand situations. This leads to high personnel costs and over capacity that must be held in order to guarantee satisfactory service levels. Production quantities have to be decided on a daily basis following uncertain demand. As a result, forecasting is very important. Accordingly, they develop a computer software package that determines forecasts on daily demands and computes optimal decision rules for the short-term tuning of the daily production rate.

3.3 Lot Sizing

Lot sizing is a very important determinant of lead times and, therefore, of the quality of items, because decisions on lot sizes determine the frequency of setup times necessary to switch from the production of one product to the production of another and in some cases the time that items wait for lot completion before moving. Consequently, capacity is increasingly consumed that, otherwise, would be available for production. Accordingly, setup times caused by lot sizing decisions affect lead times of products due to induced waiting times. Lead times increase with the number (and duration) of setups and so does the number of products that deteriorate while waiting in line to be processed. Therefore, lot sizing decisions must be made taking into consideration deteriorating rates as well as the age of inventories.

In general, lot sizing decisions are integrated into production-inventory models. For instance, the economic production quantity model regards inventory control and determines the lot sizes of a single product that is produced on a single machine so as to meet its deterministic demand over an infinite planning horizon (see [70]). Hence, the objective is to determine a production schedule that minimizes the total costs and time of inventory holding, production, and setups.

A model that spans lot sizing, production, inventory and the customer perspective is presented by [71] including deterioration considerations and imperfect quality. Since collaboration of firms is vital in order to reduce overall supply chain costs, joint strategies are emphasized. In case of production-inventory management this can be achieved by, e.g., *joint economic lot size* models that decrease total relevant inventory costs for the supply chain generating overall benefits that can be shared by all supply chain partners. In addition to deterioration, shortages, and partial backlogging, imperfect quality is considered that is independent of deterioration. The supplier's perspective is integrated by [72]. Their

integrated (single-item multi-echelon) supply chain includes a single supplier, producer, and customer. The model aims at deriving the optimal number of deliveries and order lot sizes while minimizing the joint total costs. In contrast to [71], shortages or backlogging is not allowed. This is relaxed in [73] considering shortage effects on the downstream supply chain partners. In contrast with other models, the approach of [72] considers different deterioration rates of the specific supply chain partners.

3.4 Inventory management and delivery

The main objective of inventory management for deteriorating items is to obtain optimal returns during the useful shelf-life of the product. This leads to mainly three issues: determining reasonable and appropriate methods for issuing inventory, replenishing inventory, and allocating inventory.

A single-item production-inventory model that considers stopping and restarting times of production is presented by [74]. This model generally treats the interface between production and inventory management, but can also be employed to model the situation between two machines in a production system with buffer inventory (see Figure 1). In general, certain courses of inventory are assumed. Therefore, inventory intervals are assumed to start with positive, negative or zero inventory. In the model of [74], inventory builds continuously up in the second interval until stopping of production. Thereafter, inventory begins to decrease until the occurrence of shortages. This indicates the third interval. In the fourth interval, production restarts. The model is extended by [75] considering a time-dependent backlogging rate.

Besides, there is a body of literature that addresses the situation of multiple inventory locations. For instance, a situation of price discounts for high volume (bulk) purchases is studied by [46]. Since the capacity of the decision maker's own warehouse is limited, additional items have to be stored in a rented warehouse subject to extra costs. Thus, the advantage of price discounts has to be weighed against the extra (and higher than the owned warehouse's) costs for a rented warehouse. In addition, the deterioration rate of the rented warehouse is superior to the own warehouse's rate. Shortages are allowed and completely backlogged. The assumption that unit costs of rented warehouses are higher than these of owned warehouses is relaxed by [76]. He argues that unit costs of rented warehouses are more realistically assumed to be lower than own warehouse's costs as a result of optimized warehouse operations, specialized personnel, state of the art equipment, learning effects and economies of scale resulting from higher volumes. Besides, he studies a FIFO policy and finds out that FIFO is a better policy than LIFO especially in regard of deterioration of items. The main objective of his model is to determine optimal cycle times in a two warehouse setting. A similar model is proposed by [36] especially regarding permissible delay in payments. Shortages are not allowed, replenishment lead times and transportation costs are assumed to be zero and deterioration follows an exponential distribution.

A more strategic perspective is employed by [77] who proposes a stochastic set-covering location model with deterioration as well as amelioration of items with the aim of deriving the minimum necessary number of storage facilities so that the probability of covering all customers is not less than a critical value. Therefore, he assumes a supply chain setting with supply centers and retailers.

3.5 Rework

As mentioned before, an important aspect of deterioration is the required additional resource utilization caused by reworking the deteriorated parts and items. The increased utilization of production capacity leads to increased lead times, because lead times are dependent on the load of a facility or production machine. This phenomenon is extensively reviewed in [78]. Consequently, if the load increases due to additional machine utilization caused by rework, lead times increase and with them the number of items that have to be reworked, because they passed a certain threshold of their shelf-life. Therefore, it is mandatory to consider this vicious cycle in aggregate production planning and to develop mathematical models taking these complications into consideration.

As mentioned in Section 3.1, items can be reworked either on machines of the regular production system or on other machines parallel to the production system. In the first case approaches exist to optimally plan regular production and rework on the same machines (see, e.g. [68], [79], [80], or [81]). A survey on rework in planning and control of process industries is given by [21].

4 CONCLUSIONS AND FUTURE RESEARCH

We have seen that the consideration of deterioration is widely applied in various fields of production planning such as, e.g., ordering and replenishment policies, inventory management, lot sizing and rework management. We have provided an extensive review of the incorporation of deterioration in the various aspects of production planning. There exist only a few approaches of aggregate production planning taking these production item characteristics into account, which makes it ripe for further research.

5 REFERENCES

[1] Hariga M., 1996, Optimal EOQ models for deteriorating items with time-varying demand, Journal of the Operational Research Society, 47(10), 1228-1246.

[2] Goyal S. K., Giri B. C., 2001, Recent trends in modeling deteriorating inventory, European Journal of Operational Research, 134(1), 1-16.

[3] Raafat F., 1991, Survey of literature on continuously deteriorating inventory models, Journal of the Operational Research Society, 42(1), 27-37.

[4] Nahmias S., 1982, Perishable inventory theory: A review, Operations Research, 30(4), 680-708.

[5] Naddor R., 1966, Inventory Systems, John Wiley and Sons, New York.

[6] Prastacos G. P., 1984, Blood bank inventory management: An overview of theory and practice, Management Science, 30(7), 777-800.

[7] Brodheim E., Prastacos G. P., 1979, The long island blood distribution system as a prototype for regional blood management, Interfaces, 9(5), 3-20.

[8] Chazan D., Gal S., 1977, A Markovian model for a perishable product inventory, Management Science, 23(5), 512-521.

[9] Cohen M. A., Pekelman D., 1978, LIFO inventory systems, Management Science, 24(11), 1150-1162.

[10] Cooper W. L., 2001, Pathwise properties and performance bounds for a perishable inventory system, Operations Research, 49(1), 455-466.

[11] Dumas M. B., Rabinowitz M. 1977, Policies for reducing blood wastages in hospital blood banks, Management Science, 23(10), 1124-1132.

[12] Fernández B. D., del Brío Gonzáles J. A., 2001, Modelización de un DSS para la Gestión de Productos Perecederos, Qüestiió, 25(2), 287-300.

[13] Frankfurter G. M., Kendall K. E., Pegels C. C., 1974, Management control of blood through a short-term supply-demand forecast system, Management Science, 21(4), 442-452.

[14] Goh C.-H., Greenberg B. S., Matsuo H., 1993, Two-stage perishable inventory models, Management Science, 39(5), 633-649.

[15] Jagannathan R., Sen T., 1991, Storing crossmatched blood: a perishable inventory model with prior allocation, Management Science, 37(3), 251-266.

[16] Jennings J. B., 1973, Blood bank inventory control, Management Science, 19(6), 637-645.

[17] Katasaliaki K., Brailsford S. C., 2006, Using simulation to improve the blood supply chain, Journal of the Operational Research Society, 57(4), 1-9.

[18] Kendall K. E., Lee S. M., 1990, Formulation blood rotation policies with multiple objectives, Management Science, 26(11), 1145-1157.

[19] Pierskalla W. P., Roach C. D., 1972, Optimal issuing policies for perishable inventory, Management Science, 18(11), 603-614.

[20] Pierskalla W. P., 2004, Blood banking supply chain management, in Brandeau M., Sainfort F., Pierskalla W. P., editors, Operations Research and Health Care, Kluwer Academic Publishers, New York, 103-147.

[21] Flapper S. D. P., Fransoo J. C., Broekmeulen R. A. C. M., Inderfurth K., 2002, Planning and control of rework in the process industries: A review, Production Planning & Control, 13(1), 26-43.

[22] Ferguson M., Koenigsberg O., 2003, Production and pricing decisions: How should a firm manage a product with deteriorating quality? Working Paper, DuPree College of Management, Columbia University, Atlanta.

[23] Yao M.-J., Wang Y.-C., 2005, Theoretical analysis and a search procedure for the joint replenishment problem with deteriorating products, Journal of Industrial and Management Optimization, 1(3), 359-375.

[24] Wee H.-M., 1993, Economic production lot size model for deteriorating items with partial backordering, Computers & Industrial Engineering, 24(3), 449-458.

[25] Alfares H. K., Khursheed S. N., Noman S. M., 2005, Integrating quality and maintenance decisions in a production-inventory model for deteriorating items, International Journal of Production Research, 43(1), 899-911.

[26] Jain K., Silver E. A., 1994, Lot sizing for a product subject to obsolescence or perishability, European Journal of Operational Research, 75(2), 287-295.

[27] Mehta N. J., Shah N. H., 2005, Time varying increasing demand inventory model for deteriorating items, Investigación Operacional, 26(1), 3-9.

[28] Sarkis J., Meade L. M., Talluri S., 2004, E-logistics and the natural environment, Supply Chain Management: An International Journal, 9(4), 303-312.

[29] Aggrawal S. P., Jaggi C. K., 1995, Ordering policies of deteriorating items under permissible delay in payments, Journal of the Operational Research Society, 46(5), 658-662.

[30] Chang C.-T., 2004, An EOQ model with deteriorating items under inflation when supplier credits lined to order quantity, International Journal of Production Economics, 88(3), 307-316.

[31] Chang H.-J., Dye C.-Y., 2001, An inventory model for deteriorating items with partial backlogging and permissible delay in payments, International Journal of Systems Science, 32(3), 345-352.

[32] Chang H.-J., Wu S.-J., 2003, A note on optimal payment time under permissible delay in payment for products with deterioration, Production Planning & Control, 14(5), 478-482.

[33] Chang C.-T., Teng J.-T., 2004, Retailer's optimal ordering policy under supplier credits, Mathematical Methods of Operations Research, 60(3), 471-483.

[34] Chang H.-J., Dye C.-Y., 2005, The effect of credit period on the optimal lot size for deteriorating items with time varying demand and deterioration rates, Asia Pacific Journal of Operational Research, 22(2), 211-227.

[35] Chung K.-J., Liao J.-J., 2004, Lot-sizing decisions under trade credit depending on the ordering quantity, Computers & Operations Research, 31(6), 909-928.

[36] Chung K.-J., Huang T.-S., 2006, The optimal cycle time for deteriorating items with limited storage capacity under permissible delay in payments, Asia Pacific Journal of Operational Research, 23(3), 347-370.

[37] Dye C.-Y., Chang H.-J., 2003, A replenishment policy for deteriorating items with linear trend demand and shortages when payment periods are offered, Information and Management Science, 14(2), 31-45.

[38] Goyal S. K., 1985, Economic order quantity under conditions of permissible delay in payments, Journal of the Operational Research Society, 36(4), 335-338.

[39] Jamal A. M. M., Sarker B. R., Wang S., 1997, An ordering policy for deteriorating items with allowable shortage and permissible delay in payment, Journal of the Operational Research Society, 48(8), 826-833.

[40] Liao H.-C., Chen Y.-K., 2003, Optimal payment time for retailer's inventory system, International Journal of Systems Science, 34(4), 245-253.

[41] Maddah B. S., Jaber M. Y., Abboud N. E., 2004, Periodic review (s,S) inventory model with permissible delay in payments, Journal of the Operational Research Society, 55(2), 147-159.

[42] Ouyang L.-Y., Chang C.-T., Teng J.-T., 2005, An EOQ model for deteriorating items under trade credits, Journal of the Operational Research Society, 56(6), 719-726.

[43] Ouyang L.-Y., Teng J.-T., Chen L. H., 2006, Optimal ordering policy for deteriorating items with partial backlogging under permissible delay in payments, Journal of Global Optimization, 34(2), 245-271.

[44] Sarker B. R., Jamal A. M. M., Wang S., 2000, Supply chain models for perishable products under inflation and permissible delay in payments, Computers & Operations Research, 27(1), 59-75.

[45] Shah N. H., 2004, Probabilistic order level system when items in inventory deteriorate and delay in payments is permissible, Asia Pacific Journal of Operational Research, 21(3), 319-331.

[46] Yang H.-L., 2004, Two-warehouse inventory models for deteriorating items with shortages under inflation, European Journal of Operational Research, 157(2), 344-356.

[47] Abad P. L., 1996, Optimal pricing and lot-sizing under conditions of perishability and partial backordering, Management Science, 42(8), 1093-1104.

[48] Abad P. L., 2003, Optimal pricing and lot-sizing under conditions of perishability, finite production and partial backordering and lost sales, European Journal of Operational Research, 144(3), 677-685.

[49] Balkhi Z. T., 2004, On the optimality of inventory models with deteriorating items for demand and on-

hand inventory dependent production rate, Journal of Management Mathematics, 15(1), 67-86.

[50] Chande A., Dhekane S., Hemachandra N., Rangaraj N., 2005, Perishable inventory management and dynamic pricing using RFID technology, Sadhana, 30(2-3), 445-462.

[51] Chandrashekar K., Dave N., Hemachandra N., Rangaraj N., 2003, Timing of discount offers for perishable inventories, Proceedings of the Sixth Asia Pacific Operations Research Society, New Delhi, 1-8.

[52] Lin Y., Lin C., 2006, Purchasing model for deteriorating items with time-varying demand under inflation and time discounting, International Journal of Advanced Manufacturing Technology, 27(7-8), 816-823.

[53] Mehta N. J., Shah N. H., 2003, An inventory model for deteriorating items with exponentially increasing demand and shortages under inflation and time discounting, Investigaçao Operacional, 23(1), 103-111.

[54] Moon I., Giri B. C., Ko B., 2005, Economic order quantity models for ameliorating / deteriorating items under inflation and time discounting, European Journal of Operational Research, 162(3), 773-785.

[55] Wee H.-M., Law S.-T., 2001, Replenishment and pricing policy for deteriorating items taking into account the time-value of money, International Journal of Production Economics, 71(1), 213-220.

[56] Yang H.-L., Teng J.-T., Chern M.-S., 2001, Deterministic inventory lot-size models under inflation with shortages and deterioration for fluctuating demand, Naval Research Logistics, 48(2), 144-158.

[56] Nahmias S., Wang S. S., 1979, A heuristic lot size reorder point model for decaying inventories, Management Science, 25(1), 90-97.

[57] Lawrence A. S., Sivakumar B., Arivarignan G., 2006, Perishable inventory system with random supply quantity and negative demands, Advanced Modeling and Optimization, 8(2), 151-168.

[58] Sivakumar B., Arivarignan G., 2006, A perishable inventory system at service facilities with negative customers, 17(2), 1-18.

[59] Kalpakam S., Shanthi S., 2006, A continuous review perishable system with renewal demands, Annals of Operations Research, 143(1), 211-225.

[60] Chiu H. N., 1999, A good approximation of the inventory level in a (Q,r) perishable inventory system, RAIRO Operations Research, 33(1), 29-45.

[61] Yang P.-C., Wee H.-M., 2002, A single-vendor and multiple-buyer production-inventory policy for a deteriorating item, European Journal of Operations Research, 143(3), 570-581.

[62] Yang P.-C., Wee H.-M., 2000, Economic ordering policy of deteriorated item for vendor and buyer: An integrated approach, Production Planning & Control, 11(5), 474-480.

[63] Chen J.-M., Chen L.-T., 2005, Pricing and production lot size / scheduling with finite capacity for a deteriorating item over a finite horizon, Computers & Operations Research, 32(11), 2801-2819.

[64] Wu M. Y., Wee H.-M., 2001, Buyer-seller joint cost for deteriorating items with multiple-lot-size deliveries, Journal of the Chinese Institute of Industrial Engineering, 18(1), 109-119.

[65] Emmons H., Rabinowitz G., 2002, Inspection allocation for multistage deteriorating production systems, IEE Transactions, 34(12), 1031-1041.

[66] Rajagopalan S., 1992, Deterministic capacity expansion under deterioration, Management Science, 38(4), 525-539.

[67] Rahim M. A., Ben-Daya M., 2001, Joint determination of production quantity, inspection schedule, and quality control for an imperfect process with deteriorating products, Journal of the Operational Research Society, 52(12), 1370-1378.

[68] Yum B. J., McDowell E. D., 1987, Optimal inspection policies in a serial production system including scrap rework and repair: An MILP approach, International Journal of Production Research, 25(10), 1451-1464.

[69] Bartmann D., 1983, Optimale Produktionsplanung bei Frischprodukten mit dem Programm FORTUNA – Eine Fallstudie, Zeitschrift für Operations Research, 27:B 179-B 204.

[70] Lin G. C., Kroll D. E., Lin C. J., 2006, Determining a common production cycle time for an economic lot scheduling problem with deteriorating items, European Journal of Operations Research, 173(2), 669-682.

[71] Wee H. M., Yu J. C. P., Wang K. J., 2006, An integrated production-inventory model for deteriorating items with imperfect quality and shortage backordering considerations, Lecture Notes in Computer Science, Springer, 3981, 885-897.

[72] Rau H., Wu M.-Y., Wee H.-M., 2003, Integrated inventory model for deteriorating items under a multi-echelon supply chain environment, International Journal of Production Research, 86(2), 155-168.

[73] Rau H., Wu M.-Y., Wee H.-M., 2004, Deteriorating item inventory model with shortage due to supplier in an integrated supply chain. International Journal of Systems Science, 35(5), 293-303.

[74] Yan H., Cheng T. C. E., 1998, Optimal production stopping and restarting times for an EOQ model with deteriorating items, Journal of the Operational Research Society, 49(12), 1288-1295.

[75] Skouri K., Papachristos S., 2003, Optimal stopping and restarting production times for an EOQ model with deteriorating items and time-dependent partial backlogging, International Journal of Production Economics, 81-82(1), 525-531.

[76] Lee C.-C., 2006, Two-warehouse inventory model with deterioration under FIFO dispatching policy, European Journal of Operational Research, 174(2), 861-873.

[77] Hwang H.-S., 2004, A stochastic set-covering location model for both ameliorating and deteriorating items, Computers & Industrial Engineering, 46(2), 313-319.

[78] Pahl J., Voß S., Woodruff D. L., 2005, Production planning with load dependent lead times, 4OR: A Quarterly Journal of Operations Research, 3(4), 257-302.

[79] Teunter R. H., Flapper S. D. P., 2003, Lot-sizing for a single-stage single-product production system with rework of perishable product defectives, OR Spectrum, 25(1), 85-96.

[80] Inderfurth K., Lindner G., Rachaniotis N. P., 2005, Lot sizing in a production system with rework and production deterioration, International Journal of Production Research, 43(7), 1355-1374.

[81] Inderfurth K., Lindner G., Rachaniotis N. P., 2007, Cost minimizing scheduling of work and rework processes on a single facility under deterioration of reworkables, Computers & Operations Research, 105(2), 345-356.

Discrete Lot-Sizing and Scheduling Including Deterioration and Perishability Constraints

Julia Pahl and Stefan Voß

University of Hamburg, Institute of Information Systems, D-20146 Hamburg
pahl@econ.uni-hamburg.de, stefan.voss@uni-hamburg.de
https://iwi.econ.uni-hamburg.de/IWIWeb/Default.aspx

Abstract. Constraints on the lifetime of items force organizations to carefully plan their production in cooperation with their supply chain partners up- and downstream along the chain. This is important because waiting times due to suboptimal planning give rise to increasing lead times and, consequently, to deterioration and thus decreasing quality of items so that, in the worst case, they cannot be used. Increased costs, delivery delays, quality decreases and unsatisfied customers are negative effects that can be avoided by accounting for product depreciation in the production process. We highlight the importance of including deterioration and perishability in planning decisions and present well-known discrete lot-sizing and scheduling models extended in this regard. The effects on plans derived by these models including depreciation is shown using a numerical example.

Keywords: Production Planning, Lot-Sizing, Scheduling, Deterioration, Perishability.

1 Introduction

Deterioration has a great influence not only on inventory management, but on every area of production where items are stocked or forced to wait due to uncertain demand, technical matters, variabilities or disruptions of the production process. Not surprisingly, scrap is one of the major causes of inventory loss regardless of the types of industries or products.

The aspect of deterioration and perishability of items came to prominence with the modeling of inventory in blood bank management and the distribution of blood from transfusion centers to hospitals. Later, the interest shifted also to other products, e.g., chemicals, food, various drugs, fashion clothes, technical components, newspapers etc. In accordance, the attention changed to a more general view on how to include deterioration and perishability constraints into mathematical models for production planning in different settings, e.g., regarding multiple facilities, warehouses, as well as supply chain perspectives.

In this paper we propose discrete lot-sizing and scheduling models accounting for depreciation effects of products in the production process. While depreciation effects are widely regarded in the area of stochastic and continuous

W. Dangelmaier et al. (Eds.): IHNS 2010, LNBIP 46, pp. 345–357, 2010.
© Springer-Verlag Berlin Heidelberg 2010

inventory management, only very few models have been proposed for the deterministic, discrete case. Nevertheless, it may be useful to examine depreciation effects in non-complex, straightforward settings in order to develop a better understanding for more complex situations. Therefore, we analyze the *capacitated lot-sizing problem* (CLSP) extending it to deterioration and perishability effects. Furthermore, scheduling decisions are taken into consideration, too. The effects of depreciation are emphasized using a numerical example which is certainly not representable for practical cases due to its small size, but very useful regarding the effects on planning decisions.

The remainder of the paper is organized as follows. In Section 2 we give a review on the related literature on depreciation of products (Section 2.1) and lot-sizing and scheduling models (Section 2.2). In Section 3 we present the CLSP with perishability and deterioration further extending it to scheduling decisions. Section 4 provides a numerical example used to highlight computational results for two of the proposed scheduling models and Section 5 concludes with ideas for future research.

2 Literature Review

Depreciation effects are especially interesting in situations where items are forced to wait in the production process. The task of lot-sizing is to plan quantities of items produced together so as to minimize certain cost factors such as order or setup costs and inventory holding costs. For instance, the importance of accurately accounting for setup times and costs and their effects on capacity utilization has been highlighted by the discussion of just-in-time manufacturing [1] which requires small and frequent batch sizes in order to decrease WIP inventory. This influences capacity utilization, because setup times decrease machine capacity utilization and represent idle times.

Scheduling decisions play an important role regarding overall processing or lead times of items. If products are subject to enhanced deterioration or perishability it may be preferable to schedule them so that they do not spend much time waiting in inventory. This is even more true for multi-level situations regarding multiple (sequenced) machines. Consequently, when regarding depreciation effects scheduling decisions should be taken into account.

2.1 Depreciation of Parts and Products

All products deteriorate sooner or later depending on the considered time horizon. If the time horizon is chosen to be sufficiently short there may be no need to consider such product characteristics. Nevertheless, in practice and depending on the product, the time horizon for tactical and even operational planning is long enough, so that the influence of deterioration and perishability tends to play an important role.

Deterioration or *perishability* of goods is regarded as the process of decay, damage or spoilage of items such that they can no longer be used for their original purpose, i.e., they go through a change while stored and lose their utility

partially or completely [2,3,4,5,6]. Deterioration is a continuous process so that items have a stochastic lifetime in contrast to perishable goods which are considered as items with a fixed, maximum lifetime [7]. This is true for products that become obsolete at some fixed point in time, due to various reasons or external factors, e.g., change in style, technological developments, outdates or external regulations (e.g. pharmaceuticals) that predetermine their shelf-lifes. Consequently, deterioration or perishability can be characterized by different aspects. [4] give three characteristics of decay of goods: *direct spoilage* (e.g. fresh food, vegetables), *physical depletion* (e.g. gasoline, alcohol), or the *loss of efficacy* in inventory (e.g. electronic components, medicine). A detailed discussion of a classification of deterioration and perishability is given in [8].

The lifetime of products can be regarded fixed or random: the first case implies that the products' lifetime is pre-specified and time-independent of deterioration factors. The utility of these products is given during a specific time period. When the products pass their lifetime, they perish completely. In the second case there is no specific lifetime for the products, instead there is depreciation over time. For instance, the lifetime of food items can be seasonal as, e.g., vegetables deteriorate faster in summer time. This kind of deterioration is called time-dependent. In the literature we find the integration of lifetime probabilities as, e.g., gamma distributions, Weibull, two-parameter Weibull, or exponential distributions. An alternative basis for classification is the *value loss* of the goods; see [9]. [10] emphasize the related utility loss and distinguishes products whose actual functionality deteriorates over time, e.g., fruits, vegetables, or milk, and those whose functionality does not degrade, but customers' perceived utility deteriorates over time which can be true for, e.g., fashion clothes, high technology products with a short life cycle, or items whose information content deteriorates, e.g., newspapers. For instance, [11] prove empirically that the willingness to pay for a product continuously decreases with its perceived actuality or utility; see also [12]. So this is somewhat fuzzy, as it clearly depends on customer preferences if a good still has some value, viz. if a customer still regards it as valuable or useful. A multitude of authors in the field of deterioration and perishability use the words interchangeably.

We make a clear distinction between perishable goods that cannot be used anymore and lose all their utility at once after a certain point in time whereas items that lose their utility gradually are regarded as being subject to continuous deterioration. For instance, Figure 1 depicts three functional relationships of the value of items in time where the first graph shows the course of perishability, the second the discrete and variational course of deterioration and the third three possible courses of continuous deterioration as defined above.

2.2 Simultaneous Lot-Sizing and Scheduling

Lot-sizing and scheduling decisions are generally considered in tactical and operational production planning. Often, these planning steps are executed sequentially which can transform formally feasible (tactical) lot-size plans to infeasible plans due to subsequent scheduling decisions that violate capacity constraints.

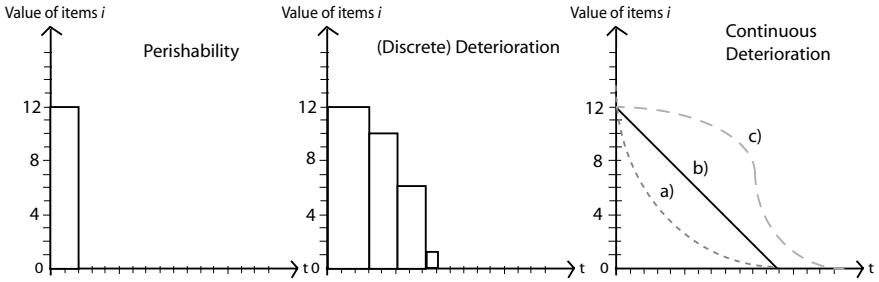

Fig. 1. Three examples for perishability and deterioration rates

Consequently, iterations of planning decisions and information are needed or, better, their simultaneous consideration. The classification of such models differentiates between single or multiple resources, single or multi-level structures of products, finite or infinite planning horizon, discrete or continuous time, and/or deterministic or stochastic input data. In this paper we concentrate on the deterministic, discrete single resource single structure case with finite planning horizon.

Discrete dynamic modeling assumes that the time dimension can be divided into "time buckets" that can be further segregated into "big bucket" and "small bucket" models. Big bucket models contain long time periods in which several different products can be set up and produced unlike small bucket models where time periods are rather short, so that startups, switch-offs and/or change-overs are modeled. Small bucket models can be further divided into those that allow at most one setup per period and those allowing more than one setup [13]. Simultaneous lot-sizing and scheduling models have been subject to extensive research for many decades. Consequently, we refer the reader to excellent surveys and reviews on (multi-level) lot-sizing and scheduling provided by [14], [15], [16], [17]. Nevertheless, discrete lot-sizing and scheduling decision models including deterioration and perishability constraints are rare. A short review on these models is given in the subsequent section.

2.3 Models Including Depreciation Constraints

Economic lot-sizing models (ELSP) including deterioration and perishability constraints are provided, e.g., by [18,19,20], where deterioration rates of inventory are dependent on the age of products. This is also true for the replenishment model of [21] where the shelf-life of items is constrained by two periods. [19] shows that the ELSP with perishability constraints is equivalent to a minimum cost network flow problem with flow loss. He analyzes special structural properties of the optimal solution in order to develop a dynamic programming algorithm able to solve the problem in polynomial time. Shortages and complete backorders are added to the model by [20]. [18] employ a Cobb-Douglas cost function in order to integrate effects of economies of scale.

Optimal utilization policies of time-varying deteriorating items are studied by [22] in the field of blood bank management. They define diverse age categories of items due to the fact that, in general, requested but unused blood preserves are returned to the blood bank after one or two days. Consequently, they do not begin to deteriorate in stock, but enter already aged. [23] studies the parallel machine scheduling problem with deteriorating jobs where the objective is to minimize the total completion time of jobs or total machine workload. Algorithms to solve the scheduling problem base on steepest descent search heuristics. [24] provides an *economic order quantity* (EOQ) model with deteriorating items and time proportional demand.

Our approach described in the next section integrates perishability that constrains the time items can be held in inventory. Thus, the inventory balance equation is subject to modifications. Furthermore, disposal costs are regarded in the objective function.

In order to consider deterioration and perishability constraints, we begin with the CLSP. We will extend it step by step in the course of our discussion considering the above mentioned aspects of lot-sizing and scheduling including sequence-dependent setup times and costs with depreciation.

3 Model Formulations

The CLSP presents a standard model for discrete, dynamic, capacitated multi-product lot-sizing where multiple products are planned on one machine or resource subject to finite capacity. Demand is deterministic and known in advance for each period t. Backlogging is not allowed and *first-in-first-out* (FIFO) is assumed regarding the stock outflow process. The objective is to find optimal production lot-sizes that minimize inventory and setup costs. Regarding perishability, we include the minimization of costs for disposal of perished items.

Given the variables, parameters, and indices in Table 1 that is valid for all presented models, the *CLSP with perishability constraints* (CLSP-P) is formulated as follows:

$$\min \sum_{i=1}^{M} \sum_{t=1}^{T} \left(K_i^s \delta_{it} + \pi_i I_{it} + \phi_i^D I_{it}^D \right) \tag{1}$$

subject to:

$$I_{it} = I_{i,t-1} + X_{it} - D_{it} - I_{it}^D \qquad \forall\, i, t \tag{2}$$

$$I_{it}^D \geq \sum_{\tau=0}^{t-\Theta_i} X_{i,\tau} - \sum_{\tau=0}^{t} D_{i\tau} - \sum_{\tau=0}^{t-1} I_{i\tau}^D \qquad \forall\, i, t \geq \Theta_i \tag{3}$$

Table 1. Variables, parameters and indices used in the models

Parameters	
K_i^s	setup costs of product i
ϕ_i^D	waste disposal cost factor for product i
π_i	inventory holding cost factor for product i
D_{it}	demand of product i in period t
ξ_i	resource consumption factor or production coefficient of product i
ϑ_i	setup time for product i
Cap_t	capacity that is available in period t
α_i^Θ	deterioration rate of product i
V	big number
Variables	
X_{it}	production of product i in period t
I_{it}	inventory holding of product i in period t
I_{it}^D	perished inventory of product i in period t
δ_{it}	setup indicator denoting if a product i is set up in period t
δ_{is}^s	setup state variable denoting if product i is set up at the end of period s ($\delta_{is}^s = 1$) or not ($\delta_{is}^s = 0$) used in scheduling models
Indices	
i	product with $i = 1, \ldots, M$
j	subset of products with $j \in M, i \neq j$
t	time period with $t = 1, \ldots, T$
τ	subset of periods
Θ_i	shelf-life of product i
s	micro-periods with $s = 1, \ldots, S$ used in scheduling models
u	subset of micro-periods used in scheduling models

$$\sum_{i=1}^{M} \xi_i X_{it} + \sum_{i=1}^{M} \vartheta_i \delta_{it} \leq \mathrm{Cap}_t \qquad \forall t \tag{4}$$

$$X_{it} \leq V \delta_{it} \qquad \forall\, i, t \tag{5}$$

$$I_{i0} = I_{iT} = 0 \qquad \forall\, i \tag{6}$$

$$X_{it}, I_{it}, I_{it}^D \geq 0 \qquad \forall\, i, t \tag{7}$$

$$\delta_{it} \in \{0, 1\} \qquad \forall\, i, t \tag{8}$$

The objective function (1) minimizes the sum of costs of setting up for product i in period t, inventory holding costs, and disposal costs of perished products. Equation (2) describes the inventory balance equation where inventory of product i in period t is composed by inventory of the previous period $t-1$ and production less products that are used to satisfy demand in that period and less the products that are deteriorated in that period. Inequality (3) presents items that need to be disposed in period t. These are calculated by the sum of produced items in the interval $[0, t - \Theta_i]$, thus production from the beginning

of the planning horizon until period t deduced by the shelf-life of items, less those items that are used to satisfy demand until period t and items that have been already disposed in the previous period $t-1$. In order to incorporate deterioration effects in the model, the parameter Θ_i is modified to Θ_{it}, so that changing deterioration rates over time are captured. Thus, the course of deterioration for items is defined via input data. This is valid also for the subsequent models. In our model formulation, we limit the mathematical presentation to the perishability case. Inequality (4) requires production and setup times to be smaller or equal available capacity in period t. Constraint (5) entails the machine to be set up for item i when production of that item takes place where V is a large number that needs to be chosen in a way that the production amount is not restricted. This constraint is also frequently formulated as follows:

$$X_{it} \leq \frac{\mathrm{Cap}_t}{\xi_i}\delta_{it} \quad \forall\, i, t \tag{9}$$

linking continuous variables X_{it} to the binary variables δ_{it} ensuring that whenever product i is produced in period t, thus $X_{it} > 0$, the related setup indicator δ_{it} is set, accordingly. Equality (6) denotes that initial and final inventory is zero. Non-negativity on the variables and the definition for the binary setup variable δ_{it} are given by (7) and (8), respectively.

Inequality (3) might be reformulated assuming that a certain amount of items in stock perish/deteriorate in a time period t independent of their age and how long they have been in stock. The inventory balance equation is reformulated, accordingly:

$$I_{it} = (1 - \alpha_i^{\Theta})I_{i,t-1} + X_{it} - D_{it} \qquad \forall\, i, t \tag{10}$$

where α_i^{Θ} denotes the deterioration rate of product i. A certain percentage of inventory of the previous period $t-1$ perishes in stock and need to be subtracted from the inventory balance. For instance, this formulation is used by [19,20].

The CLSP is a "big bucket" model and, consequently, scheduling decisions are not included. Additionally, setup states are not carried over from one period to the next, thus, regardless if a product i is planned on the machine in period $t-1$ and period t, the model induces two setups. Moreover, a lot-size cannot be split and produced in two consecutive periods, but has to be produced in one period. Neither is it possible to setup the machine twice for the same product. In other words, a product i can be produced only once in a period t which constrains the maximal possible number of setups to the number of products (types) i. Additionally, lot-sizes X_{it} are constrained by Cap_t/ξ_i.

In order to include scheduling decisions, the CLSP-P is modified to a *discrete lot-sizing and scheduling problem* with perishability (DLSP-P). It inherits the assumptions from the CLSP despite that the macro-periods of the CLSP are subdivided into micro-periods denoted by $s = 1, \ldots, S$. The DLSP-P further assumes that at most one product type i is produced in a period s, so that full capacity is utilized for that product. This is denoted as the *all-or-nothing*

assumption. Consequently, the capacity restriction of the CLSP, inequality (4), is reformulated considering micro-periods:

$$X_{is} = \frac{(\mathrm{Cap}_s - \vartheta_i)}{\xi_i}\delta_{is}^s \qquad \forall\, i, s \tag{11}$$

where δ_{is}^s is a setup state variable denoting that, e.g., no setup is required if the machine was left set up in the previous period for product i, thus $\delta_{is}^s = 0$. Consequently, the inventory balance equation is modified as (13) below. The setup state variables δ_{is}^s are linked to the setup costs $K_i^s\delta_{is}$ in the objective function by $\delta_{is} \geq \delta_{is}^s - \delta_{i,s-1}^s \quad \forall\, i, s$ denoting that if a setup in two consecutive periods takes place, a setup operation δ_{is} must have occurred. The complete DLSP-P is stated as follows:

$$\min \sum_{i=1}^{M}\sum_{s=1}^{S}\left(K_i^s\delta_{is} + \pi_i I_{is} + \phi_{is}^D I_{is}^D\right) \tag{12}$$

subject to

$$I_{is} = I_{i,s-1} + \frac{(\mathrm{Cap}_s - \vartheta_i)}{\xi_i}\delta_{is}^s - D_{is} - I_{is}^D \qquad \forall\, i, s \tag{13}$$

$$I_{is}^D \geq \sum_{u=0}^{s-\Theta_i}\frac{(\mathrm{Cap}_{s-\Theta_i} - \vartheta_i)}{\xi_i}\delta_{iu}^s - \sum_{u=0}^{s} D_{iu} - \sum_{u=0}^{s-1} I_{iu}^D \qquad \forall\, i, s \geq \Theta_i \tag{14}$$

$$\sum_{s=1}^{S}\delta_{is}^s \leq 1 \qquad \forall\, s \tag{15}$$

$$\delta_{is} \geq \delta_{is}^s - \delta_{i,s-1}^s \qquad \forall\, i, s \tag{16}$$

$$I_{i0} = I_{iS} = 0 \qquad \forall\, i \tag{17}$$

$$I_{is}, I_{is}^D \geq 0 \qquad \forall\, i, s \tag{18}$$

$$\delta_{is}^s \in \{0, 1\} \qquad \forall\, i, s \tag{19}$$

The preservation of setup states across periods gives a certain kind of continuity [25]. Nevertheless, still the all-or-nothing assumption underlies the model restricting lot-sizes to be produced within one period and not exceeding this period. This can be interpreted as a minimum lot size or batch. In practice, this might be required, e.g., due to technical issues in the production process. In fact, minimal lot sizes are the very hard constraints that lead to disposals, because enforcing production higher than demand due to capacity leading to increased inventory can cause disposals. Otherwise, disposals are always avoided independently how high setup costs might be due to the fact that these costs are due anyway within the production horizon in order to switch to production of another product type. Thus, producing at most the required demands is cost optimal. This is also highlighted in Section 4.

The problem of spoilage due to minimal lot sizes implied by the all-or-nothing assumption may be remedied by allowing production lot-sizes to take any value

up to the capacity limit [26] further carrying over setup states without incurring additional setup costs [27]. Consequently, the inventory balance equation (13) is changed as follows:

$$I_{is} = I_{i,s-1} + X_{is} - D_{is} - I_{is}^{D} \qquad \forall\, i, s \tag{20}$$

with the production lot-size being constrained as in (11). Such model formulation is denoted *continuous setup lot-sizing problem* (CSLP) in the literature, although without perishability constraints. Still, in the CSLP with perishability (CSLP-P) production capacity may be lost due to the constraint that only one product type i can be produced per period which is given by inequality (11). This problem is readdressed by the *proportional lot-sizing and scheduling problem with perishability* (PLSP-P). Therefore, the production constraints (11) and capacity constraints are reformulated:

$$X_{is} \leq \frac{(\mathrm{Cap}_i - \vartheta_i)}{\xi_i} \left(\delta_{is}^{s} + \delta_{i,s-1}^{s}\right) \qquad \forall\, i, s \tag{21}$$

$$\sum_{i=1}^{M} \xi_i X_{is} + \sum_{i=1}^{M} \vartheta_i \delta_{is} \leq \mathrm{Cap}_s \qquad \forall\, s \tag{22}$$

where the remainder of the CSLP-P remains unaltered.

4 Numerical Results

In this section we highlight the effects of the all-or-nothing assumption calculating the DLSP-P, and the PLSP-P using exemplified data sets. We consider four types of products and a planning horizon of 10 periods. The capacity of one machine is 70 units per period. The data sets used are rather small, because the aim is to provide some insights regarding the mechanisms and hard constraints of minimum lot sizes that lead to disposals. Applying larger data sets to the models leads to increased computational calculation times as variables and constraints increase exponentially. This is straightforward and experienced by extending the data set to 7 products and 50 periods.

The models are implemented in AMPL/CPLEX and are calculated on a computer with an Intel Core(R) Pentium(R) 4 CPU processor with 2.42 GHz and 1 GB RAM.

Due to the relative small numerical example the calculating time for finding the optimal solution is negligible. When extending the example to seven products and 20 time periods and thus dealing with 550 variables and 423 constraints the optimal solution is still found within seconds, but time begins to be relevant. The optimal solution for the DLSP-P is found with the production/inventory plan as given in Table 3 resulting in total costs of 6.440. The calculation employs 289 MIP simplex iterations including one branch-and-bound node.

The PLSP-P gives a plan that does not generate disposals. Therefore, disposals that are zero for all products in all periods are not included in Table 4. For

Table 2. Demand data for the small numerical example

	Cost Factors						Demand in Periods									
Products	K_i^s	π_i	ϕ_i^D	ϑ_i	Θ_i	ξ_i	1	2	3	4	5	6	7	8	9	10
A	400	2	10	10	2	1	10	10	–	–	–	40	–	–	–	60
B	400	1.5	10	10	2	1	–	–	–	5	30	30	–	–	40	–
C	400	0.5	10	10	2	1	–	–	10	–	10	–	10	–	10	–
D	400	0.5	10	10	2	1	–	–	–	–	10	20	30	–	–	60

Table 3. Results of the DLSP-P model

		Periods									
		1	2	3	4	5	6	7	8	9	10
Production	A	60	-	-	-	-	60	-	-	-	60
	B	-	60	-	60	-	-	-	-	60	-
	C	-	-	60	-	-	-	60	-	-	-
	D	-	-	-	-	60	-	-	60	-	-
Inventory	A	10	-	-	-	-	-	-	-	-	-
	B	-	5	5	60	30	-	-	-	-	-
	C	-	-	10	10	-	-	10	10	-	-
	D	-	-	-	-	50	30	-	60	60	-
Disposal	A	40	-	-	-	-	20	-	-	-	-
	B	-	55	-	-	-	-	-	-	20	-
	C	-	-	40	-	-	-	40	-	-	-
	D	-	-	-	-	-	-	-	-	-	-

the PLSP-P, the optimal solution results in 3.740 total costs. CPLEX employed 16.879 MIP simplex iterations in order to find the optimal solution including 1.187 branch-and-bound nodes. This is due to the fact that the DLSP-P contains the all-or-nothing assumption which leads to full utilization of capacities even if demands are not equally high. Therefore, inventory does not decrease due to demand, so that items perish after their lifetime and need to be disposed. This is not true for the PLSP-P where production utilization is as high as the demand, so that items do not wait in inventory and the risk of perishability is rather low. When comparing the results for the PLSP-P to the PLSP without perishability we see that the number of setups in the PLSP are relatively small, inventory higher and, therefore, total costs smaller than in the PLSP-P. This is due to the fact that products not subject to perishability can be hold longer in inventory than perishable items which is straightforward. Furthermore, when less capacity is available for production, inventory increases. This behavior

Table 4. Results of the PLSP-P model

		Periods									
		1	2	3	4	5	6	7	8	9	10
Production	A	10	10	-	-	-	40	-	-	60	
	B	-	-	-	5	60	-	-	-	40	-
	C	-	-	20	-	-	-	-	20	-	-
	D	-	-	-	30	-	-	30	40	20	-
Inventory	A	-	-	-	-	-	-	-	-	-	-
	B	-	-	-	-	30	-	-	-	-	-
	C	-	10	10	10	-	-	-	10	-	-
	D	-	-	-	30	20	-	-	40	60	-

is true for all presented models except the DLSP-P where the number of products in inventory for all periods remains the same, but spoilage decreases. This is reasonable, because less capacity decreases minimum lot sizes implied by the all-or-nothing constraint.

5 Conclusion

Aspects of deterioration and perishability are important for tactical and operational production planning, because such effects might generate increased costs if not considered in production plans. We have integrated such effects in well known discrete lot-sizing models and compared resulting plans using a numerical example. The DLSP turns out to be not suited to problems where products are subject to deterioration and perishability due to the all-or-nothing-assumption which implies minimal lot sizes, because it enforces full utilization of capacities disregarding periodic demand rates and, thus, causing spoilage and related costs. Further research will be in the direction of sequence-dependent setup times and multiple levels including diverse product structures where items depend on each other further considering rework, thus the possibility to rework items that need perish during the production process.

References

1. Trigeiro, W., Thomas, L., McClain, J.O.: Capacitated lot-sizing with setup times. Management Science 35(3), 353–366 (1989)
2. Andijani, A., Al-Dajani, M.: Analysis of deteriorating inventory/production systems using a linear quadratic regulator. European Journal of Operational Research 106(1), 82–89 (1998)
3. Dave, U.: A probabilistic scheduling period inventory model for deteriorating items with lead times. Zeitschrift für Operations Research 30(5), A 229–A 237 (1986)

4. Ghare, P.M., Schrader, G.F.: A model for exponentially decaying inventories. Journal of Industrial Engineering 14, 238–243 (1963)
5. Shah, B., Shah, N., Shah, Y.: EOQ model for time-dependent deterioration rate with a temporary price discount. Asia Pacific Journal of Operational Research 22(4), 479–485 (2005)
6. Zhao, P.X.: An EOQ model for items with Weibull distribution deterioration. In: 2nd IEEE Conference on Industrial Electronics and Applications, pp. 451–454 (2007)
7. Ketzenberg, M., Ferguson, M.: Sharing information to manage perishables (ed. 2). Working Paper (2005), `http://smartech.gatech.edu/handle/1853/7098` (last call Janaury 2010)
8. Pahl, J., Voß, S., Woodruff, D.: Production planning with deterioration constraints: A survey. In: Ceroni, J. (ed.) The Development of Collaborative Production and Service Systems in Emergent Economies, Proceedings of the 19th International Conference on Production Research. IFPR, Valparaiso (2007)
9. Goyal, S.K., Giri, B.C.: Recent trends in modeling deteriorating inventory. European Journal of Operations Research 134(1), 1–16 (2001)
10. Ferguson, M.E., Koenigsberg, O.: Production and pricing decisions: How should a firm manage a product with deteriorating quality (ed. 2). Preliminary Draft, Internet-Source (2005), `http://smartech.gatech.edu/handle/1853/7078`
11. Tsiros, M., Heilman, C.M.: The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. Journal of Marketing 69, 114–129 (2005)
12. Wee, H.M.: Economic production lot size model for deteriorating items with partial backordering. Computers & Industrial Engineering 24(3), 449–458 (1993)
13. Staggemeier, A.T., Clark, A.R.: A survey on lot-sizing and scheduling models. In: 23rd Annual Symposium of the Brazilian Operational Research Society (SOBRAPO) (2001)
14. Brahimi, N., Dauzere-Peres, S., Najid, N.M., Nordli, A.: Single item lot sizing problems. European Journal of Operational Research 168(1), 1–16 (2006)
15. Comelli, M., Gourgand, M., Lemoine, D.: A review of tactical planning models. Journal of Systems Science and Systems Engineering 17(2), 204–229 (2008)
16. Karimi, B., Ghomi, F.S., Wilson, J.: The capacitated lot sizing problem: a review of models and algorithms. Omega 31, 365–378 (2003)
17. Quadt, D., Kuhn, H.: Capacitated lot-sizing with extensions: A review. 4OR: A Quarterly Journal of Operations Research 6(1), 61–83 (2008)
18. Chu, L.Y., Hsu, V.N., Max Shen, Z.-J.: An economic lot-sizing problem with perishable inventory and economics of scale costs: Approximation solution and worst case analysis. Naval Research Logistics 52(6), 536–548 (2005)
19. Hsu, V.N.: Dynamic economic lot size model with perishable inventory. Management Science 46(8), 1159–1169 (2000)
20. Hsu, V.N.: An economic lot size model for perishable products with age-dependent inventory and backorder costs. IIE Transactions 35(8), 775–780 (2003)
21. Häselbarth, L., Scholl, A.: Dynamische Bestellmengenplanung für verderbliche Luxusgüter. Technical Report 13, Wirtschaftswissenschaftliche Fakultät, Friedrich-Schiller-Universität Jena (2003)
22. Pierskalla, W., Roach, C.: Optimal issuing policies for perishable inventory. Management Science 18(11), 603–614 (1972)

23. Gawiejnowicz, S.: Scheduling deteriorating jobs subject to job or machine availability constraints. European Journal of Operational Research 180(1), 472–478 (2007)
24. Dave, U.: On a discrete-in-time deterministic inventory model for deteriorating items with time proportional demand. Optimization 16(3), 449–461 (1985)
25. Suerie, C., Stadtler, H.: The capacitated lot-sizing problem with linked lot sizes. Management Science 49(8), 1039–1054 (2003)
26. Almada-Lobo, B., Klabjan, D., Carravilla, M.A., Oliviera, J.F.: Multiple machine continuous setup lot-sizing with sequence-dependent setups. Computational Optimization and Applications (2009)
27. Karmarkar, U.S., Kekre, S., Kekre, S.: Capacity analysis of a manufacturing cell. Journal of Manufacturing Systems 6(3), 165–175 (1987)

Discrete Lot-Sizing and Scheduling with Sequence-Dependent Setup Times and Costs including Deterioration and Perishability Constraints

Julia Pahl, Stefan Voß, David L. Woodruff

Abstract— **Deterioration and perishability constraints force organizations to carefully plan their production in cooperation with their supply chain partners up- and downstream. This is important because waiting times due to suboptimal planning give rise to increasing lead times and, consequently, to depreciation of parts and products while waiting and, thus, decreasing quality of items, so that, in the worst case, they cannot be used. Increased costs are only one problem. A more troubling aspect is unsatisfied customers waiting for their products or being concerned about quality. It is well known that reducing lot sizes leads to lower inventory holding costs, but also to increased setup costs. Therefore, lot size planning seeks to weigh the trade-off of setup costs and costs of inventory holding. The limits of lifetimes of parts and products in the production process increase the complexity of planning, especially if setup times and costs are dependent on the sequence of items. We analyze lot sizing models with sequence-dependent setup times and costs extending them to depreciation effects.**

I. INTRODUCTION

The consideration of sequence-dependent (sd) setup times and costs is important in many production settings where significant setup times influence capacities and thus lead times. Sd-setup times and costs are not uncommon. For instance, it is more time consuming and expensive to clean a coating facility from dark colors to light colors, so that it is more convenient to begin with light colors passing continuously to darker ones. The same is valid for cake mix manufacturing passing from chocolate to any other taste mix. Models that ignore such times and related costs may lead to non-optimal plans with longer lead times and higher idle times due to more frequent setups. If parts and products are subject to depreciation effects, so that the time they can spend while waiting in the production process or being stored as final products is limited, the objectives of shorter lead times and reduced waiting times gain importance.

In the early days, research on depreciation effects and their inclusion into models for inventory management focused on blood bank management and the distribution of blood from transfusion centers to hospitals. Later, other products, e.g., chemicals, food, drugs, fashion clothes, technical components, and newspapers were considered with an attention that shifted to a more general view on how to include deterioration and perishability constraints into mathematical models for aggregate production planning with different settings and assumptions, e.g., including multiple facilities and warehouses with diverse characteristics as well as supply chain perspectives; see [29], [32].

We note that including perishability in a planning model might imply just a few modifications to models created under some strong assumptions. Consider, e.g., the simple Wagner Whitin problem as a special case of the capacitated lot sizing problem (CLSP). Once perishability constraints are considered, the classical rolling horizon approach for solving the problem actually does not change much: edges that go beyond the time point for perishing are not considered. In that sense, problems tend to become a little "easier." Also for a mathematical programming formulation of the CLSP, adding perishability just reduces the number of variables.

In this paper we concentrate on discrete dynamic lot sizing and scheduling integrating deterioration and perishability constraints together with sd-setup costs and times. We consider the capacitated lot sizing problem with linked lot sizes and sd-setup costs and times (CLSD) and the general lot sizing and scheduling problem (GLSP) partly derived or, better, extended from the CLSP in order to further extend them to deterioration and perishability constraints. Accordingly, we study the effects of such constraints on the behavior/mechanisms and solutions of regarded models. To the best of our knowledge, no work integrating and studying these two important aspects has been done so far.

The remainder of the paper is organized as follows: In Section II we review the relevant literature on deterioration/perishability (Section II-A) and lot sizing with sd-setup times and costs (Section II-B). Based on these models, Section III represents their extensions. A numerical study is given in Section IV and Section V gives conclusions and further research ideas.

II. LITERATURE REVIEW

Depreciation effects come into play especially in situations where items are forced to wait. This might be due to technical matters of the production process, variabilities in, e.g., demand, disturbances of the production process or waiting times induced by the plan. The task of lot sizing is to plan quantities of items produced together so as to minimize certain cost factors such as setup costs and inventory holding costs. For instance, the importance of accurately accounting for setup times and costs and their effects on capacity utilization has been highlighted by the discussion of just-in-time manufacturing; see [23], [45]. Small and frequent batch sizes are preferred in order to decrease work in process (WIP) inventories, but frequent setups negatively influence capacity

Julia Pahl and Stefan Voß are with the Institute of Information Systems, University of Hamburg, D-20146 Hamburg, Germany pahl@econ.uni-hamburg.de, stefan.voss@uni-hamburg.de

David L. Woodruff is with the Graduate School of Management, University of California, Davis, CA 95616-8609, USA dlwoodruff@ucdavis.edu

utilization, because they represent idle times of machines. This is a trade-off that has to be considered and is highly influenced by related cost factors that express priorities of planning objectives. For instance, if lead time reduction is a prime goal, setups and thus idle times are minimized leading to increased WIP inventories. On the other hand, if high inventory holding costs are prevalent, the main objective is to reduce WIP (waiting times) and final product inventory holding at the expense of capacity utilization degraded by frequent setups. Consequently, scheduling information of jobs especially when subject to depreciation is important already on a tactical (lot sizing) basis, so that orders/items are planned in a way that minimizes their overall waiting times.

A. Depreciation of Products

Deterioration and perishability of products imply the loss of value or function, so that items cannot be used for their original purpose anymore, which can be evoked by damage, spoilage, or decay of items in a continuous process or a specific point in time; see [32], [29]. Fixed, maximum lifetimes that we define as *perishability* may be due to external factors, e.g., regulations (pharmaceuticals), technological developments, or change in style. As such they might be deterministic or stochastic. In contrast to that, *deterioration* is defined as a gradual or variable process of utility loss. For instance, lifetimes of food items are generally seasonal and vegetables mature/deteriorate faster in summer times with elevated temperatures than in winter times. Thus, deterioration is time-dependent and integrated in mathematical planning models by different probability distributions such as Gamma distributions, Weibull, two-parameter Weibull, or exponential distributions. An increasing number of papers assume Weibull distributions for the deterioration rate which is also supported empirically by fitting real/practical data to mathematical distributions. This is true for items such as frozen food, roasted coffee, breakfast cerials, cottage cheese, ice cream, pasteurized milk, or corn seeds; see [1].

The examples given so far illustrate cases where the functional capabilities of the items deteriorate. Nevertheless, items might lose their value in the eyes of the customer while their functional quality remains the same. This is especially true for fashion clothes, high technology products or newspapers whose information content deteriorates and perishes at the end of the day. For instance, see [46] for an empirical study of the willingness of customers to pay for a product subject to continuously decreasing desirability and/or utility.

Figure 1 depicts three functional relationships of the value of items over time where the first graph shows the course of perishability, the second the discrete and variational course of deterioration and the third three possible courses of continuous/gradual deterioration. This is where we draw a clear distinction between perishable goods that cannot be used after a certain point in time and, thus, lose their utility at once and deteriorating items that lose their utility/value gradually; see also [29]. It should be mentioned that many

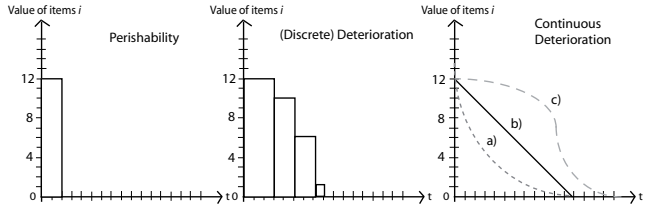

Fig. 1. Three examples for perishability and deterioration rates, see [29].

authors in the field of deterioration and perishability use the words interchangeably.

A great body of literature exists that incorporates deterioration and perishability into models considering different perspectives of production planning, e.g., order/replenishment and inventory management, lot sizing, aggregate production planning, rework, remanufacturing, or recycling where the latter refer to closed-loop supply chain management, which is part of the currently upcoming and promising research area named green supply chain management. For an overview of the literature in this field we refer to [32]. The major part of work done to integrate depreciation effects into mathematical models is in the continuous case in the field of inventory management. Models are extended by additional assumptions, e.g., shortages together with partial or complete backordering, lost sales independent or dependent on the waiting time of the customer. Very little work is provided regarding discrete dynamic modeling; see [29] for a short review and models including deterioration and perishability constraints. For an extensive review see [34].

B. Integration of Sequence-Dependent Setup Times and Costs into Lot Sizing Models

Setup times are often implicitely assumed to be zero in the literature. This might be due to complexity reduction of models (see [12]), so that they are applicable to real world problems which might be a good compromise because of insignificant setup times and costs and/or due to the application of static models, such as the economic order quantity formula. Relaxing the assumption of zero setup times and costs in (aggregate) lot sizing models implies scheduling decisions where a great body of work exists even for the static case, e.g., the economic lot scheduling problem; see, e.g., [9], [13], [15], [26], [42], [48], [53].

Assuming a (discrete) dynamic setting for integrated lot sizing and scheduling, we deal with *big bucket* and *small bucket* models where big bucket models contain rather long time (or *macro*) periods in which several different products can be set up and produced unlike small bucket models where time periods are rather short (*micro* periods), so that startups, switch-offs and/or change-overs can be modeled. Thus, small bucket models are particularly used to integrate more detailed information about the shop floor in order to determine accurate and feasible plans. Nevertheless, we are not on an operational level with "real" scheduling calculating due dates, machine utilization etc., but on a meta level between aggregate and operational planning. This is

important when translating plans determined on an aggregate level into detailed operational processes and schedules. For instance, planned lead times are considered on a tactical planning level. Therefore, buffer times used not only to hedge against uncertain events are included to surely meet due dates. However, this may increase inventory stocks or WIP further increasing realized lead times. This vicious cycle is named lead time syndrome in the literature; see [33], [43], [51], [50]. Problems may further occur if items/products are subject to deterioration or perishability. Consequently, including scheduling decisions already on a tactical level can prevent time-consuming iterations to determine viable plans; see also [10].

The integration of big and small bucket periods in order to provide planning together with some scheduling information is presented as *hybrid models* in the literature. They allow the production of multiple products per time period while preserving setup states across periods by carrying over setup states to one or more periods and, consequently, providing sequencing information of (ordered) items. Such models are the subject of extensive research; for more details see [6], [12], [18], [19], [22], [23], [25], [35], [38], [40], [44]. For excellent reviews on scheduling with lot-sizing and/or sequence-(in)dependent setup times and costs see [2], [7], [14], [36], or [47]. A more recent survey is presented by [52].

The consideration of lot sizing together with scheduling results from interdependencies of planning aspects that may be cyclic in nature. For instance, [30], [31] discuss lead times dependent on WIP inventories and capacity utilization. In this regard, [12] emphasize the cyclic nature of lot sizing, scheduling, and sd-setup costs: in order to determine optimal lot sizes, sd-setup times and costs are needed which, on the other hand, depend on the scheduling solution. However, the determination of the schedule needs information concerning types and amounts of items planned in all periods. A method to estimate initial setup costs that can be used in joint lot sizing and scheduling is proposed by [12] which shows the effects of such estimated costs on the solution.

The CLSD with zero setup times is studied by [22] allowing for continuous lot sizes and preservation of setup states over idle times. This paper provides a local search heuristic including priority rules. A branch-and-bound method with rescheduling possibilities for the CLSD with positive setup times is proposed by [23] and tested using a practical example. A heuristic method for the CLSP with positive setup times and costs on a rolling horizon basis is proposed by [20] which performs best in case that the number of product types is superior to considered planning periods. A relax-and-fix heuristic (see also [47]) for the GLSP on a rolling horizon basis is provided by [10] further allowing for backlogs in their model.

The GLSP is studied by, e.g., [14], [8], [16], and [27]. An asymmetric travelling salesman problem formulation together with an iterative solution procedure based on subtour elimination and patching is proposed by [8]. A dual reoptimization algorithm combined with local search heuristics, i.e., threshold accepting and simulated annealing are presented by [27].

There are some lot sizing and scheduling models that include multiple machines; see, e.g., [3], [4], [28], [49] for scheduling with sd-setup times including parallel machines. In fact, such an extension is interesting in terms of capacity increases or reduction of lead times. Therefore, capacity could be extended by buying another machine or outside capacity. This might not be desired, because additional machines may be expensive and/or know-how on the production process should remain company-intern; see [23] for a practical example. We consider the CLSD and the GLSP without capacity extensions. Nevertheless, it can be interesting for future work especially when regarding deterioration and perishability with rework options.

III. MODEL FORMULATIONS AND SHORT DISCUSSION

For our model extension we consider the CLSD and the GLSP that assume multiple items manufactured on a single machine subject to sd-setup times and costs. Inventory costs occur for items that are produced in advance of their demands that need to be satisfied. Neither shortages nor backlogging is allowed. We assume first-in-first-out (FIFO) for inventory holding. Furthermore, items are subject to deterioration or perishability, thus the periods they can be left in inventory are limited. The objective is to minimize the sum of sd-setup costs, inventory holding costs and costs for disposal of items regarding capacity constraints. Disposal costs of items are assumed to be relatively higher than inventory holding costs, e.g., due to environmental reasons. For instance, if spoilage is not penalized, the model produces more items than demand and throws away items not requested. Thus, spoilage costs are important to be considered in the objective function symbolizing opportunity costs of production that are, a priori, not considered in the classical GLSP. Nevertheless, in the case of spoilage costs that are linear with respect to time, they need not be included in the objective function, because they can simply be added to inventory holding costs.

Arguing that depreciation constraints are not necessary in lot sizing models, because such models plan inventory as late as possible, is not true, because constraints on lifetimes of items are hard constraints further limiting the solution space. Models without such constraints plan production and inventory as late as possible regarding demand and capacity resources, but allow for long(er) inventory holding than models including limited lifetimes of items. We come back to this in Section (IV) in more detail. Spoilage happens only if minimum lot sizes on a macro period level are imposed. These might be necessary due to technical matters, so that machines require to produce an amount of items in a batch that is superior to demand. For instance, this is accounted for in the discrete lot sizing and scheduling problem (DLSP) with perishability constraints; see [29] where the *all-or-nothing* constraint regarding capacity utilization holds. That is, if an item is produced in one period, it is required to utilize full available machine capacity regardless of item

demand in that period in order to avoid idle times; see also [22]. A similar model to the DLSP with sd-setup costs is given in [11]. The problem regarding the all-or-nothing constraint is remedied by the proportional lot sizing and scheduling problem (PLSP) with perishability constraints allowing for more than one setup in a period, so that idle times can be used for the production of another item; see [21] for the PLSP without perishability constraints. We use the following notations for our models: **Parameters:**

K_{ij}^{sd}	sd-setup cost factor setting up from product i to product j.
π_i	inventory holding cost factor of item i.
D_{it}	demand of product i in period t.
Θ_i	lifetime of product i.
V	a big number.
Cap_t	available capacity in period t.
ξ_i	resource consumption factor of product i.
ϑ_{ij}	sd-setup time setting up from product i to product j.
$X_s^{\min}$	minimum lot size for a micro period s.

Variables:

$\delta_{ijt}^{\text{sd}} \geq 0$	sd-setup variable setting up from product i to product j in period t.
$I_{it} \geq 0$	inventory holding of product i in period t.
$X_{it} \geq 0$	production amount of product i in period t.
$I_{it}^D \geq 0$	amount of spoiled items of product i in period t.
$\delta_{it} \in \{0,1\}$	setup state variable for item i in period t.
$z_{it} \geq 0$	*position* of product i in period t.

Indices:

$i,j,k \in M$	indices denoting products.
$t = 1,\ldots,T$	macro periods.
$s \in S_t$	micro periods.
s_t^f	first micro period.
s_t^l	last micro period.

A. CLSD with Depreciation

The CLSD is an extension of the CLSP where the latter includes the following assumptions:

- production takes place on a single machine restricted in its capacity,
- multiple items are produced,
- the planning cycle contains T discrete time periods,
- orders to produce X_{it} items of i in t are given in the beginning of a period t,
- products are delivered at the beginning of period t in lots X_{it},
- initial and ending inventory of the planning horizon are assumed to be zero.

Consequently, the CLSD inherits all assumptions of the CLSP further requiring that the *triangle inequalities* must be fulfiled. These conditions originate from graph/network theory and state that it is not more expensive and time consuming to setup directly from product i to product j than setting up from product i to product j via product k independently of the time period t. This is stated as follows:

$$\delta_{ijt}^{sd} \leq \delta_{ikt}^{sd} + \delta_{kjt}^{sd} \qquad \forall\, j,i,k,t \in M \tag{1}$$

$$K_{ij}^{sd} \leq K_{ik}^{sd} + K_{kj}^{sd} \qquad \forall\, j,i,k \in M \tag{2}$$

In case that the triangle inequalities are not required, *cleaning effects* of products might occur, thus setting up from product i to product j via product k is less expensive due to cleaning characteristics of product k.

We state the capacitated lot sizing problem with linked lot sizes and sd-setup costs and times with perishability (CLSD-P) implementing a perishability constraint with a lifetime limit denoted by Θ_i. In order to incorporate deterioration, this factor is changed to a time-dependent factor Θ_{it}. Thus the course of deterioration is defined via the input data. Furthermore, regarding perishability, we define that $\Theta_i = 0$ means that items cannot be stored in inventory whereas $\Theta_i = 1$ implies the possible storage of one period after production in the previous period. Thus, the period of production is, anyhow, "gratis" in terms of spoilage effects and costs. This is valid for the CLSD as well as the GLSP.

The complete CLSD-P is stated as follows:

$$\min \sum_{i,j=1}^{M}\sum_{t=1}^{T} K_{ij}^{\text{sd}}\delta_{ijt}^{\text{sd}} + \sum_{i=1}^{M}\sum_{t=1}^{T}\pi_i I_{it} + \sum_{i=1}^{M}\sum_{t=1}^{T}\phi_i I_{it}^D \tag{3}$$

subject to:

$$I_{it} = I_{i,t-1} + X_{it} - D_{it} - I_{it}^D \qquad \forall\, i,t \tag{4}$$

$$I_{it}^D \geq \sum_{\tau=1}^{t-\theta_i} X_{i\tau} - \sum_{\tau=1}^{t} D_{i\tau} - \sum_{\tau=1}^{t-1} I_{i\tau}^D \qquad \forall\, i,t \geq \Theta_i \tag{5}$$

$$X_{it} \leq V \cdot \left(\delta_{ijt}^{\text{sd}} + \delta_{i,t-1}\right) \qquad \forall\, i,j,t \tag{6}$$

$$\sum_{i=1}^{M}\sum_{t=1}^{T}\xi_i X_{it} + \sum_{j=1}^{M}\vartheta_{ji}\delta_{jit}^{\text{sd}} \leq \text{Cap}_t \qquad \forall\, i,t \tag{7}$$

$$\sum_{i=1}^{M}\delta_{it} = 1 \qquad \forall\, t \tag{8}$$

$$\sum_{j\in M}^{M}\left(\delta_{jkt}^{\text{sd}} + \delta_{k,t-1}\right) = \sum_{i\in M}^{M}\left(\delta_{kit}^{\text{sd}} + \delta_{kt}\right) \qquad \forall\, k,t \tag{9}$$

$$z_{jt} + 1 - M \cdot \left(1 - \delta_{ijt}^{\text{sd}}\right) \leq z_{it} \qquad \forall\, i,j,t \tag{10}$$

$$I_{i0} = I_{iT} = 0 \qquad \forall\, i \tag{11}$$

The objective function of the CLSD minimizes overall sd-setup costs, the inventory holding costs and costs for spoiled items in the planning horizon. The inventory balance equation is given in (4) stating that actual inventory is composed by the inventory of the previous period $t-1$ plus produced items in period t subtracted by demand D_{it} in that period and spoilage of items denoted by I_{it}^D. The amount of spoiled items in a period t is given in Inequalities (5). These are calculated by the sum of produced items in the interval $[0, t-\Theta_i]$, thus production from the beginning of the planning horizon until period t deduced by the perishability factor or liefetime of items Θ_i, less those items that are used to

satisfy demand until period t and items that have already been disposed in the previous periods $t-1$.

Inequalities (6) constrain the production quantity X_{it} by allowing production of item i in period t only if a setup operation from any product j to product i is executed in period t, so that $\sum_{j\in M}^{M} \delta_{jit}^{\text{sd}} = 1,\ \forall\, i,t$, or if the setup state for product i is carried over from period $t-1$ to period t, thus $\delta_{i,t-1} = 1$. The big number V needs to be chosen in a way that the production amount is limited, too. This is frequently reformulated using capacity utilization, thus $X_{it} \leq \text{Cap}_t/\xi_i\delta_{it}(\delta_{jit}^{\text{sd}}+\delta_{i,t-1}), \forall\, i,j,t$. Inequalities (7) state that available capacity Cap_t must not be exceeded by the sum of production of items i in period t and the sum of sd-setup times. Furthermore, Equations (8) require that the machine at the end of period t and, consequently, at the beginning of the next period $t+1$ is set up for a product i. The setup flow condition is presented in Equation (9). For instance, if a setup operation is performed from product j to product k, thus $\sum_{j\in M} \delta_{jkt}^{sd} = 1,\ \forall\, k,t$, or the setup state for product k is carried over in period t, i.e., $\delta_{k,t-1} = 1$, then the setup state for product k is carried to period $t+1$, i.e., $\delta_{kt} = 1$ or a setup from product k to a product j is done in period t, i.e., $\sum_{i\in M} \delta_{kit}^{sd} = 1\ \forall\, k,t$. In other words, if the machine is set up for product k in period t, then product k is the last product scheduled in period t ($\delta_{kt} = 1$) or an additional setup operation from product k to product j is performed; see, e.g., [22].

In order to avoid subtours, a variable z_{it} is defined in (10) which represents a kind of *position*, e.g., the higher the values for z_{it}, the later product j is scheduled. For example, if we assume the partial setup sequence $(\ldots, j, i, k, \ldots)$ of period t, thus $\delta_{jit}^{sd} = \delta_{ikt}^{sd} = 1$ and it follows from Inequalities (10):

$$\begin{aligned} z_{it} &\geq z_{jt} + 1 \\ z_{kt} &\geq z_{it} + 1 \\ z_{kt} &\geq z_{jt} + 2 \end{aligned}$$

Consequently, no subtour is possible due to $j \neq i, i \neq k, j \neq k$. Equation (11) of the CLSD require initial and ending inventory to be zero.

B. GLSP with Depreciation

The GLSP encompasses all features of the capacitated lot sizing problem with linked lot sizes (CLSPL) and the CLSD, so that they become special cases of the GLSP. For that reason, the GLSP is denoted *general*, because other presented models differ only regarding additional constraints for the time structure; see [16]: the GLSP divides a macro period t into sets of non-overlapping micro periods $s \in S_t$ of variable length. Therefore, micro period lengths might also be decision variables that are determined by the production quantities produced within. Frequently, the number of micro time periods within a macro period is defined in advance, so that it can be used in mixed integer programming (MIP) models. Consequently, a lot contains a production amount of the same item to which a sequence of micro periods is assigned that may continue over macro periods, too. There are two variations of the GLSP available in the literature, i.e., a version where setup states are preserved (GLSP-CS) and a version where they are lost (GLSP-LS); see, e.g., [16], [41]. Here, we consider setup carry overs and thus the first mentioned version of the GLSP.

The two-level time structure was first introduced by [37] for modeling dynamic production systems. External (environmental) dynamics influencing the production system are determined by the discrete macro time grid, defined by, e.g., demand data or holding cost data. Internal system dynamics are presented by system state variables dependent on production decisions that can occur anytime and independently of external dynamics, but at the beginning or end of the micro periods. Thus, the inventory balance equation (4) is reformulated as follows:

$$I_{it} = I_{i,t-1} + \sum_{s\in S_t} X_{is} - D_{it} - I_{it}^D \qquad \forall\, i,t \tag{12}$$

The all-or-nothing assumption constraining X_{is} is valid for micro periods as in the DLSP. In case that micro periods are of variable length, they do not impose minimal lot sizes that could cause spoilage at this stage. Nevertheless, we fix micro periods here in order to use them in MIPs and implement them in standard optimization software.

Production quantities are constrained as follows, so that idle times are possible:

$$X_{is} \leq \frac{\text{Cap}_t}{\xi_i}\delta_{is} \qquad \forall\, i,s,t \tag{13}$$

with the restriction that production of product i in period s only takes place if the machine is set up, respectively, thus:

$$\sum_{i=1}^{M} \delta_{is} = 1 \qquad \forall\, s \tag{14}$$

Besides, the link between setup states δ_{is} and the changeover indicators δ_{ijs}^{sd} are determined by the following equation:

$$\delta_{ijs}^{\text{sd}} \geq \delta_{i,s-1} + \delta_{js} - 1 \qquad \forall\, i,j,s \tag{15}$$

The changeover variable δ_{ijs}^{sd} is not defined as binary, but takes values of $0, 1$, respectively, in an optimal solution. Nevertheless, [16] argues that such formulation gives rise to redundancies if, e.g., a sequence of micro periods is assigned to the same product i and, consequently, its production amounts are arbitrarily distributed among them although not changing schedules nor the objective function value. Changes in inequality (13) and (15) remedy that effect forcing idle periods to be placed at the end of macro periods; for more details see [16].

The lifetime of items is assumed to be related to the macro period level. Consequently, the perishability constraint (5) is only changed in regard of the production variables defined on a micro period level:

$$I_{it}^D \geq \sum_{\tau=1}^{t-\theta_i} \sum_{s\in S_\tau} X_{i\tau} - \sum_{\tau=1}^{t} D_{i\tau} - \sum_{\tau=1}^{t-1} I_{i\tau}^D, \ \ \forall\, i, t \geq \Theta_i \tag{16}$$

Capacity consumption of X_{is} determines the length of micro periods s, so that micro periods of zero length are allowed;

see also [16]. A mimimum lot size condition is defined in case that the triangle inequalities (1)-(2) do not hold for certain setup times and costs matrices. For instance, this might be true for products that can have a *cleaning character* such that a direct setup from product i to product j may be more cost and time intensive than the detour via product k which can be the case in chemical industries; see [16]. Therefore, a minimum lot size of product i is defined that has to be produced, so that direct setup changes from product i to product k instead of product i via product j to product k are avoided; see also [8]:

$$X_{is} \geq X_i^{min} \left(\delta_{is}^s - \delta_{i,s-1}^s\right) \qquad \forall\, i, s \tag{17}$$

Nevertheless, the definition of minimum lot sizes may lead to non-zero final inventories and thus related costs as well as spoilage due to perishability. Inequalities (17) require production on a micro period level to be at least as large as the minimum lot size. This formulation admits overlapping of micro and macro periods due to setup state carryovers.

The capacity constraint of the capacitated lot sizing problem with perishability constraints (CLSP-P) (7) is reformulated regarding micro periods:

$$\sum_{i=1}^{M} \sum_{s \in S_t} \xi_i X_{is} + \sum_{i,j=1}^{M} \vartheta_{ij}^{\text{sd}} \delta_{ijs}^{\text{sd}} \leq \text{Cap}_t \qquad \forall\, s \tag{18}$$

The complete general lot sizing and scheduling problem with perishability constraint (GLSP-P) is given as follows:

$$\min \sum_{i,j=1}^{M} \sum_{s=1}^{S} K_{ji}^{sd} \delta_{jis}^{sd} + \sum_{i=1}^{M} \sum_{t=1}^{T} \pi_{it} I_{it} + \sum_{i=1}^{M} \sum_{t=1}^{T} \phi_i I_{it}^D$$

subject to:

$$I_{it} = I_{i,t-1} + \sum_{s \in S_t} X_{is} - D_{it} - I_{it}^D \qquad \forall\, i, t$$

$$I_{it}^D \geq \sum_{\tau=1}^{t-\theta_i} X_{i\tau} - \sum_{\tau=1}^{t} D_{i\tau} - \sum_{\tau=1}^{t-1} I_{i\tau}^D \qquad \forall\, i, t \geq \Theta_i$$

$$X_{is} \leq \frac{\text{Cap}_t}{\xi_i} \delta_{is} \qquad \forall\, i, t, s$$

$$X_{is} \geq X_i^{\min} \left(\delta_{is} - \delta_{i,s-1}\right) \qquad \forall\, i, s$$

$$\text{Cap}_t \geq \sum_{i=1}^{M} \sum_{s \in S_t} \xi_i X_{is} + \sum_{i,j=1}^{M} \vartheta_{ij}^{\text{sd}} \delta_{ijs}^{\text{sd}} \qquad \forall\, t$$

$$\sum_{i=1}^{M} \delta_{is} = 1 \qquad \forall\, s$$

$$\delta_{ijs}^{sd} \geq \delta_{i,s-1} + \delta_{is} - 1 \qquad \forall\, i, j, s$$

$$I_{i0} = 0 \qquad \forall\, i$$

$$\delta_{i0} = 0 \qquad \forall\, i$$

where the last two equations denote initial inventory and setup states, respectively. As has been stated in [16], the GLSP can be reduced or traced back to the CLSP using the GLSP-LS version adding

$$\delta_{i0s_t^f}^{\text{sd}} \leq \delta_{iis_t^f}^{\text{sd}} \qquad \forall\, i, t \tag{19}$$

so that setup states are lost and fixing $|S_t| = M + 1$. In a similar way, the DLSP can be derived by reformulating (13) applying the all-or-nothing assumption regarding macro periods, thus $X_{is} = \text{Cap}_t / \xi_i \delta_{is}$, further setting $|S_t| = 1$ and thus allowing only one micro period per macro period; see [16]. The same is valid for the continuous setup and lot sizing problem with perishability constraints (CSLP-P). For the proportional lot sizing and scheduling problem with perishability constraints (PLSP-P) that allows at most the production of two products per macro period, micro periods are fixed to $|S_t| = 2$ and at least two changeovers are allowed by adding the following constraint:

$$\sum_{i,j \in M} \sum_{s \in S_t} \delta_{ijs}^{\text{sd}} \leq 1 \tag{20}$$

IV. NUMERICAL STUDY

In order to present and analyze mathematical lot sizing models including sd-setup costs and times especially exploring and *visualizing* their planning mechanisms, we use in a first step a small numerical example and implement the GLSP-P in XPress-IVE calculating it on a computer with an Intel Core(TM) 2 Duo CPU processor with 1.60 GHz and 1,6 GB RAM. This is done in order to test our extension in a *proof-of-concept*-sense and because of the combinatorial nature of the proposed models that are at least as difficult as the CLSP which is already proved to be NP-hard (see, e.g., [5], [17]; the same is valid for the CLSP with positive setup times; see [24]). In a second step, we calculate 135 test instances in order to study effects of perishability.

Larger problem instances increase computational calculation time exponentially, so that exact optimization methods become impractical for use in practice. Therefore, heurstic methods are proposed by many authors; see Section II-B for related references. Nevertheless, test instance sizes with less than 10 items, but up to 20 periods might already be of practical relevance as argued by different empirical studies; see [23].

A. A First Small Numerical Example

For the small example, we consider three types of products and a planning horizon of 10 macro periods and related three equal micro periods within each. The capacity of one machine is 100 units per macro period and minimum lot sizes are fixed for each product to 30 units per micro period, thus $X_{is}^{\min} = 30$. Overlapping of micro periods regarding minimal lot sizes is addmitted, thus if in micro period $s = 15$ an amount of 30 items of a product is produced, the production of 10 further products in $s = 16$ is allowed, so that in total a lot of 40 items is produced within a setup and minimal lot sizes are respected. Extending data sets leads to the foreseen effect of increased computation times as variables and constraints increase exponentially. This is

TABLE I

COST FACTORS FOR THE SMALL NUMERICAL EXAMPLE

	Cost Factors			
Products	π_i	ϕ_i^D	Θ_i	ξ_i
A	1	$2\pi_A$	$0-3$	1
B	1	$2\pi_B$	$0-3$	1
C	1	$2\pi_C$	$0-3$	1

TABLE II

SD-COSTS AND TIMES FOR REGARDED PRODUCTS

	Products j					
Products i	A		B		C	
	K_{Ai}^{sd}	$\vartheta_{Ai}^{\text{sd}}$	K_{Bi}^{sd}	$\vartheta_{Bi}^{\text{sd}}$	K_{Ci}^{sd}	$\vartheta_{Ci}^{\text{sd}}$
A	0	0	20	1	40	2
B	10	1	0	0	50	1
C	40	2	30	1	0	0

TABLE III

DEMAND DATA FOR THE SMALL NUMERICAL EXAMPLE

	Demand in Periods									
Products	1	2	3	4	5	6	7	8	9	10
A	12	13	9	7	10	11	10	12	12	10
B	21	24	25	19	17	20	22	23	23	18
C	29	31	30	33	34	31	27	28	25	21

TABLE IV

ANALYSIS REGARDING THE VARIATION OF Θ_i

Variation Lifetime	Optimality Gap in %	Best Solution	Best Bound
$\Theta_i = 0$	$38,98$	976	$595,6$
$\Theta_i = 1$	$49,15$	606	$308,1$
$\Theta_i = 2$	$22,15$	439	$341,8$
$\Theta_i = 3$	$39,76$	537	$323,5$

straightforward and experienced by extending the data set to 7 products and 50 periods.

Table I states the cost factors of the model and lifetimes of items. We defined disposal costs in relation to inventory holding costs, so that disposal costs are always superior to inventory holding, here twice as large. The lifetime of items is varied for each product from zero to two macro periods where zero lifetime means that items cannot be stocked. Additionally, we regard sd-setup costs and times as stated in Table II where the first value denotes sd-setup costs and the second value denotes sd-setup times. Demand data are stated in Table III. We ran the calculation for the GLSP until the maximum clock time of 1000 seconds without finishing the search. At least one optimal solution was found. The calculation results varying Θ_i from $1-3$ for all products is given in Table IV. The related resulting plans displaying inventory, production and spoilage are given in Table V, VI, and VII. In case that disposal costs are assumed to be zero, the model starts to produce and dispose a higher number of items than necessary. This is shown in Table VIII. Therefore, the consideration of penalty costs for spoilage has to be included into the objective function.

The mechanisms of the GLSP shows that the incorporation of deterioration and perishability constraints is a hard constraint rendering the finding of a solution very difficult especially in case of relevant sd-times and costs. We checked the model varying the data regarding different lifetimes of items. For instance, when regarding lifetimes of one macro period for each item thus prohibiting inventory holding, the model produces items that are required due to demand. Nevertheless, due to implied minimal lot sizes in order to restrict setups on a micro period basis, disposals occur.

B. Numerical Study Including Test Instance Sets

We generated 135 test instances for the GLSP-P combining the different parameter profiles and variations using the test data instance generator described in [39].

We generate the following parameters:

- *demand time series* D_{it}: the mean demand time series are generated according to a Gamma distribution. A standardized normal distribution is not used due to the fact that such a distribution may generate negative demand. We assumed for the demand time series a rather low variability, so that no problems regarding capacity generation and production lifetime are created, because the first priority in generating the data sets is to create solvable test instances.
- *machine capacity* Cap_t: the machine capacity is fixed to 1.000 units per time. Coming from this fixed parameter, diverse load profiles are given in order to calculate the production coefficients and sd-setup times for all products. Available capacity is calculated regarding the selected load/utilization profiles,
- *setup times* ϑ_{ij}: these are randomly generated numbers regarding the product combinations and maximal setup times checked with the available capacity, so that the maximal setup times do not exceed available capacity.
- *production coefficients* ξ_i: for each setup coefficient ξ_i a random number is generated. All coefficients are summed and checked with the capacity utilization less the setup times from the available capacity utilization, so that a multiplicator is generated that corrects the production rate in a way not exceeding the capacity limit.
- *time between orders (TBO)* : mean time between orders profiles are generated using the classical economic

TABLE V

PLAN FOR THE GLSP AND THE NUMERICAL EXAMPLE WITH $\Theta_i = 2$

	Macro periods									
Products	1	2	3	4	5	6	7	8	9	10
I_A	22	9	–	21	11	–	24	12	–	20
I_B	9	15	2	17	–	10	–	7	–	–
I_C	1	30	–	–	–	–	5	46	21	–
I_A^D	–	–	–	2	–	–	–	–	–	–
I_B^D	–	–	–	–	–	–	–	–	–	–
I_C^D	–	–	–	–	–	–	–	–	–	–

TABLE VI

PLAN FOR THE GLSP AND THE NUMERICAL EXAMPLE WITH $\Theta_i = 1$

	Macro periods									
Products	1	2	3	4	5	6	7	8	9	10
I_A	18	9	–	13	3	22	12	–	18	8
I_B	9	15	–	–	20	–	8	4	18	–
I_C	1	30	–	–	–	–	28	–	5	–
$I^D(A)$	–	–	–	10	–	–	–	–	–	–
$I^D(B)$	–	–	–	–	–	–	–	–	–	–
$I^D(C)$	–	–	–	–	–	–	–	–	–	–

TABLE VII

PLAN FOR THE GLSP ASSUMING $\Theta_i = 0$

	Macro periods									
Products	1	2	3	4	5	6	7	8	9	10
I_A	–	–	–	–	–	–	–	–	–	–
I_B	–	–	–	–	–	–	–	–	–	–
I_C	–	–	–	–	–	–	–	–	–	–
$I^D(A)$	18	17	–	23	–	19	–	18	18	20
$I^D(B)$	9	6	5	11	13	10	8	7	–	12
$I^D(C)$	1	–	–	–	–	–	3	–	5	–

TABLE VIII

PLAN FOR THE GLSP ASSUMING ZERO SPOILAGE COSTS

	Macro periods									
Products	1	2	3	4	5	6	7	8	9	10
I_A	13	–	7	–	21	10	–	12	–	–
I_B	–	–	–	4	20	–	–	–	–	–
I_C	–	–	–	1	–	–	–	–	21	–
$I^D(A)$	5	–	14	–	–	–	–	6	–	71
$I^D(B)$	9	6	–	10	–	–	8	–	7	–
$I^D(C)$	1	–	–	–	–	–	–	2	–	–

order quantity (EOQ) formula, thus $\sqrt{2K^s_{ijs} \cdot D/\pi_i}$. We assume the inventory holding cost factor π_i to remain constant thus varying demand according to the demand profiles and setup costs.

We generated 135 test instances varying the perishability rate, generating fife demand profiles, assuming different machine load profiles, and varying the TBO, thus $3 \cdot 5 \cdot 3 \cdot 3 = 135$ test instances considering

- lifetime $\Theta_i := 1, 2, 3$ macro periods,
- fife different demand profiles assuming a Gamma distribution,
- machine loads $:= 0,75;\ 0,8;\ 0,9$,
- TBO profiles $:= 3, 4, 5$ where it is important to choose these profiles longer than perishability. If the lifetime of items would be longer than the TBOs, the perishability constraint would not constraint the solution space.

Considering a lifetime factor of one macro periods, thus $\Theta_i = 1$, we expect the model to create a *lot-for-lot* plan, i. e. lot sizes that match orders/demand of products in that period. In that way, inventory is not build, the machine has to be set up every time and orders are not aggregated in lot sizes. A lifetime factor $\Theta_i = 2$ is expected to give different lot sizing possibilities to the model leading to longer calculation times or a larger gap to optimality due to the fixed calculation time limit. Thus, the model/solver initiates to evaluate the various aggregations of lot sizes building up inventory for the sake to diminishing setup costs. Spoilage due to perishability is expected not to occur regarding a lifetime near the TBO length.

We tested the calculation results of the 135 test instances with a Wilcoxon signed rank test (W-SRT) for matched pairs in order to compare the difficulty to find a solution regarding lifetime constraints, the TBOs, and the machine loading; see Tables IX, X, XI, XII, XIII, and XIV. The tables give the p-value of the W-SRT which gives the probability to observe the test sample or an even rarer test sample under consideration of the null hypothesis. In other words, the p-value gives the smallest test level where the test sample is just significant. If the p-value is smaller or equal the α-value denoting the error to reject the null hypothesis although it is true, the null hypothesis is rejected. In case that the p-value is greater than α, the null hypothesis is not rejected. Values that greater than an α-value of $0,05$ are highlighted. They signify that the null hypothesis is not rejected and thus the distribution is not shifted regarding an amount $\mu = 0$. The tables are symmetric due to the matching of the lifetime factors and their corresponding optimality gaps.

For instance, Table IX shows that $\Theta_i = 1$ and $\Theta_i = 3$ have a p-value of $0,6905$, thus the median of the distribution of the optimality gap is not significant. In contrast, the p-values of $\Theta_i = 2$ and $\Theta_i = 1$ are smaller the α-value of $0,05$ and thus the median of the distribution of gap-values ist not equal. This is true when having a look in the statistics of the optimality gaps of the 135 test instances that solving the model with a lifetime

TABLE IX

WILCOXON RST, TBO $= 3$, LOAD $= 0,75$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	$0,0079$	**0,6905**
$\Theta_i = 2$	$0,0079$	$0,0000$	$0,0079$
$\Theta_i = 3$	**0,6905**	$0,0079$	$0,0000$

TABLE X

WILCOXON RST, TBO $= 3$, LOAD $= 0,9$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	**0,8413**	**0,6905**
$\Theta_i = 2$	**0,8413**	$0,0000$	**1,0000**
$\Theta_i = 3$	**0,6905**	**1,0000**	$0,0000$

TABLE XI

WILCOXON RST, TBO $= 4$, LOAD $= 0,75$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	**0,2222**	$0,0317$
$\Theta_i = 2$	**0,2222**	$0,0000$	$0,0159$
$\Theta_i = 3$	$0,0317$	$0,0159$	$0,0000$

TABLE XII

WILCOXON RST, TBO $= 4$, LOAD $= 0,9$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	**0,6905**	**0,3095**
$\Theta_i = 2$	**0,6905**	$0,0000$	**0,5476**
$\Theta_i = 3$	**0,3095**	**0,5476**	$0,0000$

factor $\Theta = 2$ gives higher optimality gaps. Test statistics and test instances are available at http://iwi.econ.uni-hamburg.de/IWIWeb/Default.aspx?tabid=174.

Nevertheless, we have to note that the sample of 135 test instances is rather small. We expect that extending test instances, especially increasing TBO values to six or more profiles and also increasing values for Θ_i, the tendency proceeds that with decreased values of Θ_i, test instances become easier to solve due to the fact that with less flexibility to store products, the model tends to employ a lot-for-lot policy.

V. CONCLUSION

Deterioration and perishability constraints of items are important to be considered in aggregate planning especially regarding sd-setup times and costs, because avoidance of waiting times is mandatory if parts and products are subject to depreciation effects. We integrated deterioration and perishability effects in hybrid lot sizing and scheduling models with sd-setup times and costs such as the CLSD and the GLSP and highlighted their effects on the resulting plan.

The inclusion of lifetime limits of items is also interesting in a mathematical sense as it reduces setup variables of items for periods exceeding their maximal lifetimes. This might render classical lot sizing models as the CLSP easier to solve in a mathematical sense. Further research leads in the direction of extending models to multiple levels such as complex production structures as well as multiple machines. Moreover, we are interested in the reformulation of general lot sizing and scheduling models with deterioration and perishability constraints into network flow problems considering hop constraints in order to model depreciation. As we discussed earlier, depreciation may occur due to waiting in the production process where lead times are load-dependent. If items can be reworked, increased utilization further increases lead times and WIP, thus waiting times. This vicious cycle and its consideration in planning models are very interesting, so that we extend our research to this interesting topic in the future.

REFERENCES

[1] A. Al-Khedhairi and L. Tadj. Optimal control of a production inventory system with Weibull distributed deterioration. *Applied Mathematical Sciences*, 1(35):1703–1714, 2007.

[2] A. Allahverdi, J. N. D. Gupta, and T. Aldowaisan. A review of scheduling research involving setup considerations. *Omega*, 27(2):219–239, 1999.

[3] B. Almada-Lobo, D. Klabjan, M.A. Carravilla, and J.F. Oliviera. Multiple machine continuous setup lot-sizing with sequence-dependent setups. *Computational Optimization and Applications*, 2009. Online first.

[4] P. Beraldi, G. Ghiani, A. Grieco, and E. Guerriero. Rolling-horizon and fix-and-relax heuristics for the parallel machine lot-sizing and scheduling problem with sequence-dependent set-up costs. *Computers & Operations Research*, 35(11):3644–3656, 2008.

[5] G.R. Bitran, E.A. Haas, and A.C. Hax. Hierarchical production planning: A two stage system. *Operations Research*, 30:232–251, 1982.

[6] D. Briskorn. A note on capacitated lot sizing with setup carry over. *IIE Transactions*, 38:1045–1047, 2006.

[7] T.C.E. Cheng, Q. Ding, and B.M.T. Lin. A concise survey of scheduling with time-dependent processing times. *European Journal of Operational Research*, 152(1):1–13, 2004.

[8] A.R. Clark, R.M. Neto, and E.A.V. Toso. Multi-period production setup-sequencing and lot-sizing through ATSP subtour elimination and patching. In *Proceedings of the 25th workshop of the UK planning and scheduling special interest group (PlanSIG 2006)*, pages 80–87, University of Nottingham, 2006.

[9] S.G. Davis. An improved algorithm for solving the economic lot size problem (ELSP). *International Journal of Production Research*, 33(4):1007–1026, 1995.

[10] S. de Araujo, M. Arenales, and A. Clark. Joint rolling-horizon scheduling of materials processing and lot-sizing with sequence-dependent setups. *Journal of Heuristics*, 13(4):337–358, 2007.

[11] M.A.H. Dempster and J.K. Lenstra, editors. *Deterministic and Stochastic Scheduling*, Nato Advanced Study Institutes Series, Dordrecht, 1982. D. Reidel Publishing Company.

[12] D.M. Dilts and K.D. Ramsing. Joint lot sizing and scheduling of multiple items with sequence-dependent setup costs. *Decision Sciences*, 20(1):120–133, 1989.

[13] G. Dobson. The cyclic lot scheduling problem with sequence-dependent setups. *Operations Research*, 40(4):736–749, 1992.

[14] A. Drexl and A. Kimms. Lot sizing and scheduling - survey and extensions. *European Journal of Operational Research*, 99:221–235, 1997.

TABLE XIII

WILCOXON RST, TBO $= 5$, LOAD $= 0,75$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	**0,6905**	**0,3095**
$\Theta_i = 2$	**0,6905**	$0,0000$	**0,1508**
$\Theta_i = 3$	**0,3095**	**0,1508**	$0,0000$

TABLE XIV

WILCOXON RST, TBO $= 5$, LOAD $= 0,9$

p	$\Theta_i = 1$	$\Theta_i = 2$	$\Theta_i = 3$
$\Theta_i = 1$	$0,0000$	**0,2222**	**0,3095**
$\Theta_i = 2$	**0,2222**	$0,0000$	**0,0952**
$\Theta_i = 3$	**0,3095**	**0,0952**	$0,0000$

[15] S.E. Elmaghraby. The economic lot scheduling problem (ELSP): Review and extensions. *Management Science*, 24(6):587–598, 1978.

[16] B. Fleischmann and H. Meyr. The general lotsizing and scheduling problem. *OR Spectrum*, 19(1):11–21, 1997.

[17] M. Florian, J. K. Lenstra, and A. H. G. Rinnooy Kan. Deterministic production planning: Algorithms and complexity. *Management Science*, 26(7):669–679, 1980.

[18] M. Gopalakrishnan, K. Ding, J.-M. Bourjolly, and S. Mohan. A Tabu-Search heuristic for the capacitated lot-sizing problem with set-up carryover. *Management Science*, 47(6):851–863, 2001.

[19] M. Gopalakrishnan, D. M. Miller, and C. P. Schmidt. A framework for modelling setup carryover in the capacitated lot sizing problem. *International Journal of Production Research*, 33(7):1973, 1995.

[20] D. Gupta and T. Magnusson. The capacitated lot-sizing and scheduling problem with sequence-dependent setup costs and setup times. *Computers & Operations Research*, 32(4):727–747, 2005.

[21] K. Haase. *Lotsizing and Scheduling for Production Planning*, volume 408 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin et al., 1994.

[22] K. Haase. Capacitated lot-sizing with sequence dependent setup costs. *OR Spectrum*, 18(1):51–59, 1996.

[23] K. Haase and A. Kimms. Lot-sizing and scheduling with sequence-dependent setup costs and times and efficient rescheduling opportunities. *European Journal of Operational Research*, 66:159–169, 2000.

[24] K.K. Haugen, A. Olstad, and B.I. Pettersen. The profit maximizing capacitated lot-size (PCLSP) problem. *European Journal of Operational Research*, 176(1):165–176, 2007.

[25] B. Karimi, SMT Fatemi Ghomi, and J. M. Wilson. A tabu search heuristic for solving the CLSP with backlogging and set-up carry-over. *Journal of the Operational Research Society*, 57:140–147, 2006.

[26] M.A. Narro Lopez and B.G. Kingsman. The economic lot scheduling problem: theory and practice. *International Journal of Production Economics*, 23(1-3):147–164, 1991.

[27] H. Meyr. Simultaneous lotsizing and scheduling by combining local search with dual reoptimization. *European Journal of Operational Research*, 120(2):311–326, 2000.

[28] I.M. Ovacik and R. Uzsoy. Rolling horizon procedures for dynamic parallel machine scheduling with sequence-dependent setup times. *International Journal of Production Research*, 33(11):3173–3192, 1995.

[29] J. Pahl and S. Voß. Discrete lot-sizing and scheduling including deterioration and perishability constraints. In *Advanced Manufacturing and Sustainable Logistics*, Lecture Notes in Business Information Processing, 46, pages 345–357. Springer Berlin Heidelberg, 2010. 10.1007/978-3-642-12494-5_31.

[30] J. Pahl, S. Voß, and D.L. Woodruff. Load dependent lead times - from empirical evidence to mathematical modeling. In H. Kotzab, S. Seuring, M. Müller, and G. Reiner, editors, *Research Methodologies in Supply Chain Management*, pages 539–554. Physica, Heidelberg, 2005.

[31] J. Pahl, S. Voß, and D.L. Woodruff. Production planning with load dependent lead times. *4OR: A Quarterly Journal of Operations Research*, 3(4):257–302, 2005.

[32] J. Pahl, S. Voß, and D.L. Woodruff. Production planning with deterioration constraints: A survey. In J.A. Ceroni, editor, *The Development of Collaborative Production and Service Systems in Emergent Economies, Proceedings of the 19th International Conference on Production Research, IFPR*, Valparaiso, Chile, 2007.

[33] J. Pahl, S. Voß, and D.L. Woodruff. Production planning with load dependent lead times: An update of research. *Annals of Operations Research*, 153(1):297–345, 2007.

[34] J. Pahl, S. Voß, and D.L. Woodruff. *Production Planning with Deterioration and Perishability*. 2010. Draft Manuscript, to be published.

[35] P. Porkka, A. P. J. Vepsäläinen, and M. Kuula. Multiperiod production planning carrying over set-up time. *International Journal of Production Research*, 41(6):1133–1148, 2003.

[36] C.N. Potts and L.N. Van Wassenhove. Integrating scheduling with batching and Lot-Sizing: a review of algorithms and complexity. *Journal of the Operational Research Society*, 43(5):395–406, 1992.

[37] D. Pressmar. *Führungsprobleme industrieller Unternehmungen*. de Gruyter, 1980.

[38] D. Quadt and H. Kuhn. Capacitated lot-sizing and scheduling with parallel machines, back-orders, and setup carry-over. *Naval Research Logistics*, 56(4):366–384, 2009.

[39] C. Seipl. *Simulation kontinuierlicher Prozesse in hierarchischen Produktionsplanungssystemen*. PhD thesis, University of Hamburg, 2009. http://www.sub.uni-hamburg.de/opus/volltexte/2010/4504/pdf/Dissertation_Christian_Seipl.pdf.

[40] C.R. Sox and Y. Gao. The capacitated lot sizing problem with setup carry-over. *IIE Transactions*, 31(2):173–181, 1999.

[41] A.T. Staggemeier and A.R. Clark. A survey on lot-sizing and scheduling models. In *23rd Annual Symposium of the Brazilian Operational Research Society (SOBRAPO)*, 2001.

[42] O. Tang and R. Teunter. Economic lot scheduling problem with returns. *Production & Operations Management*, 15(4):488–497, 2006.

[43] I.P. Tatsiopoulos and B.G. Kingsman. Lead time management. *European Journal of Operational Research*, 14:351–358, 1983.

[44] H. Tempelmeier and L. Buschkühl. A heuristic for the dynamic multi-level capacitated lotsizing problem with linked lotsizes for general product structures. *OR Spectrum*, 31:385 – 404, 2009.

[45] W.W. Trigeiro, L.J. Thomas, and J.O. McClain. Capacitated lot-sizing with setup times. *Management Science*, 35(3):353–366, 1989.

[46] M. Tsiros and C.M. Heilman. The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *Journal of Marketing*, 69:114–129, 2005.

[47] L.A. Wolsey. MIP modelling of changeovers in production planning and scheduling problems. *European Journal of Operational Research*, 99(1):154–165, 1997.

[48] M.-J. Yao. The economic lot scheduling problem without capacity constraints. *Annals of Operations Research*, 133(1-4):193–205, 2005.

[49] K.-C Ying and H.-M. Cheng. Dynamic parallel machine scheduling with sequence-dependent setup-times using an iterated greedy heuristic. *Expert Systems with Applications*, 37(4):2848–2852, 2010.

[50] G. Zäpfel and H. Missbauer. New concepts for production planning and control. *European Journal of Operational Research*, 67:297–320, 1993.

[51] G. Zäpfel and H. Missbauer. Production planning and control (PPC) systems including load-oriented order release - problems and research perspectives. *International Journal of Production Economics*, 30:107–122, 1993.

[52] X. Zhu and W.E. Wilhelm. Scheduling and lot sizing with sequence-dependent setup: A literature review. *IIE Transactions*, 38:987–1007, 2006.

[53] P.H. Zipkin. Computing optimal lot sizes in the economic lot scheduling problem. *Operations Research*, 39(1):56–63, 1991.

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Christian Seipl regarding the generation and support of the used test instances for the GLSP with perishability.

Integrating Deterioration and Lifetime Constraints in Production and Supply Chain Planning: A Survey

Julia Pahl

Abstract

Items with short life cycles that are subject to deterioration are a major part of products traded in our business world. Research has a long tradition in integrating deterioration and value loss effects into mathematical models for inventory planning and control where such effects are understood as a general loss or shrinkage of inventory. Only a few work has been done in modeling lifetime restrictions of items to prevent wastage and disposal especially in a dynamic planning context. Recent trends extend the consideration of single companies to whole supply chains exploring their increased coordination and information needs. This is important as planning decisions impact lead times and thus the quality of items in the whole supply chain. Products that pass their useful lifetime can impose high costs due to inventory loss or the need to rework them which implies enhanced utilization of (scarce) resources, e.g., machine time, (noble) metals, and/or energy increasing CO_2-levels. We give an overview on the state-of-the-art regarding depreciation effects and lifetime constraints in planning and provide a classification of models along business planning functions of the value chain.

Keywords: Review, Deterioration, Perishability, Lifetime Constraints, Classification, Production Planning, (Green) Supply Chain Management
2000 MSC: 46N10, 90B30, 90C90

1. Introduction

Constraints on the lifetime of items force organizations to carefully plan their production in cooperation with their supply chain partners. This gains importance because waiting times imply increased *work in process* (WIP) inventories that might lead to longer lead times and thus increase the possibility of depreciation of items. In the worst case, they cannot be used for their original purpose and have to be disposed. This is notably disquieting, e.g., in the food industry where significant losses occur during handling, processing, and their distribution [1]. Results of a study conducted for the International Congress "Save Food!" state that 1.3 billion tons of food per year are wasted at a global scale which is around one-third of food produced worldwide for human consumption [1]. This is particularly valid for developing countries where more than 40% of food losses are due to processing whereas total losses in industrialized countries are as high, but with more than 40% mainly occurring at the retailer and consumer level [1]. Increased supply chain costs are only one problem. A more troubling aspect are enhanced losses of resources,

Email address: pahl@econ.uni-hamburg.de (Julia Pahl)

increased energy consumption and related emission of CO_2. Certainly, unsatisfied customers waiting for their products or being concerned about quality is mainly a problem in industrialized countries.

Deterioration has a great influence not only on inventory management, but on every area of the production process where items are stocked or forced to wait due to uncertain demand, technical matters, variabilities, or disruptions of the production process. We explore the integration of deterioration effects and lifetime constraints along different business functions (see Figure 1, [2]) that are closley related, but not limited to inventory management in order complement reviews available on deterioration in inventory management [3, 4] and extended settings [5]. Regarding food production the question is how to prevent losses due to spillage or degradation of items in the (industrial) handling process. Apart from engineering issues to enhance the production process, related coordinated planning throughout the supply chain is a starting point to this problem. Despite a long tradition in the operations research area to integrate deterioration effects in models for planning, there are only a few models that integrate lifetime contraints.

The paper is organized as follows: In Section 2 we provide some general thoughts on the classification of models including depreciation effects and lifetime constraints. We present mathematical formulations that are frequently used and discuss the role of demand in such models. Section 3 reviews the literature regarding important business (planning) functions and highlights issues where further research is needed. In that sense, Figure 1 serves as an orientation.

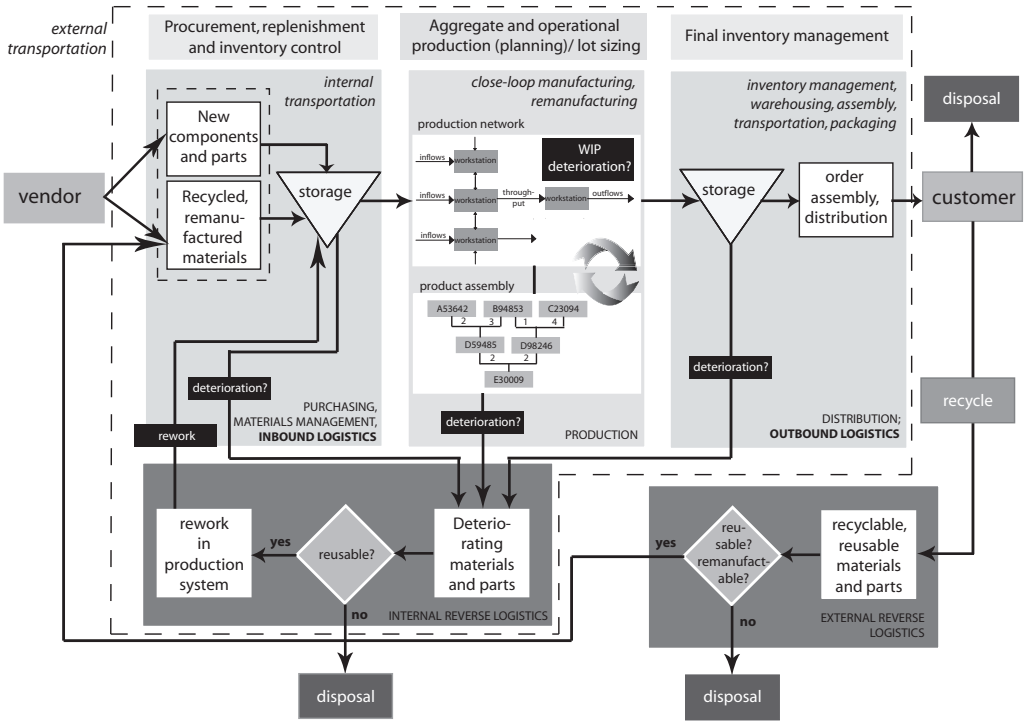

Figure 1: Material flow along the internal and external supply network; modified from [6]

Accordingly, Section 3.1 reviews procurement/replenishment models including marketing issues and inventory control whereas Section 3.2 discusses aggregate production-inventory planning.

Section 3.3 presents the current research concerning inventory of final goods and their delivery to customers which is complemented by Section 3.4 that evaluates the use of multiple warehouses. Until this point, it is assumed that deteriorated items cannot be reworked and need to be disposed which is relaxed in Section 3.5 examining rework options. Section 3.6 analyzes current research of deterioration effects and lifetime constraints in supply chains. Section 4 resumes major findings.

2. Classification

Interests of the operations research community regarding deterioration of items came to prominance with the modeling of inventory in blood bank management and the distribution of blood from transfusion centers to hospitals. Later, the research focus shifted also to other products, e.g., chemicals, food, fashion clothes, technical components, newspapers etc. In accordance, the attention changed to a more general view on how to include deterioration effects and lifetime constraints into models for inventory management. The major part of approaches consider deterioration as a kind of general shrinkage of inventory. We first discuss characteristics of deterioration, perishability, and the integration of lifetime restrictions providing definitions followed by mathematical formulations.

2.1. Discussion on Depreciation

Most products deteriorate within a certain interval. If the observed time horizon is chosen to be sufficiently short there may be no need to consider such product characteristics. Nevertheless, in practice and depending on the product, the time horizon for aggregate planning is long enough, so that the influence of deterioration starts to play an important role. Deterioration and lifetime restrictions render the planning process complex due to hard constraints on inventory holding. Therefore, restrictions are imposed to produce ahead of demand and sales realizations in order to hedge against uncertainties, e.g., employing over-production, exploit volume-discounts, or own production setup-sequence advantages. For instance, in some business fields, make-to-order approaches are not applicable, because replenishment lead times are significantly shorter than production manufacturing lead times [7], so that production plans rely on demand forecasts.

Deterioration of goods might be regarded as the process of decay, damage or spoilage of items in such a way that they can no longer be used for their original purpose, i.e., they go through a change while stored and continuously lose their utility [7–10]. Perishable goods are considered as items with fixed, maximum lifetimes [11]. This is true for products that become obsolete at some point in time, because of various reasons or external factors, e.g., laws and regulations that predetermine their shelflives [12]. Therefore, a classification basis is the *utility* of the item. In this regard, two subcategories can be distinguished: the first includes products whose functionalities deteriorate over time, e.g., fruits, vegetables or milk. The second encompasses those products whose functionality does not degrade, but customers' perceived utility and thus their demand deteriorates over time, e.g., fashion clothes, high technology products with a short life cycle, or items whose information content deteriorates, e.g., newspapers. For example, it is empirically shown that the willingness to pay for a product continuously decreases with its perceived actuality that is similar to defining deterioration as a process of continuously reduced usefulness of the comodity to the customer [13]. Other products encounter value loss related perishability induced by demand that drops when a new version or generation is introduced accompanied by significant downmarks [14]. This is valid, e.g., for computers, mobile phones, high fashion wear,

Christmas gifts, or calendars. Such value loss is indirectly integrated into mathematical models by making assumptions on products' demand patterns, e.g., time-dependent demand that is used to describe sales in different product lifetime phases in the market [15]. The examination of effects of stochastic demand rates on deterioration includes stochastic processes that may have independent increments, e.g., compound Poisson processes, or not, e.g., compound renewal processes, known or arbitrary probability distributions. As a result, we can characterize depreciation regarding different aspects: 1. physical degradation or depletion (e.g., fresh food, gasoline), 2. loss of efficacy or functional loss (e.g., medicine), and 3. perceived value loss (e.g., fashion clothes, high technology) where the first two are difficult to be exactly distinguished.

A further distinction can be made by the products' lifetime that can be *time-independent* or *time-dependent*. The first case implies a pre-specified lifetime. The utility of these products is given during a specific time period. The second case expresses deterioration over time, e.g., to model seasonality. The majority of authors working in the field of deterioration and perishability use the words interchangeably.

We make a clear distinction between perishable goods that cannot be used anymore and lose all their utility at once after a certain point in time whereas items that lose their utility gradually are regarded as being subject to deterioration.

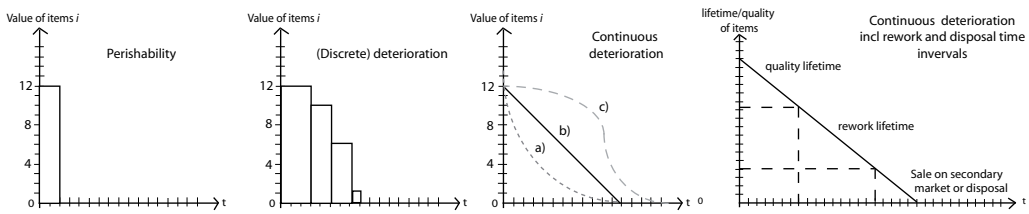

Figure 2: Courses of perishability and deterioration

Figure 2 depicts three functional relationships of item quality where the first graph shows the course of perishability, the second discrete and variational courses of deterioration and the third possible courses of continuous deterioration. The fourth graph gives three possible time intervals regarding the quality decrease of items that can be reworked: the first interval presents regular quality of items until the phase where items can be reworked, and the phase where items have passed their lifetime and might be sold on a secondary market or need to be disposed.

The definitions of deterioration, perishability, and lifetime constraints remain somewhat fuzzy as it clearly depends on the physical state/fitness, behavior of items over time, and on the production planner or the quality controller who has to decide whether to employ items or components having passed a certain age. In this paper, we focus on the integration of functional deterioration, perishability, and lifetime constraints in contrast to value loss due to customer demand variations that is expressed via demand patterns.

2.2. Mathematical Formulations of Deterioration and Lifetime Constraints of Items

Generally, there are different methods to integrate deterioration/perishability effects and lifetime constraints into mathematical models depending on the model type (deterministic or stochastic) and the considered time horizon (infinite or finite). One method that is frequently employed assumes the loss of a certain *fraction* or *part* of inventory stated as follows:

$$\frac{dI(t)}{dt} = -\Theta I(t) - D(t) \qquad t_j \leq t \leq T_j \tag{1}$$

where Θ is the shrinkage factor due to perishability, $I(t)$ the inventory holding in time t, and $D(t)$ is the time-dependent demand defined for a time interval (t_j, T_j). Models integrating this formulation assume exponential shrinkage of inventory which becomes clear when regarding the solution of Equation (1):

$$I(t) = -e^{-\Theta t} \int_t^{T_j} e^{\Theta u} D(u) du \qquad t_j \leq t \leq T_j \tag{2}$$

Such formulation is mainly used in infinite time deterministic and/or stochastic settings for replenishment and inventory control. For use in (discrete) dynamic deterministic settings, the following formulation might be employed:

$$(1 - \alpha^{\Theta}) I_t \qquad \forall\ t = 1, \cdots, T \tag{3}$$

where α^{Θ} is the percentage of inventory that is lost due to perishability. For the remainder of this paper, we adress these modeling approaches as *fraction formulation* (FF). They are mainly used in optimization models with the aim to derive optimal replenishment quantities and time lengths between orders assuming that all products in inventory undergo the same transformation independent of their age or their production period. This might be a good approximation in case of general deterioration studies [16]. Anyhow, the assumption of general inventory shrinkage is limiting if we want to create models for planning that are supposed to avoid deterioration in the first place. Therefore, we need formulations that restrict products' lifetimes. Items that lose their functional characteristics and quality in inventory cannot (or should not) be kept in inventory due to various reasons. For instance, food that deteriorates and develops biological toxines should be separated from fresh inventory and disposed immediately whereas items that lose their customers' perceived utility can be kept in inventory without such problems in order to be sold on a secondary market. It becomes clear that functional deterioration and depreciation due to value loss pertain to the same problem, but require different consequences or actions. The modeling of limited shelflives can be integrated by, e.g., modifying the inventory balance equation of related optimization models by subtracting items that passed their useful lifetime denoted by I_t^D and calculated as follows:

$$I_t^D \geq \sum_{\tau=1}^{t-\Theta_t^L} X_\tau - \sum_{\tau=1}^{t} D_\tau - \sum_{\tau=1}^{t-1} I_\tau^D,\ \forall\ t \geq \Theta_t^L \tag{4}$$

where I_t^D is the amount of items that deteriorate in a time period t that is calculated by the sum of produced items X_τ in the time interval $(\tau, t - \Theta_t^L)$, thus the production from the beginning of the planning horizon until period t reduced by the length of item lifetime Θ_t^L, less those items that are used to satisfy demand until period t and items that have already been disposed in previous

periods $t-1$. Contraint (4) can be used in all classical discrete deterministic lot sizing (and scheduling) models that are formulated as *inventory and lot size models* (I&L; see [17]) further including minimal lot size constraints [18, 19], so that we refer to this method as *I&L* formulation. However, these models become quickly computational intractable regarding practical input data sets. For example, the *capacitated lot sizing model* (CLSP) is proofed to be NP-hard [20, 21] and so are its extensions. Therefore, a third formulation is proposed in the literature that is mainly used in network flow formulations and integrated into the production constraint of related models:

$$\sum_{\tau=1}^{t} X_\tau \leq \sum_{\tau=1}^{t+\Theta_t^L} D_\tau \qquad \forall\, t = 1, \cdots, T - \Theta_t^L \tag{5}$$

which states that the sum of production until time period t must be smaller or equal the sum of demands that do not exceed time period t and the length of the variable lifetime of the product Θ_t^L. We name this method *index transformation* (IT) as it restricts the considered time periods by the lifetime of items. Network flow formulations are efficient in terms of calculation time to find the optimal solution, because they approximately describe the convex hull of the solution space. Nevertheless, the formulation given in Constraint (5) does not allow for minimal lot sizes.

Another type of index transformation is adopted especially for use in stochastic models to integrate lifetime restrictions. This type considers the periods of lifetimes from the time period an item is produced until its lifetime elapses:

$$T_{\Theta^L} - (n-1)t \geq T_{\Theta^E} \qquad \text{or} \qquad n \leq \frac{T_{\Theta^L} - T_{\Theta^E}}{t} + 1 \tag{6}$$

where T_{Θ^L} denotes the lifetime of an item, T_{Θ^E} the estimated or expected lifetime of an item, n an integer variable denoting the frequency of product shipments in a period, and t the current time period. Such formulations are adopted, e.g., by [22]. In the subsequent, we also refer to them as IT formulations. Further stochastic methods to integrate inventory shrinkage assume that the age of ordered items arriving in inventory are given by a vector of age classes of items [23] and defined to be *independent and identically distributed* (i.i.d.) random variables. Besides, the shrinkage of inventory might be assumed as a virtual outdating process similar to the virtual waiting time process in queuing theory [24].

The most frequent formulation used to integrate deterioration is FF mainly employed in infinite time horizon models where effects ofs demand patterns are added to study the influence of deterioration. The I&L formulation is recently developed and used by [18, 19] for discrete dynamic deterministic settings whereas IT-integrated lifetime constraints are mainly encountered in deterministic settings within approaches formulated as network flow models; see, e.g., [25, 26].

2.2.1. Model Extensions and Further Assumptions

Frequently, additional constraints or features are integrated into mathematical models with the aim of rendering them more realistic and applicable to real world problems. Situations might include the planning of multiple resources with or without capacity constraints that produce(s) a single or multiple product(s) adhering to multi-level product structures. In case of multiple products, setup times and costs that may be sequence-dependent play an important role especially regarding capacity utilization. Moreover, positive lead times might be included that can either be assumed for replenishment or inter-production processes treated as exogenious or endogenious input data. We find positive replenishment lead times mainly in stochastic dynamic replenishment models. The major part of reviewed references assume lead times to be negligible.

Besides, recent trends examine the possibilities and willingness of customers to wait for their ordered products. Early work includes the integration of complete backlogging which is extended to partial backlogging thus assuming that a certain number of customers is impatient which is further fine-tuned considering partial backlogging decreasing with the length of the waiting time. We highlight such extensions in the following sections. Due to limited space, we give extensions in the provided tables, but might not refer to them in the text.

3. The Consideration of Depreciation

The presentation of models incorporating functional depreciation is in line with Figure 1 considering the different areas of planning and extended perspectives. We divide approaches by their assumption on (un)certainty in deterministic and stochastic models. These categories are further separated in infinite time horizon and finite time horizon approaches.

3.1. Procurement, Replenishment, and Inventory Control

The production process is initialized by the need for raw materials calculated depending on the forecasted demand for finished goods. The aim is to obtain a sufficient amount of raw materials of the desired quality for production at the right time at minimal ordering costs. Marketing issues like discounting, allowances of credits, and lot sizing gain an increasing importance, because raw materials are generally ordered in batches taking advantages of price and volume discounts further minimizing transportation costs. Nevertheless, large order batches give rise to increased inventory costs and risks of deterioration. In case of order contracts like *vendor management inventory* (VMI) or *just in time* (JIT) agreements, such problems are avoided for the manufacturer and passed to suppliers who are responsible for the contractual partner's inventory level and thus the planning of replenishment batches. A review on dynamic pricing and inventory management is given in [27]. Models considering JIT logistics in supply chain planning are, e.g., considered in [28–34]. Additionally, the consideration of inflation and time value of money plays an important role in inventory control as risks of financial losses might increase with inventory levels. We refer to these as marketing issues, as well, because they are frequently considered with questions on pricing, replenishment, and inventory control. Besides, some approaches take *discounted cash flows* (DCF) into account, e.g., [35–38].

3.1.1. Deterministic Replenishment Approaches

This stream of literature is dominated by *economic order quantity* (EOQ) and *economic production quantity* (EPQ) models emphasizing marketing aspects like pricing/discounting, payment delays, inflation scenarios or time value of money, and studying different demand patterns; see Table 1 and Figure 3. Additional considered issues are mainly shortages and/or complete or partial backlogging of demand. Deterioration and perishability are integrated using FF in the infinite time case. Multiple products in this context are studied in [22, 39].

Infinite Time Replenishment Approaches. EOQ models integrating deterioration frequently assumed as Weibull functions prevail over EPQ models in this section. Table 1 gives approaches sorted by their consideration of marketing issues, demand patterns, and incorporation of deterioration/perishability whereas Table 2 resumes references emphasizing on other aspects. All listed approaches integrate general shrinkage of inventory employing FF given in Equation (1).

Table 1: Details on Deterministic Infinite Time Replenishment Models with Depreciation Regarding Marketing Issues

Marketing Issue	Demand	DT	References	Further Specifications			
				MF	S	BL	Comments
Pricing/	c	Wbd	[9]	EOQ	-	-	
Discounts	c	per	[40]	EOQ	-	-	p-dep oq
	l-i	per	[41]	EOQ	yes	com	cap restr
	p-d	t-v	[35]	EOQ	yes	-	DCF
	p-d	t-v	[42]	EOQ	yes	p	joint pricing
	s-d	per	[43]	EOQ	yes	p	
	t-v	Wdb	[44]	EOQ	yes	p	bdw
	p-d	t-v	[45]	EOQ	yes	p	two shops
	p-d/s-d	Wbd	[46]	EOQ	yes	p	bdw, pos TC, cap restr
Time value of	t-v	t-v	[47, 48]	EPQ	yes	p	
money	t-v	t-v,	[36]	EOQ	yes	p	DFC
	s-d	tw-p Wbd	[49]	EPQ	yes	p	
	p-d	t-v,	[50]	EOQ	yes	p	pricing
	p-d	per	[51]	EOQ	-	-	pricing, pos LT
Permissible delay	c	per	[52–54]	EOQ	-	-	pricing
in payments	c	twp-p Wbd	[55]	EOQ	yes	p	bdw
	l-i	twp-p Wbd	[56]	EOQ	-	-	-
	l-i	twp-p Wbd	[57]	EOQ	-	-	
	p-d	per	[58]	EOQ	-	-	pricing
	t-v	t-v	[59, 60]	EPQ	yes	p	mp, tvm, bdw

Abbreviations in columns

Column names	DT = depreciation type, MF = model formulation, S = shortages, BL = backlogging
Demand	c = constant, l-i = linear-increasing, p-d = price-dependent, s-d = stock-dependent, t-v = time-varying,
DT	Wbd = Weibull, per = perishability, t-v = time-varying, tw-p Wbd = two-parameter Weibull,
BL	com = completely, p = partially
Comments	p-dep oq = price dep on order quantity, cap restr = capacity restrictions, bdw = backlog dependent on waiting length, pos TC/LT = positive transportation costs/lead times, mp = multiple products, tvm = time value of money

Frequently, we find models studying effects of shortages that are partially backlogged as an extension in marketing focused approaches [36, 42–45, 47–50] whereas partial backlogging dependent on the waiting time is included rather into replenishment models without marketing issues [46, 61–69]; see Table 2. An EOQ setting together with a two-parameter Weibull distribution for the deterioration rate is introduced in [9] exploring effects of temporary price discounts on inventory and deterioration. Advantages of price discounts ordering large quantities while encountering increased inventory holding costs and risks of deterioration versus ordering smaller amounts thus not benefiting from discounts are analyzed. In fact, their results reveal that under these assumptions the model is sensitive to discounts as well as the deterioration rate. An integrated EOQ-EPQ model considering multiple items, viz. raw materials and final products linked via their bill of materials, together with considerations on permissible delay in payments, effects of inflation, and partial backlogging dependent on their waiting time is given by [59, 60]. The latter further considers variable cost parameters.

Table 2: Details on Deterministic Infinite Time Replenishment Models with Depreciation

Demand	DT	References	Further Specifications			
			MF	S	BL	Comment
p-d	l-i	[70]	EOQ	yes	com	
p-d	decr	[62]	EOQ	yes	p	bdw
r-t	t-p Wbd	[71]	EOQ	yes	com	
r-t	per, Wbd	[72, 73]	EOQ	yes	p	
r-t	t-v	[74]	EOQ	-	-	
s-d	per	[75, 76]	EPQ	-	-	
s-d	per	[61, 63]	EOQ	yes	p	bdw
t-v	t-v, tw-p Wbd	[77, 78]	EOQ	yes	com	
t-v	Wbd	[64–66]	EOQ	yes	p	bdw
t-v	per	[67, 68]	EOQ	yes	p	bdw, exp bdw
t-v	per	[79]	EOQ	yes	com	postponement
t-v	t-v	[69]	EOQ	yes	p	bdw
c	per	[39]	EOQ	-	-	cap restr, mp

Abbreviations in columns

Column names	DT = depreciation type, MF = model formulation, S = shortages, BL = backlogging
Demand	p-d = price-dependent, r-t = ramp type, s-d = stock-dependent, t-v = time-varying, c = constant
DT	l-i = linear-increasing, decr = decreasing, t-v = time-varying, two-p Wbd = two-parameter Weibull
BL	com = completely, p = partially
Comment	bdw = backlogging dependent on waiting time, exp bdw = exponential bdw, mp = multiple products

An exact formulation for overall inventory costs and mathematical methods to find the optimal solution for production of final products are proposed. Impatient customers together with permissible delay in payments are also considered in [55, 56]. In [56], the authors show that increasing permissible delays in payments result in significant decreases of average inventory costs per unit time. Demand increasing with a quadratic time-varying rate and complete backlogging is assumed in [77, 78] and ramp-type demand together with Weibull distributions for the deterioration rate considered in [71–74]. Effects of price-dependent demand functions are studied, e.g., in [35, 42, 45, 46, 50, 51, 58, 62, 70] where strategies of selling age-advanced items in a secondary shop at a lower price in order to maximize profits are analyzed in [45]. Positive transportation costs and capacity restrictions as well as stock and price-dependent demand influenced by show room advertisement are covered in [46]. Positive replenishment lead times are regarded in [51]. Otherwise, capacity restrictions are added in [39, 41, 46]. The impact of deterioration on postponement strategies in a replenishment scenario without marketing issues is studied in [79]. Postponement is defined as the possibility of a retailer to order either finished perishable products or product parts, so that the final product is assembled later according to customer requirements or a combination of both finished and unfinished products. Results show that in case of small perishability rates the postponement strategy gives decreased overall average costs. The major part of approaches assumes single products and thus abstracts from (sequence-dependent) setup times and costs, or multi-level product structures.

Finite Time Replenishment Approaches. Models might be extended to a dynamic setting by replicating the assumed procurement-inventory cycles n times. We name these models, e.g., EOQ/EPQ-based approaches.

Table 3: Details on Deterministic Finite Time Horizon Replenishment Models with Depreciation Regarding Marketing Issues

Marketing Issue	Demand	DT	References	Further Specifications			
				MF	S	BL	Comment
Pricing/Discounts	p-d	t-v	[80]	PR	yes	com	discounts
Time value of money	c	per	[81]	EOQ-b	yes	com	
	c	two-p Wbd	[37]	EOQ-b	-	-	DCF
	s-d	per	[82]	EOQ-b	yes	p	
	s-d	per	[38]	EOQ-b	yes	com	cap res, DCF
	t-v	t-v	[83]	EOQ-b	yes	p	bdw
Permissible delay in payments	s-d	per	[84]	EOQ-b	-	-	discounts
	l-i	t-v	[85]	EOQ-b	yes	p	
	c	per	[86]	EOQ-b	-	-	

Abbreviations in columns

Column names	DT = depreciation type, MF = model formulation, S = shortages, BL = backlogging
Demand	p-d = price-dependent, c = constant, s-d = stock-dependent, t-v = time-varying, l-i = linear-increasing,
DT	t-v = time-varying, per = perishability, two-p Wbd = two-parameter Weibull distribution
MF	PR = periodic review, EOQ-b = EOQ-based,
BL	com = completely, p = partially
Comment	cap res = capacity restrictions, bdw = backlogging dependent on the waiting time

The references in this section mainly employ FF in order to include shrinkage of inventory. Marketing issues in dynamic settings with their extensions are shown in Table 3. Finite time replenishment approaches emphasizing other aspects are listed in Table 4. For instance, a dynamic *periodic review* (PR) replenishment model with price-dependent demand for a time-varying deterioration rate including complete backlogging is proposed in [80] with the objective to determine periodic prices and replenishment lot sizes that maximize profits. A EOQ-based model with constant demand studying the effects of time value of money, perishability according to FF, and permissible delay in payments is proposed in [81]. DCF methods to derive optimal solution values for the net present value of total production costs regarding a two-parameter Weibull distribution for deteriorating items also subject to constant demand are examined in [37]. Approaches including stock-dependent demand are given in [38, 82] where the latter further restricts production capacity. Time-varying demand and deterioration rates are studied in [83] together with partial backlogging dependent on the waiting time. We find EOQ-based models including permissible delays in payments for demand being stock-dependent, linear-increasing, and constant in [84–86]. Models concentrating on the analysis of different demand patterns for deterioration and perishability are listed in Table 4 mainly presenting EOQ-based settings [10, 15, 87–91] or *economic lot size* (ELS) formulations [92–94] using FF to integrate shrinkage due to deterioration and/or perishability. Age-dependent inventory and partial backorders also dependent on their age are formulated in [92] using a slightly modified FF as given in Equation (3) where the age-dependent shrinkage of inventory is expressed as the difference between production in period τ and time until the current period t, viz.

$$I_{\tau t} = (1 - \alpha_{\tau,t-1})I_{\tau,t-1} - X_{\tau t} \qquad \forall\ 1 \leq \tau < t \leq T \tag{7}$$

Equation (7) is also used in [94] considering economies of scale cost functions. Complete backlogging is added in [15, 93].

Table 4: Details on Deterministic Finite Time Replenishment Models with Depreciation

Demand	DT	References	Further Specifications			
			MF	S	BL	Comment
c	two-p Wbd	[10]	EOQ-b	yes	p	
c	per	[91]	EPQ-b	-	-	multiple markets
t-v	per	[87]	EOQ-b	yes	p	
t-v	per	[88, 89]	EOQ-b	yes	p	bdw
t-v	per	[90]	DP, EOQ-b	-	-	
t-v	t-v	[92]	ELS	yes	p	
t-v	t-v	[93]	ELS	yes	com	eos
t-v	t-v	[94]	ELS	-	-	eos
l-i	per	[15]	EOQ-b	yes	com	

Abbreviations in columns

Demand	c = constant, t-v = time-varying, l-i = linear-increasing
MF	DP = dynamic programming, ELS = economic lot size model
Comment	minlot = minimum lot-sizes, bdw = backlogging dependent on waiting length, eos = economies of scale

3.1.2. Stochastic Replenishment Approaches

Replenishment approaches found in the literature considering deteriorating items are stochastic EOQ formulations that frequently serve as a starting point for more complex formulations. We find different integration methods for deterioration and perishability in stochastic infinite time approaches. There is less recent work on marketing issues in this section than in the deterministic case, but more frequent use of positive order lead times in infinite time models [95–98] and in finite time cases [97, 99–101].

Stochastic Infinite Time Formulations. In this category we find different marketing issues such as permissible delays of payments [102] and discounting [103, 104].

In [102], demand is assumed to be fuzzy whereas demand following a Poisson distribution or assumed as a Markovian arrival process is modeled in [95–97, 99, 100, 105]. A *continuous review* (CR) model including shrinkage due to perishability and defective items, so that the number of items in good conditions is a random number, further assuming demand as a Poisson process is given in [105]. Negative demand on backlogs may occur, so that the assumption of full backlogging might change to partial backlogging. Positive replenishment lead times following a phase type distribution are considered in a similar setting with demand assumed as a *Markov arrival process* (MAP). A Markov renewal process for demand for a similar setting is studied in [96]. Impatient customers assumed as negative demand are considered in [97, 105]. A queuing-based perishable inventory model is proposed by [24] featuring a virtual outdating process similar to a virtual waiting time process of traditional queuing theory; see also Section 2.2. Outdates takes place after one period.

PR models are proposed in [23, 98, 106, 107]. The value of information regarding the age of ordered items used in replenishment decision-making is analyzed in [23] where the lifetime of an item is fixed, but the age upon arrival is uncertain. The benefit of information sharing seems highest in case that the variability of demand is significant, products' lifetimes are short, and products' prices elevated. These benefits are not necessarily shared with upstream supply chain partners. Investments in equipment to maintain or extend the quality/freshness of items and thus increase products' lifetimes are compared to the benefits of information sharing.

Table 5: Details on Infinite Time Stochastic Replenishment Models with Depreciation

Demand	DT	DC	References	Further Specifications			
				MF	S	BL	Comment
fuzzy	per	FF	[102]	EOQ	-	-	pdp
MAP/Poisson	per	FF	[99, 100]	CR	-	-	positiv LT
MAP	per	FF	[95–97]	CR	yes	com/p	positive LT
Poisson	per	FF	[105]	CR	yes	com/p	
stochastic	per	VOP	[24]	QM	-	-	
stochastic	t-v	a-d v	[23]	PR	yes	-	
stochastic	per	IT	[98]	PR	yes	-	positive LT
stochastic	t-v	IT	[106]	PR	yes	com	substitution
stochastic	t-v	IT	[107]	PR	-	-	kind of substitution
stochastic	t-v	IT	[101]	CR	yes	com	positive LT substitution
stochastic	t-v	IT	[108]	M	yes	com	
s-d	per	FF	[103]	EOQ	yes	p	discounts
s/p-d	per	FF	[104]	EOQ	-	-	discounts
t-v	t-v	IT	[109]	QM	yes	-	priority demands

Abbreviations in columns

Demand	MAP = Markovian arrival process,
DC	VOP = virtual outdating process, a-d v = age-dependent vector
MF	QM = queuing model, M = Markov model,
Comment	pdp = permissible delay in payments, cap res = capacity restriction, LT = lead time,

Results show that lifetime-enhancing investments mostly outperform information sharing for certain lifetimes. Lifetime extensions are also obtained for blood products [109] determining optimal policies for priority demand levels.

Substitution of aged items with new products to satisfy random demand in a PR setting where items have a lifetime of two periods is studied by [106]. Demand for an item of a specific age can only be satisfied with items of that characteristic, so that no substitution is allowed, or excess demand for old items can be satisfied with new items (downward substitution), or excess demand for new item can be satisfied with old items (upward substitution). A forth setting allows downward as well as upward substitution. The remaining lifetime of items is included via IT. A infinite and finite time model for a seasonable good with similar characteristics as in [106] is proposed by [107]. Positive replenishment lead times are added to a similar approach in [101]. Moreover, positive replenishment lead times for a perishable product with a fixed lifetime in a PR model are considered in [98]. Perishability is integrated using IT denoting the amount of outdates at the end of the period. The consideration of random lifetimes of items in a discrete time Markov model is proposed by [108] where demand follows a discrete phase type renewal process and lifetimes a discrete phase type distribution.

Stochastic Finite Time Formulations. Marketing issues such as pricing, time value of money or inflation in EOQ-based approaches are covered in [110–113] where multiple items are considered in [111]. Stochastic demand with unknown distribution parameters that are periodically updated using a Bayesian approach relying on historical data finds integration in [114]; see Table 6.

Table 6: Details on Finite Time Stochastic Replenishment Models with Depreciation

Demand	DT	DC	References	Further Specifications			
				MF	S	BL	Comment
BM	per	portion	[110]	EOQ-b	yes	-	pricing
c	per	FF	[111]	EOQ-b	-	-	MC, tvm, multi items discounts, inflation
stochastic	per	indirect	[114]	PR	-	-	Baysian update (demand)
s-d	t-v	IT	[115]	PR	-	-	positive LT, RFID
p-d	per	IT	[112]	PR	yes	-	pricing, RFID
p-d	per	FF	[113]	EOQ-b	-	-	pricing, RFID
stochastic	per	IT	[116]	PR	yes	-	positive LT

Abbreviations in columns

Demand	BM = Brownian Motion, MAP = Markovian arrival process,
MF	PR = Periodic review, EOQ-b = EOQ-based
Comment	MC = multi-criteria objective, tvm = time value of money, LT = lead time,

Recent literature in stochastic finite time approaches often takes the use of *radio frequency identification* RFID technology into consideration in order to keep track of the age of inventory to enhance the quality of input data; see, e.g., [112, 113, 115]. In fact, traceability of items, especially food, and related systems can be regarded as a highly important operations management function [22]. However, in past years, item traceability has been considered rather separately from *supply chain management* (SCM) strategies as stated in [22]. There is a visible trend to consider traceability and thus better information on inventory. RFID controlled inventory subject to stochastic, seasonal demand in a PR model is studied in [115] where demand depend on remaining lifetimes and thus freshness of items as described in [114]. Dynamic pricing for deteriorating (food) products subject to price- and freshness-dependent demand is examined by [113]. Price adjustments based on demand forecasting gained via RFID technology enable, e.g., supermarkets to automatically mark down items near their expiration date in order to stimulate demand. Results of the optimal pricing and ordering strategy show that the deterioration rate has a significant influence on pricing as well as the order quantity.

3.2. Tactical and Operational Production Planning with Lot Sizing and Scheduling

Tactical production planning deals with questions concerning the implementation of decisions from the strategic level that are translated in mid-range decisions regarding, e.g., the production network design and layout, arrangement and number of machines, and the product mix. Lot sizing is an important determinant of lead times and, consequently, the quality of items especially in case of multiple products enforcing setups. This is because lot sizing decisions determine the frequency of setup times necessary to switch from the production of one product to the production of another and, in some cases, the time that items wait for lot completion before moving. Capacity is increasingly consumed that, otherwise, would be available for production. Accordingly, setup times affect lead times of products due to induced waiting times. Lead times increase with the number (and duration) of setups and so does the number of products that deteriorate while waiting in line to be processed. This might be significantly aggravated in case that setup times and costs are sequence-dependent, so that some scheduling information should be considered already at this planning stage.

3.2.1. Deterministic Production-Inventory Models

In general, lot sizing is regarded in aggregate production-inventory. In case that marketing issues prevail as important factors of examination, we relate approaches to Section 3.1. Here, other aspects are of major interest such as the possibility of adapting production rates to various situations, e.g., increased demand rates including ad hoc priority orders. Production rates might be regulated regarding previous or subsequent production processes where items frequently have to wait because of bottleneck machines.

Infinite Time Production-Inventory Approaches. Traceability of items subject to temporary price discounts in an EPQ approach is studied in [22]. Perishability is integrated via IT as given in Equation (6); see Table 7. Simulation results emphasize the benefits of traceability systems despite investment costs for system components, barcode systems, wireless scanners, etc.

Table 7: Details on Infinite Time Deterministic Production-Inventory Models with Depreciation

Demand	DT	DC	References	Further Specifications			
				MF	S	BL	Comment
c	per	FF	[117]	ELSP/EPQ	-	-	multiple products
c	per	IT	[22]	EPQ	-	-	discounts, traceability
t-v	t-v	FF	[118]	EPQ	yes	p	bdw
t-v	t-v	FF	[119, 120]	EPQ	yes	p	learning
t-v	t-v	FF	[121]	EPQ	-	-	(learning/forgetting)
t-v	t-v	FF	[122]	EPQ	-	-	var prod rates
s-d	per	FF	[123]	EPQ	-	-	var prod rates
p-d	per	FF	[124]	EPQ	yes	p	bdw
r-t	t-v	FF	[125]	EPQ	yes	com	

Abbreviations in columns

Comment var prod rates = variable production rates

Multiple products thus setup times and costs are considered by the authors in [117] assuming different perishability and demand rates. Production-inventory models for single items subject to perishability with and without shortages and partial backlogging dependent on the waiting time are examined in [118, 124]. Time-varying demand in such setting finds consideration in [119, 120] where the latter analyzes effects of learning on the optimal production lot size. Learning as well as forgetting influencing the production rate complicates decision making in [121]. Variable production rates are as well decision variables in [122, 123]. The latter assumes production rate-dependent production costs which is further done in [125] considering a deteriorating item subject to ramp-type demand. Complete backlogging of demand is considered in a second modeling step.

Discrete Dynamic Approaches Regarding Lot Sizing. As stated before, models assuming a finite planning horizon which is divided into n cycles of equal duration are considered as dynamic in nature. Such EPQ-based approach is considered in [126] regarding perishability of an item subject to time-varying demand that can be completely backlogged. Pricing of a deteriorating item a similar EPQ-based model together with capacity restrictions is presented in [127]. *Mixed-integer programming* (MIP) models are frequently formulated for discrete dynamic production-inventory systems.

Table 8: Details on Finite Time Deterministic Production-Inventory Models with Depreciation and Lifetime Constraints

Demand	DT	DC	References	Further Specifications			
				MF	S	BL	Comment
t-v	per	FF	[126]	EPQ-b	yes	com	
p-d	t-v	FF	[127]	EPQ-b	-	-	pricing
c	per	FF	[128]	MIP	-	-	
t-v	t-v	IT	[25]	MIP, NF	-	-	minlot, sd-st/sc, mp
t-v	t-v	IT	[129]	MIP	-	-	pricing, minlot, mp
t-v	per	I&L	[18, 19]	CLSP, GLSP			minlot
t-v	per	IT	[130]	GLSP-PL			minlot, multi-crit

Abbreviations in columns

MF MPSSP = multi-period single-sourcing problem, TSP = Travelling Salesman Problem

Comment minlot = minimum lot-sizes, sd-st/sc = sequence-dependent setup times and costs, mp = multiple products

We find an early reference considering a MIP formulation assuming that a part of produced items is lost due to deterioration in each time period (FF, see Equation (3)) by [128]. A CLSP formulated as a network flow model including lifetime constraints (IT) of perishable products, minimal lot size constraints, and sequence-dependent setup times and costs is presented in [25]. The work of [129] proposes block planning with the aim at determining the optimal profit maximization and thus pricing strategy dependent on the item quality at the sales day. A block is defined as the production of products based on the same bill of materials. Capacity restrictions and minimum batch sizes add complexity. Multiple products are considered, but only products of different blocks require setups operations of the production resource.

A study on integrating lifetime constraints in classical lot sizing models is given by [18]. They develop an I&L formulation to restrict products' lifetimes according to Equation (4). Hybrid lot sizing models including a micro-marco period structure considering sequence-dependent setup times and costs together with lifetime constraints are studied by [19]. Results show that their inclusion limits the solution space, so that models become more difficult to be solved in case that lifetime constraints are greater than one time period. For instance, if the lifetime of items is exactly one period, the model employs a *lot-for-lot* policy, i.e., lot sizes match demands of products in that period, so that no inventory is build up, the machine is set up to every requested product in that period, and no aggregation of lot sizes takes place. If lifetimes are greater than one period, evaluation of different lot sizing possibilities starts leading to longer calculation times or larger optimality gaps if calculation time is limited [19]. A multi-criteria lot sizing and scheduling model for an example in the food industry is proposed by [130] incorporating IT.

3.2.2. Stochastic Production-Inventory Models

In general, stochastic production-inventory control models can be divided into CR and PR models. In PR systems inventory levels are measured and adjusted only at specific points in time whereas in CR systems the monitoring is continuous, so that exact inventory levels are known at every point in time.

Infinite Time Horizon Approaches. Recent infinite as well as finite time stochastic models for production and inventory are proposed by [131, 132]. Periodic as well as continuous stochastic review EPQ models are formulated by [131] including a two-parameter Weibull distributions for

the deterioration rate which is modeled as shrinkage thus using FF. Inventory levels are regarded as the state variable and the production rate as the control variable. The Weibull distribution function includes some probability for deviations or expected values. The models becomes deterministic assuming fixed parameters for the Weibull distribution regarding discrete time points. A stochastic EPQ formulation reflecting an unreliable production system that might produce defective items subject to perishability in inventory is proposed in [132]. The production system shifts from an in-control state to an out-of-control state where defective items are produced characterized by fuzzy numbers or random variables that follow an exponential distribution with fuzzy parameters. Inventory shrinkage due to perishability is expressed via FF.

Finite Time Horizon Approaches. We find finite time horizon stochastic approaches in [133–135] formulated as continuous as well as periodic models with perishability. Variable production rates are studied in [133]. The work in [134] includes uncertain production preparation times modeled as fuzzy numbers as well as related setup costs. Cost minimization as well as a multi-objective functions including marketing investments to enhance demand are examined. The authors in [135] present a simulation study regarding a two-stage production planning and scheduling system including capacity and lifetime restrictions of items in intermediate storage thus WIP. They define different performance criteria, i.e., flow times, makespan, unfinished orders, and the amount of waste where the latter restricts WIP lifetimes via IT.

3.3. Final Inventory Management and Delivery

The final assembly of orders/jobs, their packaging, distribution or transportation to different warehouses is the task of the *outbound logistics system*; see Figure 1. The physical distribution to distribution centers, wholesalers, or customers is a complex problem spanning all planning levels from strategic to operational planning with decisions ranging from where to place warehouses to transshipments in order to satisfy ad hoc demand. In case of multiple shops, transshipments of items subject to perishability can be effective to avoid shortages as well as outdating. Depreciation and thus disposals might be reduced by one shop maintaining a good inventory level while preventing shortages in another shop. Additionally, if transportation of transshipments are integrated in regular deliveries, no increased transportation costs occur. Related problems are, e.g., due to excess inventories, flawed inventory transshipments, shortages due to incorrect inventory data, and/or ad hoc priority orders. Technologies like RFID may support the monitoring problem providing real-time data of inventories [136]. Therefore, distribution logistic involving inventory management is closely related to the consideration of multiple warehouses, supply chain design, and settings. Issues of managing the distribution of deteriorating items such as food regarding safety, quality, and sustainability are highlighted in [137]. Integrated planning of production and distribution is considered by [138]. The examination of JIT deliveries are mainly studied in supply chain contexts and thus are discussed in Section 3.6. However there are only a few approaches considering positive logistic times and costs together with deterioration, perishability or lifetime constraints. The next section regarding multiple warehouses might be considered a part of final inventory, assembly, and delivery. For sake of clear arrangement, we dedicate an own section to it.

3.4. Multiple Warehouses

Final assembly and delivery become more complex in case of more than one storage facility with different environmental conditions, e.g., diverse depreciation rates and/or inventory holding

costs. In the literature, prevailing practice is to consider two warehouses, the *own warehouse* (OW) with limited capacity and a *rented warehouse* (RW) that generally charges higher inventory costs. Items are first stored at the OW. Overflows go to the RW where they are retrieved with priority in order to keep renting costs reduced. Generally, no fixed renting costs are assumed, but a unit rent cost for each batch of items stored. On the other hand, different assumptions might be true in practice. For instance, RWs with specialized services could achieve and offer lower inventory charges due to learning effects of trained personnel and economies of scale deriving from high volumes or high(er) overall competition in the market. The integration of positive logistic times and costs are rare in recent model approaches considered in this section.

3.4.1. Deterministic Cases

Generally, models including two warehouses assume that the RW is used in case that items exceed the maximum capacity of the OW. Discounting and time value of money considered in a two warehouse setting are important, because they influence replenishment from partners downward the supply chain and, therefore, might contribute to the bullwhip effect. However, depreciation might degrade related price advantages, so that ordering increased quantities to ensure discounts cease to be attractive.

Infinite Time Approaches Regarding Two Warehouses. A two warehouse setting considering the net present value and effects of discounting is examined by [139] including complete backlogging. Different perishability rates or major rates for the RW are assumed. Partial backlogging dependent on the waiting time finds its consideration in [140]. LIFO and FIFO policies of a two warehouse setting with deterioration are studied by [141]. Results show that FIFO is less expensive than LIFO provided that the deterioration rate is the same for both warehouses and that holding costs at the RW are less than in the OW. Exponentially distributed deterioration rates as well as permissible delay in payments are assumed by [142].

Finite Time Approaches With Two Warehouses. A dynamic setting with shortages that are completely backlogged is proposed by [143] assuming that each cycle begins with, but ends without shortages further regarding effects of inflation and superior inventory costs in the RW than in the OW. The model operates in a less expensive way if the inflation rate is greater than zero. Partial backlogging allowed in all cycles is included by [144]. They further consider positive transportation costs of items from the RW to the OW. Nevertheless, transportation times are assumed to be negligible.

3.4.2. Stochastic Cases

We found one recent approach considering a stochastic two warehouse approach given by a set-covering location model for a deteriorating as well as an ameliorating item including positiv transportation costs [145]. The aim of the model is to determine the number of warehouses, so that the probability that each customer is covered is below a certain value.

3.5. Rework and Remanufacturing

All work reviewed so far abstract from the possibility that spoiled items can be repaired instead of being disposed. In many situations, it is economically reasonable to rework deteriorated products in order to reuse them for production, e.g., if rework times are significantly smaller than regular production times or raw materials are rather expensive. Moreover, many companies are concerned about their responsibilities with respect to sustainability driven by many factors,

e.g., reactive regulatory reasons and/or proactive strategic and competitive advantage aspects. Besides, legislation in an increasing number of countries places the responsibility for environmentally safe disposal of end-of-life products primary on the producer. Governments tend to increase taxes on raw materials in order to motivate companies to enhance recycling activities.

We distinguish between external and internal reverse logistics (see also Figure 1): the external reverse logistics system deals with used items returning from external resources, thus customers or retailers whereas the internal reverse logistics system treats parts and products that deteriorate while waiting and thus need rework. Items can be reworked either on machines of the regular production system which is named *in-line processing* or on other machines parallel to the production system named *off-line processing*. The major part of models in the literature including deterioration or perishability assume in-line processing together with a *one-by-one* policy where a batch of regular production follows a batch of rework. For a review on rework see [146]. An overview on recent research on green supply chain management is provided in [147].

3.5.1. Some Useful Definitions

All actions encompassing rework and remanufacturing pertain to the broad field of *reverse logistics*. The term *rework* includes all recovery actions required to transform products that do not meet a pre-specified quality standards (anymore) in a way that they fulfill the quality requirements. Authors using rework in that sense are, e.g., [148, 149]. In general, rework activities take place before items are distributed to the customer due to encountered problems. Reasons for rework include unreliable production processes, e.g., *out-of-control* production systems producing defective items [146]. Besides, the term *remanufacturing* encompasses industrial processes where worn out products that are returned from the customer are restored to *like-new* conditions through a series of actions, e.g., disassembly, clearing, refurbishing etc. [150, 151]. In contrast to this, the term *recycling* is used for the recovery of (raw) materials of used products, e.g., the copper of electronic parts. For that purpose, returned products are disassembled in order to regain useful materials. Hence, recycling and remanufacturing pertain rather to the external reverse logistics system whereas we define rework as a company internal process. External reverse logistic models are extensively studied in the literature and are worth a separate treatment. Thus, we refer the interested reader to [152–154].

We focus on the work that assumes deterioration on waiting items that might, but need not acquire a defect in the production system due to an out-of-control situation. While waiting in rework inventory, they are subject to deterioration.

3.5.2. Deterministic Rework

Generally, it is assumed that a constant fraction of items in a production batch acquire a defect that can be adjusted by rework. After rework, items are assumed to regain an *as-good-as-new* quality, so that demand may generally be satisfied either by newly produced or reworked items. Extensions to the classical EOQ model in regard of remanufacturing and rework are provided by several authors. We consider remanufacturing to the extent that an assumption is made of a flow of returned items that may be worth reworking or have to be disposed, but in any case increasing the storage of reworkable items that are subject to deterioration. Assumptions considering the point of origin of reworkables raises the complexity of the system, especially when amounts of return flows are unpredictable or subject to uncertainty. If they are assumed to be deterministic, the issue is less critical. Besides, returns may increase inventory costs and machine utilization, but decreases replenishment of new items and thus might be a desirable strategy.

Infinite Time Rework Systems. Defectives due to an unreliable production process that are subject to perishability of reworkable items of a batch are modeled by [148, 155] in an EPQ approach. They employ an average waiting time approach which is frequently used to avoid the consideration and monitoring of individual items resulting in less complex formulations and solution approaches. Defective items occuring uniformly with the same percentage in each production lot size are subject to perishability thus increasing rework times and costs because of the increasing amount of items waiting to be reworked. The option of disposing itmes is not allowed, thus all defective or deteriorated items have to be reworked. Therefore, the lifetime of items is not restricted in a sense that after passing a certain amount of time, deteriorated items cannot be reworked anymore and have to be disposed as depicted, e.g., in the fourth graph of Figure 2. This assumption is relaxed in [156]. Multiple products are discussed in [155] and multiple parallel machines in [157].

3.5.3. Stochastic Rework Approaches

Recent models regarding stochastic rework approaches with depreciation are rare. To our knowledge, the most recent work is provided by [149, 158]. A stochastic EPQ approach with similar assumptions as in [148, 155], but assuming probabilities for the occurrence of defectives and stochastic times and cost values, is proposed in [149]. The system is triggered by the necessity to rework items rather than by demand functions. Production batches are assumed to be of a fixed size, so that the aim is to find the optimal production-rework switching policy and optimal rework batch sizes, so that average profits are maximized. Disposal of non-reworkable items is allowed. They derive the closed-form expression for expected profits in order to calculate the optimal rework batches. Recycling in a stochastic EPQ approach where recycled materials are used as raw materials subject to a random lifetime expressed by a Weibull deterioration rate is studied in [158]. Recycled materials are assumed to arrive at a constant rate and is used according to LIFO policy. The authors state that deriving the inventory levels assuming a Weibull deterioration rate is rather complicating due to the dependence of deterioration on the age of each items. Thus, despite numerical analysis, they use an approximation formulations based on perturbation technique.

3.6. Integrated Approaches: Supply Chain Settings

The management of supply chains especially regarding their order management is an important topic that becomes more apparent with the shortening of product life cycles and inherent risks for supply partners, e.g., regarding their replenishment policies, specifically for upstream partners. The objective of SCM is the effective and efficient coordination of information, material, and financial flows so as to minimize overall system costs or maximizing overall profits in terms of reactivity, flexibility, time constraints, and quality with the aim of satisfying final customers. We define a supply chain as a relation of at least two companies or legally independent actors that exchange materials, information and financial goods in order to add value to a product or service.

3.6.1. Deterministic Approaches for Supply Chain Settings

Since collaboration of supply chain partners is vital in order to reduce overall supply chain costs, joint strategies are increasingly studied. Integrated production-inventory models considering the perspective of suppliers, producers and buyers show that reduced joint costs can be obtained. Nevertheless, it frequently occurs that the costs of one supply chain partner are higher than those of another. Consequently, incentives must be incorporated into the system to share

benefits of collaborative planning. The role of revenue-sharing contracts in supply chains is discussed in [159] analyzing dynamic settings especially regarding downstream competition. In recent literature, the prevailing part of approaches extend or link EOQ and EPQ approaches, accordingly. We refer to such integrated models as *related EOQ* or *related EPQ* models; see, e.g., Tables 9 and 10.

Infinite Time Horizon Supply Chain Approaches. Recent supply chain approaches consider mainly shrinkage of inventoried items due to deterioration or perishability employing FF; see Table 9.

Table 9: Details on Demand Rates in Deterministic Infinite Time Horizon Supply Chain Models with Deterioration and Perishability

TH	Demand	DT	DC	References	Further Specifications			
					MF	S	BL	Comment
stat	c	per	FF	[28]	rel EPQ	-	-	JIT, p-b (1:n)
stat	c	per	FF	[160]	rel EPQ	-	-	s-p-b (1:1:1)
stat	c	per	FF	[29]	rel EPQ	yes	-	JIT, p-b (1:n)
stat	c	per	FF	[30]	rel EPQ	yes	-	JIT, s-p-b (1:1:1)
stat	decr, t-v	t-v	FF	[161]	rel EPQ	yes	p	p-b (1:1)
stat	t-v	per	FF	[162]	rel EOQ	-	-	mp, cap restr, p-b (1:1), pricing
stat	c	per	FF	[163]	rel EPQ	yes	com	p-b (1:1)
stat	c	per	FF	[164]	EPQ	yes	p/com	bdw, p-b (1:1)
stat	c	per	FF	[31]	rel EPQ	-	-	JIT, p-b (1:1) positiv tc and ic
stat	s-d	per	FF	[165]	rel EPQ	-	-	p-b (1:1)
stat	c	per	FF	[166]	rel EOQ	-	-	pdp, p-b (1:1)
stat	c	per	FF	[33]	rel EPQ	-	-	JIT, prod design, p-b (1:1),
dyn	t-v	per	IT	[167]	MPSS,NF	-	-	p-b (n:n)
dyn	t-v	per	IT	[26]	MPSS,NF	-	-	p-b (n:n)
dyn	c	per	IT	[34]	MIP	yes	com	p-w-b (n:n:n)

Abbreviations in columns

TH — TH = time horizon, stat = static, dyn = dynamic

Comments — tc and ic = transportation costs and inspection costs, NF = network flow model, s-p-b = supplier-producer-buyer relation, p-w-b =producer-warehouse-buyer relation

Different supply chain settings are studied, e.g., 1) single suppliers and producers or single producers and buyers [31, 33, 161–166], 2) single suppliers with multiple buyers [28, 29], 3) multiple producers and buyers [26, 167], or three parties [34, 160]. As shown in Table 9, the most frequent setting contains two parties: a producer and a buyer.

Integrated and non-integrated approaches are studied in [164] considering a single-producer single-buyer supply chain where the buyer decides upon the number of replenishment quantitites, so that his average replenishment costs are minimized which need not lead to vendors' optimal average production costs. The objective is formulated regarding multiple objectives. Minimum average overall costs are obtained in the integrated approach. Nevertheless, pareto-optimal solutions do not necessarily give minimum costs for both the customer and the manufacturer as in the separated case. The supplier's perspective in an infinite time horizon setting is integrated by

[160]. The model aims at deriving the optimal number of predetermined amount deliveries and order lot sizes while minimizing joint total costs. However, lead times are assumed to be negligible as well as transportation lead times and costs. JIT lot splitting in an related EPQ-supply chain model is examined by [28] deriving from the idea to split a single order into multiple deliveries. The number of deliveries is inversely proportional to the lot size. Results show that the delivery quantity increases when the perishability rate decreases; see also [29]. The cost of perishability is assumed to be constant but varies for raw material and FGI. Shortages are not allowed, but considered together with JIT deliveries and backlogging effects on downstream members of the supply chain by [30]; see also [29, 161, 163] for the integration of shortages and partial or complete backlogging.

Green component-value design and remanufacturing options together with associated costs in an related EPQ approach with JIT deliveries are proposed by [33]. Multiple products subject to perishability, quantity discounts, and increasing demand are studied in [162]. Major and minor setup costs in a production/ordering cycle are examined where the latter can be interpreted as, e.g., fixed freight fees. The authors study different replenishment situations with and without coordination.

Results show that a centralized policy together with joint replenishment is always better than any decentralized or individual policy. Effects of supply chain coordination for stock-dependent demand rates of a deteriorating item are examined by [165]. A centralized supply chain policy is shown to be superior to decentralized planning in terms of increased profit which seems to be especially true for highly significant deterioration rates. A similar model with permissible delay in payments is provided in [166]. Quality control inspections for final products that are delivered assuming JIT and positive transportation costs are covert in [31]. As shown in Table 9, all cited references assume perishability modeled as shrinkage except for [161].

Finite Time Multi-Stage Supply Chain Setting. We find three recent deterministic, dynamic approaches including supply chain settings; see Table 9. Integrated distribution management on a tactical as well as an operative level of production is presented by [167]. The authors assume a multi-stage supply chain for a perishable product integrated by IT (Equation (5)) produced at multiple facilities; see also [26] for a similar approach. Facilities have the same production capabilities, but different time-variable setup, production, inventory holding, and transportation costs. Transshipments of products between facilities/vendors or inventory holding at the retailer are not allowed which is a typical restriction in food industries where retailers are supermarkets or restaurants with limited storage space. The problem is formulated first as a MIP assuming demand to be known and, second, as a network flow model based on the multi-period single source assignment property. Small test instances are solved using CPLEX. For large test instances, A primal-dual algorithm is proposed to solve this NP-hard problem considering large test instances. Three levels of a supply chain including multiple producers, warehouses, and buyers in a dynamic setting is proposed in [34]. Further features considered in the approach are JIT deliveries executed by different transportation means and related costs as well as pricing of a perishable item with a restricted lifetime. Transportation times are assumed negligible. As in [26], small test instances are solved using CPLEX and large sets by genetic algorithms.

3.6.2. Stochastic Supply Chain Settings

When demand flows are uncertain, the challenge of supply chain partners is to determine replenishment order so as to minimize excess inventory while avoiding shortages. In this section, single-producer single-buyer supply chains are prevailing [11, 32, 168–170].

Table 10: Details on Demand Rates in Stochastic Infinite/Finite Time Horizon Supply Chain Models with Deterioration and Perishability

TH	Demand	DT	DC	References	Further Specifications			
					MF	S	BL	Comment
stat	c	per	FF	[170]	MONLP	-	-	mc, pricing, cap restr, faGP, p-b (1:1)
stat	p-d/s-d	per	IT	[169]	PR	-	-	pricing, p-b (1:1)
stat	stochastic	per	IT	[11]	PR	yes	-	VMI, pos LT, p-b (1:1)
stat	c	per	FF	[32]	rel EPQ	yes	p/com	defectives, JIT, p-b (1:1)
stat	stochastic	per	FF	[168]	rel EOQ	yes	-	p-b (1:1), CPFR

Abbreviations in columns

Comment mc = multi-criteria, faGP = fuzzy additive goal programming,

Fuzzy inventory control formulated as a *multi-objective nonlinear problem* (MONLP) is examined by [170] analyzing the effects of encouraging buyers via price discounts on item perishabilty. Uncertainties in the multi-objective function are expressed using fuzzy linear membership functions. Stock-dependent as well as price-dependent demand for a lifetime restricted item is investigated in [169]. The value of information sharing in supply chains examining centralized and decentralized control is studied by [11] assuming fixed replenishment lot sizes for a item with lifetime constraints modeled via IT. Results show that supply chains have greatest benefit from centralized control or information sharing if product lifetimes are short, batch sizes large, demand uncertainty rather high, and penalty costs for deviations in demand and supply increased. Perishability considerations together with imperfect quality expressed via defectives are given in [32] using an infinite time model approach that spans lot sizing, production, inventory, JIT delivery, and the customer perspective. The probability of a percentage of items acquiring a defect in a production run is assumed to be uniformly distributed. Perishability is integrated by FF as described in Equation (1). The authors of [168] cover considerations of *collaborative planning, forecasting and replenishment* (CPFR) in a single supplier single buyer supply chain. Replenishment lead times are assumed to be positive. They provide mathematical formulations for the optimal replenishment policy that minimizes total costs.

4. Conclusions and Further Research

In this paper, we provided an extensive discussion on depreciation effects and lifetime constraints including a classification scheme of approaches that grouped mathematical models regarding the quality of information and with respect to the assumed time horizon. We considered relevant company and supply chain functions. Further interest was directed to rework or remanufacturing of items that might aquire a defect in production and deteriorate over time while waiting to be reworked. We extended the view to strategies for supply chain planning in regard of deterioration, perishability, and lifetime constraints of items.

We note that there is little research on multiple items considering setup times and costs that might be sequence-dependent especially in dynamic settings. The major part of the considered

literature integrates shrinkage due to deterioration or perishability into their models. Lifetime constraints are rarely incorporated.

The number of references is steadily decreasing going along the business functions except for supply chain approaches as shown in Figure 3 that build the structure of the paper. We see that deterministic models prevail over stochastic as well as static over dynamic approaches. Some discrete dynamic model approaches incorporating depreciation by integrating lifetime constraints exist, but such approaches are not extensively explored. We believe that these models are important, because they allow for determining plans for production as well as SCM aiming at avoiding deterioration and perishability in the first place. Due to discrete dynamic time structures they might be integrated in frequently applied planning software such as enterprise resource planning and solved using exact or heuristic algorithms.

Business Functions / **Overview**	Models including Depreciation			
Model Type	deterministic		stochastic	
Time Horizon	infinite	finite	infinite	finite
Replenishment/ Procurement/ Sales and Marketing	[9,35,36,39,40,41,42,43,44,45, 46,47,48,49,50,51,52,53,54,55, 56,57,58,59,60,61,62,63,64,65, 66,67,68,69,70,71,72,73,74,75, 76,77,78,79] main features: - bdw - capacity restrictions	[10,15,37,38,80,81,82,83,84,85,86, 87,88,89,90,91,92,93,94] main features: - bdw - capacity restrictions	[23,24,95,96,97,98,99,100,101,102, 103,104,105,106,107,108,109] main features: - positive lead times - substitutions	[110,111,112,113,114,115, 116] main features: - RFID
Aggregate production planning and scheduling	[22,117,118,119,120,121,122, 123,124,125] main features: - learning - variable production rates	[18,19,25,126,127,128,129,130] main features: - minlot - multiple products	[131,132]	[133,134,135,]
Inventory management/ distribution planning	[136,137,138]			
Multiple warehouses	[139,140,141,142]	[143,144]	[145]	
Rework/ internal reverse logistics	[148,155,156,157]		[149,158]	
Supply chain setting	[28,29,30,31,33,160,161, 162,163,164,165,166] main features: - JIT	[26,34,167]	[11,32,168,169,170]	

Figure 3: Overview on references in related planning functions

We visualized the number of recent approaches in Figure 3 that clearly shows research gaps regarding recent approaches, e.g., deterministic finite as well as stochastic infinite and finite final inventory and distribution management, deterministic or stochastic finite rework, stochastic finite approaches considering multiple warehouses, or stochastic finite supply chain approaches. Furthermore, we found only few models considering positive replenishment lead times. In this regard, most references provide stochastic infinite as well as finite time approaches. Surprisingly, positive lead times due to waiting in the production process are not examined. Queuing theory providing formulations for situations with more than on production resource might be employed, accordingly. We find approaches determining optimal production rates for perishable

or deteriorating items incorporated via FF in infinite time settings. Nevertheless, it might be useful to consider production rates that are adjusted in case of high workloads (due to demand or bottlenecks) or short product lifetimes in dynamic settings in order to avoid disposals. Recent references on such considerations are not available (yet). Besides, only a few references are available that regard multiple products and related setup times and costs. This might be due to enhanced complexity especially in dynamic settings. As such situations prevail in practice, future research should include multiple items and related problems such as sequence-dependent setup times and costs, and/or multi-level production structures.

Research on supply chains shows enhanced attention including a variety of settings regarding multiple up- and downstream partners. EOQ and EPQ-related approaches mainly examine different joint replenishment policies employing effects of pricing or JIT. Dynamic approaches are further available formulated as network flow or MPSS problems. Also in these cases, positive lead times are not considered which is a simplifying assumption. Anyhow, collaborative and joint supply chain planning regarding global partners and thus locations might include such important issues on timing in order to enhance plans and avoid wastages due to deterioraion, perishability, and lifetime constraints on items.

References

[1] J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk, A. Meybeck, Global food losses and food waste, Tech. rep., Study conducted for the International Congress SAVE FOOD! at Interpack 2011, Düsseldorf, Germany for the Food and Agriculture Organization of the United Nations (2011).

[2] J. Pahl, S. Voß, D. L. Woodruff, Production planning with deterioration constraints: A survey, in: J. Ceroni (Ed.), The Development of Collaborative Production and Service Systems in Emergent Economies, Proceedings of the 19th International Conference on Production Research, IFPR, Valparaiso, Chile, 2007, pp. 1–6.

[3] R. Li, H. Lan, J. R. Mawhinney, A review on deteriorating inventory study, Journal of Service Science and Management 3 (2010) 117–129.

[4] B. Karmakar, K. D. Choudhuri, A review on inventory models for deteriorating items with shortages, Assam University Journal of Science and Technology: Physical Sciences and Technology 6 (2010) 51–59.

[5] I. Z. Karaesmen, A. Scheller-Wolf, B. Deniz, Managing perishable and aging inventories: Review and future research directions, in: K. G. Kempf, P. Keskinocak, R. Uzsoy (Eds.), Planning Production and Inventories in the Extended Enterprise, Springer, 2011, Ch. 15, pp. 393–436.

[6] J. Sarkis, L. M. Meade, S. Talluri, E-logistics and the natural environment, Supply Chain Management: An International Journal 9 (4) (2004) 303–312.

[7] R. Darlington, S. Rahimifard, A responsive demand management framework for the minimization of waste in convenience food manufacture, International Journal of Computer Integrated Manufacturing 19 (8) (2006) 751–761.

[8] M. E. Ferguson, O. Koenigsberg, How should a firm manage deteriorating inventory?, Production and Operations Management 16 (3) (2007) 306–312.

[9] B. J. Shah, N. H. Shah, Y. K. Shah, EOQ model for time-dependent deterioration rate with a temporary price discount, Asia Pacific Journal of Operational Research 22 (4) (2005) 479–485.

[10] P.-X. Zhao, An EOQ model for items with Weibull distribution deterioration, in: 2nd IEEE Conference on Industrial Electronics and Applications, 2007, pp. 451–454.

[11] M. E. Ketzenberg, M. Ferguson, Sharing information to manage perishables, available at: http://smartech.gatech.edu/handle/1853/10768, last call August 2011 (2005).

[12] E. K. Boukas, Z. K. Liu, Manufacturing systems with random breakdowns and deteriorating items, Automatica 27 (3) (2001) 401–408.

[13] M. Tsiros, C. M. Heilman, The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories, Journal of Marketing 69 (2005) 114–129.

[14] S. C. H. Leung, W.-L. Ng, A stochastic programming model for production planning of perishable products with postponement, Production Planning & Control 18 (3) (2007) 190–202.

[15] N. J. Mehta, N. H. Shah, Time varying increasing demand inventory model for deteriorating items, Investigação Operacional 26 (1) (2005) 3–9.

[16] H. Waterer, Lot-sizing with inventory gains, Operations Research Letters 35 (6) (2007) 759–766.

[17] H. Stadtler, Mixed integer programming model formulations for dynamic multi-item multi-level capacitated lot-sizing, European Journal of Operational Research 94 (3) (1996) 561–581.
[18] J. Pahl, S. Voß, Discrete lot-sizing and scheduling including deterioration and perishability constraints, in: Advanced Manufacturing and Sustainable Logistics, no. 46 in Lecture Notes in Business Information Processing, Springer, Berlin Heidelberg, 2010, pp. 345–357.
[19] J. Pahl, S. Voß, D. L. Woodruff, Discrete lot-sizing and scheduling with sequence-dependent setup times and costs including deterioration and perishability constraints, in: Hawaiian International Conference on Systems Sciences (HICSS), 2011, pp. 1–10.
[20] G. R. Bitran, E. A. Haas, A. C. Hax, Hierarchical production planning: A two stage system, Operations Research 30 (1982) 232–251.
[21] M. Florian, J. K. Lenstra, A. H. G. R. Kan, Deterministic production planning: Algorithms and complexity, Management Science 26 (7) (1980) 669–679.
[22] X. Wang, D. Li, C. O'Brien, Optimisation of traceability and operations planning: an integrated model for perishable food production, International Journal of Production Research 47 (11) (2009) 2865–2886.
[23] M. Ferguson, M. E. Ketzenberg, Information sharing to improve retail product freshness of perishables, Production and Operations Management 15 (1) (2006) 57–73.
[24] S. Nahmias, D. Perry, W. Stadje, Perishable inventory systems with variable input and demand rates, Mathematical Methods of Operations Research 60 (1) (2004) 155–162.
[25] A. Förster, K. Haase, M. Tönnies, Ein modellgestützter Ansatz zur mittelfristigen Produktions- und Ablaufplanung für eine Brauerei, Zeitschrift für Betriebswirtschaft 76 (12) (2006) 1255–1274.
[26] R. K. Ahuja, W. Huang, H. E. Romeijn, D. Romero-Morales, A heuristic approach to the multi-period single-sourcing problem with production and inventory capacities and perishability constraints, INFORMS Journal on Computing 19 (1) (2007) 14–26.
[27] W. Elmaghraby, P. Keskinocak, Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions, Management Science 49 (10) (2003) 1287–1309.
[28] P.-C. Yang, H.-M. Wee, A single-vendor and multiple-buyers production-inventory policy for a deteriorating item, European Journal of Operational Research 143 (3) (2002) 570–581.
[29] P.-C. Yang, H.-M. Wee, An integrated multi-lot-size production inventory model for deteriorating items, Computers & Operations Research 30 (5) (2003) 671–682.
[30] H. Rau, M.-Y. Wu, H.-M. Wee, Deteriorating item inventory model with shortage due to supplier in an integrated supply chain, International Journal of Systems Science 35 (5) (2004) 293–303.
[31] C.-J. Chung, H.-M. Wee, Optimal replenishment policy for an integrated supplier-buyer deteriorating inventory model considering multiple JIT delivery and other cost functions, Asia-Pacific Journal of Operational Research 24 (1) (2007) 125–145.
[32] H. M. Wee, J. C. P. Yu, K. J. Wang, An integrated production-inventory model for deteriorating items with imperfect quality and shortage backordering considerations, in: Computational Science and Its Applications - ICCSA 2006, International Conference, Glasgow, UK, May 8-11, 2006, Proceedings, Part III, Vol. 3982, Springer, Heidelberg, 2006, pp. 885–897.
[33] H.-M. Wee, C.-J. Chung, Optimizing replenishment policy for an integrated production inventory deteriorating model considering green component-value design and remanufacturing, International Journal of Production Research 47 (5) (2009) 1343–1368.
[34] S. A. Ghasimi, R. Gohdsi, Improvement and solving the three new supply chain inventory control models for perishable items using just-in-time logistic, in: UKSim: 11th International Conference on Computer Modelling and Simulation, IEEE, 2009, pp. 279–286.
[35] C.-Y. Dye, L.-Y. Ouyang, T.-P. Hsieh, Inventory and pricing strategies for deteriorating items with shortages: A discounted cash flow approach, Computers & Industrial Engineering 52 (2007) 29–40.
[36] I. Moon, B. C. Giri, B. Ko, Economic order quantity models for ameliorating/deteriorating items under inflation and time discounting, European Journal of Operational Research 162 (3) (2005) 773–785.
[37] H.-M. Wee, S.-T. Lo, J. Yu, H. C. Chen, An inventory model for ameliorating and deteriorating items taking account of time value of money and finite planning horizon, International Journal of Systems Science 39 (8) (2008) 801–807.
[38] K.-L. Hou, An inventory model for deteriorating items with stock-dependent consumption rate and shortages under inflation and time discounting, European Journal of Operational Research 168 (2) (2006) 463–474.
[39] B. Zhang, X. Wang, Optimal policy and simgle algorithm for a deteriorated multi-item EOQ problem, American Journal of Operations Research 1 (2011) 46–50.
[40] L.-Y. Ouyang, C.-T. Yang, H.-F. Yen, Optimal order policy for deteriorating items in response to temporary price discount linked to order quantity, Tamkang Journal of Mathematics 40 (4) (2009) 383–400.
[41] C.-T. Chang, S. K. Goyal, J.-T. Teng, On "an EOQ model for perishable items under stock-dependent selling rate and time-dependent partial backlogging" by Dye and Ouyang, European Journal of Operational Research 174 (2)

(2006) 923–929.
[42] C.-Y. Dye, Joint pricing and ordering policy for a deteriorating inventory with partial backlogging, Omega 35 (2007) 184–189.
[43] A. K. Pal, A. K. Bhunia, R. N. Mukherjee, Optimal lot size model for deteriorating items with demand rate dependent on displayed stocklevel (DSL) and partial backordering, European Journal of Operational Research 175 (2006) 977–991.
[44] S. Papachristos, K. Skouri, An inventory model with deteriorating items, quantity discount, pricing and time-dependent partial backlogging, International Journal of Production Economics 83 (3) (2003) 247–256.
[45] S. K. Mondal, J. K. Ondal, J. K. Dey, M. Maiti, A single period inventory model of a deteriorating item sold from two shops with shortage via genetic algorithm, Yugoslav Journal of Operations Research 17 (1) (2007) 75–94.
[46] A. K. Bhunia, A. A. Shaikh, A deterministic model for deteriorating items with displayed inventory level dependent demand rate incorporating marketing decisions with transportation costs, International Journal of Industrial Engineering Computations 2 (2011) Online–First.
[47] Z. T. Balkhi, On the optimality of inventory models with deteriorating items for demand and on-hand inventory dependent production rate, Journal of Management Mathematics 15 (1) (2004) 67–86.
[48] Z. T. Balkhi, An optimal solution of a general lot size inventory model with deteriorated and imperfect products, taking into account inflation and time value of money, International Journal of Systems Science 35 (2) (2004) 87–96.
[49] F. Jolai, R. Tavakkoli-Moghaddam, M. Rabbani, M. R. Sadoughian, An economic production lot size model with deteriorating items, stock-dependent demand, inflation, and partial backlogging, Applied Mathematics and Computation 181 (2006) 380–389.
[50] M.-S. Chern, H.-L. Yang, J.-T. Teng, S. Papachristos, Partial backlogging inventory lot-size models for deteriorating items with fluctuating demand under inflation, European Journal of Operational Research 191 (1) (2008) 127–141.
[51] H. Sabahno, Optimal policy for a deteriorating inventory model with finite replenishment rate and with price dependent demand rate and cycle length dependent price, in: Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, San Francisco, USA, 2008, pp. 1053–1057.
[52] J.-J. Liao, An EOQ model with noninstantaneous receipt and exponentially deteriorating items under two-level trade credit, International Journal of Production Economics 113 (2) (2008) 852–861.
[53] C. Sugapriya, K. Jeyaraman, Determining a common production cycle time for an EPQ model with non instantaneous deteriorating items allowing price discount using permissible delay in payments, Journal of Engineering and Applied Sciences 3 (2) (2008) 26–30.
[54] L.-Y. Ouyang, J.-T. Teng, S. Goyal, C.-T. Yang, An economic order quantity model for deteriorating items with partially permissible delay in payments linked to order quantity, European Journal of Operational Research 194 (2) (2009) 418–431.
[55] H.-J. Chang, C.-Y. Dye, An inventory model for deteriorating items with partial backlogging and permissible delay in payments, International Journal of Systems Science 32 (3) (2001) 345–352.
[56] H.-J. Chang, C.-H. Hung, C.-Y. Dye, An inventory model for deteriorating items with linear trend demand under the condition of permissible delay in payments, Production Planning & Control 12 (3) (2001) 274–282.
[57] C.-T. Chang, Inventory models with stock-dependent demand and nonlinear holding costs for deteriorating items, Asia Pacific Journal of Operational Research 21 (4) (2004) 435–446.
[58] Y.-C. Tsao, G.-J. Sheen, Joint pricing and replenishment decisions for deteriorating items with lot-size and time-dependent purchasing cost under credit period, International Journal of Systems Science 38 (7) (2007) 549–561.
[59] Z. T. Balkhi, On the optimality of multi item integrated production inventory systems, in: Proceedings of the 3rd International Conference on Applied Mathematics, Simulation, Modelling (ASM'09), ASMCSS'09, World Scientific and Engineering Academy and Society (WSEAS), Vouliagmeni, Athens, Greece, 2009, pp. 143–154.
[60] Z. T. Balkhi, On the optimality of multi item integrated production inventory systems with variable parameters, in: Proceedings of the 2010 American conference on Applied mathematics, AMERICAN-MATH'10, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 2009, pp. 584–596.
[61] C.-Y. Dye, L.-Y. Ouyang, An EOQ model for perishable items under stock-dependent selling rate and time-dependent partial backlogging, European Journal of Operational Research 163 (3) (2005) 776–783.
[62] C.-Y. Dye, T.-P. Hsieh, L.-Y. Ouyang, Determining optimal selling price and lot size with a varying rate of deterioration and exponential partial backlogging, European Journal of Operational Research 181 (2007) 668–678.
[63] K.-S. Wu, L.-Y. Ouyang, C.-T. Yang, An optimal replenishment policy for non-instantaneous deteriorating items with stock-dependent demand and partial backlogging, International Journal of Production Economics 2 (2) (2006) 369–384.
[64] K.-S. Wu, EOQ inventory model for items with Weibull distribution deterioration, time-varying demand and partial backlogging., International Journal of Systems Science 33 (5) (2002) 323–329.
[65] S.-P. Wang, An inventory replenishment policy for deteriorating items with shortages and partial backlogging,

Computers & Operations Research 29 (14) (2002) 2043–2051.
[66] S.-P. Wang, Optimal lot-sizing policy for time-varying demand with shortages and deterioration, Journal of the Chinese Institute of Industrial Engineers 20 (5) (2003) 449–456.
[67] M.-S. Chern, Y.-L. Chan, J.-T. Teng, A comparison among various inventory shortage models for deteriorating items on the basis of maximizing profits, Asia Pacific Journal of Operational Research 22 (1) (2005) 121–134.
[68] K. Skouri, S. Papachristos, Four inventory models for deteriorating items with time varying demand and partial backlogging: A cost comparison, Optimal Control Applications and Methods 24 (6) (2003) 315–330.
[69] Z. T. Balkhi, L. Tadj, A generalized economic order quantity model with deteriorating items and time varying demand, deterioration, and costs, International Transactions in Operational Research 15 (4) (2008) 509–517.
[70] A. Roy, An inventory model for deteriorating items with price dependent demand and time-varying holding cost, Advanced Modeling and Optimization 10 (1) (2008) 25–37.
[71] B. C. Giri, A. K. Jalan, K. S. Chaudhuri, Economic order quantity model with Weibull deterioration distribution, shortage and ramp-type demand, International Journal of Systems Science 34 (4) (2003) 237–243.
[72] S. Roy, A. K. Bhunia, S. Mukhopadhyay, A genetic algorithmic approach on a deterministic inventory model for deteriorating items with shortages, Information Systems and Operational Research (INFOR) 43 (3) (2005) 271–282.
[73] K. Skouri, I. Konstantaras, S. Papachristos, I. Ganas, Inventory models with ramp type demand rate, partial backlogging and Weibull deterioration rate, European Journal of Operational Research 192 (1) (2009) 79–92.
[74] S. Panda, S. Senapati, M. Basu, Optimal replenishment policy for perishable seasonal products in a season with ramp-type time dependent demand, Computers & Industrial Engineering 54 (2) (2008) 301–314.
[75] Z. T. Balkhi, L. Benkherouf, On an inventory model for deteriorating items with stock dependent and time-varying demand rates, Computers & Operations Research 31 (2) (2004) 223–240.
[76] J.-T. Teng, M.-C. Cheng, L.-Y. Ouyang, An EOQ model for deteriorating items with power-form stock-dependent demand, Information and Management Sciences 16 (1) (2005) 1–16.
[77] S. K. Ghosh, K. S. Chaudhuri, An EOQ model with a quadratic demand, time-proportional deterioration and shortages in all cycles, International Journal of Systems Science 37 (10) (2006) 663–672.
[78] S. K. Ghosh, K. S. Chaudhuri, An order-level inventory model for a deteriorating item with Weibull distribution deterioration, time-quadratic demand and shortage, Advanced Modeling and Optimization 6 (1) (2004) 21–35.
[79] J. Li, T. C. E. Cheng, S. Wang, Analysis of postponement strategy for perishable items by EOQ-based models, International Journal of Production Economics 107 (2007) 31–38.
[80] J.-M. Chen, L.-T. Chen, Pricing and lot-sizing for a deteriorating item in a periodic review inventory system with shortages, Journal of the Operational Research Society 55 (2004) 892–901.
[81] N. H. Shah, Inventory model for deteriorating items and time value of money for a finite time horizon under the permissible delay in payments, International Journal of Systems Science 37 (1) (2006) 9–15.
[82] L.-Y. Ouyang, T.-P. Hsieh, C.-Y. Dye, H.-C. Chang, An inventory model for deteriorating items with stock-dependent demand under the conditions of inflation and time-value of money, The Engineering Economist 48 (1) (2003) 52–68.
[83] Y. Lin, C. Lin, Purchasing model for deteriorating items with time-varying demand under inflation and time discounting, The International Journal of Advanced Manufacturing Technology 27 (7) (2006) 816–823.
[84] H.-C. Liao, Y.-K. Chen, Optimal payment time for retailer's inventory system, International Journal of Systems Science 34 (4) (2003) 245–253.
[85] C.-Y. Dye, H.-J. Chang, A replenishment policy for deteriorating items with linear trend demand and shortages when payment periods are offered, Information and Management Science 14 (2) (2003) 31–45.
[86] C.-T. Chang, An EOQ model with deteriorating items under inflation when supplier credits linked to order quantity, International Journal of Production Economics 88 (3) (2004) 307–316.
[87] Y.-W. Zhou, H.-S. Lau, S.-L. Yang, A new variable production scheduling strategy for deteriorating items with time-varying demand and partial lost sales, Computers & Operations Research 30 (12) (2003) 1753–1776.
[88] J.-T. Teng, H.-L. Yang, Deterministic economic order quantity models with partial backlogging when demand and cost are fluctuating with time, Journal of the Operational Research Society 55 (2004) 495–503.
[89] C.-Y. Dye, H.-J. Chang, J.-T. Teng, A deteriorating inventory model with time-varying demand and shortage-dependent partial backlogging, European Journal of Operational Research 172 (2006) 417–429.
[90] H.-Y. Kang, Optimal replenishment policies for deteriorating control wafers inventory, The International Journal of Advanced Manufacturing Technology 35 (7) (2008) 736–744.
[91] Y. He, S.-Y. Wang, K. K. Lai, An optimal production-inventory model for deteriorating items with multiple-market demand, European Journal of Operational Research 203 (2010) 593–600.
[92] V. N. Hsu, An economic lot size model for perishable products with age-dependent inventory and backorder costs, IIE Transactions 35 (8) (2003) 775–780.
[93] Q. Bai, J. Xu, Y. Zhang, An approximation solution to the ELS model for perisable inventory with backlogging, in: The 7th International Symposium on Operations Research and Its Applications (ISORA '08), Lijiang, China,

2008, pp. 66–73.
[94] L. Y. Chu, V. N. Hsu, Z.-J. M. Shen, An economic lot-sizing problem with perishable inventory and economics of scale costs: Approximation solution and worst case analysis, Naval Research Logistics 52 (6) (2005) 536–548.
[95] B. Sivakumar, G. Arivarignan, An inventory system with postponed demands, Stochastic Analysis & Applications 26 (1) (2008) 84–97.
[96] S. Kalpakam, S. Shanthi, A continuous review perishable system with renewal demands, Annals of Operations Research 143 (1) (2006) 211–225.
[97] P. Manuel, B. Sivakumar, G. Arivarignan, Perishable inventory system with postponed demands and negative customers, Journal of Applied Mathematics and Decision Sciences (2007) 1–12.
[98] C. L. Williams, B. E. Patuwo, Analysis of the effect of various unit costs on the optimal incoming quantity in a perishable inventory model, European Journal of Operational Research 156 (1) (2004) 140–147.
[99] B. Sivakumar, G. Arivarignan, A perishable inventory system with service facilities and negative customers, Advanced Modeling and Optimization 7 (2) (2005) 193–210.
[100] B. Sivakumar, G. Arivarignan, A perishable inventory system at service facilities with negative customers, Information and Management Science 17 (2) (2006) 1–18.
[101] V. S. Yadavalli, C. van Schoor, J. J. Strasheim, S. Udayabaskaran, A single product perishing inventory model with demand interaction, ORiON, Journal of the Operations Research Society of South Africa (ORSSA) 20 (2) (2004) 109–124.
[102] G. C. Mahata, A. Goswami, An EOQ-model for deteriorating items under trade credit financing in the fuzzy sense, Production Planning & Control 18 (8) (2007) 681–692.
[103] L.-Y. Ouyang, K.-S. Wu, C.-T. Yang, Retailer's orderling policy for non-instantaneous deterioration items with quantity discount, stock-dependent demand and stochastic backorder rate, Journal of the Chinese Institute of Industrial Engineers 25 (1) (2008) 62–72.
[104] A. Roy, S. Kar, M. Maiti, Coordinated sale of differential units with promotional cost and units price through different shops in fuzzy environment, Advanced Modeling and Optimization 7 (1) (2005) 39–56.
[105] A. S. Lawrence, B. Sivakumar, G. Arivarignan, Perishable inventory system with random supply quantity and negative demands, Advanced Modeling and Optimization 8 (2) (2006) 151–168.
[106] B. Deniz, A. Scheller-Wolf, I. Karaesmen, Managing inventories of perishable goods: The effect of substitution, available at: http://faculty.fuqua.duke.edu/seminarscalendar/Scheller-WolfSeminar.pdf, last call August 2011 (2004).
[107] Z. Li, F. Tao, D. Sun, On jointly optimising the changes of seasonable goods and inventory replenishment, International Journal of Systems Science (2011) 1–13.
[108] Z. Lian, L. Liu, M. F. Neuts, A discrete-time model for common lifetime inventory systems, Mathematics of Operations Research 30 (3) (2005) 718–732.
[109] R. Kopach, B. Balcioglu, M. Carter, Tutorial on constructing a red blood cell inventory management system with two demand rates, European Journal of Operational Research 185 (2008) 1051–1059.
[110] Z. Liao, Q. Xie, Inventory management for depreciating items with stochastic demand, in: Proceedings of the International Conference on Services Systems and Services Management (ICSSSM '05), 2005, pp. 347–351.
[111] K. Maity, M. Maiti, A numerical approach to a multi-objective optimal inventory control problem for deteriorating multi-items under fuzzy inflation and discounting, Computers & Mathematics with Applications 55 (8) (2008) 1794–1807.
[112] A. Chande, S. Dhekane, N. Hemachandra, N. Rangaraj, Perishable inventory management and dynamic pricing using RFID technology, Sādhanā 30 (2-3) (2005) 445–462.
[113] X. Liu, O. Tang, P. Huang, Dynamic pricing and ordering decision for the perishable food of the supermarket using RFID technology, Asia Pacific Journal of Marketing and Logistics 20 (1) (2008) 7–22.
[114] X. Lu, J.-S. Song, K. Zhu, Analysis of perishable-inventory systems with censored demand data, Operations Research 56 (4) (2008) 1034–1038.
[115] R. A. Broeckmeulen, K. H. van Donselaar, A replenishment policy for a perishable inventory system based on estimated aging and retrival behavior, Internet Source: http://cms.ieis.tue.nl/Beta/Files/WorkingPapers/Beta_wp218.pdf, BETA Working Paper Nr. 218, last call August 2011 (2007).
[116] S. Minner, S. Transchel, Periodic review inventory-control for perishable products under service-level constraints, OR Spectrum 32 (4) (2010) 979–996.
[117] G. C. Lin, D. E. Kroll, C. J. Lin, Determining a common production cycle time for an economic lot scheduling problem with deteriorating items, European Journal of Operational Research 173 (2006) 669–682.
[118] K. Skouri, S. Papachristos, Optimal stopping and restarting production times for an EOQ model with deteriorationitems and time-dependent partial backlogging, International Journal of Production Economics 81-82 (1) (2003) 525–531.
[119] S. K. Goyal, B. C. Giri, The production-inventory problem of a product with time varying demand, production

and deterioration rates, European Journal of Operational Research 147 (3) (2003) 549–557.
[120] Z. T. Balkhi, The effects of learning on the optimal production lot size for deteriorating and partially backordered items with time varying demand and deterioration rates, Applied Mathematical Modelling 27 (10) (2003) 763–779.
[121] A. Alamri, Z. T. Balkhi, The effects of learning and forgetting on the optimal production lot size for deteriorating items with time varying demand and deterioration rates, International Journal of Production Economics 107 (2007) 125–138.
[122] L. Tadj, M. Bounkhel, Y. Benhadid, Optimal control of a production inventory system with deteriorating items, International Journal of Systems Science 37 (15) (2006) 1111–1121.
[123] S. Sana, K. S. Chaudhuri, On a volume flexible stock-dependent inventory model, Advanced Modeling and Optimization 5 (3) (2003) 197–210.
[124] J.-T. Teng, L.-Y. Ouyang, L.-H. Chen, A comparison between two pricing and lot-sizing models with partial backlogging and deteriorated items, International Journal of Production Economics 105 (1) (2007) 190–203.
[125] S. K. Manna, K. S. Chaudhuri, An EOQ model with ramp type demand rate, time dependent deterioration rate, unit production cost and shortages, European Journal of Operational Research 171 (2006) 557–566.
[126] S. Sana, K. Chaudhuri, On a volume flexible production policy for a deteriorating item with time-dependent demand and shortages, Advanced Modeling and Optimization 6 (1) (2004) 57–73.
[127] J.-M. Chen, L.-T. Chen, Market-driven production lot-size and scheduling with finite capactiy for a deteriorating item, Production Planning & Control 17 (1) (2006) 44–53.
[128] V. Hsu, Dynamic economic lot size model with perishable inventory, Management Science 46 (8) (2000) 1159–1169.
[129] M. L. Entrup, H. Günther, P. van Beek, M. Grunow, T. Seiler, Mixed-Integer linear programming approaches to shelf-life-integrated planning and scheduling in yoghurt production, International Journal of Production Research 43 (23) (2005) 5071–5100.
[130] P. Amorim, C. H. Antunes, B. Almada-Lobo, Multi-objective lot-sizing and scheduling dealing with perishability issues, Industrial & Engineering Chemistry Research 50 (6) (2011) 3371–3381.
[131] A. Al-Khedhairi, L. Tadj, Optimal control of a production inventory system with Weibull distributed deterioration, Applied Mathematical Sciences 1 (35) (2007) 1703–1714.
[132] K. A. Halim, B. C. Giri, K. S. Chaudhuri, Fuzzy EPQ models for an imperfect production system, International Journal of System Science 40 (1) (2009) 45–52.
[133] M. Bounkhel, L. Tadj, R. Hedjar, Nonlinear receding horizon control of production inventory systems with deteriorating item, Yugoslav Journal of Operations Research 18 (1) (2008) 37–45.
[134] N. K. Mahapatra, M. Maiti, Production-inventory model for a deteriorating item with imprecise preparation time for production in a finite time horizon, Asia Pacific Journal of Operational Research 23 (2) (2006) 171–192.
[135] R. Akkerman, D. P. van Donk, G. Gaalman, Influence of capacity- and time-constrained intermediate storage in two-stage food production systems, International Journal of Production Research 45 (13) (2007) 2955–2973.
[136] M. Kärkkäinen, Increasing efficiency in the supply chain for short schelf life goods using RFID tagging, International Journal of Retail & Distribution Management 31 (10) (2003) 529–536.
[137] R. Akkerman, P. Farahani, M. Grunow, Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges, OR Spectrum 32 (4) (2010) 863–904.
[138] B. Bilgen, H.-O. Günther, Integrated production and distribution planning in the fast moving consumer goods industry: a block planning application, OR Spectrum 32 (4) (2010) 927–955.
[139] T.-P. Hsieh, C.-Y. Dye, L.-Y. Ouyang, Determining optimal lot size for a two-warehouse system with deterioration and shortages using net present value, European Journal of Operational Research 191 (1) (2008) 182–192.
[140] C.-Y. Dye, L.-Y. Ouyang, T.-P. Hsieh, Deterministic inventory model for deteriorating items with capacity constraint and time-proportional backlogging rate, European Journal of Operational Research 178 (2007) 789–807.
[141] C.-C. Lee, Two-warehouse inventory model with deterioration under FIFO dispatching policy, European Journal of Operational Research 174 (2) (2006) 861–873.
[142] K.-J. Chung, T.-S. Huang, The optimal cycle time for deteriorating items with limited storage capacity under permissible delay in payments, Asia Pacific Journal of Operational Research 23 (3) (2006) 347–370.
[143] H.-L. Yang, Two-warehouse inventory models for deteriorating items with shortages under inflation, European Journal of Operational Research 157 (2) (2004) 344–356.
[144] P. Pal, C. B. Das, A. Panda, A. K. Bhunia, An application of real-coded genetic algorithm (for mixed integer non-linear programming in an optimal two-warehouse inventory policy for deteriorating items with a linear trend in demand and a fixed planning horizon), International Journal of Computer Mathematics 82 (2005) 163–175.
[145] H.-S. Hwang, A stochastic set-covering location model for both ameliorating and deteriorating items, Computers & Industrial Engineering 46 (2) (2004) 313–319.
[146] S. D. P. Flapper, J. C. Fransoo, R. A. C. M. Broekmeulen, K. Inderfurth, Planning and control of rework in the process industries: a review, Production Planning & Control 13 (2002) 26–34.

[147] S. K. Srivastava, Green supply-chain management: A state-of-the-art literature review, International Journal of Management Reviews 9 (1) (2007) 53–80.
[148] K. Inderfurth, G. Lindner, N. P. Rachaniotis, Lot sizing in a production system with rework and production deterioration, International Journal of Production Research 43 (7) (2005) 1355–1374.
[149] R. H. Teunter, S. D. P. Flapper, Lot-sizing for a single-stage single-product production system with rework of perishable product defectives, OR Spectrum 25 (1) (2003) 85–96.
[150] A. J. Clegg, D. J. Williams, R. Uzsoy, Production planning for companies with remanufacturing capability, Proceedings of the 1995 IEEE International Symposium onElectronics and the Environment (ISEE) (1995) 186–191.
[151] D. R. Guide, M. S. Spencer, Rough-cut capacity planning for remanufacturing firms, Production Planning & Control 8 (3) (1997) 237–244.
[152] R. Dekker, K. Inderfurth, L. N. van Wassenhove, M. Fleischmann, Reverse Logistics: A Quantitative Approach, 1st Edition, Springer, 2004.
[153] G. Fandel, J. Reese, Reverse Logistics, no. 3 in Zeitschrift für Betriebswirtschaftslehre, Special Issue, Gabler, 2005.
[154] R. S. Tibben-Lembke, Life after death: Reverse logistics and the product life cycle, International Journal of Physical Distribution & Logistics Management 32 (3) (2002) 223 – 244.
[155] K. Inderfurth, M. Y. Kovalyov, C. T. D. Ng, F. Werner, Cost minimizing scheduling of work and rework processes on a single facility under deterioration of reworkables, International Journal of Production Economics 105 (1) (2007) 345–356.
[156] K. Inderfurth, A. Janiak, M. Y. Kovalyov, F. Werner, Batching work and rework processes with limited deterioration of reworkables, Computers & Operations Research 33 (6) (2006) 1595–1605.
[157] F. S. Al-Anzi, A. Allahverdi, M. Y. Kovalyov, Batching deteriorating items with applications in computer communication and reverse logistics, European Journal of Operational Research 182 (2007) 1002–1011.
[158] Y. J. Kim, H. Hwang, C. S. Ko, Inventory level of recycled material for deteriorating items under lifo policy, in: P. U. V. Kachitvichyanukul, U. Purintrapiban (Ed.), Proceedings of the 2005 International Conference on Simulation and Modelling, 2005, pp. 1–6.
[159] H. Krishnan, R. A. Winter, On the role of revenue-sharing contracts in supply chains, Operations Research Letters 39 (2011) 28–31.
[160] H. Rau, M.-Y. Wu, H.-M. Wee, Integrated inventory model for deteriorating items under a multi-echelon supply chain environment, International Journal of Production Economics 86 (2) (2003) 155–168.
[161] C. Lin, Y. Lin, A joint EOQ model for supplier and retailer with deteriorating items, Asia Pacific Journal of Operational Research 21 (2) (2004) 163–178.
[162] J.-M. Chen, T.-H. Chen, Effects of joint replenishment and channel coordination for managing multiple deteriorating products in a supply chain, Journal of the Operational Research Society 56 (10) (2005) 1224–1234.
[163] C. Lin, Y. Lin, A cooperative inventory policy with deteriorating items for a two-echelon model, European Journal of Operational Research 178 (1) (2007) 92–111.
[164] N. Mahapatra, K. Das, M. Maiti, Production-inventory policy for a deteriorating item with a single vendor-buyer system, Optimization and Engineering 8 (4) (2007) 431–448.
[165] C. Wang, Supply chain decision model for deteriorating goods with stock-dependent demand rate, in: Wireless Communications, International Conference on Networking and Mobile Computing, 2007, pp. 3758–3761.
[166] N. H. Shah, A. S. Gor, H. M. Wee, Optimal joint vendor-buyer inventory strategy for deteriorating items with salvage value, in: J. Harland, P. Manyem (Eds.), Fourteenth Computing: The Australasian Theory Symposium (CATS 2008), Vol. 77 of CRPIT, ACS, Wollongong, NSW, Australia, 2008, pp. 63–66.
[167] S. D. Ekşioğlu, M. Jin, Cross-facility production and transportation planning problem with perishable inventory, in: M. Gavrilova (Ed.), Computational Science and Its Applications (ICCSA), Lecture notes in computer science (LNCS) Nr. 3982, Springer, Heidelberg, 2006, pp. 708–717, workshop on Optimization: Theories and Applications (OTA 2006).
[168] D. Shen, K. K. Lai, S. C. H. Leung, L. Liang, Modelling and analysis of inventory replenishment for perishable acgricultural products with buyer-seller collaboration, International Journal of System Science iFirst (2010) 1–11.
[169] K. H. Hahn, H. Hwang, S. W. Shinn, A returns policy for distribution channel coordination of perishable items, European Journal of Operational Research 152 (3) (2004) 770–780.
[170] K. Das, T. K. Roy, M. Maiti, Buyer-seller fuzzy inventory model for a deteriorating item with discount, International Journal of Systems Science 35 (8) (2004) 457–466.

Production Planning with Load Dependent Lead Times and Sustainability Aspects

Institute of Information Systems
Department of Business Sciences
University of Hamburg

Technical Report

Julia Pahl

(22. March 2012)

Part I

Production Planning with Load Dependent Lead Times and Sustainability Aspects: Technical Report

Contents

I Production Planning with Load Dependent Lead Times and Sustainability Aspects: Technical Report **VI**

Table of Contents **VII**

List of Figures **X**

List of Tables **XII**

Acronyms **XV**

List of Symbols **XVII**

1 Introduction and Outline **2**
1.1 Introduction 2
1.2 Outline 4

2 Basic Concepts and Definitions **5**

3 Production and Supply Chain Planning **6**
3.1 Production Planning 6
3.1.1 Organizational Issues 7
3.1.2 Time Horizons and Hierarchical Planning 8
3.1.2.1 Strategic Planning Level 9
3.1.2.2 Tactical Planning Level 9
3.1.2.3 Operational Planning Level 10
3.1.3 The Modeling of Time 11
3.2 Supply Chain Management 12

4 Depreciation and Lifetime Constraints **14**
4.1 Characteristics and Definition 14
4.1.1 Discussion on Depreciation Characteristics 15
4.1.2 Definitions of Depreciation 16
4.2 Mathematical Issues 16

5 Lead Times and Capacity Management **19**
5.1 Lead Time Definitions and Related Considerations 19
5.2 Planning Circularities 20
5.2.1 Fixed Planned Lead Times 22
5.2.2 Issues Regarding Production Mix 23
5.2.3 Issues Regarding Depreciation 23
5.3 Capacity Management and Utilization 23
5.4 Load Dependent Lead Times 26

6 Green Operations Management and Sustainability 29
6.1 Strategies and Areas of GSCM 30
6.2 Sustainability 31
6.2.1 Strategies for Sustainability 33
6.2.2 The Formulation of Sustainability Objectives 34
6.3 Rework and Closed-Loop Supply Chain Management 36
6.3.1 Classification and Organization 36
6.3.2 Literature Review of Related Models 37
6.3.3 Mathematical Issues Regarding Rework and Remanufacturing . . . 38

7 Models Incorporating Load-Dependent Lead-Times 42
7.1 Production-Performance Models 42
7.1.1 The Model of Lautenschläger 42
7.1.1.1 Assumptions of the Basic Model of Lautenschläger . . . 43
7.1.1.2 Mathematical Formulation 44
7.1.1.3 Determination of Parameters for the Production Forward-Shifting Curve 44
7.1.1.4 Integration of Overtime 47
7.1.2 LDLT by Voß and Woodruff (2004) 48
7.2 Release-Production Models 50
7.2.1 The Concept of Clearing Functions 51
7.2.2 Lot Sizing Model of Karmarkar (1989) 52
7.2.3 Model of Asmundsson et al. (2003) 54
7.3 Discussion on Presented Approaches 56

II Combining LDLT and Lot Sizing with Sustainability Aspects 58

8 Production Performance Models with Perishability 59
8.1 CLSP-P with LDLT 59
8.2 Including Overtime and Minimal Lot Sizes 63
8.3 Inclusion of Setup Carry Overs 65
8.4 Load-Dependent Capacity Restrictions and Production Smoothing 68
8.4.1 Literature Review on Production Smoothing 69
8.4.2 Integration of Production Smoothing via PSCU 71
8.4.3 Production Smoothing via Targeting Utilization Sectors 75
8.5 Some Excursion: Forward Shifting or Extended CLSP-P 78

9 Remanufacturing and Rework 83
9.1 Remanufacturing of Returns 83
9.2 Integration of In-Line Rework 88

10 Numerical Study and Results 90
10.1 Analysis of Resulting Costs and Optimality Gaps 92
10.1.1 Test Instances with Return Rates of 94% 92
10.1.2 Test Instances with Return Rates of 6% 96
10.2 Influences of Lifetime Constraints 97
10.3 Influence of Minimal Lot Size Constraints 100
10.4 Influences of the Load on the Optimality Gap 101
10.5 Influences of the TBO 103
10.6 W-SRT Regarding for Pairwise Parameter Combinations 104

III Conclusions **107**

11 Conclusions and Further Research Directions **108**

IV Appendix **VI**

12 Data for Testing Behavior and Mechanisms of Models: Results **VII**

13 Results for Return Rates 94% **VIII**

14 Results for Return Rates 6% **XVIII**

References **XXIV**

List of Figures

3.1 Interaction of management and performance systems as a regulating system; see Reusch (2006) and Zäpfel (1989). 7
3.2 Time horizons of planning; see Domschke et al. (1997). 8
3.3 Time horizons, planning tasks and company functions including remanufacturing/recycling; modified from Schneeweiß (1999). 9
3.4 Discrete and continuous representations of time. 11
3.5 Decisions regarding timely position of lot sizes; see also Meyr (1999). . . . 13

4.1 Courses of perishability and deterioration; see Pahl (2011). 16

5.1 Different process times and denotation of compositions regarding lead times and other process time related terms; based on Obenauf (2011) 21
5.2 Extended planning circularities and lead time syndrome. 24
5.3 Visualization of the different capacity definitions in a sense of capability. . . 25
5.4 Relation between WIP and resource utilization and its integration in clearing functions; see Pahl et al. (2005). 26
5.5 Simple queuing model with one workstation (server); modified from Hopp and Spearman (2001). 27

6.1 Fields of green supply chain management; modified from Klütsch (2009), based on Srivastava (2007). 30
6.2 Comparison of reverse logistics and green logistic; see Rogers and Tibben-Lembke (2001). 31
6.3 Production system with deterioration and rework; based on Sarkis et al. (2004), Hervani et al. (2005); see also Pahl et al. (2007b). 32
6.4 Three pillars of sustainability; see Garetti and Taisch (2012). 32
6.5 Trade-offs and synergies regarding different production and delivery options; modified from Akkerman et al. (2009). 34
6.6 In-Line production and rework system including external returns; see, e.g., Yum and McDowell (1987), Richter and Sombrutzki (2000), Richter and Weber (2001) for similar illustrations. 38

7.1 Example of a forward shifting curve; modified from Lautenschläger (1999). 45
7.2 Parts and amounts of production in mode $m = 1$ and $m = 2$ in % for $\alpha = 1,2$ and $\beta = 1,0$ with parameters $e = 0,8$ and $b = 0,6$ for utilization values $30 - 100\%$. 46
7.3 Extension of the FS curve to overtime. 48
7.4 Course of WIP and FGI inventory. 51
7.5 Clearing function of Karmarkar (1989) for different k_2-values. 52
7.6 Clearing function of Srinivasan et al. (1988) for different k_2-values. 53

8.1 Ideal and actual cumulative production quantities; see Yavuz and Akçali (2007). 70
8.2 Approximation of load-dependent resource utilization costs for use in piece-wise linear cost function term in objective function. 75
8.3 Total costs of the CLSP-P-LDCap-5. 80
8.4 Total costs of the CLSP-P-5. 81
8.5 Differences in total costs of the CLSP-P-5 versus the CLSP-P-5 for the 5-product case. 81

9.1 Depiction of the remanufacturing/rework/production process applied on the CLSP-P-R-6. 86

10.1 Lowest overall costs regarding test instances with return rates of 94%. . . . 93
10.2 Highest overall costs regarding test instances with return rates of 94%. . . . 95

List of Tables

7.1 Parameter combinations to determine FS curves 47

8.1 List of time-infinite parameters 61
8.2 Numerical example to show load effects for varying v-values 62
8.3 Numerical example to show load effects including overtime 65
8.4 Numerical example to show load effects including overtime and variations of product lifetimes employing CLSP-P-LDCap-2 66
8.5 Numerical example to show load effects employing PSCU in Model CLSP-P-LDCap-4 73
8.6 Numerical example to show load effects employing PSCU in Model CLSP-P-LDCap-4 and varying v-values 74
8.7 Numerical example to show load effects including overtime and variations of product lifetimes employing CLSP-P-LDCap-5 78
8.8 List of time-infinite parameters for the comparison of CLSP-P-LDCap-5 and CLSP-P-5 79
8.9 Cost parameters of utilization intervals 80
8.10 Differences of overall costs comparing the CLSP-P-LDCap-5 and the CLSP-P-5 80
8.11 Comparison of results of the CLSP-P-LDCap-5 and the CLSP-P-5 regarding their overall costs with $st_i = 0$ 82

10.1 List of time-infinite remanufacturing/rework parameters for all products i . 90
10.2 Cost parameters of utilization intervals 90
10.3 Test instances with lowest costs and return rates of 94% 92
10.4 Test instances with highest costs and return rates of 94% 92
10.5 Test instances with lowest overall costs with detailed cost categories for test instances with return rates of 94% 94
10.6 Test instances with highest overall costs with detailed cost categories for test instances with return rates of 94% 94
10.7 Test instances with lowest optimality gaps and return rates of 94% 95
10.8 Test instances with highest optimality gaps and return rates of 94% 95
10.9 Test instances with lowest costs and return rates of 6% 96
10.10 Test instances with highest costs and return rates of 6% 96
10.11 Test instances with highest optimality gap and return rates of 6% 97
10.12 Test instances with lowest overall costs with detailed cost categories for test instances with return rates of 6% 98
10.13 Test instances with highest overall costs with detailed cost categories for test instances with return rates of 6% 98
10.14 Mean optimality gap of different lifetimes for test instances with return rates of 94% 99

10.15 Analysis regarding the variations of lifetimes regarding test instances with return rates of 94% . . . 99
10.16 W-SRT regarding Lifetimes and their influence on the optimality gap for test instances with return rates of 94% . . . 99
10.17 W-SRT regarding Lifetimes and their influence on the optimality gap for test instances with return rates of 6% . . . 99
10.18 Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 50$ and $x_i^{\min} = 100$ . . . 100
10.19 Mean optimality gap of different minimal lot sizes for test instances with return rates of 94% . . . 101
10.20 Analysis regarding the variations of minimal lot sizes for test instances with return rates of 94% . . . 101
10.21 W-SRT analyzing the influences of the minimal lot size constraint on the optimality gap retarding test intances with 94% return rates . . . 101
10.22 W-SRT analyzing the influences of the minimal lot size constraint on the optimality gap retarding test intances with 6% return rates . . . 102
10.23 Average gap in % regarding test instances with return rates of 94% sorted by minimal lot sizes, lifetimes and loads . . . 102
10.24 Influences of the load on the optimality gap retarding test intances with 94% return rates . . . 103
10.25 Influences of the load on the optimality gap retarding test intances with 6% return rates . . . 103
10.26 Analysis regarding the variations of *TBO* for test instances with return rates of 94% . . . 103
10.27 Analysis regarding the variations of TBOs for test instances with return rates of 94% . . . 104
10.28 Influences of the TBO on the optimality gap retarding test intances with 94% return rates . . . 104
10.29 Influences of the TBO on the optimality gap retarding test intances with 6% return rates . . . 104
10.30 Average gap in % regarding test instances with return rates of 94% sorted by TBOs, lifetimes and minimal lot sizes . . . 105

12.1 Numerical example to show load effects: demand data . . . VII

13.1 W-SRT on pairwise lifetime and minimal lot sizes and their influence on the optimality gap for test instances with return rates of 94% for *p*-values until pair (10;50) . . . IX
13.2 W-SRT on pairwise lifetime and minimal lot sizes and their influence on the optimality gap for test instances with return rates of 94% for *p*-values from pair (10;50) . . . X
13.3 Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 0$ and $x_i^{\min} = 20$ . . . XI
13.4 Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 50$ and $x_i^{\min} = 100$ . . . XII
13.5 W-SRT on pairwise TBO and lifetime and their influence on the optimality gap for test instances with return rates of 94% until (2;5) . . . XIII
13.6 W-SRT on pairwise TBO and lifetime and their influence on the optimality gap for test instances with return rates of 94% from (2;4) . . . XIV
13.7 W-SRT on pairwise TBO and minimal lot size and their influence on the optimality gap for test instances with return rates of 94% . . . XV

13.8 W-SRT on pairwise Load and minimal lot size and their influence on the optimality gap for test instances with return rates of 94% until values (0.8;50)XVI
13.9 W-SRT on pairwise Load and minimal lot size and their influence on the optimality gap for test instances with return rates of 94% from values (0.8;50)XVII

14.1 W-SRT on pairwise Lifetime and minimal lot size and their influence on the optimality gap for test instances with return rates of 6% XIX
14.2 W-SRT on pairwise Lifetime and minimal lot size and their influence on the optimality gap for test instances with return rates of 6% XX
14.3 W-SRT on pairwise TBO and Lifetime and their influence on the optimality gap for test instances with return rates of 6% XXI
14.4 W-SRT on pairwise TBO and Lifetime and their influence on the optimality gap for test instances with return rates of 6% XXII
14.5 W-SRT on pairwise TBO and minimal lot size and their influence on the optimality gap for test instances with return rates of 6% XXIII

Acronyms

APS	advanced planning system
AtO	assemble to order
BOM	bill of materials
CF	clearing function
CLSP	capacitated lot sizing problem
CLSP-P	capacitated lot sizing problem with perishability constraints
CSLP	continuous setup and lot sizing problem
EFQM	European Foundation of Quality Management
ELSP	economic lot scheduling problems
EMS	environmental management system
EOQ	economic order quantity
EPQ	economic production quantity
ERP	enterprise resource planning
FGI	finished goods inventory
FIFO	first-in-first-out
FS	forward shifting
GLSP	general lot sizing and scheduling problem
GSCM	green supply chain management
i.i.d.	independent and identically distributed
I&L	inventory and lot size model
IT	information technology
JIT	just-in-time
LDLT	load-dependent lead time
LP	linear programming
MIP	mixed integer program
MLCLSP	multi-level capacitated lot sizing model

MLCLSP-FS	multi-level capacitated lot sizing model with forward shifting times
MPS	master planning schedule
mrp	material requirements planning
MRP II	manufacturing resources planning
MtO	make-to-order
MtS	make-to-stock
NFP	network flow problem
OVR	output rate variation
PLSP	proportional lot sizing and scheduling problem
PP	production-performance model
PRV	product rate variation
PSCU	planned smoothed capacity utilization
QU	quantity unit
RFID	radio frequency identification
ROI	return on investment
RP	release and production model
SC	supply chain
SCM	supply chain management
SDST	sequence-dependent setup time
SKU	stock-keeping unit
SOS	special ordered set
SOS1	special ordered set type 1
TBO	time between orders
TU	time unit
W-SRT	Wilcoxon signed rank test
WIP	work in process
WW	Wagner-Whitin

List of Symbols

Constants

ac_i Marginal costs that apply for using an alternative *stock-keeping unit (SKU) i*, first used in eq. (7.24) 49

a_{ij} Number of units of item i required to produce one unit of the immediate successor item j, first used in eq. (7.8) 44

α_w X-axis intercept of the *forward shifting (FS)* time curve of resource w, first used in eq. (7.8) 44

α_i^R Fraction of item i that returns from customers, first used in eq. (9.18) 85

α^Θ Percentage of inventory lost due to perishability, first used in eq. (4.2) 17

$A_I(p,s)$ Transfer time from two workstations where $A_I(p,s)$ denotes the inflow from a workstation p to a workstation s, first used in eq. (7.61) 55

$A_O(p,s)$ Transfer time from two workstations where $A_O(p,s)$ denotes the outflow from a workstation p to a workstation s, first used in eq. (7.61) 55

$A(t)$ Number of jobs arriving during a time interval $(0,t]$, first used in eq. (5.2) . . . 27

bc_i Tardiness or backorder cost factor for SKU i, first used in eq. (7.24) 49

β_w Slope of the FS time curve of resource w, first used in eq. (7.8) 44

BP_{ir} A list indexed by R utilization breakpoints for SKU i, first used in eq. (7.21) . 49

b_w Unitless parameter that denotes the percentage of utilization at which forward shifting of products in production mode $m = 2$ begins, first used in eq. (7.8) . 45

C Cycle time at a workstation, determined by $C = st + x/\dot{p}$ following Karmarkar (1989) where st is the setup time and $\dot{p}$ the nominal production rate, first used in eq. (7.44) 53

c Individual lines of an approximation for a *clearing function (CF)*, first used in eq. (7.61) 55

c_w The number of workstations in the system, first used in eq. (5.7) 27

d^k Cumulative demand, i.e., $\sum_{t=1}^{T}\sum_{i=1}^{N} d_{it}$, first used in eq. (8.21) 64

$D(t)$ Demand in period t, first used in eq. (4.1) 16

$d(t)$ Number of jobs departing during a time interval $(0,t]$, first used in eq. (5.3) . . 27

e_w Unitless parameter that denotes the amount of items that are shifted forward at an utilization of 100%, first used in eq. (7.8) 45

$\Gamma_{i\gamma t}$ Tansfer or transportation distance denoted in time units, first used in eq. (7.56) . . 55

G_t External returns assumed to be known in period t, first used in eq. (6.8) 40

h_i^e Echelon inventory holding cost parameter, first used in eq. (7.8) 44

hw_i Work in process holding costs for a SKU i, first used in eq. (7.24) 49

L_{ir} Estimated lead time value for SKU i corresponding to breakpoint r, first used in eq. (7.20) 48

λ Arrival rate of jobs in the production system or at a workstation frequently assumed as a stochastic input parameter that has an *independent and identically distributed (i.i.d.)* distribution; defined in *quantity unit (QU)*/*time unit (TU)*, first used in eq. (5.1) 26

$\overline{\lambda}$ Average job arrival rate, first used in eq. (5.12) 28

$\overline{\mu}$ Average service rate, first used in eq. (5.12) 28

$\overline{p}$ Average job processing or residence time in [TU], first used in eq. (5.1) 26

$P(\cdot)$ Probability value, first used in eq. (5.0) 20

$\dot{p}$ Nominal production rate, i.e., the maximal value that is theoretically achievable in case of 100% efficiency or, in other words, if not disruptions takes place, first used in eq. (7.44) 53

pc Processing cost factor for regular production, first used in eq. (6.8) 40

pcr Processing cost factor for rework processing, first used in eq. (6.8) 40

pcu_b^f Fixed cost term of interval/sector b of utilization, first used in eq. (8.36) 76

pcu_b^f Variable cost term of interval/sector b of utilization, first used in eq. (8.36) . . 76

ϕ_i^r Returned item disposal of item i in period t, first used in eq. (9.37) 87

psc_t Cost factor of planned/smoothed capacity utilization *planned smoothed capacity utilization (PSCU)*, first used in eq. (8.30) 71

r Parameter denoting how much rework is faster than regular processing. It is required that $r > 0$. Values higher than $r = 1$ express that rework employs more time than regular production, first used in eq. (9.0) 83

rc Release cost factor, first used in eq. (7.50) 54

rsc Rework setup cost factor, first used in eq. (6.8) 40

R_t Amount of releases in period t, first used in eq. (7.50) 54

Θ_{it}^r Lifetime of returned items i in inventory in time period t, first used in eq. (9.37) . 87

$tc_{i\gamma t}$ Transportation cost factor, first used in eq. (7.56) 55

Θ Shrinkage factor expressing perishability, first used in eq. (4.1) 16

Θ_i^L Lifetime of item i, first used in eq. (4.4) 17

Θ_t^L Time variable length of item lifetime, first used in eq. (4.4) 17

V Large number, first used in eq. (7.8) 44

ϑ Backordering cost factor, first used in eq. (7.50) 54

w_i Critical resource or bottleneck of the production system dependent on SKU i, first used in eq. (7.22) 49

ξ_{iw} Resource capacity consumption factor producing product i on production entity w, first used in eq. (7.8) 44

Indices

A_i Set of alternatives or substitutes of SKUs i with $a \in A_i$, first used in eq. (7.37) 49

b Set of breakpoints $b = 1, \ldots, B$ for piecewise linear terms, first used in eq. (8.31) 75

w Index denoting machines or production resources with $w = 1, \ldots, W$, first used in eq. (7.8) 44

i Index denoting products with $i = 1, \ldots, N$, first used in eq. (7.0) 43

r Index denoting numbers of lead time breakpoints for a SKU i with $r = 1, \ldots, R$, first used in eq. (7.20) 48

$t = 1, \ldots, T$ Time periods; synonmys are *macro periods* or *big time buckets*, first used in eq. (4.1) 16

Variables

B_t Amount of backorders in period t, first used in eq. (7.50) 54

δ Job departure rate, first used in eq. (5.6) 27

I_{at}^+ Positive inventory of SKU i in time period t, first used in eq. (7.24) 49

I_{at}^- Negative inventory denoting shortages for SKU i in time period t, first used in eq. (7.24) 49

I_{it}^{Dr} Returned item disposal of product i in period t, first used in eq. (9.37) 87

I_t^D Spoiled inventory due to deterioration in time period t, first used in eq. (4.4) . 17

I_{it}^r Rework inventory of product i in period t, first used in eq. (9.37) 87

I_t^r Inventory holding of rework/remanufacturing in period t, first used in eq. (6.8) . . 40

$I(t)$ Inventory holding in time period t used in static model approaches, first used in eq. (4.1) 16

$\overline{L}$ Average lead time, first used in eq. (5.13) 28

$1/\mu$ Average service rate, first used in eq. (5.7) 27

O_{wt} Overtime of production resource w in period t, first used in eq. (7.16) 47

p Variable or operational production rate (*synonym*: processing time of a product), determined by $p = x/c$ where x is the actual amount of products and c is the cycle time, first used in eq. (7.44) 53

$\dot{p}$ Nominal production processing rate in a sense of *best* processing rate that does not include stochastic (negative) influences. Consequently, it is the rate at which a machine operates at its less resource-consuming level, first used in eq. (7.45) 53

p_{irt} Variable defined as *special ordered set (SOS)* that becomes 1 for a SKU i of one breakpoint r in time period t, so that the corresponding estimated lead time value for SKU i of breakpoint r is chosen, first used in eq. (7.20) 48

PSCU_t Planned smoothed capacity utilization that is integrated as a quadratic term in the objective function of models including production smoothing, first used in eq. (8.30) 71

ρ_w Utilization of production resource w, first used in eq. (7.9) 45

S_n Service rate of n jobs during a given time interval, first used in eq. (5.8) 27

$\overline{W}$ Average number of jobs waiting in a queue in front of a production system or a workstation, first used in eq. (5.1) 26

W Number of jobs waiting in a queue in front of a production system or a workstation, first used in eq. (5.1) 26

x_{it}^m Production mode with set M with $m = 1,2$ thus two modes where mode 1 produces product i in period t that is also available in period t and mode 2 denotes production of product i in period $t-1$ and is available in period t, first used in eq. (7.8) 44

x_{it}^r Rework of returned items (remanufacturing) of product i in period t, first used in eq. (9.37) 87

x_t Amount of production in time period t, first used in eq. (4.4) 17

x_{it}^{total} Total amount of production of product i in period t, first used in eq. (7.8) 44

y_{bt} Binary variable for interval selection regarding load-dependent utilization costs with breakpoints $b = 1,\ldots,B$ for every time period t, first used in eq. (8.31) . 75

y_{it}^r Rework setup variable for product i in period t, first used in eq. (9.37) 87

y_t^r Setup variable for rework lot sizes in period t, first used in eq. (6.8) 40

Z_{it} Fraction of maximum possible output of product i in period t that is defined by the CF allocated to production i in period t. These fractions are required to sum up to 1, first used in eq. (7.67) 55

z_{it} Setup state defined as a binary variable, i.e., $z_{it} \in \{0,1\}$, first used in eq. (8.23) .. 66

$\widehat{\delta}_{it}(x)$ Approximation to resource utilization, first used in eq. (7.22) 49

Chapter 1

Introduction and Outline

1.1 Introduction

Dynamics and globalization of markets intensify the pressure on companies' performances in terms of price, costs and time based factors such as flexibility or responsiveness to customers especially in case of variable demand patterns. Charney (1991) observed that "how to do more" was the question in the 1960ies. "How to do it cheaper" became central in the 1970ies. In the 1990ies importance rose on "how to do it quicker." This has induced considerable interest in understanding lead times and their role in production planning as well as supply chains until present. An increasing consciousness for the environment and quality of products give us reason to pursuit sustainability not only regarding products and related (raw) materials, but also and especially regarding overall business processes as a main goal of nowadays. While the focus of managing supply chains has undergone a drastic change as a result of improved *information technology (IT)*, integrated and coordinated production planning remains a critical issue.

Constraints on lifetimes of products force organizations to carefully plan their production in cooperation with their supply chain partners in order to prevent decreased quality or wastages of production parts and products. This gains importance, because waiting times imply longer lead times charging increasingly the system with *work in process* (*work in process (WIP)*). Longer waiting/lead times lead to increased quality loss due to depreciation. In the worst case, parts and products cannot be used for their original purpose, anymore, and have to be discarded or reworked if technically possible. This is notably disquieting in the food industry where significant losses occur during handling, processing, and their distribution; see Gustavsson et al. (2011). Results of a study conducted for the International Congress "Save Food!" state that 1.3 billion tons of food per year are wasted at a global scale which is around one-third of food produced worldwide for human consumption. This is particularly valid for developing countries where more than 40% of food losses are due to processing whereas total losses in industrialized countries are as high, but with more than 40% mainly occurring at the retailer and consumer level; see Gustavsson et al. (2011). Despite increased supply chain costs due to losses in terms of depreciated goods, increased energy consumption and related emissions of CO_2, and other forms of pollution are a troubling aspect. Certainly, unsatisfied customers waiting for their products or being concerned about quality is mainly a problem in industrialized countries.

Classical planning models treat lead times as static, exogenous input data, but in most situations the output of a planning model implies capacity utilization which, in turn, influences lead times. It has been shown empirically that lead times increase exponentially with the workload measured by WIP as well as resource utilization long before 100% is reached. This can be a reason for significant differences in planned and realized lead times. Abstrac-

tion from such nonlinearities frequently takes place mainly in favour of complexity reduction for models that might be used in practice by implementing them into standard software such as *advanced planning system (APS)* or *enterprise resource planning (ERP)*-systems. As a consequence, variabilities in lead times may become significant with increasing WIP and system utilization. Thus, the frequently employed target to maximize system utilization should be critically reviewed. Additionally, empirical evidence shows that "flustering" due to the load of a production entity takes place, so that outputs are not produced at a specific, constant rate, but at a fraction of this rate that decreases with increasing WIP; see Elmaghraby (2011).

There are only a few models that respect lead times dependent on resource utilization (*load-dependent lead times (LDLTs)*) as endogenous to the model, so that lead times are a result of the plan. The same is valid for models linking order releases, planning and capacity decisions to lead times taking into account the system workload, batching/lot sizing and sequencing decisions. Therefore, one important goal of this work is to formulate optimization models for aggregate production planning especially focussing on lot sizing decisions that account for capacity utilization. Reducing lot sizes leads to lower inventory holding costs, but also to increased setup costs together with increased capacity occupation due to setups. On the other hand, increasing lot sizes force items to wait which increases WIP. In case that items have limited lifetimes and their quality degrades while waiting, items might pass their valuable lifetime and need to be discarded or reworked. In case that rework is executed on production resources employed for regular manufacturing, resource utilization is further raised due to setups and processing of rework batches, so that a negative process (*downward spiral*) on lead times and item quality might take place. "Variations in production mix play havoc with the utilization of the available capacity of individual shops" as Elmaghraby (2011) states in his analysis on production capacity and its measurement(s). Limits of lifetimes of parts and products in the production process increase the complexity of planning which is especially true in case that setup times and costs are sequence dependent and rework processing becomes longer with increased waiting times. Scheduling decisions can be further important regarding lead times, system workload, and depreciation, so that some scheduling information are already regarded on the tactical planning level.

Aspects of sustainability encompass every action from wastage/disposal reduction also in terms of energy usage that implies reductions of CO_2-levels to long-term recovery actions of exploited lands and environments. Reductions of CO_2-levels can be achieved by adapting means and/or frequencies of shipments, "greening" transportation technologies, e.g., the implementation of filters for carbon dioxides in vehicles/motors. Decreasing transportation distances may decrease lead times and risks of congestion on the way. Reducing wastages is further possible by choosing packaging of items in relation to the transportation mode or vehicle while maintaining high quality of products. For instance, the quality of chilled food meals can be enhanced or maintained by transportation in less environmentally friendly polystyrene boxes while shipping, because they maintain low temperatures and thus freshness of food better than, e.g., cardboards; see, e.g., Akkerman et al. (2009). Consequently, decreasing transportation distances (also named "food miles"; see van der Vorst et al. (2009)) and times render the use of environmentally friendly cardboards practicable while maintaining quality. Besides, delivery planning and thus batching/lot sizing of *finished goods inventory (FGI)* transportation has the same trade-offs as aggregated production planning: the reduction of shipment frequences and thus planning greater *transportation lot sizes* force final products to wait for transportation which increases lead times while quality decreases. Jaegler and Burlat (2012) state that transport and storage of items cause 50% of environmental impacts; see also Cholette and Venkat (2009) for an analysis of wine distribution and Rizet and Keïta (2005) for a study on yoghurt and jeans. Further greening actions include the minimization of usage/wastage regarding raw/supporting materials in

the production process, e.g., water. The substitution of environmental unfriendly materials or pollutive wastages that are by-products of the production process, is another important aspect of sustainability.

In this work, discrete dynamic models are developed and analyzed that are formulated as *mixed integer programs (MIPs)* for production planning including LDLTs together with lifetime constraints on products, and rework options of items that have passed their lifetime limits. The concentration is on classical *inventory and lot size model (I&L)* also named *production-performance model (PP)* formulations. Aspects of sustainability are integrated, such as minimization of wastage or disposals as well as in-line rework of internally passed as well as externally returned items. It is shown that lifetime constraints are hard constraints of models. Furthermore, overtime and minimal lot sizes constraints are integrated. The developed models aim at avoiding disposals of passed items in the first place. As shown elsewhere, passed items in such models are mainly due to minimal lot size constraints that require to produce more than a certain number of items in each regarded period than demand might be, so that items that can only be stored for a short time interval or not at all cannot be used for demand satisfaction and are immediately discarded. Therefore, minimal lot size constraints are added in order to study internally passed items, their rework, and their influence on LDLT.

1.2 Outline

This study is organized as follows: in Part 2 basic concepts on the treated subject and definitions are provided. Chapter 3 gives an overview on production and supply chain planning with special attention to mathematical modeling. Depreciation effects of parts and products are discussed in Chapter 4 that further gives definitions as well as an overview on mathematical modeling methods. Lead time and capacity management are discussed in Chapter 5 providing some thoughts on planning circularities, lead time considerations together with capacity and utilization management lead time LDLT and their determination and implications to (production) planning. In Chapter 6, the literature regarding *green supply chain management (GSCM)* and sustainability is reviewed as well as rework and closed-loop *supply chain management (SCM)*. Mathematical models including LDLT are discussed in Chapter 7. Models of the PP type are analyzed in this regard as well as *work-in-process* (WIP)-based or *release and production models (RPs)* in order to compare their application to further extensions regarding lifetime constraints, rework, and other sustainability-related issues. Part II contains the development of modeling approaches that include discussed aspects. PP-based models are developed and explored. General extensions include, e.g., minimal lot sizes, overtime, production smoothing to account for LDLT issues, and setup carry overs. Models are tested extensively in Chapter 10. In Part III, conclusions are drawn and ideas for further research proposed.

Chapter 2

Basic Concepts and Definitions

Chapter 3

Production and Supply Chain Planning

The framework of the analysis is presented in the following chapters. definitions on planning regarding production and supply chains are provided. Special attention is on organizational issues of production planning including time horizons, hierarchical organization of planning, and mathematical issues of optimization models.

3.1 Production Planning

The framework for production is generally given by product designs and their physical production process from an engineering point of view. Planning is the anticipation and evaluation of alternatives for future actions. It provides answers on what to do and when to do it in order to achieve a desired result expressed by, prevalently, economical objectives. The latter are frequently formulated include high service levels, short lead times, high capacity utilization, low inventory levels, and/or a high flexibility to changes; see Hopp and Spearman (2001), Monden (2011). Here, planning is defined as the anticipation of future alternatives; see Domschke et al. (1997). These are evaluated by means of their contribution to the achievement of predefined targets. As such, planning includes the preparation and support of decision making using abstraction as a method to analyze and understand a given (more or less) complex problem. Figure 3.3 shows the plan as the outcome of the planning process that defines input information for more detailed planning on hierarchically lower levels. While executing the plan, continuous monitoring and re-evaluation takes place employing deviation analysis, so that corrective actions can be undertaken if deviations exceed a certain corridor.

Production is the creation of products or services that satisfy human/social needs and supports the quality of human life. Therefore, it contains the combination of elemental production factors such as work, (raw) materials, and production resources including planning and organization as dispositive factors in order to add value and satisfy demand. Garetti and Taisch (2012) distinguish between manufacturing and production where the latter is limited to “the process of making goods.” Instead, manufacturing comprises all industrial activities from the supplier via the manufacturer to the customer. They further divide manufacturing into discrete manufacturing, process industries, and services. In this work, the terms is used synonymously regarding all activities including supportive actions as manufacturing or production. Additionally, *supply chain management* (SCM) is refered to when regarding two or more companies although abstraction is often made to denote SCM regarding the planning, control, and coordination of multiple production entities that might also be multiple department or, more generally, “planning functions” of the same company.

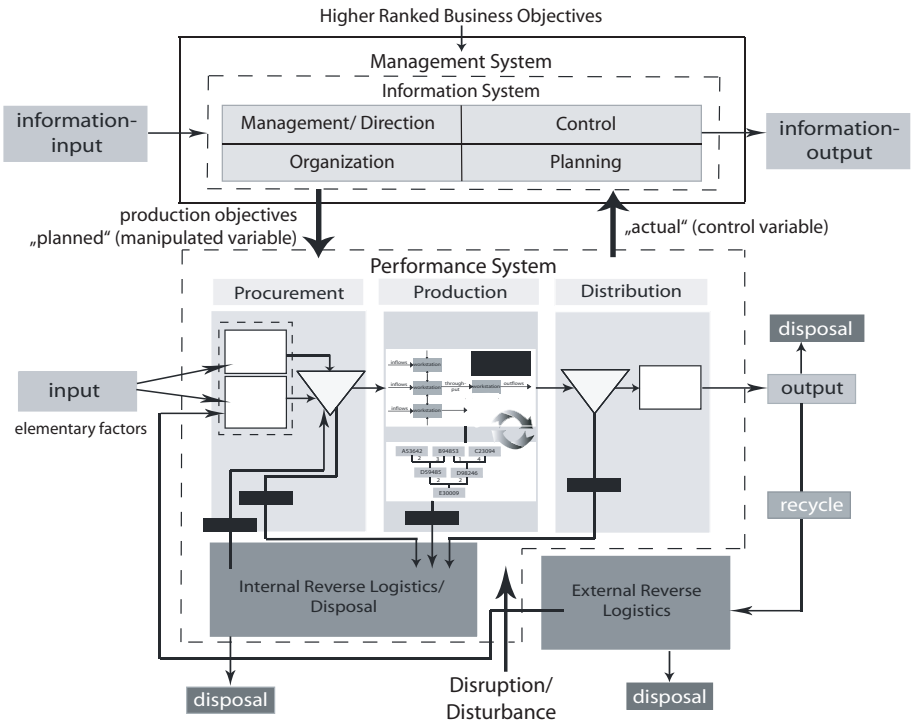

Figure 3.1: Interaction of management and performance systems as a regulating system; see Reusch (2006) and Zäpfel (1989).

Different activities in the production process are necessary that contain direct or supportive functions depicted in Figure 3.1 where, e.g., the *management system* includes the top management that provides visions and missions of the overall company. The *organization* function defines hierarchies, processes, and, in general, the structure of work including IT. The *planning and control* function provides guideline information to the performance system. The *performance system* is composed by the procurement of input factors such as raw materials or work, the physical production process, and the distribution of FGI. The *internal reverse logistic* system deals with rework of internally passed or damaged parts and products whereas the external *reverse logistic system* regards recycling/remanufacturing flows from external customers. All these parts of the performance system are highly dynamic research fields on their own. Here, the focus is on the planning function of the management system regarding the physical production process and the internal reverse logistics system especially emphasizing on lead time and workload management.

3.1.1 Organizational Issues

Lead times and the routing of products within the production system are highly dependent on the design and organization of the production process that, in turn, depend on the related industry. Process designs/organizations range from completely automated systems to job shops. Classification schemes and their categorization have been proposed, e.g., by Adam (1998), Domschke et al. (1997) including the degree of automation and repetition, the organizational form and arrangement of production facilities, the type of production structure and/or requirement explosion, and market-driven characteristics. The degree of repetition in manufacturing implies job shop, serial or batch production, and mass production. For instance, *make-to-order* and *make-to-stock* productions are characteristical types that are usually attributed to market-driven characteristics. Mass production is an example for standardization where highly standardized products are made on a “make-to-stock” basis.

Economies of scale and learning effects permit to produce on an efficient basis. Significant automation through conveyor belts speeds up production. Generally, such production systems are precision-synchronized, so that disruptions stop the whole manufacturing process. Besides, this type of production is capital intensive, especially regarding process or product changes and flexibility requirements. This is a problem not only due to intensification of customer needs for individual or tailored products require differentiation and customization of products and services. The concept of mass customization tries to circumvent these problems: it combines the production of standardized products and postponed individualization to final assembly through modularization. In other words: standardized products are customized in the end of the assembly by combining them regarding customer requirements and present orders; see also Caridi and Cavalieri (2004). This is somewhat a fictitious individualization, because products are customized within the limits of standardized parts. Nevertheless, it is a solution and a compromise to decrease coordination, related complexity and to mitigate effects of demand variabilities.

Organizational forms have their strengths and weaknesses. For instance, flow shops allow for advantages like short lead times, small transportation distances, and constant capacity utilization due to assembly lines with conveyor belts, predefined processes, and frequencies for product transportation. Drawbacks of such designs include high investment costs and capital commitment in equipment, little flexibility to changes in products or processes and high susceptibility to disturbances. This is not the case with job shops that allow for flexible reaction to changes and disturbances by variable machine setups or routings. Nevertheless, they are prevalent to longer transportation times and distances, higher inventory, and variable or unbalanced resource utilization that influence lead times due to the lack of fixed production processes and routines. Pure versions of these organizational forms exist, but their combinations to hybrid forms are more often applied to exploit their advantages and reduce their shortcomings.

3.1.2 Time Horizons and Hierarchical Planning

The planning and control function that has been introduced in Section 3.1.1 can be segregated into *strategic*, *tactical*, and *operational* time levels; see Figure 3.2. Some authors abstract from the tactical level allocating its tasks to the operational level; see, e.g., Günther and Tempelmeier (2000). These planning levels are discussed in more details in the following.

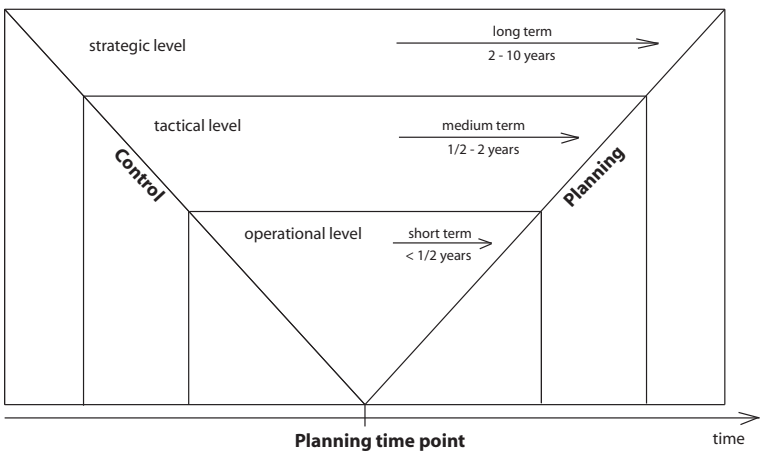

Figure 3.2: Time horizons of planning; see Domschke et al. (1997).

3.1.2.1 Strategic Planning Level

The *strategic planning level* includes long term decisions for, among others, the production program and allocation of resources to production. These decisions involve the top management level of a company. Key performance indicators include the rentability or the *return on investment (ROI)*; see also Figure 3.3. Questions that are to be answered at this level are mainly what to produce, where to produce, and how to produce. The resulting production program includes an approximate idea of production amounts that encompasses the choice of production resources and technologies together with the selection of manufacturing and warehouse locations. Possibilities of recycling and remanufacturing are also determined on this level, because certain product structures, designs and materials facilitate the recycling process in terms of time and costs. For that reason, facilities for disassembly are regarded on this planning level. Remanufacturing and rework decisions imply choices on *in-line* versus *off-line* remanufacturing/rework where in-line remanufacturing describes rework on the same production entities as regular production and off-line rework the use of different production entities; see also Section 6.3 for more details. As depicted in Figure 3.2, the horizon regarding strategic decisions ranges from two up to ten years.

Decisions are significantly influenced by company politics and markets. Classical rough strategic company objectives include *cost or price leadership* or *differentiation*; see Porter (1985, 1986) for a detailed discussion on company strategies. Key performance indicators on this level are rentabilities, e.g., ROI. Further activities assigned to different functions include, e.g., forecasts on product life cycles and market analysis for current or planned products.

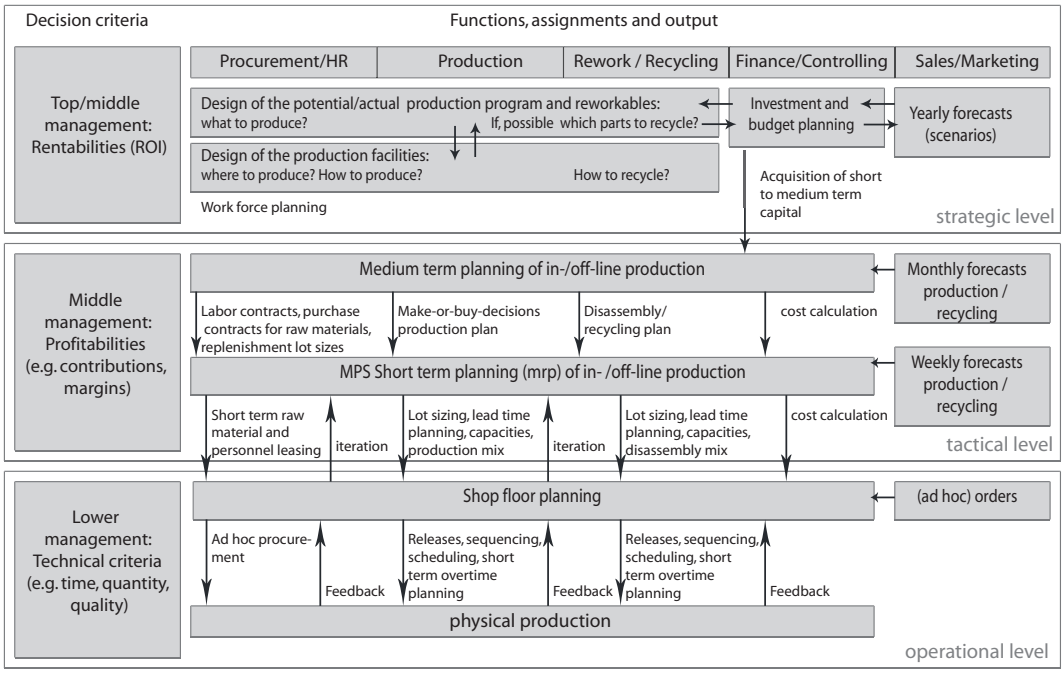

Figure 3.3: Time horizons, planning tasks and company functions including remanufacturing/recycling; modified from Schneeweiß (1999).

3.1.2.2 Tactical Planning Level

The *medium or tactical level* embraces decisions with a time range from half a year up to two years; see Figure 3.2. On this level, plans are made to implement decisions taken

on the strategic level, which means that a detailed medium term production program including product types, amounts and qualities based on strategic assessments is established. This process might include capacity decisions and/or potential adjustments examining the profitability of in-house production or external procurement, named *make-or-buy*-decisions or *vertical integration*. Some authors assume resource capacity to be fixed at the tactical level; see, e.g., Woodruff and Voß (2004), so that the objective of tactical planning is to best use limited resources and inventory. In case that make-or-buy decisions are included on the tactical level, the determination of global raw material supplies would be included here. In case that demand requires higher capacities than those given by the regular amount, a difference is made between fixed company-internal capacities, named *overtime*, and externally procured capacity extensions, named *subcontracting* or *external procurement*. Monthly forecasts support related decision making also regarding recycling and remanufacturing flows. Key performance indicators for the middle management/tactical level are profitabilities, e.g., contributions of products to overall rentability or margins. At this level, products and production resources (synonym *workstations*) are aggregated regarding similar characteristics to product or workstation types or families/groups; see also Missbauer and Uzsoy (2011). Further decisions regard the time horizon and related time units. The outcome of these planning steps is the *master planning schedule (MPS)* which encompasses lot sizing, due dates, lead times, production mix, and capacity decisions. This is also referred to as *aggregate production planning* in the literature. Frequently, the goal of aggregate production planning is the minimization of total production costs encompassing inventory holding costs, fixed and variable production costs, overtime, and backlogging. A review of these models with special emphasis on depreciation is given in Pahl (2011) or Pahl and Voß (2011). Details on necessary raw material supplies is given by *material requirements planning (mrp)*[1] where supplies are based on the quantities established in the MPS and determined by their *bill of materials (BOM)* which is the composition list of raw materials and parts together with the quantities that are needed to produce one final product; see, e.g., Hopp and Spearman (2001). Literature on planning tools and methods like mrp, *manufacturing resources planning (MRP II)*, ERP, APS etc. is substantial; see, e.g., Stadtler et al. (2008) for an overview and research directions. McKay (2011) and Ovacik (2011) provide recent thoughts on historical evolution and future developments of such tools.

3.1.2.3 Operational Planning Level

The *operational level* implements the plan of the higher planning levels. Its assignment comprises the planning and execution of the most efficient deployment of production resources regarding the objectives defined on the strategic and tactical level. Production scheduling takes place in the middle between production planning and real-time dispatching including product or job assembly; see Maimon et al. (1998). In fact, the decision horizon includes daily and/or weekly planning of production amounts and short term procurement of materials as well as the production control. The latter encompasses detailed scheduling, sequencing and order releases, monitoring, validation, and feedback of production tasks regarding demand and due dates, quality and costs. A *schedule* is defined according to Pinedo (2002) as a set of start and end times for setup operations and production of jobs or item. A *sequence* is denoted as an ordering of jobs. Missbauer and Uzsoy (2011) go further and state that the link or *essential interface* between the tactical and operational level is provided by order releases. Assumptions on lead times regarding the operational execution of production are taken on the tactical level to determine production plans. This implies assumptions on order releases frequently using the concept of constant lead times; see Missbauer and Uzsoy

[1] See Voß and Woodruff (2006) for definitions and abbreviation issues.

(2011). The constant lead time based on an order release plan must reflect and anticipate the behavior and possible events on the operational level in order to predict and account for its consequences already on the tactical level. There exist a great body of literature regarding workload control and performance efficiency through better order release mechanisms; see, e.g., Missbauer and Uzsoy (2011), Pahl et al. (2007a) and the references therein. Problems arising from using estimates or fixed lead times are subject to an in-depth discussion on lead times and planning circularities in Section 5.

Key performance indicators on the operational stage contain technical criteria providing data on times, quality, and quantity. The use of *radio frequency identification (RFID)* can give valuable support providing current (practical) data on inventory levels, delivery times, or whereabouts of items in transit or transportation. This can increase the product visibility, process transparency, and opportunities for integrated and automated data capture; see Chande et al. (2005), Liu et al. (2008).

3.1.3 The Modeling of Time

Continuous modeling of production planning decisions allows to study complex planning situations including uncertainties. If changing input parameters are regarded, modeling extends to continuous dynamic settings mainly by iteration methods to determine changes in plans in the long run; see Pahl and Voß (2011). Related classical models especially regarding lot sizing and production decisions on an aggregated basis are, e.g., the *economic order quantity (EOQ)*, *economic lot scheduling problems (ELSP)*, or, in the stochastic case, the multi-period newsvendor (*flowergirl*) problem; see, e.g., Pahl and Voß (2011) for an overview of static/dynamic continuous/discrete modeling including depreciation.

Dynamic processes in mathematical models are frequently reflected by a discrete representation of time: the time horizon is divided into a number of time intervals of – generally but not necessarily – uniform length, also named “time buckets.” Planning decisions are associated to these intervals or time blocks; see Figure 3.4. Such a presentation of the time

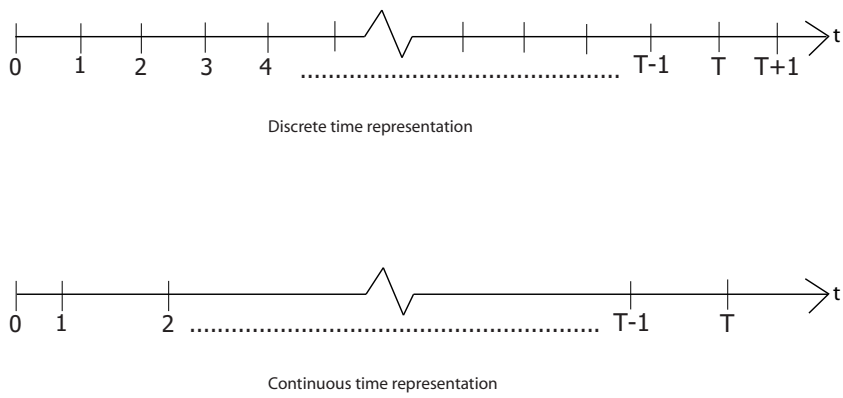

Figure 3.4: Discrete and continuous representations of time.

horizon bears some drawbacks, especially when dealing with practical problems of considerable size, because an increased number of time buckets may lead to large combinatorial problems implying increased complexity requiring significant computational calculation times. This is surely one of the reasons why related discrete dynamic models formulated as MIPs have limited practical application and research concentrates on the development of efficient (meta) heuristics; see also Floudas and Lin (2005). Nevertheless, commercial solvers[2] for optimization problems together with enhanced computational power regarding

[2]For instance, Xpress-IVE^{TM}, ILOG/CPLEX^{TM}, or Gurobi^{TM}are frequently employed solvers.

processor performance and memory evolve fast, so that larger and more difficult test instances are solvable in more and more acceptable calculation times. As a result, decision regarding the right planning horizon and number of observed time periods is important, but not only from a computational point of view. Activities need to be placed at a specific instance of these time intervals that can be crucial for the resulting plan in terms of accuracy and costs. For instance, a rough presentation of time might lead to undesired effects on decision variables such as increased production and inventory holding. On the other hand, attributing rather short time intervals to time buckets significantly increases their number and thus complexity. Besides, the decision of the granularity of the time scale has implications on the regarded events. If a short time horizon and related short time buckets are assumed, products' lifetimes may be longer and thus do not have an effect. Consequently, a model including deterioration and/or perishability should have a time horizon that is greater than the lifetimes of items, so that depreciation becomes an issue and, consequently, a binding constraint; see also Pahl and Voß (2011). On the other hand, the decision on the planning horizon should be chosen according to the assumed demand forecast accuracy and quality that may decrease the farther future periods are taken into account; see Silver et al. (1998).

Discrete dynamic time models are generally divided into *big bucket* and *small bucket* models where big bucket models contain rather long time periods also named *macro* periods in which several different products can be set up for or produced. Small bucket models have rather short time periods (*micro* periods), so that startups, switch-offs and/or change-overs can be taken into account. Micro periods are generally used to integrate detailed information about the shop floor in order to determine accurate and feasible plans. A further distinction can be made for small bucket models that allow for one setup per period and those allowing for more than one; see, e.g., Staggemeier and Clark (2001). Models exist that integrate macro as well as micro periods. In aggregated production and lot sizing planning these aim at including some scheduling information in order to provide more consistent plans to the operational level. Such models are named *hybrid* models presenting a meta level between aggregated and operational planning. For instance, Figure 3.5 describes time structures and related "positions" of lot sizes: they can be fixed on the outer grid or macro periods which is true for big bucket models. Small bucket models divide macro periods in micro periods and, consequently, lot sizes are fixed according to the micro period time grid. They may also be freely determined in the third case disregarding macro as well as micro period grid limits; see, e.g., Fleischmann and Meyr (1997), Meyr (1999). Detailed scheduling information determining due dates applying priority orders on jobs are regarded at later planning steps. Hybrid models are subject to dynamic research activities, see, e.g., Briskorn (2006), Haase and Kimms (2000), Karimi et al. (2006), Porkka et al. (2003), Quadt and Kuhn (2009), Tempelmeier and Buschkühl (2009).

3.2 Supply Chain Management

Extending the perspective to multiple companies leads to issues of SCM. Networks of companies that are linked to each other through up- and/or downstream flows of materials, financial resources, and information with the aim of producing value-added products and services for final customers are considered as *supply chains (SCs)*.

SCM emerged from business practice mainly focussing on logistics two decades ago and still is one of the global operations strategies for achieving organizational competitiveness in the 21st century; see Gunasekaran and Ngai (2004), Li and Wang (2007). Due to its emergence, SCM encompass various aspects and business functions (see Ayers (2000)) that is reflected in heterogenious definitions. For instance, Simchi-Levi et al. (2002) define SCM as "a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses,

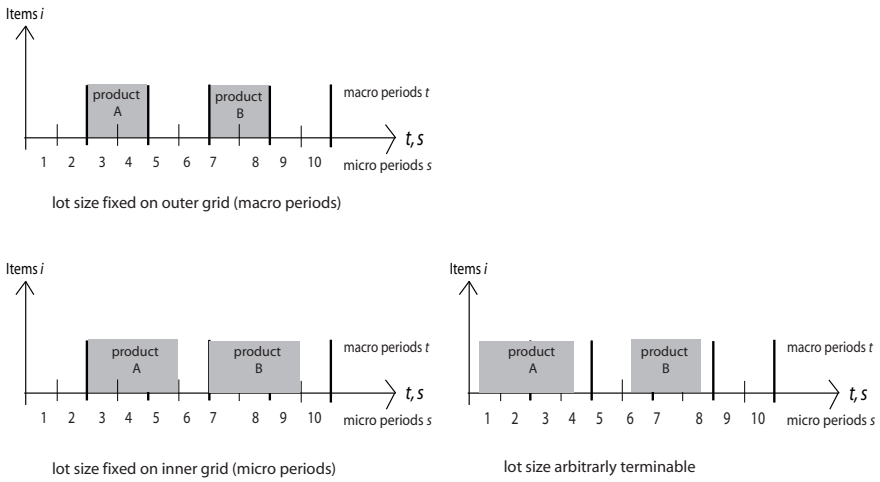

Figure 3.5: Decisions regarding timely position of lot sizes; see also Meyr (1999).

and stores, so that merchandise is produced and distributed at the right quantities, to the right locations at the right time in order to minimize system wide costs while satisfying service level requirements.” Gunasekaran and Ngai (2004) put it succinctly by stating that “SCM is based on the integration of all activities that add value to customers starting from product design to delivery" or “from the source of supply to the point of consumption” (see also Li and Wang (2007)). As a result, SCM is a broad topic not easy to grasp. It has a wide range of influence that extends from organizational structures, e.g., cooperations, over IT questions to selections and implementations of software applications. The perspective on multiple companies can be relaxed in order to regard different departments of the same company that act sometimes with different objectives that may be conflicting. The focus here is on production planning that assumes deterministic demand thus involving more than one department, i.e., marketing and sales and production planning. Striktly speaking, this relates to pure production planning, but in the relaxed sense involves concepts of SCM.

Chapter 4

Depreciation and Lifetime Constraints

The interest regarding deterioration and perishability comes into prominance with the modeling of inventor in blood bank management and the distribution of blood from transfusion centers to hospitals where a decent amount of literature exists; see, e.g., Pierskalla and Roach (1972), Jennings (1973), Frankfurter et al. (1974), Chazan and Gal (1977), Dumas and Rabinowitz (1977), Cohen and Pekelman (1978), Brodheim and Prastacos (1979), Kaspi and Perry (1984), Perry (1999), Perry and Stadje (1999, 2001), Deniz et al. (2004). Overviews on the subject are provided by Naddor (1966), Prastacos (1984), Pierskalla (2004). Later, the research focus shifted also to other products, e.g., chemicals, fashion clothes, technical components, newspapers, or food. In accordance, the attention changed to a more general view on how to include deterioration, perishability, and lifetime constraints into models for inventory management. The major part of approaches treat deterioration and perishability as a kind of general shrinkage of inventory that happens independently of planning or demand functions. Furthermore, most mathematical approaches incorporating deterioration or perishability effects are time-independent (static) deterministic inventory control models studying replenishment strategies and extending them to take into account marketing related issues, e.g., pricing/discounting, time value of money/inflation, permissible delay in payments. An in-depth discussion on depreciation is given in Pahl (2011), Pahl and Voß (2011). For the sake of self-completeness, important results are presented regarding characteristics of depreciation and product lifetimes focussing on deterministic modeling due to the focus of this work.

4.1 Characteristics and Definition

Most products deteriorate within a certain interval. If the observed time horizon is chosen to be sufficiently short, there may be no need to regard such product characteristics, because values loss does not take place; see also Section 3.1.3. Nevertheless, in practice and dependening on the product, the time horizon for aggregated planning is long enough, so that the influence of deterioration starts to play an important role. Deterioration and lifetime restrictions are hard constraints on inventory holding that render the planning process complicated. Restrictions are imposed to produce ahead of demand and sales realizations in order to hedge against uncertainties, e.g., employing over-production, exploit volume-discounts, or own production setup sequence advantages. For instance, in case that manufacturing lead times are significantly longer than replenishment lead times, *make-to-order (MtO)* manufacturing is not applicable, so that production plans rely on demand forecasts; see, e.g., Darlington and Rahimifard (2006).

4.1.1 Discussion on Depreciation Characteristics

Deterioration of products includes the process of decay, damage, or spoilage of items due to various reasons, so that they cannot be employed for their original purpose. Generally, they go through a change, physically or due to demand, so that they use their utility while stored; see Darlington and Rahimifard (2006), Ferguson and Koenigsberg (2007), Shah et al. (2005), Zhao (2007), Pahl (2011).

Perishable goods are regarded as items with fixed, maximum lifetimes, that are named *best-before date* in practice. This applies for products that become obsolete at some point in time, because of various reasons or external factors, e.g., laws and regulations that predetermine their shelflives; see Boukas and Liu (2001). Products including services as a main component incur similar problems, e.g., airline seats or hotel rooms, because they are subject to highly price-dependent demand; see Ferguson et al. (2007). Companies selling these kinds of products set multiple prices and related quantities on the time before their expiration. After expiration they receive zero revenue; for references on revenue or yield management see Cattani et al. (2007), Chiang et al. (2007), Petrick et al. (2010). Therefore, a classification basis is the *utility* of the item.

Two sub categories are distinguished where the first includes products whose *functionalities* deteriorate over time, e.g., fruits, vegetables or milk. A flight seat or hotel room might be considered in this category as the service cannot be provided anymore after a certain time, e.g., the airplane lifts off or the hotel room is taken. The second encompasses those products whose functionality does not degrade, but *customers' perceived utility* and thus their demand deteriorates over time, e.g., fashion clothes, high technology products with a short life cycle, or items whose information content deteriorates, e.g., newspapers. For example, it is empirically shown that the willingness to pay for a product continuously decreases with its perceived actuality that is similar to defining deterioration as a process of continuously reduced usefulness of the comodity to the customer; see Tsiros and Heilman (2005). Other products encounter value loss related perishability induced by demand that drops when a new version or generation is introduced accompanied by significant downmarks; see Leung and Ng (2007). This is valid, e.g., for computers, mobile phones, high fashion wear, Christmas gifts, or calendars. Such value loss is indirectly integrated into mathematical models by making assumptions on products' demand patterns, e.g., time-dependent demand that is used to describe sales in different product lifetime phases in the market; see Mehta and Shah (2005). For example, newly introduced products generally have an increasing demand rate at the beginning of their product life cycle which comes along with a growing market. Decreasing demand rates are usually a sign of a product at the end of its life cycle. Exponential decreasing demand is experienced by products where a new version or competitive product comes into the market or a change in customer's preferences takes place; see Su and Lin (2001), Chu and Chen (2002), Lin et al. (2000). Moreover, ramp type demand rates are regarded for new brands of consumer goods coming in the market; see, e.g., Giri et al. (2003), Manna and Chaudhuri (2006). As a result, depreciation is characterized regarding different aspects:

1. physical degradation or depletion (e.g., fresh food, gasoline),
2. loss of efficacy or functional loss (e.g., medicine), and
3. perceived value loss (e.g., fashion clothes, high technology)

where the first two are difficult to be exactly distinguished. A further distinction can be made by the products' lifetime that can be *time-independent* or *time-dependent*. The first case implies a pre-specified lifetime. The utility of these products is given during a specific time period. The second case expresses deterioration over time, e.g., to model seasonality.

The majority of authors working in the field of deterioration and perishability use the words interchangeably; see Pahl (2011).

4.1.2 Definitions of Depreciation

A clear distinction is made between perishable goods that cannot be used anymore and lose all their utility at once after a certain point in time whereas items that lose their utility gradually are regarded as being subject to deterioration; see Pahl (2011).

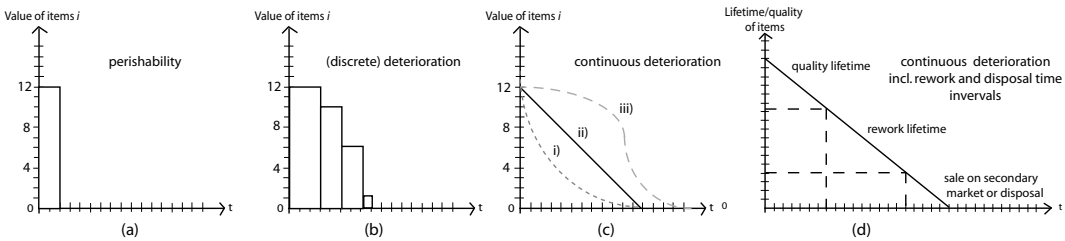

Figure 4.1: Courses of perishability and deterioration; see Pahl (2011).

Figure 4.1 depicts three functional relationships of item quality where the graph (a) shows the course of perishability, (b) the discrete and variational courses of deterioration and (c) possible courses of continuous deterioration where i) is nonlinear decreasing, ii) linear decreasing, and iii) arbitrary decreasing. The graph in (d) gives three possible time intervals regarding the quality decrease of items that can be reworked: the first interval presents regular quality of items until the phase where items can be reworked, and the phase where items have passed their lifetime and might be sold on a secondary market or need to be disposed; see Pahl (2011).

The definitions of deterioration, perishability, and lifetime constraints remain somewhat fuzzy as it clearly depends on the physical state/fitness, behavior of items over time, and on the production planner or the quality controller who has to decide whether to employ items or components having passed a certain age. Here, the focus is on the integration of functional deterioration, perishability, and lifetime constraints in contrast to value loss due to customer demand variations that is expressed via demand patterns.

4.2 Mathematical Issues

Generally, there are different methods to integrate deterioration, perishability, and lifetime constraints into mathematical models depending on the one hand on the intention of the model approach, e.g., general study of depreciation effects on replenishment policies or avoidance of wastages thus integration of lifetime constraints. On the other hand, their integration depends on the model type including the treatment of uncertainty, i.e., deterministic or stochastic input data and processes, and regarded time horizons, i.e., time-independent (static) or time-dependent (dynamic); see Pahl (2011).

Studies on depreciation effects frequently assume a general loss or shrinkage of inventory stated as follows:

$$\frac{dI(t)}{dt} = -\Theta I(t) - D(t) \qquad t_j \leq t \leq T_j \tag{4.1}$$

where Θ is the shrinkage factor expressing to perishability, $I(t)$ the inventory holding in time period t, and $D(t)$ is the time-dependent demand defined for a time interval (t_j, T_j) used in

static model approaches. Models integrating this formulation assume exponential shrinkage of inventory which becomes clear when regarding the solution of Equation (4.1):

$$I(t) = -e^{-\Theta t} \int_t^{T_j} e^{\Theta u} D(u)\, du \qquad t_j \leq t \leq T_j. \tag{4.2}$$

Such formulation is mainly used in infinite time deterministic and/or stochastic settings for replenishment and inventory control. The formulation given in (4.3) might be employed for use in (discrete) dynamic deterministic settings where α^{Θ} is the percentage of inventory that is lost due to perishability.

$$(1-\alpha^{\Theta})\, I_t \qquad \forall\, t = 1, \cdots, T \tag{4.3}$$

Pahl (2011) refers to these methods as fraction formulations. They are mainly used in optimization models with the aim to derive optimal replenishment quantities and time lengths (intervals) between orders assuming that all products in inventory undergo the same transformation independent of their age or their production period. As stated before, this can be a good approximation in case of general studies on depreciation effecs; see Waterer (2007).

The assumption of general shrinkage of inventory is limiting if the objective is to create plans that avoid deterioration and related wastages in the first place. To achieve this, lifetime restrictions of products are integrated. Items that lose their functional chracteristics and quality in inventory cannot (should not) be kept in inventory due to various reasons which is, e.g., obvious in case of food that develops biological toxins while deteriorating after a certain time. Consequently, separation from fresh products is necessary to avoid contamination and respect hygiene standards. On the contrary, items that lose their customer's perceived utility can be kept in inventory without such problems in order to be sold on a secondary market. It becomes clear that functional deterioration and depreciation due to value loss pertain to the same problem, but require different consequences or actions.

The modeling of limited shelflives can be integrated, e.g., by modifying the inventory balance equation of discrete dynamic production-inventory or lot sizing optimization models formulated as I&L formulation (see Stadtler (1996)), by subtracting items that pased their useful lifetime calculated as follows:

$$I_t^D \geq \sum_{u=1}^{t-\Theta_t^L} x_u - \sum_{u=1}^{t} D_u - \sum_{u=1}^{t-1} I_u^D,\ \forall\, t \geq \Theta_t^L \tag{4.4}$$

where I_t^D is the amount of items that deteriorate in a time period t that is calculated by the sum of produced items x_u in the time interval $(u, t-\Theta_t^L)$, thus the production from the beginning of the planning horizon until period t reduced by the length of item lifetime Θ_t^L, less those items that are used to satisfy demand until period t and items that have already been disposed in previous periods $t-1$. Spoilage costs need to be added in the objective function if no production costs are regarded which is mostly the case in classical discrete deterministic lot sizing models. If neither production costs or spoilage costs are taken into account, the model might produce items and immediately dispose them to prevent inventory costs; see Pahl et al. (2011).

Constraint (4.4) can be used in all classical discrete deterministic lot sizing (and scheduling) approaches that are formulated as *inventory and lot size models* (I&L; see Stadtler (1996)); see Pahl and Voß (2010), Pahl et al. (2011). However, these models become quickly computational intractable regarding practical input data sets; see also Section 3.1.3. For example, the *capacitated lot sizing model* (CLSP) is proofed to be NP-hard (see Bitran et al. (1982), Florian et al. (1980)) which is also true for its extensions. Therefore, a third formulation is proposed in the literature that is mainly used in network flow formulations and

integrated into the production constraint:

$$\sum_{t=1}^{t} x_t \leq \sum_{u=1}^{t+\Theta_t^L} D_u \qquad \forall\, t = 1, \cdots, T - \Theta_t^L \tag{4.5}$$

of related models which states that the sum of production until time period t must be smaller or equal the sum of demands that do not exceed time period t and the length of the variable lifetime of the product Θ_t^L. This method is named *index transformation* as it restricts the regarded time periods by the lifetime of items; see Pahl (2011). Network flow formulations are efficient in terms of calculation time to find the optimal solution, because they approximately describe the convex hull of the solution space. Nevertheless, the formulation given in Constraint (4.5) does not allow for minimal lot sizes and needs modification that can be done with the following restriction:

$$\sum_{t=1}^{T} \left(x_t + I_t^D\right) \leq \sum_{u=t}^{t+\Theta^L} D_u \qquad \forall\, t = 1, \ldots, T - \Theta^L \tag{4.6}$$

The variable denoting perished products is added to x_t, so that minimal lot size values that are higher than demand until that period and cannot be used to satisfy demand are immediately discarded. Otherwise, the right hand side of Inequality (4.6) might be violated without variable I_t^D. This can be easily modified to integrate rework actions, so that item disposal need not be the only option. Minimal lot size requirements can be due to the physical characteristics of a production resource that can or should always produce a certain number of items in a batch. Examples might be the filling of pallets to transport parts or products to the next production step or packing a certain amount of items together in a package. A cutting machine might be required to always cut the maximal possible number of parts to minimize raw material wastage or WIP. Minimal lot sizes are also named technical lot sizes in the literature; see Flapper et al. (2002). There is a difference to require always exactly the same batch or lot size or to constrain production batches or lot sizes to be equal or higher a certain amount. Technical reasons constraining lot sizes might be, e.g., a machine that is set up to a product cannot technically produce less than a certain amount, but any other quantity above this limit whereas in other settings the exact amount of party and products is binding, e.g., packaging or pallets.

The most frequent formulations used to integrate deterioration found in the literature are Constraints (4.1) and (4.3) that assume a general shrinkage. These are mainly employed in infinite time horizon models where effects ofs demand patterns are added to study the influence of deterioration. The I&L formulation is recently developed and used by Pahl and Voß (2010), Pahl et al. (2011) for discrete dynamic deterministic settings whereas the use of index transformation to limite items' lifetimes are mainly encountered in deterministic settings within approaches formulated as network flow models; see, e.g., Förster et al. (2006), Ahuja et al. (2007).

Chapter 5

Lead Times and Capacity Management

In this chapter, the understanding of lead times are discussed (Section 5.1), planning circularities (Section 5.2), and capacity management emphasizing on the relation of WIP and resource utilization (Section 5.3) leading to lead times that are load-dependent (Section 5.4).

5.1 Lead Time Definitions and Related Considerations

Various forms and definitions of lead times and capacity definitions exist and are used in literature frequently describing different issues or perspectives. They range from *minimal*, *worst case*, *planned*, *actual* or *realized* to *effectuation* lead times where the latter encompasses the time length where a decision is made until the time the consequences of this decision can be observed throughout the supply chain; see de Kok and Fransoo (2003). Besides, a minimal or *best case* lead time might express the shortest possible time a job needs from an initial point in production passing a certain routing until a specific end (inventory) point provided that no unforseen disruptive events or waiting induced additional times take place by, e.g., setups. The opposite is denoted as worst case lead times. If order lead times are regarded, planned minimal lead times may differ by planning induced waiting times, e.g., setups due to the production of multiple products that further increase if setup times are sequence-dependent. In semiconductor manufacturing and assembly line balancing, the term *cycle time* is frequently used to denote lead times; see Hopp and Spearman (2001), Domschke et al. (1997). This might be regarded as an expression of realized lead times that may differ from planned lead times due to various reasons, often being of stochastic nature. Additionally, Missbauer and Uzsoy (2011) refer to *production lead times* as cycle or flow times where *flow times* present the time span of the (actual) manufacturing perspective, i.e., processing of items, that can be interpreted as realized lead times. On the other hand, *lead times* reflect the planning perspective; see Missbauer and Uzsoy (2011). Besides various forms of lead times, there is a great number of terms expressing inclusions of various process steps where subsets of process steps may have further denominations as well; see Figure 5.1 for an overview. Yücesan and de Groote (2000) regard lead times as the time between the authorization of production to the completion of processing where material is available to fill a customer order. As a result, lead times include queuing time, processing time, lot sizing or batching time, transportation and handling time. This is referred to as *flow time* by Hung and Hou (2001). Similarily, Pahl et al. (2005, 2007a) define lead times as the time between the release of an order to the shop floor or to a supplier and the receipt of the items. This definition of lead times is equal to the *total order lead time* in Figure 5.1 including all process step times from replenishment of related (raw) materials until the final

delivery of the order. Hopp and Spearman (2001) define the lead time "of a given routing as the anticipated or maximal time allotted for production of a part on that routing given a certain service level. As such, it is a management constant" whereas the cycle time is regarded as its random version. Note that an anticipated or maximal time value is also named *nominal*. Depending on the definition of the end of routing, the definition of lead time and cycle time according to Hopp and Spearman (2001) are similar to the definition of the *manufacturing lead time* or *cycle/throughput time* in Figure 5.1. The authors further distinguish *customer* and *manufacturing lead times* where the first denotes the time that is allowed to satisfy a customer order from order acceptance to delivery at the customer where different routings can be regarded whereas manufacturing lead times regard a specific routing. In different production environments such as MtO, *make-to-stock (MtS)*, or *assemble to order (AtO)*, lead times may be quite different and their variations can have impacts. In a MtS system, customer lead times defined as by Hopp and Spearman (2001) are zero, because items are on stock or not whereas in MtO, lead times encompass ordering, manufacturing, and processing, so that variabilities might take place and average cycle times must be smaller than lead times in order to guarantee a certain service level, i.e.

$$\text{service level} = P\left(\text{cycle time} \leq \text{lead time}\right)$$

where $P(\cdot)$ denotes the probability value; see Hopp and Spearman (2001). Reductions of variations help to improve estimates of lead times in order to guarantee service levels. Other options to reduce customer lead times is the application of mixed MtO-MtS systems, also named AtO systems, where certain parts of products are produced on stock, so that customer orders are assembled on a MtO basis. MtS systems are also denoted *push systems*, because items are produced on a *just in case*-basis whereas MtO systems produce on a *just-in-time (JIT)*-basis or, in other words, when required by, e.g., demand, so that FGI/WIP inventories are minimized. Hopp and Spearman (2001) define these systems differently due to the change in perspectives: push systems release jobs in the system based on demand (forecasts) whereas pull systems release jobs due to a signal comes from within/inside the system dependent on its status. Here, the first definition is applied that pull systems are triggered by company-external demand. Classical lot sizing models pertain mainly to MtO or pull systems. Nevertheless and as in practice, hybrid forms are prevailing with both push and pull components. For instance, the mechanism of mrp or MRP II that are pull systems releases jobs due to demand. If production smoothing is regarded so as to keep resource utilization at a certain limit (see also Section 8.4) in order to reduce lead times, jobs might be release in previous periods which denotes a kind of forward shifting (see also Section 7.1.1), so that items are produced on stock which gives a hybrid system.

In this analysis, the definition given by Pahl et al. (2007a) is used where the lead time is the time that elapses between the release of an order to the shop floor or to a supplier starting from replenishment and ending at the receipt of the items. This is consistent to the definition of total order lead times in Figure 5.1. Note that in classical lot sizing models, demand is given as the required amounts per product, so that generally no specific orders are taken into account. Therefore, items are not labeled individually and/or explicitely attributed to orders. Consequently, lead times of FGI are regarded abstracting from specific orders including different amounts of multiple orders.

5.2 Planning Circularities

Interdependencies in complex production situations render planning a cumbersome task. Assumptions to reduce complexity abstract from interdependencies, but might lead to infeasible or simply wrong plans. Some of those planning circularities that influence lead times

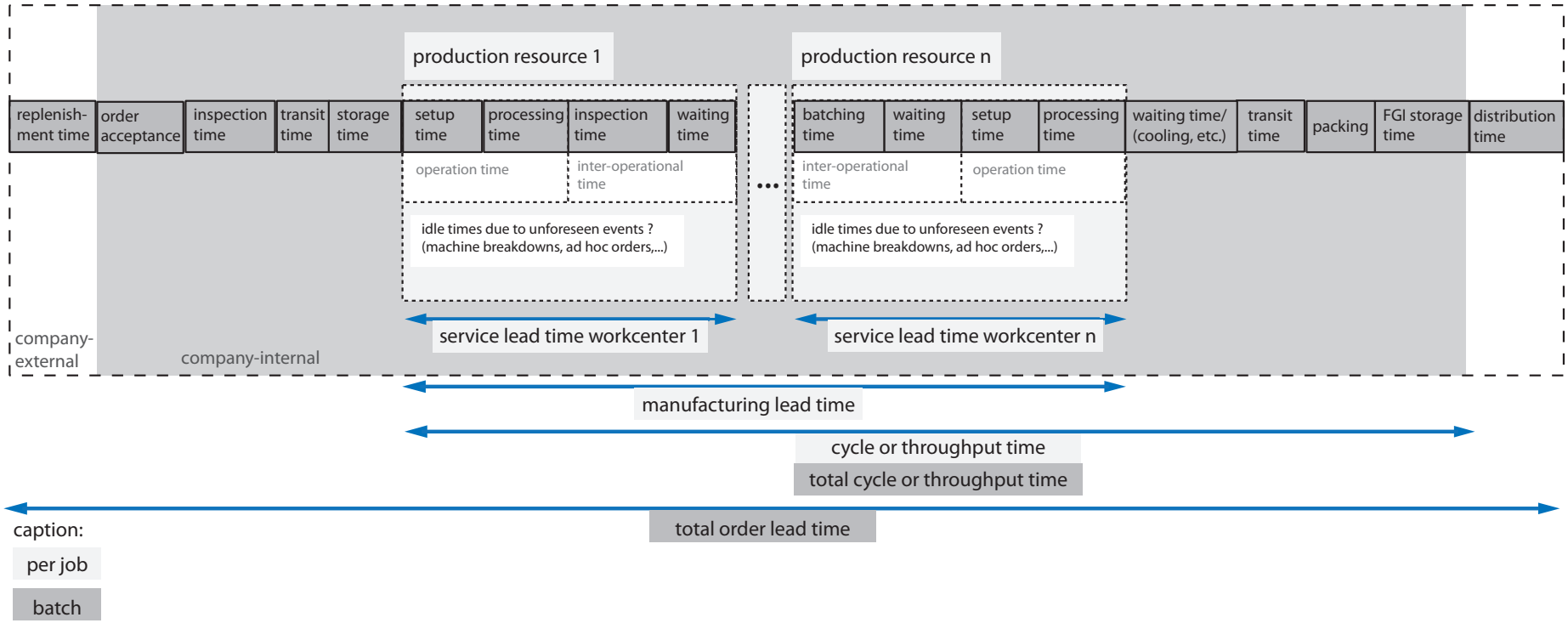

Figure 5.1: Different process times and denotation of compositions regarding lead times and other process time related terms; based on Obenauf (2011)

and thus should be accounted for are discussed in the next sections. First, planning circularities regarding lead times, utilization, and WIP are examined in Section 5.2.1. Afterwards, problems due to the consideration of multiple products are taken into account in Section 5.2.2. Influences of depreciation on LDLT are highlighted in Section 5.2.3.

5.2.1 Fixed Planned Lead Times

Most models for aggregate (production) planning treat lead times as static, fixed input parameters that are estimated upon experience or fixed with a rule of thumb. In most situations, the output of a planning model determines planned capacity utilization which, in turn, influences planned lead times; see Pahl et al. (2005, 2007a) and the references therein. For instance, *worst case* lead times might be used in order to have enough "buffer time" to fulfill accepted orders regarding a certain time interval and thus satisfy demand in time; see, e.g., Enns (2001), Enns and Suwanruji (2004). Besides, orders may be released earlier into the system than necessary to ensure delivery due dates, so that WIP builds up way ahead and resource utilization increases. Items are forced to wait, so that lead times increase, too, compromising the initial idea to assure due dates and service levels. This over-reaction that corrupts lead times leading to high variabilities becomes a self-fulfilling prophecy and is adressed in the literature as the *lead time syndrome*. It shows the results if the relations between average lead times, WIP, workload, are ignored; see Tatsiopoulos and Kingsman (1983), Zäpfel and Missbauer (1993a,b), Teo et al. (2011) and Selcuk et al. (2006, 2009) for extensive studies. Lead times should be a result of planning than input data. In fact, they depend on the overall available capacity and workload of the system that can vary significantly over time not only according to demand requirements; see, e.g., Karmarkar (1993). This is not reflected in common planning models like mrp or MRP II systems that treat lead times as attributes of parts; see, e.g., Dauzère-Pérès and Lasserre (2002) or Pahl et al. (2005, 2007a).

In the literature, various researchers concentrate on the estimation of planned lead times (see, e.g., Hackman and Leachman (1989), Vepsalainen and Morton (1988), Lu et al. (1994), Hung and Hou (2001)), so that lead times are unrelated to system workload, but system utilization is indirectly regarded by, e.g., testing and iterating estimated lead times regarding their influence on the system using simulation studies. However, the relationship between system workload, resource utilization, and lead time(s) (distributions) is not taken into account. Teo et al. (2011) argue that not much work is provided that analyze planned lead time and workload in MtO environments subject to highly varying demands. They study related production smoothing methods with the aim to reduce subcontracting and overtime as options to deal with demand peaks and variabilities. They propose smoothing strategies on the MPS and planned lead time control at multiple workstations. Decisions focus on the optimal planning windows of the planning horizon and optimal planned lead times that minimize subcontracting and overtime costs. Rao et al. (2005) propose a model including maximum lead time guarantees for a MtO production system as a tool to hedge against overall supply chain risks. They further highlight the impact on customer demand and production planning. Customers are attracted by short lead times that increase profits of firms. On the other hand, short(er) lead times imply more complex coordination of production and its planning including tight production plans deprived of buffer times. The authors show in a stochastic setting that the optimal lead time is similar to a newsvendor structure with a closed form solution; see Rao et al. (2005) for more details. Other authors incorporate the relation between lead times and production (system) workload in their optimization models linearizing the resulting nonlinear relationship, so that WIP-related exponential increasing lead times are captured; see, e.g., Asmundsson et al. (2003, 2006a,b), Missbauer (1999, 2002, 2006b,a), Missbauer and Uzsoy (2011), Yano (1987), Zijm and Buitenhek (1996).

For more details on such models see Section 7.

5.2.2 Issues Regarding Production Mix

Lead times are lengthened through setup times in case of multiple products. The number of setups are an output of the planning function dependent on the requested production mix which generally varies in dynamic planning settings. Setups consume capacity that is not available for processing. Consequently, an increased number of setups leads to an increment of resource utilization and a decrease of available capacity worsening or leading to infeasible plans in case that setups are sequence-dependent (*sequence-dependent setup time (SDST)*) in times and costs and not accounted for in planning, accordingly; see Pahl et al. (2011). Large lot sizes decrease the number of setups, but lead to increases in WIP/FGI further implying additional waiting times for or after processing. Consequently, lead times do not only depend on the workload and utilization of the production system, but also on the work assignments of resources that are, in turn, determined by the planning function. Besides, the determination of optimal lot sizes requires SDST and costs that, in turn, depend on the scheduling solution; see Dilts and Ramsing (1989). On the other hand, information concerning types and amounts of products in every period is necessary for the determination of the optimal schedule. Dilts and Ramsing (1989) propose a method to estimate initial setup costs employable in lot sizing and scheduling models and analyzes the effects of estimated costs on the solution.

5.2.3 Issues Regarding Depreciation

Limited lifetimes of products and related quality decreases aggravate the problems regarding LDLT, because their increases can lead to significant quality losses of items until they pass their useful lifetime. Related rework actions further occupy resource capacities enhancing workload, utilization, and thus lead times; see also Figure 5.2. Certainly, this is less problematic if passed products are scraped, sold on a secondary market, or reworked in an off-line manner, i.e., employing production resources different to regular production. Anyhow, the solution to scrap items is environmentally undesirable and off-line rework requires investments in additional capacity leading to enhanced maintenance and operational costs. Rework actions need to be included into planning, because they consume capacity. Thus, lot sizing and sequencing needs to be adapted, accordingly. In case that depreciation-dependent rework is prevalent and items continue to depreciate while waiting for rework, actions become even more time and resource-consuming. Related mathematical modeling is addressed in Section 9.2.

5.3 Capacity Management and Utilization

Definitions of capacity and measurements are numerous. Terms to describe capacity range from *design* capacity, *theoretical maximum* capacity, *potential* capacity to *true, actual, effective* capacity; see, e.g., Elmaghraby (2011). He states that, frequently, the term “capacity” is used in the sense of *capability* which is associated to *performance* or *output* of production entities. In fact, in classical lot sizing, the capacity input parameter denotes the maximum time units a production resource is available. A resource consumption parameter attributed to each product denotes how many capacity units the production of a single product employs. Consequently, the capacity parameter denotes capacity as capability of the production resource to produce a certain number of output of the regarded products. Resource capacity is greatly influenced by many factors. Fowler and Robinson (1995) analyze such influencing factors that are grouped the following way:

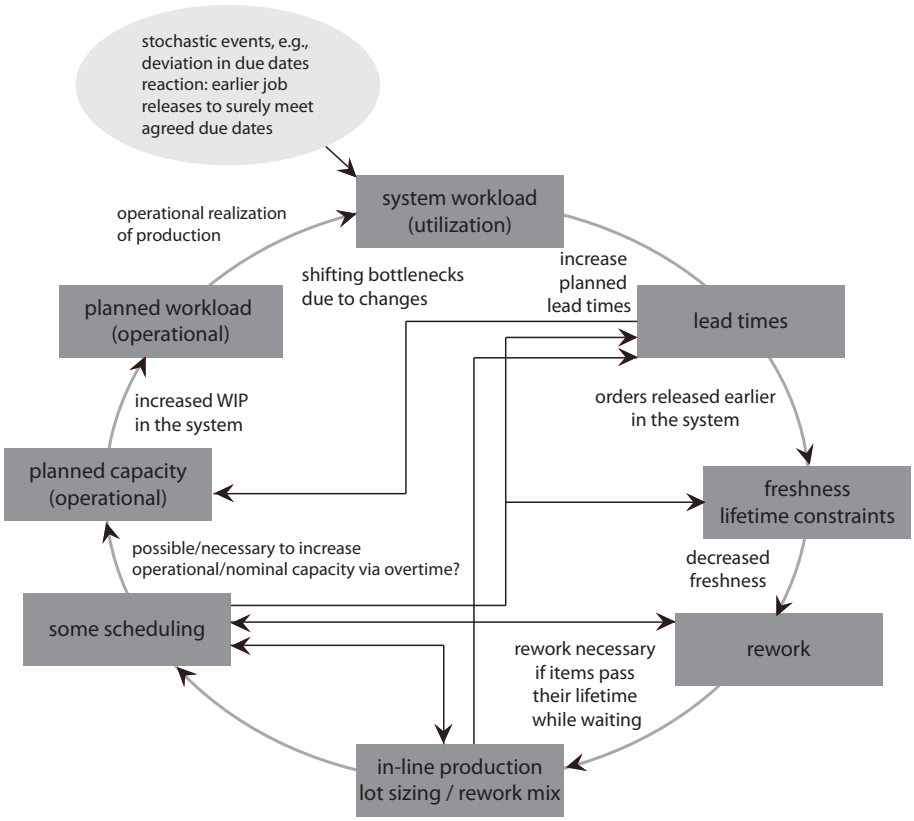

Figure 5.2: Extended planning circularities and lead time syndrome.

- Technically required and process-related (waiting) times; e.g., tool dedication, *hot lots*, time bound sequences,
- Maintenance-related processes, e.g., regular inspections, operator cross-training,
- Planning-related (waiting) times, e.g., production mix-related batching/lot sizing, dispatching/sequencing, setups, rework,
- Unforseen events, e.g., breakdowns, unplanned inspection and maintenance due to yields/disruptions,
- Human factors, e.g., end-of-shift effects, operator availability.

Certainly, these factors might also be classified in a different way regarding the causality of appearance. Note that rework is assigned to planning-related (waiting) times, because rework options for items that passed their lifetime limits are regarded and thus need to be reworked in order to gain required quality standards for re-used in production. In case of defectives and their rework, this influencing factor would be classified as unforeseen events.

For the measurement of capacity, Elmaghraby (2011) proposes four forms of capacity, i.e., *nominal, operational, planned*, and *utilized capacity*; see also Figure 5.3. The *nominal capacity* denotes a theoretical, maximal capacity of a production facility, thus the capability of a productive entity in regard of a single standard product or process assuming perfect support activities including optimal maintenance or overhaul actions. Such *anticipated* and *unavoidable* change-over losses of productivity are dependent on the general (physical) state of the regarded production entity. Other change-over losses, e.g., those due to the production mix, decrease the nominal capacity to the operational capacity. As already stated elsewhere, the determination of the *operational capacity* depends, on the one hand, on the performance

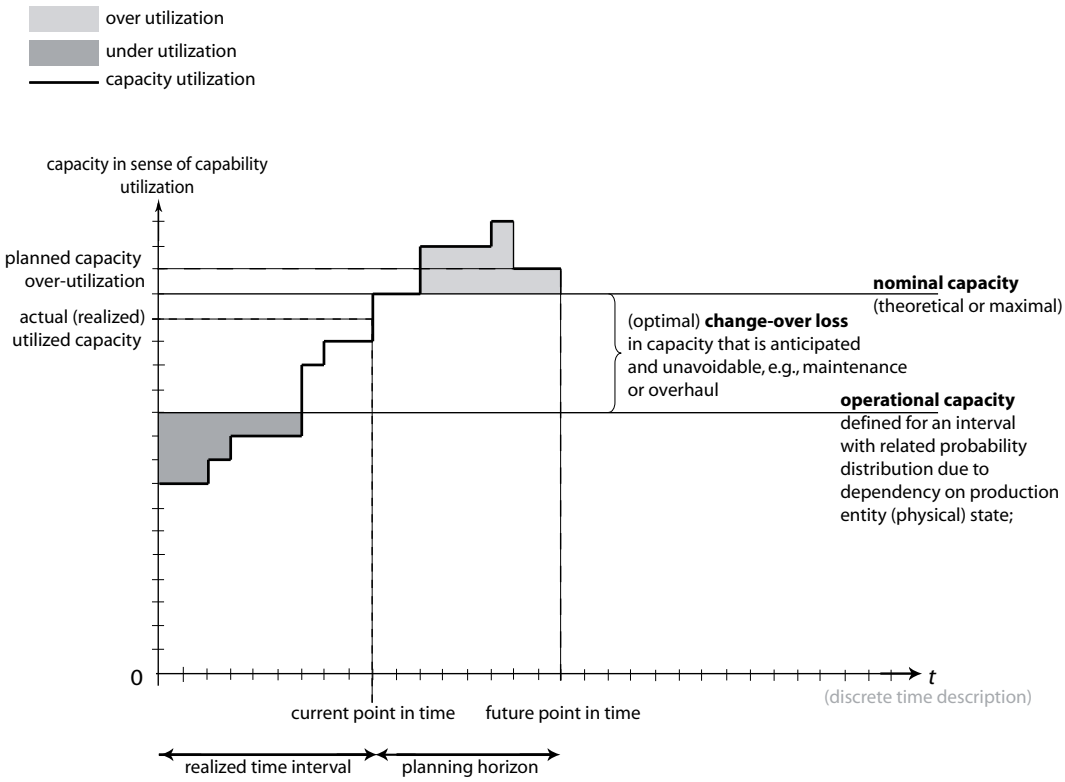

Figure 5.3: Visualization of the different capacity definitions in a sense of capability.

of the production scheduling function to calculate optimal setups in case of multiple products and, on the other hand, on the standard (average) rate of rejects, i.e., defectives or waste. Benchmarks regarding related or similar industries might further help to determine operational capacity or reasons for deviations; see Elmaghraby (2011). For instance, idle times due to poor performance of support activities, e.g., labor absenteeism, shortages of equipment, raw materials, or electricity blackouts etc., can or should be included into the determination of operational capacity.

In order to give an example regarding this issue from a different, but similar field of application, project management is regarded: it is common practice to use average operational capacities and resources utilization of project workforces with ca. 80% utilization.[1] This is straightforward, because people need personal time to concentrate, re-issue, arrange work, and take pauses. Personal leave due to holidays and other allowances are assumed to be already subtracted from nominal capacity. Furthermore, illnesses might take place, so that these capacity variabilities need to be taken into account as well, so that unforeseen events do not compromize projects. One might argue that this is not the same with machines in manufacturing. But similarities can be drawn: machines might need time to power up in case that they are not operated on a non-stop basis, so that end-of-shift or begin-of-shift effects exist. Besides, machines need regular maintainance, so that their effective employable time or operational capacity is neither 100%. Disruptive events of the production process such as machine breakdowns or externally induced waiting times increase variabilities in capacities, so that calculating plans assuming 100% resource utilization is hazardous in terms of securing plan fulfillments. Applying this logic, the *planned capacity utilization* denotes a portion of the operational capacity over a planning horizon. This might not be equal to the operational capacity due to a variety of reasons, e.g., a lack of demand in relation to nominal

[1]Operational capacity in sense of capability; see also the proposed method of Elmaghraby (2011).

capacity in the planning horizon (planned under-utilization), or bottlenecks regarding other process-dependent production entities etc.; see Figure 5.3. Planned over-utilization realizable by overtime or adjustments of work paces (or production rate) might take place in case of foreseen peaks in seasonal demands.

For the remainder, the term "capacity" is used to express capacity in a sense of capability of the regarded production entity. If not differently stated, Here, capacity is refered to as the nominal capacity which is decreased, among others, through the planning function regarding lot sizing and related setups to operational capacity. Because deterministic optimization models for planning are analyzed with the output being a plan, realized forms of capacity or lead times are not stated as this can only be measured ex post and would take place on the execution level of production that is not subject to this analysis.

5.4 Load Dependent Lead Times

System workload, i.e., WIP, increases in a nonlinear fashion with capacity utilization long before the capacity is reached which implies lead times dependent on the workload of the system, named LDLT. The workload measured by WIP inventory and its relation to resource utilization can be depicted as in Figure 5.4.

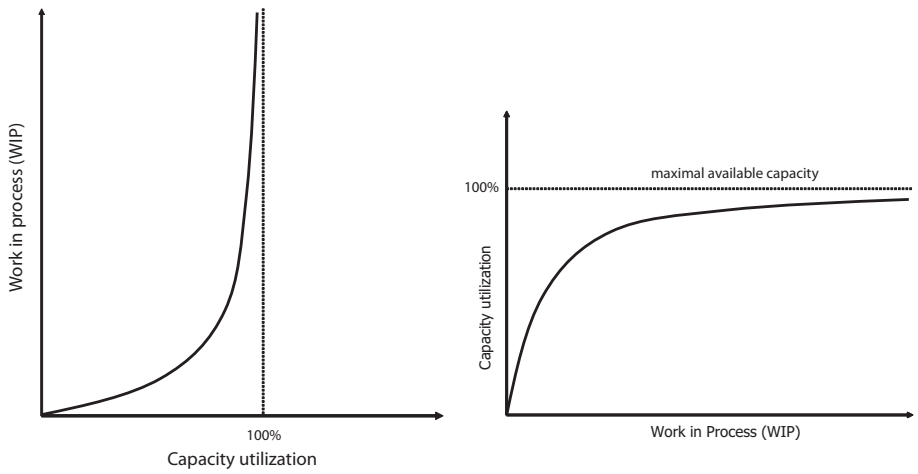

Figure 5.4: Relation between WIP and resource utilization and its integration in clearing functions; see Pahl et al. (2005).

This relationship is supported by empirical evidence that suggests an exponential increase in lead times with the utilization of the production system; Riaño (2002), Elmaghraby (2011). This can be also mathematically shown using queuing theory, especially applying Little's law which constitutes a fundamental law in production; see Hopp and Spearman (2001). It derives from queuing theory and states that the WIP of a production entity (see Figure 5.5) is determined by the arrival rate of jobs and the average processing or residence time of jobs within that production entity.

In the steady state, the input into a system (synonym: *server*) must equal its output. If the processing of a job consumes on average a processing time denoted by $\bar{p}$ stated in TU and the number of jobs arrive with a rate λ [QU/TU], the length of the queue can be expressed via the average number of jobs waiting denoted by $\overline{W}$ as follows:

$$\overline{W}\,[\mathrm{QU}] = \lambda\,[\mathrm{QU/TU}] \cdot \bar{p}\,[\mathrm{TU}] \tag{5.1}$$

This relation is named Little's law; see Hopp and Spearman (2001), Buzacott and Shantikumar (1993), Elmaghraby (2011). Buzacott and Shantikumar (1993) derive Little's Law

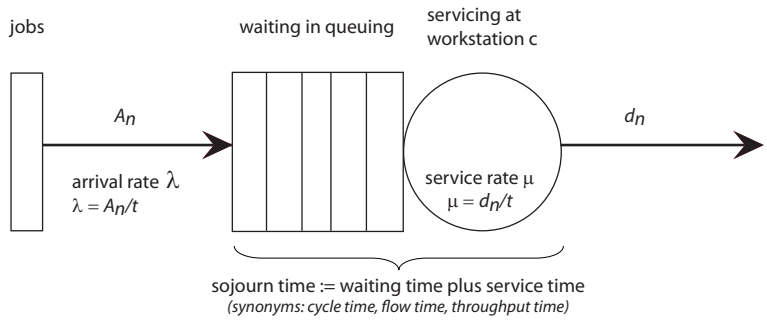

Figure 5.5: Simple queuing model with one workstation (server); modified from Hopp and Spearman (2001).

assuming that the number of arriving jobs in any finite time interval is uniformly finite. The number of jobs arriving and departing from the system is given by A_n and D_n, respectively. The number of jobs arrived in a time interval $(0,t]$ is given, accordingly:

$$A(t) = \sup\{n : A_n \leq t, n = 0,1,\ldots\} \tag{5.2}$$

with $A_0 = 0$. The number of jobs departed during time period $(0,t]$ is given by

$$d(t) = \sup\{n : d_n \leq t, n = 0,1,\ldots\} \tag{5.3}$$

where d_0 is set to $d_0 = 0$ and $D(t)$ includes all jobs that might depart at time t. The number of jobs in the system is given by the following equation at time t (see Buzacott and Shantikumar (1993)) as follows:

$$W(t) = W(0) + A(t) - d(t), \qquad t \geq 0 \tag{5.4}$$

Assuming that the job arrival rate is uniform as the number of time periods approach infinity, i.e.:

$$\frac{A(t)}{t} \to \lambda \text{ uniformly as } t \to \infty \tag{5.5}$$

as well as the job departure rate denoted by δ:

$$\delta = \lim_{t\to\infty} \frac{d(t)}{t}, \tag{5.6}$$

these two rates λ and δ are the same in case that the number of jobs in the system is uniformly finite for all time $t \geq 0$. This might imply that also the number of jobs in the system need to be uniformly finite, i.e., $\lim_{t\to\infty} W(t) = 0$; see Buzacott and Shantikumar (1993). Nevertheless, they show that $W(t) \to \infty$ as $t \to \infty$. The sufficient condition to this is the following inequality

$$\lambda \leq c_w \cdot \mu \tag{5.7}$$

where c_w is the number of workstations in the system and $1/\mu$ is the average service rate described as follows:

$$\frac{1}{k}\sum_{n=1}^{k} S_n \to \frac{1}{\mu} \text{ uniformly as } k \to \infty \tag{5.8}$$

If the time $p = d_{K(n)} - A_n$ that the nth job stays in the system is given and supposing that the average time the job spend in the system is determined by the difference in time of departure and arrival rate, i.e.

$$\overline{p} = \lim_{t\to\infty} \frac{1}{k}\sum_{n=1}^{k} \left(d_{K(n)} - A_n\right) \tag{5.9}$$

further assuming that the average number of jobs in the system exists:

$$\overline{W} = \lim_{t \to \infty} \frac{1}{t} \int_0^t W(\tau) d\tau \tag{5.10}$$

Little's Law can be derived as given in (5.1):

$$\overline{W} = \lambda \cdot \overline{p}. \tag{5.11}$$

The average capacity utilization is stated, accordingly (see also Woodruff and Voß (2004)):

$$\rho = \frac{\overline{\lambda}}{\overline{\mu}} \tag{5.12}$$

where $\overline{\lambda}$ denotes an average job arrival rate and $1/\overline{\mu}$ the average service rate or level. As Woodruff and Voß (2004) state, $\overline{\mu}$ might be assumed as fixed in order to determine $\overline{\lambda}$ by planning decisions that is proportional to the utilization denoted by ρ in tactical planning. The capacity of the system is determined by $\text{Cap} = w \cdot \mu$, so that any arrival rate below the capacity can be handled in a way that the number of jobs in the system is always finite. Due to the fact that this capacity is related to the average service time, it is assumed that this presents the operational capacity which is straightforward, because this capacity definition is not the same as a maximum rate at which jobs can depart from the system; see Buzacott and Shantikumar (1993). In case that the arrival rate is higher than capacity, thus $\lambda > w \cdot \mu$, an infinite number of jobs might be allowed in the system, so that also lead times would approach infinity; see Buzacott and Shantikumar (1993). Finally, the average lead time can be expressed by the following equation assuming a constant average WIP in the steady state (see Elmaghraby (2011)):

$$\overline{L} = \frac{\overline{p} \cdot W}{2}. \tag{5.13}$$

If release rates determining job arrivals are assumed to be decided externally which is the case assuming deterministic demand in production and lot sizing models, methods to reduce WIP include, e.g., an increase in capacity through quantitative or intensity adjustments of production entities which implies an increase of the number of production entities (workers or machines) or an increase of the processing rate (speed up work); see Adam (1998). Elmaghraby (2011) argues that empirical evidence shows some kind of "flustering" of production resources at high loads, so that they cease to clear an average of $1/\overline{p}$ jobs assuming that there is enough WIP to keep the production resource occupied. Consequently, approaching a certain workload leads to significant decreases in productivity of the production resource. As a consequence, it might be a misbelief to think that increasing average intensities, i.e., the average processing rate, might decrease WIP and therefore, lead times. On the other hand, there is psychological proof that work intensities are intensified if a significant amount of work/jobs awaits for processing. But this intensification may come with an increased amount of errors, so that rework needs to be performed and lead times become longer on average even though an increased processing rate has been employed. Such intensification option might lead to enhancements of lead times in the short run. This might not be true in the long run.

Chapter 6

Green Operations Management and Sustainability

Greening the supply chain in terms of reducing waste of various forms and emissions of environmentally harmful materials along the whole supply chain to protect our living place for us and future generations is adressed by GSCM. This is an urging objective when thinking of globalization and the dynamics of competition required to deliver less costly products in a fast and flexible way to continuously decreasing cost requirements. Shorter product life cycles as well as increased rates of depreciation not only due to perceived reduced customer value add to this dynamism; see Garetti and Taisch (2012), Jaber and Rosen (2008). The pressure to offer products at minimal costs in order to remain competitive increases, so that companies are forced to search for alternative ways that might come at the expense of the environment. For instance, it may be economically reasonable to relocate production to countries with low wage structures despite higher transportation costs of materials and products. Increased transportation rise emissions of carbon dioxides and other harmful gases damaging the environment. Consequently, actions to avoid or minimize increased emissions range from rethinking production/storage locations to minimize the necessity of transportation to rendering transportation and vehicles less pollutive by, e.g., employing other sources of energy than fuel or implementing appropriate filters. The action plan for energy efficiency of the European Council (2011) states that energy efficiency in transportation has the second largest potential regarding energy saving. Buildings have the largest energy saving potential; see European Council (2011). The third largest energy saving potential lies in industrial production processes that causes 20% of the primary energy consumption of the European Union whereas transportation accounts for 32%; see European Council (2011).

Another target of GSCM regards the preservation of slowly or non-regenerative resources. In case of destructive exploitation, such resources do not regenerate and disappear forever, e.g., the rainforest or natural oil wells. Consequently, the development of alternative resources, technologies, and related strategies is mandatory in order to secure availability and use of resources in an intelligent way. However, the problem of carbon dioxides is not only due to transportation requirements, but also present in diverse parts of the supply chain such as manufacturing. Therefore, requirements to use renewable energy in the production process become a central interest of law making of many countries also due to consumer pressure as one of the main drivers of GSCM; see Ho et al. (2009), Srivastava (2007). As stated by the European Council (2011) in their energy efficiency plan, progress in terms of energy efficiency has been greatest in the industrial production sector with 30% improvement over the past 20 years. Further improvements are targeted by eliminating obstacles, e.g., lack of information, insufficient access to financial resources for energy saving investments especially for small and medium enterprises and short term pressures of busi-

ness environments; see European Council (2011). Moreover, the European Council (2011) anounces the definition of strategies to enhance the transportation concept in all its facets by introducing an advanced traffic management system as well as an infrastructure investment plan to create a "Single European Transport Area." This includes *smart pricing*, efficiency standards for all vehicles and measures to enhance innovations regarding vehicle technologies; see European Council (2011).

6.1 Strategies and Main Areas of Green Supply Chain Management

The strategies of GSCM range from reactive monitoring of general environmental programs mainly required by law to proactive environmental management including active avoidance of pollution, wastages, over-utilization of scarce or long-term regenerative resources, rework, recycling, as well as remanufacturing of used items. These strategies can be divided into different fields as depicted in Figure 6.1.

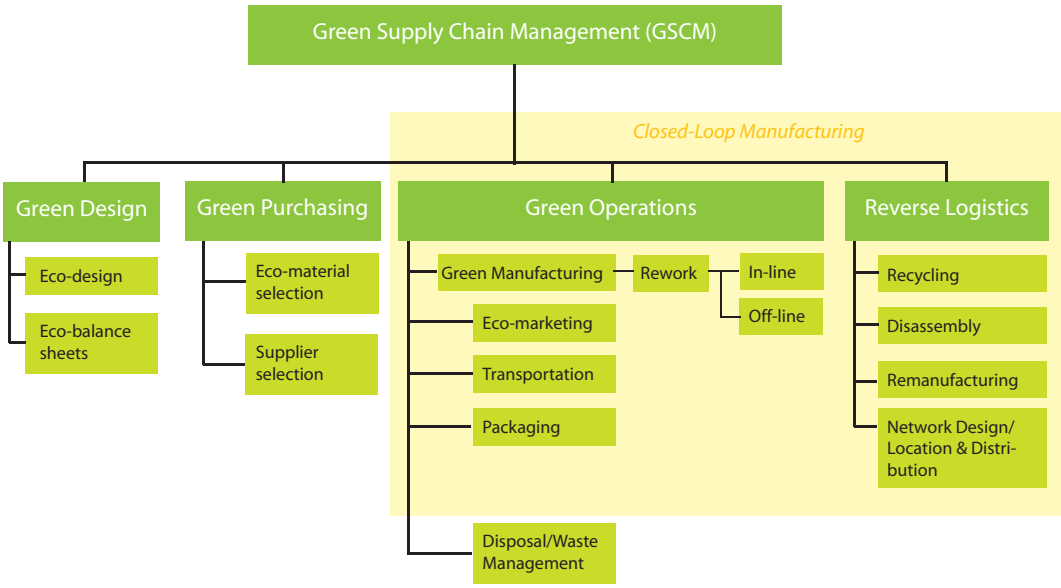

Figure 6.1: Fields of green supply chain management; modified from Klütsch (2009), based on Srivastava (2007).

For instance, *green (eco) design* comprises the development and configuration of (new) products and related production processes taking into account ecological criteria ranging from the selection of appropriate (raw) materials to the configuration of low energy production processes; see Nikbakhsh (2009), Srivastava (2007). Related analysis regarding lifecycles of products to reveal aspects of ecological improvement are presented in *eco-balance-sheets* and published on a regular basis by companies in order to document related efforts. The integration of such documentation into *energy management systems* according to EN 16001 norms[1] might be worth a thought that includes reviewing and auditing and may become mandatory for larger companies in the future; see European Council (2011).

[1]These are the norms regarding energy management systems that are oriented on ISO 14001 norms. The latter mentioned cover all aspects regarding waste/disposals, effluents, emissions, and energy. EN 16001 concentrate on energy management; see http://www.iso.org/iso/home.html.

Green purchasing comprises all actions of replenishment concentrating on the selection of suppliers according to diverse environmental factors, e.g., locations/transportation, quality and usage of materials. Naturally, green purchasing is closely related to green design which defines, but also constrains the usage of materials. Additionally, green purchasing is also closely related to disassembly and recycling issues. For instance, strategies of green design might lead to the production of a few congenerous product families to reduce product and production process complexity. The usage of a few standardized recyclable materials further decreases replenishment and transportation costs also due to the possibility to exploit quantity discounts.

Green operations pertain together with *reverse logistics* to *closed-loop manufacturing* and contain the whole set of actions regarding green production, eco-marketing, transportation, packaging, disposal management, and rework. Besides, reverse logistics include recycling from company-external sources, disassembly, and remanufacturing of recycled products. Recycling issues are very important regarding transportation costs due to the fact that logistics represent up to 95% of total costs; see Srivastava (2007). Rogers and Tibben-Lembke (2001) state that reverse logistics activities and related costs are difficult to determine, because such activities are not tracked. Nevertheless, reverse logistics of studied firms accounted for approximately 4% of total logistics costs. A great body of work exist in the field of reverse logistics; see, e.g., Fleischmann et al. (1997), Angell and Klassen (1999), Dekker et al. (2004), Inderfurth et al. (2006) or Fleischmann (2001), Flapper et al. (2002) for reviews. Rogers and Tibben-Lembke (2001) examine reverse logistics and highlight similarities and differences to green logistics. They define reverse logistics as the "process of eficiently and effectively planning, implementing, and controlling the flow of raw materials, WIP, FGI, and related information from the point of consumption to the point of origin in order to recapture or recreate value or to properly dispose products." Activities may pertain to both reverse and green logistics, such as the usage of recyclable packaging to avoid one-way cardboards whereas package reduction activities are rather distributable to green than reverse logistics; see Figure 6.2.

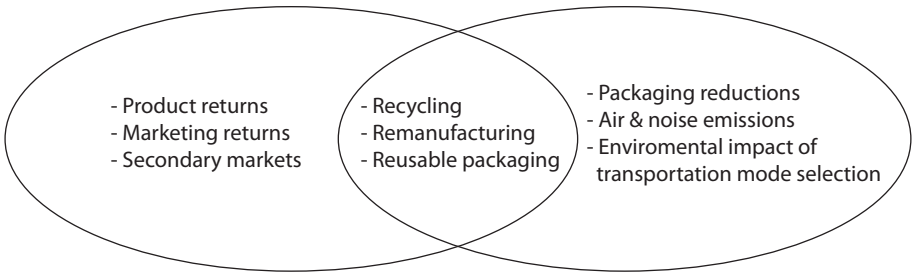

Figure 6.2: Comparison of reverse logistics and green logistic; see Rogers and Tibben-Lembke (2001).

Whitin the focus on depreciation effects of parts and products in tactical and operational production and inventory management, rework and remanufacturing issues are taken into account in the production planning process. The *internal rework system* (see also Figure 6.3) and the *external recycling system* are studied where the latter assumes external returns and their in-line remanufacturing/rework for reuse in regular demand satisfaction.

6.2 Sustainability

The term "sustainability"can be widely understood as the "quality that permits to preserve, to keep, to maintain something;" see Duque Ciceri et al. (2010). Therefore, "if something

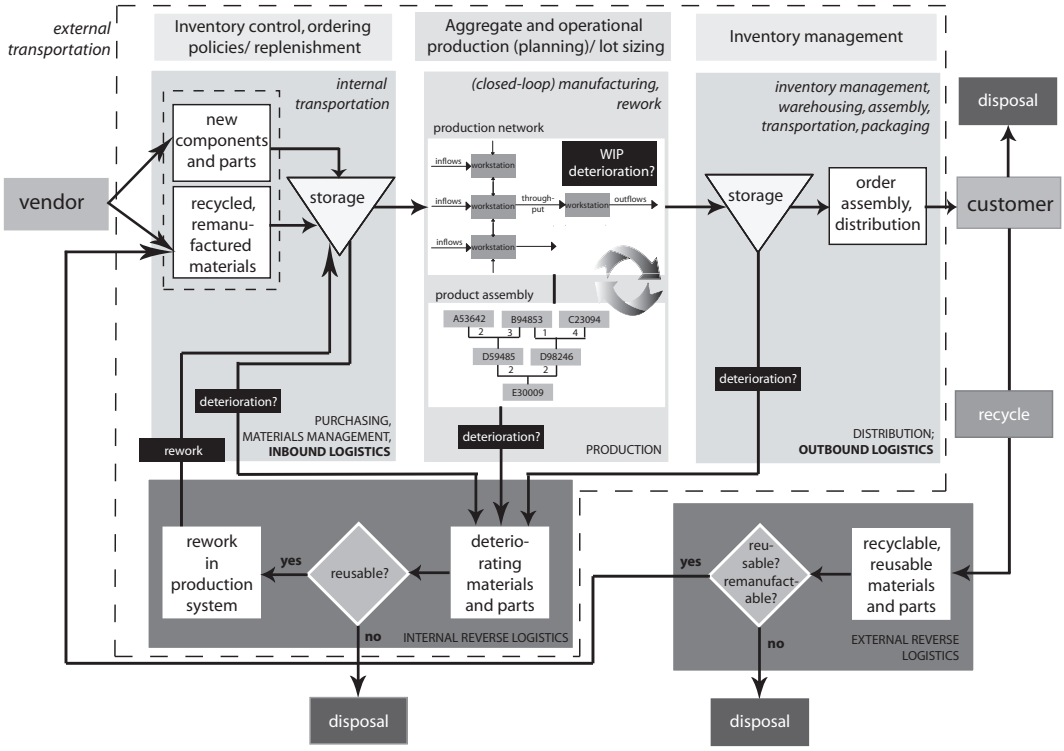

Figure 6.3: Production system with deterioration and rework; based on Sarkis et al. (2004), Hervani et al. (2005); see also Pahl et al. (2007b).

is sustainable, it is able to be kept;" see Garetti and Taisch (2012). The World Commission on Environment and Development (WCED) defines sustainability as "using resources to meet the present needs without compromising the ability of future generations to meet their own needs;" see World Commission on Environment and Development (WCED) (1987). Chaabane et al. (2011) put it simple and state that sustainability implies to avoid permanent environmental damage. For an overview of sustainability in supply chains we refer to Linton et al. (2007).

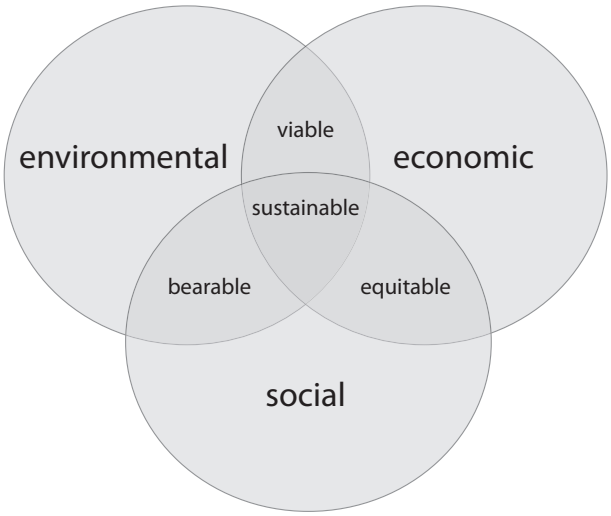

Figure 6.4: Three pillars of sustainability; see Garetti and Taisch (2012).

Production impact all pillars of sustainability; see Figure 6.4. This becomes clear when

regarding its impact on global warming, waste generation, toxic releases and water/air emissions, floating plastic and product end-of-life issues; see Ho et al. (2009). Technology is stated to be one critical factor in environmental preservation, especially regarding sustainable design of manufacturing processes as well as information and communication technologies that might help to change work and life concepts, accordingly, which certainly goes along with an corresponding education in sustainability and economical implications (see Garetti and Taisch (2012)) influencing the customer-supplier relations; see Hall (2000). As shown by Beamon (1999), 75% of consumers stated that they were influenced by the environmental reputation of firms and 80% of them are willing to encounter higher prices for environmentally friendly products. In fact, Hall (2000) reveals that firms modify their approaches to sustainability according to preassure from various groups.

6.2.1 Strategies for Sustainability

Different strategies for sustainability/environmentally friendly practices for internal operations have been developed by, e.g., Vachon and Klassen (2006) that include, e.g., the implementation of an *environmental management system (EMS)*[2]. These might have a positive impact on the operations performance of companies especially in case that they are certified, e.g., according to ISO 14001 standards. Certified companies have a process-related working basis that helps to concentrate on the analysis of creation, management, and elimination of pollution or waste instead of purely measuring output; see Melnyk et al. (2003) and the references therein. Positive impact that comes along with waste and pollution reduction may further enhance key performance indicators such as quality, cost factors, and lead times. EMSs have an impact on every area of GSCM shown in Figure 6.1 by evaluation and audit on the organizational level as well as on product life cycle and process level. Positive impacts can be spread to the whole supply chain by motivating or require contractors and suppliers to establish their own EMS. Environmental performance indicators to evaluate sustainability performance of activities, processes, etc. are found in the ISO 14031[3] guidelines specific to environmental performance evaluation; see Hervani et al. (2005). Cagno et al. (2012) propose environmental costing methods for measuring eco-efficiency sustainable production based on activity-based costing methods. Tarí and Molina-Azorín (2009) draw similarities to the model of the *European Foundation of Quality Management (EFQM)*. Other strategies incorporate lean management principals together with pollution preventing technologies, and *product stewardship* which includes activities such as reverse logistics, product recovery, and remanufacturing; see Vachon and Klassen (2006). Kyrö et al. (2011) study similarities between lean management principals and energy management especially regarding building management whereas King and Lenox (2001) examine the relationship between lean production and environmental performance based on empirical data of U.S. manufacturers. They reveal a linkage between the adoption of ISO 9000 standards[4] and the likelihood to adopt ISO 14000 environmental quality management standards[5]. Consequently, there is a positive correlation between lean production principles and the aim to

[2] An EMS is defined as a formal system that documents processes and procedures to summarize, control, report, and train personnel regarding environmental performance for internal and external use. The internal use regards the minimization or avoidance of waste and pollution based on (process) design, training, and reporting to the management who sets goals that might concentrate on compliance meaning to avoid sanctions by maintaining the minimal legally accepted pollution standards or waste reduction. The latter encompasses all activities to maximize reductions of negative environmental impacts; see Melnyk et al. (2003). The external use includes reporting to stakeholders and shareholders in order to enhance the company image; see Melnyk et al. (2003).

[3] See http://www.iso.org/iso/home.html.

[4] See ibidem.

[5] See ibidem.

avoid/reduce overall wastages and pollution; see also Simpson and Power (2005).

6.2.2 The Formulation of Sustainability Objectives

Economic objectives are dominant in practice that are often contrary/detrimental to sustainability and environmental considerations. Conventional economic objectives regard profits and costs of economic activities omitting social and environmental effects frequently owed to competitive preassure; see Ho et al. (2009). Objectives of companies need to be analyzed and reformulated according to sustainability and environmental aspects, so that trade-offs are taken into account and eliminated or equilibrated; see also Figure 6.5. As stated by Ho

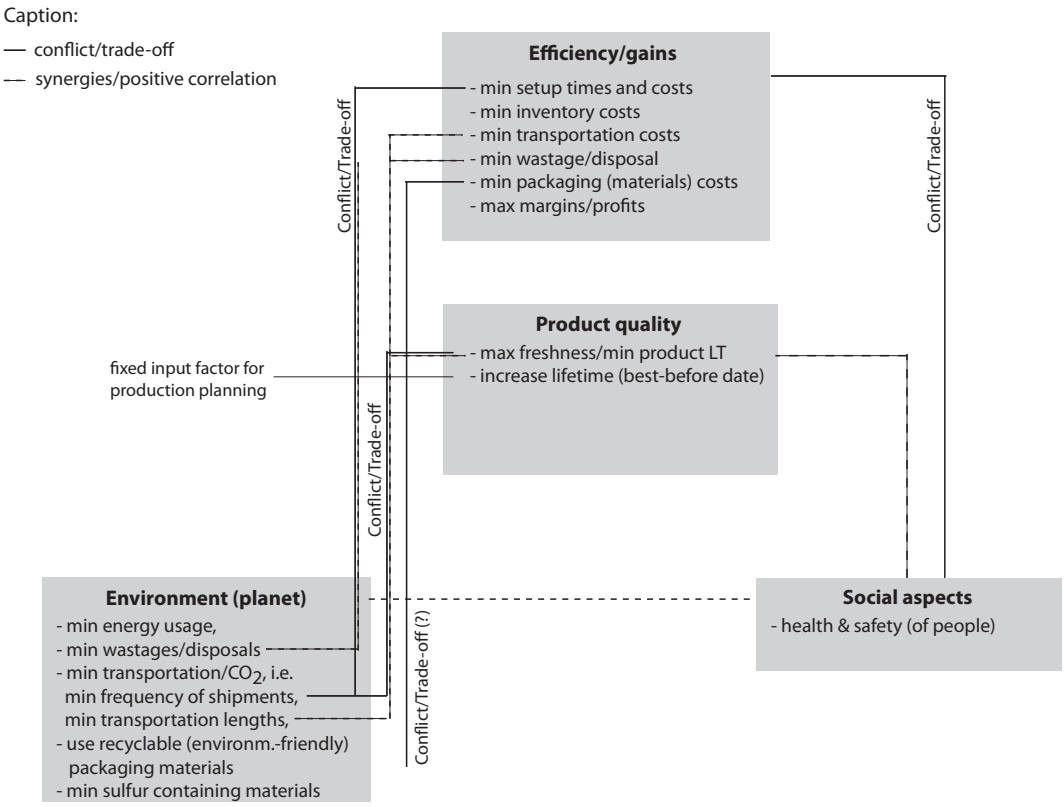

Figure 6.5: Trade-offs and synergies regarding different production and delivery options; modified from Akkerman et al. (2009).

et al. (2009), innovations and optimal planning of the green supply chain may close the gap regarding higher manufacturing costs of green supply chains in comparison to conventional supply chains.

Objectives to greening the supply chain are closely related to lean management principles emphasizing the elimiation of every kind of waste (see also Melnyk et al. (2003), Ho et al. (2009)). They further include

- *Substitution* implying the replacement of environmental harmful materials with unproblematic ones (see also Ho et al. (2009)),
- *Internal consumption* including, e.g., the use of returned packaging and wood pallets employed in shipping to attain energy or to reuse them in regular production; see also Golany et al. (2001), Despeisse et al. (2011). This includes enhancing the relation of resource utilization and their productivity including preferences to consume less

materials in manufacturing by increasing productivity of materials and energy and the reduction of unwanted outputs as well as the employment of the restant as inputs which implies recycling in every production area.

- *Waste minimization* including energy savings, water savings, minimization of CO_2-levels, minimization of every kind of pollution (water, air), process-related waste/disposal of materials; see also Ho et al. (2009), Jaegler and Burlat (2012).
- *Supplier selection* according to ecological criteria; see Ho et al. (2009),
- *Development of innovative products* that employ environmentally friendly materials and allow for a resource conserving production process; see also Beamon (1999).

Simpson and Power (2005) confirm the positive influence regarding the application of lean principles and pollution prevention actions. Packaging and related transportation are important aspects to promote sustainability, because it significantly influences the distribution of products regarding quality and loading patterns that give possibilities to reduce material usage optimizing transportation space utilization of warehouses and transportation vehicles as well as handling; see Ho et al. (2009), Sarkis (1999, 2003).

Akkerman et al. (2009) develop a MIP formulation to plan the production and distribution of prepared meals in the catering sector with the objective to maintain high quality of products while improving environmental aspects, e.g., regarding packaging and transportation, see also Figure 6.5 depicting possible conflicts or trade-offs regarding the environment, social aspects such as health and safety, and efficiency that is comparable to Figure 6.4. Akkerman et al. (2009) focus on the distribution *cold* supply chain including packaging decisions to ensure a constant temperature of meals. Energy usage and waste in the production process are further taken into account. Social aspects, e.g., health safety are indirectly regarded by providing high quality meals. As shown in Figure 6.5, there exist various trade-offs in objectives regarding sustainability and efficiency as already stated before. For instance, the quality of items may be increased by enhancing the frequency of shipments leading to decreased overall lead times, because products do not need to wait for shipping. This increases transportation frequencies and, consequently, CO_2-levels. It further increases setup times and costs and thus overall lead times, so that waiting times are shifted forward from the FGI to the production process. Packaging of items, especially chilled ones, is important regarding the time between FGI and distribution to customers after the chilling process, because it slows down the defrosting process (see Akkerman et al. (2009)) where polystyrene boxes permit better temperature retention than cardboxes, but at an expense of packaging costs and sustainability. Akkerman et al. (2009) propose a mathematical model for production and distribution planning that take into account delivery structures, packaging decisions as well as production costs where the latter are constrained by quality requirements including the maintenance of a certain temperature level of chilled products. Customer demand is presented by requested amounts as well as distribution network information that denote the frequency of deliveries that are chosen from several possible delivery structures. The objective function is reformulated in a second step to account for environmental aspects such as the production of an item with a certain temperature level, its shipping from the production facility via the distribution center to the customer, and its packaging. These aspects are measured in kg CO_2-consumption. Both objective functions might be combined to denote costs as well as CO_2-levels. However, CO_2-consumption must be translated into costs in order to combine them. Jaegler and Burlat (2012) develop a discrete event simulation model to examine supply chains and their CO_2 emissions thus determining carbon footprints for various scenarios with the aim of evaluating best practice methods to align sustainability/green performance and efficiency objectives. For measuring and controlling the carbon footprint of products across supply chains see also Sundarakani et al. (2010).

Wang et al. (2009) propose an optimization model for production planning regarding food safety risks. They integrate a risk factor related to recalls of products which requires the quantification of related risks for every production part in order to give support at critical control stages of the production process and related supplier selection. Wang et al. (2011) study a food safety risk assessment model using fuzzy set theory. A discrete event simulation environment is developed by van der Vorst et al. (2009) to study food quality changes with the aim to redesign supply chains according to increasing sustainability requirements.

Digiesi et al. (2012) present an EOQ model including the selection of transportation means that minimizes logistic and environmental costs by calculating them as a function of a *loss factor* that denotes the energy consumption related to the chosen transportation mode which is defined as the ratio of work required to overcome the frictional resistance during transport and the weight multiplied by the transportation distance. Using a numerical example, they reveal that optimal decisions on transportation means and order quantities depend on the reorder level for short transportation distances. For long distances, reorder levels affect solely optimal reorder quantities. Chaabane et al. (2011) present a methodology for designing sustainable supply chains including carbon emissions trading as regulatory constraints within classical economic objectives. They further take into account decisions on supplier and sub-contractor selections, technology investments, and transportation modes and present a MIP approach. Two objective functions are regarded where one includes classical economic objectives, i.e., the minimization of transportation costs of the overall supply chain and the minimization of overall carbon emissions. They are combined leading to a multi-objective optimization model. Abdallah et al. (21012) propose an EOQ model including shipping costs, inventory management, and reverse logistics. Swenseth and Godfrey (2002) study an EOQ with the inclusion of a cost function of shipping.

6.3 Rework and Closed-Loop Supply Chain Management

Items that passed their lifetime in the production process due to various reasons might be subject to rework actions in order to re-use them in the production process. Rework pertains to the field of GSCM and present a sustainability option. It includes all recovery actions required to transform products that do not meet pre-specified quality (anymore) in a way that they regain quality standards; see, e.g., Guide et al. (1997), Teunter and Flapper (2001), Flapper et al. (2002), Teunter and Flapper (2003), Inderfurth et al. (2005).

6.3.1 Classification and Organization

Generally, rework actions might be due to company-internal problems in production, e.g., unplanned waiting times or out-of-control production systems leading to defective items. Rework takes place before items are distributed to the customer in contrast to recycling or remanufacturing; see Flapper et al. (2002) or Pahl (2011). Besides, *remanufacturing* encompasses industrial processes where worn out products that are returned from the customer are restored to *like-new conditions* through a series of actions, e.g., disassembly, clearing, refurbishing etc.; see Clegg et al. (1995), Guide and Spencer (1997), Savaskan (2004), Zhao (2011). In contrast to this, the term *recycling* is used for the recovery of (raw) materials of used products, e.g., the copper of electronic parts. For that purpose, returned products are disassembled. Recycling and remanufacturing pertain rather to the external reverse logistics system whereas rework is related to company internal actions or processes; see also Figure 6.3 for a graphical presentation of system components. Lund (1983) provides criteria and classification for remanufacturing options regarding products; see also Teunter et al. (2006). For reviews on production planning and control for remanufacturing, the reader is refered to Guide (2000), Junior and Filho (2011).

In order to study depreciation effects and rework, *in-line* is differentiated from *off-line* and *single-stage* versus *multi-stage* rework; see Pahl (2011). Off-line production employs different production resources. Moreover, in a single-stage production-rework system, defectives or reworkables become available for rework after passing the restant production process. In a multi-stage production-rework system, items may be available after passing only a part of the production process and thus might need to revisit just some parts of the production process; see Flapper et al. (2002).

6.3.2 Literature Review of Related Models

Rework and remanufacturing issues are often approached using continuous time formulations such as EOQ/*economic production quantity (EPQ)* and its stochastic version where it is assumed that a constant fraction of production batches acquire a defect that can be adjusted reworking the items of that batch; see Pahl and Voß (2011). Deterioration may increase while parts and products wait for their turn to be reworked which increases costs, accordingly; see Teunter and Flapper (2003). Priority rules regarding reworking and regular production are developed for production planning and control; see Srivastava (2007).

Discrete dynamic models for production-inventory planning including lot sizing decisions and remanufacturing or rework options are very scarce to date; see also Teunter et al. (2006). There is an increasing number of models for reverse logistics that are formulated as network flow problems; see Srivastava (2007) as well as Fleischmann et al. (1997), Guide (2000), Dekker et al. (2004) for overviews.

Dekker et al. (2004) provide a review on reverse logistics and Flapper et al. (2002) on planning and control of rework in the process industry. Optimal replenishment and disposal policies for product recovery including lead time considerations are presented by Inderfurth (1997). Studies on stochastic effects on remanufacturing are provided by Inderfurth (2005). Moreover, Zargar (1995) develop queuing models to study the effect of rework policies on the cycle time. In general, rework increases cycle times and processing costs, but decreases wastages. He uses simulation to analyze the effectiveness of their proposed policies. His simulation study shows that for a facility with three (re-) workstations that work at a 85% utilization level decisions on rework may increase the cycle time by 10%. Therefore, policies that aim at minimizing variability and the variance resulting from different rework policies are the most effective. The planning problem of regular production lot sizes and rework lot sizes is studied by Inderfurth et al. (2005) proposing an EOQ formulation including deterioration of reworkables whereas Inderfurth et al. (2006) regard scheduling decisions implying operative (short term) planning and control rather than the tactical planning level. The same is valid for Inderfurth et al. (2007). The problem of setting inventory holding cost rates in a production environment with reverse logistics that accounts for average costs is examined by Teunter et al. (2000). Forcasting issues regarding reverse logistics are discussed in Toktay et al. (2003). A discrete dynamic and deterministic mathematical model including remanufacturing is presented by van den Heuvel (2004) extending the classical *Wagner-Whitin (WW)* model, accordingly, so that demand can also be satisfied by remanufactured returned items that have the same quality after remanufacturing. The amount of returned items that need remanufacturing is known in advance. He shows that already this problem is generally NP-hard.

Clegg et al. (1995) present a discrete-dynamic linear programming model including processes of 1) assembling new products with new parts, 2) partial disassembling and reuse returned/recycled parts of products, 3) remanufacturing including partial disassembly of returned/recycled products and replacement with new parts which may include “upgrading” remanufactured products for new functionalities and disposing replaced parts, and 4) complete disassembly for reuse or disposal of parts. The model includes the various inventory

balance equations and capacity contraints for regular production, disassembly, disposal, and assembly. Besides, the objective function regard the maximization of profits including sales revenues and total costs for overall inventory holding, disposal, partial/complete disassembly, assembly, and raw material procurement. The model allows for an analysis regarding the effects of demand variabilities, cost structures, and their changes with respect to legislation concerning remanufacturing, recycling, and disposal schemes in order to develop and implement efficient remanufacturing strategies.

A network flow formulation for a production and remanufacturing model is proposed by Golany et al. (2001). The assumed situation includes deterministic demand that can be satisfied by regular or remanufactured products in a finite macro period time horizon where three options for returned items are available, i.e., 1) remanufacture items of the returned item storage, 2) keep returned items in returned item storage, or 3) dispose returned items. The first option might require actions of inspection/testing, cleaning, disassembly, replacement of parts, reassembly to regain "as good as new" conditions of remanufactured products, and associated costs. The second option entails costs regarding transportation in/out storage spaces, special requirements for inventory holding (cooling, etc.), control/monitoring of items etc. whereas disposal causes transportation out the storage space to collection stations or landfills, disassembly of hazardous materials and distinct disposal etc. Revenues need to be regarded in case that disposed items are sold on a secondary market; see Golany et al. (2001). Deterioration or perishability effects are not included, so that returned items can be hold in inventory to an unlimited extend and used as many times as necessary or desirable.

Corominas et al. (2012) present a single product dynamic production and off-line remanufacturing planning model further regarding lost demands, overtime within certain limits, and limited storage capacities. They account for transportation costs, end-of-lifetime for estimating returns and condition-dependent remanufacturing. Credits are taken into account, too, so that the objective is to maximize net earnings where nonlinear relations are regarded between returned items and their prices. The latter is linearized using a piecewise linear function.

6.3.3 Mathematical Issues Regarding Rework and Remanufacturing

Extensions to the classical EOQ model in regard of remanufacturing and rework are provided by several authors. Remanufacturing is accounted for to the extent that the assumption is made of a flow of returned items that may be worth reworking or must be disposed, in some way increasing the storage of reworkable items. This is depicted in Figure 6.6.

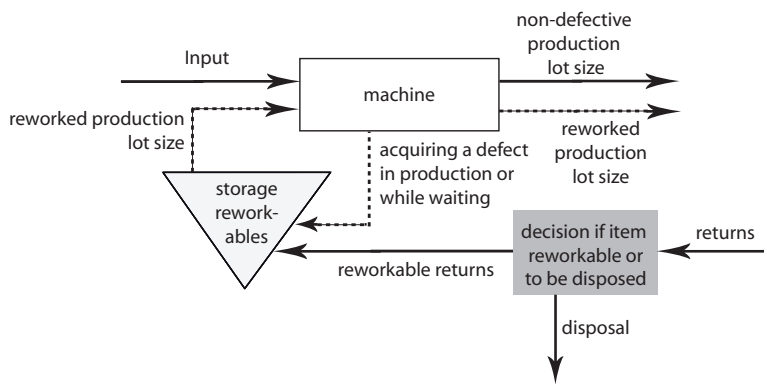

Figure 6.6: In-Line production and rework system including external returns; see, e.g., Yum and McDowell (1987), Richter and Sombrutzki (2000), Richter and Weber (2001) for similar illustrations.

The storage of reworkables is filled with items that acquire a defect in production or deteriorate while waiting and handled within the production system together with external returns of used items that might be reworked or have to be finally disposed. For that reason, a point is introduced where decisions regarding rework or final disposal of returned items are taken. Assumptions on the point of origin of reworkables raises the complexity of the system, especially when amounts of return flows are unpredictable or uncertain. If they are assumed to be deterministic, the issue is less critical, but still needs to be integrated in production planning. Besides, returns may increase inventory costs and machine utilization, but decreases replenishment of new items and, therefore, is a strategy not only in line with sustainability, but also regarding increases in profits due to decreases of procurement costs.

A discrete-dynamic and deterministic situation assuming that items are returned externally and remanufactured on the same single workstation is studied by van den Heuvel (2004) who proposes the following model formulation. The model assumes a single production workstation that produces a single product regarding a planning horizon divided into discrete macro periods. Backordering is not allowed and initial and ending inventory is assumed to be zero.

The model is stated as follows:

$$\min \sum_{t=1}^{T} (sc \cdot y_t + pc \cdot x_t + h \cdot I_t + rsc \cdot y_t^r + pc^r \cdot x_t^r + h \cdot I_t^r) \tag{6.1}$$

subject to

$$I_t = I_{t-1} + x_t + x_t^r - d_t \qquad \forall\, t = 1, \ldots, T \tag{6.2}$$

$$I_t^r = I_{t-1}^r + G_t - x_t^r \qquad \forall\, t = 1, \ldots, T \tag{6.3}$$

$$x_t \leq \sum_{t=1}^{T} d_t \cdot y_t \qquad \forall\, t = 1, \ldots, T \tag{6.4}$$

$$x_t^r \leq \sum_{t=1}^{T} d_t \cdot y_t^r \qquad \forall\, t = 1, \ldots, T \tag{6.5}$$

$$I_0 = I_0^r = I_T = I_T^r = 0 \qquad \forall\, t = 1, \ldots, T \tag{6.6}$$

$$x_t, x_t^r, I_t, I_t^r \geq 0 \qquad \forall\, t = 1, \ldots, T \tag{6.7}$$

$$y_t, y_t^r \in \{0, 1\} \qquad \forall\, t = 1, \ldots, T \tag{6.8}$$

where *pc* denotes the processing cost factor for regular production, *rsc* the rework setup cost factor, *pcr* the processing cost factor for rework processing, y_t^r states the setup variable for rework lot sizes in period t, and I_t^r the inventory holding cost factor of rework or returned items in period t.

The objective function (6.1) minimizes total costs for setups for regular production, processing, and inventory holding as well as the same total costs for returns and their remanufacturing. Equations (6.2) give the inventory balance stating that inventory in period t must be equal inventory holding from previous periods plus production in period t and remanufactured items denoted by x_t^r in that period less those items that are taken for demand satisfaction d_t. Equations (6.3) state the inventory balance regarding rework/remanufacturing requiring that inventory holding of reworked items in period t denoted by I_t^r is equal to inventory holding of reworked items in previous periods $t-1$ plus those items that are returned from an external source, e.g., customers or other supply chain partners, denoted by G_t less those items that are reworked (x_t^r) in period t. A setup is required if regular production or rework processing takes place that is invoked by Constraints (6.4) and (6.5), respectively. Initial and ending inventories of regular and reworked items are required to be zero by Equation (6.6) whereas non-negativity restrictions and definitions of binary variables are provided by Constraints (6.7) and (6.8), respectively. The model is an extension of the WW model, consequently, no capacity restrictions are included; see also Richter and Sombrutzki (2000), Richter and Weber (2001) for WW formulations with remanufacturing options. A similar model is proposed by Golany et al. (2001) for remanufacturing. They propose the employment of the following relation to demand for a single product with a fixed τ-period cycle of utilization and a fixed non-usability fraction α^{Θ}:

$$G_t = (1 - \alpha^{\Theta}) \cdot D_t \qquad \forall\, t = 1, \ldots, T \tag{6.9}$$

The advantage lies in the substitution of input parameters to the model. Furthermore, choosing carefully the relation between demand and return flows can lead to realistic practical situations, so that the model might be employed; see Golany et al. (2001). Estimates for return flow are determined in practice by using forcasting techniques that are implemented in APS and MRP II software solutions; see also Guide and van Wassenhove (2009). Zhao (2011) use forcasting techniques based on regression analysis of historical data in their stochastic dynamic model. Yang et al. (2005) give a proof that the model is NP-hard in case

of all cost factors being stationary. They propose a heuristic for the problem with minimum concave costs.

Teunter et al. (2006) study a joint setup cost model for regular production and remanufacturing as well as a separate cost model for both production types abstracting from capacity restrictions. Their second model is equal to the model of van den Heuvel (2004). They propose modified heuristics regarding *Silver-Meal*, *Least Unit Cost*, and *Part Period Balancing* and compare their performance. They further mention that a disposal option for returned items might not lead to decreased costs if the return rate is not significantly high in relation to the demand rate, i.e., over 90% while the demand rate is rather low.

Chapter 7

Models Incorporating Load-Dependent Lead-Times

There are various models accounting for the nonlinear relationship between lead times and utilization or workload of production systems. Performance-related models also named production-performance models (PP)[1] take this relationship into account by determining lead times according to the workload of the system; see, e.g., the models of Lautenschläger (1999), Voß and Woodruff (2006); see Section 7.1.

Another branch of models differentiate between WIP and FGI that explicitely represent WIP flows through a production resource, so that output is related to some measure of WIP levels within the planning period. The nonlinear relationship between these decision variables that might be subject to congestion is captured by related constraints in the models. We refer to these models as *release-production* (RP) models. They use *clearing functions* (CFs) to integrate the relation between WIP, lead times, and FGI (system output); see Missbauer and Uzsoy (2011). These models are discussed in Section 7.2. An evaluation of all presented approaches is given in Section 7.3.

7.1 Production-Performance Models

Classical discrete dynamic production and lot sizing models that pertain to PP models are triggered by known (deterministic) demand in discrete time periods that must be fulfilled. That means that products must be produced in time, but as late as possible to avoid increasing inventory holding costs. In case of complex BOM structures or limited nominal capacity, parts may be produced ahead. They are produced as late as possible to prevent higher inventory holding costs, as well. Such systems are also referred to as pull systems; see Section 5.2 which is valid for MtO environments. Consequently, order releases are triggered by externally given demand, so that they are not decision variables. This directly influences capacity utilization/workload and thus lead times. Utilization of capacities may be, to a limited extend, subject to management decisions, so that mathematical models may contain some constraint to limit high resource utilization; see, e.g., Section 8.4. Besides, capacity constraints are modeled as dependent on the workload, so that production is limited, accordingly.

7.1.1 The Model of Lautenschläger

The PP model of Lautenschläger (1999) restricts production regarding increased utilization, so that the amount of production that can be produced in a period is limited according to

[1] The name has been proposed by Lautenschläger (1999).

resource utilization. Increasing utilization permits the production of a decreasing amount of products in a period, so that a specific amount is shifted forward to previous periods. At 100% utilization, a predefined maximal amount of products is shifted foreward; see Lautenschläger (1999). Lautenschläger (1999) distinguish WIP/FGI regarding *physical production* and the *availability* of products where physical production takes place in previous periods, so that items are available only in subsequent periods. This can be compared with "inventory reservation" where parts are "booked" or assigned to FGI orders. The resulting plan of the model of Lautenschläger (1999) gives inventory holding as a consequence of system utilization, workload, or performance that becomes the evaluation variable of the system. In the following, the model of Lautenschläger (1999) is examined in more details.

7.1.1.1 Assumptions of the Basic Model of Lautenschläger

The model of Lautenschläger (1999) is based on the classical assumptions of the *multi-level capacitated lot sizing model (MLCLSP)* that are modified and extended to incorporate capacity-related forward shifting (see also Derstroff (1995)) that are stated as follows:

1. A discrete dynamic model is developed subject to deterministic data within a finite planning horizon that is divided in T discrete time periods of equal length ("time buckets;" see Suerie and Stadtler (2003), Suerie (2005));
2. Multiple products i with $i = 1, \ldots, N$ that might be linked via a BOM given by a matrix of predecessor/successor and denoted by a_{ij} are produced on production resources w with $w = 1, \ldots, W$ subject to capacity constraints given by parameter Cap_{wt} for production resource w in period t;
3. Customer demand D_{it} for FGI of product i in period t is deterministic and varies from period to period;
4. Backlogging is not allowed;
5. Initial and ending inventory holding can be pre-defined and need not necessarily be zero; see also Stadtler (1996);
6. Inventory holding costs are proportional to inventory holding at the end of the regarded time period;
7. A difference is made between parts and products in the production system that are phyically produced and their "disposition;" see also Figure 7.4 on Page 51. This is expressed and modeled by different production modes denoted $m = 1, 2$ where $m = 1$ denotes both production and disposition in period t. The second mode $m = 2$ denotes pre-production in period $t - 1$ for disposition and demand satisfaction in period t or subsequent periods. Due to the distinction of physical and "disponible" WIP inventory, the inventory holding costs are attributed to physical inventory using an echelon inventory holding cost factor.
8. Variable per unit production and (raw) material costs a assumed to be constant;
9. Every product i can be produced in every time period t;
10. Setup costs and times take place in period t if product i is produced. Setup states are not carried over to the next period.
11. The objective of the basic model is to minimize inventory holding and setup costs;
12. The capacity constraint, also named FS-curve, defines the relation between production and utilization of a production resource w is assumed to be linear.

7.1.1.2 Mathematical Formulation

Based on the above stated assumptions, the *multi-level capacitated lot sizing model with forward shifting times (MLCLSP-FS)* of Lautenschläger (1999) is stated as follows:

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T} sc_i \cdot y_{it} + \sum_{i=1}^{N} T \cdot h_i \cdot I_{i0} + \sum_{i=1}^{N}\sum_{t=1}^{T} (T-t+1) \cdot \left[h_i^e \cdot x_{it}^{\text{total}} - h_i \cdot D_{it}\right] \tag{7.1}$$

subject to

$$x_{it}^{\text{total}} = x_{it}^{m=1} + x_{i,t+1}^{m=2} \qquad \forall\, i,t \tag{7.2}$$

$$I_{it} = I_{i,t-1} + \sum_{m=1}^{M} x_{it}^{m} - \sum_{j=1}^{N} a_{ij} \cdot x_{jt}^{\text{total}} - D_{it} \qquad \forall\, i,t \tag{7.3}$$

$$\sum_{i=1}^{N} \xi_{iw} \cdot x_{it}^{\text{total}} \leq \text{Cap}_{wt} \qquad \forall\, w,t \tag{7.4}$$

$$\sum_{i=1}^{N} \xi_{iw} \cdot x_{it}^{m=1} \leq \alpha_w \cdot \text{Cap}_{wt} - \beta_w \cdot \sum_{i=1}^{N} \xi_{iw} \cdot x_{it}^{\text{total}} \qquad \forall\, w,t \tag{7.5}$$

$$x_{it}^{\text{total}} \leq V \cdot y_{it} \qquad \forall\, i,t \tag{7.6}$$

$$x_{it}, x_{it}^{\text{total}}, x_{it}^{m} \geq 0 \qquad \forall\, i,t \tag{7.7}$$

$$y_{it} \in \{0,1\} \qquad \forall\, i,t \tag{7.8}$$

The objective function (7.1) minimizes the costs for setups, initial (physical) inventory holding at the beginning of the time horizon, echelon inventory holding costs due to production of (physical) items (x_{it}^{total}) less inventory due to customer demand D_{it}. Equations (7.2) state the total disponible production quantity of product i in period t which is composed by the sum of produced items in period t for disposition and demand satisfaction in period t (mode $m = 1$) and the amount of items produced in period t for use in period $t+1$ (mode $m = 2$) thus produced in advance in period t. The inventory balance equations presented by (7.3) state that inventory of product i in period t is composed by inventory of product i of the previous period $t-1$ plus the sum of produced items i in period t less the sum of items i required to produce (immediate successor) item j in period t and less customer demand of item i in period t. Inequalities (7.4) require that the capacity used to produce the sum of items i in period t is less or equal operational capacities Cap_{wt} of production resource w in period t.[2] Inequalities (7.5) restrict production amounts of item i in period t produced in mode $m = 1$ where α_w and β_w are the shape and the slope of the FS curve for production resource w, respectively. In other words, production in mode $m = 1$ is restricted by the operational capacity that is not used by production and setups. Lautenschläger (1999) assumes that operational production in t is already consumed by "booked" production, thus production in mode $m = 2$ that is shifted forward in $t-1$. Therefore, production in mode $m = 2$ can be interpreted as WIP where releases take place already in $t-1$ as in the models of Asmundsson et al. (2002, 2006a); see also Missbauer and Uzsoy (2011). The total amount of production is linked to setups via Inequalities (7.6). Inequalities (7.7) denote the non-negativity constraints and Constraints (7.8) require the setup variable y_{it} to be binary. In the following, we examine in more details the FS capacity constraint.

7.1.1.3 Determination of Parameters for the Production Forward-Shifting Curve

Lautenschläger (1999) determines the course of the FS curve by two unitless parameters b_w and e_w where b_w denotes the utilization of a workstation w when forward shifting of

[2] The term *operational capacity* is discussed and defined in Section 5.3.

production to $t-1$ starts. The parameter e_w gives the part of production that is shifted forward at 100% utilization; see Figure 7.1.

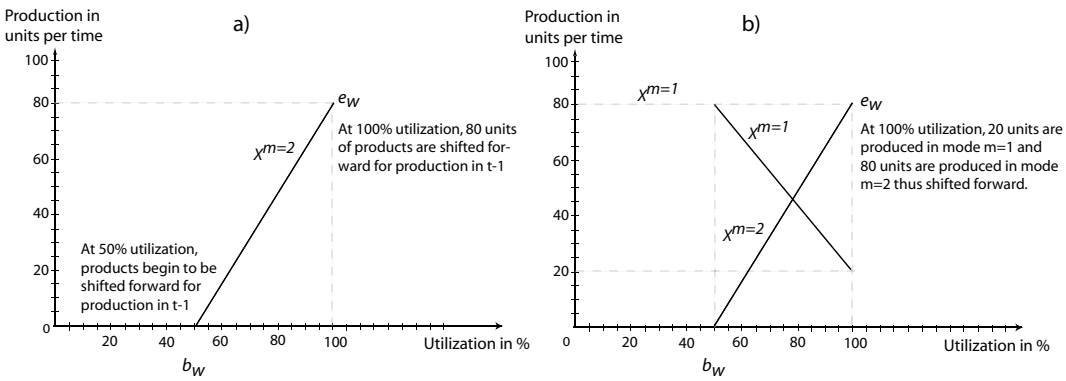

Figure 7.1: Example of a forward shifting curve; modified from Lautenschläger (1999).

These parameters serve to determine an appropriate or assumed FS function with parameters α_w and β_w that limit production in mode $m=1$, thus production in period t available in t given in Inequalities (7.5). Parameter b_w is defined for an interval $]0,1[$ and parameter e_w for an interval $]0,1]$; see Lautenschläger (1999). In order to determine parameters b_w and e_w, Inequalities (7.5) are simplified assuming a single product, time period and workstation:

$$\xi x^{m=1} \leq \alpha \cdot \text{Cap} - \beta \cdot \xi x^{\text{total}} \tag{7.9}$$

The capacity utilization of a production resource is given by

$$\rho = \frac{\xi \cdot x^{\text{total}}}{\text{Cap}},$$

so that the shifting forward constraint (7.9) can be reformulated as

$$x^{m=1} \leq \frac{\text{Cap}}{\xi}\left[\alpha - \beta \cdot \rho\right].$$

As stated in Equation (7.2), the overall sum of production is composed by the production of items manufactured in mode $m=1$ thus without being shifted forward and items shifted forward and thus produced in mode $m=2$. Consequently, the part of items that need to be shifted forward due to a certain level of utilization is given by

$$\begin{aligned} \text{part}^{\text{FS}} &= \frac{x^{m=2}}{x^{\text{total}}} = 1 - \frac{x^{m=1}}{x^{\text{total}}} \\ \rightarrow \frac{x^{m=2}}{x^{\text{total}}} &= 1 - \frac{\alpha}{\rho} + \beta \end{aligned} \tag{7.10}$$

and the part of production that is produced in mode $m=1$ is reduced, accordingly

$$\frac{x^{m=1}}{x^{\text{total}}} = \frac{\alpha \cdot \text{Cap}}{x^{\text{total}}} - \beta \tag{7.11}$$

where these parts follow a nonlinear function. This becomes clear as shown in the upper graph in Figure 7.2.

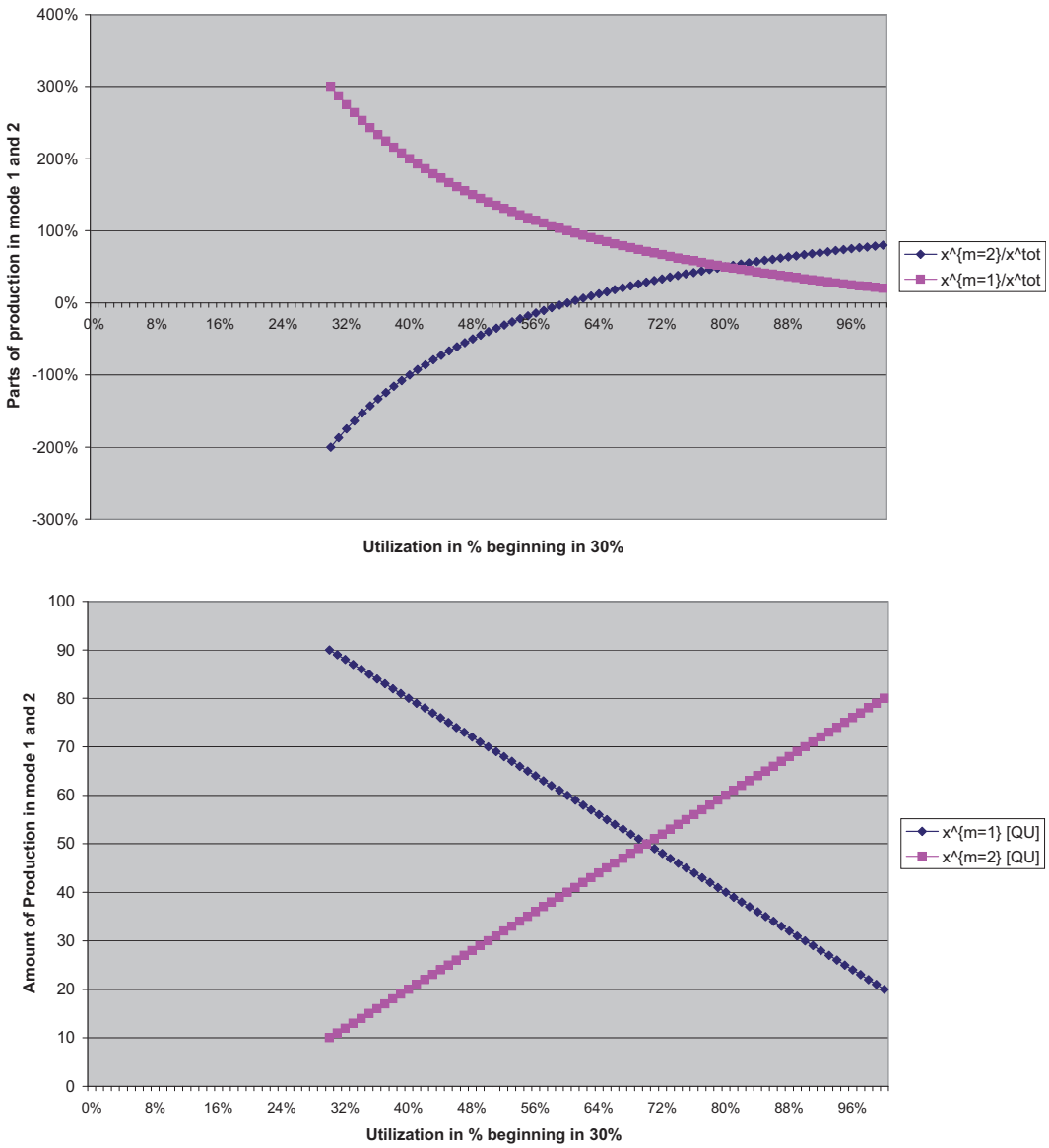

Figure 7.2: Parts and amounts of production in mode $m = 1$ and $m = 2$ in % for $\alpha = 1,2$ and $\beta = 1,0$ with parameters $e = 0,8$ and $b = 0,6$ for utilization values $30 - 100\%$.

Assuming that at 100% utilization the part of items that are shifted forward is denoted by e (see also Figure 7.1) and that the shifting forward of items begins at an utilization given by b, the following two inequalities can be formulated:

$$\frac{x^{m=2}}{x^{\text{total}}}(b) = 0 \qquad \Rightarrow 1 + \beta - \frac{\alpha}{b} \leq 0$$

$$\frac{x^{m=2}}{x^{\text{total}}}(1) = e \qquad \Rightarrow 1 + \beta - \alpha \leq e$$

Hence, the parameters of the FS curve are calculated with

$$\alpha = \frac{e \cdot b}{1 - b} \qquad \forall\, b \geq \rho; \text{provided } b \neq 1 \tag{7.12}$$

$$\beta = \frac{e + b - 1}{1 - b} \qquad \forall\, b \geq \rho; \text{provided } b \neq 1. \tag{7.13}$$

Nevertheless, as the parameter b denotes the utilization when forward shifting of production begins, the parameters and thus the FS curve are defined begining at an utilization denoted by b. Therefore, the amounts of production in mode $m = 1$ and $m = 2$ are determined for each utilization value

$$x^{m=1} = \frac{\text{Cap}}{\xi}\left[\alpha - \beta \cdot \rho\right] \tag{7.14}$$

$$x^{m=2} = x^{\text{total}} - x^{m=1} \tag{7.15}$$

where the amount of $x^{m=2}$ increases linearly with an increasing utilization and $x^{m=1}$ decreases, accordingly. These definition of the parameters α and β present a linear FS graph; see also Figure 7.2. Table 7.1 gives parameter combinations on which α and β are calculated.

Table 7.1: Parameter combinations to determine FS curves

		e values		
b values	FS values	0.6	0.8	1.0
0.3	α	0.26	0.34	0.43
	β	−0.14	0.14	0.43
0.6	α	0.90	1.20	1.50
	β	0.50	1.00	1.50
0.8	α	2.40	3.20	4.00
	β	2.00	3.00	4.00

The restrictions regarding production in mode $m = 1$ might lead to problems finding a solution for data sets, especially those with an increased demand in initial periods. The model of Lautenschläger (1999) shifts forward production in relation to resource utilization. However, it does not prevent high utilization in any time period as forward shifting of production leads to high utilization up to 100% in previous periods. Therefore, it addresses the problem, but does not really solve it; see also Section 8 for a more detailed analysis on forward shifting and production modes.

7.1.1.4 Integration of Overtime

The integration of overtime allows for flexibility in case of high utilization of production resources. Lautenschläger (1999) integrates overtime by modifying the capacity restriction given in (7.5) as follows:

$$\sum_{i=1}^{N} \xi_{iw} \cdot x_{it}^{m=1} \leq \text{Cap}_{wt} + O_{wt} \qquad \forall\, w,t \tag{7.16}$$

where O_{wt} is the overtime of production resource w in period t. The objective function presented in (7.1) is extended, accordingly, to account for increased costs due to overproduction by the term $\sum_{w=1}^{W}\sum_{t=1}^{T} oc_w O_{wt}$, $\forall\, w,t$. Figure 7.3 depicts the case that 70% of products are shifted forward in case that utilization increases by 40% to 140%.

In order to capture the additional amount of items that are produced, a new production mode $m = 3$ is defined that is positive in case that the FS curve would restrict an increase over 100% utilization. Thus, the increase of the amount of produced items due to overtime is given by $x_{it}^{m=3}$ and added to the overall production amount given in (7.2) as follows

$$\check{x}_{it}^{\text{total}} = x_{it}^{m=1} + x_{i,t+1}^{m=2} + x_{i,t+1}^{m=3} \qquad \forall\, i,t. \tag{7.17}$$

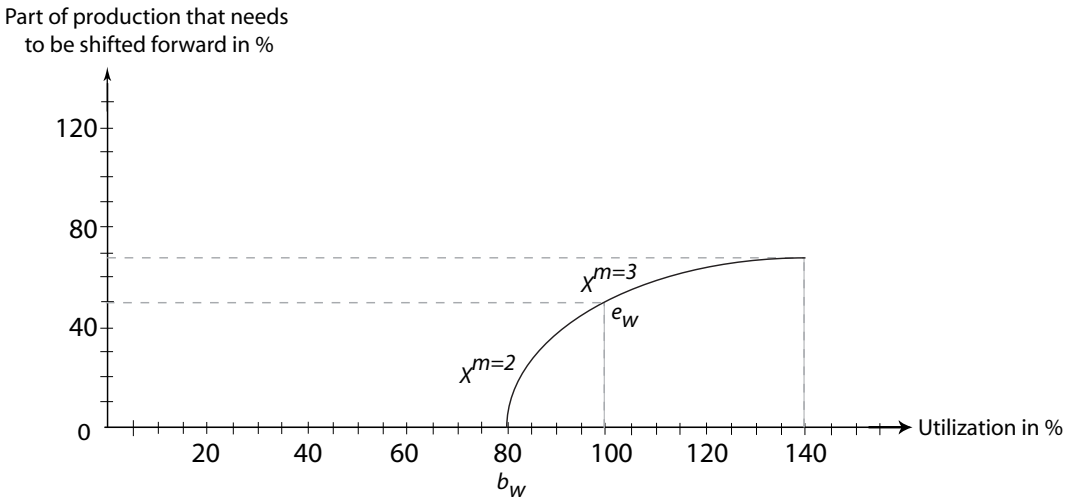

Figure 7.3: Extension of the FS curve to overtime.

The inventory balance equation presented in (7.3) is modified including the additional production amount:

$$I_{it} = I_{i,t-1} + \sum_{m=1}^{M} x_{it}^{m} + x_{it}^{m=3} - \sum_{j=1}^{N} a_{ij} x_{it}^{\text{total}} - D_{it} \qquad \forall\, i,t \tag{7.18}$$

The nonlinear FS curve is employed that can be approximated by using n straight lines and, consequently, breakpoints to define the FS curve presented as follows:

$$\sum_{i=1}^{N} \xi_{iw} x^{m=1} \leq \alpha_{wn} \cdot \text{Cap}_{wt} - \sum_{i=1}^{N} \beta_{wn} \cdot \xi_{iw} x_{it}^{\text{total}} \qquad \forall\, w,n,t \tag{7.19}$$

Additional items due to overtime that are shifted forward need to be stocked and thus cause additional inventory holding costs, so that the model should prefer for an optimal solution to increase the amount of items produced in mode $m = 1$ in order to avoid inventory costs.

7.1.2 LDLT by Voß and Woodruff (2004)

Woodruff and Voß (2004) regard LDLT together with routing alternatives in an aggregate planning model. Routing alternatives and choices regarding subcontractors are defined for each part, also denoted *stock keeping unit* (SKU), via input data denoting related substitutes for items. Alternatives are grouped according to the master SKU which is the only FGI item subject to external demand. Estimates of lead times are integrated using piecewise estimates of utilization, so that the differences in lead times from one breakpoint of this piecewise function to the next is one macro period; see Woodruff and Voß (2004). They discretize the utilization-lead time curve using integer variables to formulate constraints that denote which segment or breakpoint of the curve is active in a time period. Estimated lead times are integrated in the model using the following terms:

$$L_t = \sum_{r=1}^{R} p_{irt} \cdot L_{ir} \qquad \text{with} \qquad \sum_{r=1}^{R} p_{irt} = 1 \qquad \forall\, i,t \tag{7.20}$$

The Equations (7.20) denote active lead time variables L_{ir} for a SKU i at breakpoint r where the selection variable p_{irt} sums up to one modeled as *special ordered set type 1 (SOS1)*-variables. Furthermore, Woodruff and Voß (2004) add the constraint

$$\sum_{r=1}^{R} BP_{ir} \cdot p_{irt} \geq \rho_{it}(x) \qquad \forall\, i,t = 1,\ldots,T \tag{7.21}$$

where BP_{ir} is the set of utilization breakpoints for SKU i and lead time breakpoints r. The approximation to resource utilization $\rho_{it}(x)$ is stated with

$$\rho_{it}(x) \equiv \sum_{i=1}^{N} (\xi_{iw_i} x_{it} + st_{iw_i} y_{it}) \qquad \forall\, i,t = 1,\ldots,T \tag{7.22}$$

where the variable $\rho_{it}(x)$ denoted resource utilization by Woodruff and Voß (2004) gives the planned used capacity resources. In the following, the term "utilization" is rather used as the planned employed resource time divided nominal capacity in a time period. Inequalities (7.21) require that the lead time selection takes place according to the workload. The index w_i denotes the bottleneck of the production system dependent on SKU i. Finally, lead times are prevented to be longer than one time period by

$$\sum_{r=1}^{R} p_{ir,t+1} \cdot L_{ir} - \sum_{r=1}^{R} p_{irt} \cdot L_{ir} \leq 1 \qquad \forall\, t = 1,\ldots,T-1 \tag{7.23}$$

This constraint prevent that production in a period could be modeled as being completed after job releases in the next period; see Woodruff and Voß (2004). The complete model is stated as follows:

$$\min \sum_{t=1}^{T} \sum_{i=1}^{N} \left[sc_i y_{it} + \left(hw_i \cdot \sum_{r=1}^{R} p_{irt} + ac_i \right) \cdot x_{it} + h_i I_{it}^{+} + bc_i \cdot I_{it}^{-} \right] \tag{7.24}$$

subject to

$$I_{irt}(x,y) + \sum_{u=1}^{t} d(i,u) \geq (p_{irt} - 1) \cdot V \qquad \forall\, i,r,t \tag{7.25}$$

$$\sum_{i=1}^{N} (\xi_{ij} x_{it} + st_{ij} y_{it}) \leq 1 \qquad \forall\, t,j \tag{7.26}$$

$$y_{it} \geq \frac{x_{it}}{V} \qquad \forall\, i,t \tag{7.27}$$

$$\sum_{r=1}^{R} p_{irt} = 1 \qquad \forall\, i,t \tag{7.28}$$

$$\sum_{r=1}^{R} BP_{ir} \cdot p_{irt} \leq \rho_{it}(x) \qquad \forall\, i,t \tag{7.29}$$

$$\sum_{r=1}^{R} p_{irt} \cdot L_{ir} \geq \sum_{r=1}^{R} p_{ir,t+1} \cdot L_{ir} - 1 \qquad \forall\, t = 1,\ldots,T-1 \tag{7.30}$$

$$p_{irt} \ \text{ SOS1 across } r \qquad \forall\, i,t \tag{7.31}$$

$$I_{a,t}^{+} = 0 \qquad \forall\, t, a \in A_i \tag{7.32}$$

$$I_{a,t}^{-} = 0 \qquad \forall\, t, a \in A_i \tag{7.33}$$

$$I_{it}^{+} - I_{it}^{-} - I_{it}(x,y) \geq (1 - p_{irt}) \cdot V \qquad \forall\, i,r,t \tag{7.34}$$

$$I_{it}^{+} - I_{it}^{-} - I_{it}(x,y) \leq (p_{irt} - 1) \cdot V \qquad \forall\, i,r,t \tag{7.35}$$

$$y_{it} \in \{0,1\} \qquad \forall\, i,t \tag{7.36}$$

$$I_{it}^{+}, I_{it}^{-} \geq 0 \qquad \forall\, t, a \neq A \tag{7.37}$$

where inventories $I_{it}(x,y)$ for SKUs i employed in Inequalities (7.25) are determined provided that $A_i \neq 0$

$$I_{it}(x,y) \equiv \sum_{u=1}^{t} \sum_{a \in A_i} x_{a,t-L_a} + I_{i,0} - \sum_{u=1}^{t} \left(d_{iu} + \sum_{j=1}^{N} a_{ij} \cdot x_{ju} \right) \qquad \forall\, i,t = 1,\ldots,T \tag{7.38}$$

The objective function (7.24) minimizes the sum of setup costs, WIP holding costs, FGI or master SKU holding costs, and backordering costs. The latter also expresses tardiness costs, so that waiting time-independent backordering is allowed. WIP holding costs are enhanced by marginal costs ac_i using an alternative SKU i. The objective function includes WIP holding costs based on Little's Law extracted as follows:

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\left(hw_i \cdot \sum_{r=1}^{R} p_{irt} \cdot L_{ir}\right) \cdot x_{it} \tag{7.39}$$

Inequalities (7.25) require that inventory and the sum of external demands are greater than the decision variables $p_{irt} - 1$ multiplied by a big number, so that related lead times are prevented from jumping down more than one period. Inequalities (7.26) force the sum of processing and setup times for SKU i in period t to be smaller or equal 100% ressource utilization. Inequalities (7.27) require binary (see (7.36)) setup variables to be greater than zero in case of production, thus $x_{it} \geq 0$. Inequalities (7.28), (7.29), and (7.30) state the constraints regarding LDLT that prevents the lead time from being longer than one time period. This would lead to passing of jobs, so that production of a SKU in one time period could be modeled as completed after its release in the next period; see Woodruff and Voß (2004). Equalities (7.32) and (7.33) require zero production for master SKUs that is given for each substitution group through related input data. Inequalities (7.34) and (7.35) restrict backorders and inventory holding, respectively. They form together with Statement (7.38) an inventory balance. Inventory for SKU i in period t is given by the sum of production of SKU-alternative a in time period $t - L_a$ where L_a denotes the estimated lead time of a SKU-alternative a, initial inventories of SKU i in time period $t = 0$ less external demand for SKU i up to period $u = 1, \ldots, t$ and SKUs i that are needed to build one unit of SKU j. Inequalities (7.37) and (7.36) give non-negativity constraints on inventory and setup variables, respectively.

In classical LP-formulations, it is frequently assumed that production equals releases of period $t - L_i$, thus $x_{it} = R_{i,t-L_i}$ where L_i denotes some fixed lead time (Missbauer and Uzsoy (2011)) integrated using index transformation. The objective function of Woodruff and Voß (2004) becomes nonlinear due to $p_{irt} \cdot x_{it}$ where p_{irt} are decision variables modeled as SOS1, so that the model becomes hard to be computationally solved; see also Missbauer and Uzsoy (2011).

7.2 Release-Production Models

Replease-production (RP) models also denoted WIP-based models (see Missbauer and Uzsoy (2011)) differentiate between production of parts that are in the system as WIP and FGI that are available for demand satisfaction. These are captured by different inventory balance equations in mathematical model formulations. While FGI is subject to planning and the output of the planning process, WIP depends on releases and is significantly influenced by the realized lead times; see Section 5.1 and the (throughput) performance of the production resource; see Missbauer and Uzsoy (2011). Note that Missbauer and Uzsoy (2011) refer to this as the *flow time*. These models relate resource output in a planning period to some measure of WIP levels at the resource during this time period. For instance, Figure 7.4 depicts an order release of 20 units of product A in period $t = 0$ of a discrete time scale ("time bucket"; see Suerie and Stadtler (2003)). In period $t = 4$, 60 units WIP of product B are released, so that a total of 80 units of parts for products are present in the production system. Physical production starts in period $t = 9$ producing 20 units of product A and, in period $t = 15$, 60 units of product B that are put in stock to serve demand in a later period. In period $t = 20$, 15 units of product A are taken from FGI inventory to satisfy demand. In period $t = 25$, the

remainder of FGI of product A and B satisfy demand, so that the production system including inventory is emptied. The amount of WIP inventory or "released orders" in the system waiting for production is the *descriptive or evaluation variable* of system performance; see Lautenschläger (1999).

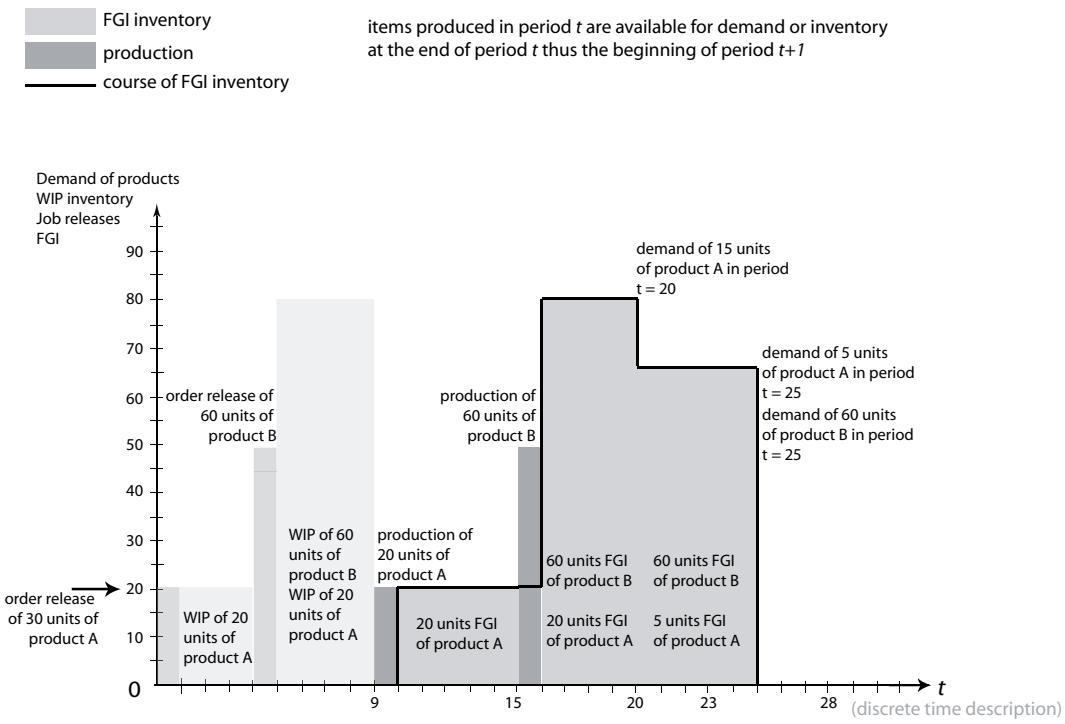

Figure 7.4: Course of WIP and FGI inventory.

CFs present the expected output of a production resource or, in case of aggregate production resources, workcenters, as a function of a suitable measure of aggregate WIP of all regarded products, e.g., the average WIP over a period, two or more periods, or total available WIP or load meaning the sum of initial WIP of the planning horizon and WIP resulting in that period; see Missbauer and Uzsoy (2011). For instance, Asmundsson et al. (2002, 2006a) use a kind of average production quantity to get some WIP measure, i.e., $1/2(x_{it} + x_{i,t-1})$. LDLT are taken into account by the nonlinear clearing function and are calculated using Little's Law, but not explicitely modeled, so that the problem of assuming some fixed lead time for use in, e.g., index transformations, is avoided. Missbauer and Uzsoy (2011) note that CFs resemble some concepts given in dynamic traffic assignment literature that is interesting when thinking of lot sizing models formulated as *network flow problem (NFP)*; see also Carey and Subrahmanian (2000), Carey (2001), Nahapetyan and Lawphongpanich (2007).

7.2.1 The Concept of Clearing Functions

The concept of clearing functions has been introduced by Graves (1986) using a constant proportional factor to denote the clearing of WIP of a production resource in relation to its utilization. Graves (1986) uses a linearly increasing relation that allows for unlimited capacity. This concept is modified by Karmarkar (1989) and Srinivasan et al. (1988) who regards the nonlinear relation between lead times and workload/utilization employing "clearing factors" $\alpha(W)$ that are nonlinear functions of WIP yielding the following form:

$$\text{Capacity} = \alpha(W) \cdot W = f(W) \tag{7.40}$$

where $f(\cdot)$ denotes the clearing function that can be derived using a variety of functional forms or relations, e.g.:

$$g_1(W) = \frac{k_1 \cdot W}{k_2 + W} \tag{7.41}$$

$$g_2(W) = k_2(1 - e^{-k_2 \cdot W}). \tag{7.42}$$

The first functional form has been proposed by Karmarkar (1989) and the second form by Srinivasan et al. (1988) where k_1 denotes maximum capacity and k_2 gives the slope of the curve, i.e. the degradation of productivity; see Figure 5.4 for a presentation of a clearing function. Smaller values of k_2 lead to steeper curves; see Figures 7.5 and 7.6.

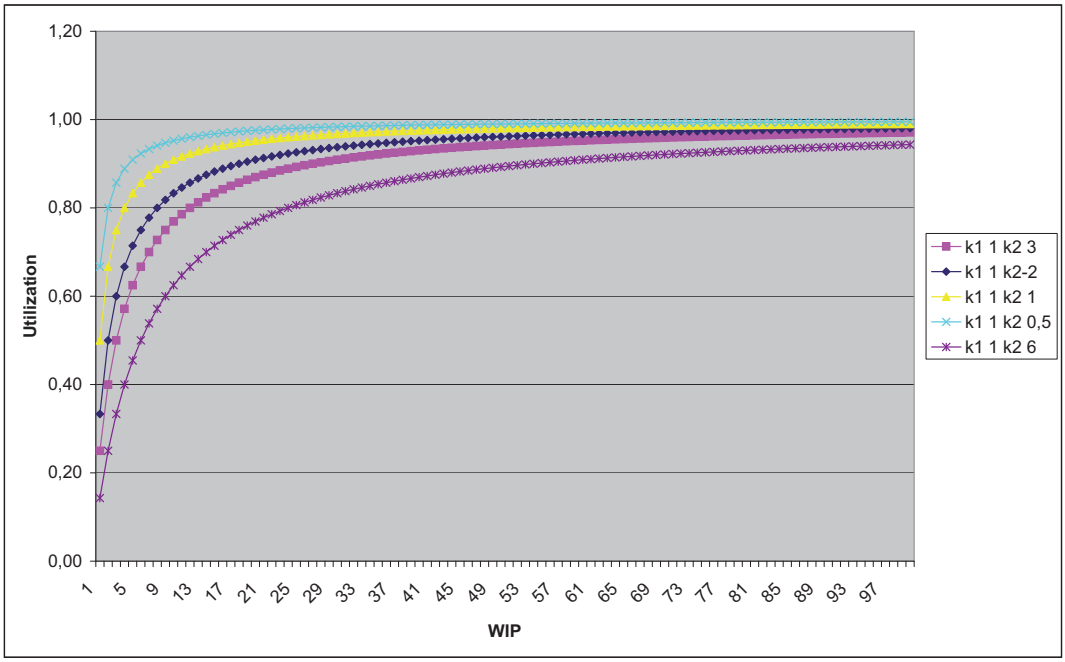

Figure 7.5: Clearing function of Karmarkar (1989) for different k_2-values.

Elmaghraby (2011) proposes another functional form in order to mitigate the effect that capacity utilization approaches 100% only in case of an infinite WIP, or, in other words, a significantly high amount of WIP in the system

$$g_3(W) = k_1 W^a e^{-bW} \qquad \text{with } a, b > 0,\ 0 \leq W \leq \overline{W} < \infty \tag{7.43}$$

where $\overline{W}$ denotes a finite upport bound on WIP, thus $W^{\text{opt}} = a/b$, so that a maximum value is reached and does not increase infinitely high while approaching operational capacity. Various researchers integrate this concept in their optimization models, e.g., Asmundsson et al. (2002, 2003, 2006b), Hwang and Uzsoy (2004), Kim and Uzsoy (2006), Missbauer (1998, 2006b), Caramanis and Anli (1999), Orcun et al. (2007), Riaño (2002), Selcuk et al. (2008).

7.2.2 Lot Sizing Model of Karmarkar (1989)

Karmarkar (1989) uses a nonlinear clearing function to represent the output as a function of the average WIP in the production system. As pointed out by Karmarkar (1989), the form of the curve is also valid for synchronous deterministic flow lines with batched flows.

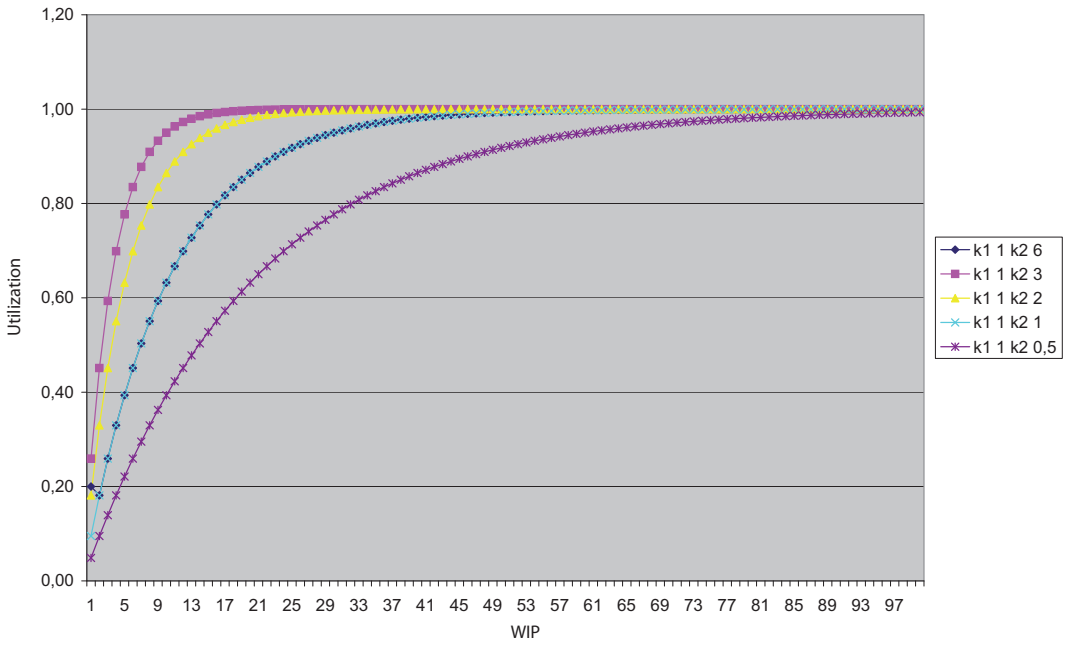

Figure 7.6: Clearing function of Srinivasan et al. (1988) for different k_2-values.

Assuming production of a single product on a single production resource and abstracting from multiple periods, he proposes the following function for the actual or *operational* production rate p:

$$p = \frac{x}{C} \tag{7.44}$$

Karmarkar (1989) defines the cycle time C as the sum of setup times and production processing times regarding production amount x divided by the nominal production processing rate, thus

$$C = \frac{st + x}{\dot{p}} \tag{7.45}$$

where $\dot{p}$ denotes the nominal processing rate in a sense of maximal or best processing rate that does not include stochastic negative influences and states the rate at which a machine operates at its less resource-consuming level, e.g., regarding energy and other auxiliary production means. Moreover, Karmarkar (1989) substitutes lot sizes x in terms of WIP over all workstations thus using $x = W/N$ where N the sum of workstations. The formulation for the operational production rate p is given by

$$p = \frac{\dot{p} \cdot W}{N \cdot \dot{p} \cdot st + W}. \tag{7.46}$$

The utilization for a $M/M/1$ queuing system is stated as follows (see also Buzacott and Shantikumar (1993), Karmarkar (1989), Zijm and Buitenhek (1996)):

$$\rho = \lambda p = \frac{p}{\dot{p}} + \frac{p \cdot st}{x} \tag{7.47}$$

Therefore, the average lead time $\overline{L}$ can be formulated with

$$\overline{L} = \frac{st + \frac{x}{\dot{p}}}{1 - \frac{p}{\dot{p}} - \frac{p \cdot st}{x}} \tag{7.48}$$

abstracting from setup times st to keep things simple for deriving the CF, writing total WIP as $W = \overline{L}p$, and substituting $\overline{L} = W/p$ results in

$$p = \frac{p \cdot W}{x + W}. \tag{7.49}$$

This CF can be depicted as shown in Figure (7.5) assuming the form given in (7.41). Karmarkar (1989) implements the CF given in (7.49) in a discrete dynamic optimization model assuming a single product. The model is stated with

$$\min \sum_{t=1}^{T} (hw \cdot W_t + h \cdot I_t + pc \cdot x_t + \vartheta \cdot B_t + rc \cdot R_t) \tag{7.50}$$

subject to

$$W_t = W_{t-1} + R_t - x_t \qquad \forall\, t \tag{7.51}$$

$$I_t = I_{t-1} + x_t + B_{t-1} - d_t \qquad \forall\, t \tag{7.52}$$

$$x_t \leq f(W_{t-1}, R_t, p_t) \qquad \forall\, t \tag{7.53}$$

$$x_t, R_t, W_t, I_t, B_t \geq 0 \qquad \forall\, t \tag{7.54}$$

where the objective function given in (7.50) minimizes the sum of costs for WIP inventory holding, FGI inventory holding, production of amount x_t in period t, backordering, and releases. Release costs encompass, e.g., costs for raw materials. Equations (7.51) denote the WIP balance equation and (7.52) the FGI inventory balance equation regarding FGIs. Inequalities (7.53) denote the CF that is dependent on the WIP of previous periods $t-1$, releases in t, and the processing rate in period t. Inequalities (7.54) state the usual non-negativity constraints on decision variables. Missbauer and Uzsoy (2011) propose the use of the following CF form in such a deterministic model

$$\sum_{t=1}^{T} \xi_t x_t \leq f_t \left(\sum_{t=1}^{T} \xi_t \hat{W}_t \right) \qquad \forall\, t \tag{7.55}$$

where $\hat{W}_t$ denotes a WIP measure given, e.g., as $\hat{W}_t = W_{t-1} + R_t$ that might be expressed by time units or workload. If multiple products are modeled, setup times and costs need to be integrated into the model and the CF modified, accordingly. A multi-product model is provided by Asmundsson et al. (2003, 2006a, 2009), abstracting from setup times and costs assuming them as negligible.

7.2.3 Model of Asmundsson et al. (2003)

Asmundsson et al. (2006a, 2009) employs a clearing function to determine the nonlinear relationship between lead times and workload as Karmarkar (1989). They use a G/G/1 queuing model for derivation of WIP and resource utilization and special properties of the CF to formulate a *linear programming (LP)* model, so that the model remains computational tractable. In the literature, there are two methods available for the determination of the CF. The first is the analytical derivation from queuing network models and the second is the empirical approximation using functional forms that can be fitted to empirical data. Because of the size of practical systems the complete identification of the clearing function is an endeavour, approximations are used; see Asmundsson et al. (2006a, 2009).

Asmundsson et al. (2006a, 2009) integrate the estimated clearing function in a mathematical programming model where the framework is based on the production model of Hackman and Leachman (1989) with an objective function that minimizes the overall costs.

It is assumed that backorders do not occur and that all demand must be met on time. One may assume a representation as a graph with N nodes referring to the workstations and a set A of transportation arcs between these nodes. Predecessors of workstations are indicated by A_I and successors by A_O, respectively.

The multi-product multi-stage model of Asmundsson et al. (2002, 2006a) is given differently with

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\sum_{w=1}^{W} (hw_{iwt} W_{iwt} + h_{iwt} I_{iwt} + pc_{iwt} x_{iwt} + rc_{iwt} R_{iwt}) + \sum_{\gamma \in A}^{\Gamma} tc_{i\gamma t} \Gamma_{i\gamma t} \right] \tag{7.56}$$

subject to

$$W_{iwt} = W_{iw,t-1} + \frac{1}{2}(x_{iwt} + x_{iw,t-1} + R_{iwt}) + \sum_{\gamma \in A_I(p,s)}^{\Gamma} \qquad \forall\, i,w,t \tag{7.57}$$

$$I_{iwt} = I_{iw,t-1} + \frac{1}{2}(x_{iwt} + x_{iw,t-1}) - d_{iwt} - \sum_{\gamma \in A_O(p,s)}^{\Gamma} \qquad \forall\, i,w,t \tag{7.58}$$

$$\xi_{iwt} x_{\cdot wt} \leq \alpha_{wt}^{c} \xi_{iwt} W_{iwt} + Z_{iwt} \beta_{wt}^{c} \qquad \forall\, i,w,t,c \tag{7.59}$$

$$\sum_{i=1}^{N} Z_{iwt} = 1 \qquad \forall\, w,t \tag{7.60}$$

$$x_{iwt}, W_{iwt}, I_{iwt}, R_{iwt}, \Gamma_{i\gamma t} \geq 0 \qquad \forall\, i,w,t,\gamma \tag{7.61}$$

where the objective function (7.56) minimizes the sum of overall costs. Equations (7.57) and (7.58) state the workload and inventory balance equation. Inequalities (7.59) give the CF for multiple products and machines, so that a variable Z_{iwt} is introduced defining the partitions of utilization for the products that have to sum up to 1 which is required in Equation (7.60). Inequalities (7.61) denote the usual non-negativity constraints.

Missbauer and Uzsoy (2011) present the model of Asmundsson et al. (2006a, 2009) in a more compact form with

$$\min \sum_{t=1}^{T} \sum_{i=1}^{N} (hw_i W_{it} + h_i I_{it} + pc_i x_{it} + rc_i R_{it}) \tag{7.62}$$

subject to

$$W_{it} = W_{i,t-1} - x_{it} + R_{it} \qquad \forall\, i,t \tag{7.63}$$

$$I_{it} = I_{i,t-1} + x_{it} - d_{it} \qquad \forall\, i,t \tag{7.64}$$

$$\xi_{it} x_{it} \leq Z_{it} f_t \left(\frac{\xi_{it} \hat{W}_{it}}{Z_{it}} \right) \qquad \forall\, i,t \tag{7.65}$$

$$\sum_{i=1}^{N} Z_{it} = 1 \qquad \forall\, t \tag{7.66}$$

$$x_{it}, W_{it}, I_{it}, R_{it}, Z_{it} \geq 0 \qquad \forall\, i,t \tag{7.67}$$

where $\hat{W}_{it}$ is defined as in the model of Karmarkar (1989). The objective function (7.62) minimizes the sum of WIP holding costs, FGI holding costs, processing costs, and release costs over the planning horizon $t = 1, \ldots, T$. Note that Asmundsson et al. (2009), Missbauer and Uzsoy (2011) do not sum over all products N in the objectiv function. Furthermore, no setups are included despite the consideration of multiple products.

7.3 Discussion on Presented Approaches

Previously, various discrete deterministic model approaches are presented that account for LDLT. They have some advantages and disadvantages that are discussed in this section starting from RP models.

LDLT are due to high utilization of production resources that can be also measured by resulting WIP/workload in the system. RP models incorporate the relationship between lead times and workload by the use of CFs using analytically derived functional forms or approximations using empirical data or simulation. Consequently, the CF reflects system behavior, so that the planning model releases orders/jobs in the production system with the aim to reduce overall system costs. As a result, due dates are known for specific demand that are assumed to be fixed in adcance, so that a possibility to decrease the utilization of production resources is to release jobs earlier in the system. LDLT are indirectly incorporated using queuing models to determine input data regarding the nonlinear relationship between system workload and lead times. The queue length in such approaches is fixed and a certain system performance is reached. WIP builds up due to the difference in order releases and production periods. Lot sizing decisions are assumed to be made on a previous planning step, so that setup and batching plans are fixed input data to the RP model. This is different to the PP approach of Lautenschläger (1999) that takes into account lot sizing decisions and their influence on system utilization, inventories, and waiting times.

On the other hand, PP models that assume estimates of lead times, so that lead times become an input parameter to the optimization model (see, e.g., the model of Woodruff and Voß (2004)), approach the problem of modeling the behavior of the production resource by computing lead time distributions. This is integrated into the model by constraint sets accounting for lead time distributions across different time periods; see also Missbauer and Uzsoy (2011). The model of Woodruff and Voß (2004) further assume routings of items as input data to the system. Such data is frequently unavailable for production systems. Furthermore, a large matrix of routing decisions is build up for the solution process that is likely to decrease the solution finding process.

The PP model of Lautenschläger (1999) takes into account capacity decisions due to the production mix thus addressing another planning circularity; see Section 5.2.2. Workload due to WIP in the system and their processing including required setup times is modeled differentiating between physical and available production which is a kind of reservation of pre-produced items. They can be regarded as being an equal idea to distinguish between releases (R_{it}), WIP (W_{it}), and production (x_{it}): Missbauer and Uzsoy (2011) state that in lot sizing, production and WIP have the relation

$$x_{it} = R_{i,t-L_i} \tag{7.68}$$

so that releases take place L_i periods before production. In the model of Lautenschläger (1999), the release time $L_i = 1$ is assumed if we draw the analogy to the regarded production mode $m = 2$, so that we may state the following: $x_{it}^{m=2} = R_{i,t-L_i}$. Nevertheless, Lautenschläger (1999) abstract from setup times in the load-dependent capacity restrictions.

The parameters involved in establishing the FS curve in the model of Lautenschläger (1999) need to be set very carefully, because it is easy to highly restrict production in mode $m = 1$, so that high amounts of production need to be shifted forward causing increased inventory costs. Moreover, production in mode $m = 2$ thus forward shifting is limited by a classical LP upper bound, so that the capacity constraint is not load-dependent and permits high utilization up to 100%.

In the subsequent, capacity management and LDLT are combined further accounting for effects of the production mix with sustainability aspects like wastage avoidance and rework. Therefore, basic ideas from the model of Lautenschläger (1999) are regarded and

extensions integrated until reaching an overall model including LDLT in form of production smoothing and sustainability options such as limited lifetimes of items, rework of passed items, and remanufacturing of returned items. This is due to the objective to minimize the number of fixed input parameters to the model and to apply data from classical lot sizing models.

Part II

Combining LDLT and Lot Sizing with Sustainability Aspects

Chapter 8

Production Performance Models with Perishability

This analysis concentrates on discrete parts production and dynamic planning where a finite time horizon is modeled according to the concept of its division into time buckets. It is assumed that overtime takes place, but capacity expansion and, consequently, investment in new production entities are not regarded as such decisions are attributed to the strategic planning level; see Section 3.1.2.

PP formulations are developed that integrate various features regarding LDLT and aspects of sustainability as presented in Sections 4 and 6. Proposed models are analyzed regarding their behaviour and resulting plans, their complexity and thus their applicability in practice.

8.1 CLSP with Lifetimes and Load-Dependent Capacity Constraint

We adopt basic ideas of Lautenschläger (1999) to model load-dependent capacity constraints using different production modes and the concept of available/disponible inventory containing WIP/releases in form of pre-ordered items and physical produced inventory. Therefore, the basic model assumptions of Lautenschläger (1999) are still valid despite the assumption of items being connected via BOMs, so that the single stage case is regarded adding the following assumptions:

13. Items have a limited lifetime Θ_i^L. If items pass their lifetimes, they are discarded at a disposal cost ϕ_i. This assumption is relaxed in Section 9.2 that extends models to account for rework options;

14. The storage space of the production system is not restricted;

15. The order in which products are processed follows the *first-in-first-out (FIFO)* rule.

The basic model is based on the classical *capacitated lot sizing problem with perishability constraints (CLSP-P)* (see Pahl and Voß (2010)) and the ideas of Lautenschläger (1999) regarding production modes named CLSP-P-LDCap-1 and can be stated as follows:

CLSP-P-LDCap-1

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left(sc_i y_{it} + h_i P_{it} + \phi_i \cdot I_{it}^D \right) \tag{8.1}$$

subject to

$$x_{it}^{\text{dispo}} = x_{it}^{m=1} - I_{it}^D \qquad \forall\, i,t = 1 \tag{8.2}$$

$$x_{it}^{\text{dispo}} = x_{it}^{m=1} + x_{i,t-1}^{m=2} \qquad \forall\, i,t = 2,\ldots,T \tag{8.3}$$

$$P_{it} = P_{i,t-1} + \sum_{m=1}^{M} x_{it}^m - d_{it} - I_{it}^D \qquad \forall\, i,t = 1,\ldots,T \tag{8.4}$$

$$I_{it} = I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} \qquad \forall\, i,t = 1,\ldots,T \tag{8.5}$$

$$I_{it}^D \geq \sum_{m=1}^{M} \sum_{u=1}^{t-\Theta_i^L} x_{iu}^m - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D \qquad \forall\, i,t \geq \Theta_i^L \tag{8.6}$$

$$\sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) \leq \text{Cap}_t \qquad \forall\, t = 1,\ldots,T \tag{8.7}$$

$$\sum_{i=1}^{N} \left(\xi_i x_{it}^{m=1} + st_i y_{it} \right) \leq \alpha \cdot \text{Cap}_t - \beta \sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) \qquad \forall\, t = 1,\ldots,T \tag{8.8}$$

$$\sum_{m=1}^{M} x_{it}^m \leq V \cdot y_{it} \qquad \forall\, i,t = 1,\ldots,T \tag{8.9}$$

$$I_{j0} = I_{jT} = 0 \qquad \forall\, j \tag{8.10}$$

$$y_{it} \in \{0,1\} \qquad \forall\, i,t \tag{8.11}$$

$$I_{it}, P_{it}, x_{it}^m \geq 0 \qquad \forall\, m,j,t \tag{8.12}$$

The objective function given in (8.1) minimizes overall costs for setups, (physical) inventory holding, and disposals. Equations (8.2) define available/disponible products (x_{it}^{dispo}) for the first period which is limited to production mode $m = 1$, because forward shifting (mode $m = 2$) to previous periods cannot take place due to the start of the planning horizon at $t = 1$. If this is not ascertained for the first period, the model would be allowed to take products "out of nowhere" to fill inventory at the expense of inventory holding costs that are lower than other regarded cost factors. Moreover, disposals might be necessary if there is no demand for products in $t = 1$, so that I_{it}^D is added. Equation (8.3) determines available production in period t: this can be achieved by production in mode $m = 1$, so that items are produced and available for demand satisfaction in t, or by mode $m = 2$ that produce items in $t-1$ for use in t. Equations (8.4) and (8.5) give the inventory balance equations for available inventory and physical inventory, respectively. These two equations are necessary for inventory to restrict and calculate the amounts of production and inventory at the correct time period given by the production mode definitions. Available inventory and physical inventory differ exactly by those products that are shifted forward in time to $t-1$ (production mode $m = 2$). Inequalities (8.6) determine disposals due to passed lifetimes Θ_i^L; see also Pahl and Voß (2010) or Section 4.2. Restrictions (8.7) limit the sum of production and setups in both modes in period t to the maximal nominal capacity. This is necessary, because the load-dependent capacity restrictions presented in Inequalities (8.8) restrict only production in mode $m = 1$. Applying an LP upper bound is somewhat inconsequent as the stated requirements for models accounting for LDLT are to limit production in accordance to WIP inventories to minimize lead times. Therefore, LDLT are only regarded indirectly via production mode $m = 1$. This is modified in Section 8.4. Inequalities (8.8) restrict production in production mode $m = 1$ by the resting operational capacity available for production, i.e., Cap_t multiplied with α less the already used capacity in period t regarding processing

and setup times, i.e., $\beta \cdot \sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it})$. Inequalities (8.9) set the setup variables according to production of items in all modes M. Consequently, products that are shifted forward are allowed to be batched in order to produce them together and the machine is set up, accordingly. The classical assumptions of the *capacitated lot sizing problem (CLSP)* include assumptions on inital and ending inventory required to be zero given in Equalities (8.10). This can be changed assuming initial positive inventories also with different initial ages; see Pahl and Seipl (2011) for I&L as well as *network flow problem* (NFP) formulations. Constraints (8.11) define setup variables as binary and Inequalities (8.12) state the usual non-negativity requirements on decision variables.

The load-dependent capacity constraint restricting production in mode $m = 1$ stated in (8.8) calculates how much nominal capacity is available for production in mode $m = 1$ in period t which is limited by planned operational capacity for overall production in period t. Production of items triggered by demand requirements that are higher than allowed by load-dependent capacity restrictions (8.8) are shifted forward which is likewise restricted by an LP upper bound given in (8.7), so that high utilization in a period t up to 100% is not avoided. It could be required that the forward shifting of a maximal amount e before 100% utilization is reached, so that the determination of α and β are modified as follows:

$$\alpha = \frac{v \cdot e \cdot g}{v - g} \tag{8.13}$$

$$\beta = \frac{v \cdot (e + g - 1)}{v - g} \tag{8.14}$$

where v denotes the parameter at which utilization level the maximal amount e is shifted forward. The resulting production smoothing is studied employing the following small test example: a planning horizon is assumed that has $T = 10$ time periods. Nominal capacity is 100 TU. A single product A is taken into account that has the characteristics regarding processing and costs given in Table 8.1.

Table 8.1: List of time-infinite parameters

Product	sc_i	st_i	h_i	ϕ_i	ξ_i	Θ_i^L
A	1	0	1	10	1	10

Forward shifting is assumed to begin at an utilization of 70%, i.e., $g = 0.7$ and that, at 100% utilization ($v = 1.0$), 80% of jobs is shifted forward, i.e., $e = 0.8$. Varying this parameter gives adjustment flexibility regarding production mode $m = 1$. For instance, assuming $e = 0.0$ eliminates the load-dependent capacity limit of production mode $m = 1$, so that the LP upper bound on all production modes is the next bound that applies. Besides, Varying v-values are applied to demonstrate their effects on capacity loading and forward shifting. Demand is assumed to be zero despite the last period $t = 10$ that has a high value, so that shifting forward of production is necessary. Lower values for demand that exceed period capacity show the same forward shifting behavior. Higher values of demand, e.g., $d_{A,10} = 1200$ require overtime that is discussed in Section 8.2.

Table 8.2 shows results of planning in the two production modes for a single product A. The values for $X1_{At}$ and $X2_{At}$ define the "position" on the FS curve in relation to the planned workload where $X1_{At} = \text{Cap}_t \cdot (\alpha - \beta \cdot \rho_t)$ and $X2_{At} = \sum_{m=1}^{M} \sum_{i=1}^{N} x_{it}^m - \text{Cap}_t \cdot (\alpha - \beta \cdot \rho_t)$. In other words, $X1_{At}$ that are used as statistical values present planned production in mode $m = 1$ that are restricted by the load-dependent capacity constraints given in (8.8) and $X2_{At}$ denote resting capacity of items that can be employed to produce items in mode $m = 2$.

Table 8.2: Numerical example to show load effects for varying v-values

v-values	Data and PI†	Time periods 1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	300
$v=1$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	0	0	0	0	0	0	**30**	100	100	**0**
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.3	1.0	1.0	0.7
	$X1_{At}$	187	187	187	187	187	187	137	20	20	70
	$X2_{At}$	−187	−187	−187	−187	−187	−187	−107	80	80	0
	d_{At}	0	0	0	0	0	0	0	0	0	900
$v=1$	$x_{At}^{m=1}$	**0**	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	**30**	100	100	100	100	100	100	100	100	**0**
	ρ_t	0.3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.7
	$X1_{At}$	137	20	20	20	20	20	20	20	20	70
	$X2_{At}$	−107	80	80	80	80	80	80	80	80	0
	d_{At}	0	0	0	0	0	0	0	0	0	900
$v=0.8$	$x_{At}^{m=1}$	**0**	0	0	0	0	0	0	0	0	**90**
	$x_{At}^{m=2}$	**10**	100	100	100	100	100	100	100	100	**0**
	ρ_t	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
	$X1_{At}$	406	48	48	48	48	48	48	48	48	90
	$X2_{At}$	−396	52	52	52	52	52	52	52	52	0
	d_{At}	0	0	0	0	0	0	0	0	0	900
$v=1.4$	$x_{At}^{m=1}$	**0**	0	0	0	0	0	0	0	0	**56**
	$x_{At}^{m=2}$	**44**	100	100	100	100	100	100	100	100	**0**
	ρ_t	0.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.6
	$X1_{At}$	68	12	12	12	12	12	12	12	12	56
	$X2_{At}$	−24	88	88	88	88	88	88	88	88	0

† PI: performance indicators

These values become negative if planned utilization is low. In case of demand peaks implying high or full utilization, the position of $X2_{At}$ gives the capacity limit for production in mode $m = 2$ as shown, e.g., in Table 8.2, row for $v = 1$ and demand $d_{A,10} = 900$, position values $X1_{A,1} = 20$ and $X2_{A,1} = 80$. Note that the higher the value of v, the higher the amount of items shifted forward from the "beginning" period to previous periods. Capacity utilization in periods where forward shifting takes place is at 100% which is due to the LP upper bound limit an all production modes; see Table 8.2. Production in mode $m = 1$ is given only in the *parent* period $t = 10$ and forward shifting to *child* periods $t - 1$ fully occupies nominal capacity. This is not satisfying regarding the goal to minimize LDLT. Moreover, problem instances that assume that products cannot be stored which implies a lifetime of $\Theta_i^L = 0$ and high workload due to demand data are frequently infeasible, because forward shifting, i.e., processing in mode $m = 2$ is not allowed. Moreover, demand peaks require short term capacity extensions, e.g., overtime or subcontracting that can be interpreted as some kind of production outsourcing. This is discussed in more details in Section 8.2. Therefore, we integrate overtime into the model to diminish these drawbacks. Moreover, different formulations are investigated that may help to limit high utilizations dependent on the (planned) workload of the system which is some kind of production smoothing; see Section 8.4.

8.2 Including Overtime and Minimal Lot Sizes

Overtime options can be interpreted as additional working hours in a period t needed to extend the nominal capacity of that period. Subcontracting is another option to deal with high demand in peak situations that imply production outsourcing of processing that pertains to procurement actions. Job outsourcing is only feasible if a supply chain partner exists that can provide required quality and amounts of parts or services. This frequently is not true for specific parts and products where a company and its customers have special quality requirements. Therefore, job outsourcing might only be possible for standardized products. Teo et al. (2011) discuss the options of overtime and subcontracting in case of peak demand and variabilities in order to keep promised delivery dates which is similar to the work of Rao et al. (2005) and an extension to the work of Graves (1986). They combine two time scales which might be similar to hybrid lot sizing models, e.g., *general lot sizing and scheduling problem (GLSP)* where micro periods allow for multiple item movements within that micro period. The proposed stochastic optimization model includes an objective function that concentrates on minimizing expected total subcontracting costs.

In order to integrate overtime, it would be straightforward to extend the load-dependent capacity constraints (8.7) as follows:

$$\sum_{i=1}^{N}\left(\xi_i x_{it}^{m=1} + st_i y_{it}\right) \leq \alpha \cdot (\mathrm{Cap}_t + O_t) - \beta \sum_{m=1}^{M}\sum_{i=1}^{N}(\xi_i x_{it}^{m} + st_i y_{it}) \qquad \forall\, t = 1,\ldots,T \quad (8.15)$$

where O_t is the decision variable regarding overtime, so that the overall capacity K_t is extended by overtime. However, in case that the lifetime of items is Θ_i^L is zero, these constraints lead to infeasibility, because the use of production in mode $m = 2$ is not possible and $m = 1$ is restricted. Furthermore, due to the α and β-values, the use of overtime is over-proportional, e.g., in order to produce 1.000 entities of product A in period $t = 10$, an overtime of 1.329 is employed. To remedy these effects, a third production mode is introduced that denotes overtime. Therefore, Equations (8.2) and (8.3) that determine disponible inventory have to be extended with the third production mode as shown in (8.16) and (8.17), accordingly. The LP upper bound on capacity adds overtime given in (8.18).

$$x_{it}^{\mathrm{dispo}} = x_{it}^{m=1} + x_{it}^{m=3} - I_{it}^{D} \qquad \forall\, i,t = 1 \quad (8.16)$$

$$x_{it}^{\mathrm{dispo}} = x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} - I_{it}^{D} \qquad \forall\, i,t = 2,\ldots,T \quad (8.17)$$

$$\sum_{m=1}^{M}\sum_{i=1}^{N}(\xi_i x_{it}^{m} + st_i y_{it}) \leq \mathrm{Cap}_t + O_t \qquad \forall\, t = 1,\ldots,T \quad (8.18)$$

Moreover, production mode $m = 3$ is defined as overtime, accordingly, in (8.19).

$$\sum_{i=t}^{N}\left(\xi_i \cdot x_{it}^{m=3} + st_i \cdot y_{it}\right) = O_t \qquad \forall\, 1 = 1,\ldots,T \quad (8.19)$$

Furthermore, the constraints on minimal lot sizes are added givne in (8.20).

$$\sum_{m=1}^{M} x_{it}^{m} \geq x_i^{\min} \cdot y_{it} \qquad \forall\, i,t \quad (8.20)$$

This equals the requirements for setups presented in Inequalities (8.9) where products processed in the same production mode are regarded together in order to build lot sizes for equal product types. In other words, if items are shifted forward and these items are also produced in mode $m = 1$, the machine should be set up only once for that product even if they are produced in two different modes. Besides, the setup constraints presented in (8.9)

need some modification, because an upper bound given by a big number V which is also frequently substituted by $(\text{Cap}_t - st_j)/\xi_i$ can be limiting especially in case of integrated overtimes, so that the solution process becomes awkward or the constraints cause infeasibility. An upper bound on production as proposed by Wagner and Whitin (1958), but slightly modified is given with

$$\sum_{m=1}^{M} x_{it}^m \leq \left(d^k + \text{Cap}_t\right) \cdot y_{it} \qquad \forall\, i,t = 1,\ldots,T \tag{8.21}$$

where d^k presents the cumulative demand over all products i and periods t. Constraints (8.21) require that production is always less than or equal the sum of cumulative demand and nominal period capacity. Finally, the objective function is extended regarding overtime, so that the complete model is stated as follows:

CLSP-P-LDCap-2

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T} \left(sc_i y_{it} + h_i P_{it} + \phi_i \cdot I_{it}^D + oc_i \cdot O_t\right) \tag{8.22}$$

subject to

$$\begin{aligned}
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 1\\
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 2,\ldots,T\\
P_{it} &= P_{i,t-1} + \sum_{m=1}^{M} x_{it}^m - d_{it} - I_{it}^D && \forall\, i,t = 1,\ldots,T\\
I_{it} &= I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} && \forall\, i,t = 1,\ldots,T\\
I_{it}^D &\geq \sum_{m=1}^{M}\sum_{u=1}^{t-\Theta_i^L} x_{iu}^m - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D && \forall\, i,t \geq \Theta_i^L\\
\sum_{m=1}^{M}\sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) &\leq \text{Cap}_t + O_t && \forall\, t = 1,\ldots,T\\
\sum_{i=1}^{N} \left(\xi_i x_{it}^{m=1} + st_i y_{it}\right) &\leq \alpha \cdot \text{Cap}_t - \beta \sum_{m=1}^{2}\sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) && \forall\, t = 1,\ldots,T\\
\sum_{i=t}^{N} \left(\xi_i x_{it}^{m=3} + st_i y_{it}\right) &= O_t && \forall\, 1 = 1,\ldots,T\\
\sum_{m=1}^{M} x_{it}^m &\leq \left(d^k + \text{Cap}_t\right) \cdot y_{it} && \forall\, i,t = 1,\ldots,T\\
\sum_{m=1}^{M} x_{it}^m &\geq x_i^{\min} \cdot y_{it} && \forall\, i,t = 1,\ldots,T\\
I_{j0} = P_{i0} = I_{jT} = P_{iT} &= 0 && \forall\, j\\
y_{it} &\in \{0,1\} && \forall\, i,t = 1,\ldots,T\\
I_{it}, P_{it}, x_{it}^m &\geq 0 && \forall\, m,j,t = 1,\ldots,T
\end{aligned}$$

Here, overtime is interpreted as an additional shift, so that utilization ρ_t can get values above $\rho_t \geq 1.0$ that is shown, e.g., in Figure 8.3. In doing so and to be consistent with the interpretation, overtime should be limited to two or three times of the regular capacity. This constraint can be easily added to allow for, e.g., at most three shifts:

$$O_t \leq 3 \cdot \text{Cap}_t \qquad \forall\, t \tag{8.23}$$

For the sake of always ensuring feasibility, overtime is not restricted. Subsequently, these extensions are tested using the small test example introduced in the previous Section 8.1 to the overtime cost parameter. Because overtime should be the last option in planning, overtime is assumed to be highly expensive in comparison to the other cost factors, thus $oc_t = 100$.

Table 8.3: Numerical example to show load effects including overtime

	Data and	Time periods									
v-values	PI†	1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	1200
$v = 1$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	100	100	100	100	100	100	100	100	100	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**230**
	ρ_t	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	3.0
	$X1_{At}$	20	20	20	20	20	20	20	20	20	−313
	$X2_{At}$	80	80	80	80	80	80	80	80	80	613

† PI: performance indicators

Table 8.3 shows the effects of demand peaks on load-dependent capacity restrictions and overtime that is used as late as possible, i.e., in period $t = 9$, 230 units of product A are produced using overtime for demand satisfaction in period $t = 10$ and 100 items are shifted forward from $t = 10$. Positions $X1_{At}$ and $X2_{At}$ are moved, accordingly. The difference between these positions denote the amount of items that is shifted forward. For instance, $X1_{At} = 187$ and $X2_{At} = -187$ with $\Delta = 0$ shows that no forward shifting takes place whereas $X1_{At} = 137$ and $X2_{At} = -107$ with $\Delta = 30$ presents a forward shifting of 30 items. The model behaves well when varying the lifetime factor $\Theta_i^L = 10$ until $\Theta_i^L = 0$ limiting more and more the possibility to produce in mode $m = 2$ that implies inventory holding until lifetime of items is zero. In this case, 70 units are still produced in production mode $m = 1$ and the resting 1.130 units are produced in overtime in the same period. Table 8.4 presents the effects of varying lifetimes of items on the production modes and, therefore, overall production. A v-value of $v = 1.0$ and a single product A are assumed for this example which is restricted to product A for ease of presentation. It is easy to see that in case of overall capacity exceeding demand ($d_{At} = 1.200$ over $\sum_{t=1}^{T} \text{Cap}_t = 1.000$), production mode $m = 1$ produces maximal amounts allowed by the load-dependent capacity constraint ($x_{A,10}^{m=1} = 70$) and the exceeding amounts are shifted forward viz. produced in production mode $m = 2$. Production mode $m = 2$ is restricted by the LP upper bound, so that amounts up to the nominal capacity are produced. This is possible as setup times are assumed to be zero for this small testing example also for ease of presentation.

8.3 Liberate Resource Capacity: Inclusion of Setup Carry Overs

Setting up the machine anew every period even though the same product is produced subsequently may have its reasons if, e.g., maintenance actions such as cleaning need to be performed after every period. In case that such maintenance actions are rather short and the machine can be assumed to remain set up for a specific product, we account for setup carry overs to the next period in order to avoid unnecessary costs and resource utilization. Setup carry overs are also named linked lot sizes. These are subject to dynamic research, see, e.g.,

Table 8.4: Numerical example to show load effects including overtime and variations of product lifetimes employing CLSP-P-LDCap-2

Θ_A-values	Data and PI	Time periods 1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	1200
$\Theta_A = 10$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	100	100	100	100	100	100	100	100	100	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**230**
	ρ_t	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	3.0
	$X1_{At}$	20	20	20	20	20	20	20	20	20	−313
	$X2_{At}$	80	80	80	80	80	80	80	80	80	613
$\Theta_A = 5$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	0	0	0	0	100	100	100	100	100	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**630**
	ρ_t	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	7.0
	$X1_{At}$	187	187	187	187	20	20	20	20	20	−980
	$X2_{At}$	−187	−187	−187	−187	80	80	80	80	80	1680
$\Theta_A = 1$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	100	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**1030**
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	11.0
	$X1_{At}$	187	187	187	187	187	187	187	187	20	−1647
	$X2_{At}$	−187	−187	−187	−187	−187	−187	−187	−187	80	2747
$\Theta_A = 0$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**70**
	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	0	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**1130**
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	11.0
	$X1_{At}$	187	187	187	187	187	187	187	187	187	−1813
	$X2_{At}$	−187	−187	−187	−187	−187	−187	−187	−187	−187	3013

Gopalakrishnan et al. (1995), Sox and Gao (1999), Gopalakrishnan et al. (2001), Porkka et al. (2003), Suerie and Stadtler (2003), Briskorn (2006), Karimi et al. (2006), Quadt and Kuhn (2008, 2009).

For the integration of setup carry overs, we orientate the mathematical formulation on the *continuous setup and lot sizing problem (CSLP)* and *proportional lot sizing and scheduling problem (PLSP)* including perishability/lifetime constraints while regarding macro periods instead of micro periods although we are aware that the CSLP and the PLSP are micro period (small bucket) models that are intended for use in short term planning; see also Drexl and Haase (1995), Kimms (1996), Marinelli et al. (2007) for the mathematical formulation or Pahl and Voß (2010) for a CSLP and PLSP including lifetime constraints.

Setup carry overs are modeled by the introduction of binary setup state variables $z_{it} \in \{0,1\}$ linking the setup variables as follows (see Buschkühl et al. (2010)):

$$\sum_{m=1}^{M} x_{it}^{m} \cdot \xi_i \leq \left(d^k + \text{Cap}_t\right) \cdot (y_{it} + z_{it}) \qquad \forall\, i,t = 1,\ldots,T \tag{8.24}$$

$$\sum_{i=1}^{N} z_{it} \leq 1 \qquad \forall\, t = 1,\ldots,T \tag{8.25}$$

$$z_{it} \leq y_{i,t-1} \qquad \forall\, i,t = 1,\ldots,T \tag{8.26}$$

Inequalities (8.24) ensure that production of an item i takes place in period t only if the production resource is set up for item i in this period, i.e., $y_{it} = 1$, or if the production is already in the right setup state for the item, i.e., $z_{it} = 1$. The resource can only be in one setup state of a specific item i in a period t which is enforced by Inequalities (8.25). Inequalities (8.26) denotes that setup states for an item i can only be carried over in t if the corresponding setup has been performed in $t-1$; see Tempelmeier and Buschkühl (2009). This implies that the setup state is "lost" or "forgotten" in the subsequent period $t+1$ and an additional setup needs to be performed even if the machine has been set up to the related product. We may interprete that also as some cleaning, maintenance, or re-powering up if the machine has not been used for some time, i.e., one period. If the retention of a setup state for more than one period is desirable, Inequalities (8.25) are modified as follows (see also Porkka et al. (2003) to

$$z_{it} \leq y_{i,t-1} + z_{i,t-1} \qquad \forall\, i,t = 1,\ldots,T \tag{8.27}$$

If a setup has been performed in $t-1$ ($y_{i,t-1} = 0$), then the setup state has been set, accordingly, in $t-1$ ($z_{i,t-1} = 1$) or is carried over to t ($z_{it} = 1$). Besides, the constraints (8.25) are required to be $\sum_{i=1}^{N} z_{it} = 1, \forall\, t = 1,\ldots,T$ in the literature; see, e.g., Drexl and Haase (1995), Kimms (1996) which enforces a setup state. We prefer not to enforce a setup state and assume that the production resource in the beginning of the planning horizon is not set up to a specific product, but has to be prepared which is decided by a plan, so that $z_{i0} = 0, \forall\, i,t = 0$ is assumed that may be interpreted as some general warming up of the machine or a setup after maintenance actions.

The complete model including load-dependent capacity constraints is stated as follows:

CLSP-P-LDCap-3

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left(sc_i y_{it} + h_i P_{it} + \phi_i \cdot I_{it}^D + oc_i \cdot O_t \right)$$

subject to

$$\begin{aligned}
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 1 \\
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 2,\ldots,T \\
P_{it} &= P_{i,t-1} + \sum_{m=1}^{M} x_{it}^m - d_{it} - I_{it}^D && \forall\, i,t = 1,\ldots,T \\
I_{it} &= I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} && \forall\, i,t = 1,\ldots,T \\
I_{it}^D &\geq \sum_{m=1}^{M} \sum_{u=1}^{t-\Theta_i^L} x_{iu}^m - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D && \forall\, i,t \geq \Theta_i^L \\
\sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) &\leq \text{Cap}_t + O_t && \forall\, t = 1,\ldots,T \\
\sum_{i=1}^{N} \left(\xi_i x_{it}^{m=1} + st_i y_{it} \right) &\leq \alpha \cdot \text{Cap}_t - \beta \sum_{m=1}^{2} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) && \forall\, t = 1,\ldots,T \\
\sum_{i=t}^{N} \left(\xi_i \cdot x_{it}^{m=3} + st_i \cdot y_{it} \right) &= O_t && \forall\, t = 1,\ldots,T \\
\sum_{m=1}^{M} x_{it}^m &\leq \left(d^k + \text{Cap}_t \right) \cdot (y_{it} + z_{it}) && \forall\, i,t = 1,\ldots,T \\
\sum_{i=1}^{N} z_{it} &\leq 1 && \forall\, t = 1,\ldots,T \\
z_{it} &\leq y_{i,t-1} + z_{i,t-1} && \forall\, i,t = 1,\ldots,T \\
\sum_{m=1}^{M} x_{it}^m &\geq x_i^{\min} \cdot (y_{it} + z_{it}) && \forall\, i,t = 1,\ldots,T \\
I_{i0} = P_{i0} = I_{iT} = P_{iT} &= 0 && \forall\, j \\
y_{it} &\in \{0,1\} && \forall\, i,t = 1,\ldots,T \\
I_{it}, P_{it}, x_{it}^m &\geq 0 && \forall\, m,i,t = 1,\ldots,T
\end{aligned}$$

8.4 Load-Dependent Capacity Restrictions and Production Smoothing

Inventory holding is one of the classical tools to smooth production or, in other words, to buffer manufacturing against demand peaks; see, e.g., Allen (1999). If products have limited lifetimes, inventory holding is limited and buffer options to smooth production non-optimal if accounting for quality issues. Smoothing resource utilization of bottleneck machines imply inventory holding, but the focus is on workload that may be penalized by the objective function using, e.g., a piecewise linear function on workload costs. This is a frequently described method in the production smoothing literature; see Yavuz and Akçali (2007), Kubiak and Yavuz (2008).

The lead time syndrome states that releasing jobs earlier increases in the system WIP and thus lead times. Consequently, production smoothing should have a negative effect on lead times. On the other hand, lead times rise exponentially with resource utilization. As shown by queuing models and also confirmed in practice (see Schulte (2012)), there is a desirable *operation sector* of utilization regarding the load-dependent capacity constraint (*operation curve* or "Kennlinie" in German; see Nyhuis (1991), Breithaupt et al. (2002)) of production resources where slight reductions of utilization levels can lead to great reductions in lead times. This does not say much on WIP, but if workload measured in WIP is another performance indicator of utilization, then lead times can be reduced by holding the same amount/level of WIP. Therefore, these levels or this operation sector should be targeted by production smoothing actions.

8.4.1 Literature Review on Production Smoothing

Classical lot sizing models are triggered by demand and thus pertain to pull systems as production, lot sizing, and setups are invoked due to demand. Such pull systems might apply MtO and JIT principles in order to minimize WIP/FGI inventories and to maintain quality standards by reducing overall sojourn or lead times and thus maximizing freshness (minimizing ages) of items. These ideas and principles are subsumed to *lean management*. The application of JIT principles requires a smooth production plan that implies no significant variations in demands in order to maintain a certain performance level of the production system where many methods or tools exist; see, e.g., Hopp and Spearman (2001), Özkavukcu and Durmuşoğlu (2010), Monden (2011). For instance, the Toyota production system is the famous example of the application of the JIT philosophy together with lean manufacturing principles; see, e.g., Yavuz and Akçali (2007), Kubiak and Yavuz (2008), Monden (2011). These concepts advice employ pull-oriented production control systems that are triggered by demand (see, e.g., Yavuz and Akçali (2007)) which is the same for classical lot sizing approaches. In this context, production smoothing at the tactical planning level may aim at reducing the production's (consumption, i.e. demand) rate variability at the final manufacturing stage in order to determine a robust/stable demand flow from previous manufacturing stages that is necessary for a JIT system; see Yavuz and Akçali (2007). This becomes more complicating with the increase of production mix and related product varieties required by customers.

Production smoothing has mainly been taken into account in synchronized assembly line systems; see, e.g., Yavuz and Akçali (2007), Kubiak and Yavuz (2008) and the references therein. In this context, analysis concentrates, on the one hand, on the output rate of FGI at the final assembly stage workstations also named *output rate variation (OVR)* and, on the other hand, on the *product rate variation (PRV)* which refers to the pull rate of WIP or sub-assembly parts. Yavuz and Akçali (2007) state that the major part of research is on the PRV, possibly due to raised complexity regarding the settings and planning problem where OVR is relevant. Setup times are frequently assumed as negligible in these production environments and processing times allow to produce exactly one unit in a time period; see also Yavuz and Akçali (2007), Kubiak and Yavuz (2008). Furthermore, multi-product scheduling with the aim of smoothing workload in job shops has not been studies so far although there is a great scheduling literature examining various manufacturing environments also including identical parallel machines.

Yavuz and Tufekçi (2004) regard the PRV approach in a single (final stage) workstation system. Setup and processing time requirements are assumed to be arbitrary and time buckets to be of equal length. Time buckets can be used for setups or processing of multiple units of final products, or left idle. A two phase solution method is applied that determines, in the first phase, setups, batch sizes, and the lengths of time buckets and, in the second

phase, sequences of end products. The authors report that the first phase is NP-hard and the second phase relatively easy to solve due to its structure of discrete time periods that allows for solving it as an assignemt problem; see Yavuz and Akçali (2007). They further report that the solution process of large test instances applying exact optimization is very time consuming, so that heuristics are proposed by, e.g., Yavuz et al. (2006). Kubiak and Yavuz (2008) adopt the two phase methodology of Yavuz and Tufekçi (2004) and apply it to JIT manufacturing for multiple products with a single bottleneck machine.

Production smoothing is frequently integrated in mathematical models by modifying the objective function according to some criterion. Monden (2011) emphasize on two issues regarding production smoothing: the first is named *usage goal* and take into account the deviation of the *actual* amount of cumulative production and consumption of production parts that are assembled to the final product and the pre-specified or *ideal* production/consumption amounts; see also Yavuz and Akçali (2007). The second issue regards the *loading goal* of a production system that emphasize on the deviation between actual workload of a production resource and its ideal or predefined workload. These deviations are integrated into the objective function of related optimization models, e.g., using *absolute* of *squared* values for the deviation that can take positive or negative values; see, e.g., Figure 8.1 for a presentation of ideal and cumulative production quantities in a discrete time horizon. As shown in Figure

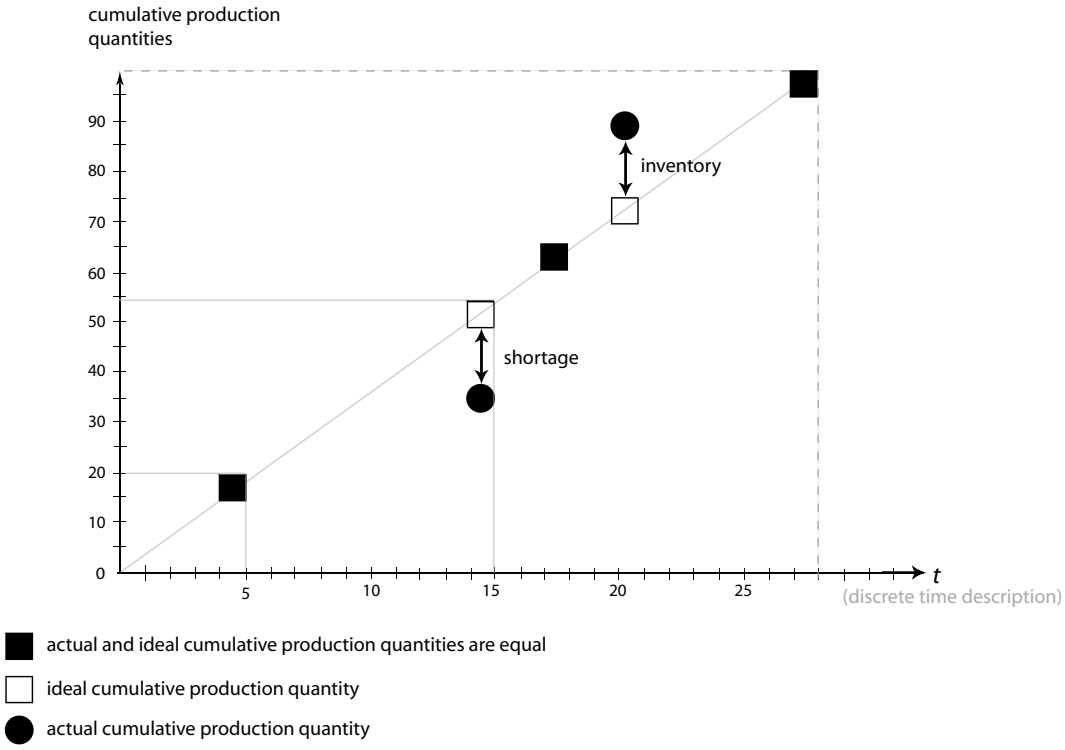

Figure 8.1: Ideal and actual cumulative production quantities; see Yavuz and Akçali (2007).

8.1, demand is assumed to be uniformly distributed over the regarded planning horizon, so that production smoothing implies the evenly partition of production in order to minimize FGI; see Kubiak and Yavuz (2008). Deviations from the ideal cumulative production quantities and demand such as inventory holding and shortages are included into the objective function and thus subject to minimization.

Decision regarding the integration of production smoothing criteria into the objective function remains arbitrary according to Kubiak and Yavuz (2008) as there does not seem to be any criterion superior over another. In fact, they include four decision criteria into the

objective function in their study, i.e., the minimization of total production rate variations, WIP, the maximization of system utilization, and responsiveness. According to our analysis so far, the two goals "minimization of WIP" and the "maximization of system utilization" are conflicting. The goal "maximization of responsiveness" entails the minimization of average times of idle time periods, so that these might serve as buffer time that can be used for unexpected events, maintenance, and the like. The multi-objective problem is solved applying Pareto optimization; see Kubiak and Yavuz (2008). The authors define inequalities regarding the time buckets that must be sufficiently long to allow for setting up the workstation for the respective product and processing of the items, so that the available capacity is not exceeded. Setup and processing times are assumed to be sequence-independent. The inequalities that must be satisfied by the time bucket lengths are nonlinear and solved via enumeration by Kubiak and Yavuz (2008). The goal of this model is to determine the optimal length of time buckets regarding multiple criteria integrated in the objective function.

In classical lot sizing, the lengths of time buckets are assumed to be fixed and not subject to decisions. The integration of a piecewise linear term in the objective function that penalize high workloads or, in other words, the deviation from an ideal workload that allows for significant lead time reductions while slightly reducing utilization, seems a promising direction that we further proceed.

8.4.2 Integration of Production Smoothing via PSCU

The literature on production smoothing emphasizes production smoothing issues in the objective function. We follow this path and define a term that determines planned occupied capacity named PSCU which is used in the objective function together with its cost factor and assumed to be quadratic. We use the CLSP-P-LDCap-3 for implementation that uses setup carry overs:

$$\mathrm{PSCU}_t = \sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i \cdot x_{it} + st_i \cdot y_{it})^2 \tag{8.28}$$

In case that overtime is interpreted as an extra shift thus an extension to regular production, it should be included into the PSCU as follows:

$$\mathrm{PSCU}_t = \left(\sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i \cdot x_{it} + st_i \cdot y_{it}) + O_t \right)^2 \qquad \forall\, t \tag{8.29}$$

If overtime is interpreted rather as subcontracting where orders/jobs and related production are outsourced, the component should not be included in the PSCU. The objective function given in (8.22) is modified as follows:

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left(h_i P_{it} + \phi_i \cdot I_{it}^D + psc_t \cdot \mathrm{PSCU}_t^2 \right) \tag{8.30}$$

where the parameter psc_t denotes the cost factor of planned smoothed capacity utilization named PSCU, so that the costs attributed to processing, setups, and overtime are equal. In practice, overtime is more costly than regular production due to night shifts. Furthermore, overtime should be the last option for production, so that we integrate additional overtime costs in the objective function expressing the overtime (tarif) rate:

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left(h_i P_{it} + \phi_i \cdot I_{it}^D + psc_t \cdot \mathrm{PSCU}_t^2 + oc_i \cdot O_t \right) \tag{8.31}$$

The complete CLSP-P-LDCap-4 is given as follows:

CLSP-P-LDCap-4

$$\min \sum_{i=1}^{N} \sum_{t=1}^{T} \left(h_i P_{it} + \phi_i \cdot I_{it}^D + psc_t \cdot \mathrm{PSCU}_t^2 + oc_i \cdot O_t \right)$$

subject to

$$\begin{aligned}
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 1 \\
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 2,\ldots,T \\
P_{it} &= P_{i,t-1} + \sum_{m=1}^{M} x_{it}^m - d_{it} - I_{it}^D && \forall\, i,t = 1,\ldots,T \\
I_{it} &= I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} && \forall\, i,t = 1,\ldots,T \\
I_{it}^D &\geq \sum_{m=1}^{M} \sum_{u=1}^{t-\Theta_i^L} x_{iu}^m - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D && \forall\, i,t \geq \Theta_i^L \\
\sum_{m=1}^{M} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) &\leq \mathrm{Cap}_t + O_t && \forall\, t = 1,\ldots,T \\
\sum_{i=1}^{N} \left(\xi_i x_{it}^{m=1} + st_i y_{it} \right) &\leq \alpha \cdot \mathrm{Cap}_t - \beta \sum_{m=1}^{2} \sum_{i=1}^{N} (\xi_i x_{it}^m + st_i y_{it}) && \forall\, t = 1,\ldots,T \\
\sum_{i=t}^{N} \left(\xi_i \cdot x_{it}^{m=3} + st_i \cdot y_{it} \right) &= O_t && \forall\, t = 1,\ldots,T \\
\sum_{m=1}^{M} x_{it}^m &\leq \left(d^k + \mathrm{Cap}_t \right) \cdot (y_{it} + z_{i,t}) && \forall\, i,t = 1,\ldots,T \\
\sum_{i=1}^{N} z_{it} &\leq 1 && \forall\, t = 1,\ldots,T \\
z_{it} &\leq y_{i,t-1} + z_{i,t-1} && \forall\, i,t = 1,\ldots,T \\
\sum_{m=1}^{M} x_{it}^m &\geq x_i^{\min} \cdot (y_{it} + z_{it}) && \forall\, i,t = 1,\ldots,T \\
I_{i0} = P_{i0} = I_{iT} = P_{iT} &= 0 && \forall\, i \\
y_{it} &\in \{0,1\} && \forall\, i,t = 1,\ldots,T \\
I_{it}, P_{it}, x_{it}^m &\geq 0 && \forall\, m,i,t = 1,\ldots,T
\end{aligned}$$

Table 8.5 presents the effect of the PSCU on the utilization. We use the model CLSP-P-LDCap-4 and the example stated in Section 8.2 using cost parameters stated in Table 8.1 on Page 63 assuming a time horizon of $T = 8$.

As Table 8.5 shows, utilization ρ_t is relatively evenly distributed and so is production over both production modes. Setups and related costs are not relevant in this example as the machine is set up once for product A, thus relevant costs are processing costs and inventory holding costs where processing costs have a high relevance due to the quadratic term in the objective function.

Production smoothing integrated via the quadratic term in the objective function always takes place, so that even if utilization is not significantly high, production is distributed/disseminated evenly on other time periods. High utilizations should be penalized differently, e.g., beginning at 75% where lead times begin to increase significantly, in a

Table 8.5: Numerical example to show load effects employing PSCU in Model CLSP-P-LDCap-4

Θ_A-values	Data and PI	Time periods 1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	300
$\Theta_A = 10$	$x_{At}^{m=1}$	13	8	12	14	14	14	15	15	16	32
	$x_{At}^{m=2}$	14	20	16	16	16	16	16	16	16	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**	**0.3**
	$X1_{At}$	140	140	139	138	137	136	135	135	134	133
	$X2_{At}$	−113	−111	−110	−109	−107	−106	−105	−103	−102	−101
	d_{At}	0	0	0	0	0	0	0	0	0	900
$\Theta_A = 10$	$x_{At}^{m=1}$	19	16	10	9	10	10	12	13	13	70
	$x_{At}^{m=2}$	71	75	81	83	82	83	81	81	81	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.9**	**0.7**
	$X1_{At}$	36	35	35	34	33	32	31	30	30	70
	$X2_{At}$	54	55	57	58	59	61	62	63	65	0
	d_{At}	0	0	0	0	0	0	0	0	0	1200
$\Theta_A = 10$	$x_{At}^{m=1}$	9	10	6	7	7	8	9	9	10	70
	$x_{At}^{m=2}$	91	90	94	93	93	92	91	91	90	0
	$x_{At}^{m=3}$	21	21	21	21	21	22	22	22	22	37
	ρ_t	**1.2**	**1.2**	**1.2**	**1.2**	**1.2**	**1.2**	**1.2**	**1.2**	**1.2**	**1.1**
	$X1_{At}$	−15	−15	−15	−16	−16	−16	−16	−16	−17	8
	$X2_{At}$	136	136	137	137	137	138	138	138	139	99
$\Theta_A = 5$	$x_{At}^{m=1}$	0	0	0	0	12	11	12	13	8	70
	$x_{At}^{m=2}$	0	0	0	0	88	89	88	87	92	0
	$x_{At}^{m=3}$	0	0	0	0	102	102	102	103	103	118
	ρ_t	0.0	0.0	0.0	0.0	**2.0**	**2.0**	**2.0**	**2.0**	**2.0**	**1.9**
	$X1_{At}$	187	187	187	187	−150	−151	−151	−151	−151	−126
	$X2_{At}$	−187	−187	−187	−187	353	353	353	354	354	314
$\Theta_A = 1$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	12	70
	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	88	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	507	523
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**2.0**	**6.1**	**5.9**
	$X1_{At}$	187	187	187	187	187	187	187	187	−826	−801
	$X2_{At}$	−187	−187	−187	−187	−187	−187	−187	−187	1433	1394
$\Theta_A = 0$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	70
	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	0	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	1130
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0
	$X1_{At}$	187	187	187	187	187	187	187	187	187	−1813
	$X2_{At}$	−187	−187	−187	−187	−187	−187	−187	−187	−187	3013

Table 8.6: Numerical example to show load effects employing PSCU in Model CLSP-P-LDCap-4 and varying v-values

Θ_A-values v-values	Data and PI	Time periods 1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	900
$\Theta_A = 10$	$x_{At}^{m=1}$	**19**	**16**	**10**	**9**	**10**	**10**	**12**	**13**	**13**	**70**
$v = 1.0$	$x_{At}^{m=2}$	71	75	81	83	82	83	81	81	81	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	**0.7**
	$X1_{At}$	36	35	35	34	33	32	31	30	30	70
	$X2_{At}$	54	55	57	58	59	61	62	63	65	0
$\Theta_A = 10$	$x_{At}^{m=1}$	**23**	**19**	**24**	**28**	**34**	**32**	**29**	**28**	**28**	**90**
$v = 0.8$	$x_{At}^{m=2}$	65	69	65	62	56	59	62	63	64	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	**0.9**
	$X1_{At}$	96	94	92	90	88	86	84	82	80	90
	$X2_{At}$	−8	−5	−3	0	2	5	7	10	12	0
$\Theta_A = 10$	$x_{At}^{m=1}$	**12**	**7**	**9**	**8**	**7**	**5**	**8**	**8**	**8**	**56**
$v = 1.4$	$x_{At}^{m=2}$	80	85	84	85	87	89	87	87	88	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0	1.0	**0.6**
	$X1_{At}$	20	20	19	19	18	18	17	17	16	56
	$X2_{At}$	72	73	74	75	76	77	78	79	80	0
$\Theta_A = 5$	$x_{At}^{m=1}$	0	0	0	0	6	6	4	4	6	**56**
$v = 1.4$	$x_{At}^{m=2}$	0	0	0	0	94	94	96	96	94	0
	$x_{At}^{m=3}$	0	0	0	0	53	53	54	54	54	76
	ρ_t	0.0	0.0	0.0	0.0	1.5	1.5	1.5	1.5	1.5	1.3
	$X1_{At}$	112	112	112	112	−41	−41	−42	−42	−42	−20
	$X2_{At}$	−112	−112	−112	−112	195	195	195	195	196	152
$\Theta_A = 1$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	**6**	**56**
$v = 1.4$	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	**94**	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	**361**	**383**
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.6	4.4
	$X1_{At}$	112	112	112	112	112	112	112	112	−349	−327
	$X2_{At}$	−112	−112	−112	−112	−112	−112	−112	−112	810	766
$\Theta_A = 0$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	**56**
$v = 1.4$	$x_{At}^{m=2}$	0	0	0	0	0	0	0	0	0	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	**844**
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0
	$X1_{At}$	112	112	112	112	112	112	112	112	112	−788
	$X2_{At}$	−112	−112	−112	−112	−112	−112	−112	−112	−112	1688

harder way than distribute low capacity utilization. Therefore, we change the quadratic PSCU to a piecewise linear functional term in the objective function.

8.4.3 Production Smoothing via Targeting Utilization Sectors

The PSCU smoothed production evenly on periods disregarding its utilization levels. There is an operative utilization sector where utilization is relatively high, but lead times relatively short that is reported to be around 85% of utilization; see also Section 5.4 and Figure 5.4 on Page 26 or Figure 8.2. Therefore, it is desirable to smooth production or capacity utilization around this value avoiding higher values as well as lower values.

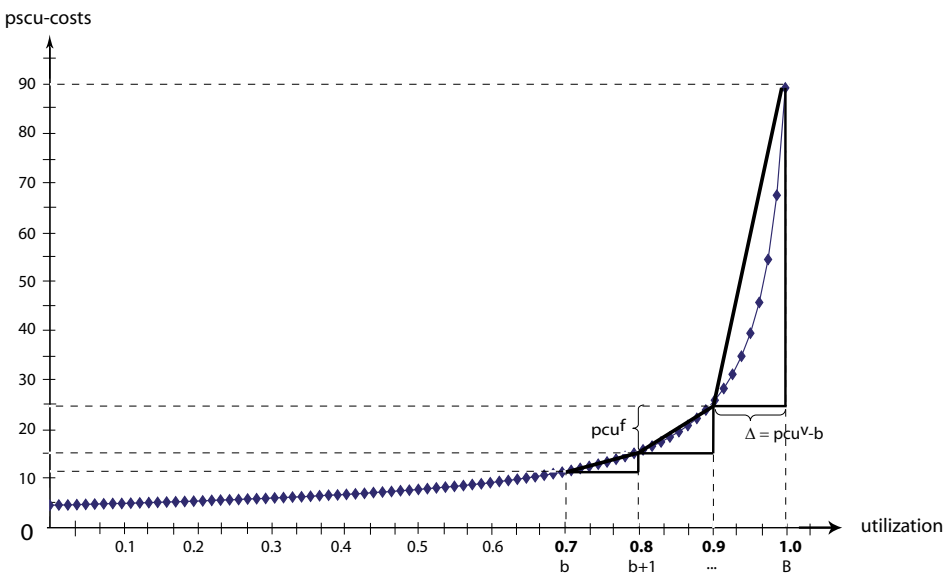

Figure 8.2: Approximation of load-dependent resource utilization costs for use in piecewise linear cost function term in objective function.

The integration of a piecewise linear cost term that is used to motivate planned capacity utilization to be in the described sector, i.e. between $70-85\%$ might be done via the insertion of SOS1, but we model the interval selection according to planned capacity utilization employing binary variables, denoted by $y_{bt} \in \{0,1\}$. The breakpoints give the interval borders regarding the utilization sectors where, depending on the planned processing and setup times, cost terms are selected, accordingly; see also Figure 8.2. In order to correctly model the cost terms, we differentiate between fixed utilization costs when a subsequent interval (or sector) is reached and variable costs according to the planned utilized capacity units. In each time period t, an interval/sector according to the planned capacity utilization needs to be selected:

$$\sum_{b=1}^{B} y_{bt} = 1 \qquad \forall\, t = 1,\ldots,T \tag{8.32}$$

The planned capacity utilizataion in a time period t is attributed to the load-dependent cost

term as follows:

$$\sum_{i=1}^{N}\sum_{m=1}^{3}(\xi_i \cdot x_{it}^m + st_i \cdot y_{it}) = \sum_{b=1}^{B} \text{pscu}_{bt} \qquad \forall\, t \tag{8.33}$$

where the parameter pscu_{bt} denotes the planned smoothed capacity utilization attributed to breakpoint b in time period t. Moreover, we define the intervals as follows:

$$\text{pscu}_{bt} \leq y_{bt} \cdot b_B \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T \tag{8.34}$$

$$\text{pscu}_{bt} \geq b_b \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T \tag{8.35}$$

$$\text{pscu}_{bt} \leq b_{b+1} \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B-1, t = 1,\ldots,T \tag{8.36}$$

Inequalitites (8.34) gives the border of the last interval denoted by breakpoint B. Inequalities (8.35) denotes the lower limit of pscu_{bt} and (8.36) gives the upper limit. The borders of the intervals/sectors b_b are parameters and thus input data. In order to integrate the load-dependent processing and setup costs into the objective function, we define a fixed cost term pcu_b^f and a variable cost term pcu_b^v, so that the objective function is modified as follows:

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_i P_{it} + \phi_i \cdot I_{it}^D + \left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \text{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t\right) \tag{8.37}$$

The load-dependent capacity utilization costs are chosen according to the binary variable y_{bt} that becomes 1 if the utilization is in a certain interval. We attribute fixed costs for an utilization interval pcu_b^f and variable costs psc_b^v that give additional costs according to the amount of items that utilize capacity.

The complete CLSP-P-LDCap-5 is stated as follows:

CLSP-P-LDCap-5

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_i P_{it} + \phi_i \cdot I_{it}^D + \left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \text{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t\right)$$

subject to

$$\begin{aligned}
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 1\\
x_{it}^{\text{dispo}} &= x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} - I_{it}^D && \forall\, i,t = 2,\dots,T\\
P_{it} &= P_{i,t-1} + \sum_{m=1}^{M} x_{it}^m - d_{it} - I_{it}^D && \forall\, i,t = 1,\dots,T\\
I_{it} &= I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} && \forall\, i,t = 1,\dots,T\\
I_{it}^D &\geq \sum_{m=1}^{M}\sum_{u=1}^{t-\Theta_i^L} x_{iu}^m - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D && \forall\, i,t \geq \Theta_i^L\\
\sum_{m=1}^{M}\sum_{i=1}^{N}(\xi_i x_{it}^m + st_i y_{it}) &\leq \text{Cap}_t + O_t && \forall\, t = 1,\dots,T\\
\sum_{i=1}^{N}\left(\xi_i x_{it}^{m=1} + st_i y_{it}\right) &\leq \alpha \cdot \text{Cap}_t - \beta \sum_{m=1}^{2}\sum_{i=1}^{N}(\xi_i x_{it}^m + st_i y_{it}) && \forall\, t = 1,\dots,T\\
\sum_{i=t}^{N}\left(\xi_i \cdot x_{it}^{m=3} + st_i \cdot y_{it}\right) &= O_t && \forall\, t = 1,\dots,T\\
\sum_{m=1}^{M} x_{it}^m &\leq \left(d^k + \text{Cap}_t\right)\cdot(y_{it} + z_{i,t}) && \forall\, i,t = 1,\dots,T\\
\sum_{i=1}^{N} z_{it} &\leq 1 && \forall\, t = 1,\dots,T\\
z_{it} &\leq y_{i,t-1} + z_{i,t-1} && \forall\, i,t = 1,\dots,T\\
\sum_{m=1}^{M} x_{it}^m &\geq x_i^{\min}\cdot(y_{it} + z_{it}) && \forall\, i,t = 1,\dots,T
\end{aligned}$$

$$\begin{aligned}
\sum_{b=1}^{B} y_{bt} &= 1 && \forall\, t = 1,\dots,T\\
\text{pscu}_{bt} &\leq y_{bt}\cdot b_B && \forall\, b = 1,\dots,B, t = 1,\dots,T\\
\text{pscu}_{bt} &\geq b_b \cdot y_{bt} && \forall\, b = 1,\dots,B, t = 1,\dots,T\\
\text{pscu}_{bt} &\leq b_{b+1}\cdot y_{bt} && \forall\, b = 1,\dots,B-1, t = 1,\dots,T\\
I_{i0} = P_{i0} = I_{iT} = P_{iT} &= 0 && \forall\, i\\
y_{it} &\in \{0,1\} && \forall\, i,t = 1,\dots,T\\
I_{it}, P_{it}, x_{it}^m &\geq 0 && \forall\, m,i,t = 1,\dots,T
\end{aligned}$$

with $\sum_{i=1}^{N}\sum_{m=1}^{3}(\xi_i \cdot x_{it}^m + st_i \cdot y_{it}) = \sum_{b=1}^{B}\text{pscu}_{bt}\ \forall\, t = 1,\dots,T$. In order to show the resulting production smoothing, we use the example given in Table 8.1. We further assume a capacity of $\text{Cap}_t = 100$ TU and that at 70%, forward shifting beginns ($g = 0,7$). Moreover, at 100%, 80% are shifted forward ($e = 0,8$) and $v = 1.0$.

Table 8.7: Numerical example to show load effects including overtime and variations of product lifetimes employing CLSP-P-LDCap-5

	Data and	Time periods									
Θ_A-values	PI	1	2	3	4	5	6	7	8	9	10
	d_{At}	0	0	0	0	0	0	0	0	0	500
$\Theta_A = 10$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	70
	$x_{At}^{m=2}$	0	0	0	35	70	70	85	85	85	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	0.0	0.0	0.0	0.4	0.7	0.7	0.8	0.8	0.8	0.7
	$X1_{At}$	187	187	187	128	70	70	45	45	45	70
	$X2_{At}$	−187	−187	−187	−93	0	0	40	40	40	0
$\Theta_A = 5$	$x_{At}^{m=1}$	0	0	0	0	0	0	45	0	0	70
	$x_{At}^{m=2}$	0	0	0	0	85	85	40	85	90	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	0
	ρ_t	0.0	0.0	0.0	0.0	0.8	0.8	0.8	0.8	0.9	0.7
	$X1_{At}$	187	187	187	187	45	45	45	45	37	70
	$X2_{At}$	−187	−187	−187	−187	40	40	40	40	53	0
$\Theta_A = 3$	$x_{At}^{m=1}$	0	0	0	0	0	0	0	0	0	70
	$x_{At}^{m=2}$	0	0	0	0	0	0	0	100	100	0
	$x_{At}^{m=3}$	0	0	0	0	0	0	0	0	0	130
	ρ_t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	2.0
	$X1_{At}$	187	187	187	187	187	187	20	20	20	−147
	$X2_{At}$	−187	−187	−187	−187	−187	−187	80	80	80	347

As shown in Table 8.7, the capacity utilization is allocated, so that the target sectors for production are met whose costs are lower than those of the higher utilizations. The lifetime of item A forces production and utilization to take place in the useful lifetime periods.

The forward shifting and thus the restriction on the first production mode including production in period t is limiting. Similarities to the CLSP-P regarding production in the previous period if capacity utilization in the current period t is higher than nominal capacity are clear, so that we compare the behavior and results of the CLSP-P-LDCap-5 to a CLSP-P with the same extensions, i.e., overtime, minimal lot size requirements, setup carry overs, and production smoothing given by the piecewise linear objective function term on capacity utilization.

8.5 Some Excursion: Forward Shifting or Extended CLSP-P

The CLSP-P-LDCap-5 includes a hard constrain on production mode $m = 1$ that restricts production in this mode according to the load in the system. The forward shifting parameters can be adjusted regarding a specific production system, so that the load-dependent capacity restriction becomes less restrictive as in the previous examples assuming a high forward shifting with parameter $e = 0.8$, so that 80% of items are shifted at 100% and only 20% can be produced in the regarded period which is very restrictive. The CLSP shifts production forward in case that the LP upper bound on capacity is reached, so that LDLT are not taken into account. This is changed using the target utilization sector approach including a piecewise linear cost term on utizilization. As a consequence, restrictions on production in period t need not be restrictive as in the CLSP-P-LDCap-5. In the following, we state the CLSP-P-5 and compare behavior and results of some numerical example with those of the CLSP-P-LDCap-5 given on Page 77. The extended CLSP-P-5 is stated as follows:

CLSP-P-5

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_i P_{it} + \phi_i \cdot I_{it}^D + \left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \text{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t\right) \quad (8.38)$$

subject to

$$I_{it} = I_{i,t-1} + x_{it} - d_{it} - I_{it}^D \qquad \forall\, i,t = 1,\ldots,T \quad (8.39)$$

$$I_{it}^D \geq \sum_{u=1}^{t-\Theta_i^L} x_{iu} - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D \qquad \forall\, i,t = 1,\ldots,T \quad (8.40)$$

$$\sum_{i=1}^{N}(\xi_i \cdot x_{it} + st_i \cdot y_{it}) \leq \text{Cap}_i + O_t \qquad \forall\, i,t = 1,\ldots,T \quad (8.41)$$

$$z_{it} \leq y_{i,t-1} + z_{i,t-1} \qquad \forall\, i,t = 1,\ldots,T \quad (8.42)$$

$$\sum_{i=1}^{N} z_{it} \leq 1 \qquad \forall\, i,t = 1,\ldots,T \quad (8.43)$$

$$x_{it} \leq (d^k + \text{Cap}_t) \cdot (y_{it} + z_{it}) \qquad \forall\, i,t = 1,\ldots,T \quad (8.44)$$

$$x_{it} \geq x_i^{\min} \geq (y_{it} + z_{it}) \qquad \forall\, i,t = 1,\ldots,T \quad (8.45)$$

$$\text{pscu}_{bt} \leq y_{bt} \cdot b_B \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T \quad (8.46)$$

$$\text{pscu}_{bt} \geq b_b \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T \quad (8.47)$$

$$\text{pscu}_{bt} \leq b_{b+1} \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B-1, t = 1,\ldots,T \quad (8.48)$$

with $\text{pscu}_{bt} = \sum_{i=1}^{N}(\xi_i \cdot x_{it} + st_i \cdot y_{it})\ \forall t = 1,\ldots,T$.

In order to analyze/study the formulations, we take an example including $i = 5$ products, $T = 10$ macro periods, minimal lot sizes $x_i^{\min} = 20$, and varying lifetimes 1) $\Theta_i = 10$, 2) $\Theta_i = 10$, 3) $\Theta_i = 3$, 4) $\Theta_i = 2$, and 5) $\Theta_i = 1$. The demand data is given in Table 12.1 in the Appendice, Page VII where we take demand data of the first five products. Nominal capacity is assumed to be $\text{Cap}_t = 250$, $\phi_i = 12$, and $oc_i = 10$. Moreover, cost parameters as given in Table 8.8 are used.

Table 8.8: List of time-infinite parameters for the comparison of CLSP-P-LDCap-5 and CLSP-P-5

Product	ξ_i	sc_i	h_i	st_i
A	0.282	4.201	1.887	0.944
B	0.942	7.340	5.169	0.241
C	0.491	8.267	0.553	0.471
D	0.213	5.189	0.250	0.241
E	0.232	7.809	2.872	0.701
F	0.981	4.201	6.798	0.344
G	0.951	7.340	2.661	0.196
H	0.741	8.267	4.740	0.767
I	0.324	5.189	5.628	0.342
J	0.543	7.809	4.669	0.931

Furthermore, the costs of utilization for the utilization intervals are stated in Table 8.9. For the load-dependent capacicty constraint, we use parameters $e = 0.4, g = 0.7, v = 1.0$.

Table 8.11 presents total costs of the objective for all six cases regarding lifetimes of items. The relatively high overall costs for the CLSP-P-LDCap-5 are mainly caused by the

Table 8.9: Cost parameters of utilization intervals

Cost factor	interval borders b				
	0	175	213	225	2000
pcu_b^f	1	42	83	238	523
pcu_b^v	1	8	17	48	65

use of overtime due to the restrictions in production modes $m = 1$ and $m = 2$. Moreover, inventory costs are higher for the CLSP-P-LDCap-5 in relation to those of the CLSP-P-5. The same is valid for setups and their costs. The utilization costs of both models is the same due to the same amount of items produced or disposed.

Table 8.10: Differences of overall costs comparing the CLSP-P-LDCap-5 and the CLSP-P-5

	Lifetimes				
Differences	$\Theta_i = 10$	$\Theta_i = 5$	$\Theta_i = 3$	$\Theta_i = 2$	$\Theta_i = 1$
Objective	3.419	3.419	3.421	3.431	3.460
Inventory costs	506	506	406	502	364
Setup costs	−36	−36	−20	−33	−8
Disposal costs	0	0	0	0	−34
Overtime costs	3.156	3.156	3.240	3.160	3.289
Utilization costs	−208	−208	−205	−197	−149

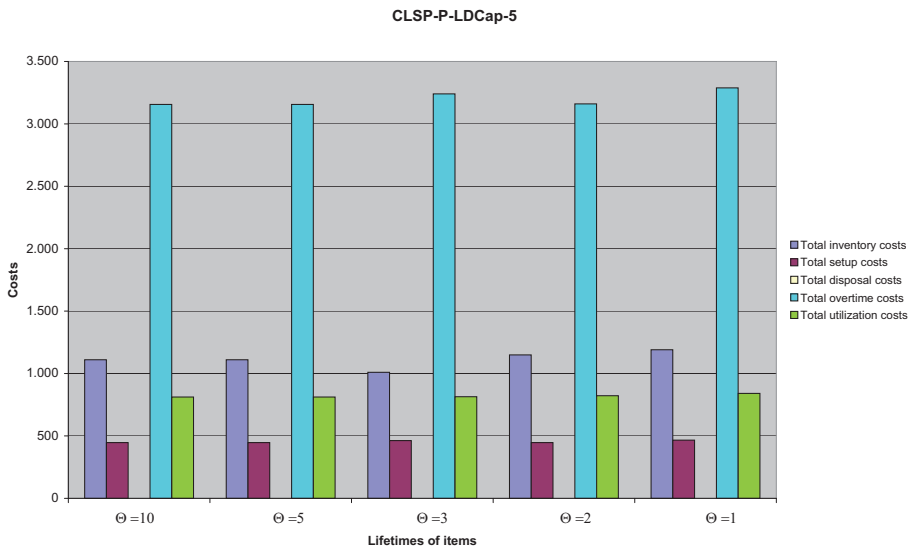

Figure 8.3: Total costs of the CLSP-P-LDCap-5.

Figure 8.5 highlights the cost differences of the results of both models. Thus, positive values show higher costs of the CLSP-P-LDCap-5 and negative values imply less costs. It clearly shows that the CLSP-P-LDCap-5 results in significantly higher overtime costs due to restrictions on production mode $m = 1$. Moreover, the CLSP-P-LDCap-5 has slightly lower setup costs and less disposal costs, but higher inventory holding costs.

The CLSP-P-LDCap-5 has three dimensions for production where the classical CLSP needs two. This rises complexity and slows down the solution process.

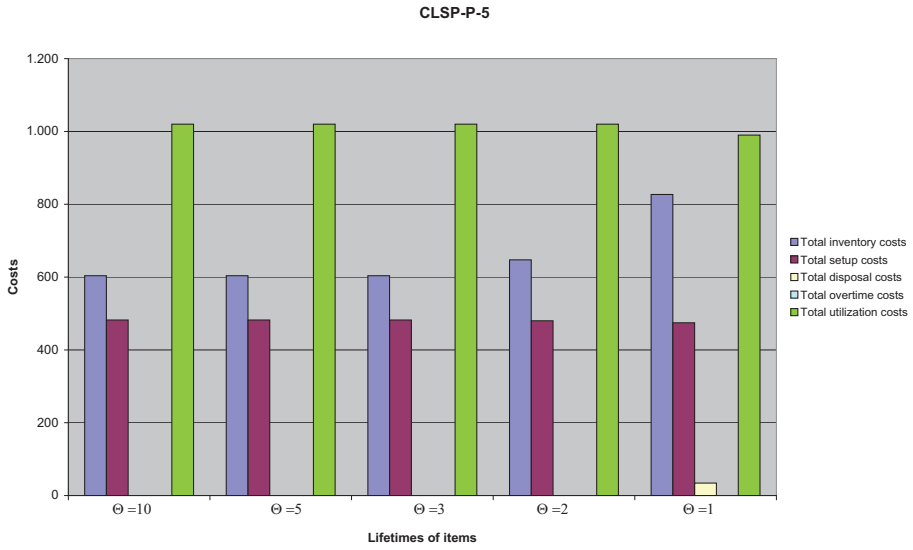

Figure 8.4: Total costs of the CLSP-P-5.

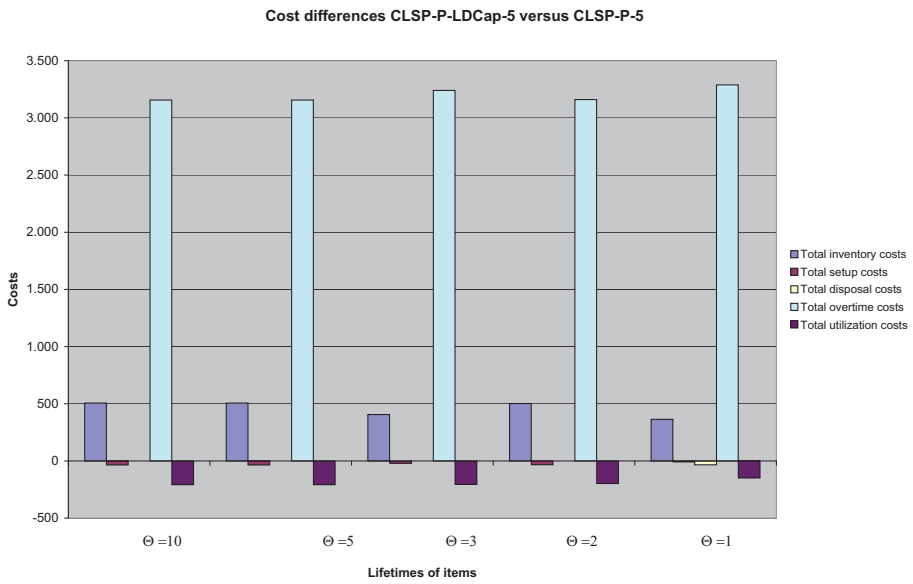

Figure 8.5: Differences in total costs of the CLSP-P-5 versus the CLSP-P-5 for the 5-product case.

Table 8.11: Comparison of results of the CLSP-P-LDCap-5 and the CLSP-P-5 regarding their overall costs with $st_i = 0$

	Lifetimes					Lifetimes				
Total costs	$\Theta_i = 10$	$\Theta_i = 5$	$\Theta_i = 3$	$\Theta_i = 2$	$\Theta_i = 1$	$\Theta_i = 10$	$\Theta_i = 5$	$\Theta_i = 3$	$\Theta_i = 2$	$\Theta_i = 1$
Objective	5.524	5.524	5.527	5.578	5.786	2.106	2.106	2.106	2.147	2.326
Inventory costs	1.110	1.110	1.010	1.149	1.191	604	604	604	647	827
Setup costs	447	446.53	462.60	446.53	466.10	482	482	482	480	475
Disposal costs	0	0	0	0	0	0	0	0	0	34
Overtime costs	3.156	3.156	3.240	3.160	3.289	0	0	0	0	0
Utilization costs	812	812	815	822	841	1.020	1.020	1.020	1.020	990

Chapter 9

Remanufacturing and Rework

In this section, we introduce rework and remanufacturing of externally returned items and internally passed items. Therefore, in Section 9.1, we develop models that take into account remanufacturing and modify them to include rework of internally passed items in Section 9.2.

9.1 Remanufacturing of Returns

Remanufacturing is the disassembly, restoring/refurbishing of worn out products that are returned from the customer. Therefore, returned items G_{it} from customers are regarded as the source of items that need to be reworked. Remanufacturing is integrated into the planning and scheduling of regular products. A fourth production mode $m = 4$ is defined that represents remanufacturing, i.e., the rework of externally returned items. The following assumptions are added to those stated previously:

16. Returns presented by parameter G_{it} are externally given and push returned items in the system requiring production mode $m = 4$ to process or discard them, accordingly;

17. remanufactured items reacquire an "as-good-as-new" quality, so that their lifetime fully restored or, in other words, reset to zero;

18. The remanufacturing rate of items is given in relation to regular processing, i.e., ξ_i/r. We assume that the remanufacturing is smaller or equal regular production, so that $r \leq \xi_i$ with $r > 0$. This is done in order to favor in-line rework. Certainly, it gives flexibility also to define rework more time-consuming than regular processing. Values higher than $r = 1$ express that rework employs more time than regular production. This assumption should be proofed or discarded by empirical data.

19. Neither regular storage space of FGI or returned items are restricted;

20. Returned items that may not serve for demand satisfaction are discarded immediately, so that the pull character of the system is still prevalent. A setup is enforced in case of positive returns even if disposal for all returned items in that period takes place. This might be interpreted as some fixed cost encountered due to handling of returns.

21. Returned items can be processed together with other production modes regarding the same product i, so that no additional setup for remanufacturing is necessary.

We integrate remanufacturing in the CLSP-P-LDCap-5 that name CLSP-P-LDCap-R1. The CLSP-P-LDCap-R1 is stated as follows:

CLSP-P-LDCap-R1

$$\min \sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_i P_{it} + \phi_i \cdot I_{it}^{D} + \left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \text{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t\right) \tag{9.1}$$

subject to

$$x_{it}^{\text{dispo}} = x_{it}^{m=1} + x^{m=3} + x^{m=4} - I_{it}^{D} \qquad \forall\, i,t=1 \tag{9.2}$$

$$x_{it}^{\text{dispo}} = x_{it}^{m=1} + x_{i,t-1}^{m=2} + x_{it}^{m=3} + x_{it}^{m=4} - I_{it}^{D} \qquad \forall\, i,t=2,\ldots,T \tag{9.3}$$

$$P_{it} = P_{i,t-1} + \sum_{m=1}^{M} x_{it}^{m} - d_{it} - I_{it}^{D} \qquad \forall\, i,t=1,\ldots,T \tag{9.4}$$

$$I_{it} = I_{i,t-1} + x_{it}^{\text{dispo}} - d_{it} \qquad \forall\, i,t=1,\ldots,T \tag{9.5}$$

$$I_{it}^{D} \geq \sum_{m=1}^{M}\sum_{u=1}^{t-\Theta_i^L} x_{iu}^{m} - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^{D} \qquad \forall\, i,t \geq \Theta_i^L \tag{9.6}$$

$$\sum_{m=1}^{M}\sum_{i=1}^{N}\left(\xi_i x_{it}^{m} + st_i y_{it}\right) \leq \text{Cap}_t + O_t \qquad \forall\, t=1,\ldots,T \tag{9.7}$$

$$\sum_{i=1}^{N}\left(\xi_i x_{it}^{m=1} + st_i y_{it}\right) \leq \alpha \cdot \text{Cap}_t - \beta \sum_{m=1}^{3}\sum_{i=1}^{N}\left(\xi_i x_{it}^{m} + st_i y_{it}\right) \qquad \forall\, t=1,\ldots,T \tag{9.8}$$

$$G_{it} = \xi_i \cdot x_{it}^{m=4} \qquad \forall\, i,t=1,\ldots,T \tag{9.9}$$

$$\sum_{i=t}^{N} \xi_i \cdot x_{it}^{m=3} = O_t \qquad \forall\, t=1,\ldots,T \tag{9.10}$$

$$\sum_{m=1}^{M} x_{it}^{m} \leq \left(d^k + \text{Cap}_t\right) \cdot \left(y_{it} + z_{i,t}\right) \qquad \forall\, i,t=1,\ldots,T \tag{9.11}$$

$$\sum_{i=1}^{N} z_{it} \leq 1 \qquad \forall\, t=1,\ldots,T \tag{9.12}$$

$$z_{it} \leq y_{i,t-1} \qquad \forall\, i,t=1,\ldots,T \tag{9.13}$$

$$\sum_{m=1}^{M} x_{it}^{m} \geq x_i^{\min} \cdot \left(y_{it} + z_{it}\right) \qquad \forall\, i,t=1,\ldots,T \tag{9.14}$$

$$I_{i0} = I_{iT} = 0 \qquad \forall\, i \tag{9.15}$$

$$y_{it} \in \{0,1\} \qquad \forall\, i,t=1,\ldots,T \tag{9.16}$$

$$I_{it}, P_{it}, x_{it}^{m} \geq 0 \qquad \forall\, m,i,t=1,\ldots,T \tag{9.17}$$

Inequalities (9.9) and (9.10) do not contain setup times, because these are accounted for in (9.7) and (9.8). Modifications are performed in Equalities (9.2) and (9.3) that adds production mode $m = 4$. This formulation has several drawbacks additionally to the complexity of the model that is further increased.

- Returns of items are not allowed to be higher than nominal capacity in a time period Cap_t. We might argue that allowing for returns higher than available capacity is not reasonable if no additional capacity for remanufacturing is available,

- No other production modes may be used for remanufacturing, because production mode $m = 4$ for remanufacturing is an input parameter, so that returns are forced to be reworked in the period they arrive or discarded.

Due to these additional drawbacks, we try a more flexible formulation regarding the processing of returned items. Furthermore, we adopt the CLSP-P for extension regarding remanufacturing and rework. Therefore, we subsequently distinguish between setup and remanufacturing of returned items and regular/overtime production and integrate remanufactured items in the inventory balance equations while adding an inventory balance equation for remanufactured/returned items according to van den Heuvel (2004).

We add the following assumptions for the remanufacturing approach that we name CLSP-P-R-6 which is based on the CLSP-P-5 given in Section 8.5 on Page 79.

22. remanufacturing/rework of items is faster than regular processing assuming that passed products do not have to return to every step of the production process;
23. Reworking items and setting up for them is less costly and time consuming than for regular production;
24. The machine is either set up for regular production or for reworking. These setup states are carried over or changed, accordingly;
25. The machine can only be in one setup state, thus either producing regular products or reworking passed/returned items. In other words, parallel processing is not allowed;
26. Setting up for rework and reworking increases machine utilization, so that these costs are added in the utilization cost term in the objective function;
27. Minimal lot sizes do not apply for rework batches because they should be reworked immediately when returned, so that they do not need to wait before enough items are returned and their quality decreases further while waiting;
28. Returned or passed items may wait Θ_i^r TU in rework inventory. If this time elapses or returned/passed items are not in time to serve for demand satisfaction, they are discarded;
29. Returns are assumed to be a fraction of demanded items.

The relation of returns to demand is given using the following form to calculate them:

$$G_{it} = (1 - \alpha_i^R) \cdot d_{it} \tag{9.18}$$

In practice, forecasts of returns apply that are rather difficult to be established; see Toktay et al. (2003).

The underlying decision process of the model is depicted in Figure 9.1: returned items that are a percentage of items previously demanded and stated in (9.18) are collected in the rework inventory. Dependening on the deterministic demand that triggers the system and, consequently, returned items, it is decided whether to rework or dispose returned items in a first step.

The constraint regarding disposed returned items is given as follows:

$$I_{it}^{Dr} \geq \sum_{u=1}^{t-\Theta_i^r} x_{iu}^r - \sum_{u=1}^{t} G_{iu} - \sum_{u=1}^{t-1} I_{iu}^{Dr} \qquad \forall\, i,t = 1,\ldots,T \tag{9.19}$$

The possibility to rework items is restricted by the returned item lifetime Θ_i^r as shown in the first term of (9.19). We further integrate a rework/remanufacturing inventory balance equation to characterize the flow of remanufacutring stated as follows:

$$I_{it}^r = I_{i,t-1}^r + G_t - x_{it}^r - I_{it}^{Dr} \qquad \forall\, i,t = 1,\ldots,T \tag{9.20}$$

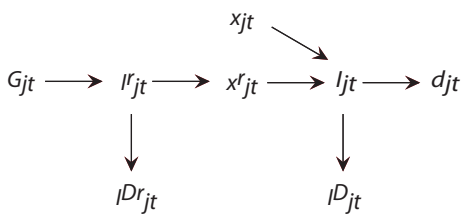

Figure 9.1: Depiction of the remanufacturing/rework/production process applied on the CLSP-P-R-6.

The returned items/rework inventory balance is given by returned items that were already in inventory since previous periods $t-1$, by new external returns G_{it} less those items that are reworked/remanufactured to serve for demand satisfaction and those that are discarded, because there is no demand in the range of their lifetime restrictions. After remanufacturing of items, they regain full quality standards, so that they are indistinguishable of regular produced products and enter FGI. Rework/remanufacturing of returned items is further restricted by the amount of returns:

$$x^r_{it} \leq G_{it} \qquad \forall\, i,t = 1,\ldots,T \tag{9.21}$$

Constraint (9.21) prevents from *phantom* reworked items are taken from "out of nowhere" from the model and might further serve as a cutting plane to speed up the solution finding process. Reworked/remanufactured items either serve to satisfy demand d_{it} in period t or they enter regular inventory I_{it} until they are used for demand together with regularly produced items x_{it} stated in the following constraint:

$$I_{it} = I_{i,t-1} + x_{it} + x^r_{it} - d_{it} - I^D_{it} \qquad \forall\, i,t = 1,\ldots,T \tag{9.22}$$

Disposals at this stage are due to minimal lot size constraints on regular production, because these models avoid wastage in the first place as shown before; see also Pahl and Voß (2010), Pahl et al. (2011). As stated elsewhere, we solely constrain regular production to comply to a specific minimal lot size. Disposals of regular and/or reworked items on that stage are determined as follows:

$$I^D_{it} \geq \sum_{u=1}^{t-\Theta_i} (x_{iu} + x^r_{it}) - \sum_{u=1}^{t} d_{ju} - \sum_{u=1}^{t-1} I^D_{iu} \qquad \forall\, i,t = 1,\ldots,T \tag{9.23}$$

The modification to the disposal constraint is done in the first term where reworked items that are now as good as new are restricted by the regular lifetime of item i.

The complete extended CLSP-P-R-6 is stated as follows:

CLSP-P-R-6

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(sc_i \cdot y_{it} + rsc_i \cdot y_{it}^r + h_i \cdot I_{it} + h_i^r \cdot I_{it}^r + \phi_i \cdot I_{it}^D + \phi_i^r \cdot I_{it}^{Dr}\right) \tag{9.24}$$
$$+\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \mathrm{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t$$

subject to

$$I_{it} = I_{i,t-1} + x_{it} + x_{it}^r - d_{it} - I_{it}^D \qquad \forall\, i,t = 1,\ldots,T \tag{9.25}$$
$$I_{it}^r = I_{i,t-1}^r + G_t - x_{it}^r - I_{it}^{Dr} \qquad \forall\, i,t = 1,\ldots,T \tag{9.26}$$
$$x_{it}^r \le G_{it} \qquad \forall\, i,t = 1,\ldots,T$$
$$I_{it}^D \ge \sum_{u=1}^{t-\Theta_i}(x_{iu} + x_{it}^r) - \sum_{u=1}^{t} d_{iu} - \sum_{u=1}^{t-1} I_{iu}^D \qquad \forall\, i,t = 1,\ldots,T$$
$$I_{it}^{Dr} \ge \sum_{u=1}^{t-\Theta_i^r} x_{iu}^r - \sum_{u=1}^{t} G_{iu} - \sum_{u=1}^{t-1} I_{iu}^{Dr} \qquad \forall\, i,t = 1,\ldots,T$$
$$\mathrm{Cap}_t + O_t \ge \sum_{i=1}^{N}(\xi_i \cdot x_{it} + st_i \cdot y_{it} + t \cdot \xi_i^r \cdot x_{it}^r + rst \cdot y_{it}^r) \qquad \forall\, t = 1,\ldots,T \tag{9.27}$$
$$x_{it} \le \left(d^k + \mathrm{Cap}_t\right) \cdot (y_{it} + z_{it}) \qquad \forall\, i,t = 1,\ldots,T \tag{9.28}$$
$$x_{it}^r \le \left(d^k + \mathrm{Cap}_t\right) \cdot (y_{it}^r + z_{it}^r) \qquad \forall\, i,t = 1,\ldots,T \tag{9.29}$$
$$1 \ge \sum_{i=1}^{N}(z_{it} + z_{it}^r) \qquad \forall\, t = 1,\ldots,T \tag{9.30}$$
$$1 \ge z_{it} - y_{i,t-1} - z_{i,t-1} \qquad \forall\, i,t = 1,\ldots,T \tag{9.31}$$
$$1 \ge z_{it}^r - y_{i,t-1}^r - z_{i,t-1}^r \qquad \forall\, i,t = 1,\ldots,T \tag{9.32}$$
$$x_{it} \ge x_i^{\min} \cdot (y_{it} + z_{it}) \qquad \forall\, i,t = 1,\ldots,T \tag{9.33}$$
$$I_{i0} = I_{iT} = 0 \qquad \forall\, i \tag{9.34}$$

$$I_{i0}^r = I_{iT}^r = 0 \qquad \forall\, i \tag{9.35}$$
$$\mathrm{pscu}_{bt} \le y_{bt} \cdot b_B \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T$$
$$\mathrm{pscu}_{bt} \ge b_b \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B, t = 1,\ldots,T$$
$$\mathrm{pscu}_{bt} \le b_{b+1} \cdot y_{bt} \qquad \forall\, b = 1,\ldots,B-1, t = 1,\ldots,T$$
$$y_{it}, y_{it}^r, z_{it}, z_{it}^r, y_{bt} \in \{0,1\} \qquad \forall\, i,t = 1,\ldots,T \tag{9.36}$$
$$x_{it}, x_{it}^r, I_{it}, I_{it}^r, I_{it}^D, I_{it}^{Dr}, O_t \ge 0 \qquad \forall\, i,t = 1,\ldots,T \tag{9.37}$$

The objective function minimizes total costs of regular production and rework setups, inventory and rework inventory holding where the latter is less costly, disposal of regularly produced items and returned items, load-dependent utilization costs, and overtime. Setup costs are taken into account additionally to the unit times of nominal capacity included in the utilization term, so that setup variables are set correctly. Constraints (9.25) denote the inventory balance equation of regular production with the modification of the rework of returned items i that are available for demand satisfaction in period t. Constraints (9.26) state the returned items inventory balance equation that is composed by returned items t that were

in inventory in previous periods $t-1$, returned items G_{it}, less those items that are reworked x^r_{it} and disposal of returned items I^{Dr}_{it}. Constraints (9.18) require that the amount of reworked items can never be greater than returns. Inequalities (9.23) and (9.19) determine the items that are dispose due to lifetime restriction of regular or returned items and according to demand requirements. Returned items can stay in returned inventory until their *rework lifetime* given by Θ^r_i expires; see also Figure 4.1 in Section 4.1.2 on Page 16. Constraints (9.27) present the regular capacity restriction including capacity consumption due to rework and rework setups. These are stated as LP upper bounds as in the classical CLSP. LDLT are accounted for regarding the utilization term in the objective function that is penalized by load-dependent costs. Constraints (9.28) and (9.29) give the setups for regular or rework batches that can be carried over from/to the next period in both cases. Nevertheless, the machine state requires being set up exactly for one product of regular production of rework or not at all which is presented by Constraints (9.30). Inequalities (9.31) and (9.32) determine the setup carry over of machine states z_{it}. Inequalities (9.33) present the minimal lot size requirements that are required only for regular production. This can be easily modified for the specific practical case.

In case that rework/remanufacturing of returned items is waiting dependent, the capacity restriction can be modified as follows:

$$\text{Cap}_t + O_t \geq \sum_{i=1}^{N} (\xi_i \cdot x_{it} + st_i \cdot y_{it} + t \cdot \xi^r_i \cdot x^r_{it} + rst \cdot y^r_{it}) \qquad \forall\, t = 1, \ldots, T \tag{9.38}$$

so that rework/remanufacturing processing times become longer the more returns stay in rework inventory.

The resting Inequalities (9.34), (9.35), (9.36), and (9.37) denote initial and ending inventory levels, binary variables regarding setups, and non-negativity constraints.

9.2 Integration of In-Line Rework

Parts and products in the production system might be required to be reworked due to misconform to production or quality standards, if technically possible.

The following rework model includes returned items (recycling) as well as rework of passed items (interally passed). Therefore, assumptions are modified or added as follows:

30. Items that pass their useful (quality) lifetime while being in stock before the serve for demand satisfaction are passed to the internal rework system to regain quality standards;

31. Passed items need not necessarily be rework if there is not enough demand, so that such items are disposed immediately. Rework in that sense is some kind of virtual prolungation of lifetime.

Rework of internally passed items make sense in case that there is a demand in the near future for these items and rework is less costly than regular production or, in our case as this is included, remanufacturing of externally returned items.

The complete extended CLSP-P with minimal lot sizes, setup carry overs, production smoothing (CLSP-P-5) and remanufacturing/rework of returned items and internally passed items is stated as follows:

CLSP-P-R-7

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(sc_i \cdot y_{it} + rsc_i \cdot y_{it}^r + h_i \cdot I_{it} + h_i^r \cdot I_{it}^r + \phi_i \cdot I_{it}^D + \phi_i^r \cdot I_{it}^{Dr}\right)$$
$$+\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{b=1}^{B} pcu_b^f \cdot y_{bt} + \mathrm{pscu}_{bt} - pcu_b^v \cdot y_{bt} \cdot b_b\right] + oc_i \cdot O_t$$

subject to

$$\begin{aligned}
I_{it} &= I_{i,t-1} + x_{it} + x_{it}^r - d_{it} - I_{it}^D && \forall\, i,t = 1,\ldots,T\\
I_{it}^r &= I_{i,t-1}^r + G_t + I_{it}^D - x_{it}^r - I_{it}^{Dr} && \forall\, i,t = 1,\ldots,T\\
I_{it}^D &\geq \sum_{u=1}^{t-\Theta_i^L}(x_{iu} + x_{it}^r) - \sum_{u=1}^{t} d_{ju} - \sum_{u=1}^{t-1} I_{iu}^D && \forall\, i,t = 1,\ldots,T\\
I_{it}^{Dr} &\geq \sum_{u=1}^{t-\Theta_i^{Lr}}\left(G_{it} + I_{it}^D\right) - \sum_{u=1}^{t-\Theta_i^L} x_{iu}^r - \sum_{u=1}^{t-1} I_{iu}^{Dr} && \forall\, i,t = 1,\ldots,T\\
\mathrm{Cap}_t + O_t &\geq \sum_{i=1}^{N}(\xi_i \cdot x_{it} + st_i \cdot y_{it} + t \cdot \xi_i^r \cdot x_{it}^r + rst \cdot y_{it}^r) && \forall\, t = 1,\ldots,T\\
x_{it} &\leq \left(d^k + \mathrm{Cap}_t\right) \cdot (y_{it} + z_{it}) && \forall\, i,t = 1,\ldots,T\\
x_{it}^r &\leq \left(d^k + \mathrm{Cap}_t\right) \cdot (y_{it}^r + z_{it}^r) && \forall\, i,t = 1,\ldots,T\\
1 &\geq \sum_{i=1}^{N}(z_{it} + z_{it}^r) && \forall\, t = 1,\ldots,T\\
1 &\geq z_{it} - y_{i,t-1} - z_{i,t-1} && \forall\, i,t = 1,\ldots,T\\
1 &\geq z_{it}^r - y_{i,t-1}^r - z_{i,t-1}^r && \forall\, i,t = 1,\ldots,T\\
x_{it} &\geq x_i^{\min} \cdot (y_{it} + z_{it}) && \forall\, i,t = 1,\ldots,T\\
I_{i0} &= I_{iT} = 0 && \forall\, i\\
I_{i0}^r &= I_{iT}^r = 0 && \forall\, i
\end{aligned}$$

$$\begin{aligned}
\mathrm{pscu}_{bt} &\leq y_{bt} \cdot b_B && \forall\, b = 1,\ldots,B, t = 1,\ldots,T\\
\mathrm{pscu}_{bt} &\geq b_b \cdot y_{bt} && \forall\, b = 1,\ldots,B, t = 1,\ldots,T\\
\mathrm{pscu}_{bt} &\leq b_{b+1} \cdot y_{bt} && \forall\, b = 1,\ldots,B-1, t = 1,\ldots,T\\
y_{it}, y_{it}^r, z_{it}, z_{it}^r, y_{bt} &\in \{0,1\} && \forall\, i,t = 1,\ldots,T\\
x_{it}, x_{it}^r, I_{it}, I_{it}^r, I_{it}^D, I_{it}^{Dr}, O_t &\geq 0 && \forall\, i,t = 1,\ldots,T
\end{aligned}$$

Chapter 10

Numerical Study and Results

In order to analyze the behavior of the CLSP-P-R-7, the model is implemented[1] in Xpress-IVE calculating test instances on a computer with an Intel Core(TM) 2 Duo CPU processor with 1.60 GHz and 1.6 GB RAM. Relatively large test instances are calculated employing 20 products and 50 time periods. As stated by Haase and Kimms (2000), test instance sizes with less than 10 items and up to 20 periods may already be of practical relevance.

In Table 10.1 parameters are stated that are equal to all test instances mainly defining relations of parameters of returns. Returns are calculated as in Equation (9.18) where a fraction of $\alpha_i^R = 0.06$ is assumed, so that the return rates are high with 94%. Return rates for fashion clothes at catalogue retailers may be less high, but in comparison to other products still high (de Koster et al. (2002)) with around 60%. Other authors report return rates from customers to be approximately 6%, but such rates significantly vary by industries; see Rogers and Tibben-Lembke (2001). Therefore, we study two cases employing very high returns rates of 94% and very low rates of 6%.

Table 10.1: List of time-infinite remanufacturing/rework parameters for all products i

Product	rsc_i	rpc_i	rh_i	ϕ_i^r	ξ_i^r	Θ_i^r
i	$sc_i/10$	$pc_i/100$	h_i	ϕ_i	$\xi_i/10$	4

The intervals/borders and costs for utilization are given in Table 10.2. The borders need to be modified according to capacity and expected use of overtime.

Table 10.2: Cost parameters of utilization intervals

Cost factor	Interval borders b				
	0	700	850	900	5000
pcu_b^f	1	42	83	238	523
pcu_b^v	1	8	17	48	65

A number of $6 \cdot 4 \cdot 4 \cdot 4 \cdot 5 = 1.920$ test instances were generated combining different parameter profiles and variations using the test data generator described in Seipl (2009); see also Pahl et al. (2011) for a study with SDST and costs. Lifetimes are assumed to be equal for all products. They are varied assuming different parameter values. Other variations are executed for machine load profiles, *time between orderss (TBOs)*, minimal lot sizes, and demand profiles as follows:

[1] This is valid for all presented models.

- number of products $:= 20$,
- number of periods $:= 50$,
- five different demand profiles assuming a Gamma distribution,
- lifetimes $\Theta_i^L := 0, 1, 2, 3, 5, 10$,
- machine loads $:= 0.75, 0.8, 0.9, 1.0$,
- TBO profiles $:= 1, 2, 3, 4$,
- minimal lot sizes $:= 0, 20, 50, 100$,

The distribution of demand profiles on products is assumed to be rather low in order to create solvable test instances and in regard of the assumptions on used item returns. Nominal capacities are fixed to 1.000 TU that are used by the generator to calculate product coefficients and setup times for all products. Setup times are randomly generated numbers that take into account available capacity that is not exceeded in order to generate solvable test instances. The same is valid for the generation of production coefficients that checks capacity utilization in order to create multiplicators that correct processing rates, accordingly. The mean TBO is further determined using the classical EOQ formula. It expresses the frequency of demands for specific products. For instance, the $TBO = 1$ denotes that the related product is demanded in every time period.

The maximum solving time for each test instance is fixed to 60 seconds, so that optimality gaps are prevalent. The optimality gap denotes the differences between the best bound and the best incumbent solution.[2] Additionally, all test instance are feasible and at least one solution is found.

Calculation results are further tested with the *Wilcoxon signed rank test (W-SRT)* for matched pairs to compare, e.g., the diffculty to find a solution regarding a certain characteristic. The W-SRT is a statistical test that abstracts from a specific distribution for its test samples. It gives information if the location parameters of the distribution of two samples are shifted regarding a certain value generally denoted by μ. The formulation of the null hypothesis assumes that distributions are equal, i.e., $\mu = 0$. The test statistic and an error probability α give information if the null hypothesis ("the medians of the distributions are equal") can be rejected. The W-SRT is employed regarding single characteristics, e.g., TBO, item lifetimes, minimal lot sizes, loads, as well as matched pairs of liftimes/minimal lot sizes, TBO/lifetimes Θ_i^L, TBO/minimal lot sizes, and loads/minimial lot sizes. The tables show the resulting p-values of the W-SRT. These denote the probability that a certain test sample or an even rarer test sample is observed under the consideration of the null hypothesis. In other words, the p-value states the smallest test level where the regarded sample is still significant. Consequently, if the p-value is smaller or equal a chosen α-value that gives the error to reject the null hypothesis although it is true, the null hypothesis is rejected; see Duller (2008). If the p-value is greater than the α-value, the null hypothesis is not rejected implying that the distribution is not shifted. Values smaller than an α-value of 0.05 are highlighted. This α-value is employed for all tables and interpretations. The tables are symmetric due to the matching of characterizations and their corresponding optimality gaps.

[2] See also http://www.gams.com/dd/docs/solvers/xpress.pdf.

10.1 Analysis of Resulting Costs and Optimality Gaps

In a first study, the resulting costs of the test instances are analyzed that permit conclusions on the optimal plans and their characteristics. Therefore, test instances with solutions implying overall costs that are lowest are compared with those who result in highest costs are analyzed regarding their underlying parameter combinations. We expect test instances with lowest costs to be those with no minimal lot size constraints, long lifetimes of items, low loads, and high values of TBOs.

10.1.1 Test Instances with Return Rates of 94%

Table 10.3 confirms the expectations for load values and TBOs, but surprisingly not for lifetimes that are rather arbitrary with values ranging from $\Theta_i^L = 2$ to $\Theta_i^L = 5$. Furthermore, test instances $1010 - 1014$ and $1510 - 1514$ have minimal lot size constraints of $x_i^{\min} = 20$. The number of rework setups is significantly high over those of regular production which is due to the high return rates of 94% and the preference on rework implied by cost parameters.

Table 10.3: Test instances with lowest costs and return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
1010-1014	6.03	3	3	0.75	20	9.788.120	92	444
1510-1514	6.47	5	4	0.75	20	9.799.570	69	452
1495-1499	6.27	3	4	0.75	0	9.801.950	96	457
1478	7.11	2	4	0.75	0	9.803.960	124	464
1475-1479	7.39	2	4	0.75	0	9.803.960	124	464
1035-1039	6.55	5	3	0.75	0	9.807.570	98	444

On the other hand, it is expected that test instances with highest costs are those that include lifetimes of $\Theta_i^L = 0$ which implies a *lot-for-lot* policy and minimal lot size constraints with high values which is confirmed by results given in Table 10.4. A lot-for-lot policy is required due to zero lifetimes of items resulting in a storage inflexibility. It denotes a plan that includes lot sizes that match demand of products in the regarded period, so that inventory is not build up and the machine has to be set up every time. Lot sizes are not aggregated. The model does not need to evaluate various aggregations of lot sizes and inventory combinations neither for regular production nor for rework, so that these test instances are solved rather fast and close to optimality, but to the expense of highly increased costs. A zero lifetime is included in this study to demonstrate the well behavior of the model. Nevertheless, a lifetime of $\Theta_i = 0$ is rarely of practical relevance and there is no need for a mathematical model to find a solution.

Table 10.4: Test instances with highest costs and return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
222-224	35.28	0	1	0.8	100	16.305.400	244	453
220-221	35.29	0	1	0.8	100	16.305.400	244	453
1420-1424	9.42	0	3	1	100	17.105.500	257	441
940-944	10.03	0	2	1	100	17.218.800	256	438
460-464	13.37	0	4	1	100	17.717.300	254	433

Results regarding TBOs are arbitrary with values ranging from $TBO = 1$ to $TBO = 4$. Test instances with highest costs show mainly highest loads except for test instances

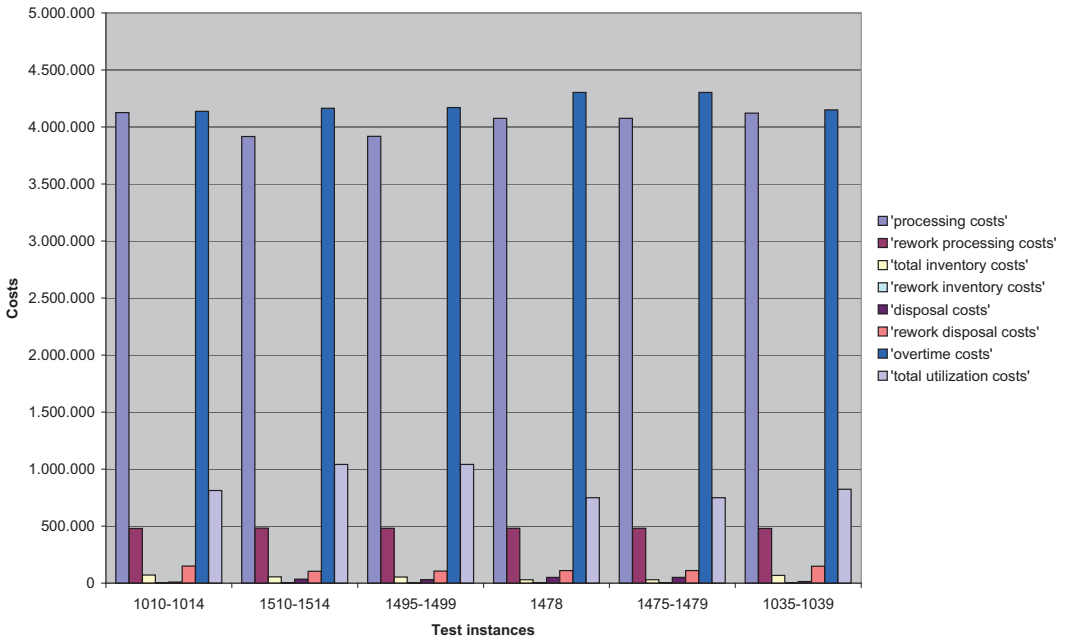

Figure 10.1: Lowest overall costs regarding test instances with return rates of 94%.

$222-224$ and $220-221$ with loads of 0.8%. Related detailed costs are depicted in Figures 10.1 and 10.2 and their values given in Tables 10.5 and 10.6.

Table 10.6 shows that inventory holding is not permitted for intances with zero lifetimes regarding regular production. Minimal lot size constraints with value $x_i^{\min} = 100$ characterizing test instances with highest costs cause increased processing costs in contrast to those with lowest costs that are not subject to such constraints or that contain rather low requirements. High values for minimal lot size constraints further lead to enhanced disposal costs which is also valid for rework disposals in comparison to test instances with lowest overall costs. Overtime costs are superior for test instances with highest overall costs than for those with lowest overall costs that might also be due to minimal lot size constraints. Utilization costs for test instances with both lowest and highest costs are significantly increased for some test instances with a $TBO = 4$ except for test instances 1478 and 1475-1479.

Figures 10.1 and 10.2 demonstrates that the significantly highest cost categories for test instances with both lowest and highest overall costs are the same, e.g., processing and overtime costs. Nevertheless, this might be also due to cost parameters assumed for the numerical study, so this has to be relativized.

The influences of parameters and thus test instance combinations on the solution process expressed by the optimality gap due to solution time restrictions are analyzed subsequently. Test instances that experience lowest and highest optimality gaps have characteristics as shown in Tables 10.7 and 10.8.

Test instances with lowest optimality gaps are those that have a zero lifetime with $\Theta_i^L = 0$ and zero minimal lot size constraints. In fact, we expect the minimal lot size to have a high impact on the solution process that has been already experienced while testing the model behavior with small numerical examples. Although the load of test instances resulting in lowest overall costs is highest with 1.0%, it does not have a great influence the solution process. This is confirmed by the W-SRT in Section 10.4.

Test instances with highest optimality gaps are those that have rather high values of minimal lot size constraints. However, characteristics regarding lifetimes, loads, and TBOs are arbitrary ranging from $\Theta_i^L = 0$ to $\Theta_i^L = 10$, $TBO = 1$ to $TBO = 4$, loads of 0.750.9% and minimal lot sizes from $x_i^{\min} = 20$ to $x_i^{\min} = 100$. High values for minimal lot size constraints

Table 10.5: Test instances with lowest overall costs with detailed cost categories for test instances with return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Costs regarding processing	rework processing	inventory	rework inventory	disposal	rework disposal	overtime	utilization
1010-1014	6.03	3	3	0.75	20	9.788.120	4.126.300	478.724	71.331	4.108	9.523	149.731	4.136.650	811.751
1510-1514	6.47	5	4	0.75	20	9.799.570	3.917.110	482.036	54.870	1.871	33.923	104.651	4.163.740	1.041.370
1495-1499	6.27	3	4	0.75	0	9.801.950	3.918.760	481.837	53.254	1.870	30.273	105.024	4.170.480	1.040.460
1478	7.11	2	4	0.75	0	9.803.960	4.075.520	481.314	30.642	2.468	51.155	110.799	4.302.120	749.943
1475-1479	7.39	2	4	0.75	0	9.803.960	4.075.520	481.314	30.642	2.468	51.155	110.799	4.302.120	749.943
1035-1039	6.55	5	3	0.75	0	9.807.570	4.122.190	478.960	67.325	3.807	13.411	148.865	4.149.370	823.644

Table 10.6: Test instances with highest overall costs with detailed cost categories for test instances with return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Costs regarding processing	rework processing	inventory	rework inventory	disposal	rework disposal	overtime	utilization
222-224	35.28	0	1	0.8	100	16.305.400	7.014.760	480.220	0	35.722	617.122	522.488	7.456.730	178.314
220-221	35.29	0	1	0.8	100	16.305.400	7.014.760	480.220	0	35.722	617.122	522.488	7.456.730	178.314
1420-1424	9.42	0	3	1	100	17.105.500	8.608.960	448.999	0	14.726	311.547	576.057	6.445.420	699.820
940-944	10.03	0	2	1	100	17.218.800	8.960.950	441.603	0	27.111	234.033	628.141	5.934.740	992.207
460-464	13.37	0	4	1	100	17.717.300	9.110.410	439.109	0	10.480	214.049	629.947	6.210.230	1.103.090

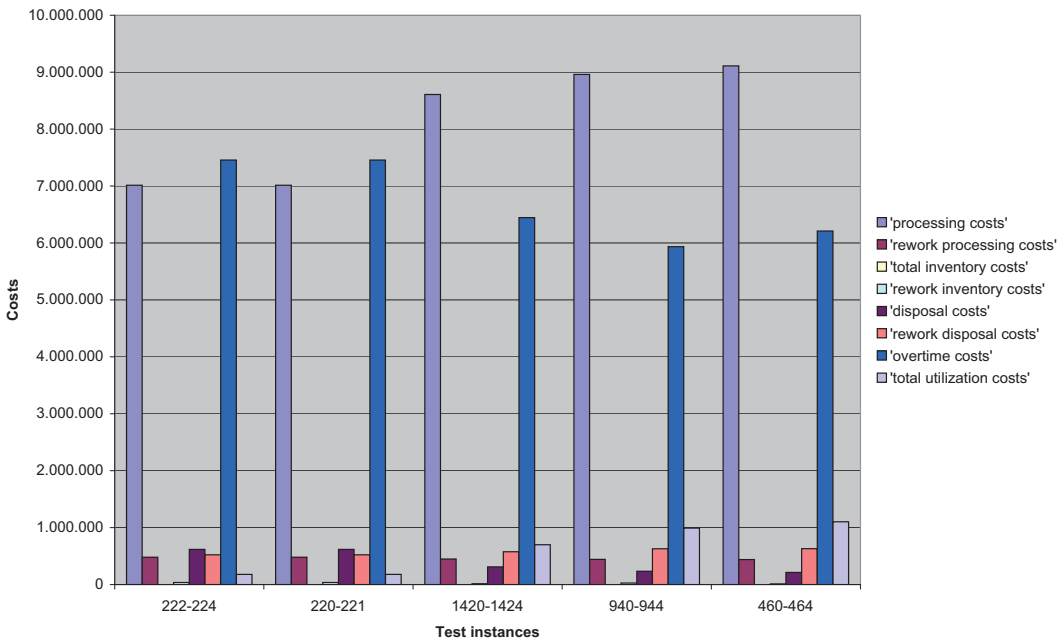

Figure 10.2: Highest overall costs regarding test instances with return rates of 94%.

Table 10.7: Test instances with lowest optimality gaps and return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
1919	0.15	0	4	1	0	15.928.300	412	421
1435-1439	0.71	0	3	1	0	15.933.700	418	422
1916-1918	0.81	0	4	1	0	15.928.300	412	421
475	0.83	0	1	1	0	16.014.300	412	423
1915	0.84	0	4	1	0	15.928.300	412	421
476-479	0.86	0	1	1	0	16.019.000	418	425
955	0.99	0	2	1	0	15.944.600	421	422
956-959	1.02	0	2	1	0	15.944.600	421	422
355-359	1.45	0	1	0.9	0	13.666.100	437	441

Table 10.8: Test instances with highest optimality gaps and return rates of 94%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
222-224	35.28	0	1	0.8	100	16.305.400	244	453
220-221	35.29	0	1	0.8	100	16.305.400	244	453
1765-1769	35.99	10	4	0.9	50	12.260.900	36	411
1540	40.94	0	4	0.75	100	16.007.800	278	421
810-814	41.06	10	2	0.9	20	12.362.700	90	423
1541-1544	41.32	0	4	0.75	100	16.007.800	278	421
1290-1294	41.38	10	3	0.9	20	12.400.100	43	423

are expected to cause significant optimality gaps, but short lifetimes and $TBOs = 3,4$ are expected to facilitate the solution process. Detailed analysis is provided subsequently.

10.1.2 Test Instances with Return Rates of 6%

The test instances that experience return rate of 6% demonstrate optimality gaps of 0% as shown in the following tables. Overall costs are significantly higher than those for return rates of 94%, because regular production is more expensive than rework of returned items or internally passed products.

Table 10.9: Test instances with lowest costs and return rates of 6%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
0	0.02	1	1	0.75	100	48.617.600	491	467
1465-1469	0	2	4	0.75	50	48.875.200	490	482
1480-1484	0	3	4	0.75	100	48.875.300	493	481
970-973	0	1	3	0.75	20	48.875.300	489	481
1475-1479	0	2	4	0.75	0	48.875.400	493	470
1485-1488	0	3	4	0.75	50	48.875.400	491	473

These results are similar to those for return rate of 94%. Nevertheless, the minimal lot sizes show higher values and the number of setups for regular production are enhanced due to an increased regular production, because returns are lower. Surprisingly, test instances with $\Theta_i^L = 1$ pertain to those with the lowest overall costs which is unexpected.

Table 10.10: Test instances with highest costs and return rates of 6%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
940-944	0	0	2	1	100	49.787.600	491	442
469-467	0	0	1	1	50	49.787.700	491	443
471-473	0	0	1	1	50	49.787.800	491	443
476-477	0	0	1	1	50	49.788.000	491	443
460-463	0	0	1	1	50	49.788.300	491	445

Table 10.10 presents a similar picture regarding the test instances with highest resulting costs including return rates of 94%. Highest costs of test instances are caused by a lot-for-lot policy due to zero lifetimes. Furthermore, loads are highest. Surprisingly, minimal lot sizes are not necessarily at the highest value which is true only for some test instances. However, the resting value are high with $x^{\min_i} = 50$.

Table 10.11 gives the results of those test instances that demonstrate highest optimality gaps. In comparison to those test instances with return rates of 94%, optimality gaps are very low which is due to the return rate of 6%. Consequently, the return rate has a great influence on the solution process.

Table 10.12 and Table 10.13 show detailed costs of the test instances that result in plans with lowest and highest costs. The processing costs are the greatest part of overall costs that are higher than for those test instances with 94% return rates. This is due to enhanced regular production due to low return rates of 6%. Disposals and rework disposals are for all test instances near to zero despite for test instance 0 and $970-973$. No overtime takes place for those test instances with lowest overall costs that might be due to a rather low load of 75%.

Table 10.11: Test instances with highest optimality gap and return rates of 6%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	Nr. of setups	Nr. of rework setups
1856-1859	0.53	3	4	1	0	49.650.900	373	441
842	0.54	1	2	1	100	49.684.400	403	455
1810-1814	0.54	1	4	1	20	49.656.300	447	443
840-844	0.56	1	2	1	100	49.692.300	456	441
856-859	0.56	1	2	1	0	49.694.900	381	423
850-853	0.57	1	2	1	20	49.684.900	478	445
855-858	0.57	1	2	1	0	49.697.400	453	442
1320-1324	0.58	1	3	1	100	49.688.200	422	441

10.2 Influences of Lifetime Constraints

The lifetime is expected to have a significant influence on the solution process and thus on the optimality gap. Surprisingly, results shown in Table 10.14 reveal that the mean optimality gap is rather small for test instances with a lifetime of $\Theta_i^L = 2$ and $\Theta_i^L = 3$ which was not expected. Test instances with lifetimes of $\Theta_i^L = 0$ have a rather high mean optimality gap. The same is valid for test instances with $\Theta_i^L = 10$ lifetimes.

Table 10.15 shows another picture regarding the influence of lifetimes of the optimality gap where, e.g., test instances with $\Theta_i^L = 0$ experience optimality gaps between $0.15 - 41.32\%$, so that it might be a reasonable assumption that another parameter may cause these significant differences which is expected to be the minimal lot size. Therefore, further analysis including a pairwise W-SRT regarding lifetimes and minimal lot sizes is executed and results are given in Tables 13.1 and 13.2 in the appendices.

The p-values determined by executing the W-SRT regarding the influence of lifetimes on the optimality gap are given in Table 10.16 revealing that the median of distributions is not equal for lifetimes $\Theta_i^L = 2$ and $\Theta_i^L = 5$ as well as for $\Theta_i^L = 1$ and $\Theta_i^L = 10$. This is also confirmed by some results given in Table 10.15. These show great differences between resulting optimality gaps for test instances with lifetimes $\Theta_i^L = 5, 10$, but also for $\Theta_i^L = 0$ which is not confirmed by the p-values in Table 10.16.

Results in Table 10.16 further show that the medians of distributions for $\Theta_i^L = 2$ and $\Theta_i^L = 3$ are not significantly shifted, so that the probability to observe a certain test sample or an even rarer one is rather high with 88.7%. The same is valid for $\Theta_i^L = 0$ and $\Theta_i^L = 2$, but only with a probability of 48.9%.

Table 10.17 presents the p-values regarding the lifetime of test instances with 6% return rate which gives a different picture of the influence of the lifetime on the solution process where the p-values are significantly smaller than the error probability of $\alpha = 0.05$ for test instances with $\Theta_i^L = 1$ and all other lifetimes despite $\Theta_i^L = 0$. For lifetimes $\Theta_i^L = 2$ and $\Theta_i^L = 3$, the p-values are high both for return rate 6% as well as 94% with 88.7% and 99.2%, respectively.

The W-SRT on pairwise parameters regarding lifetime and minimal lot sizes given in Tables 13.1 and 13.2. Despite some exceptions, it is observable that the differences in the medians of the distributions regarding the test instances are significant for lifetimes $\Theta_i^L = 1$ and $\Theta_i^L = 10$ versus $\Theta_i^L = 2$ and $\Theta_i^L = 3$, but regarding all minimal lot sizes values, so that this characteristic seems not be responsible for the differences regarding the optimality gap for test instances with the same lifetime. Nevertheless, while closer examining the characteristics of test instances with return rates of 94% and their mean optimality gaps, it becomes obvious that high mean optimality gaps are due to a high value of the minimal lot size constraints independent of the other characteristics such as the load or the TBO as shown in Table 10.18. Values greater than 10% are highlighted. The complete Tables 13.3 and 13.4

Table 10.12: Test instances with lowest overall costs with detailed cost categories for test instances with return rates of 6%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	processing	Costs regarding rework processing	inventory	rework inventory	disposal	rework disposal	overtime	utilization
0	0.02	1	1	0.75	100	48.617.600	48.566.500	31.001	7.617	1	29	0	0	12.404
1465-1469	0	2	4	0.75	50	48.875.200	48.834.000	31.171	3.840	17	0	0	0	6.150
1480-1484	0	3	4	0.75	100	48.875.300	48.834.000	31.171	3.499	9	0	0	0	6.615
970-973	0	1	3	0.75	20	48.875.300	48.834.000	31.180	3.675	19	188	0	0	6.180
1475-1479	0	2	4	0.75	0	48.875.400	48.834.000	31.171	3.345	13	0	0	0	6.789
1485-1488	0	3	4	0.75	50	48.875.400	48.834.000	31.171	4.181	61	0	0	0	5.914

Table 10.13: Test instances with highest overall costs with detailed cost categories for test instances with return rates of 6%

Nr. of test instance	optimality gap in %	Θ_i^L	TBO	load	$x_i^{\min}$	objective	processing	Costs regarding rework processing	inventory	rework inventory	disposal	rework disposal	overtime	utilization
940-944	0	0	2	1	100	49.787.600	48.834.000	31.171	0	273	0	0	365.474	556.601
469-467	0	0	1	1	50	49.787.700	48.834.000	31.171	0	518	0	0	365.412	556.541
471-473	0	0	1	1	50	49.787.800	48.834.000	31.171	0	547	0	0	365.461	556.589
476-477	0	0	1	1	50	49.788.000	48.834.000	31.171	0	586	0	0	365.537	556.661
460-463	0	0	1	1	50	49.788.300	48.834.000	31.171	0	480	0	0	365.771	556.886

Table 10.14: Mean optimality gap of different lifetimes for test instances with return rates of 94%

Mean optimality gap in %	Lifetime	Mean best solution
5.84	$\Theta_i^L = 2$	11.989.084
5.91	$\Theta_i^L = 3$	11.904.251
6.95	$\Theta_i^L = 5$	11.833.043
6.99	$\Theta_i^L = 1$	12.271.117
8.01	$\Theta_i^L = 0$	13.413.630
11.12	$\Theta_i^L = 10$	11.849.873

Table 10.15: Analysis regarding the variations of lifetimes regarding test instances with return rates of 94%

Lifetime	Optimality gap in %		Best solution	
	lowest	highest	lowest	highest
$\Theta_i^L = 0$	0.15	41.32	15.928.290	16.007.751
$\Theta_i^L = 1$	3.48	17.55	15.169.797	11.066.587
$\Theta_i^L = 2$	3.03	9.34	14.792.317	9.979.803
$\Theta_i^L = 3$	1.92	11.25	14.573.429	10.310.650
$\Theta_i^L = 5$	2.44	29.25	14.405.461	14.436.627
$\Theta_i^L = 10$	2.33	41.38	14.411.491	12.400.745

Table 10.16: W-SRT regarding Lifetimes and their influence on the optimality gap for test instances with return rates of 94%

	$\Theta_i^L = 0$	$\Theta_i^L = 1$	$\Theta_i^L = 10$	$\Theta_i^L = 2$	$\Theta_i^L = 3$	$\Theta_i^L = 5$
$\Theta_i^L = 0$	0.00000	0.00000	0.00000	0.48957	0.23532	0.00642
$\Theta_i^L = 1$	0.00000	0.00000	**0.00140**	0.00000	0.00000	0.00000
$\Theta_i^L = 10$	0.00000	**0.00140**	0.00000	0.00000	0.00000	0.00000
$\Theta_i^L = 2$	0.48957	0.00000	0.00000	0.00000	0.88728	0.17042
$\Theta_i^L = 3$	0.23532	0.00000	0.00000	0.88728	0.00000	0.08047
$\Theta_i^L = 5$	**0.00642**	0.00000	0.00000	0.17042	0.08047	0.00000

Table 10.17: W-SRT regarding Lifetimes and their influence on the optimality gap for test instances with return rates of 6%

	$\Theta_i^L = 0$	$\Theta_i^L = 1$	$\Theta_i^L = 10$	$\Theta_i^L = 2$	$\Theta_i^L = 3$	$\Theta_i^L = 5$
$\Theta_i^L = 0$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\Theta_i^L = 1$	0.00000	0.00000	**0.01318**	**0.00528**	**0.01171**	**0.00030**
$\Theta_i^L = 10$	0.00000	**0.01318**	0.00000	0.87080	0.79250	0.30999
$\Theta_i^L = 2$	0.00000	**0.00528**	0.87080	0.00000	0.99241	0.21444
$\Theta_i^L = 3$	0.00000	**0.01171**	0.79250	0.99241	0.00000	0.19892
$\Theta_i^L = 5$	0.00000	**0.00030**	0.30999	0.21444	0.19892	0.00000

including all minimal lot size constraints is given in the Apendice.

The p-values for test instances with return rates of 6% demonstrate different results; see Table 14.1 and 14.2 in the Appendice. For instance, the p-value regarding $\Theta_i^L = 0$ and all values of minimal lot sizes for all other pairs regarding lifetimes are significantly lower than the error probability assumed with $\alpha = 0.05$, despite pairs (5;50) versus (0;100), (0;20), and (0;50), but these have a rather low probability to observe such sample or an even rarer sample ranging from 58% to 9.24%. Therefore, there is a great influence regarding a zero lifetime on the solution process. Another observation that can be made regards the return rates that cause very different results. Consequently, return rates are important aspects of the model leading to very different results and inexpected behaviors.

Table 10.18: Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 50$ and $x_i^{\min} = 100$

			Lifetimes					
$x_i^{\min}$	Load	TBO	$\Theta_i^L = 0$	$\Theta_i^L = 1$	$\Theta_i^L = 2$	$\Theta_i^L = 3$	$\Theta_i^L = 5$	$\Theta_i^L = 10$
100	0.75	1	22.81	9.64	7.77	11.25	9.91	10.78
		2	17.86	10.27	8.88	8.04	7.72	9.25
		3	20.89	8.75	7.51	7.34	7.76	8.19
		4	41.24	10.03	7.21	7.52	8.39	7.48
	0.8	1	35.28	8.09	7.51	7.48	7.21	10.01
		2	16.86	7.15	5.65	5.81	6.82	6.10
		3	17.35	8.11	7.66	5.65	6.24	8.11
		4	30.39	8.60	6.77	6.22	6.59	6.93
	0.9	1	15.39	5.98	6.47	5.21	4.62	5.44
		2	13.45	5.54	3.91	4.64	4.75	5.09
		3	13.64	4.98	4.00	5.23	5.33	28.74
		4	13.39	6.03	5.96	5.73	5.58	6.22
	1.0	1	10.03	3.48	3.94	2.11	2.44	2.33
		2	10.03	3.80	3.03	3.77	4.59	24.69
		3	9.42	4.19	3.36	3.49	4.76	22.00
		4	13.37	4.48	3.06	4.09	4.50	5.53

10.3 Influence of Minimal Lot Size Constraints

As already reviously examined and confirmed, minimal lot size constraints significantly influence the solution process, so that test instances with rather high minimal lot sizes have a higher optimality gap which is already demonstrated regarding lifetime constraints in Section 10.2. Table 10.19 presents the mean optimality gaps and mean best solutions of the objective regarding the minimal lot size constraints for test instances with return rates of 94%. The mean optimality gap is smallest for test instances that have a minimal lot size constraint of $x_i^{\min} = 0$ and highest for test instances with $x_i^{\min} = 100$. Nevertheless, optimality gaps vary between $0.15 - 29.22\%$ for $x_i^{\min} = 0$ and $1.92 - 41.32\%$ for $x_i^{\min} = 100$ as shown in Table 10.20 which is not surprising due to previous analysis.

Mean optimality gaps for minimal lot size constraints of $x_i^{\min} = 20$ and $x_i^{\min} = 50$ are rather close that is also shown in Table 10.20 and Table 10.21 where the latter presents the p-values of the related W-SRT. The p-values given in Table 10.21 confirm no significant differences for these minimal lot size pairs. However, p-values of minimal lot sizes $x_i^{\min} = 20$ and $x_i^{\min} = 0$, $x_i^{\min} = 50$ and $x_i^{\min} = 0$ as well as $x_i^{\min} = 50$ and $x_i^{\min} = 100$ denote significant differences in the medians of the distributions regarding test instances with such characteristics. Moreover, specific optimality gaps of test instances with these charac-

Table 10.19: Mean optimality gap of different minimal lot sizes for test instances with return rates of 94%

Mean optimality gap in %	Minimal lot size	Mean best solution
6.56	$x_i^{\min} = 0$	12.103.680
7.11	$x_i^{\min} = 50$	12.154.666
7.25	$x_i^{\min} = 20$	12.112.402
8.97	$x_i^{\min} = 100$	12.469.918

teristics significantly vary as shown in Table 10.20, so further analysis on their underlying characteristics is provided in Table 10.23.

Table 10.20: Analysis regarding the variations of minimal lot sizes for test instances with return rates of 94%

Minimal lot size	Optimality gap in %		Best solution	
	lowest	highest	lowest	highest
$x_i^{\min} = 0$	0.15	29.22	15.928.290	9.952.864
$x_i^{\min} = 20$	2.33	41.38	15.930.762	12.400.745
$x_i^{\min} = 50$	2.44	35.99	15.977.751	12.260.896
$x_i^{\min} = 100$	1.92	41.32	14.573.429	16.007.751

Table 10.21: W-SRT analyzing the influences of the minimal lot size constraint on the optimality gap retarding test intances with 94% return rates

	$x_i^{\min} = 0$	$x_i^{\min} = 100$	$x_i^{\min} = 20$	$x_i^{\min} = 50$
$x_i^{\min} = 0$	0.00000	0.00000	**0.04069**	**0.00012**
$x_i^{\min} = 100$	0.00000	0.00000	0.00000	**0.00003**
$x_i^{\min} = 20$	**0.04069**	0.00000	0.00000	0.10581
$x_i^{\min} = 50$	**0.00012**	**0.00003**	0.10581	0.00000

In contrast to Table 10.21, Table 10.22 presenting the p-values for test instances with return rates of 6% show no significant differences between the medians of the distributions regarding the difficulty to find the optimality gap. Nevertheless, probabilities to observe the same or an even rarer test sample are with 11.7% for $x_i^{\min} = 100$ versus $x_i^{\min} = 0$ and 28.46% for $x_i^{\min} = 50$ versus $x_i^{\min} = 0$.

Table 10.18 also provides information on the influence of the minimal lot size and lifetimes on the mean optimality gap of test instances with return rates of 94%.

Table 10.23 demonstrates that for a lifetime $\Theta_i^L = 0$, a minimal lot size constraint $x_i^{\min} = 100$, and all load-values, the mean optimality gap is higher than 10%. Mean optimality gaps greater than 10% are highlighted in the table revealing high values for $\Theta_i^L = 10$, $x_i^{\min} = 100$, and loads $0.9 - 1.0$ as well as $\Theta_i^L = 10$ and load 1.0 for all minimal lot size parameters.

10.4 Influences of the Load on the Optimality Gap

The load profiles are taken in order to calculate the production coefficients and setup times. They might have an influence on the optimality gap which is examined using the W-SRT.

Table 10.22: W-SRT analyzing the influences of the minimal lot size constraint on the optimality gap retarding test intances with 6% return rates

	$x_i^{\min} = 0$	$x_i^{\min} = 100$	$x_i^{\min} = 20$	$x_i^{\min} = 50$
$x_i^{\min} = 0$	0.00000	0.11744	0.44639	0.28465
$x_i^{\min} = 100$	0.11744	0.00000	0.43796	0.64644
$x_i^{\min} = 20$	0.44639	0.43796	0.00000	0.78432
$x_i^{\min} = 50$	0.28465	0.64644	0.78432	0.00000

Table 10.23: Average gap in % regarding test instances with return rates of 94% sorted by minimal lot sizes, lifetimes and loads

		Minimal lot sizes			
Θ_i^L	Load	$x_i^{\min} = 0$	$x_i^{\min} = 20$	$x_i^{\min} = 50$	$x_i^{\min} = 100$
0	0.75	5.46	4.22	7.37	25.70
	0.8	4.78	6.00	7.09	24.97
	0.9	2.03	4.23	4.32	13.97
	1	0.82	2.93	3.51	10.71
1	0.75	8.38	9.09	11.04	9.67
	0.8	7.46	7.86	7.94	7.99
	0.9	6.44	6.17	7.10	5.63
	1	4.62	4.43	4.04	3.99
2	0.75	7.62	8.17	7.80	7.85
	0.8	6.21	7.20	7.49	6.90
	0.9	4.64	4.72	5.13	5.09
	1	4.12	3.73	3.46	3.35
3	0.75	7.26	7.79	8.13	8.54
	0.8	6.84	7.16	6.13	6.29
	0.9	4.54	5.51	6.55	5.20
	1	3.86	3.81	3.67	3.37
5	0.75	12.27	7.48	7.65	8.44
	0.8	6.54	6.66	6.69	6.71
	0.9	5.36	5.06	4.96	5.07
	1	9.61	4.36	10.31	4.07
10	0.75	7.75	14.65	8.64	8.92
	0.8	6.05	7.36	6.87	7.79
	0.9	4.83	23.20	13.24	11.37
	1	19.93	12.15	11.47	13.64

However, the resulting *p*-values are given in Table 10.24 and show that the load profiles have no influence on the optimality gap.

Table 10.24: Influences of the load on the optimality gap retarding test intances with 94% return rates

	(0.75)	(0.8)	(0.9)	(1.0)
(0.75)	0.00000	0.00000	0.00000	0.00000
(0.8)	0.00000	0.00000	0.00000	0.00000
(0.9)	0.00000	0.00000	0.00000	0.00000
(1.0)	0.00000	0.00000	0.00000	0.00000

Table 10.25: Influences of the load on the optimality gap retarding test intances with 6% return rates

	(0.75)	(0.8)	(0.9)	(1.0)
(0.75)	0.00000	0.00000	0.00000	0.00000
(0.8)	0.00000	0.00000	0.75125	0.00000
(0.9)	0.00000	0.75125	0.00000	0.00000
(1.0)	0.00000	0.00000	0.00000	0.00000

Resulting *p*-values in Table 10.25 for return rates of 6% denote a probability to observe the same or an even rarer sample with 75.12% for loads 0.9 and 0.8.

10.5 Influences of the TBO

The mean time between orders (TBO) is calculated by the EOQ formula. Table 10.26 presents the mean optimality gaps and mean best solutions of test instances with the same TBO-values. A TBO-value of 1 results in lowest mean optimality gaps which is also confirmed by Table 10.27. Nevertheless, optimality gaps vary for this TBO-value between $0.83 - 35.29\%$. Differences between the lowest and highest optimality gap are largest for $TBO = 4$ which might be due to minimal lot size constraints.

Table 10.26: Analysis regarding the variations of *TBO* for test instances with return rates of 94%

TBO	Mean optimality gap in %	Mean best solution
$TBO = 1$	6.81	12.297.877
$TBO = 3$	7.67	12.150.428
$TBO = 2$	7.69	12.204.705
$TBO = 4$	7.72	12.187.655

The resulting *p*-values for the W-SRT on the influence of the TBOs on the optimality gap is given in Table 10.28. They are above the error probability $\alpha = 0.05$, so that there is no significant differences between the medians of the distributions regarding the optimality gap of related test instances. Nevertheless, for TBO-pairs (2;1) and (4;1), the probabilities

Table 10.27: Analysis regarding the variations of TBOs for test instances with return rates of 94%

TBO	Optimality gap in %		Best solution	
	lowest	highest	lowest	highest
TBO=1	0.83	35.29	16.014.317	16.305.356
TBO=2	0.99	41.06	15.944.802	12.364.183
TBO=3	0.71	41.38	15.933.709	12.400.745
TBO=4	0.15	41.32	15.928.290	16.007.751

Table 10.28: Influences of the TBO on the optimality gap retarding test intances with 94% return rates

	TBO=1	TBO=2	TBO=3	TBO=4
TBO=1	0.00000	0.11776	**0.60146**	**0.26619**
TBO=2	0.11776	0.00000	**0.60568**	**0.94673**
TBO=3	**0.60146**	**0.60568**	0.00000	**0.53889**
TBO=4	**0.26619**	**0.94673**	**0.53889**	0.00000

to observe a certain test sample or an even rarer one are 11% and 26%, respectively. The differences in the optimality gaps presented in Table 10.27 are examined more closely in Table 10.30.

In contrast to this, the p-values for test instances with a return rate of 6% given in Table 10.29 demonstrate significant differences in the medians of distributions regarding TBOs. This is rather surprising as greater differences with greater TBO-differences are expected.

Table 10.29: Influences of the TBO on the optimality gap retarding test intances with 6% return rates

	TBO=1	TBO=2	TBO=3	TBO=4
TBO=1	0.00000	**0.00625**	0.00000	0.00000
TBO=2	**0.00625**	0.00000	**0.00162**	**0.00001**
TBO=3	0.00000	**0.00162**	0.00000	0.17445
TBO=4	0.00000	**0.00001**	0.17445	0.00000

The results given in Table 10.30 reveal that the differences of optimality gaps presented in Table 10.27 result mainly from minimal lot size constraints, especially from a value $x_i^{\min} = 100$ for all TBO-values where the optimality gap is highest for $TBO = 4$ and load 0.75% with 24.6%. Furthermore, mean optimality gaps are larger than 10% for test instances with lifetime $\Theta_i^L = 10$ and TBO-values $2-4$. The test instances with $\Theta_i^L = 10$ and $TBO = 3$ resulting in a mean optimality gap of 25.97% have additional characteristics, i.e. a minimal lot size constraint of $x_i^{\min} = 20$, load of either 0.75 or 0.9 that give highest optimality gaps of 41.38% and 33.88% which is rather unexpected regarding the minimal lot size value.

10.6 W-SRT Regarding for Pairwise Parameter Combinations

The influence of the combination regarding lifetime and minimal lot sizes are already studied in Section 10.2. In the following, the combined influences of the TBO/lifetime, the load/minimal lot size, and the TBO/minimal lot size are examined. Related tables are given

Table 10.30: Average gap in % regarding test instances with return rates of 94% sorted by TBOs, lifetimes and minimal lot sizes

		Mean time between orders (TBO)			
Θ_i^L	$x_i^{\min}$	TBO=1	TBO=2	TBO=3	TBO=4
0	0	2.95	4.04	2.94	3.16
	20	4.57	4.67	3.83	4.31
	50	5.92	5.81	5.82	4.75
	100	20.88	14.55	15.33	24.60
1	0	6.43	6.97	6.79	6.69
	20	6.49	6.87	7.52	6.65
	50	6.86	8.41	7.26	7.59
	100	6.80	6.69	6.51	7.29
2	0	5.72	5.64	5.36	5.88
	20	5.99	5.86	5.95	6.01
	50	6.10	6.49	5.33	5.95
	100	6.42	5.37	5.63	5.75
3	0	5.59	5.70	5.48	5.73
	20	6.31	5.89	5.68	6.39
	50	6.11	5.99	5.75	6.63
	100	6.51	5.57	5.43	5.89
5	0	10.22	5.74	5.57	12.25
	20	5.89	5.78	5.99	5.90
	50	6.17	11.68	5.98	5.78
	100	6.04	5.97	6.02	6.26
10	0	5.67	10.93	10.54	11.42
	20	6.30	19.03	25.97	6.06
	50	6.28	10.27	6.65	17.01
	100	7.14	11.28	16.76	6.54

in the Appendice. An error probability of $\alpha = 0.1$ is regarded and p-values highlighted, accordingly.

Tables 13.5 presents the p-values regarding the pairwise tested TBO and lifetime and their influence on the optimality gap. Significant differences in the means of the distributions regarding the test instances exist, e.g., for pairs (2;1) and (1;0) with p-value 6%. The same is valid for (2;10) and (1;0) with 0.001%. Moreover, (4;1) and (4;10) versus (1;0) demonstrate significantly different medians given by low p-values, i.e., 4.88% and 0.12$. The p-values regarding pair (3;1) and all other pairs show also significant differences despite pair (2;1), (4;1), and (4;10); see row of (3;1) in Table 13.5. The latter is not surprising, because the lifetimes are equal despite for (4;10). Furthermore, the pairs (1;1) and all pairs for the lifetimes and $TBO = 2$ despite $\Theta_i^L = 1$ as well as all liftimes and $TBO = 3$ show significant differences. The same is valid for pairs (4;0) and (1;1) with a p-value of 0.76% or pair (4;10) versus all pairs (1;all) despite lifetime $\Theta_i^L = 10$, so that the differences in the medians regarding the distributions of the test instances might derive from the different TBO-values. Surprisingly, the differences are not significant for lifetimes $\Theta_i^L = 1$ versus $\Theta_i^L = 10$. We expect these lifetimes to create significant differences. This migh derive from a rework lifetime $\Theta_i^r = 4$, but further test are needed on this issue. Regarding the pairs (4;1) versus pairs with $TBO = 2$ including all lifetimes despite $\Theta_i^L = 1$, the p-values are also very low ranging from 0.003% to 0.499%. The same is valid for pairs (4;10) despite pair (2;3) with p-values ranging from 0.013% to 4.96%. Therefore, $TBO = 2$ and $TBO = 4$ demonstate differences. Surprisingly, the same is valid for (1;1) versus pair with $TBO = 2$ and $TBO = 3$ including all liftime values despite (2;1) that has probability of 76.4% to observe a certain or even lower sample, so that also $TBO = 1$, $TBO = 2$, and $TBO = 3$ have significant differences including nearly all lifetimes which is not expected. Tables 14.1 and 14.2 presents different results where clear differences exist between $TBO = 1$ and all other TBO-values including

all liftime values despite $\Theta_i^L = 0$.

Table 13.7 give the p-values for pairs regarding the TBO and minimal lot sizes. Also here, the influence of the minimal lot size seem prevalent, e.g., regarding pairs (2;100) versus (1;20), (1;50), and (2;0). For instance, pairs (2;100) and (1;100) with 51.8% have a higher probability to observe this or an even rarer sample, so that the minimal lot size might have a higher influence on the solution process than the TBO. The same is valid for pairs (3;0) versus (2;100), (2;20), and (2;50) where pairs (3;0) and (2;0) have a probability of 12.98% which is rather low. Additionally, this pattern is repeated for pairs (3;20) and (3;50) versus (3;0) as well as (3;100) clearly demonstrating low p-values. Where p-values for 94% return rates do not show a clear picture, Table 14.5 denoting results for 6% return rates confirms the expectance that higher and lower TBO-values show significant differences in their medians of distributions. For instance, $TBO = 3$ and $TBO = 4$ demonstrate low p-values in comparison to $TBO = 1$ including all minimal lot sizes, so that the latter do not influence the solution process. The same is valid regarding $TBO = 2$ versus $TBO = 4$.

Tables 13.8 and 13.9 present the p-values for pairs of loads and minimal lot sizes regarding their influence on the solution process. It shows that significant differences in the medians exist for pairs (0.75;50) and (0.75;0) as well as (0.75;50) and (0.75;100), so that it seems that these differences derive rather from the minimal lot size constraints clearly demonstrating differences for $x_i^{\min} = 0$, $x_i^{\min} = 50$, and $x_i^{\min} = 100$. This is also true regarding the p-values for loads of 0.9 where significant differenes exist for these pairs regarding $x_i^{\min} = 0$-$x_i^{\min} = 20$, $x_i^{\min} = 0$-$x_i^{\min} = 50$, and $x_i^{\min} = 20$-$x_i^{\min} = 100$, and $x_i^{\min} = 50$-$x_i^{\min} = 100$. Nevertheless, significant differences are also shown by pairs (0.8;20) and (0.75;20), so that these differenes derive from the different loads.

Part III

Conclusions

Chapter 11

Conclusions and Further Research Directions

In this work, we developed optimization models for aggregate production planning based on classical lot sizing models that take into account the problem of lead times that are dependent on the workload of production systems. This phenomenon is prevailing in practice and should be regarded in planning. Classical methods like mrp, MRP II that are implemented in ERP systems do not provide related tools or planning methods. An increase of workload in the system leads to increases of lead times, so that items have to wait at various steps of the production system. There is empirical evidence that lead times increase exponentially with the workload long before 100% resource utilization is reached, so that planned and realized lead times vary significantly. Additionally empirical evidence shows that flustering due to the workload of a production resource takes place, so that outputs are not produced at a specific constant production rate, but at a fraction of the original rate, so that planned outputs are not realized. There are few model approaches accounting for the link between order releases, planning and capacity decisions to lead times taking into account the system workload as well as lot sizing and sequencing decisions. We studied different model formulations that directly approach the problem with lead times that are load dependent using diverse methods. These have their advantages and disadvantages. Therefore, the choice was made to pursuit with the proposed approach that account for lot sizing decisions in order to directly capture the planning circularity regarding lot sizes, setups, and capacity utilization. We developed different formulations for production smoothing for the tactical planning level. Up to date, production smoothing is rather accounted for the operational planning level regarding scheduling decisions mainly in the environment of synchronized assembly line systems. No model approaches have been proposed for the tactical planning level regarding lot sizing, so far. Our production smoothing approaches distribute production jobs evenly or in a way that a certain target utilization sector is achieved, so that LDLT are avoided.

Besides, a major objective of this work is to regard effects of LDLT on the quality of items. If items have a certain lifetime, i.e., time interval in which they are useful for further employment in production or consumption, lead times that are load-dependent might lead to enhanced quality losses of waiting items. In the worst case, they cannot be used for demand satisfaction and need to be discarded or reworked if technically possible. The consideration of limited lifetimes is a hard constraint on inventory management and its integration into mathematical models for lot sizing, production, and inventory management an important issue. One may argue that longer storage times are avoided by mathematical models by accounting for inventory holding costs in the objective function. This is true, but there is a difference between being able to store products for long time periods and not being able

due to lifetime constraints. In this regard, we further regarded aspects of sustainability such as avoidance of disposals caused by limited lifetimes of items, rework and remanufacturing issues. We proposed related models further including features, e.g., overtime, minimal lot size constraints, and setup carry overs in order to minimize their effects on resource utilization thus already taking into account some scheduling at this stage. Two models are developed based on the CLSP-P that include possibilities of remanufacturing of externally returned items and rework of internally passed items that is waiting dependent. An extensive numerical study on this model is executed analyzing functionality and characteristics regarding complexity and its implications on the solution finding process. The influence of constraints, e.g., minimal lot size constraints, lifetimes, workloads, etc. are studied. Two different returns rates of items were assumed that significantly influenced results, so that return rates turn out to be a very important input factor to the model which confirms conclusions from other authors that the return or recycling rates significantly influences the planning process. Besides, minimal lot size constraints are a complicating factor of planning, not only regarding lifetime constraints, but the overall planning process. Therefore, more research is needed on this influencing factor.

Further research directions that are pursuited are the development of network flow formulations for proposed models and their comparison regarding complexity and efficiency of solution finding. Besides, we are interested in integrating FGI distribution decisions to the customer accounting for aspects of packaging and decision making on transportation modes in order to maintain quality while minimizing negative influences on the environment. This includes the reformulation of different objectives in the objective function of mathematical models that are dominated by profit maximization or cost minimization up to date. Additionally, stochastic influences are prevalent regarding LDLT, so that another research direction is to study the influence of stochastic events on the model behavior and the resulting plan using szenario techniques and deterministic/stochastic decomposition approaches. With the introduction of the GSCM and related research, a change is perceptible in the research literature. We pursuit this important topic by enhancing consciousness and sensitivity regarding sustainability in the production (planning) process.

Part IV

Appendix

Chapter 12

Data for Testing Behavior and Mechanisms of Models: Results

The example to show the load effects is taken from Table 8.1 on Page 61 with the following demand for $i = 10$ products.

Table 12.1: Numerical example to show load effects: demand data

Data and performance indicators	Time periods 1	2	3	4	5	6	7	8	9	10
Demand d_{At}	79	74	72	48	16	72	19	73	73	52
Demand d_{Bt}	47	7	85	72	85	55	32	14	65	70
Demand d_{Ct}	11	58	85	12	7	42	50	33	80	47
Demand d_{Dt}	9	60	61	90	30	59	73	69	72	24
Demand d_{Et}	81	85	39	54	20	70	96	1	83	21
Demand d_{Ft}	64	74	56	29	66	87	65	60	90	76
Demand d_{Gt}	99	3	82	31	53	66	33	53	89	20
Demand d_{Ht}	61	33	33	80	80	54	70	60	99	67
Demand d_{It}	50	95	82	99	64	93	49	30	30	11
Demand d_{Jt}	0	0	13	25	42	87	15	54	41	30

Chapter 13

Results for Return Rates 94%

Table 13.1: W-SRT on pairwise lifetime and minimal lot sizes and their influence on the optimality gap for test instances with return rates of 94% for *p*-values until pair (10;50)

	(0;0)	(0;100)	(0;20)	(0;50)	(1;0)	(1;100)	(1;20)	(1;50)	(10;0)	(10;100)	(10;20)	(10;50)
(0;0)	0.00000	0.00000	0.00122	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;100)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;20)	0.00122	0.00000	0.00000	0.00027	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;50)	0.00000	0.00000	0.00027	0.00000	0.00011	0.00008	0.00001	0.00001	0.00262	0.00000	0.00000	0.00001
(1;0)	0.00000	0.00000	0.00000	0.00011	0.00000	0.74699	0.74440	0.19511	0.96324	0.00239	0.00171	0.11671
(1;100)	0.00000	0.00000	0.00000	0.00008	0.74699	0.00000	0.91842	0.39244	0.89273	0.00515	0.00634	0.09137
(1;20)	0.00000	0.00000	0.00000	0.00001	0.74440	0.91842	0.00000	0.38022	0.94149	0.01857	0.00747	0.16883
(1;50)	0.00000	0.00000	0.00000	0.00001	0.19511	0.39244	0.38022	0.00000	0.47445	0.09468	0.05186	0.29067
(10;0)	0.00000	0.00000	0.00000	0.00262	0.96324	0.89273	0.94149	0.47445	0.00000	0.02227	0.00328	0.08499
(10;100)	0.00000	0.00000	0.00000	0.00000	0.00239	0.00515	0.01857	0.09468	0.02227	0.00000	0.53437	0.50667
(10;20)	0.00000	0.00000	0.00000	0.00000	0.00171	0.00634	0.00747	0.05186	0.00328	0.53437	0.00000	0.09951
(10;50)	0.00000	0.00000	0.00000	0.00001	0.11671	0.09137	0.16883	0.29067	0.08499	0.50667	0.09951	0.00000
(2;0)	0.00000	0.00000	0.00000	0.81641	0.00002	0.00040	0.00001	0.00003	0.00306	0.00000	0.00000	0.00000
(2;100)	0.00000	0.00000	0.00004	0.81109	0.00078	0.00231	0.00020	0.00008	0.03161	0.00001	0.00000	0.00006
(2;20)	0.00000	0.00000	0.00000	0.28911	0.00475	0.00711	0.00118	0.00046	0.01647	0.00000	0.00000	0.00067
(2;50)	0.00000	0.00000	0.00000	0.40371	0.00861	0.00951	0.00778	0.00028	0.07064	0.00003	0.00000	0.00088
(3;0)	0.00000	0.00000	0.00000	0.74695	0.00001	0.00129	0.00010	0.00003	0.00105	0.00000	0.00000	0.00000
(3;100)	0.00000	0.00000	0.00000	0.48933	0.00063	0.00685	0.00015	0.00014	0.01050	0.00000	0.00000	0.00000
(3;20)	0.00000	0.00000	0.00000	0.06132	0.01472	0.06984	0.01158	0.00126	0.06549	0.00001	0.00000	0.00015
(3;50)	0.00000	0.00000	0.00000	0.24365	0.01175	0.04637	0.00670	0.00434	0.06600	0.00002	0.00000	0.00005
(5;0)	0.00000	0.00000	0.00000	0.05698	0.13798	0.47237	0.11911	0.03871	0.52208	0.00109	0.00006	0.00385
(5;100)	0.00000	0.00000	0.00000	0.10263	0.01276	0.06426	0.01444	0.00434	0.04883	0.00000	0.00000	0.00016
(5;20)	0.00000	0.00000	0.00000	0.39713	0.00091	0.01733	0.00131	0.00031	0.01486	0.00000	0.00000	0.00000
(5;50)	0.00000	0.00000	0.00000	0.15703	0.00836	0.15212	0.03030	0.01130	0.14257	0.00016	0.00000	0.00038

Table 13.2: W-SRT on pairwise lifetime and minimal lot sizes and their influence on the optimality gap for test instances with return rates of 94% for p-values from pair (10;50)

	(2;0)	(2;100)	(2;20)	(2;50)	(3;0)	(3;100)	(3;20)	(3;50)	(5;0)	(5;100)	(5;20)	(5;50)
(0;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;100)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;20)	0.00000	0.00004	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0;50)	0.81641	0.81109	0.28911	0.40371	0.74695	0.48933	0.06132	0.24365	0.05698	0.10263	0.39713	0.15703
(1;0)	0.00002	0.00078	0.00475	0.00861	0.00001	0.00063	0.01472	0.01175	0.13798	0.01276	0.00091	0.00836
(1;100)	0.00040	0.00231	0.00711	0.00951	0.00129	0.00685	0.06984	0.04637	0.47237	0.06426	0.01733	0.15212
(1;20)	0.00001	0.00020	0.00118	0.00778	0.00010	0.00015	0.01158	0.00670	0.11911	0.01444	0.00131	0.03030
(1;50)	0.00003	0.00008	0.00046	0.00028	0.00003	0.00014	0.00126	0.00434	0.03871	0.00434	0.00031	0.01130
(10;0)	0.00306	0.03161	0.01647	0.07064	0.00105	0.01050	0.06549	0.06600	0.52208	0.04883	0.01486	0.14257
(10;100)	0.00000	0.00001	0.00000	0.00003	0.00000	0.00000	0.00001	0.00002	0.00109	0.00000	0.00000	0.00016
(10;20)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00006	0.00000	0.00000	0.00000
(10;50)	0.00000	0.00006	0.00067	0.00088	0.00000	0.00000	0.00015	0.00005	0.00385	0.00016	0.00000	0.00038
(2;0)	0.00000	0.80318	0.38213	0.30251	0.99319	0.66585	0.06305	0.25634	0.02393	0.13351	0.33226	0.17696
(2;100)	0.80318	0.00000	0.31058	0.26415	0.76641	0.76641	0.44949	0.49795	0.23948	0.39429	0.96732	0.27157
(2;20)	0.38213	0.31058	0.00000	0.78607	0.26423	0.64611	0.63025	0.87926	0.25207	0.71363	0.52656	0.82307
(2;50)	0.30251	0.26415	0.78607	0.00000	0.18075	0.31055	0.90624	0.96052	0.57438	0.73408	0.69585	0.94692
(3;0)	0.99319	0.76641	0.26423	0.18075	0.00000	0.71488	0.10896	0.28296	0.03455	0.09071	0.24432	0.13798
(3;100)	0.66585	0.76641	0.64611	0.31055	0.71488	0.00000	0.24642	0.24641	0.05064	0.36470	0.70093	0.12914
(3;20)	0.06305	0.44949	0.63025	0.90624	0.10896	0.24642	0.00000	0.86581	0.42336	0.93334	0.80716	0.56282
(3;50)	0.25634	0.49795	0.87926	0.96052	0.28296	0.24641	0.86581	0.00000	0.42830	0.72383	0.80188	0.54226
(5;0)	0.02393	0.23948	0.25207	0.57438	0.03455	0.05064	0.42336	0.42830	0.00000	0.40773	0.26939	0.82838
(5;100)	0.13351	0.39429	0.71363	0.73408	0.09071	0.36470	0.93334	0.72383	0.40773	0.00000	0.72511	0.89677
(5;20)	0.33226	0.96732	0.52656	0.69585	0.24432	0.70093	0.80716	0.80188	0.26939	0.72511	0.00000	0.47134
(5;50)	0.17696	0.27157	0.82307	0.94692	0.13798	0.12914	0.56282	0.54226	0.82838	0.89677	0.47134	0.00000

Table 13.3: Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 0$ and $x_i^{\min} = 20$

			Lifetimes					
$x_i^{\min}$	Load	TBO	$\Theta_i^L = 0$	$\Theta_i^L = 1$	$\Theta_i^L = 2$	$\Theta_i^L = 3$	$\Theta_i^L = 5$	$\Theta_i^L = 10$
0	0.75	1	5.21	7.88	7.91	8.18	25.34	8.53
		2	5.94	8.65	7.54	7.22	9.21	8.56
		3	5.72	8.46	7.69	7.36	6.55	6.59
		4	4.95	8.52	7.33	6.27	7.97	7.33
	0.8	1	4.27	8.10	6.99	7.66	7.70	7.24
		2	6.89	7.37	6.60	6.57	6.10	5.30
		3	3.68	7.67	5.46	5.82	5.82	5.25
		4	4.30	6.70	5.80	7.31	6.54	6.40
	0.9	1	1.45	5.30	4.51	3.91	5.26	4.41
		2	2.31	7.15	4.63	4.65	4.69	4.75
		3	1.63	6.24	4.02	4.72	5.42	5.32
		4	2.71	7.05	5.40	4.86	6.07	4.83
	1.0	1	0.85	4.45	3.46	2.62	2.58	2.52
		2	1.01	4.72	3.79	4.34	2.96	25.12
		3	0.71	4.81	4.26	4.01	4.49	24.98
		4	0.68	4.48	4.98	4.46	28.42	27.11
20	0.75	1	4.62	8.12	8.78	9.42	8.97	8.56
		2	4.43	9.62	8.33	8.30	7.73	8.08
		3	3.98	10.13	7.62	6.03	6.75	34.88
		4	3.84	8.48	7.93	7.41	6.47	7.08
	0.8	1	6.75	7.16	6.84	8.11	6.71	8.40
		2	6.09	8.23	7.10	6.14	6.24	7.76
		3	4.63	8.09	7.59	6.38	7.22	6.81
		4	6.52	7.94	7.27	8.01	6.47	6.47
	0.9	1	4.36	6.41	3.95	4.74	4.56	4.98
		2	4.14	5.67	4.65	5.60	4.88	41.06
		3	3.92	6.61	5.03	6.06	5.00	41.38
		4	4.50	5.97	5.24	5.62	5.79	5.36
	1.0	1	2.56	4.28	4.39	2.95	3.31	3.25
		2	4.02	3.95	3.37	3.52	4.28	19.23
		3	2.78	5.27	3.57	4.25	4.98	20.80
		4	2.36	4.21	3.58	4.51	4.85	5.32

Table 13.4: Average gap in % regarding lifetimes of test instances with return rates of 94% and minimal lot sizes $x_i^{\min} = 50$ and $x_i^{\min} = 100$

			Lifetimes					
$x_i^{\min}$	Load	TBO	$\Theta_i^L = 0$	$\Theta_i^L = 1$	$\Theta_i^L = 2$	$\Theta_i^L = 3$	$\Theta_i^L = 5$	$\Theta_i^L = 10$
50	0.75	1	7.88	8.23	7.61	9.36	9.51	9.60
		2	7.80	17.55	8.68	9.03	7.62	9.62
		3	7.29	9.27	6.62	6.55	6.51	7.59
		4	6.52	9.12	8.28	7.58	6.95	7.75
	0.8	1	7.80	8.70	7.07	7.05	7.61	7.35
		2	7.19	6.92	7.80	5.27	5.84	6.74
		3	8.01	8.50	7.57	6.27	6.79	6.79
		4	5.35	7.64	7.52	5.92	6.52	6.58
	0.9	1	4.56	6.68	5.38	4.66	4.61	5.71
		2	3.70	5.51	6.43	5.94	4.00	4.26
		3	5.52	6.66	4.01	5.97	6.28	6.99
		4	3.48	9.53	4.70	9.62	4.95	35.99
	1.0	1	3.42	3.83	4.35	3.36	2.94	2.45
		2	4.53	3.67	3.06	3.73	29.25	20.47
		3	2.44	4.60	3.12	4.19	4.33	5.23
		4	3.65	4.08	3.29	3.40	4.71	17.72
100	0.75	1	22.81	9.64	7.77	11.25	9.91	10.78
		2	17.86	10.27	8.88	8.04	7.72	9.25
		3	20.89	8.75	7.51	7.34	7.76	8.19
		4	41.24	10.03	7.21	7.52	8.39	7.48
	0.8	1	35.28	8.09	7.51	7.48	7.21	10.01
		2	16.86	7.15	5.65	5.81	6.82	6.10
		3	17.35	8.11	7.66	5.65	6.24	8.11
		4	30.39	8.60	6.77	6.22	6.59	6.93
	0.9	1	15.39	5.98	6.47	5.21	4.62	5.44
		2	13.45	5.54	3.91	4.64	4.75	5.09
		3	13.64	4.98	4.00	5.23	5.33	28.74
		4	13.39	6.03	5.96	5.73	5.58	6.22
	1.0	1	10.03	3.48	3.94	2.11	2.44	2.33
		2	10.03	3.80	3.03	3.77	4.59	24.69
		3	9.42	4.19	3.36	3.49	4.76	22.00
		4	13.37	4.48	3.06	4.09	4.50	5.53

Table 13.5: W-SRT on pairwise TBO and lifetime and their influence on the optimality gap for test instances with return rates of 94% until (2;5)

	(1;0)	(1;1)	(1;10)	(1;2)	(1;3)	(1;5)	(2;0)	(2;1)	(2;10)	(2;2)	(2;3)	(2;5)
(1;0)	0.00000	0.09604	0.55136	0.77299	0.88464	0.80846	0.89812	0.06066	0.00001	0.94556	0.76256	0.42538
(1;1)	0.09604	0.00000	0.74957	0.02414	0.16942	0.21336	0.19512	0.76643	0.00013	0.01181	0.00441	0.03745
(1;10)	0.55136	0.74957	0.00000	0.27538	0.49152	0.72383	0.87925	0.29068	0.00004	0.29853	0.18360	0.58961
(1;2)	0.77299	0.02414	0.27538	0.00000	0.79134	0.88060	0.95508	0.04415	0.00000	0.35842	0.39435	0.88464
(1;3)	0.88464	0.16942	0.49152	0.79134	0.00000	0.94420	0.69336	0.07843	0.00000	0.51988	0.59080	0.84438
(1;5)	0.80846	0.21336	0.72383	0.88060	0.94420	0.00000	0.95236	0.07756	0.00000	0.65101	0.61218	0.72384
(2;0)	0.89812	0.19512	0.87925	0.95508	0.69336	0.95236	0.00000	0.26794	0.00000	0.40865	0.46084	0.76255
(2;1)	0.06066	0.76643	0.29068	0.04415	0.07843	0.07756	0.26794	0.00000	0.00040	0.00971	0.00545	0.12915
(2;10)	0.00001	0.00013	0.00004	0.00000	0.00000	0.00000	0.00000	0.00040	0.00000	0.00000	0.00000	0.00000
(2;2)	0.94556	0.01181	0.29853	0.35842	0.51988	0.65101	0.40865	0.00971	0.00000	0.00000	0.96188	0.42338
(2;3)	0.76256	0.00441	0.18360	0.39435	0.59080	0.61218	0.46084	0.00545	0.00000	0.96188	0.00000	0.27840
(2;5)	0.42538	0.03745	0.58961	0.88464	0.84438	0.72384	0.76255	0.12915	0.00000	0.42338	0.27840	0.00000
(3;0)	0.46401	0.02330	0.33142	0.25776	0.54455	0.32713	0.24368	0.01472	0.00000	0.62058	0.32635	0.05974
(3;1)	0.00819	0.18077	0.21336	0.00031	0.01000	0.02902	0.02007	0.47026	0.00192	0.00023	0.00002	0.00092
(3;10)	0.00000	0.00009	0.00004	0.00000	0.00000	0.00000	0.00000	0.00032	0.28294	0.00000	0.00000	0.00000
(3;2)	0.21083	0.00003	0.07472	0.08013	0.26716	0.18814	0.12405	0.00353	0.00000	0.55590	0.30816	0.04962
(3;3)	0.70853	0.00003	0.10523	0.02227	0.25347	0.23279	0.35752	0.00190	0.00000	0.43829	0.91977	0.11279
(3;5)	0.39907	0.00299	0.18138	0.34000	0.59790	0.73153	0.91166	0.00981	0.00000	0.84705	0.24294	0.96052
(4;0)	0.47239	0.00759	0.32719	0.03109	0.45367	0.30819	0.17696	0.00316	0.00000	0.33141	0.13754	0.03444
(4;1)	0.04884	0.11279	0.10522	0.00029	0.01884	0.05723	0.03531	0.34869	0.00098	0.00028	0.00003	0.00499
(4;10)	0.00119	0.47656	0.06676	0.00952	0.01306	0.00924	0.00222	0.24093	0.04964	0.00098	0.00000	0.00063
(4;2)	0.56281	0.00514	0.26352	0.69587	0.74052	0.93334	0.75994	0.01579	0.00000	0.82170	0.56281	0.85240
(4;3)	0.46397	0.02900	0.78606	0.75994	0.87252	0.82571	0.75994	0.24987	0.00000	0.41352	0.13795	0.81244
(4;5)	0.05248	0.26351	0.96052	0.38574	0.42535	0.41452	0.20347	0.30492	0.00003	0.11276	0.00852	0.12487

Table 13.6: W-SRT on pairwise TBO and lifetime and their influence on the optimality gap for test instances with return rates of 94% from (2;4)

	(3;0)	(3;1)	(3;10)	(3;2)	(3;3)	(3;5)	(4;0)	(4;1)	(4;10)	(4;2)	(4;3)	(4;5)
(1;0)	0.46401	0.00819	0.00000	0.21083	0.70853	0.39907	0.47239	0.04884	0.00119	0.56281	0.46397	0.05248
(1;1)	0.02330	0.18077	0.00009	0.00003	0.00003	0.00299	0.00759	0.11279	0.47656	0.00514	0.02900	0.26351
(1;10)	0.33142	0.21336	0.00004	0.07472	0.10523	0.18138	0.32719	0.10522	0.06676	0.26352	0.78606	0.96052
(1;2)	0.25776	0.00031	0.00000	0.08013	0.02227	0.34000	0.03109	0.00029	0.00952	0.69587	0.75994	0.38574
(1;3)	0.54455	0.01000	0.00000	0.26716	0.25347	0.59790	0.45367	0.01884	0.01306	0.74052	0.87252	0.42535
(1;5)	0.32713	0.02902	0.00000	0.18814	0.23279	0.73153	0.30819	0.05723	0.00924	0.93334	0.82571	0.41452
(2;0)	0.24368	0.02007	0.00000	0.12405	0.35752	0.91166	0.17696	0.03531	0.00222	0.75994	0.75994	0.20347
(2;1)	0.01472	0.47026	0.00032	0.00353	0.00190	0.00981	0.00316	0.34869	0.24093	0.01579	0.24987	0.30492
(2;10)	0.00000	0.00192	0.28294	0.00000	0.00000	0.00000	0.00000	0.00098	0.04964	0.00000	0.00000	0.00003
(2;2)	0.62058	0.00023	0.00000	0.55590	0.43829	0.84705	0.33141	0.00028	0.00098	0.82170	0.41352	0.11276
(2;3)	0.32635	0.00002	0.00000	0.30816	0.91977	0.24294	0.13754	0.00003	0.00000	0.56281	0.13795	0.00852
(2;5)	0.05974	0.00092	0.00000	0.04962	0.11279	0.96052	0.03444	0.00499	0.00063	0.85240	0.81244	0.12487
(3;0)	0.00000	0.00237	0.00000	0.75994	0.30249	0.11673	0.95780	0.00237	0.00013	0.28603	0.06930	0.00836
(3;1)	0.00237	0.00000	0.00069	0.00000	0.00000	0.00007	0.00047	0.89273	0.84705	0.00015	0.00062	0.01789
(3;10)	0.00000	0.00069	0.00000	0.00000	0.00000	0.00000	0.00000	0.00262	0.00349	0.00000	0.00000	0.00000
(3;2)	0.75994	0.00000	0.00000	0.00000	0.92655	0.24923	0.44953	0.00000	0.00034	0.53997	0.07897	0.00818
(3;3)	0.30249	0.00000	0.00000	0.92655	0.00000	0.07114	0.14727	0.00000	0.00000	0.17802	0.03744	0.00034
(3;5)	0.11673	0.00007	0.00000	0.24923	0.07114	0.00000	0.00767	0.00019	0.00000	0.46399	0.44543	0.08908
(4;0)	0.95780	0.00047	0.00000	0.44953	0.14727	0.00767	0.00000	0.00083	0.00000	0.07065	0.02227	0.00062
(4;1)	0.00237	0.89273	0.00262	0.00000	0.00000	0.00019	0.00083	0.00000	0.97413	0.00009	0.00213	0.03162
(4;10)	0.00013	0.84705	0.00349	0.00034	0.00000	0.00000	0.00000	0.97413	0.00000	0.00119	0.00353	0.00794
(4;2)	0.28603	0.00015	0.00000	0.53997	0.17802	0.46399	0.07065	0.00009	0.00119	0.00000	0.38953	0.28289
(4;3)	0.06930	0.00062	0.00000	0.07897	0.03744	0.44543	0.02227	0.00213	0.00353	0.38953	0.00000	0.28291
(4;5)	0.00836	0.01789	0.00000	0.00818	0.00034	0.08908	0.00062	0.03162	0.00794	0.28289	0.28291	0.00000

Table 13.7: W-SRT on pairwise TBO and minimal lot size and their influence on the optimality gap for test instances with return rates of 94%

	(1;0)	(1;100)	(1;20)	(1;50)	(2;0)	(2;100)	(2;20)	(2;50)	(3;0)	(3;100)	(3;20)	(3;50)	(4;0)	(4;100)	(4;20)	(4;50)
(1;0)	0.00000	0.00004	0.15696	0.03090	0.25579	0.00007	0.00606	0.00403	0.83571	0.00000	0.02962	0.15911	0.06848	0.00001	0.20066	0.00430
(1;100)	0.00004	0.00000	0.00025	0.00360	0.00030	0.51808	0.04418	0.12288	0.00001	0.73849	0.05377	0.00112	0.00216	0.68650	0.00009	0.10188
(1;20)	0.15696	0.00025	0.00000	0.32382	0.71962	0.00107	0.24243	0.04408	0.30240	0.00005	0.14764	0.67491	0.34858	0.00051	0.92960	0.03004
(1;50)	0.03090	0.00360	0.32382	0.00000	0.25464	0.02425	0.54925	0.29372	0.02281	0.00143	0.80895	0.22348	0.48728	0.03734	0.16334	0.28405
(2;0)	0.25579	0.00030	0.71962	0.25464	0.00000	0.00563	0.28570	0.04994	0.12980	0.00004	0.21033	0.50490	0.49901	0.00108	0.89862	0.03959
(2;100)	0.00007	0.51808	0.00107	0.02425	0.00563	0.00000	0.19295	0.43910	0.00064	0.27656	0.30855	0.04253	0.08419	0.44076	0.00551	0.26927
(2;20)	0.00606	0.04418	0.24243	0.54925	0.28570	0.19295	0.00000	0.70713	0.01781	0.01715	0.93256	0.70782	0.79672	0.02581	0.23102	0.63663
(2;50)	0.00403	0.12288	0.04408	0.29372	0.04994	0.43910	0.70713	0.00000	0.00162	0.06081	0.60317	0.22701	0.24280	0.14581	0.05105	0.93330
(3;0)	0.83571	0.00001	0.30240	0.02281	0.12980	0.00064	0.01781	0.00162	0.00000	0.00000	0.01423	0.00209	0.02125	0.00000	0.07314	0.00176
(3;100)	0.00000	0.73849	0.00005	0.00143	0.00004	0.27656	0.01715	0.06081	0.00000	0.00000	0.01217	0.00035	0.00077	0.55049	0.00004	0.03827
(3;20)	0.02962	0.05377	0.14764	0.80895	0.21033	0.30855	0.93256	0.60317	0.01423	0.01217	0.00000	0.81471	0.61294	0.04617	0.23356	0.60708
(3;50)	0.15911	0.00112	0.67491	0.22348	0.50490	0.04253	0.70782	0.22701	0.00209	0.00035	0.81471	0.00000	0.55359	0.01689	0.15608	0.14710
(4;0)	0.06848	0.00216	0.34858	0.48728	0.49901	0.08419	0.79672	0.24280	0.02125	0.00077	0.61294	0.55359	0.00000	0.00355	0.56612	0.13171
(4;100)	0.00001	0.68650	0.00051	0.03734	0.00108	0.44076	0.02581	0.14581	0.00000	0.55049	0.04617	0.01689	0.00355	0.00000	0.00012	0.21822
(4;20)	0.20066	0.00009	0.92960	0.16334	0.89862	0.00551	0.23102	0.05105	0.07314	0.00004	0.23356	0.15608	0.56612	0.00012	0.00000	0.03321
(4;50)	0.00430	0.10188	0.03004	0.28405	0.03959	0.26927	0.63663	0.93330	0.00176	0.03827	0.60708	0.14710	0.13171	0.21822	0.03321	0.00000

Table 13.8: W-SRT on pairwise Load and minimal lot size and their influence on the optimality gap for test instances with return rates of 94% until values (0.8;50)

	(0.75;0)	(0.75;100)	(0.75;20)	(0.75;50)	(0.8;0)	(0.8;100)	(0.8;20)	(0.8;50)
(0.75;0)	0.00000	0.00000	0.21856	0.00040	0.00000	0.81832	0.00061	0.00040
(0.75;100)	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000
(0.75;20)	0.21856	0.00000	0.00000	0.11712	0.00000	0.38355	0.00003	0.00003
(0.75;50)	0.00040	0.00001	0.11712	0.00000	0.00000	0.00147	0.00000	0.00000
(0.8;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.8;100)	0.81832	0.00000	0.38355	0.00147	0.00000	0.00000	0.02253	0.01697
(0.8;20)	0.00061	0.00000	0.00003	0.00000	0.00000	0.02253	0.00000	0.94069
(0.8;50)	0.00040	0.00000	0.00003	0.00000	0.00000	0.01697	0.94069	0.00000
(0.9;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.9;100)	0.00000	0.00000	0.00000	0.00000	0.00121	0.00000	0.00000	0.00000
(0.9;20)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.9;50)	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000	0.00000
(1.0;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(1.0;100)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(1.0;20)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(1.0;50)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table 13.9: W-SRT on pairwise Load and minimal lot size and their influence on the optimality gap for test instances with return rates of 94% from values (0.8;50)

	(0.9;0)	(0.9;100)	(0.9;20)	(0.9;50)	(1.0;0)	(1.0;100)	(1.0;20)	(1.0;50)
(0.75;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.75;100)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.75;20)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.75;50)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.8;0)	0.00000	0.00121	0.00000	0.00002	0.00000	0.00000	0.00000	0.00000
(0.8;100)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.8;20)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.8;50)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
(0.9;0)	0.00000	0.00000	0.00047	0.00007	0.00010	0.11136	0.00003	0.00001
(0.9;100)	0.00000	0.00000	0.00937	0.08723	0.00000	0.00000	0.00000	0.00000
(0.9;20)	0.00047	0.00937	0.00000	0.42062	0.00000	0.00000	0.00000	0.00000
(0.9;50)	0.00007	0.08723	0.42062	0.00000	0.00000	0.00001	0.00000	0.00000
(1.0;0)	0.00010	0.00000	0.00000	0.00000	0.00000	0.29033	0.92960	0.80462
(1.0;100)	0.11136	0.00000	0.00000	0.00001	0.29033	0.00000	0.23172	0.15995
(1.0;20)	0.00003	0.00000	0.00000	0.00000	0.92960	0.23172	0.00000	0.47626
(1.0;50)	0.00001	0.00000	0.00000	0.00000	0.80462	0.15995	0.47626	0.00000

Chapter 14

Results for Return Rates 6%

Table 14.1: W-SRT on pairwise Lifetime and minimal lot size and their influence on the optimality gap for test instances with return rates of 6%

	(0;0)	(0;100)	(0;20)	(0;50)	(1;0)	(1;100)	(1;20)	(1;50)	(10;0)	(10;100)	(10;20)	(10;50)
(0;0)	0.00000	0.04243	0.04124	0.02842	0.00000	0.00000	0.00000	0.00000	0.00007	0.00000	0.00000	0.00000
(0;100)	0.04243	0.00000	0.91766	0.78235	0.00070	0.00005	0.00001	0.00001	0.01600	0.00027	0.00321	0.00233
(0;20)	0.04124	0.91766	0.00000	0.78916	0.00116	0.00009	0.00002	0.00002	0.01843	0.00044	0.00461	0.00304
(0;50)	0.02842	0.78235	0.78916	0.00000	0.00438	0.00048	0.00011	0.00021	0.02956	0.00119	0.01367	0.00653
(1;0)	0.00000	0.00070	0.00116	0.00438	0.00000	0.39484	0.31578	0.34246	0.23906	0.96486	0.55999	0.54712
(1;100)	0.00000	0.00005	0.00009	0.00048	0.39484	0.00000	0.86430	0.97419	0.06907	0.46881	0.16359	0.20643
(1;20)	0.00000	0.00001	0.00002	0.00011	0.31578	0.86430	0.00000	0.80764	0.04134	0.31639	0.09669	0.12675
(1;50)	0.00000	0.00001	0.00002	0.00021	0.34246	0.97419	0.80764	0.00000	0.03808	0.29991	0.14321	0.14269
(10;0)	0.00007	0.01600	0.01843	0.02956	0.23906	0.06907	0.04134	0.03808	0.00000	0.27022	0.54782	0.59871
(10;100)	0.00000	0.00027	0.00044	0.00119	0.96486	0.46881	0.31639	0.29991	0.27022	0.00000	0.69181	0.66417
(10;20)	0.00000	0.00321	0.00461	0.01367	0.55999	0.16359	0.09669	0.14321	0.54782	0.69181	0.00000	0.93407
(10;50)	0.00000	0.00233	0.00304	0.00653	0.54712	0.20643	0.12675	0.14269	0.59871	0.66417	0.93407	0.00000
(2;0)	0.00000	0.00342	0.00459	0.00881	0.42967	0.13932	0.10168	0.10627	0.53354	0.61066	0.98648	0.93684
(2;100)	0.00000	0.00143	0.00192	0.00472	0.69847	0.24822	0.19371	0.23677	0.28148	0.97196	0.74696	0.57561
(2;20)	0.00005	0.01566	0.01892	0.02878	0.20889	0.04895	0.03344	0.03723	0.84806	0.38907	0.60486	0.70468
(2;50)	0.00000	0.00342	0.00460	0.00882	0.48605	0.13934	0.10957	0.13934	0.42628	0.73243	0.94894	0.80042
(3;0)	0.00000	0.00217	0.00285	0.00652	0.67709	0.24109	0.17797	0.27060	0.41761	0.87614	0.72229	0.74435
(3;100)	0.00001	0.00328	0.00418	0.00895	0.53362	0.16972	0.11618	0.16587	0.44157	0.77871	0.96837	0.78532
(3;20)	0.00007	0.01511	0.01742	0.02925	0.27880	0.06907	0.04801	0.06393	0.68874	0.49628	0.65417	0.81177
(3;50)	0.00000	0.00143	0.00192	0.00472	0.69166	0.24828	0.17393	0.14553	0.34468	0.76580	0.94472	0.79401
(5;0)	0.00007	0.01511	0.01742	0.02925	0.24360	0.06907	0.04781	0.05289	0.84280	0.34505	0.59986	0.75279
(5;100)	0.00001	0.00328	0.00418	0.00896	0.47424	0.16976	0.10400	0.09710	0.65874	0.54040	0.86223	0.90364
(5;20)	0.00007	0.01511	0.01742	0.02925	0.24283	0.06907	0.04865	0.03808	0.83053	0.30078	0.53518	0.70944
(5;50)	0.00033	0.05799	0.06702	0.09238	0.07690	0.01423	0.00768	0.00815	0.62824	0.14375	0.23007	0.29875

Table 14.2: W-SRT on pairwise Lifetime and minimal lot size and their influence on the optimality gap for test instances with return rates of 6%

	(2;0)	(2;100)	(2;20)	(2;50)	(3;0)	(3;100)	(3;20)	(3;50)	(5;0)	(5;100)	(5;20)	(5;50)
(0;0)	0.00000	0.00000	0.00005	0.00000	0.00000	0.00001	0.00007	0.00000	0.00007	0.00001	0.00007	0.00033
(0;100)	0.00342	0.00143	0.01566	0.00342	0.00217	0.00328	0.01511	0.00143	0.01511	0.00328	0.01511	0.05799
(0;20)	0.00459	0.00192	0.01892	0.00460	0.00285	0.00418	0.01742	0.00192	0.01742	0.00418	0.01742	0.06702
(0;50)	0.00881	0.00472	0.02878	0.00882	0.00652	0.00895	0.02925	0.00472	0.02925	0.00896	0.02925	0.09238
(1;0)	0.42967	0.69847	0.20889	0.48605	0.67709	0.53362	0.27880	0.69166	0.24360	0.47424	0.24283	0.07690
(1;100)	0.13932	0.24822	0.04895	0.13934	0.24109	0.16972	0.06907	0.24828	0.06907	0.16976	0.06907	0.01423
(1;20)	0.10168	0.19371	0.03344	0.10957	0.17797	0.11618	0.04801	0.17393	0.04781	0.10400	0.04865	0.00768
(1;50)	0.10627	0.23677	0.03723	0.13934	0.27060	0.16587	0.06393	0.14553	0.05289	0.09710	0.03808	0.00815
(10;0)	0.53354	0.28148	0.84806	0.42628	0.41761	0.44157	0.68874	0.34468	0.84280	0.65874	0.83053	0.62824
(10;100)	0.61066	0.97196	0.38907	0.73243	0.87614	0.77871	0.49628	0.76580	0.34505	0.54040	0.30078	0.14375
(10;20)	0.98648	0.74696	0.60486	0.94894	0.72229	0.96837	0.65417	0.94472	0.59986	0.86223	0.53518	0.23007
(10;50)	0.93684	0.57561	0.70468	0.80042	0.74435	0.78532	0.81177	0.79401	0.75279	0.90364	0.70944	0.29875
(2;0)	0.00000	0.61689	0.60055	0.90971	0.82669	0.96376	0.67528	0.79524	0.67247	0.91857	0.68372	0.21206
(2;100)	0.61689	0.00000	0.30830	0.80097	0.96254	0.76418	0.40091	0.78025	0.36947	0.49861	0.33493	0.10570
(2;20)	0.60055	0.30830	0.00000	0.53456	0.43927	0.57190	0.99376	0.53668	0.99220	0.73085	0.93145	0.44680
(2;50)	0.90971	0.80097	0.53456	0.00000	0.85176	0.89602	0.55909	1.00000	0.49022	0.72760	0.47472	0.19401
(3;0)	0.82669	0.96254	0.43927	0.85176	0.00000	0.77803	0.51622	0.93860	0.45847	0.67178	0.46314	0.17461
(3;100)	0.96376	0.76418	0.57190	0.89602	0.77803	0.00000	0.61053	0.99398	0.56925	0.71974	0.52571	0.22207
(3;20)	0.67528	0.40091	0.99376	0.55909	0.51622	0.61053	0.00000	0.69498	0.93583	0.95540	0.82441	0.44027
(3;50)	0.79524	0.78025	0.53668	1.00000	0.93860	0.99398	0.69498	0.00000	0.50688	0.63212	0.43397	0.20741
(5;0)	0.67247	0.36947	0.99220	0.49022	0.45847	0.56925	0.93583	0.50688	0.00000	0.81849	0.98590	0.55041
(5;100)	0.91857	0.49861	0.73085	0.72760	0.67178	0.71974	0.95540	0.63212	0.81849	0.00000	0.80352	0.36326
(5;20)	0.68372	0.33493	0.93145	0.47472	0.46314	0.52571	0.82441	0.43397	0.98590	0.80352	0.00000	0.61572
(5;50)	0.21206	0.10570	0.44680	0.19401	0.17461	0.22207	0.44027	0.20741	0.55041	0.36326	0.61572	0.00000

Table 14.3: W-SRT on pairwise TBO and Lifetime and their influence on the optimality gap for test instances with return rates of 6%

	(1;0)	(1;1)	(1;10)	(1;2)	(1;3)	(1;5)	(2;0)	(2;1)	(2;10)	(2;2)	(2;3)	(2;5)
(1;0)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.75328	0.00000	0.00038	0.00003	0.00005	0.00002
(1;1)	0.00000	0.00000	0.80079	0.38300	0.41961	0.12695	0.00000	0.33648	0.00075	0.00540	0.00318	0.01909
(1;10)	0.00000	0.80079	0.00000	0.98873	0.34528	0.53899	0.00000	0.19990	0.02729	0.13020	0.06603	0.11853
(1;2)	0.00000	0.38300	0.98873	0.00000	0.75372	0.26229	0.00000	0.07164	0.00677	0.03750	0.02161	0.04537
(1;3)	0.00000	0.41961	0.34528	0.75372	0.00000	0.16912	0.00000	0.30440	0.00276	0.01132	0.00724	0.01609
(1;5)	0.00000	0.12695	0.53899	0.26229	0.16912	0.00000	0.00002	0.01186	0.09047	0.25264	0.19435	0.28559
(2;0)	0.75328	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000	0.00050	0.00006	0.00009	0.00007
(2;1)	0.00000	0.33648	0.19990	0.07164	0.30440	0.01186	0.00000	0.00000	0.00020	0.00107	0.00078	0.00246
(2;10)	0.00038	0.00075	0.02729	0.00677	0.00276	0.09047	0.00050	0.00020	0.00000	0.22651	0.44426	0.73121
(2;2)	0.00003	0.00540	0.13020	0.03750	0.01132	0.25264	0.00006	0.00107	0.22651	0.00000	0.51350	0.53882
(2;3)	0.00005	0.00318	0.06603	0.02161	0.00724	0.19435	0.00009	0.00078	0.44426	0.51350	0.00000	0.76634
(2;5)	0.00002	0.01909	0.11853	0.04537	0.01609	0.28559	0.00007	0.00246	0.73121	0.53882	0.76634	0.00000
(3;0)	0.30467	0.00000	0.00000	0.00000	0.00000	0.00013	0.63003	0.00000	0.00555	0.00083	0.00123	0.00073
(3;1)	0.07303	0.00000	0.00764	0.00862	0.01005	0.09264	0.05531	0.00003	0.44933	0.18576	0.22558	0.21852
(3;10)	0.00061	0.00088	0.09766	0.09901	0.05909	0.49323	0.00076	0.00033	0.52028	0.95681	0.85290	0.95677
(3;2)	0.02610	0.00002	0.01247	0.00800	0.00697	0.08766	0.02146	0.00002	0.69549	0.24101	0.39618	0.37032
(3;3)	0.02618	0.00002	0.01615	0.02147	0.01994	0.17213	0.02153	0.00003	0.69561	0.33701	0.39637	0.37051
(3;5)	0.13575	0.00000	0.00444	0.00357	0.00614	0.05539	0.09911	0.00000	0.31430	0.11632	0.14440	0.14464
(4;0)	0.43987	0.00000	0.00000	0.00000	0.00000	0.00004	0.56200	0.00000	0.00217	0.00023	0.00037	0.00027
(4;1)	0.02620	0.00002	0.01616	0.02075	0.01994	0.16861	0.02154	0.00003	0.69564	0.33705	0.39640	0.37055
(4;2)	0.13574	0.00000	0.00444	0.00165	0.00259	0.03322	0.09910	0.00000	0.31429	0.11180	0.14440	0.14463
(4;2)	0.13568	0.00000	0.00372	0.00220	0.00302	0.03521	0.09905	0.00000	0.31424	0.08903	0.14435	0.14459
(4;3)	0.13574	0.00000	0.00444	0.00441	0.00614	0.05790	0.09910	0.00000	0.31429	0.11631	0.14440	0.14463
(4;5)	0.13574	0.00000	0.00195	0.00019	0.00048	0.00897	0.09910	0.00000	0.26807	0.04015	0.09537	0.13148

Table 14.4: W-SRT on pairwise TBO and Lifetime and their influence on the optimality gap for test instances with return rates of 6%

	(3;0)	(3;1)	(3;10)	(3;2)	(3;3)	(3;5)	(4;0)	(4;1)	(4;10)	(4;2)	(4;3)	(4;5)
(1;0)	0.30467	0.07303	0.00061	0.02610	0.02618	0.13575	0.43987	0.02620	0.13574	0.13568	0.13574	0.13574
(1;1)	0.00000	0.00000	0.00088	0.00002	0.00002	0.00000	0.00000	0.00002	0.00000	0.00000	0.00000	0.00000
(1;10)	0.00000	0.00764	0.09766	0.01247	0.01615	0.00444	0.00000	0.01616	0.00444	0.00372	0.00444	0.00195
(1;2)	0.00000	0.00862	0.09901	0.00800	0.02147	0.00357	0.00000	0.02075	0.00165	0.00220	0.00441	0.00019
(1;3)	0.00000	0.01005	0.05909	0.00697	0.01994	0.00614	0.00000	0.01994	0.00259	0.00302	0.00614	0.00048
(1;5)	0.00013	0.09264	0.49323	0.08766	0.17213	0.05539	0.00004	0.16861	0.03322	0.03521	0.05790	0.00897
(2;0)	0.63003	0.05531	0.00076	0.02146	0.02153	0.09911	0.56200	0.02154	0.09910	0.09905	0.09910	0.09910
(2;1)	0.00000	0.00003	0.00033	0.00002	0.00003	0.00000	0.00000	0.00003	0.00000	0.00000	0.00000	0.00000
(2;10)	0.00555	0.44933	0.52028	0.69549	0.69561	0.31430	0.00217	0.69564	0.31429	0.31424	0.31429	0.26807
(2;2)	0.00083	0.18576	0.95681	0.24101	0.33701	0.11632	0.00023	0.33705	0.11180	0.08903	0.11631	0.04015
(2;3)	0.00123	0.22558	0.85290	0.39618	0.39637	0.14440	0.00037	0.39640	0.14440	0.14435	0.14440	0.09537
(2;5)	0.00073	0.21852	0.95677	0.37032	0.37051	0.14464	0.00027	0.37055	0.14463	0.14459	0.14463	0.13148
(3;0)	0.00000	0.24298	0.00791	0.11883	0.11901	0.37041	0.97468	0.11905	0.37040	0.37034	0.37040	0.37040
(3;1)	0.24298	0.00000	0.52220	0.64197	0.75973	0.25211	0.18234	0.98475	0.25210	0.41436	0.32676	0.25210
(3;10)	0.00791	0.52220	0.00000	0.41304	0.61304	0.14211	0.00329	0.53516	0.12879	0.21760	0.23636	0.03344
(3;2)	0.11883	0.64197	0.41304	0.00000	0.86114	0.36075	0.07895	0.71737	0.36074	0.76442	0.55477	0.23706
(3;3)	0.11901	0.75973	0.61304	0.86114	0.00000	0.27779	0.07910	0.87775	0.29730	0.62935	0.49206	0.14676
(3;5)	0.37041	0.25211	0.14211	0.36075	0.27779	0.00000	0.29837	0.26388	1.00000	0.63089	0.69609	0.46024
(4;0)	0.97468	0.18234	0.00329	0.07895	0.07910	0.29837	0.00000	0.07913	0.29836	0.29829	0.29836	0.29836
(4;1)	0.11905	0.98475	0.53516	0.71737	0.87775	0.26388	0.07913	0.00000	0.25931	0.45867	0.39625	0.16244
(4;10)	0.37040	0.25210	0.12879	0.36074	0.29730	1.00000	0.29836	0.25931	0.00000	0.63089	0.78073	0.50214
(4;2)	0.37034	0.41436	0.21760	0.76442	0.62935	0.63089	0.29829	0.45867	0.63089	0.00000	0.79789	0.48789
(4;3)	0.37040	0.32676	0.23636	0.55477	0.49206	0.69609	0.29836	0.39625	0.78073	0.79789	0.00000	0.37048
(4;5)	0.37040	0.25210	0.03344	0.23706	0.14676	0.46024	0.29836	0.16244	0.50214	0.48789	0.37048	0.00000

Table 14.5: W-SRT on pairwise TBO and minimal lot size and their influence on the optimality gap for test instances with return rates of 6%

	(1;0)	(1;100)	(1;20)	(1;50)	(2;0)	(2;100)	(2;20)	(2;50)	(3;0)	(3;100)	(3;20)	(3;50)	(4;0)	(4;100)	(4;20)	(4;50)
(1;0)	0.00000	0.27075	1.00000	0.60925	0.06378	0.25047	0.30312	0.43504	0.00100	0.04503	0.00532	0.00957	0.00025	0.00103	0.00224	0.00097
(1;100)	0.27075	0.00000	0.35485	0.53093	0.01307	0.06197	0.09313	0.15487	0.00023	0.00664	0.00082	0.00138	0.00003	0.00025	0.00030	0.00023
(1;20)	1.00000	0.35485	0.00000	0.69361	0.13173	0.40202	0.50108	0.62004	0.00798	0.10748	0.02171	0.03217	0.00205	0.00918	0.01162	0.00800
(1;50)	0.60925	0.53093	0.69361	0.00000	0.03106	0.13682	0.17808	0.27886	0.00048	0.01652	0.00209	0.00367	0.00008	0.00050	0.00076	0.00044
(2;0)	0.06378	0.01307	0.13173	0.03106	0.00000	0.51299	0.34833	0.35823	0.11183	0.82537	0.23330	0.36382	0.06439	0.11564	0.21048	0.09760
(2;100)	0.25047	0.06197	0.40202	0.13682	0.51299	0.00000	0.79390	0.77262	0.01781	0.32593	0.05314	0.10134	0.00922	0.01943	0.04309	0.01771
(2;20)	0.30312	0.09313	0.50108	0.17808	0.34833	0.79390	0.00000	0.94815	0.01403	0.28412	0.04460	0.08318	0.00714	0.01603	0.03681	0.01043
(2;50)	0.43504	0.15487	0.62004	0.27886	0.35823	0.77262	0.94815	0.00000	0.01968	0.25356	0.05064	0.08431	0.00694	0.02144	0.04075	0.01778
(3;0)	0.00100	0.00023	0.00798	0.00048	0.11183	0.01781	0.01403	0.01968	0.00000	0.11314	0.69409	0.51472	0.81602	1.00000	0.61850	0.92788
(3;100)	0.04503	0.00664	0.10748	0.01652	0.82537	0.32593	0.28412	0.25356	0.11314	0.00000	0.29950	0.39406	0.04725	0.12374	0.24920	0.11060
(3;20)	0.00532	0.00082	0.02171	0.00209	0.23330	0.05314	0.04460	0.05064	0.69409	0.29950	0.00000	0.89997	0.45324	0.61214	0.92208	0.55486
(3;50)	0.00957	0.00138	0.03217	0.00367	0.36382	0.10134	0.08318	0.08431	0.51472	0.39406	0.89997	0.00000	0.32704	0.51472	0.79271	0.47755
(4;0)	0.00025	0.00003	0.00205	0.00008	0.06439	0.00922	0.00714	0.00694	0.81602	0.04725	0.45324	0.32704	0.00000	0.86417	0.40689	0.86417
(4;100)	0.00103	0.00025	0.00918	0.00050	0.11564	0.01943	0.01603	0.02144	1.00000	0.12374	0.61214	0.51472	0.86417	0.00000	0.64810	0.87944
(4;20)	0.00224	0.00030	0.01162	0.00076	0.21048	0.04309	0.03681	0.04075	0.61850	0.24920	0.92208	0.79271	0.40689	0.64810	0.00000	0.60188
(4;50)	0.00097	0.00023	0.00800	0.00044	0.09760	0.01771	0.01043	0.01778	0.92788	0.11060	0.55486	0.47755	0.86417	0.87944	0.60188	0.00000

Bibliography

T. Abdallah, A. Diabat, and D. Simchi-Levi. Sustainable supply chain design: A closed-loop formulation and sensitivity analysis. *Production Planning & Control*, 23(2–3):120–133, 21012.

D. Adam. *Produktions-Management*. Gabler, 9 edition, 1998.

R. K. Ahuja, W. Huang, H. E. Romeijn, and D. Romero-Morales. A heuristic approach to the multi-period single-sourcing problem with production and inventory capacities and perishability constraints. *INFORMS Journal on Computing*, 19(1): 14–26, 2007.

R. Akkerman, Y. Wang, and M. Grunow. MILP approaches to sustainable production and distribution of meal elements. *Computers and Industrial Engineering (IEEE)*, 1:973–987, 2009. IEEE International Conference on Computers and Industrial Engineering (CIE 2009).

D.S. Allen. Seasonal production smoothing. *Federal Reserve Bank of St. Louis Review*, 81(5):21–40, 1999.

L. C. Angell and R. D. Klassen. Integrating environmental issues into the mainstream: An agenda for research in operations management. *Journal of Operations Management*, 17(5):575–598, 1999.

J. Asmundsson, R. L. Rardin, and R. Uzsoy. Tractable nonlinear capacity models for aggregate production planning. Working paper, School of Industrial Engineering, Purdue University, West Lafayette, 2002.

J. Asmundsson, R. L. Rardin, and R. Uzsoy. An experimental comparison of linear programming models for production planning utilizing fixed lead time and clearing functions. Working paper, School of Industrial Engineering, Purdue University, West Lafayette, 2003.

J. Asmundsson, R. L. Rardin, C. H. Turkseven, and R. Uzsoy. Tractable nonlinear capacity models for production planning. Working paper, School of Industrial Engineering, Purdue University, West Lafayette, 2006a.

J. Asmundsson, R. L. Rardin, and R. Uzsoy. Tractable nonlinear production planning models for semiconductor wafer fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing*, 19(1):95–111, 2006b.

J. M. Asmundsson, R. L. Rardin, C. H. Turkseven, and R. Uzsoy. Production planning models with resources subject to congestion. *Naval Research Logistics*, 56(2):142–157, 2009.

J. B. Ayers. *Handbook of Supply Chain Management*. Auerbach Publications - Taylor & Francis Group, Los Angeles, California, 2000.

B. Beamon. Designing the green supply chain. *Logistics Information Management*, 12(4):332–342, 1999.

G. R. Bitran, E. A. Haas, and A. C. Hax. Hierarchical production planning: A two stage system. *Operations Research*, 30:232–251, 1982.

E. K. Boukas and Z. K. Liu. Manufacturing systems with random breakdowns and deteriorating items. *Automatica*, 27(3):401–408, 2001.

J.-W. Breithaupt, M. Land, and P. Nyhuis. The workload control concept: Theory and practical extensions of load oriented order release. *Production Planning & Control*, 13(7):625–638, 2002.

D. Briskorn. A note on capacitated lot sizing with setup carry over. *IIE Transactions*, 38:1045–1047, 2006.

E. Brodheim and G. P. Prastacos. The long island blood distribution system as a prototype for regional blood management. *Interfaces*, 9(5):3–20, 1979.

L. Buschkühl, F. Sahling, S. Helber, and H. Tempelmeier. Dynamic capacitated lot-sizing problems: A classification and review of solution approaches. *OR Spectrum*, 32:231–261, 2010.

J. A. Buzacott and J. G. Shantikumar. *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, New York, 1993.

E. Cagno, G.J. Micheli, and P. Trucco. Eco-efficiency for sustainable manufacturing: An extended environmental costing method. *Production Planning & Control*, 23 (2–3):134–144, 2012.

M. C. Caramanis and O. M. Anli. Dynamic lead time modeling for JIT production planning. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 2, pages 1450–1455, 1999.

M. Carey. Dynamic traffic assignment with more flexibility modelling within links. *Networks and Spatial Economics*, 1:349–375, 2001.

M. Carey and E. Subrahmanian. An approach to modelling time-varying ows on congested networks. *Transportation Research Part B*, 34:157–183, 2000.

M. Caridi and S. Cavalieri. Multi-agent systems in production planning and control: an overview. *Production Planning & Control*, 15(2):106–118, 2004.

K. Cattani, O. Perdikaki, and A. Marucheck. The perishability of online grocers. *Decision Sciences*, 38(2):329–355, 2007.

A. Chaabane, A. Ramudhin, and M. Parquet. Designing supply chains with sustainability considerations. *Production Planning & Control*, 22(8):727–741, 2011.

A. Chande, S. Dhekane, N. Hemachandra, and N. Rangaraj. Perishable inventory management and dynamic pricing using RFID technology. *Sādhanā*, 30(2-3):445–462, 2005.

C. Charney. Time to market: Reducing production lead time, 1991.

D. Chazan and S. Gal. A markovian model for a perishable product inventory. *Management Science*, 23(5):512–521, 1977.

W.-C. Chiang, J. C. H. Chen, and X. Xu. An overview of research on revenue management: Current issues and future research. *International Journal of Revenue Management*, 1(1):97–128, 2007.

S. Cholette and K. Venkat. The energy and carbon intensity of wine distribution: A study of logistical options for delivering wine to consumers. *Journal of Cleaner Production*, 17:1401–1413, 2009.

P. Chu and P. S. Chen. A note on inventory replenishment policies for deteriorating items in an exponentially declining market. *Computers & Operations Research*, 29(13):1827–1842, 2002.

A. J. Clegg, D. J. Williams, and R. Uzsoy. Production planning for companies with remanufacturing capability. *Proceedings of the 1995 IEEE International Symposium onElectronics and the Environment (ISEE)*, pages 186–191, 1995.

M. A. Cohen and D. Pekelman. LIFO inventory systems. *Management Science*, 24 (11):1150–1162, July 1978.

A. Corominas, A. Lusaa, and J. Olivella. A manufacturing and remanufacturing aggregate planning model considering a non-linear supply function of recovered products. *Production Planning & Control*, 23(2–3):194–204, 2012.

R. Darlington and S. Rahimifard. A responsive demand management framework for the minimization of waste in convenience food manufacture. *International Journal of Computer Integrated Manufacturing*, 19(8):751–761, 2006.

S. Dauzère-Pérès and J.-B. Lasserre. On the importance of sequencing decisions in production planning and scheduling. *International Transactions in Operational Research*, 9(6):779–793, 2002.

T. G. de Kok and J. C. Fransoo. Planning supply chain operations: Definition and comparison of planning concepts. In A.G. de Kok and S.C. Graves, editors, *Supply Chain Management: Design, Coordination and Operation*, Handbooks in Operations Research and Management Science, pages 597–675. North-Holland, Amsterdam, 2003.

R.B.M. de Koster, M.P. de Brito, and M.A. van de Vedel. Return handling: An exploratory study with nine retailer warehouses. *International Journal of Retail & Distribution Management*, 30(8):407–421, 2002.

R. Dekker, K. Inderfurth, L. N. van Wassenhove, and M. Fleischmann. *Reverse Logistics: A Quantitative Approach*. Springer, 1 edition, 2004.

B. Deniz, A. Scheller-Wolf, and I. Karaesmen. Managing inventories of perishable goods: The effect of substitution, 2004. available at: http://faculty.fuqua.duke.edu/seminarscalendar/Scheller-WolfSeminar.pdf, last call March 2012.

M. C. Derstroff. *Mehrstufige Losgrößenplanung mit Kapazitätzbeschränkungen*. Physica-Verlag, Bremthal, Heidelberg, 1995.

M. Despeisse, F. Mbaye, P. D. Ball, and A. Levers. The emergence of sustainable manufacturing practices. *Production Planning & Control*, 23(5):354–376, 2011.

S. Digiesi, G. Mossa, and G. Mummolo. A loss factor based approach for sustainable logistics. *Production Planning & Control*, 23(2–3):160–170, 2012.

D.M. Dilts and K.D. Ramsing. Joint lot sizing and scheduling of multiple items with sequence-dependent setup costs. *Decision Sciences*, 20(1):120–133, 1989.

W. Domschke, A. Scholl, and S. Voß. *Produktionsplanung: Ablauforganisatorische Aspekte*. Springer, Berlin, 2 edition, 1997.

A. Drexl and K. Haase. Proportional lotsizing and scheduling. *International Journal of Production Economics*, 40:73–87, 1995.

C. Duller. *Einführung in die nichtparametrische Statistik mit SAS und R - Ein anwendungsorientiertes Lehr- und Arbeitsbuch*. Physica, Heidelberg, 2008.

M. B. Dumas and M. Rabinowitz. Policies for reducing blood wastage in hospital blood banks. *Management Science*, 23(10):1124–1132, 1977.

N. Duque Ciceri, M. Garetti, and S. Sperandio. From product end-of-life sustainable considerations to design management. In B. Vallespir and T. Alix, editors, *APMS 2009, IFIP AICT 338*, pages 152–159. IFIP International Federation for Information Processing, 2010.

S. E. Elmaghraby. Production capacity: Its bases, functions and measurement. In K. G. Kempf, P. Keskinocak, and R. Uzsoy, editors, *Planning Production and Inventories in the Extended Enterprise*, volume 151 of *International Series in Operations Research & Management Science*, chapter 7, pages 119–166. Springer Science and Business Media, 2011.

S. T. Enns. MRP performance effects due to lot size and planned lead time settings. *International Journal of Production Research*, 39(1):461–480, 2001.

S. T. Enns and P. Suwanruji. Work load responsive adjustment of planned lead times. *Journal of Manufacturing Technology Management*, 15(1):90–100, 2004.

European Council. Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Energy efficiency plan 2011. Technical report, European Commission, Brussels, 2011.

M. Ferguson, V. Jayaraman, and G. C. Souza. Note: An application of the EOQ model with nonlinear holding cost to inventory management of perishables. *European Journal of Operational Research*, 180:485–490, 2007.

M. E. Ferguson and O. Koenigsberg. How should a firm manage deteriorating inventory? *Production and Operations Management*, 16(3):306–312, 2007.

S. Douwe P. Flapper, J. C. Fransoo, R. A. C. M. Broekmeulen, and K. Inderfurth. Planning and control of rework in the process industries: a review. *Production Planning & Control*, 13:26–34, 2002.

B. Fleischmann and H. Meyr. The general lotsizing and scheduling problem. *OR Spectrum*, 19(1):11–21, 1997.

M. Fleischmann. *Quantitative Models for Reverse Logistics*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 2001.

M. Fleischmann, J. M. Bloemhof-Ruwaard, R. Dekker, E. van der Laan, J. A. E. E. van Nunen, and L. N. van Wassenhove. Quantitative models for reverse logistics: A review. *European Journal of Operational Research*, 103(1):1–17, 1997.

M. Florian, J. K. Lenstra, and A. H. G. Rinnooy Kan. Deterministic production planning: Algorithms and complexity. *Management Science*, 26(7):669–679, 1980.

C. Floudas and X. Lin. Mixed integer linear programming in process scheduling: Modeling, algorithms, and applications. *Annals of Operations Research*, 139(1): 131–162, 2005.

A. Förster, K. Haase, and M. Tönnies. Ein modellgestützter Ansatz zur mittelfristigen Produktions- und Ablaufplanung für eine Brauerei. *Zeitschrift für Betriebswirtschaft*, 76(12):1255–1274, 2006.

J. W. Fowler and J. K. Robinson. Measurement and improvement of manufacturing capacity. MIMAC project final report, SEMATECH Technology Transfer 95062861A-TR, 1995. URL {http://www.fabtime.com/files/MIMDOE.PDF}.

G. M. Frankfurter, K. E. Kendall, and C. C. Pegels. Management control of blood through a short-term supply-demand forecast system. *Management Science*, 21 (4):442–452, 1974.

M. Garetti and M. Taisch. Sustainability manufacturing: Trends and research challenges. *Production Planning & Control*, 23(2-3):83–104, 2012.

B. C. Giri, A. K. Jalan, and K. S. Chaudhuri. Economic order quantity model with Weibull deterioration distribution, shortage and ramp-type demand. *International Journal of Systems Science*, 34(4):237–243, 2003.

B. Golany, J. Yang, and G. Yu. Economic lot-sizing with remanufacturing options. *IIE Transactions*, 33(11):995–1004, 2001.

M. Gopalakrishnan, D. M. Miller, and C. P. Schmidt. A framework for modelling setup carryover in the capacitated lot sizing problem. *International Journal of Production Research*, 33(7):1973–1988, 1995.

M. Gopalakrishnan, K. Ding, J.-M. Bourjolly, and S. Mohan. A Tabu-Search heuristic for the capacitated lot-sizing problem with set-up carryover. *Management Science*, 47(6):851–863, 2001.

S. C. Graves. A tactical planning model for job shops. *Operations Research*, 34: 522–533, 1986.

V.D.R. Guide. Production planning and control for remanufacturing: industry practice and research needs. *Journal of Operations Management*, 18(4):467–483, 2000.

V.D.R. Guide and M. S. Spencer. Rough-cut capacity planning for remanufacturing firms. *Production Planning & Control*, 8(3):237–244, 1997.

V.D.R. Guide and L. N. van Wassenhove. Evolution of Closed-Loop supply chain research. *Operations Research*, 57(1):10–18, 2009.

V.D.R. Guide, M. E. Kraus, and R. Srivastava. Scheduling policies for remanufacturing. *International Journal of Production Economics*, 48(2):187–204, 1997.

A. Gunasekaran and E. W. T. Ngai. Information systems in supply chain integration and management. *European Journal of Operational Research*, 159:269–295, 2004.

H.-O. Günther and H. Tempelmeier. *Produktion und Logistik*. Springer-Lehrbuch, Berlin, Heidelberg, New York, 5 edition, 2000.

J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk, and A. Meybeck. Global food losses and food waste. Technical report, Study conducted for the International Congress SAVE FOOD! at Interpack 2011, Düsseldorf, Germany for the Food and Agriculture Organization of the United Nations, 2011.

K. Haase and A. Kimms. Lot-sizing and scheduling with sequence-dependent setup costs and times and efficient rescheduling opportunities. *European Journal of Operational Research*, 66:159–169, 2000.

S. T. Hackman and R. C. Leachman. A general framework for modeling production. *Management Science*, 35:478–495, 1989.

J. Hall. Environmental supply chain dynamics. *Journal of Cleaner Production*, 8: 455–471, 2000.

A.A. Hervani, M.M. Helms, and J. Sarkis. Performance measurement for green supply chain management. *Benchmarking: An International Journal*, 12(4):330–353, 2005.

J. C. Ho, M. K. Shalishali, T.-L. Tseng, and D. S. Ang. Opportunities in green supply chain management. *The Coastal Business Journal*, 8(1):18–31, 2009.

W. Hopp and M. Spearman. *Factory Physics - Foundations of Manufacturing Management*. McGraw-Hill, Boston, 2 edition, 2001.

Y.-F. Hung and M.-C. Hou. A production planning approach based on iterations of linear programming optimization and flow time prediction. *Journal of the Chinese Institute of Industrial Engineers*, 18(3):55–67, 2001.

S. Hwang and R. Uzsoy. A single-product dynamic lot sizing model with work in process and congestion. Technical report, School of Industrial Engineering, Purdue University, West Lafayette, 2004.

K. Inderfurth. Simple optimal replenishment and disposal policies for a product recovery system with leadtimes. *OR Spectrum*, 19:111–122, 1997.

K. Inderfurth. Impact on uncertainties on recovery behavior in a remanufacturing environment. *International Journal of Physical Distribution and Logistics Management*, 35(5):318–336, 2005.

K. Inderfurth, G. Lindner, and N. P. Rachaniotis. Lot sizing in a production system with rework and production deterioration. *International Journal of Production Research*, 43(7):1355–1374, 2005.

K. Inderfurth, A. Janiak, M. Y. Kovalyov, and F. Werner. Batching work and rework processes with limited deterioration of reworkables. *Computers & Operations Research*, 33(6):1595–1605, 2006.

K. Inderfurth, M. Y. Kovalyov, C. T. D. Ng, and F. Werner. Cost minimizing scheduling of work and rework processes on a single facility under deterioration of reworkables. *International Journal of Production Economics*, 105(1):345–356, 2007.

M. Y. Jaber and M. A. Rosen. The economic order quantity repair and waste disposal model with entropy cost. *European Journal of Operational Research*, 188: 109–120, 2008.

A. Jaegler and P. Burlat. Carbon friendly supply chains: A simulation study of different scenarios. *Production Planning & Control*, 23(4):269–278, 2012.

J. B. Jennings. Blood bank inventory control. *Management Science*, 19(6):637–645, 1973.

M.L. Junior and M.G. Filho. Production planning and control for remanufacturing: Literature review and analysis. *Production Planning & Control*, iFirst:1–17, 2011.

B. Karimi, S.M.T. Fatemi Ghomi, and J. M. Wilson. A tabu search heuristic for solving the CLSP with backlogging and set-up carry-over. *Journal of the Operational Research Society*, 57:140–147, 2006.

U. S. Karmarkar. Batching to minimize flow times on parallel heterogeneous machines. *Management Science*, 35:607–613, 1989.

U. S. Karmarkar. Manufacturing lead times, order release and capacity loading. In S.C. Graves, A. Rinnooy Kan, and P. Zipkin, editors, *Logistics of Production and Inventory*, volume 4 of *Handbooks in Operations Research and Management Science*, pages 287–329. North-Holland, Amsterdam, 1993.

H. Kaspi and D. Perry. Inventory systems for perishable commodities with renewal input and poisson output. *Advances in Applied Probability*, 16(2):402–421, 1984.

S. Kim and R. Uzsoy. Exact and heuristic algorithms for capacity expansion problems. Technical report, Laboratory for Extended Enterprises at Purdue, School of Industrial Engineering, Purdue University, West Lafayette, 2006.

A. Kimms. Multi-level, single-machine lot sizing and scheduling (with initial inventory). *European Journal of Operational Research*, 89(1):86–99, 1996.

A.A. King and M.J. Lenox. Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production and Operations Management*, 10(3):244–265, 2001.

A. Klütsch. Green supply chain management - strategies for environmental protection: State-of-the-art mathematical models, 2009. Final year thesis.

W. Kubiak and M. Yavuz. Just-in-time smoothing through batching. *Manufacturing & Service Operations Management*, 10(3):506–518, 2008.

R. Kyrö, T. Luoma, S. Junnila, and E. Määttänen. Linking lean to green – energy efficiency as a value stream. In *10th EuroFM Research Symposium*, Vienna, Austria, 2011. EFMC2011.

M. Lautenschläger. *Mittelfristige Produktionsprogrammplanung mit auslastungsabhängigen Vorlaufzeiten.* Phd thesis, University of Darmstadt, Darmstadt, 1999.

S. C. H. Leung and W.-L. Ng. A stochastic programming model for production planning of perishable products with postponement. *Production Planning & Control*, 18(3):190–202, 2007.

X. Li and Q. Wang. Coordination mechanisms of supply chain systems. *European Journal of Operational Research*, 179(1):1–16, 2007.

C. Lin, B. Tan, and W.-C. Lee. An EOQ model for deteriorating items with time-varying demand and shortages. *International Journal of Systems Science*, 31(3): 391–400, 2000.

J.D. Linton, R. Klassen, and V. Jayaraman. Sustainable supply chains: An introduction. *Journal of Operations Management*, 25(6):1075–1082, 2007.

X. Liu, O. Tang, and P. Huang. Dynamic pricing and ordering decision for the perishable food of the supermarket using RFID technology. *Asia Pacific Journal of Marketing and Logistics*, 20(1):7–22, 2008.

S. C. H. Lu, D. Ramaswamy, and P. R. Kumar. Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants. *IEEE Transactions on Semiconductor Manufacturing*, 7:374–388, 1994.

R. Lund. Remanufacturing: United States experience and implications for developing nations. The World Bank, Washington DC, 1983.

O. Maimon, E. Khmelnitsky, and K. Kogan. *Optimal Flow Control in Manufacturing Systems, Production Planning and Scheduling.* Kluwer Academic Publisher, 1998.

S. K. Manna and K. S. Chaudhuri. An EOQ model with ramp type demand rate, time dependent deterioration rate, unit production cost and shortages. *European Journal of Operational Research*, 171:557–566, 2006.

F. Marinelli, M. Elena Nenni, and A. Sforza. Capacitated lot sizing and scheduling with parallel machines and shared buffers: A case study in a packaging company. *Annals of Operations Research*, 150:177–192, 2007.

K. N. McKay. The historic foundations of manufacturing planning and control systems. In K. G. Kempf, P. Keskinocak, and R. Uzsoy, editors, *Planning Production and Inventories in the Extended Enterprise*, volume 1 of *International Series in Operations Research and Management Science*, chapter 2, pages 21–31. Springer, 2011.

N. J. Mehta and N. H. Shah. Time varying increasing demand inventory model for deteriorating items. *Investigação Operacional*, 26(1):3–9, 2005.

S. A. Melnyk, R. P. Sroufe, and R. Calantone. Assessing the impact of environmental management systems on corporate and environmental performance. *Journal of Operations Management*, 21:329–351, 2003.

H. Meyr. *Simultane Losgrößen- und Reihenfolgeplanung für kontinuierliche Produktionslinien: Modelle und Methoden im Rahmen des Supply Chain Management.* PhD thesis, University of Augsburg, Wiesbaden, 1999.

H. Missbauer. *Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen.* PhD thesis, University of Heidelberg, Heidelberg, 1998. Physica Verlag.

H. Missbauer. Die Implikationen durchlauforientierter Losgrößenbildung für die Komplexität der Produktionsplanung und -steuerung. *Zeitschrift für Betriebswirtschaft*, 69(2):245–265, 1999.

H. Missbauer. Lot sizing in workload control systems. *Production Planning & Control*, 13(7):649–664, 2002.

H. Missbauer. Models of the transient behaviour of production units to optimize the aggregate material flow. *International Journal of Production Economics*, 118(2): 387–397, 2006a.

H. Missbauer. Lead time management and order release in manufacturing planning and control systems. Working Paper, University of Innsbruck, 2006b.

H. Missbauer and R. Uzsoy. Optimization models of production planning problems. In K. G. Kempf, P. Keskinocak, and R. Uzsoy, editors, *Planning Production and Inventories in the Extended Enterprise*, chapter 15, pages 437–507. Springer, 2011.

Y. Monden. *Toyota Production System: An Integrated Approach to Just-In-Time*. Productivity Press, 4 edition, 2011.

R. Naddor. *Inventory Systems*. John Wiley and Sons, New York, 1966.

A. Nahapetyan and S. Lawphongpanich. Discrete-time dynamic traffic assignment models with periodic planning horizon: System optimum. *Journal of Global Optimization*, 38:41–60, 2007.

E. Nikbakhsh. Supply chain and logistics in national, international and governmental environments. In R.Z. Farahani, N. Asgari, and H. Davarzani, editors, *Green Supply Chain Management*, pages 195–220. Physica, 2009.

P. Nyhuis. *Durchlauforientierte Losgrößenbestimmung*. PhD thesis, University of Hannover, Hannover, Düsseldorf, 1991.

F. Obenauf. Production planning with depreciation and rework/remanufacturing options - modeling and analysis with respect to load dependent lead times. Technical report, Institute of Information Systems, University of Hamburg, 2011.

S. Orcun, J. M. Asmundsson, R. Uzsoy, J. P. Clement, J. F. Pekny, and R. L. Rardin. Supply chain optimization and protocol environment (SCOPE) for rapid prototyping and analysis of complex supply chains. *Production Planning & Control*, 18(5): 388–406, 2007.

I. M. Ovacik. Advanced planning and scheduling systems: The quest to leverage erp for better planning. In K. G. Kempf, P. Keskinocak, and R. Uzsoy, editors, *Planning Production and Inventories in the Extended Enterprise*, volume 1 of *International Series in Operations Research and Management Science*, chapter 2, pages 33–43. Springer, 2011.

A. Özkavukcu and M.B. Durmuşoğlu. Production smoothing decisions with seasonal variation through value stream. In *40th International Conference on Computers and Industrial Engineering (CIE)*, pages 1–6, 2010.

J. Pahl. Integrating deterioration and lifetime constraints in production and supply chain planning: A survey. Invited survey; under review at the European Journal of Operations Research, 2011.

J. Pahl and C. Seipl. Efficient formulations for lot sizing models with perishability constraints. Working paper, University of Hamburg, 2011.

J. Pahl and S. Voß. Discrete lot-sizing and scheduling including deterioration and perishability constraints. In *Advanced Manufacturing and Sustainable Logistics*, number 46 in Lecture Notes in Business Information Processing, pages 345–357. Springer, Berlin Heidelberg, 2010.

J. Pahl and S. Voß. Production planning with deterioration and perishability. Draft Manuscript, 2011.

J. Pahl, S. Voß, and D. L. Woodruff. Load dependent lead times - from empirical evidence to mathematical modeling. In H. Kotzab, S. Seuring, M. Müller, and G. Reiner, editors, *Research Methodologies in Supply Chain Management*, pages 539–554. Physica, Heidelberg, 2005.

J. Pahl, S. Voß, and D. L. Woodruff. Production planning with load dependent lead times: An update of research. *Annals of Operations Research*, 153(1):297–345, 2007a.

J. Pahl, S. Voß, and D. L. Woodruff. Production planning with deterioration constraints: A survey. In J.A. Ceroni, editor, *The Development of Collaborative Production and Service Systems in Emergent Economies, Proceedings of the 19th International Conference on Production Research, IFPR*, pages 1–6, Valparaiso, Chile, 2007b.

J. Pahl, S. Voß, and D. L. Woodruff. Discrete lot-sizing and scheduling with sequence-dependent setup times and costs including deterioration and perishability constraints. In *Hawaiian International Conference on Systems Sciences (HICSS)*, pages 1–10, 2011.

D. Perry. Analysis of a sampling control scheme for a perishable inventory system. *Operations Research*, 47(6):966–973, 1999.

D. Perry and W. Stadje. Perishable inventory systems with impatient demands. *Mathematical Methods of Operations Research*, 50(1):77–90, 1999.

D. Perry and W. Stadje. Disasters in a markovian inventory system for perishable items. *Advanced Applied Probability*, 33(1):61–75, 2001.

A. Petrick, C. Steinhardt, J. Gänsch, and R. Klein. Dynamic control mechanisms for revenue management with flexible products. *Computers and Operations Research*, 37(11):2027–2039, 2010.

W. P. Pierskalla. Blood banking supply chain management. In M. Brandeau, F. Sainfort, and W.P. Pierskalla, editors, *Operations Research and Health Care*, pages 103–147. Kluwer Academic Publishers, New York, 2004.

W. P. Pierskalla and C. D. Roach. Optimal issuing policies for perishable inventory. *Management Science*, 18(11):603–614, 1972.

M. Pinedo. *Scheduling: Theory, Algorithms and Systems*. Prentical Hall, Englewood Cliffs, NJ, 2. nd. edition, 2002.

P. Porkka, A. P. J. Vepsäläinen, and M. Kuula. Multiperiod production planning carrying over set-up time. *International Journal of Production Research*, 41(6): 1133–1148, 2003.

M. E. Porter. *Competitive Strategy.* Free Press, New York, 1985.

M. E. Porter. *Competition in Global Industries.* Research Colloquium/Harvard Business School, Boston, Massechussets, 1986.

G. P. Prastacos. Blood inventory management: An overview of theory and practice. *Management Science*, 30(7):777–800, 1984.

D. Quadt and H. Kuhn. Capacitated lot-sizing with extensions: A review. *4OR: A Quarterly Journal of Operations Research*, 6(1):61–83, March 2008.

D. Quadt and H. Kuhn. Capacitated lot-sizing and scheduling with parallel machines, back-orders, and setup carry-over. *Naval Research Logistics*, 56(4):366–384, 2009.

U.S. Rao, J.M. Swaminathan, and J. Zhang. Demand and production with uniform guaranteed lead time. *Production & Operations Management*, 14(4):400–412, 2005.

P. Reusch. *Abstimmungsmechanismen zwischen Programmplanung und Mengenplanung in der mehrstufigen Produktionsplanung.* PhD thesis, Universität Duisburg-Essen, 2006.

G. Riaño. *Transient Behavior of Stochastic Networks: Application to Production Planning with Load-Dependent Lead Times.* PhD thesis, Georgia Institute of Technology, 2002.

K. Richter and M. Sombrutzki. Remanufacturing planning for the reverse wagner/whitin models. *European Journal of Operational Research*, 121:304–315, 2000.

K. Richter and J. Weber. The reverse Wagner Whitin model with variable manufacturing and remanufacturing costs. *International Journal of Production Economics*, 71:447–456, 2001.

C. Rizet and B. Keïta. Chaîne logistique et consommation d'énergie: cas du jaourt et du jean. Internet Source: http://www2.ademe.fr, 2005. last call March 2012.

D.S. Rogers and R. Tibben-Lembke. An examination of reverse logistics practices. *Journal of Business Logistics*, 22(2):129–148, 2001.

J. Sarkis. How green is the supply chain? Practice and research. *Social Science Research Network (SSRN): eLibrary*, 1999. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=956620.

J. Sarkis. A strategic decision framework for green supply chain management. *Journal of Cleaner Production*, 11(4):397, 2003.

J. Sarkis, L. M. Meade, and S. Talluri. E-logistics and the natural environment. *Supply Chain Management: An International Journal*, 9(4):303–312, 2004.

R.C. Savaskan. Closed-loop supply chain models with product remanufacturing. *Management Science*, 50(2):239–252, 2004.

C. Schneeweiß. *Hierarchies in Distributed Decision Making.* Springer, 1999.

F. Schulte. Analysis, planning and control of load-dependent lead times: A case study. Technical report, Institute of Information Systems, University of Hamburg, 2012.

C. Seipl. *Simulation kontinuierlicher Prozesse in hierarchi-schen Produktionsplanungssystemen*. PhD thesis, University of Hamburg, 2009. http://www.sub.uni-hamburg.de/opus/volltexte/2010 /4504/pdf/Dissertation_Christian_Seipl.pdf.

B. Selcuk, J. C. Fransoo, and A. G. De Kok. The effect of updating lead times on the performance of hierarchical planning systems. *International Journal of Production Economics*, 104(2):427–440, 2006.

B. Selcuk, J. C. Fransoo, and A. G. De Kok. Work-in-process clearing in supply chain operations planning. *IIE Transactions*, 40(3):206–220, 2008.

B. Selcuk, I. J. B. F. Adan, T. G. De Kok, and J. C. Fransoo. An explicit analysis of the lead time syndrome: Stability condition and performance evaluation. *International Journal of Production Research*, 47(9):2507–2529, 2009.

B. J. Shah, N. H. Shah, and Y. K. Shah. EOQ model for time-dependent deterioration rate with a temporary price discount. *Asia Pacific Journal of Operational Research*, 22(4):479–485, 2005.

E. A. Silver, D. F. Pyke, and R. Peterson. *Inventory Management and Production Planning and Scheduling*. John Wiley & Sons, New York, 3 edition, 1998.

D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi. *Designing and Managing the Supply Chain*. Irwin McGraw-Hill, Boston, 2 edition, 2002.

D.F. Simpson and D.J. Power. Use the supply relationship to develop lean and green suppliers. *Supply Chain Management: An International Journal*, 10(1):60–68, 2005.

C.R. Sox and Y. Gao. The capacitated lot sizing problem with setup carry-over. *IIE Transactions*, 31(2):173–181, 1999.

A. Srinivasan, M. Carey, and T. E. Morton. Resource pricing and aggregate scheduling in manufacturing systems. Working paper, GSIA, 1988.

S. K. Srivastava. Green supply-chain management: A state-of-the-art literature review. *International Journal of Management Reviews*, 9(1):53–80, 2007.

H. Stadtler. Mixed integer programming model formulations for dynamic multi-item multi-level capacitated lotsizing. *European Journal of Operational Research*, 94 (3):561–581, 1996.

H. Stadtler, C. Kilger, and H. Meyr. *Supply Chain Management and Advanced Planning*. Springer, Berlin, 4 edition, 2008.

A.T. Staggemeier and A.R. Clark. A survey on lot-sizing and scheduling models. In *23rd Annual Symposium of the Brazilian Operational Research Society (SOBRAPO)*, 2001.

C.-T. Su and C.-W. Lin. A production inventory model which considers the dependence of production rate on demand and inventory level. *Production Planning & Control*, 12(1):69–75, 2001.

C. Suerie. *Time Continuity in Discrete Time Models*. PhD thesis, University of Darmstadt, 2005.

C. Suerie and H. Stadtler. The capacitated lot-sizing problem with linked lot sizes. *Management Science*, 49(8):1039–1054, 2003.

B. Sundarakani, R. de Souza, M. Goh, S.M. Wagner, and S. Manikandan. Modeling carbon footprints across the supply chain. *International Journal of Production Economics*, 128:43–50, 2010.

S.R. Swenseth and M.R. Godfrey. Incorporating transportation costs into inventory replenishment decision. *International Journal of Production Economics*, 77(2): 113–130, 2002.

J.J. Tarí and J.F. Molina-Azorín. Integration of quality management and environmental management systems. *The TQM Journal*, 22(6):687–701, 2009.

I. P. Tatsiopoulos and B. G. Kingsman. Lead time management. *European Journal of Operational Research*, 14:351–358, 1983.

H. Tempelmeier and L. Buschkühl. A heuristic for the dynamic multi-level capacitated lotsizing problem with linked lotsizes for general product structures. *OR Spectrum*, 31:385 – 404, 2009.

C.-C. Teo, R. Bhatnagar, and S.C. Graves. An application of master schedule smoothing and planned lead time control. *Production and Operations Management*, pages 1–13, 2011.

R. H. Teunter and S. D. P. Flapper. Logistic planning and control of reworking perishable production defectives, 2001. available at: http://ideas.repec.org/p/dgr/eureir/1765001704.html, last call March 2012.

R. H. Teunter and S. D. P. Flapper. Lot-sizing for a single-stage single-product production system with rework of perishable product defectives. *OR Spectrum*, 25(1):85–96, 2003.

R. H. Teunter, E. van der Laan, and K. Inderfurth. How to set the holding cost rates in average cost inventory models with reverse logistics? *Omega*, 28(1):409–415, 2000.

R. H. Teunter, Z. P. Bayindir, and W. Van Den Heuvel. Dynamic lot sizing with product returns and remanufacturing. *International Journal of Production Research*, 44:4377–4400, October 2006.

L. B. Toktay, E. A. van der Laan, and M. P. de Brito. Managing product returns: The role of forecasting. In R. Dekker, K. Inderfurth, L. van Wassenhove, and M. Fleischmann, editors, *Reverse Logistics. Quantitative Models for Closed-Loop Supply Chains*, pages 1–27. Springer-Verlag, 2003.

M. Tsiros and C. M. Heilman. The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *Journal of Marketing*, 69:114–129, 2005.

S. Vachon and R. D. Klassen. Extending green practices across the supply chain. *International Journal of Operations & Production Management*, 26(7):795–821, 2006.

W. van den Heuvel. On the complexity of the economic lot-sizing problem with remanufacturing options. Technical report, Econometric Institute and Erasmus Research Institute of Management, Erasmus University Rotterdam, 2004. Last call April 2012.

J. van der Vorst, S. Tromp, and D.J. Zee. Simulation modelling for food supply chain redesign; integrated decision making on product quality, sustainability and logistics. *International Journal of Production Research*, 47(3):6611–6631, 2009.

A. P. Vepsalainen and T. E. Morton. Improving local priority rules with global lead-time estimates: A simulation study. *Journal of Manufacturing and Operations Management*, 1:102–118, 1988.

S. Voß and D. L. Woodruff. *An Introduction to Computational Optimization Models for Production Planning in a Supply Chain*. Springer, Berlin, 2 edition, 2006.

S. Voß and D. L. Woodruff. *An Introduction to Computational Optimization Models for Production Planning in a Supply Chain*. Springer, Berlin, 2 edition, 2006.

H. M. Wagner and T. M. Whitin. Dynamic version of the economic lot size model. *Management Science*, 5(1):212–219, 1958.

X. Wang, D. Li, and C. O'Brian. Optimization of traceability and operations planning: An integrated model for perishable food production. *International Journal of Production Research*, 47(11):2865–2886, 2009.

X. Wang, D. Li, and X. Shi. A fuzzy model for aggregative food safety risk assessment in food supply chains. *Production Planning & Control*, 23(5):377–395, 2011.

H. Waterer. Lot-sizing with inventory gains. *Operations Research Letters*, 35(6): 759–766, 2007.

D. L. Woodruff and S. Voß. A model for multi-stage production planning with load dependent lead times. In R.H. Sprague, editor, *Proceedings of the 37th Annual Hawaii International Conference on System Science*, pages 1–9. IEEE, Piscataway, 2004.

World Commission on Environment and Development (WCED). Report of the world commission on environment and development: Our common future. Internet Source: http://www.un-documents.net/wced-ocf.htm, 1987. last call March 2012.

J. Yang, B. Golany, and G. Yu. A concave-cost production planning problem with remanufacturing options. *Naval Reserach Logistics*, 52(1):443–458, 2005.

C. A. Yano. Setting planned leadtimes in serial production systems with tardiness costs. *Management Science*, 33:95–106, 1987.

M. Yavuz and E. Akçali. Production smoothing in just-in-time manufacturing systems: A review of the models and solution approaches. *International Journal of Production Research*, 45(16):3579–3597, 2007.

M. Yavuz and S. Tufekçi. The single-level batch production smoothing problem: An analysis and a heuristic solution. Technical report, Department of Industrial and Systems Engineering, University of Florida, 2004.

M. Yavuz, E. Akçali, and S. Tufekçi. Optimizing production smoothing decisions via batch selection for mixed-model just-in-time manufacturing systems with arbitrary setup and processing times. *International Journal of Production Research*, 44(1): 3061–3081, 2006.

E. Yücesan and X. de Groote. Lead times, order release mechanisms, and customer service. *European Journal of Operational Research*, 120:118–130, 2000.

B. J. Yum and E. D. McDowell. Optimal inspection policies in a serial production system including scrap rework and repair: An MILP approach. *International Journal of Production Research*, 25(10):1451–1464, 1987.

G. Zäpfel. *Taktisches Produktions-Management*. De-Gruyter-Lehrbuch, Berlin, New York, 1989.

G. Zäpfel and H. Missbauer. Production planning and control (PPC) systems including load-oriented order release - problems and research perspectives. *International Journal of Production Economics*, 30:107–122, 1993a.

G. Zäpfel and H. Missbauer. New concepts for production planning and control. *European Journal of Operational Research*, 67:297–320, 1993b.

A.M. Zargar. Effect of rework strategies on cycle time. *Computers and Industrial Engineering*, 29(1–4):239–243, 1995.

P.-X. Zhao. An EOQ model for items with Weibull distribution deterioration. In *2nd IEEE Conference on Industrial Electronics and Applications*, pages 451–454, 2007.

Z. Zhao. Optimization of the production planning for remanufacturing. In *2011th International Conference on Business Management and Electronic Information (BMEI)*, 2011.

W. H. M. Zijm and R. Buitenhek. Capacity planning and lead time management. *International Journal of Production Economics*, 46/47:165–179, 1996.