Statistische Methoden der industriellen Prozessmodellierung zur Echtzeitqualitätskontrolle am Beispiel einer kontinuierlichen Produktion von Faserplatten

DISSERTATION

zur Erlangung des Doktorgrades an der Universität Hamburg Fachbereich Biologie

> vorgelegt von Jörg Hasener

Hamburg 2004

Genehmigt vom Fachbereich Biologie der Universität Hamburg auf Antrag von Herrn Professor Dr. A. FRÜHWALD Weitere Gutachter der Dissertation: Herr Professor Dr. D. NOACK

Tag der Disputation: 21. Dezember 2004

Hamburg, den 07. Dezember 2004



Professor Dr. Arno Frühwald Dekan

Vorwort

Die vorliegende Arbeit entstand am Zentrum Holzwirtschaft der Universität Hamburg in Zusammenarbeit mit dem Institut für Holzphysik und mechanische Technologie des Holzes der Bundesforschungsanstalt für Forst- und Holzwirtschaft in der Zeit von Oktober 2001 bis Dezember 2004. Teile dieser Arbeit wurden finanziell mit Mitteln des Bundeswirtschaftsministeriums durch die AiF über die Deutsche Gesellschaft für Holzforschung gefördert.

Herrn Professor Arno Frühwald danke ich für die finanzielle und vertauensvolle Unterstützung dieser Arbeit. Besonders möchte ich mich jedoch dafür bedanken, dass ich Teile dieser Arbeit in Schweden schreiben konnte.

Herrn Professor Detlef Noack danke ich für die freundliche Übernahme des Koreferats.

Die Firma Kronotex, Heiligengrabe, stellte mir die Daten für die vorliegenden Untersuchungen zur Verfügung. Dafür möchte ich mich insbesondere bei Herrn Gunnar Thielecke und Herrn Jürgen Matthes bedanken. Ohne diese Daten wäre diese Arbeit nicht in dieser Form zustande gekommen. Es ist nicht selbstverständlich, industrielle Informationen für wissenschaftliche Untersuchungen zu überlassen. Umso mehr bedanke ich mich für diese Offenheit.

Herrn Dr. Hans Lobenhoffer möchte ich für seine unermüdliche Anteilnahme an der Entwicklung dieser Arbeit und für seine Diskussionsbereitschaft danken.

Herrn Gregor Bernardy und Herrn Bernd Rußkamp, von der Firma atr, danke ich für ihre Unterstützung in Fragen der industriellen Prozessmodellierung.

Herrn Dr. Eckard Moll danke für die Möglichkeit der Teilnahme an von ihm sorgsam organisierten SAS Fortbildungen. Dort erhielt ich viele nützliche Hinweise im Umgang mit dieser umfangreichen Software. Frau Carina Ortseifen danke ich für Organisation des Experten-Forums "SAS-EDU". In diesem Forum konnte ich in entscheidenden Fällen immer Hilfe erhalten.

Frau Karin Gembruch und Herrn Klaus Rosenbusch danke ich für ihre engagierte und zuverlässige Mitarbeit in den verschiedenen Projekten.

Den Mitarbeitern des Zentrums Holzwirtschaft und des Instituts für Holzphysik und mechanische Technologie des Holzes, hier insbesondere Frau Dörte Bielenberg, danke ich für ihre Hilfsbereitschaft.

Bei Herrn Dr. Andreas Lang möchte ich mich für die Korrektur des Manuskripts bedanken.

Gliederung

1	Einle	eitung	1
2	Keni	ntnisstand und Problemstellung	3
3	Ziele	e der Untersuchungen	10
4	Proz	ess der Holzwerkstoffherstellung	11
	4.1	Verfahrenstechnik	11
	4.2	Prozessparameter	12
	4.2.1	Spanaufbereitung	12
	4.2.2	Faseraufbereitung	13
	4.2.3	Trocknung und Sichtung	14
	4.2.4	Beleimung	14
	4.2.5	Streuung und Formband	15
	4.2.6	Presse	16
	4.2.7	Fertigplatte/Endfertigung	16
5	Stati	stische Prozessmodelle in der Literatur	17
	5.1	Spanplatten-Modelle	17
	5.2	MDF-Modelle	24
	5.3	OSB-Modell	27
6	Meth	oden der Prozessmodellierung	28
	6.1	Regressionsanalyse	28
	6.1.1	Grundlagen	28
	6.1.2	Gleichungssystem und Parameterschätzverfahren	29
	6.1.3	Modellannahmen	31
	6.1.4	Daten- und Prozessanalyse	32
	6.1.5	Modellbildung	33
	6.1.6	Modellaktualisierung und ex-ante-Vorhersagen	33
	6.1.7	Variablenauswahlverfahren	33
	6.2	Faktorenanalyse	35
	6.2.1	Grundlagen	35
	6.2.2	Faktorenmodell	36
	6.2.3	Modellannahmen	37
	6.2.4	Ladungsmatrix	37
	6.2.5	Rotationsverfahren	38
	6.2.6	Faktorenwertematrix	39
	6.3	PLS - Partial Least Square	39
	6.3.1	Grundlagen	39
	6.3.2	PLSR-Modell	40
	6.3.3	Modellannahmen	42
7	Keni	nzahlen der Modellgüte	43
	7.1	Kennwerte der ex-post-Prognose	43

	7.2	Kennwerte der ex-ante-Prognose (Validierung)	49
	7.3	Beurteilung der Modellgüte	51
8	Date	nerfassung	53
9	Stati	stische Prozessmodellierung	54
	9.1	Daten- und Prozessanalyse	54
	9.1.1	Ziele der Untersuchungen	54
	9.1.2	Statistische Kennzahlen	54
	9.1.3	Prüfung auf Stationärität	55
	9.1.4	Grafische Darstellungen (Line Plots)	56
	9.1.5	Plausibilitätsprüfung	57
	9.1.6	Schwankungsanalyse	58
	9.1.7	Ausreißeranalyse	60
	9.1.8	Test auf Normalverteilung	62
	9.1.9	Korrelationsanalyse und Bewertungszahlen	65
	9.1.1	0 Analyse der Varianzen	71
	9.1.1	1 Pressprogramm	73
	9.1.1	2 Hilfsvariablen	73
	9.2	Regressionsanalyse	76
	9.2.1	Ziele der Untersuchungen	76
	9.2.2	Modelle nach der Methode der optimalen Eingänge	76
	9.2.3	Modelle unter Berücksichtigung des Varianzinflationsfaktors VIF	
	9.2.4	Modelle nach klassischer Variablenauswahl	98
	9.2.5	Untersuchung der Residuen	102
	9.2.6	Endogene Regressoren	106
	9.2.7	Einfluss der Modellaktualisierung in den Parametern	108
	9.2.8	Langfristige Vorhersagen	111
	9.2.9	Technologische Betrachtungen	116
	9.3	Faktorenanalyse	127
	9.3.1	Ziele der Untersuchungen	127
	9.3.2	Faktoranalytische Untersuchungen des Pressprogramms	128
	9.3.3	Einfluss der Fortschreibung des Datensatzes	137
	9.3.4	Regressionsmodelle mit Pressenfaktoren der PCA-Analyse	139
	9.4	PLS-Analyse	142
	9.4.1	Ziele der Untersuchungen	142
	9.4.2	Modellwahl (Einzel- oder Gesamtmodell)	142
	9.4.3	Anzahl der PLS-Faktoren	144
	9.4.4	Diagnose und Interpretation	147
	9.4.5	Multivariate PLS-Modelle	157
	9.4.6	PLS-Einzelmodelle	162
	9.4.7	Langfristige Vorhersagen	175
	9.4.8	Technologische Zusammenhänge	179
10	Disk	ussion und Schlussfolgerungen	185
	10.1	Modellannahmen und -nutzen	185

10.2	Variablenauswahl	187
10.3	Vorhersagegenauigkeit	190
10.4	Industrielle Relevanz und Bewertung der Ergebnisse	192
10.5	Weiterer Forschungs- und Entwicklungsbedarf	194
11 Zu	sammenfassung	196
12 Ta	bellenverzeichnis	203
13 Ab	bildungsverzeichnis	206
14 Lit	eratur	210
15 An	hang	219

Abkürzungsverzeichnis

BWZ	Bewertungszahl
CV	Variationskoeffizient (Coefficient of Variation)
DW	Durbin-Watson-Koeffizient
FG	Freiheitsgrad
HDF	Harte Faserplatte (High Density Fibreboard)
IB	Querzugfestigkeit (Internal Bond)
К	Anzahl an Regressoren
k. A.	keine Angabe
Kz	Kennziffer
LVR	Latente Variablen Regression
Mean	Mittelwert
MLR	Multiple Lineare Regression
MOR	Biegefestigkeit (Modul of Rupture)
MOE	E-Modul (Modul of Elasticity)
MD	Mittlere Rohdichte (Mean Density)
MDF	Mitteldichte Faserplatte (Mean Density Fibreboard)
MUF-Harz	Melamin-Harnstoff-Formaldehydharz
OPTSEL	Optimal Selection (Variablenauswahlverfahren)
OSB	Oriented Strand Board
PCA	Hauptkomponenten-Analyse (Principal Component Analysis)
PCR	Hauptkomponenten-Regression (Principal Component Regression)
PLSR	Projektion auf latente Strukturen (Partial Least Square Regression)
R²	Bestimmtheitsmaß R ²
RMSE	Root Mean Square Error (im Beobachtungsraum)
$RMSE_{P}$	Root Mean Square Error of Prediction (im Vorhersageraum)
RMSE_Prog	RMSE _P
SS	Abhebefestigkeit (Surface Soundness)
S _R	Reststandardabweichung (RMSE)
STD	empirische Standardabweichung (Standard Deviation)
TS	Dickenquellung (Thickness Swelling)
UF-Harz	Harnstoff-Formaldehydharz
Х	Prozessdaten, Einflussvariablen
Y	Platteneigenschaften, Zielvariablen
VIF	Varianzinflationsfaktor

Abkürzungen innerhalb der PLS-Modelle

A	Anzahl der Komponenten im PLS-Modell, Index: a=1,2,A
К	Anzahl der X-Variablen, Index: k=1,2,K
Μ	Anzahl der Y-Variablen, Index: m=1,2,M
N	Anzahl der Beobachtungen, Index: i=1,2,N
Х	Matrix der Einflussgrößen, eine N×K Matrix
Y	Matrix der Zielgrößen, eine N $ imes$ M Matrix
b _m	Vektor der Regressionskoeffizienten des m-ten y, ein K $ imes$ 1 Vektor
В	Matrix der Regressionskoeffizienten aller Y's, eine K $ imes$ M Matrix
С	PLSR Y-Gewichtsmatrix mit \mathbf{c}_{a} als Spalten, eine M $ imes$ A Matrix
E	Matrix der X-Residuen, eine N×K Matrix
F	Matrix der Y-Residuen, eine N×M Matrix
Р	Ladungsmatrix mit p _a als Spalten
т	Score Matrix mit \mathbf{t}_{a} als Spalten, eine N×A Matrix
U	Y-Score Matrix mit \mathbf{u}_{a} als Spalten, eine N×A Matrix
W	Matrix der X-Gewichte mit ${f w}_a$ als Spalten, eine K $ imes$ A Matrix
W*	modifizierte Matrix der X-Gewichte mit \mathbf{w}_{a}^{*} als Spalten, eine K×A Matrix
VIP	Variable Importance for Projection

1 Einleitung

Im Jahr 2003 wurden in Europa 55,6 Mio. m³ Holzwerkstoffe produziert. Davon entfielen ca. 65% auf Spanplatten, 20% auf Faserplatten und 4% auf OSB. Die wirtschaftlichen Rahmenbedingungen der letzten Jahre sind durch ungünstige Preis- und Kostenentwicklungen, verstärkten Konkurrenzdruck, Überkapazitäten, Unternehmenskonzentrationen, Internationalisierung der Märkte und zunehmender Nachfragekonkurrenz um den Rohstoff Holz bestimmt worden. Eine Marktkonsolidierung ist seit Anfang 2000 vor allem im Bereich der Spanplatte festzustellen. Die europäische Spanplattenproduktion ist seit diesem Jahr stagnierend bzw. rückläufig. Um im Wettbewerbsumfeld bestehen zu können, ist vor allem eine ständig verbesserte Produktqualität bei möglichst gleichen bzw. sinkenden Herstellkosten anzustreben. Dieses Ziel ist nur durch eine Senkung der Rohstoffkosten und eine ständige Optimierung des Herstellprozesses zu erreichen.



Abbildung 1: Produktion von Holzwerkstoffen in Europa (nach EPF 2003)

Der Herstellprozess von Holzwerkstoffen ist jedoch von sehr komplexen Wechselwirkungen zwischen den Anlagenparametern und den Rohstoffen sowie zwischen verschiedenen Teilprozessen geprägt. Die Regelung des Gesamtprozesses basiert im wesentlichen auf Erfahrung und auf der Überwachung und Kontrolle der Teilprozesse. Die Qualitätskontrolle beruht auf einer Stichprobennahme und auf der Kontrolle der in den entsprechenden EN-Normen gestellten Anforderungen durch eine i.d.R. zerstörende Materialprüfung.

Nach der europäischen Norm zur Fertigungskontrolle von Holzwerkstoffen EN 326-2 sind neben konventionellen Laborprüfungen auch sogenannte alternative Verfahren zugelassen, wenn ein ausreichender statistischer Zusammenhang zwischen den etablierten und neuentwickelten Verfahren nachgewiesen werden kann. Eine systematische Verknüpfung der Informationen der Teilprozesse, der Prozessparameter mit den Qualitätsdaten erfolgt nicht. Eine durchgängige optimale Prozess- und Produktkontrolle kann demnach nicht gewährleistet werden. Die Schwankungen im Prozess werden nicht systematisch erfasst. Um die 5%-Fraktile der technischen Spezifikationen in den EN-Normen einzuhalten, müssen

relativ hohe Eigenschaftsmittelwerte in der Produktion gefahren werden (Noack, 2003). Erhebliches Einsparpotenzial liegt demnach in der Reduzierung der Schwankungen zwischen den Mittelwerten der nach Norm zu prüfenden Eigenschaften. Geringere Schwankungen im Prozess ermöglichen gleichzeitig die Herabsetzung des langfristigen Mittelwertes der Platteneigenschaften, um die 5%-Fraktil-Anforderungen der technischen Spezifikationen zu erfüllen. Dafür ist es jedoch notwendig, jederzeit die Schwankungen im Prozess und deren Auswirkungen auf die Platteneigenschaften zu kennen. Eine Qualitätskontrolle auf Stichprobenbasis kann diese Anforderung nicht erfüllen.

Es besteht die Möglichkeit einer Verknüpfung der Prozess- mit den Qualitätsdaten, um die Zusammenhänge in einem statistischen Modell zu beschreiben. Besitzt das Modell eine ausreichende Genauigkeit, kann zukünftig aus den online messbaren Prozessdaten die Produktqualität zuverlässig vorhergesagt bzw. aus gewünschten Produkteigenschaften auf die Soll-Prozessdaten geschlossen werden. Dieses Verfahren ist zur Echtzeitqualitätskontrolle zu verwenden. Dadurch liegen nach dem Verlassen der Platten aus der Heißpresse laufend Erkenntnisse über die produzierte Plattenqualität vor. Systematische Schwankungen der Plattenmerkmale durch Änderungen in den Prozesseinstellungen sind erkennbar und quantifizierbar. Mit dem quantifizierbaren Wissen der Bedeutung einzelner Prozessvariablen auf die Qualitätsmerkmale sind neben den Vorhersagen auch Prozessoptimierungen möglich.

Die mathematischen Grundlagen der multivariaten regressionsanalytischen Prozessmodellierung in der Holzwerkstoffindustrie wurden von Lobenhoffer (1990) beschrieben. Durch neuere Entwicklungen der Mess- und Regeltechnik stehen heute bis zu 2500 Prozessvariablen (Young, 2004) aus der gesamten Produktion zur Verfügung. Diese Variablen sind vor allem innerhalb der Teilprozesse aber auch zwischen den Prozessabschnitten miteinander korreliert. Es existieren derzeit keine systematischen Untersuchungen zur optimalen Variablenauswahl korrelierter Variablen für Echtzeitvorhersagemodelle auf Basis der Regressionsanalyse. Auch stehen keine fundierten Informationen zur Verfügung, welche zukünftige und langfristige Vorhersagegenauigkeiten durch statistische Prozessmodelle zu erwarten sind.

Stark korrelierte Variablen führen in Regressionsmodellen zu Problemen in der Schätzung. Die Vorzeichen der Regressionskoeffizienten korrelierter Regressoren sind technologisch nicht interpretierbar. Dies führt i.d.R. weniger zu Problemen in der Schätzgenauigkeit dieser Modelle als viel mehr zu technologischen Fehlinterpretationen. Die Variablenauswahlverfahren müssen die Problematik der Korrelationen zwischen den Prozessgrößen berücksichtigen.

Die kontinuierliche Heißpresse weist innerhalb der Teilabschnitte zur Herstellung von Holzwerkstoffen die höchste Anzahl an korrelierten Variablen auf. Mit einer Hauptkomponenten- bzw. Faktorenanalyse können die Variablen zu unkorrelierten übergeordneten Strukturen zusammengefasst werden.

Um den Einfluss korrelierter Variablen auf die Zielgrößen rechenbar zu machen, wird die Methode der latenten Variablen Regression (LVR) bzw. Partial-Least-Square-Regression (PLSR) angewendet.

2 Kenntnisstand und Problemstellung

Für die Modellierung verfahrenstechnischer Prozesse stehen grundsätzlich zwei Möglichkeiten zur Verfügung. Diese sind:

- Analytische (rigorose) Verfahren auf der Basis physikalischer, chemischer bzw. verfahrenstechnischer Beziehungen
- Empirische Verfahren unter Verwendung von Messdaten

Die Verwendung des einen oder des anderen Verfahrens ist von unterschiedlichen Bedingungen und von der Zielstellung der Untersuchung abhängig. Nach Box und Hunter (1978) ist ein analytisches Modell dem empirischen Modell immer dann vorzuziehen, wenn (a) das grundlegende Verständnis des Systems unabdingbar für das Ziel der Untersuchung ist und (b) wenn der Stand der Technik so weit fortgeschritten ist, dass ein nützliches analytisches Modell einfach zu erreichen ist. Empirische Modelle werden jedoch auch zukünftig Bestand haben, da sie zum einem, Teil der Versuchsplanung sind und zum anderen in vielen Fragestellungen die kostengünstigere Alternative darstellen. Ob jemals hochkomplexe verfahrenstechnische Prozesse analytisch modelliert werden können, wird sich zukünftig zeigen müssen. Mit dem heutigen Kenntnisstand ist es nicht möglich. Auf dem Weg zur analytischen Modellierung wird jedoch ohne Zweifel auf die Erkenntnisse der empirischen Modellierung zurückgegriffen. Eine Lösung komplexer Fragestellungen werden möglicherweise koexistierende analytisch-empirische Modelle bieten.

Analytische Modelle basieren auf Differentialgleichungen und sind gewöhnlich nichtlinear in den Parametern. Die Gleichungsfunktionen sind im Gegensatz zur empirischen Modellierung bekannt. Als Unbekannte müssen nur die Parameter des verwendeten Materials geschätzt werden. Die Schätzung der Parameter erfolgt i.d.R. auf Basis einer experimentellen Versuchsplanung.

Nach Bärmann (2001) weisen beide Methoden Vor- und Nachteile auf. Analytische Modelle führen zu einem tieferen Verständnis und es besteht die Möglichkeit zur Extrapolation der Erkenntnisse auf neu zu errichtende Anlagen, womit implizit auch schon der entscheidende Nachteil empirischer Modelle genannt wird: die jeweils nur anlagenspezifische Modellierung. Die Nachteile der analytischen Verfahren bestehen dagegen darin, dass für viele interessierende Größen keine physikalischen bzw. chemischen Beziehungen bekannt sind und somit die Formulierung von mehreren Qualitätsgrößen eines komplexen Prozesses auf Basis aller Einflussfaktoren mit physikalischen und chemischen Beziehungen nicht möglich ist. Die erreichbare Modellierungsgenauigkeit analytischer Verfahren liegt bei komplexen Prozessen heute noch i.d.R. unter der Genauigkeit empirischer Modelle, wobei eine geeignete Datenbasis Voraussetzung für den Aufbau zuverlässiger, empirischer Modelle ist. Die Modellierungsgenauigkeit analytischer Modelle ist dagegen weitgehend vom Kenntnisstand der physikalischen und chemischen Gesetzmäßigkeiten abhängig.

Tabelle 1 fasst die Vor- und Nachteile von analytischen und empirischen Modellen zusammen.

	Analytische Modelle	Empirische Modelle
Vorteile	 tieferes Anlagenverständnis Übertragbarkeit der Erkenntnisse Extrapolationen möglich 	 höhere Modellierungsgenauigkeit geringerer Modellierungsaufwand Kostengünstig
Nachteile	 weniger zu schatzende Parameter aufwendige Forschungsarbeit Beziehungen für viele qualitätsbeeinflussende Größen unbekannt komplexe Prozesse nicht modellierbar 	 - Komplexe Prozesse modelilerbal - Erkenntnisse sind anlagenspezifisch - Interpretierbarkeit in Abhängigkeit der Methode zum Teil schwierig

	Tabelle 1: Vor-	und Nachteile von	analytischen und e	empirischen Modellen
--	-----------------	-------------------	--------------------	----------------------

Eine analytische Modellierung des gesamten Herstellprozesses von Holzwerkstoffen gibt es derzeit nicht. Arbeiten zur analytischen Modellierung des Teilprozesses Heißpresse bzw. des Dichteprofils kamen u.a. von Humphrey (1982), Harless et al. (1987), Humphrey und Bolton (1989), Sou und Bowyer (1994), Haselein (1998), Carvalho und Costa (1998), Dai et al. (2000), Thömen (2001), Zombori (2001) und Garcia (2002).

Von Haas (1998) bestimmte die notwendigen Materialparameter für die analytische Modellierung des Dichteprofils von Fasern, Spänen und OSB-Strands. Honvongjirawat (2003) entwickelte für die Bestimmung der horizontalen und vertikalen Permeabilität von Holzwerkstoffen ein Schnellverfahren.

Die analytische Modellierung des Temperatur-, Druck- und Gasverlaufes in der Matte und daraus abgeleitet die Entwicklung des Rohdichteprofils in diskontinuierlichen bzw. kontinuierlichen Pressen ist heute weitgehend für die Holzwerkstoffe Spanplatten, MDF und OSB möglich. Modellsimulationen für die Herstellung von MDF zeigen Thömen und Humphrey (2003). Die Integration der Klebstoffaushärtung in die Simulation zur Darstellung der Querzugentwicklung in der Presse ist der folgende Entwicklungsschritt. Heinemann (2003) stellte dazu die Entwicklung der Querzugfestigkeit als Funktion der Mattentemperatur, Presszeit und der mittleren Rohdichte dar. Mit dieser empirisch ermittelten Regressionsfunktion, lässt sich für einen definierten Klebstofftyp und Klebstoffmenge, anhand der simulierten Mattentemperatur und Rohdichte die Querzugfestigkeit in Abhängigkeit zur Presszeit simulieren. Anhand dieses Vorgehens wird deutlich, dass die Querzugfestigkeit, als stochastische Variable, nicht rein deterministisch erklärt werden kann. Es kommt an dieser Stelle zu einer Vermengung eines analytischen und eines empirischen Modells.

Die empirischen Modellierungsverfahren lassen sich untergliedern in Klassifikationsverfahren, Projektionsverfahren, regressionsanalytische Verfahren, Fuzzy-Verfahren und künstliche neuronale Netze. Des Weiteren existieren verschiedene Kombinationsmöglichkeiten dieser Verfahren.

Für die empirische Modellierung des Prozesses zur Herstellung von Holzwerkstoffen wurden bislang das Klassifikationsverfahren (z.B. Landmesser et al. 1988; Schweitzer und Ritter 1992), die regressionsanalytische Modellierung (z.B. Lobenhoffer 1990; Kruse et al. 1997;

Bernardy et al. 1999; Steffen et al. 2001) und Projektionsverfahren (Johnsson et al. 2000, Carlsson et al. 2002) angewendet. Reine (1998) stellt einen Ansatz zur Integration eines Prozessmodells in den Optimierungsvorgang einer Spanplattenproduktionsanlage auf Basis konkurrierender neuronaler Netze vor.

Bis auf regressionsanalytische Verfahren (und eine industrielle Anwendung des Projektionsverfahrens PLS) ist der industrielle Einsatz der oben erwähnten Verfahren in der Holzwerkstoffindustrie derzeit nicht bekannt. Die einzelnen Methoden werden im folgenden mit ihren Vor- und Nachteilen dargestellt.

Die Vorteile des **Klassifikationsverfahrens** werden nach Schweitzer und Ritter (1992) in den geringeren Anforderungen an das Datenmaterial, in der Aussonderung nicht relevanter Merkmale unter fachspezifischen Gesichtspunkten, in einer ausreichenden Darstellung der Genauigkeit einer Qualitätsklasse, in der multivariaten Zielgrößenbetrachtung und in der einfachen Struktur der Modelle gesehen. Sowohl von Landmesser et al. (1988) als auch von Schweitzer und Ritter (1992) wird die These vertreten, dass das Erreichen einer Qualitätsklasse die entscheidende Rolle spielt. Die genaue wertmäßige Realisierung bestimmter Qualitätsmerkmale ist nach diesen Autoren von untergeordneter Bedeutung.

Die multivariate Zielgrößenbetrachtung kann nicht als Vorteil der Klassifikationsverfahren gesehen werden, da diese mit den Methoden der Regressionsanalyse und der Projektionsmethoden ebenfalls möglich sind. Ebenso kann eine Aussonderung nicht relevanter Merkmale unter fachspezifischen Gesichtspunkten mit den letztgenannten Methoden durchgeführt werden. Eine Klassifizierung von Prozesszuständen und entsprechenden Qualitätsmerkmalen vereinfacht die Modellerstellung und Interpretation. Ein Vorteil einer Vorhersage einer bestimmten Qualitätsklasse gegenüber einer genauen wertmäßigen Vorhersage kann nicht gesehen werden. Die Darstellung des Einflusses direkter Prozessgrößen auf die einzelnen Qualitätsmerkmale ist informationstechnisch höher einzustufen als eine Beschreibung des Einflusses auf Qualitätsklassen.

Das **PLS-Verfahren** (partial least square) wurde in erster Linie von Chemometrikern aus dem ursprünglich von Wold (1966) entwickelten Pfadmodellen weiterentwickelt, um komplexe Signale, bestehend aus sehr vielen und hochkorrelierten Einzeldaten verarbeiten zu können. Eine anschaulichere Interpretation von PLS ergibt sich mit der Übersetzung **P**rojektionen auf Latente Strukturen. Die PLS-Regression kann als Verallgemeinerung der multiplen linearen Regression verstanden werden (Wold et al. 1984). Korrelierte Messwerte werden zu sogenannten latenten Variablen zusammengefasst. Die latenten Variablen sind Linearkombinationen der Regressoren. In der PLS-Analyse erklären diese Linear-kombinationen die maximal mögliche Variation in den Prozessvariablen als auch in den Qualitätsvariablen. In der PCR-Analyse (Principal Component Regression) dagegen erklären die latenten Variablen nur die Streuungen in den Prozessvariablen. Dies ist nichts anderes als eine Regressionsanalyse mit den Hauptkomponenten der Prozessvariablen.

Die PLS-Prozessmodellierung erfolgt mittels linearer stochastischer Gleichungen. Die Parameterschätzung basiert auf das OLS- (ordinary least square) Verfahren. Vorteile der PLS-Analyse werden in der Verarbeitung von korrelierten, fehlerbehafteten und nicht vollständigen Daten gesehen (Burnham et al., 1999 und Wold et al., 2001). Die Verfahren

setzen zudem nicht wie die klassische Regressionsanalyse voraus, dass die Datenreihenlänge größer ist als die Anzahl an Regressoren. Simultane Zielgrößen können modelliert werden, wobei vorhandene Interdependenzen zwischen den Zielgrößen nicht berücksichtigt werden. Der Nachteil einer simultanen Betrachtung von Zielgrößen ist die Erklärung dieser durch nur einen "X-Variablenblock". Die Interpretation der latenten Variablen erweist sich in vielen Fällen als schwierig. Die durch latente Variablen am besten repräsentierten Einflussgrößen können durch sogenannte Gewichte aufgezeigt werden. Einflussgrößen mit kleinen absoluten Gewichten sind für das Modell weniger bedeutend als die mit großen Gewichten.

Fuzzy regelbasierte Systeme kommen zum Einsatz, wenn Wissen über das Ein- und Ausgangsverhalten des betrachteten Prozesses zwar vorhanden ist, aber dieses nicht geschlossen mathematisch formuliert werden kann. Dieses Wissen wird in "wenn...dann..."-Regeln abgelegt. Diese Abhängigkeiten als Regeln zu formulieren, erfordert Expertenwissen. Eine herausragende Eigenschaft der Fuzzy-Regelbasis liegt in deren sehr guten Interpretierbarkeit, da das Wissen in Form sprachlich formulierter Regeln dargestellt ist. Somit dürften Fuzzy-Systeme für den Herstellungsprozess von Holzwerkstoffen grundsätzlich geeignet sein. Beschreibbar wären jedoch nur bekannte Beziehungen zwischen Kontrollvariablen und Ausgangsvariablen. Nach Lobenhoffer (1997) mangelt es zudem an einer statistischen Sicherung der Aussage, und diese ist für eine Qualitätskontrolle konstruktiv eingesetzter Holzwerkstoffe notwendig. Des weiteren sind die Lösungen von Fuzzy-Systemen unscharf.

Neuronale Netze eignen sich neben der Klassierung von Daten auch zur Abbildung funktionaler Zusammenhänge zwischen Ein- und Ausgangsgrößen. Neben sehr vielfältigen Netzstrukturen und -algorithmen unterscheiden sich die Netztypen generell im Verfahren des überwachten und des unüberwachten Lernens. Alle Netze, die mit unüberwachtem Lernen erstellt werden, können im wesentlichen nur für Klassifikationsaufgaben verwendet werden. Die dazu verwendeten Daten enthalten nur beschreibende Informationen über den Prozess (Eingangsdaten), keine Handlungsdaten. Netze, die aus Ein- und Ausgangsdaten mittels überwachtem Lernen erstellt werden, können sowohl für die Modellbildung als auch für die Regelung genutzt werden. Nach Lobenhoffer (1997) kann der mittlere quadratische Ausgangsfehler bei überwachtem Lernen minimiert werden.

Als ein entscheidender Nachteil der neuronalen Netze wird oft die fehlende Interpretierbarkeit des Modells genannt. Die gelernten funktionalen Zusammenhänge lassen sich nicht ohne weiteres aus dem Modell interpretieren. In diesem Sinne wird von einem "Black-Box-Verfahren" gesprochen.

Als methodischer Nachteil der **regressionsanalytischen Verfahren** wird von Ritter und Schweitzer (1992) für die Holzwerkstoffindustrie der Aspekt der hohen Korrelationen zwischen den Einflussgrößen genannt. Zu hohe Korrelationen zwischen den Einflussgrößen verletzen die Anwendung der Regressionsanalyse. Des weiteren ist es schwierig, ein ausgewogenes Verhältnis zwischen Einflussgrößen und Datensätzen, bei ausreichend hohem Bestimmtheitsmaß, zu erreichen. Die univariate Betrachtung der miteinander korrelierten Zielgrößen wird als methodischer Nachteil genannt und zugleich auf die Lösung durch die multivariate Behandlung des Problems hingewiesen.

Die Bedeutung dieser Aussagen hat sich in den letzten Jahren aufgrund technischer Entwicklungen und neuer Untersuchungen sehr stark relativiert. Die z.T. hohen Korrelationen einzelner Prozessparameter untereinander sind zwangsläufig vorhanden. Diese gilt es durch entsprechende statistische Verfahren zu identifizieren, um anschließend eine Auswahl als Grundlage für die nachfolgende Modellbildung zu treffen. Durch Prozessleitsysteme automatisierte Datenerhebungen können heute eine ausreichende Anzahl von Datensätzen innerhalb kürzerer Zeiträume erhoben werden, um zuverlässige und statistisch gesicherte Modelle zu bilden. Die Bestimmtheitsmaße neuerer Untersuchungen liegen i.d.R. deutlich höher als in älteren Untersuchungen (Bernardy, Scherff 1997).

Der herausragende Vorteil der regressionsanalytischen Modellierung ist deren direkte Nutzung zur Steuerung des Prozesses. Dem Anlagenführer werden konkrete Vorhersagen der Qualitätsmerkmale gegeben. Diese kann er bei Berücksichtigung entsprechender statistischer Unsicherheiten für die Regelung der Anlage verwenden. Des weiteren erlaubt dieses Verfahren eine validierte Echtzeitqualitätskontrolle nach ISO 9001:2000 sowie eine Optimierung des Prozesses.

Die mathematischen Grundlagen der regressionsanalytischen Prozessmodellierung für die Herstellung von Holzwerkstoffen wurden von Lobenhoffer (1990) beschrieben. Er stellt die Herstellung von Holzwerkstoffen dabei als stochastischen Prozess interdependenter Systeme dar. Die Produkteigenschaften werden durch die vorhandenen Rohstoffe und Maschinen realisiert. Der Prozess unterliegt systematischen und natürlichen (zufälligen) Schwankungen, weshalb in der Vorhersagegleichung die Störgröße enthalten ist. Lobenhoffer unterstellt einen stationären Prozess und die Unabhängigkeit der Störfunktionen von den vorbestimmten Variablen (Regressoren). Interdependenzen zwischen den Zielgrößen und mögliche Korrelationen der Störfunktionen innerhalb einer Periode werden durch eine dreistufige Kleinste-Quadrate-Schätzung (3-SLS) berücksichtigt. Gesicherte Korrelationen zwischen den Zielgrößen benennt Lobenhoffer (1982) zwischen der Querzugfestigkeit V20 und V100, zwischen der Querzugfestigkeit und der Biegefestigkeit und zwischen Biegefestigkeit und Biegeelastizitätsmodul.

Lobenhoffer weist die Gültigkeit seiner Überlegungen durch Untersuchungen in der Spanplattenindustrie nach (Lobenhoffer 1991). Die Datenerhebung und die Zeitsynchronisation erfolgten dabei noch weitgehend manuell. Durch den besonders großen Zeitaufwand zur manuellen Datenerhebung konnten jeweils nur 18 bis 47 Datensätze erhoben und für die Modellbildung verwendet werden. Eine echte Modellvalidierung über einen längeren Zeitraum, konnte aufgrund der wenigen Datensätze, nicht durchgeführt werden.

Es ist hinlänglich bekannt, dass Rohstoffeigenschaften eine besonders wichtige Rolle bei der Ausbildung der Platteneigenschaften besitzen (z.B. Rackwitz 1963, Liiri 1977). Kruse et al. (1997) erfassen daher, zusätzlich zu den vorhandenen Prozessparametern, weitere Größen zur besseren Charakterisierung der Rohstoffeigenschaften (z.B. Spandimensionen, pH-Wert und Pufferkapazität von Spänen und Klebstoff). Es wird nachgewiesen, dass einige dieser, nur mit hohem Aufwand, messbaren Größen die Platteneigenschaften in starkem Maße beeinflussen. In Untersuchungen von Groom et al. (1999) konnte z.B. ein signifikanter Einfluss des Verhältnisses zwischen juvenilen und adulten Fasern und der Faserlänge auf den Biege-E-Modul und der Biegefestigkeit von mitteldichten Faserlatten nachgewiesen werden. Ein zunehmender Feinanteil führte unabhängig von der Faserart zu einer Abnahme in der Biegefestigkeit, im Biege-E-Modul und in der Querzugfestigkeit.

Untersuchungen von Carlsson (2002) und Mathiasson (2003) zeigen, dass mit Hilfe der NIR-Spektroskopie und der PCA- bzw. PLS-Analyse sowohl verschiedene Holzfaserarten diskriminierbar sind als auch die Faserlängen bzw. Faserlängenverteilung gut vorherzusagen sind. Mit der gleichen Methode konnte die Leimmenge auf den Fasern bestimmt werden. Beide Versuche wurden in einer industriellen Anwendung bestätigt. Die aus der NIR-Spektroskopie online ableitbaren Variablen ermöglichen damit erstmals Schwankungen des Faserqualität und –beleimung messbar zu machen und in die Prozessmodelle zu integrieren.

Untersuchungen von Kessler et al. (2000, 2002) zeigen, dass mittels einer online NIR-Spektroskopie sowohl Änderungen in der Holzart und Rindenanteil als auch in den Kocherund Refinereinstellungen mit Methoden der multivariaten Datenanalyse (PCA, PLS) zu rekonstruieren sind. Die Messungen erfolgten an Hackschnitzeln auf einem Förderband vor dem Kocher als auch an Fasern in der Blowline nach dem Refiner einer Nassfaser-Plattenproduktion.

Rials et al. (2002) messen mittels der NIR-Spektroskopie sowohl die Oberfläche der Deckschicht als auch die Bruchfläche nach dem Querzugtest an im Labor hergestellten MDF-Proben mit verschiedenen Harztypen (UF, MUF), Harzmengen (5%, 10%) und Wachsmengen (0%, 1%). Mittels der NIR-Spektren und einer PCA-Analyse konnten die Proben nach dem verwendeten Harztyp, der Harzmenge und der Wachsmenge klassifiziert werden. Mit PLS-Modellen konnten die Querzugfestigkeit, die Biegefestigkeit als auch der Biege-E-Modul vorhergesagt werden. Die Bestimmtheitsmaße der Kalibriermodelle liegen dabei zwischen R²=0,76 und 0,82. Der durch die Cross-Validierung ermittelte Fehler RMSEP (Root Mean Square Error of Prediction) lag im Verhältnis zum Mittelwert bei der Querzugfestigkeit mit 18% am höchsten. Die Autoren erwähnen das herausragende Potential der NIR-Spektroskopie für die zukünftige Prozess- und Qualitätskontrolle.

Stahl (1993), Bernardy und Scherff (1997) und Steffen et al. (2001) wenden die regressionsanalytische Prozessmodellierung auf die Herstellung von MDF an. Auch hier wird zunächst die Eignung der Prozessmodelle für die zuverlässige Vorhersage der Platteneigenschaften nachgewiesen. Eine Validierung der Ergebnisse an einem unabhängigen Testdatensatz wurde auch in diesen Untersuchungen nicht durchgeführt. In den Untersuchungen wird aufgezeigt, dass durch eine Erfassung weiterer Größen (z.B. Außentemperatur, Faserlänge), neben den reinen Maschinenparametern, der Fehler des Modells verkleinert werden kann.

Scherff et al. (1999) beschreiben eine durch ein Prozessmodell abgeleitete Geschwindigkeitserhöhnung einer kontinuierlichen Presse um 6% durch die Anpassung von Temperaturen bestimmter Heizfelder in der Heißpresse, bei gleichbleibenden Qualitätsmerkmalen.

Die Anwendbarkeit der statistischen Prozessmodellierung für die Herstellung von OSB ist bislang noch nicht untersucht worden. Es fehlen heute noch entsprechende Datenerfassungssysteme in diesen (wenigen) Anlagen, die die Daten automatisch zeitlich in ihrer Entstehung zurückverfolgen.

Kruse (1997) weist darauf hin, dass die Vorhersagegenauigkeit der Querzugfestigkeit von Spanplatten bei Berücksichtigung einer neu entwickelten Ultraschalltechnik verbessert werden kann. Es ist davon auszugehen, dass neue Messtechniken, wie Dichteprofilmessung, verbesserte Spaltererkennung mit Hinweisen zur Klebfestigkeit, Feuchte- und Temperaturmessungen von Spänen und Fasern im Bunker, einen bedeutenden Informationsgehalt besitzen.

Als ein schwieriger Aspekt der statistischen Prozessmodellierung hat sich die Erstellung und Pflege der Modelle herausgestellt. Für die Erstellung der Modelle sind zunächst ausreichend viele Datensätze zu erfassen. Dabei werden die vorhandenen Maschinenparameter mit den Ergebnissen der Laborprüfung der Stichproben verknüpft. Bei der regressionsanalytischen Modellbildung muss aus statistischen Gründen die Anzahl der Datensätze immer größer sein als die Anzahl der ins Modell aufgenommenen Variablen. Da das Sammeln von etwa 30-50 Datensätzen in Abhängigkeit vom Produktionsprogramm zwischen 2 und 4 Monaten in Anspruch nimmt, beschränkt deren Anzahl die Anzahl der zu berücksichtigenden Prozessparameter. Auf der anderen Seite werden in modernen Anlagen häufig mehrere hundert Prozessparameter laufend erfasst. Somit muss bei der Modellbildung eine Auswahl der zu berücksichtigenden Prozessparameter getroffen werden. In der Vergangenheit zeigte sich, dass durch die alleinige Auswahl nach technologischen oder nach statistischen Aspekten selten stabile Modelle gebildet werden konnten. Eine Verknüpfung beider Auswahlkriterien war zumeist erfolgreicher (Bernardy, Scherff 1997). Die Vorgehensweise bei dieser Art der Modellbildung ist bei den bisherigen Untersuchungen unsystematisch und wurde in den Veröffentlichungen nicht oder nur ungenügend beschrieben.

Es ist grundsätzlich bekannt, dass die Gültigkeit und Genauigkeit der Modelle von Zeit zu Zeit zu überprüfen ist. So erkannte Schweitzer (1992) bei Untersuchungen in einer Spanplattenanlage, dass einige Messdaten eindeutig einer Drift unterlagen, konnte aber nicht darlegen, wie die Gültigkeit eines Modells überprüft werden kann und wie eine Modellanpassung durchgeführt werden soll. Die Drifterscheinungen ergaben sich dabei z.B. durch Veränderung der Klimabedingungen, der Holzfeuchte, der Zusammensetzung der Holzsortimente oder durch produktionsbedingte Umbauten der Anlage. Weitere Ursachen einer Drift können in der begrenzten Aufnahme von messbaren Einflussgrößen in das Modell sowie in unbeobachtbaren Störvariablen liegen (Lobenhoffer et al. 1999; Greubel et al. 1999). Zur Zeit bestehen keine Kenntnisse darüber, anhand welcher Kriterien die Überprüfung des bestehenden Modells zu erfolgen hat und vor allem, bei welchen Bedingungen und wie eine eventuell notwendige Anpassung durchzuführen ist.

3 Ziele der Untersuchungen

Anhand eines aus der Industrie zur Verfügung gestellten umfangreichen Datensatzes einer kontinuierlichen Faserplattenproduktion sollen die Methoden der multivariaten Regressionsanalyse und die Partial-Least-Square-Analyse (PLS) zur Vorhersage der Platteneigenschaften angewendet werden. Zur Erstellung der Modelle wird dabei auf einen 9 Monate umfassenden Lerndatensatz zurückgegriffen. Erstmals werden beiden Methoden anhand eines identischen Testdatensatzes validiert. Der sich zeitlich dem Lerndatensatz anschließende Testdatensatz wird dabei nicht zur Auswahl der Variablen und zur Erstellung der Modelle verwendet. Die Validierung der Modelle basiert auf einen Vergleich der ex-ante-Vorhersagen mit den Ergebnissen aus dem Labor.

Die Variablenauswahl zur Erstellung linearer Regressionsmodelle zur Qualitätsvorhersage erwies sich in den bisherigen Untersuchungen als problematisch. Eine Kombination statistischer und technologischer Auswahl wird von Kruse (1997) als am erfolgreichsten beschrieben. Eine echte Validierung der erstellten Modelle konnte aufgrund zu geringer Datensätze in den bisherigen Arbeiten jedoch nicht durchgeführt werden. Ein Schwerpunkt dieser Arbeit wird daher auf die Variablenauswahl und auf die Validierung der Vorhersagen der erstellten Regressionsmodelle gelegt. Es soll eine Methode zur Modellerstellung entwickelt werden, die in der Praxis anwendbar ist und eine ausreichende Vorhersage-genauigkeit leisten kann.

Die PLS-Methode ermöglicht es, auch den Einfluss korrelierter Prozessvariablen auf die Platteneigenschaften zu beschreiben. Die Vorhersagegenauigkeiten der Regressionsmodelle sollen mit denen der PLS-Modelle verglichen werden. Ein direkter Vergleich der Methoden wird durch die Verwendung des gleichen Lern- und Testdatensatzes gewährleistet. Die unterschiedlichen Möglichkeiten zur technologischen Interpretation der Modellzusammenhänge werden aufgezeigt.

Weitere Aspekte dieser Arbeit sind:

- Darstellung verschiedener Methoden der Daten- und Prozessanalyse
- Beschreibung von Kennzahlen zur Beurteilung der Modelle
- Untersuchungen zum Einfluss der Modellaktualisierung und endogener Regressoren
- Untersuchungen zur Langzeitstabilität der Modelle
- Untersuchungen zur Überprüfung der Modellgültigkeit
- Faktorenanalytische Untersuchungen (PCA-Analyse) der Heißpresse
- Technologische Interpretation besonders bedeutender Prozessvariablen in den Modellen

4 Prozess der Holzwerkstoffherstellung

4.1 Verfahrenstechnik

Im Folgendem wird eine kurze Beschreibung der Verfahrenstechnik zur Herstellung von Holzwerkstoffen gegeben. In erster Linie wird auf den MDF-Prozess eingegangen, da die nachfolgenden Darstellungen zur Prozessmodellierung auf einem Datensatz aus der MDF-Industrie basieren. Der interessierte Leser sei für ausführlichere Informationen der Verfahrenstechnik auf Deppe und Ernst (1996 und 2000) verwiesen.

Grundlage des MDF-Prozesses sind Hackschnitzel bzw. Sägespäne. In Deutschland werden überwiegend Nadelhölzer und zu geringen Anteilen Laubhölzer als Ausgangsmaterial verwendet. Die Aufbereitung der Rundhölzer bzw. der Industrieresthölzer zu Hackschnitzeln erfolgt über Scheiben- bzw. Trommelhacker. Je nach Standort und Marktpreisen werden zu großen Teilen Fremdhackschnitzel bzw. Sägespäne aus der Sägeindustrie eingekauft.

Die Aufbereitung der Hackschnitzel und der Sägespäne zu Fasern kann über zwei getrennte Linien mit jeweils einem Refiner erfolgen.

Die Hackschnitzel werden je nach Verschmutzungsgrad gesiebt, durch Magnetwalzen von Metallbestandteilen getrennt und durch Hackschnitzelwaschanlagen geführt. In einem Vorbunker direkt über dem eigentlichen Kocher können die Hackschnitzel bei Bedarf vorgedämpft werden.

Durch eine Stopfschnecke werden die Hackschnitzel in den Kocher überführt. Nach einer Kochzeit von 3-4 Minuten bei 160-180°C und ca. 8-9 bar werden die Hackschnitzel durch eine Speise- und Austragsschnecke in den Refiner transportiert und zerfasert. Zwischen Speiseschnecke und Refiner kann Paraffin und Harnstoff zugegeben werden. Die im Refiner erzeugten Fasern werden über eine Blasleitung (Blowline) dem Trockner zugeführt. In der Blowline wird die Leimflotte eingedüst. Die Fasern verbleiben je nach Trocknungstemperatur und Trocknerlänge zwischen 5 und 30 Sekunden im Trockner. Die beleimten Fasern werden dabei auf ungefähr 10% Holzfeuchte getrocknet. Nach der Trocknung der Fasern erfolgt eine Zwischenlagerung der Fasern im Faserbunker.

Vom Faserbunker gelangen die Fasern zu dem Streubunker. Durch Streuwalzen werden die Fasern auf das umlaufende Formband gestreut. Die Höhe der Fasermatte wird durch Abkämmwalzen (Skalper) eingestellt. Die Regelung des Flächengewichtes der gestreuten Matte erfolgt durch eine Flächengewichtsmessung und über die entsprechende Einstellung der Abkämmwalzen.

Die Fasermatte wird durch eine nicht beheizte Vorpresse vorverdichtet. In einer kontinuierlichen Heißpresse wird die Matte auf ihre gewünschte Enddicke gepresst. Dabei kommt es zu einer Vernetzungsreaktion des Klebstoffes, deren Effizienz die Eigenschaften der Platten entscheidend bestimmt. Durch ein sogenanntes Pressprogramm kann das Rohdichteprofil der Platte in bestimmten Grenzen eingestellt werden. Der Endlosstrang wird hinter der Heißpresse durch Diagonalsägen auf die geforderte Plattenlänge gesägt. Als Qualitätskontrolle werden im Anschluss der Heißpresse die Plattendicke, das Plattengewicht

und bei vorhandener Messtechnik das Rohdichteprofil und Ultraschalldämpfung online gemessen. Zum Konditionieren werden die Platten in Kühlsternwender gefahren.

Nach der Konditionierung der Platten erfolgt entweder eine Zwischenlagerung oder sie werden direkt durch die Schleifmaschinen geführt, um danach ins Reifelager zu gelangen. In den Schleifstrassen werden jeweils die vorausgehärteten, nicht wesentlich verdichteten Deckschichten abgeschliffen.

Generelle Unterschiede zur MDF-Herstellung bestehen in der Span- und OSB-Herstellung in der Art der Holzaufbereitung und in der Beleimung. Die Ansprüche an die Rohstoffqualität zur Aufbereitung der jeweiligen Typen nimmt von der Spanplatte über die MDF zur OSB zu. In der Regel wird für die OSB-Herstellung auf Rundholz zurückgegriffen. Erste Ansätze zur Verwendung von Altholz für die OSB-Herstellung wurden jedoch schon vorgestellt (Suzuki, 2000 und Thole 2002). Es ist aber weiterhin zu erwarten, dass Rundholz das Haupt-ausgangsprodukt für die Produktion von OSB bleibt.

Wird in der MDF-Herstellung heute zum größten Teil eine Nassbeleimung (Blowline) durchgeführt, so ist die Trockenbeleimung für die Spanplatten- und OSB-Herstellung das Standardverfahren. Das heißt, die Trocknung des Materials erfolgt, im Unterschied zur Nassbeleimung, vor der Beleimung. Der bedeutendste Nachteil der Nassbeleimung von Fasern liegt in der Voraushärtung des Klebstoffes während des nachfolgenden Trocknungsprozesses. Aufgrund der damit verbundenen höheren Kosten scheint der Trend in der MDF-Industrie heute zu einer Mischbeleimung zu führen. Die neuesten Entwicklungen in der Trockenbeleimung für MDF zeigen Gotsmann (2004) und Krüzner (2004) auf.

4.2 Prozessparameter

Die Prozessparameter spiegeln in Form von Messwerten den gesamten Herstellungsprozess von der Holzaufbereitung bis zur Endfertigung wieder und lassen sich in die folgenden Teilprozesse untergliedern:

- Span bzw.- Faseraufbereitung
- Trocknung und Sichtung
- Beleimung
- Streuung und Formband
- Heißpresse
- Endfertigung

Im nachfolgenden Kapiteln werden die Einflussgrößen und die entsprechenden Regelgrößen in der Reihenfolge der einzelnen Prozessabschnitte aufgeführt.

4.2.1 Spanaufbereitung

Einfluss- und Messgrößen

- Holzarten
- Holzqualität
- Rindenanteil

- Holzfremde Bestandteile
- Holzartenmischung
- Spangeometrie (Schlankheitsgrad)
- Spanform
- Spanfraktion
- Spanfeuchte
- Schüttgewicht
- Durchsatzmenge
- Stromaufnahme Hacker, Zerspaner und Mühlen
- Standzeit der Messer
- Messervorstand
- Schnittgeschwindigkeit
- Vorschubgeschwindigkeit
- Bunkerfüllstände
- Bunkeraustragsmengen

Bis auf die Stromaufnahme der Spanaufbereitungsmaschinen und zum Teil die Bunkerfüllstände bzw. -austragsmengen werden die oben aufgeführten Messdaten bisher nicht online erfasst. Zur automatischen Erfassung der Spangeometrie ist bis heute keine online Messtechnik erhältlich. Im Bereich der MDF-Herstellung sind dagegen erste online Messtechniken, wie die NIR-Spektrokopie (Carlsson, et al., 2002) oder optische Faserabmessungssysteme, z.B. PQM, Pulmac, BauerMcnett und QualScan (Goroyias and McCarthy, 2002) verfügbar. Mittels Laser-Scanner und Beugungsspektrometer konnten Benker et al. (2003) Veränderungen der technologischen Parameter wie Temperatur, Verweildauer und Mahlscheibenabstand im Refiner direkt im Prozess überwachen. Die Spanabmessungen dagegen werden im Labor per Siebanalyse offline erfasst.

4.2.2 Faseraufbereitung

Die Aufbreitung der Fasern umfasst dabei Kocher und Refiner.

Regelgröße

• Kocher Füllstand, Drehzahl der Stopf- bzw. Austragsschnecke, Differenzdruck

Einfluss- und Messgrößen

- Dampfmengen
- Dampfdrücke
- Temperaturen
- Differenzdrücke
- Kocher Füllstand
- Stromaufnahme Rührwerk Kocher
- Mahlspalt Refiner
- Vibration Refiner
- Stromaufnahme Refiner
- Stromaufnahme und Drehzahlen der Stopf-, Beschick- und Austragschnecken
- Ventilstellungen

4.2.3 Trocknung und Sichtung

Regelgröße

Ausgangsfeuchten

Einfluss- und Messgrößen

- Holzarten
- Mischungsverhältnisse
- Ein- und Ausgangsfeuchte Material
- Spanfraktionen
- Temperaturen Heizmedium und Material (Ein- und Ausgang)
- Verweilzeiten
- Luftgeschwindigkeiten
- Materialeintrag
- Füllstände Dosierbunker
- Dampfdruck
- Primärenergieeinsatz
- Sauerstoff-Kohlendioxid-Gehalt
- Gaszusammensetzung
- Gasmenge
- Emissionen
- Bunkerfüllstände (Sichtung)

Auf Grund der Regelgröße der Trocknung steht die Ausgangsfeuchte des Materials online zur Verfügung. Des weiteren werden die Temperaturen am Trocknerein- und -ausgang und weitere Temperaturmessungen im Sichtungsprozess standardmäßig bereitgestellt. Die Emissionen werden überwacht, stehen aber nicht kontinuierlich als Messdaten zur Verfügung. Bis auf die Bunkerfüllstände werden die restlichen Prozessgrößen selten kontinuierlich gemessen und aufgezeichnet. Es ist bekannt, dass die Materialfeuchte eine bedeutende Rolle für die Plattenqualität besitzt (Hague et al., 1999). Auf Grund der hohen Temperaturen am Ausgang des Trockners sind die Messsysteme jedoch oft störanfällig. Es empfiehlt sich eine stetige Überwachung des Messsystems.

4.2.4 Beleimung

Regelgröße

• Beleimungsgrad

Einfluss- und Messgrößen

- Temperaturen
- Materialmengen
- Fraktionen
- Geometrie
- Materialfeuchten
- pH-Wert
- Pufferkapazität
- Leimtemperaturen
- Leimzusammensetzung

- Feststoff- bzw. Wassergehalt
- Viskosität
- Gelierzeit
- Beleimungsart (Dosiergenauigkeit, Nachzerkleinerung)
- Tröpfchengröße
- Leimverteilung (je Spanfraktion)
- Verweildauer
- Leistungsaufnahme
- Austragsmengen

Die Beleimung zeigt sich neben der Holzaufbereitung als am geringsten messtechnisch überwacht. Standardmäßig stehen die mengenmäßige Dosierung des Klebstoffgemisches und die Leistungsaufnahme der Mischer (Trockenverfahren) online zur Verfügung. Die Messung der Span- und Leimtemperaturen setzt sich zunehmend durch. Es besteht großer Forschungsbedarf, die Leimverteilung, Tröpfchengröße und die chemische Charakterisierung des Klebstoffes messtechnisch online zu erfassen und in das Prozessleitsystem einzubinden. Es ist aus der Literatur bekannt, dass diese Größen einen bedeutenden Einfluss auf die Plattenqualität besitzen. Erste Ansätze zur online Charakterisierung der Leimverteilung für MDF sind mit der NIR-Spektroskopie bzw. mit der konfokalen Fluoreszenz-Laserscan-Mikroskopie gezeigt worden (Johnsson et al., 2000; Loxton et al., 2003).

4.2.5 Streuung und Formband

<u>Regelgröße</u>

• Flächengewicht

Einfluss- und Messgrößen

- Materialfeuchten
- Temperaturen
- Schüttdichte
- Art der Streuung
- Verteilung des Materials im Bunker
- Füllstand Dosierbunker
- Geschwindigkeit des Bodenbandes
- Dosierbunkeraustrag
- Fallhöhe
- Streubreite
- Formbandgeschwindigkeit
- Vliesgewicht
- Vlieshöhe
- Vliesfeuchte
- Vorverdichtung
- Fehlschüttungen

Die meisten der oben aufgeführten Größen werden kontinuierlich messtechnisch überwacht. Für die Plattenqualität bedeutsam sind neben dem Flächengewicht vor allem die Spantemperaturen und -feuchten.

4.2.6 Presse

<u>Regelgröße</u>

• Pressprogramm und Presszeit

Einfluss- und Messgrößen

- Presstemperaturen
- Spezifische Drücke
- Schließzeit
- Verdichtungszeit
- Entlüftung
- Distanzen
- Pressprogramm
- Multipotdrücke
- Permeabilität
- Dampfdruck
- Mattentemperatur
- Temperaturniveau je Zeit

Die kontinuierliche Heißpresse liefert messtechnisch die meisten Prozessvariablen. Für jeden Rahmen und Multipot werden die Distanzen bzw. Pressdrücke aufgezeichnet. Für jedes Heizfeld steht online die Öltemperatur zur Verfügung. Die Matte betreffende bedeutende Größen wie Dampfdruck, Permeabilität oder Temperatur werden online nicht erfasst.

4.2.7 Fertigplatte/Endfertigung

Online Messgrößen

- Dicke (Springback)
- Plattengewicht
- Abmessungen
- mittlere Dichte bzw. Flächengewicht (errechnet)
- Spaltererkennung
- Schallschwächung (Ultraschall)
- Rohdichteprofil

Bis auf die Schallschwächung und das Rohdichteprofil gehören die aufgeführten Messungen zum Standard einer Produktionsanlage. Die Messung des Rohdichteprofils auf Basis von Röntgenstrahlung setzt sich in der MDF-Industrie zunehmend durch. Zukünftig werden aus dem Rohdichteprofil abgeleitete Kennzahlen dem Prozessleitsystem zur Verfügung stehen. Die über die Spaltererkennung hinausgehende Auswertung des Ultraschallsignals ist heute für jede einzelne Messspur in Form von gemittelten Schallschwächungskoeffizienten möglich (Fuchs, 2003, Mays und Kleinschmidt, 2003). Es liegen bisher noch keine Erfahrungen vor, ob und wie diese Messgrößen die Vorhersagegenauigkeiten der Platteneigenschaften verbessern können.

Erste Untersuchungen zur zerstörungsfreien Messung des Feuchteprofils der Platten wurden von Wolter et al. (2002) vorgestellt.

5 Statistische Prozessmodelle in der Literatur

Die ersten statistischen Prozessmodelle sind Ende der 80er, Anfang der 90er Jahre für die Spanplattenproduktion auf Mehretagenpressen entwickelt worden. Die Arbeiten von Lobenhoffer sowie Schweitzer und Ritter sind hierbei die Pionierarbeiten für die Holzwerkstoffbranche. Prozessmodelle zur Beschreibung des MDF-Prozesses wurden, entsprechend des späteren Markteintrittes dieses Werkstoffes, Anfang bis Mitte der 90er Jahre veröffentlicht. Statistische Modelle zur Beschreibung des OSB-Prozesses sind bislang nicht publiziert worden.

Im Folgenden werden verschiedene statistische Modelle mit den ausgewählten Einflussgrößen dargestellt, die in die Literatur Eingang gefunden haben. Diese Modelle sind alle anlagenspezifisch erstellt und aufgrund unterschiedlicher Ansätze im Modelltyp (linear und quadratisch, univariat und multivariat) und verschiedener Methoden der Variablenauswahl und Verfahren der Parameterbestimmung nur eingeschränkt miteinander vergleichbar. Eine technologische Interpretation der in die Modelle aufgenommenen Variablen steht im Vordergrund.

Nachfolgend einige Erklärungen zum Verständnis der Modelle.

- + der Regressionskoeffizient ist im angegebenen Modell positiv
- der Regressionskoeffizient ist im angegebenen Modell negativ
- +/- hier wurden von den Autoren quadratische Modellgleichungen entwickelt, das erste Vorzeichen steht für den Regressionskoeffizienten 1. Ordnung und das zweite Vorzeichen für den Koeffizienten 2. Ordnung
- # keine Angabe über Koeffizientenvorzeichen
- ? vom Autor wird die sachlogische Interpretation angezweifelt bzw. verworfen

5.1 Spanplatten-Modelle

Modell nach Lobenhoffer (1990)

	IB V20	IB V100	TS	MOR
Feuchte nach Trockner	#	#	#	#
Fraktionsverteilung Späne	#	#	#	#
Temperatur Späne nach Beleimung	#	#	#	#
Temperatur Leim vor Beleimung	#	#	#	#
Spez. Beleimungsarbeit	#		#	#
Beleimungsfaktoren			#	#
Gelierzeit Harnstoffharzleim	#			
Feuchte beleimter Späne	#		#	#
Liegezeit beleimter Späne	#		#	#
Einstreumenge DS	#		#	#
Gesamte Einstreumenge	#		#	#
Presstemperatur	#	#	#	
Max. Pressdruck			#	#

Schließzeit Presse	#	#	#	#
Heizzeit Presse	#			#
Gewicht Rohplatte				#
Dicke Rohplatte		#	#	#
Dichte Rohplatte	#	#	#	#
Feuchte Rohplatte				#
	R² _d =74%	R ² =71%	R ² =68%	R ² =81%

Signifikante Einflussgrößen bei 3-SLS Schätzung mit t-Werten >1,8 aus 3 Versuchen, Versuch 1: 18 Datensätze und 11 Variablen; Versuch 2: 34 Datensätze und 26 Variablen; Versuch 3: 47 Datensätze und 41 Variablen, V100 basiert nur auf Versuch 1; Mehretagenpresse, R²_d: durchschnittliches Bestimmtheitsmaß aus 3 Versuchen

Modell nach León-Méndez und Thömen (1996)

	IB V20	IB V20	IB V20
	(Modell 1)	(Modell 2)	(Modell 3)
Länge (MS Fraktion >2mm)	+	+	+
Länge (MS Fraktion >1mm)	+	+	
Volumen (MS Fraktion >1mm)	-	-	
Volumen (MS Fraktion >2mm)			-
Oberfläche (MS Fraktion >1mm)	-	-	
Oberfläche (MS Fraktion >2mm)			-
Beleimungsfaktor (DS)	+		
Feuchte nach Beleimung (MS)	+	+	+
Feuchte nach Beleimung (DS)			+
Spantemperatur (MS)	+	+	
Spez. Druck HD05	+		+
Spez. Druck ND33		+	+
Temperatur HD06	+	+	+
Temperatur MD21	-		
Distanz HF02		+	+
Multipot HD07			-
Spez. Druck MD25		+	+
Leim pH (DS)	-	-	-
Bandgeschwindigkeit			-
	R ² >80%	R ² >80%	R²=85%

Multiple Regression, 44 Datensätze, auf 37 reduzierte Variablen, kontinuierliche Presse

Modell 1 = die 11 häufigsten Variablen aus 44 Modellen, Modell nach max. Bestimmtheitsmaß

Modell 2 = die 11 häufigsten Variablen aus 44 Modellen, Stepwise Forward-Variablenauswahl, p-Wert für F-Wert kleiner 0,15 ist Abbruchkriterium, einzelne Modelle haben unterschiedliche Variablenanzahl

Modell 3 = Vorauswahl von 18 Variablen nach technologischen Gesichtspunkten als auch Variablen, die nach stat. Auswahlverfahren gehäuft berücksichtigt wurden, danach Modellbildung Backward-Verfahren, Eliminierung bis alle Variablen einen p-Wert <0,1 besitzen

Modell nach Schweitzer (1992)

	TS_2h	IB	MOR
Blechdurchlaufzeit Streustation	+		+
Rohplattenmasse	-	+	+
Deckspanmassen in Streukästen DS1	-	-	
Deckspanmassen in Streukästen DS2	-	-	-?
Feingutanteil der DS	-		
Feingutanteil der MS			+
Staubanteil der MS	+	-	+?
Staubanteil der DS		-	
Scheinbare Extinktion der DS	-		
Scheinbare Extinktion der MS		+	
Presszeit	+?		
Zeit der ersten Druckabbauphase	-	+	
Verdichtungszeit			-
Rohplattendicke	+		-
Liegezeit im Beschickkorb		+?	
Liegezeit in geschlossener Presse			+
Massenspez. Rückfederung des Vlieses		+?	+?
Maximaler Pressdruck		+	+
Mittlere Rohdichte der Platte		+	+
Viskosität MS-Leimflotte		+?	
	R ² =82%	R ² =81%	R ² =83%
	S _R =0,49%	S _R =0,04MPa	S _R =0,9MPa

Multiple Regression nach max. Bestimmtheitsmaß, Aufnahme der Variablen mit S=95% statistisch abgesichert, 50 Datensätze, Reduktion auf 12 Variablen, Mehretagenpresse

(Diese Arbeit baut u.a. auf die Ergebnisse von Belimow (1978), Geyer u. Tietze (1984), Friedrich (1987), Landmesser (1990), Andres (1989), Sahling (1989) und Trinks (1989) auf.)

Modell nach Brade (1986)

	IB	MOR	MOE
Spandickenverhältnis d _{MS} /d _{DS}		-	
Verhältnis Holzart Kiefer/Eiche	+/-	+	+
Massenanteilverhältnis m _{DS} /m _{MS}	-	+	+/-
Festharzanteilverhältnis FH _{DS} /FH _{MS}	-?	+/-	+
Spanfeuchtenverhältnis u _{DS} /u _{MS}	-	-??	
Pressenschließzeit	+		-
	R ² =55%	R ² =70%	R ² =80%
	S _R =0,07 MPa	S _R =2.68 MPa	S _R =0,25 GPa

Labor, Versuchsplan mit 6 strukturellen und prozesstechnischen Einflussgrößen mit jeweils 3 Faktorstufen, Modellbildung nach Regressionsansatz mit linearen und quadratischen Termen aus Mittelwerten aus 4 Platten (n=28 Varianten), Signifikanz der Einflussgrößen im Modell S ≥ 80%

	TS_2h MS	TS_2h DS	MOR DS
Verdichtungszeit	-	-	-
maximaler Pressdruck	+	+	
Spanfeuchte	-	_/+	_/+
Vorpresszeit	/+	+/-	/+
Vorpressdruck	/-		
Temperatur Rücklauf	/+	-	
Presszeit	/-		/-
pH-Wert	/-	+	/-
Dauer bis Belüftung		/+	/+
Geschwindigkeit Tabletts			+/+
Vliesmasse			+/+
Liegezeit in Presse			+/+
Schließzeit Presse			/+
	R ² =85%	R ² =76%	R ² =61%
	S _R =1,13%	S _R =1.39%	S _R =20,42 MPa

Modell nach Andres (1989)

/+ : nur das quadratische Glied ist in der Gleichung berücksichtigt worden

Multiple Regression mit quadratischen Termen getrennt nach Deck- und Mittelschicht, Signifikanz der Einflussgrößen im Modell S ≥95%, Modelle mit R²<60% wurden nicht betrachtet; Reduzierung der Einflussgrößen auf 15 durch Streichung aller Variablen mit Korrelationskoeffizienten >0,8, Anzahl Datensätze nicht angegeben, Mehretagenpresse

Modell nach Schubert (1990)

	TS_2h	IB	MOR
Blechdurchlaufzeit Streustation	+?	-	+
Rohplattenmasse	-		+
Feuchte DS oben	-		
Feuchte DS unten		+?	+
Feuchte MS			-
Masseverhältnis DS/MS	-?		
Masseverhältnis DSu/DSo	-?		+
Presszeit	+?		
Zeit der 1. Druckabbaustufe	-		-?
Liegezeit im Beschickkorb			-
Verdichtungszeit			-
Rohplattendicke	+		-
Staubanteil MS		-	
Feingutanteil MS			+?
Festharzanteil MS		+	
Festharzanteil DS			-?
Rohdichte		+	+
	R ² =76%	R ² =81%	R ² =86%
	S _R =0,9%	S _R =0,035 MPa	S _R =0,75 MPa

Multiple lineare Regression, 50 Datensätze, Signifikanz der Einflussgrößen im Modell nicht angegeben, Mehretagenpresse, Härteranteil nicht erfasst

	MOR längs	MOR längs	MOR	IB	TS_2h
Rohdichte	+/+*	+/-*	+/-*	+/-*	-/+*
Feuchte Deckschichtspäne	+	-?*	+		+*
Feuchte Mittelschichtspäne	-	-*	-	+/-	-*
Verhältnis DS/MS-Späne	+	+	+		+/-
Plattenfeuchte	+	+	+	+	_*
Plattendicke	+?	-	-	+*	+
Plattenmasse	-?	+	-?		-
	N=18	N=18	N=18	N=25	N=19
	R ² =51%	R ² =64%	R ² =55,5%	R ² =49%	R ² =79%
	S _R =2,59	S _R =2,27	S _R =2,4	S _R =0,055	S _R =1,47
	N/mm²	N/mm²	N/mm²	N/mm²	%

Modell nach Mehlhorn und Thole (1992)

Multiple Regression mit quadratischen Termen, mit * gekennzeichnete Koeffizienten Signifikanz S \ge 80%

Daten stammen aus einem chinesischen Unternehmen, Versuchsdauer 3 Monate, keine Daten und Angaben über Heißpresse

Weitere Untersuchungen zur statistischen Prozessmodellierung erfolgten von Haslbeck (1994), Schütte (1998) und Greubel (1999). Die Schwerpunkte der Arbeiten von Haslbeck und Schütte lagen in der Überprüfung der Anwendbarkeit der Programme aus dem Softwareprogramm OPIC von Lobenhoffer und in Untersuchungen zur Vorhersagegenauigkeit verschiedener Modelle. Es werden verschiedene multivariate Modelle für die Querzugfestigkeit, Abhebefestigkeit, Dickenquellung und Biegefestigkeit erstellt. Eine Diskussion der ausgewählten Regressoren und deren technologische Relevanz wird jedoch nicht geführt.

Greubel (1999) verwendete einen multivariaten Ansatz mit 22 Prozessgrößen, die jeweils 47 aufeinanderfolgende Beobachtungen zur Berechnung der Regressionsmodelle einbezogen. Insgesamt lagen 77 bereinigte Beobachtungen vor. Die Modelle werden jeweils um eine Beobachtung verschoben, um die Änderung der Regressionskoeffizienten zu beobachten. Die Auswahl der Prozessgrößen erfolgte weitgehend nach technologischen Gesichtspunkten. Unter anderem umfassten die Modelle folgende Prozessgrößen:

Stochastische Größen: Temperaturen und Feuchten der Deckschichten, Stromaufnahme der Mischer und Formbandgeschwindigkeit.

Steuerbare Prozessgrößen: Leimverbrauch, Wasserverbrauch, Gelierzeit des Leims, Plattendichte und Flächengewicht.

Von Greubel (1999) werden keine genauen Angaben über die Prozessgrößen hinsichtlich der Auswahl und der Koeffizientenvorzeichen gemacht. Eine technologische Interpretation der einzelnen Prozessgrößen ist nicht möglich. Es wird von Greubel jedoch festgehalten, dass die Vorzeichen einzelner Variablen bei schrittweiser Verschiebung der Beobachtungen zum Teil wechselten.

Das Bestimmtheitsmaß liegt bei neueren Untersuchungen mit $R^2 \ge 80\%$ i.d.R. höher als in den älteren Untersuchungen. Mit zunehmendem Einsatz und Verbesserung der Messtechnik wird der Herstellprozess genauer abgebildet. Damit erhöht sich automatisch die Anzahl der Regressoren im Modell und damit das Bestimmtheitsmaß. Aus der alleinigen Angabe des Bestimmtheitsmaßes lässt sich keine Aussage über die ex-ante-Vorhersagegenauigkeit der Modelle ableiten. Nachteilig an allen Untersuchungen ist die geringe Anzahl an Datensätzen und damit die fehlende Möglichkeit einer umfassenden Validierung des Modells im Vorhersageraum.

León-Méndez und Thömen (1996) stellen fest, dass eine Aussage über die Modellgüte nur anhand einer neuen Stichprobe erfolgen kann. In der Arbeit der Autoren wurde der Versuch unternommen, die Vorhersagequalität der Modelle, aufgrund mangelnder Datensätze, durch eine Cross-Validierung zu bestimmen. Für weitere Untersuchungen wird jedoch empfohlen, die gebildeten Modelle anhand zusätzlich erhobener Datensätze zu überprüfen.

Auffallend ist, dass bedeutende Variablen, wie Beleimungsgrad und Presszeit sehr oft nicht in den oben dargestellten Modellen enthalten sind. Das lässt sich aufgrund zumeist kleiner Datenräume damit begründen, dass diese beiden Variablen in dieser Zeitspanne zu geringe Schwankungen aufwiesen und somit keinen Erklärungsanteil an den Zieleigenschaften besitzen. Dies ist bei Vorhaben der qualitätsbedingten Optimierung ein entscheidender methodischer Nachteil, der durch aktive Versuche mit größeren Schwankungsbreiten oder durch längere Datenräume passiv ausgeglichen werden muss.

Auch wird deutlich, dass oft Variablen im Modell enthalten sind, die technologisch als solche oder aber in ihrer Wirkungsrichtung nicht zu erklären sind. Dagegen lassen sich andere scheinbar unlogische Zusammenhänge sachlogisch erklären. Dazu ist jedoch i.d.R. eine genaue Kenntnis der Umstände in der jeweiligen Anlage notwendig. Schweitzer (1992) und Schubert (1990) fanden z. B. den Zusammenhang, dass mit zunehmender Presszeit auch die 2 Stunden-Dickenquellung zunimmt. Sie führten diesen Effekt auf eine Verlängerung der Verdichtungszeit und dem daraus folgenden weniger ausgeprägten Rohdichteprofil zurück. Bei zugleich anteiliger Verkürzung der Druckabbauzeit erfolgte kein vollständiger Abbau des Platteninnendrucks. Damit kam es zu einer höheren Rückfederung der Platte, die wiederum in einer erhöhten Dickenquellung resultierte. Grundsätzlich können innerhalb von Regressionsmodellen technologisch zweifelhafte Zusammenhänge durch Multikollinearitäten verursacht werden. Multikollinearitäten, hervorgerufen durch Korrelationen einzelner oder mehrerer Prozessvariablen im Modell, können Vorzeichenwechsel in den Variablen bewirken (siehe Kapitel 9.2.3).

Die Prozessvariable Spanfeuchte ist in fast allen Modellen enthalten. Die Spanfeuchte und die Verdichtungszeit in der Presse zeigen ihre gegensätzliche Bedeutung für die Querzugfestigkeit bzw. Biegefestigkeit.

Die Rohplattendicke zeigt in allen Modellen die gleiche Bedeutung. Mit zunehmender Dicke steigt die Dickenquellung (2 Stunden) und fällt die Biegefestigkeit. Die Querzugfestigkeit wird in den Untersuchungen dagegen nicht von dieser Größe beeinflusst. Die Plattendicke kann als Ersatzgröße für die Rückfederung interpretiert werden. So lässt sich erklären, dass mit zunehmender Dicke, auch die Dickenquellung zunimmt, da die Platte mit zunehmender

Rückfederung für das Wasser zugänglicher ist. Gleiches gilt für die Biegefestigkeit. Warum die Querzugfestigkeit davon nicht beeinflusst wird, bleibt in allen Untersuchungen offen.

Die Rohplattenmasse zeigt ähnlich der Plattendicke als Einflussgröße die gleiche Konstanz über alle Modelle hinweg. So steigt mit zunehmender Plattenmasse die Biegefestigkeit und Querzugfestigkeit und sinkt die 2-Stunden-Dickenquellung. Eine höhere Plattenmasse steht bei gleicher Dicke für eine höhere Verdichtung der Platte. Damit ist die aufgezeigte Tendenz technologisch unwidersprüchlich. Diese Beziehung kann sich bei längerer Wasserlagerung wieder umkehren, da bei einer vollständigen Wasserdurchdrängung der Probe die Quellung mit zunehmender Quellmasse ansteigt.

Wenn Variablen zur Charakterisierung der Späne als auch des Klebstoffes zusätzlich offline gemessen worden sind, so sind diese auch als Variablen mit ins Modell aufgenommen worden. Dies zeigt deutlich die Notwendigkeit einer "online-Einbindung" dieser Größen zur Erhöhung der Modellgenauigkeit. Hier besteht weiterer Forschungsbedarf.

5.2 MDF-Modelle

Modell nach Stahl (1993)

	IB	MOR
Außentemperatur		+
Refiner-Stromstärke		-
Cooking level	+	
Temperatur Dampfbehälter	+	
Durchsatz		
Refinereinspeiseschnecke	-	
Druck auf Rotor Segmentscheibe	+	-
Druck auf Stator Segmentscheibe	+	-
Druck R 1	-	+*
Druck R 2	_*	-
Druck R 3		_*
Druck R 14/15l	+*	+
Druck R 16/17r	+*	
Akkumulator zw. R 12/13	+	
Akkumulator zw. R 16/17		-
Höhe Egalisierwalze Pedistor 2	+	
Temp. Vorerwärmer	_*	
Temp. Heizplatte Sektor 1		-
Temp. Heizplatte Sektor 2	+	+*
Temp. Heizplatte Sektor 3	-	-
Temp. Heizplatte Sektor 4		+
Distanz R 1/2r		-
Distanz R 5/6r		+*
Distanz R 8/9r		-
Distanz R 12/13r	+*	
Distanz R 14/15r	+	-
Distanz R 16/17r	-	+
Rohplattendicke Mitte	+	+*
Rohplattendicke Rechts	-	
Formlingstemperatur	+	+
Fixe Kontrollvariable	+	
Refiner Leistungsaufnahme	-	+
Beleimungsfaktor	-	+
Bandgeschwindigkeit	+*	-
Feuchte Formband	+	_*
Pedistor-Einstreumenge	+	+
Rohplattengewicht		+
	R ² =91%	R ² =84%

Versuchsdauer 3 Monate, Doppelbandpresse, 48 Beobachtungen, 3SLS-Schätzung, Parameterauswahl nach max. Bestimmtheitsmaß; * t-Wert ist kleiner als 1,8

Modell nach Janssen (1997)

	IB	IB	IB	TS_2h
	Modell 1	Modell 2	Modell 3	Modell 4
Spez. Druck R 9/10			#	#
Spez. Druck R 22			#	
Spez. Druck R 25/26	+	#		
Spez. Druck R 27		#		
Spez. Druck R 5		#	#	#
Distanz R19			#	#
Distanz Vorpresse		#		
Schüttdichte	-			#
Flächengewicht			#	
Faserfeuchte	-	#	#	
Position Mahlscheiben	-			
Stromaufnahme Stopfschnecke	-			
Dampfzugabe Hackschnitzelbunker	-?			
Dampfdruck Kocher			#	
Dampfdruck Refiner		#	#	
Dampf Mahlscheiben		#		
Hackschnitzelfeuchte	+	#		
Höhe der Fasermatte	+?		#	
Faseranteil grob	-			
Faseranteil mittel				#
Düse 1/2			#	#
Düse 3			#	
Temperatur Fasermatte			#	
Temperatur Rohleim				#
Drehzahl Austragsschnecke				#
Füllstand Hackschnitzelbunker				#
Geschwindigkeit Pressband			#	
Standardabweichung der Testdaten	0,052	0,052	0,052	0,069
RMSE _P im Vorhersageraum	0,036	0,046	0,054	0,076

Multiple univariate Regressionsrechnung, Beobachtungsraum: 43 Datensätze, Testdatenraum: 10 Datensätze, Variablenanzahl auf 42 reduziert, 19 mm Platten,

Modell 1: Variablenauswahl mit Forward-Stepwise-Verfahren, auf 9 Variablen festgelegt

Modell 2: Variablenauswahl nach größten F-Wert, auf 9 Variablen festgelegt

Modell 3: Variablenauswahl nach technologischen Gesichtspunkten

Modell 4: Variablenauswahl nach größten F-Wert, auf 9 Variablen festgelegt

Die bereits erwähnte Arbeit von Greubel (1999) umfasste auch Untersuchungen zur statistischen Modellbildung des MDF-Prozesses. Der Datenvektor bestand aus 45 Prozessgrößen inklusive der Produkteigenschaften. Es wurden unterschiedlich umfangreiche Modelle für 8 und 19 mm dicke Platten gebildet. Dabei wurden die in die Modellbildung einbezogenen Datensätze zwischen 21 und 61 variiert. Die Variablenauswahl erfolgte wieder nach technologischen Gesichtspunkten und nach dem visuellen Eindruck der Zeitreihendarstellung, wobei vor allem auf offensichtliche Ausreißer geachtet wurde. Angaben für die verwendeten Einflussgrößen finden sich nur für ein kleineres Modell. Dabei wurden folgende 8 Parameter verwendet:

Festharzanteil Spezifische Mahlenergie Schüttdichte Flächengewicht Presszeitfaktor Feuchte der Fasern auf Bandwaage Mittlerer Druck im Rahmen 4 (Hochdruckbereich) Mittlerer Druck in den Rahmen 9 und 10

Kennwerte über die Modellgüte sind bei Greubel (1999) nicht angegeben. Eine Validierung der Modelle erfolgte nicht.

Das MDF-Modell von Stahl (1993) weist Bestimmtheitsmaße von R²=0,91 für die Querzugfestigkeit und R²=0,84 für die Biegefestigkeit auf. Die Parameterauswahl erfolgte nach der statistischen Methode der Maximierung des Bestimmtheitsmaßes. Das Modell zeigt typische Vorzeichenwechsel in Variablen, die technologisch den gleichen Einfluss auf die Zieleigenschaft ausüben sollten. So weist die in der Plattenmitte gemessene Dicke einen entgegengesetzten Effekt auf die Querzugfestigkeit wie die rechts in Produktionsrichtung gemessene Plattendicke. Dieses Problem wird durch die Korrelation beider Messungen verursacht und ist technologisch nicht zu interpretieren. Des weiteren ist die Bandgeschwindigkeit im Modell der Querzugfestigkeit nicht signifikant, weist aber mit einem positiven Vorzeichen im Regressionskoeffizienten eine falsche technologische Beziehung auf. Eine Erhöhung des signifikanten Beleimungsfaktors führt nach dem Modell von Stahl zu einer Verringerung in der Querzugfestigkeit. Es kann vermutet werden, dass dieses Modell, ohne Berücksichtigung der Korrelationen zwischen den Prozessvariablen, zu stark an die Beobachtungsdaten angepasst worden ist und keine Vorhersagekraft aufweist. Eine Validierung des Modells erfolgte nicht.

Bei den Untersuchungen von Janssen (1997) wurden erstmals die Modelle an externen Testdaten validiert. Es wurde erkannt, dass die auf Basis einer beschränkten Anzahl von Beobachtungsdaten ermittelte Regressionsfunktion überprüft werden muss, ob sie auf die Grundgesamtheit übertragbar ist. Dazu wurden jeweils 10 Datensätze aus der Mitte der Beobachtungsperiode isoliert und nicht bei der Bildung der Modelle berücksichtigt. Dieser Vorhersageraum wurde anschließend dazu genutzt, die Modelle anhand dieser "fremden" Datensätze zu überprüfen. Dabei wurde festgestellt, dass mit zunehmendem Bestimmtheitsmaß der Modelle der Vorhersagefehler (RMSE_P) (vom Autor wird der Begriff Abweichungsmaß verwendet) im Testdatenraum größer wurde. Das bedeutet, dass Prozessmodelle mit hohen Bestimmtheitsmaßen die Beobachtungsdaten gut anpassen, aber keine Aussage über deren Prognosefähigkeit im Vorhersageraum zulassen.
5.3 OSB-Modell

Ein klassisches statistisches und umfassendes Prozessmodell einer OSB-Produktion ist in der Literatur nicht beschrieben. Barnes (2000) beschreibt ein integriertes Modell zur Vorhersage des Biegemoduls und der Biegefestigkeit als Produkt von Holzeigenschaften (F1) und Prozessparametern (F2). Dieses Modell beruht auf Gleichungen, entwickelt aus experimentellen Daten, in denen physikalische Beziehungen unterstellt werden. Die experimentellen Daten entstammen Versuchen mit Phenol-Formaldehydharz (Pulver). Die Holzeigenschaften der Strands (F1) Biegemodul, Biegefestigkeit und die Dichte der verwendeten Holzart und die Dichte der OSB-Platte werden als Eingangsparameter im Modell berücksichtigt. Als relevante Prozessparameter (F2) gehen in das Modell ein:

Beleimungsgrad (PF) Strandlänge Stranddicke Feinanteil Strandorientierung Falldistanz der Strands Scheibenabstand Verdichtung

Der Variationskoeffizient der Abweichung der berechneten Werte von den Messwerten lag für den Biegemodul bei 4,9% und für die Biegefestigkeit bei 7,9%. Interessant an dieser Arbeit ist, dass neben Prozessparametern Materialkennwerte der Strands berücksichtigt werden.

6 Methoden der Prozessmodellierung

6.1 Regressionsanalyse

6.1.1 Grundlagen

Mit der **Regressionsanalyse** werden Zusammenhänge zwischen Variablen erforscht. Ziel ist es eine Funktion zu finden, die bei Kenntnis der unabhängigen Variablen (Regressor, Einflussgröße, Prozessvariable) die Schätzung abhängigerer Variablen (Regressanden, Zielgröße, Platteneigenschaften) erlaubt.

Der **univariate Ansatz** einer Regression erklärt dabei eine einzelne Platteneigenschaft durch i.d.R. mehrere (multiple) Prozessgrößen.

Der **multivariate Ansatz** kann dagegen Dependenzen bzw. Interdependenzen zwischen verschiedenen Platteneigenschaften als auch Korrelationen zwischen den Störvariablen der einzelnen Gleichungen berücksichtigen. Korrelationen zwischen den verschiedenen Platteneigenschaften sind bei Holzwerkstoffplatten bekannt. Die gebildeten Mehrgleichungssysteme stellen eine Betrachtung der Platteneigenschaften in ihrer Gesamtheit dar.

Als erster wendete Lobenhoffer (1990) eine multivariate Regressionsanalyse an, um eine komplette Produktionsstraße einer Spanplattenfertigung zu beschreiben. Bei der Modellierung des Gesamtprozesses hat sich Lobenhoffer auf lineare Ansätze beschränkt. Dabei bedeutet linear, dass die Koeffizienten und Störvariablen linear in die Regressionsgleichung eingehen. Die erklärenden und erklärten Variablen können dabei das Ergebnis jeglicher Funktionen sein. Der Verzicht auf polynomiale Glieder, selbst zweiter Ordnung, wird mit der Anzahl der zusätzlich zu schätzenden Koeffizienten begründet. Des weiteren ist von vielen Einflussgrößen ein rein linearere Zusammenhang mit den Zielgrößen bekannt. Diese Tatsache ist aber oft entscheidend von dem zu betrachtenden Datenraum abhängig. Je enger ein Datenraum, desto weniger werden sich quadratische oder kubische Glieder auf das Ergebnis auswirken. Da es sich bei den Zielgrößen in der Holzwerkstoffherstellung um ausgeregelte Größen handelt, ist i.d.R. dieser Datenraum eng begrenzt. Damit ist der Verzicht auf polynomiale Glieder nachvollziehbar und gerechtfertigt.

Weitere Gründe für die Verwendung von linearen Regressionsmodellen sind nach Gruber (1994) die im linearen Modell theoretisch besser fundierten Verfahren als die im entsprechenden nichtlinearen Verfahren. Zudem wird die Operationalität des linearen Regressionsmodells von keinem anderen Modelltyp übertroffen.

Modelle setzen immer bestimmte Annahmen für ihre Gültigkeit voraus. Diese werden ausführlich im Kapitel 6.1.3 dargestellt. Diese Annahmen sind in aller Regel nur unter kontrollierten Bedingungen im Labor einzuhalten. In der Industrie werden die Annahmen u.a. dadurch verletzt, dass nur Variablen ins Modell aufgenommen werden können, die zum einen messbar sind und zum anderen eine genügend hohe Genauigkeit aufweisen. In der Praxis wird damit ein Modell nie vollständig bestimmt sein. In der Holzwerkstoffindustrie fallen als erstes die fehlenden Informationen der Holz- und Beleimungsqualität auf. Zukünftig werden diese Informationen verstärkt durch den Einsatz der NIR-Spektroskopie zur

Verfügung stehen. Trotz des zunehmenden Einsatzes von neuen Messtechniken werden die industriellen Datenmatrizen immer unvollständig und fehlerbehaftet sein. Dessen ungeachtet kann ein Regressionsmodell die Zielvariablen (Regressanden) gut vorhersagen, wenn das Modell sorgfältig aus einer großen Stichprobe erstellt wird. Die Interpretation der Statistik und die der Regressionskoeffizienten kann dagegen problematisch sein.

Der Beweis einer technologischen Kausalität bzw. eines Effektes auf Basis einer statistischen Signifikanz der Variablen in Regressionsmodellen ist nur innerhalb eines kontrollierten Experiments möglich. Im kontrollierten Experiment können die zu untersuchenden Einflussgrößen unabhängig voneinander fest eingestellt werden. In Modellen, die auf Basis von passiven Beobachtungen beruhen und damit keine festen einzustellenden, sondern zufällige Variablen beinhalten, unterstützt eine statistische Signifikanz einer Variable im Modell die Hypothese, dass ein Effekt bzw. ein kausaler Zusammenhang besteht. Eine zwingende Kausalität kann in diesen Modellen aus einer statistischen Beziehung nicht abgeleitet werden (Younger, 1979). Box (1966) weist ausdrücklich darauf hin, dass in Modellen, die auf passiven Beobachtungen beruhen, Zusammenhänge durch nicht messbare Variablen ("lurking" variables) überlagert sein können. Um sichere technologische Zusammenhänge zu erfassen, muss ein geplantes Experiment durchgeführt werden.

6.1.2 Gleichungssystem und Parameterschätzverfahren

Die strukturelle Form eines aus G Gleichungen bestehenden Systems für T Perioden lautet in Matrixform:

YΓ + XB +U =0.

Gleichung 1

- Y : Matrix der gemeinsam abhängigen Variablen, ist eine T \times G Matrix
- Γ : strukturelle Koeffizienten der gemeinsam abhängigen Variablen, ist eine G×G Matrix
- X : Matrix der K prädeterminierten Variablen (verzögert endogene, unverzögert oder verzögerte exogene Variablen), ist eine T×K Matrix
- B : Matrix der strukturellen Koeffizienten der prädeterminierten Variablen, ist eine K×G Matrix
- U : Matrix der Störvariablen in den G Gleichungen in T Perioden, ist eine T×G Matrix.

Die Form der Matrix Γ (diagonal, blockdiagonal, dreieckig, blockdreieckig, allgemein) ist unter ökonomischen Gesichtspunkten das wichtigste Kriterium für die Klassifikation verschiedener Typen struktureller ökonometrischer Modelle (Gruber, 1995).

Ein zweites Kriterium zur Klassifikation struktureller ökonometrischer Modelle ist die Form der Varianz-Kovarianz-Matrix des G×1 Störvariablenvektors u (t). Dieser Störvariablenvektor ist die t-te Zeile der T×G Matrix U. Bei Unterstellung von Homoskedastie der Störvariablen lassen sich 3 Formen der Varianz-Kovarianz-Matrix unterscheiden, die diagonale, blockdiagonale und die allgemeine Varianz-Kovarianz-Matrix.

Ein drittes Kriterium umfasst die "Autokorrelation der Störvariablen". Hierbei wird die i-te Spalte der T×G Matrix U betrachtet. Dabei wird zwischen der Kovarianzmatrix der Störvariablen einer strukturellen Gleichung und der Matrix der Kovarianzen zwischen den Störvariablen zweier Gleichungen unterschieden. Beide Varianz-Kovarianz-Matrizen können diagonal als auch nicht diagonal sein. Diagonale Varianz-Kovarianz-Matrizen sind ein Kennzeichen für "frei von Autokorrelation", nicht diagonale Matrizen dagegen sind ein Kennzeichen für Vorhandensein von "Autokorrelation". Beide Formen lassen sowohl Homoskedastie als auch Heteroskedastie der Störvariablen zu.

Es ist zu beachten, dass autoregressive Prozesse nur für Modelle auf Basis von Zeitreihendaten geeignet sind. Zeitreihen zeichnen sich darin aus, dass sie konstante Zeitabstände zwischen den Beobachtungen aufweisen. Dies ist in der Qualitätskontrolle mit ständig wechselnden Produktionen in der Holzwerkstoffindustrie nicht gegeben. Autoregressive Prozesse werden im folgenden nicht berücksichtigt. Nach Lobenhoffer (1990) können Autoregressionskoeffizienten verwendet werden, um die Schätzung zu verbessern. Eine direkte Interpretation ist dabei nicht möglich.

Da Abhängigkeiten zwischen den einzelnen Plattenqualitäten bestehen, kann auch davon ausgegangen werden, dass die Störvariablen der Gleichungen korreliert sind. Das Parameterschätzverfahren kann diese Korrelation der Störvariablen, in Form einer allgemeinen (Cross)-Varianz-Kovarianz-Matrix berücksichtigen. Das Seemingly-Unrelated-Regression-Verfahren (SUR, auch Zellner- oder joint generalized least squares (JGLS) - Verfahren) ist ein Mehrgleichungsschätzverfahren, dass die Gleichungsparameter gemeinsam unter Berücksichtigung der Varianz-Kovarianz-Matrix der Störgrößen berechnet. Dazu benötigt die SUR-Schätzung die Residuen einer ersten OLS-Schätzung. Mit diesen Residuen wird die (Cross)-Varianz-Kovarianz-Matrix berechnet. Zur effektiven Schätzung müssen die Gleichungen sich in den Variablen unterscheiden.

Sind zusätzlich Wechselwirkungen endogener Regressoren zu berücksichtigen, wird aus einer diagonalen eine allgemeine Matrix Γ. Als Parameterschätzverfahren wird hierfür das 3-SLS (3-Step-Least-Square)-Verfahren verwendet, welches neben der Korrelation der Störvariablen die Wechselwirkungen der endogenen Variablen untereinander berücksichtigt (Schönfeld 1971). Dabei werden im ersten Schritt die Werte der endogenen Regressoren geschätzt, im zweiten die Korrelationen der Störvariablen und erst im dritten Schritt die eigentlichen Regressionskoeffizienten. Es ist ein Parameterschätzverfahren für interdependente simultane stochastische lineare Gleichungen. Da das 3-SLS-Verfahren das allgemeinere Verfahren darstellt, kann dieses auch bei nicht verwendeten endogenen Regressoren benutzt werden. In diesem Fall kommt automatisch das SUR-Parameterschätzverfahren zur Anwendung. Es ist dann ein Parameterschätzverfahren für scheinbar unabhängige simultane stochastische lineare Gleichungen. Die statistisch erwünschten Eigenschaften der Koeffizienten dieser Schätzverfahren sind asymptotisch gegeben. Ab welcher Datenreihenlänge diese asymptotischen Eigenschaften zum Tragen kommen, ist im allgemeinen unbekannt (Gruber, 1981 und 1994).

Zur Vorhersage der gemeinsam abhängigen Zielvariablen muss die strukturelle Gleichung in die reduzierte Form (=Vorhersageform) überführt werden. Diese lautet dann:

Y=X×П + Vmit Π = B× Γ^{-1} Gleichung 2V= U× Γ^{-1}

Zur Modellbildung und Validierung wurde die Software SAS 8.2 genutzt. Zur Berechnung der multivariaten Modelle wurden die Prozeduren SYSLIN und SIMLIN verwendet.

6.1.3 Modellannahmen

In Abbildung 2 werden die wichtigsten Annahmen des klassischen linearen Modells der Normalregression und einige Verallgemeinerungen dargestellt. Diese Übersicht ist aus Gruber (1994) entnommen.

Annahmen über den	Annahmen über die	Annahmen über den	Annahmen über den	Annahmen über die
Störvariablenvektor	Regressormatrix	Regressions-	Regressandenvektor Y	gesamte Gleichung
U	x	koeffizientenvektor β	0	Y= X β + u
1	2	3	4	5
Annahme 1	Annahme 4	Annahme 9	Annahme 13 KA	Annahme 10
Die Störvariable hat	X hat den vollen Spaltenrang	Die Regressionsko-	Der Regressand ist eine	Die Regressions-
den Erwartungswert	(u.a. die Regressoren sind	effizienten sind feste	metrisch messbare	gleichung ist
Null.	frei von extremer	und in der Zeit bzw.	Variable, die keinen	identifiziert.
	Kollinearität).	im Querschnitt	besonderen Beschränk-	
		konstante Größen.	ungen unterliegt.	
			Verallgemeinerungen	
<u>Annahme 2</u>	Annahme 5		Annahme 13 OR	Annahme 11
$\sum u - \sigma^2 I$	Die Datenreihenlänge ist		Der Regressand ist	Das lineare
$\sum u = O_u I$	größer als die Anzahl der		ordinal skaliert.	Regressionsmodell
Die Störvariable hat	Regressoren.			ist korrekt
konstante Varianz				spezifiziert.
(=Homoskedastie)				
und ist frei von	Annahme 6 FN		Annahme 13 NO	Annahme 12
Autokorrelation.	Die Regressoren sind feste		Der Regressand ist	Die Datenmatrix
	(nichtstochastische) Großen.		nominal skallert.	enthalt die gesamten
Verallgemeinerung				zur Parameter-
$\sum u = \sigma^2 \Omega$			Annahma 12 DI	schatzung
$\sum u = u$			Annanme 13 DI	Informationan
Es liegt eine	Die Regressoren sind im		diabatama Variabla	inionnationen.
allgemeine				
Kovarianzmatrix vor.	Die Pegressoren sind im			
Annahme 3	Die Regressoren sind im Regressionsmodell endogen			
Die Störvariable ist	Annahme 7			
mehrdimensional	X ist frei von Fehlern in den			
normalverteilt.	Variablen			
	Annahme 8			
	X enthält alle wichtigen			
	Bestimmungsfaktoren des			
	Regressanden.			

Abbildung 2: Annahmen des klassischen linearen Modells der Normalregression und einige Verallgemeinerungen (Gruber, 1994)

Von einem verallgemeinerten linearen Regressionsmodell wird in der Regel dann gesprochen, wenn die Annahme 2, die Kovarianzmatrix ist diagonal und mit lauter gleich großen Elementen auf der Diagonale versehen (gleichbedeutend mit der Formulierung: die Störvariable ist homoskedastisch und frei von Autokorrelation), fallengelassen wird. Verallgemeinerte lineare Modelle zeichnen sich durch eine allgemeine Kovarianzmatrix aus.

Im Unterschied zum klassischen linearen Modell der Normalregression wird im verallgemeinerten linearen Modell weiterhin die Annahme, die Regressoren sind feste nichtstochastische Größen, aufgegeben. Stochastische Regressoren sind Zufallsvariablen

und haben damit eine Wahrscheinlichkeitsverteilung. Dies trifft ausnahmslos für die Prozessvariablen in der Holzwerkstoffherstellung zu. Drei Arten von stochastischen Regressoren sind zu unterscheiden:

- 1. exogene Regressoren
- 2. verzögert endogene Regressoren
- 3. unverzögert endogene Regressoren

Exogen sind Regressoren, wenn sie von der Störvariablen stochastisch unabhängig sind. Als erste Annäherung, (ist mathematisch aber nicht korrekt) sind exogene Variablen nicht mit der Störvariablen korreliert. Exogene Variablen werden außerhalb des Gleichungssystems erklärt. Bei Verwendung von exogenen und verzögert endogenen Regressoren sind noch Aussagen über erwünschte asymptotische Eigenschaften der Schätzer möglich. Das bedeutet, die Verwendung dieser Art von Regressoren erfordert noch kein verallgemeinertes Parameterschätzverfahren.

Ein Regressor ist endogen, wenn er in einer anderen Gleichung eines interdependenten Modells eine abhängige Variable (Regressand) ist oder abhängige Variable einer zusätzlichen Gleichung wäre, wenn alle vom Gleichungssystem beeinflussten Größen durch eine eigene Gleichung erklärt würden (Gruber, 1994). Die Nutzung endogener Regressoren in der Regressionsgleichung erfordert dagegen ein verallgemeinertes Parameterschätzverfahren, wie z.B. die 2-SLS-Schätzung, um eine Interdependenz-Verzerrung zu vermeiden und eine konsistente Parameterschätzung zu erreichen.

Alle weiteren in Abbildung 2 genannten Annahmen des klassischen linearen Regressionsmodells werden auch für die Mehrgleichungsmodelle vorausgesetzt. Es ist bei näherer Betrachtung dieser Annahmen ersichtlich, dass in der Praxis nicht alle diese Annahmen erfüllt werden können. So werden in aller Regel die Annahmen 7, 8, 11 und 12 implizit nicht erfüllt sein. Die Annahmen über die Störvariablen können während der Modellierung weitgehend überprüft werden. Auf die Auswirkungen der Verletzung verschiedener Annahmen wird im weiteren Text eingegangen.

6.1.4 Daten- und Prozessanalyse

Ausreißer-, Fehlstellen- sowie Plausibilitätsüberprüfungen der Daten sind Schritte der Datenanalyse. Methoden der Mathematik, Signalverarbeitung und der Statistik finden hier Anwendung. Das Bilden einfacher statistischer Kenngrößen wie Minimum, Maximum, Mittelwert, Variations- und Korrelationskoeffizient sowie grafische Darstellungen geben einen ersten Einblick in die Datensituation. Auf Grundlage der Datenanalyse wird entschieden, ob der jeweilige Datensatz bzw. Parametersatz zur Modellierung verwendet werden kann.

Die Prozessanalyse ist eine systematische Untersuchung des Prozessverhaltens mit dem Ziel der Identifizierung einer Struktur in den zur Verfügung stehenden Daten. Relevante Informationen, die Ursache-Wirkungs-Zusammenhänge abbilden, werden extrahiert. Modelle setzen Annahmen voraus, die mit verschiedenen Tests an den vorliegenden Daten auf ihre Richtigkeit zu überprüfen sind.

6.1.5 Modellbildung

Elementarer Schritt ist die Herausfilterung von Prozessvariablen, die signifikanten Einfluss auf das Verhalten des Prozesses besitzen. Dabei ist begrifflich statistisch signifikant von technologisch signifikant (bedeutsam) zu unterscheiden. Eine Variable kann statistisch auf Grund des vorliegenden Datenmaterials nicht signifikant, aber durchaus technologisch bedeutsam sein. Es ist nach Lobenhoffer (1990) aus fertigungstechnischer Sicht richtig, auch statistisch schwach gesicherte Regressoren in der Regression zu behalten, in sofern ihre Bedeutung technologisch gesichert ist.

Modellanwendung bedeutet im wissenschaftlichen Sinne die Validierung des Modells. Hierauf wird in dieser Untersuchung besonderen Wert gelegt. Die Größe des Vorhersagefehlers im Vorhersageraum (RMSE_P) ist ein entscheidendes Kriterium für die Güte des Modells. Es wird das Modell mit dem minimalsten und stabilsten Vorhersagefehler im Vergleich zu der Standardabweichung der Testdaten, gesucht. Die Schritte Modellbildung und Modellanwendung werden mehrmals durchlaufen.

6.1.6 Modellaktualisierung und ex-ante-Vorhersagen

Modellanwendung in der Praxis bedeutet neben der eigentlichen Vorhersage kurzfristig die Aktualisierung des Datenfiles und die Neuberechnung der Regressionskoeffizientenmatrix und langfristig die Überprüfung der Modellgültigkeit mit der bestehenden Variablenauswahl.

Standardmäßig wird, wenn nicht gesondert im Text angegeben, der Datensatz nach jeder ex-ante-Vorhersage aktualisiert und die Koeffizientenmatrix neu berechnet. Das bedeutet, dass der Freiheitsgrad des Fehlers mit fortschreitender Vorhersage ansteigt.

6.1.7 Variablenauswahlverfahren

Methoden zur Auswahl der Variablen für Regressionsmodelle sind in der Literatur ausreichend beschrieben (siehe u.a. Lobenhoffer 1990).

Im Programmpaket SAS 8.2. stehen verschiedenen Variablenauswahlverfahren zur Verfügung. Alle Verfahren beziehen sich auf univariate multiple Regressionsprobleme, multivariate Variablenauswahlverfahren stehen nicht zur Verfügung.

Das Programmpaket OPIC von Lobenhoffer und Ballüer beinhaltet ein Programm OPTSEL (optimal selection) zur multivariaten Auswahl von Variablen (2-SLS), welches Interdependenzen der endogenen Regressoren berücksichtigt. Das Auswahlkriterium bei OPTSEL ist die Summe der quadrierten Residuen. Derjenige Regressorsatz wird gewählt, bei dem diese Summe am geringsten ist. Durch die Angabe von minimaler und maximaler Regressorenanzahl, wird jeweils von der minimalen Anzahl beginnend ein Regressor hinzugefügt. Es wird die Anzahl an Regressoren zur Modellierung verwendet, die das größte Bestimmtheitsmaß und zugleich die minimalste Varianz des Gleichungsfehlers aufweisen. Dieses erlaubt nur eine maximale Auswahl aus 30 Variablen. Damit ist dieses Programm nicht in der Lage, eine suboptimale Auswahl unter einer größeren Anzahl an Regressoren zu treffen.

6.1.7.1 Stepwise-Selection

Die Stepwise-Prozedur beginnt mit einen Ein-Variablen Modell mit dem größten partialen F-Wert. Diese Variable ist zugleich am meisten mit dem Regressanden korreliert. Jede weitere Variable muss ein frei gewähltes oder voreingestelltes Signifikanzniveau unterschreiten, um in das Modell aufgenommen zu werden. Die Variable mit dem nächst größeren F-Wert bzw. kleinsten p-Wert wird in das Modell genommen. Nach Eintritt in das Modell werden alle im Modell vorhandenen Variablen wieder darauf hin geprüft, ob sie eine frei gewähltes oder voreingestelltes Signifikanzniveau überschreiten. Bei Überschreitung dieses Niveaus wird diese Variable wieder aus dem Modell heraus genommen, bevor die nächste Variable ins Modell genommen werden kann. Die Variablenauswahl wird abgebrochen, wenn keine der Variablen außerhalb des Modells einen p-Wert unterhalb der gewählten Grenze und keine der in das Modell aufgenommenen Variablen den p-Wert oberhalb der gewählten Grenze aufweist. In SAS sind für beide p-Werte 0,15 voreingestellt.

6.1.7.2 MAXR-Selection

Der Auswahlvorgang startet wie bei der Stepwise-Auswahl mit einem Ein-Variablenmodell. Es wird dabei diejenige Variable ausgewählt, die zu einem maximalen Bestimmtheitsmaß führt. Dieses Ein-Variablenmodell stimmt mit dem der Stepwise-Auswahl überein. Als zweite Variable wird diejenige ins Modell aufgenommen, die zum maximalen Anstieg des Bestimmtheitsmaßes führt. Danach wird jede im Modell enthaltene Variable überprüft, ob ein Austausch zu einer höheren Anpassung führt. Erst nach der Überprüfung aller Möglichkeiten, wird die Variable ausgetauscht. Dieser Austauschprozess wird solange fortgeführt, bis das "beste" Zwei-Variablenmodell mit dem höchsten Bestimmtheitsmaß gefunden wurde. Im nächsten Schritt wird erneut eine weitere Variable aufgenommen und der Vergleichs- und Austauschprozess durchlaufen. Dieser Auswahlprozess führt im Vergleich zur Stepwise-Prozedur bei gleicher Variablenanzahl zu höheren Bestimmtheitsmaßen.

6.1.7.3 Backward-Selection

Die grundlegenden Schritte der Backward-Methode sind:

- 1. Ein Regressionsmodel mit allen Variablen wird gebildet.
- 2. Der partiale F-Test wird für jede Regressor-Variable gebildet, unter der Annahme, sie sei die letzte ins Modell aufgenommene Variable.
- 3. Der geringste partiale F-Wert (F_P) wird mit einem gewählten oder voreingestellten Signifikanz-Level F_o bzw. p₀ (in SAS p=0,10) verglichen, wobei gilt:
 - a. Ist $F_P < F_o$, wird die dazugehörige Variable Z_p entfernt, die Regressionsgleichung erneut gebildet und Schritt 2 und 3 erneut durchlaufen,
 - b. Ist $F_P > F_o$, die Regressionsgleichung entspricht den Auswahlkriterium.

Die Auswahlmethoden befinden sich im Softwareprogramm SAS unter der Prozedur "Proc Reg". Die Ergebnisse dieser Auswahlprozeduren sind anschließend in die Prozeduren SYSLIN/SIMLIN übertragen worden, um eine Mehrgleichungsschätzung durchzuführen.

6.1.7.4 Optimale Eingänge

Das Programmpaket NN-Tool 2000 (Bärmann 2002) erlaubt es, aus einer großen Zahl an Eingangsparameter (=Regressoren) systematisch diejenigen auszuwählen, die zur Prognose der Ausgänge (=Regressanden) innerhalb von neuronalen Netzen besonders geeignet und dabei untereinander möglichst unkorreliert sind. Dabei wird für jeden Eingangsparameter eine Bewertung angegeben, wie wichtig dieser Parameter zur Prognose der ausgewählten Ausgangsparameter ist. Der Eingangsparameter mit der höchsten Bewertung ist für die Prognose am wichtigsten. In der Bewertung gehen drei Kriterien ein:

1. Ausgangskorrelation (AusCorr.): mittlere absolute Korrelation des betreffenden Parameters zu den Ausgängen. Die Ausgangskorrelation sollte möglichst hoch sein.

2. Eingangskorrelation (EinCorr.): maximale oder mittlere absolute Korrelation eines Eingangs zu denjenigen anderen Eingängen, die in der Bewertung bzgl. der anderen beiden Kriterien weiter oben stehen. Die Eingangskorrelation eines Eingangsparameters sollte möglichst niedrig sein. Sie geht deshalb negativ in die Gesamtbewertung ein.

3. Anzahl: Anzahl des Auftretens des Parameters in den Datensätzen. Ein geeigneter Eingangsparameter sollte in möglichst vielen Datensätzen vorkommen. Ein Vorkommen in allen Datensätzen ergibt eine Anzahlbewertung von 100, ein Vorkommen in 20% der Datensätze eine Anzahlbewertung von 20 (aus Handbuch NN-Tool 2000, Bärmann 2002).

Die Gewichtungen der einzelnen Kriterien kann frei gewählt werden. Folgende Gewichtung der Kriterien ist vorgenommen worden (Standard):

Ausgangskorrelation.:+1Eingangskorrelation (maximal):-1Anzahl Parameter:+1

Die Berücksichtigung der Korrelationen zwischen den Eingangsparametern stellt für korrelierte Prozessdaten einen großen Vorteil gegenüber den klassischen statistischen Auswahlmethoden dar.

6.2 Faktorenanalyse

6.2.1 Grundlagen

Das Ziel der Faktorenanalyse ist es, eine Vielzahl beobachtbarer Merkmale durch latente Größen, sogenannte Faktoren, möglichst gut zu erklären. Die Faktorenanalyse unterstellt, dass die Korrelation der Ausgangsdaten durch eine oder mehrere hinter den Variablen stehenden Faktoren beschreibbar sind. Diese latente Größen sind empirisch nicht messbar, sondern stellen das Ergebnis des faktoranalytischen Ansatzes dar (Hartung und Elpelt, 1984).

6.2.2 Faktorenmodell

Es wird in der Faktoranalyse davon ausgegangen, dass sich jeder standardisierte Merkmalswert $y_{st ij}$ als Linearkombination der Realisationen von q' standardisierten Faktoren F₁,..., F_{q'} darstellen lässt:

$$Y_{st} = \widetilde{F} \times \widetilde{L}'$$
 Gleichung 3

Gleichung 3 beschreibt den unterstellten linearen Zusammenhang zwischen den hypothetischen Faktoren und den Variablen der standardisierten Datenmatrix. Die Elemente der Matrix F heißen Faktorwerte des k-ten Faktors F_k beim i-ten beobachteten Objekt, die Matrix selbst wird die Faktorenwertematrix genannt. Die Faktorwerte beschreiben den Zusammenhang zwischen Faktoren und Objekten. Die Elemente der Matrix L heißen Faktorladungen des k-ten latenten Faktors F_k bzw. des j-ten standardisierten Merkmals Y_j . Die Matrix L wird Ladungsmatrix (factor pattern) genannt. Die Faktorladung ist eine Maßzahl für den Zusammenhang zwischen Merkmalen und Faktoren.

Sowohl die Matrix F als auch die Matrix L sind unbekannt und müssen geschätzt werden. Die Faktoren lassen sich in Einzelrestfaktoren (unique factor), gemeinsame Faktoren (common factor) und in allgemeine Faktoren (general factor) unterscheiden. Ist nur eine Ladung des kten Faktor F_k beträchtlich von Null verschieden, so erklärt F_k im wesentlichen nur eines der p Merkmale und heißt Einzelrestfaktor. Sind dagegen mehrere Ladungen des Faktors F_k von Null verschieden, so heißt F_k gemeinsamer Faktor. Sind alle Ladungen des Faktors F_k von Null verschieden, so handelt es sich um den speziellen gemeinsamen Faktor und heißt allgemeiner Faktor. Hauptsächlich interessieren bei der Faktorenanalyse die gemeinsamen Faktoren. Die Einzelrestfaktoren bilden in der gemeinsamen Faktoren unkorreliert sind. Das Modell der gemeinsamen Faktorenanalyse beschreibt die folgende Gleichung:

$$y_{ij} = f_{i1}l_{1j} + f_{i2}l_{2j} + f_{ik}l_{kj} + e_{ij}$$

Gleichung 4

- y_{ij} standardisierter Wert der i-ten Beobachtung und der j-ten Variable
- f_{ik} Faktorwert der i-ten Beobachtung des k-ten gemeinsamen Faktors
- l_{ki} Regressionskoeffizient des k-ten Faktors zur Vorhersage der j-ten Variable
- e_{ii} Wert der i-ten Beobachtung des j-ten Einzelrestfaktor (Residium)
- *k* Zahl der gemeinsamen Faktoren.

In der Matrixschreibweise lautet das Modell entsprechend:

Y = FL' + E

Gleichung 5

6.2.3 Modellannahmen

Der Ausgangspunkt der Berechnungen bildet die standardisierte Merkmalsmatrix, so dass jede ihrer Spalten den Mittelwert 0 und die empirischen Varianz von 1 besitzt. Diese Standardisierung dient der Vereinfachung der weiteren Berechnung der empirischen Korrelationsmatrix und aller weiteren Berechnungen.

Es wird unterstellt, dass der Zusammenhang zwischen den Faktoren linear ist und die Faktoren unkorreliert sind. Unkorrelierte Faktoren haben eine Einheitsmatrix als Korrelationsmatrix.

Das gemeinsame Faktorenmodell setzt voraus, das die Einzelrestfaktoren untereinander und zu den gemeinsamen Faktoren unkorreliert sind. In der Hauptkomponentenanalyse dagegen sind die Residuen generell miteinander korreliert.

Es wird weiterhin angenommen, dass jeder gemeinsame Faktor mindestens 2 Variablen beschreibt, anderenfalls handelt es sich um ein Einzelrestfaktor.

Die einzelnen Verfahren zur Schätzung der Ladungsmatrizen formulieren stets eigene zusätzliche Forderungen an die Ladungsmatrix, die dann eine eindeutige Schätzung dessen erlauben.

6.2.4 Ladungsmatrix

Nach dem "Fundamentaltheorem der Faktorenanalyse" (Thurstone, 1947), ist die Korrelationsmatrix R der Merkmalswerte durch die Ladungsmatrix L und die Korrelationsmatrix R_F der Faktoren reproduzierbar. Sind die Faktoren orthogonal, so ist deren Korrelationsmatrix die Einheitsmatrix I.

Für gemeinsame orthogonale Faktoren lässt sich mit Hilfe der Ladungsmatrix L eine reduzierte empirische Korrelationsmatrix \widetilde{R} berechnen.

$$\widetilde{R} = R - U^2 = LL'$$

Gleichung 6

Die Diagonalwerte der reduzierten empirischen Korrelationsmatrix werden Kommunalität des j-ten Merkmals genannt. Die Kommunalität gibt an, welcher Anteil der Varianz des j-ten standardisierten Merkmals durch gemeinsame Faktoren erklärt wird. Die Elemente der Matrix U² sind dagegen die sogenannten merkmalseigenen Varianzen, d.h. die Varianzanteile der Merkmale, die sich nicht durch gemeinsame Faktoren erklären lassen. Die Ladungsmatrix lässt sich nur berechnen, wenn U bekannt ist. Die Bestimmung von U ist analytisch nicht durchführbar, sondern nur mittels numerischer Iterationsverfahren. Als Startwerte für die Schätzung der Kommunalitäten werden bei der Hauptfaktorenanalyse u.a. die multiplen Bestimmtheitsmaße der jeweiligen Variable zu allen anderen Variablen verwendet (in SAS: priors=SMC). Sind die Kommunalitäten groß bzw. die merkmalseigenen Varianzen klein, so unterscheiden sich die Ergebnisse der Hauptfaktorenanalyse wenig von denen der Hauptkomponentenanalyse.

Zur Schätzung einer Ladungsmatrix existieren verschiedene Verfahren, die durch zusätzliche Annahmen an die Ladungsmatrix zu einer eindeutigen Lösung gelangen. Genannt seien hier die Lösungen durch die Maximum-Likelihood-Methode, kanonische Faktorenanalyse, Zentroid- und Jöreskog-Methode bzw. die Alpha- und die Minimax-Faktorenanalyse. Im Ergebnisteil angewendet werden die Hauptfaktoren- bzw. die Hauptkomponentenmethode. Letztere ist im eigentlichen Sinne keine Faktorenanalyse, da hier nicht von der reduzierten sondern von der empirischen Korrelationsmatrix ausgegangen wird. Die Hauptfaktorenmethode ist eine faktoranalytische Abwandlung der Hauptkomponentenmethode.

Die Hauptkomponentenanalyse (principal components analysis, PCA) arbeitet mit der ursprünglichen empirischen Korrelationsmatrix R, d.h., es verzichtet auf die Annahme merkmalseigener Varianzen. Die Kommunalität wird gleich eins gesetzt.

Das Prinzip der Hauptkomponentenanalyse liegt in einer Verschiebung des Nullpunktes in den Schwerpunkt der Datenwolke und in einer Verschiebung der Koordinatenachsen. Im Fall einer $n \times 2$ Datenmatrix erfolgt die Drehung der Achsen dermaßen, dass die erste Hauptachse (Hauptkomponente) ein Maximum der Gesamtvarianz und die zweite Komponente ein Maximum der Restvarianz ausschöpft. Diese Transformation des Koordinatensystems entspricht einem Eigenwertproblem. Dabei ist der Ladungsvektor I_k der k-ten Hauptkomponente der Eigenvektor zum k-größten Eigenwert λ_k der Matrix R, normiert auf die Länge λ_k . Da unendlich viele Eigenvektoren zu einem Eigenwert existieren, ist die Ladungsmatrix nur eine von unendlich vielen möglichen Matrizen. Dabei ist es in den meisten Fällen schwer möglich zu einer Interpretation der Faktoren zu gelangen.

Die Hauptfaktorenanalyse (principal factor analysis) geht dagegen von merkmalseigenen Varianzen aus. Daraus leitet sich die reduzierte Korrelationsmatrix ab (Gleichung 6). Zur Berechnung der Ladungsmatrix wird eine Hauptkomponentenanalyse für die reduzierte Korrelationsmatrix durchgeführt.

6.2.5 Rotationsverfahren

Anhand einer rotierten Ladungsmatrix lässt sich veranschaulichen, welche ursprüngliche Variablen durch die Faktoren beschrieben werden. Diese Zuordnung ist nicht immer eindeutig möglich. Die bestmöglichste Interpretation entscheidet letztendlich über die Auswahl des Rotationsverfahren.

Durch eine Faktorrotation, bei der die ursprünglichen Faktoren durch eine nichtlineare Transformation in neue Faktoren verwandelt werden, soll eine optimale Interpretation erreicht werden. Bei der Rotation der Faktoren wird zwischen orthogonaler und schiefwinkliger Rotation unterschieden. Werden alle Faktoren um den gleichen Winkel gedreht, bleibt die Orthogonalität der Faktoren erhalten, dreht man sie um unterschiedliche Winkel, so geht die Orthogonalität verloren und es wird von einer schiefwinkligen Rotation gesprochen.

Eine optimale Interpretierbarkeit wurde von Thurstone (1947) mit dem Begriff der Einfachstruktur einer Ladungsmatrix in Verbindung gebracht. Da dieser Begriff sehr komplex ist und nicht durch eine einzige Maßzahl beschrieben werden kann, haben sich mehrere Rotationsverfahren herausgebildet.

Häufig verwendete orthogonale Rotationsverfahren sind die Equamax-, Varimax- und die Quartimax-Methoden. Das Quartimaxverfahren liefert die beste Approximation an eine Einfaktor-Lösung, d.h., dass durch einen Faktor ein Maximum an Informationen der ursprünglichen Variablen erklärt werden soll. Die Equamax-Methode versucht die erklärte Varianz möglichst gleichmäßig auf die rotierten Faktoren zu verteilen. Die Varimax-Methode dagegen verteilt die Varianzen maximal auf die ersten rotierten Faktoren.

6.2.6 Faktorenwertematrix

Die Faktorenwertematrix ist eine $n \times k$ Datenmatrix für die n Objekte (hier Datensätze). Diese Matrix ist gegenüber der Ausgangsdatenmatrix (n= 1...p) eine um p-k reduzierte Matrix, die die n beobachteten Objekte in Abhängigkeit von den k extrahierten Faktoren beschreibt. Es besteht damit die Möglichkeit eine große und zugleich stark korrelierte Datenmatrix mit einigen wenigen und unkorrelierten Faktoren zu beschreiben. Auf Basis dieser Faktoren ist eine Regressionsanalyse unproblematischer durchzuführen als auf Basis der Ausgangsdatenmatrix, die zu Multikollinearitätsproblemen führen kann. Verfahren zur Berechnung der Faktorenwertematrix sind in Hartung und Elpelt (1984) beschrieben.

6.3 PLS - Partial Least Square

6.3.1 Grundlagen

Die Kleinste-Quadrate-Schätzung der klassischen Regressionsanalyse minimiert die Summe der Quadrate der Störvariablen. Das Ziel liegt im auffinden von linearen Funktionen, die soviel Variationen wie möglich in den Zielvariablen erklären können. Die PLS-Technik dagegen verfolgt das Ziel, so viel Variation wie möglich sowohl in den Regressoren als auch in den Zielvariablen zu erklären. PLS-Modelle modellieren die "Struktur" sowohl der X- als auch der Y-Matrix. Im Gegensatz zur Regressionsanalyse wird davon ausgegangen, dass die Regressoren miteinander korreliert sind. Die PLS-Technik ist die konsequente Erweiterung der Faktorenanalyse durch Einbeziehung des y-Variablen-Blocks. Latente Variablen beschreiben in der Faktorenanalyse den x-Raum, in PLS-Modellen dagegen sowohl den x- als auch den y-Raum.

Wold et al. (2001) bezeichnen die PLS-Regression als eine Verallgemeinerung der multiplen linearen Regression (MLR). Im Spezialfall einer diagonalen X-Matrix (X-Variablen zeigen keine Korrelationen) und einer einzelnen Y-Variable, sind die PLS- und die MLR-Regressionskoeffizienten identisch.

Klassische Regressionsverfahren und deren Verallgemeinerungen (2-SLS; 3-SLS) setzten voraus, dass die Regressoren frei von extremer Kollinearität sind, die X-Matrix muss den vollen Spaltenrang haben. Mit zunehmenden Einsatz von Messinstrumenten zur Überwachung bzw. Regelung des Produktionsprozesses werden die X-Matrizen größer und die Variablen stärker kollinear. Die Notwendigkeit Datenmatrizen zu reduzieren, um sie regressionstechnisch verarbeiten zu können, wird in der Zukunft stark ansteigen. Projektionsverfahren bieten die Möglichkeit der Reduzierung einer Datenmatrix zu wenigen latenten Variablen.

Im vorliegenden Fall der Prozessmodellierung sind die x-Variablen alle bekannt und interpretierbar. Die x-Variablen im gegenwärtigen Holzwerkstoffprozess stammen aus verschiedenen Messsensoren im Produktionsprozess. Die Interpretation der latenten Variablen ist soweit irrelevant, da diese aufgrund der bekannten x-Variablen eindeutig zurücktransferiert werden können. Die latenten Variablen fassen Gruppen von x-Variablen zusammen, die mit der bzw. den Zielgrößen korreliert sind. Diese Zusammenfassung korrelierter Variablen kann als ein Zwischenschritt verstanden werden, der die Probleme der Multikollinearität in der Regressionsanalyse überwindet.

6.3.2 PLSR-Modell

Vor der eigentlichen PLS-Analyse werden die X- und Y-Variablen skaliert und zentriert. Die Skalierung hat das Ziel, die unterschiedlichen Wertebereiche der Variablen auszugleichen bzw. Variablen durch geeignete Skalierungsmethoden ein höheres Gewicht in der Analyse zu verschaffen. Liegt kein Vorwissen über die relative Bedeutung der Variablen vor, werden die Variablen auf eine einheitliche Varianz skaliert. Dabei wird jede Variable durch die entsprechende Standardabweichung dividiert. Damit wird jeder Variable das gleiche Gewicht bzw. Bedeutung in der Analyse verliehen. Die Zentrierung der Variablen stellt dagegen sicher, dass das Kriterium zur Auswahl der Faktoren darauf basiert, wieviel Variation diese erklären. Zur Zentrierung wird jeweils der Mittelwert von jeder Variablen subtrahiert. Dieses Vorgehen nennt sich Auto-Skalierung und stellt sicher, dass jeder Variable den Mittelwert 0 und die Standardabweichung 1 besitzt. Diese Transformierung ist aus der Faktorenanalyse bekannt und wird in dieser Arbeit standardmäßig angewendet.

Die Erläuterungen und die verwendeten Bezeichnungen zum PLSR-Modell basieren auf Wold et al. (2001).

- A Anzahl der Komponenten im PLS-Modell, Index: a=1,2,...A
- M Anzahl der Y-Variablen, Index: m=1,2,...M
- K Anzahl der X-Variablen, Index: k=1,2,....K
- N Anzahl der Beobachtungen, Index: i=1,2,...N
- X Matrix der Einflussgrößen, eine N×K Matrix
- Y Matrix der Zielgrößen, eine N×M Matrix
- **b**_m Vektor der Regressionskoeffizienten des m-ten y, ein K×1 Vektor
- **B** Matrix der Regressionskoeffizienten aller Y's, eine K×M Matrix
- **C** PLSR Y-Gewichtsmatrix mit \mathbf{c}_a als Spalten, eine M×A Matrix
- **E** Matrix der X-Residuen, eine N×K Matrix
- F Matrix der Y-Residuen, eine N×M Matrix
- P Ladungsmatrix mit **p**_a als Spalten
- **T** Score Matrix mit \mathbf{t}_a als Spalten, eine N×A Matrix
- U Y-Score Matrix mit u_a als Spalten, eine N×A Matrix
- **W** Matrix der X-Gewichte mit \mathbf{w}_a als Spalten, eine K×A Matrix
- **W**^{*} modifizierte Matrix der X-Gewichte mit \mathbf{w}_{a}^{*} als Spalten, eine K×A Matrix

Die durch das PLSR-Modell extrahierten "neuen" Variablen werden X-Scores genannt und mit \mathbf{t}_a (a=1,2,...A) bezeichnet. Die X-Scores sind die neuen Regressoren zur Schätzung von Y und modellieren gleichzeitig X. Die X-Scores sind, wie die Faktoren in der PCA-Analyse orthogonal. Sie werden berechnet als Linearkombinationen aus den originalen Variablen x_k und den Gewichten w^*_{ka} (a=1,2,...A). Gleichung 7 zeigt die Matrixschreibweise zur Berechnung der X-Scores.

Die PLSR-Methode schätzt sowohl X als auch Y durch eine Regression mit den X-Scores. Die Residuen des X-Prozessmodells sind e_{ik} und die des Y-Prozessmodells f_{im} . Die X-Scores multipliziert mit den Ladungen p_{ak} bzw. mit den Gewichten c_{am} ergeben X bzw. Y (Gleichung 8 und Gleichung 9).

Die Residuen f_{im} drücken die Abweichung zwischen der beobachteten und der modellierten Zielgröße aus. Diese Residuen sind sowohl im Lerndatenraum als auch im Vorhersageraum (Testdaten) mit denen der multivariaten Regressionsmodellen (MVR-Modelle) vergleichbar.

Bei multivariaten Zielgrößen Y (M>1) ergeben die Y-Scores \mathbf{u}_{a} multipliziert mit den Gewichten c_{am} eine gute Zusammenfassung der Y-Variablen. Die Residuen dieses Modells werden mit g_{im} bezeichnet (Gleichung 10). Modelle, die multivariate Zielgrößen berücksichtigen, werden PLS2-Modelle genannt. PLS1-Modelle berücksichtigen dagegen nur eine Zielgröße. Algorithmen zur Lösung dieser beider Modelltypen finden sich in Manne (1987).

Wird in Gleichung 9 die Score Matrix T mit der Beziehung aus Gleichung 7 ersetzt, ergibt sich Gleichung 11. Diese Beziehung spiegelt ein Regressionsmodell wieder, wobei b_{mk} (**B**) die PLS-Regressionskoefizienten darstellen.

Y= **XW*** **C**' + **F** = **XB** +**F** mit **B** = **W*****C**'

Die PLS-Regressionskoeffizzienten sind nicht unabhängig voneinander, es sei denn die Anzahl der Komponenten A entspricht der Anzahl K an X-Variablen. Die Größe und das Vorzeichen des skalierten und zentrierten PLS-Regressionskoeffizienten zeigt den Einfluss der Prozessvariable auf das Modell. Unterschiede in den Vorzeichen zwischen MLR- und PLS-Regressionskoeffizienten bei stark korrelierten X-Variablen werden in Wold et al. (2001) anhand eines Beispiels aufgezeigt. Die MLR-Regressionskoeffizienten sind in diesem Fall irreführend und nicht interpretierbar.

Nach jeder Komponente a wird die X-Matrix durch Subtraktion des Produktes $\mathbf{t}_{a} \, \mathbf{p}_{a}$, "deflationiert". Die modifizierten Gewichte w* kombinieren dabei die ursprünglichen x-

Gleichung 11

Gleichung 10

Gleichung 8

Gleichung 9

X = TP' + E

Y = TC' + F

Variablen, um die Scores t zu bestimmen. Die Gewichte w dagegen kombinieren die X-Residuen, um die Scores t zu formen. Die Beziehung zwischen den Gewichten zeigt Gleichung 12 (Manne, 1987).

$W^* = W(P'W)^{-1}$

Gleichung 12

Die erste PLS-Komponente ist geometrisch interpretiert eine Gerade im x-Raum, t_1 und eine Gerade, u_1 , im y-Raum. Diese Geraden werden dabei so in das Koordinatensystem gelegt, dass diese eine maximale Kovarianz aufweisen. Des weiteren bilden diese Graden eine gute Approximation des Punkteschwarms sowohl im x- als auch im y-Raum. Die Projektionskoordinaten der beiden Räume sind damit durch eine innere Relation verbunden und miteinander korreliert. Die zweite PLS-Komponente wird durch Geraden im x- und y-Raum vertreten, die senkrecht auf der ersten Komponente stehen. Dabei sind die Vektoren w_a als auch t_a jeweils orthogonal zueinander, aber nicht die Ladungsvektoren p_a und nicht die Y-Scores u_a . Die zweite PLS-Komponente verbessert die Approximation an die Punktschwärme so gut wie möglich. Die Projektionskoordinaten t_2 und u_2 sind jedoch weniger stark miteinander korreliert, als das erste Paar latenter Variablen. Mit jeder weiteren PLS-Komponente nimmt die Korrelation zwischen den x-und y-Scores weiter ab.

Alle Parameter **t**, **u**, **w**, (**w***), **p**, und **c** wurden durch den NIPALS-Algorithmus von Wold et al. (1984) berechnet. In der Software SAS 8.2. ist dies innerhalb der Prozedur "Proc PLS" der Standardalgorithmus.

6.3.3 Modellannahmen

Die Hauptannahme der PLS-Modelle besteht darin, dass das untersuchte System nur durch wenige zugrundeliegenden, latenten Variablen beeinflusst wird. Die Anzahl an latenten Variablen ist gewöhnlich nicht bekannt und deren Bestimmung ein Ziel der PLS-Analyse. Es wird angenommen, dass sowohl die x- als auch die y-Variablen Realisationen dieser latenten Variablen sind. Die tatsächliche Anzahl an latenten Variablen ist jedoch abhängig von der vollständigen Erfassung der Realisationen der zugrunde liegenden Prozesse. Als Realisationen der latenten Variablen werden Korrelation zwischen den Variablen unterstellt. In den meisten praktischen Situationen sind Korrelationen zwischen den Variablen zu beobachten. Mit korrelierten Variablen kann jedoch nicht ein "korrekter" Wert einem individuellen Regressionskoeffizienten zugewiesen werden. Es wird der gemeinsame Einfluss der Variablen auf Y geschätzt.

Die Annahme, die Datenlänge ist größer als die Anzahl an Regressoren (Variablen) ist in der PLS-Analyse aufgehoben.

Aufgrund der Modellierung des X-Prozesses muss auch nicht mehr vorausgesetzt werden, dass die x-Variablen fehlerfrei sind. Es wird dagegen angenommen, dass ein Teil der x-Variablen nicht in Beziehung zu Y stehen. Dieser Teil kann ein "Rauschen" durch Messfehler bzw. im vorliegenden Fall durch eine ungenaue zeitliche Datenzuordnung beinhalten. Das "Rauschen" wird durch die X-Residuen-Matrix E gekennzeichnet.

7 Kennzahlen der Modellgüte

Kennzahlen zur Überprüfung der Modellgüte werden im Folgenden eindeutig in zwei Bereiche unterschieden, einmal in Kennzahlen, die die Anpassung im Beobachtungsraum, und einmal im echten Vorhersageraum beurteilen. Erstere sind Kennzahlen, die die reine Anpassung des Modells an die zur Modellerstellung berücksichtigten Beobachtungsdaten letztere sind Kennzahlen der eigentlichen Validierung ausdrücken. im echten Vorhersageraum. In der Literatur werden diese beiden Bereiche oft nicht eindeutig voneinander getrennt bzw. aufgrund der schlechten Datensituation (zu wenig Datensätze) die Modelle ausschließlich mit Kennzahlen aus dem Beobachtungsraum beurteilt. Diese Beurteilung fällt immer zu optimistisch aus (Janssen 1997). Im Zusammenhang von Prognose-Kennwerten wird im folgendem immer eindeutig die ex-post-Prognose (=Schätzung der Zielvariablen im Beobachtungsraum) von der ex-ante-Prognose (=Schätzung der Zielvariablen in der Zukunft) unterschieden.

7.1 Kennwerte der ex-post-Prognose

Im Laufe der Regressionsanalyse muss in den meisten Anwendungsfällen aus einer Vielzahl von Gleichungsvarianten, die sich vor allem in den Regressoren unterscheiden, eine brauchbare Regressionsfunktion ausgewählt werden. Es gibt kein Kriterium, nach dem eine eindeutige Auswahl erfolgen kann. Je nach Anwendungsgebiet und Zielvorstellungen müssen verschiedene Kriterien herangezogen werden.

Die im Folgenden aufgeführten Kennzahlen, Erläuterungen und Formeln entstammen aus Gruber (1993), Draper und Smith (1988), Farnum et al. (1992) und der SAS-Dokumentation zur Prozedur SYSLIN/SIMLIN Version 8.2.

Summe der Störvariablen ut

Bei einer effizienten Schätzung ist die Summe der einfachen Fehler gleich Null. Damit gestattet diese Kennzahl eine einfache Kontrollmöglichkeit, ob bei der Berechnung der Schätzwerte von y und u Fehler gemacht wurden.

Gleichung 13

$$\hat{u} = y - \hat{y}$$
$$\sum_{i=1}^{T} \hat{u}_i = 0$$

Mittelwert der Störvariablen ut (Mean Error)

Diese Kennzahl ist das einfache arithmetische Mittel der Störvariablen u_t. Diese sollte nahe bei Null liegen. Auf Grund von mathematischen Operationen bzw. Rundungsfehlern kann dieser Wert leicht von der Null abweichen.

$$\overline{\hat{u}} = \frac{1}{n} \sum (y - \hat{y})$$
 Gleichung 14

Mittelwert des Betrages der Störvariablen | ut | (Mean absolut deviation = MAD)

Diese Kennzahl ist das einfache arithmetische Mittel der betragsmäßigen Störvariablen ut.

$$MAD = \frac{1}{n} \sum \left| y - \hat{y} \right|$$
 Gleichung 15

Mittlerer absoluter prozentualer Fehler (Mean absolute percent error = MAPE)

Diese Kennzahl ermöglicht ein Vergleich, unabhängig von der Einheit, zwischen den verschiedenen Zielvariablen. Der prozentuale Fehler sollte möglichst gering sein.

$$MAPE = \frac{\sum |\hat{u}_t / y_t|}{n} \times 100\% \qquad y_t \neq 0$$
 Gleichung 16

Summe der Quadrate der Störvariablen ut

т

 σ

Die Summe der Quadrate der Störvariablen ist die Minimierungsfunktion der Kleinsten-Quadrate-Schätzung. Diese Kennzahl wird in der Literatur auch Summe der Fehlerquadrate (SFQ) bzw. Sum of Squares Errors (SSE), Error Sum of Squares (ESS) bzw. Residual Sum of Squares (RSS) oder SS about the Regression) genannt. Da diese Kennzahl keine obere Grenze besitzt, ist der Vergleich bei unterschiedlicher Periodenzahl T (Datensatzanzahl) nicht möglich. Dieser Nachteil wird durch das Bestimmtheitsmaß vermieden.

$$SFQ = \sum_{i=1}^{I} (y_i - \hat{y}_i)^2$$
 Gleichung 17

Varianz der Störvariablen (Means Square Error of Model = MSE Model)

Die geschätzte Varianz der Störvariablen u_t in der Regressionsgleichung ist eine sehr einfache Kennzahl des Regressionsmodells. Sie soll möglichst klein sein.

$$\sigma_u^2 = \frac{SFQ}{T-K}$$
 Gleichung 18

Standardabweichung der Störvariablen des Modells (RMSE Model)

Diese Kennzahl ist die Schätzung der Standardabweichung der Störvariablen in der Regressionsgleichung. Sie kennzeichnet eine Maßzahl für den Fehler, mit dem jedes beobachtete Y vorhergesagt werden kann, wenn dafür die entwickelte Regressionsgleichung herangezogen wird.

$$_{u} = \sqrt{\frac{SFQ}{T-K}}$$
 Gleichung 19

In der Literatur sind die drei letztgenannten Kennzahlen nicht immer eindeutig benannt. In Schweitzer (1992), Brade (1986), Andres (1989) und Schubert (1990) wird für die Kennzahl RMSE der Begriff Reststandardabweichung verwendet und mit S_R abgekürzt. In der

englischen Literatur wird unter dieser Abkürzung dagegen nur die Summe der Fehlerquadrate (SFQ) verstanden (siehe Box und Hunter, 1978).

Das Verhältnis aus der empirischen Standardabweichung der Lerndaden und der geschätzten Standardabweichung des Modells (RMSE) erlaubt es, den Fehler und damit die Schätzgenauigkeit im Beobachtungsraum anderen Modellgleichungen gegenüberzustellen. Diese genormte Kennzahl erlaubt damit die Schätzgenauigkeit im Beobachtungsraum zwischen den einzelnen Eigenschaften mit verschiedenen physikalischen Einheiten zu vergleichen. Des weiteren können mit dieser Kennzahl die Modellergebnisse verschiedener Anlagen bzw. Holzwerkstoffhersteller verglichen werden.

Es sei an dieser Stelle darauf hingewiesen, dass die SAS-Prozedur SIMLIN unter "Statistics of Fit" den RMSE abweichend von der oben aufgeführten Formel berechnet. Als Nenner wird die Anzahl der Beobachtungen n anstatt des Freiheitsgrades T-K verwendet. Die exakte Berechnung des RMSE findet sich unter der Kennzahl "Sigma". Diese wird durch den Befehl "outest" ausgegeben. Im Ergebnisteil werden beide Berechnungen aufgeführt.

Root mean square percent error (RMSPE)

Diese Kennzahl ist die Schätzung des Variationskoeffizienten des Fehlers in der Regressionsgleichung. Auch hier verwendet die Prozedur SIMLIN die Anzahl der Beobachtungen nals Divisor.

$$RMSPE = \sqrt{\frac{100}{n} \sum_{t=1}^{n} ((y_t - \hat{y}_t) / y_t)^2}$$
 Gleichung 20

Das Bestimmtheitsmaß R²

Das Bestimmtheitsmaß R² ist das am häufigsten zur Beurteilung der Güte der Anpassung der empirischen Regressionsfunktion an das Datenmaterial verwendete Kriterium. Das Bestimmtheitsmaß gibt den Anteil der durch die Regression erklärten Varianz an. Anders ausgedrückt erklärt es, zu welchem Anteil die Streuung der Zielgröße durch die Streuungen der Einflussgrößen erklärt werden kann. Die Werte können zwischen 0 und 1 liegen, wobei Werte nahe 1 eine gute Anpassung der Beobachtungen durch die Regressionsgleichung bedeutet.

Das Bestimmtheitsmaß berechnet sich nach:

$$R^{2} = 1 - \frac{SFQ}{SGQ} \text{ oder } R^{2} = \frac{SRQ}{SGQ}$$

$$SRQ = \sum_{i=1}^{T} (\hat{y}_{i} - \overline{y})^{2}$$

SRQ = Summe Regressionsquadrate (im Englischen: SSR oder SS due to regression)

$$SGQ = \sum_{i=1}^{T} (y_{i} - \overline{y})^{2}$$

$$\overline{i}$$

SGQ = Summe der Gesamtquadrate (im Englischen: SST oder TSS oder SS about mean)

$$SFQ = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2$$

SFQ = Summe der Fehlerquadrate (im Englischen: Sum of Squares Errors (SSE) oder Error Sum of Squares (ESS) bzw. Residual Sum of Squares (RSS) oder SS about the Regression)

Das Bestimmtheitsmaß ist im strengen Sinn nur für den Vergleich von Modellen mit gleicher Anzahl von Regressoren verwendbar, da das Bestimmtheitsmaß von K+1-Regressoren stets größer ist als das von K-Regressoren. Dieser Nachteil wird vom korrigierten Bestimmtheitsmaß vermieden, da dieses den Verlust des Freiheitsgrades durch Aufnahme eines weiteren Regressors berücksichtigt.

Bei Schätzmethoden, die die Verwendung von endogenen Regressoren berücksichtigen (2-SLS, 3-SLS), werden die endogenen Regressoren durch die OLS-Schätzwerte ersetzt. Dies hat zur Folge, dass sich die SRQ und die SFQ nicht mehr zu den SGQ aufaddieren. Für diesen Fall gibt es verschiedene Möglichkeiten das Bestimmtheitsmaß zu berechnen. Die SYSLIN-Procedur verwendet folgende Definition:

$$R^{2} = \frac{SRQ}{SRQ + SFQ}$$
 Gleichung 22

Diese Definition ist mit dem F-Test, dass die wahren Regressionskoeffizienten gleich Null sind, konsistent. Jedoch ist dieses R² kein gutes Anpassungsmaß für die Güte des Modells (SAS Online Doc 8.0. S.1087).

Korrigierte Bestimmtheitsmaße

Zum Vergleich von Modellen mit unterschiedlicher Anzahl an Regressoren ist das korrigierte Bestimmtheitsmaß als Gütemaß dem einfachen Bestimmtheitsmaß vorzuziehen. Das korrigierte Bestimmtheitsmaß berücksichtigt den Verlust des Freiheitsgrades mit jedem zusätzlichen Regressor. Für t- und F-Tests und für die Berechnung von Konfidenz- und Prognoseintervallen ist es aber vorteilhaft, möglichst viele Freiheitsgrade zu haben. Daher kann ein zusätzlicher Regressor in statistischer Hinsicht nachteilig sein (Gruber 1993).

Es gibt verschiedene berechnete korrigierte Bestimmtheitsmaße. Hier seien die von Theil (1971) und Amemiya (1980) dargestellt.

$\overline{R}_T^2 = 1 - (1 - R^2) \cdot \frac{T - 1}{T - K}$	(Theil)	Gleichung 23
$\overline{R}_A^2 = 1 - (1 - R^2) \cdot \frac{T + K}{T - K}$	(Amemiya)	Gleichung 24

Das korrigierte Bestimmtheitsmaß nach Amemiya bestraft den Verlust eines Freiheitsgrades, der mit der Aufnahme eines zusätzlichen Regressors verbunden ist, mehr als das Bestimmtheitsmaß nach Theil. Damit wird man ceteris paribus Gleichungsvarianten mit weniger Regressoren bevorzugen, wenn man das Bestimmtheitsmaß nach Amemiya als Auswahlkriterium benutzt.

Das Statistikprogramm SAS 8.2 verwendet das korrigierte Bestimmtheitsmaß nach Theil und wird im Ergebnisteil entsprechend aufgeführt.

Die bisher dargestellten Kenngrößen helfen bei der Beurteilung der Güte von Regressionsvarianten, insbesondere von konkurrierenden Modellen. Sie beziehen sich auf die Regressionsgleichung als Gesamtheit.

Die folgenden Kennzahlen eignen sich bei der Gütebeurteilung hinsichtlich der fachwissenschaftlich interessanten Teile des Regressionsmodells: die Regressionskoeffizienten.

Standardisierte Regressionskoeffizienten

Der standardisierte Regressionskoeffizient $\hat{\beta}_i^s$ gibt an, wie groß ceteris paribus der typische Effekt des i-ten Regressors im Verhältnis zum typischen Effekt des Regressanden ist. Je größer $\hat{\beta}_i^s$ dem Betrage nach ist, um so bedeutender ist der i-te Regressor nach der zugrunde liegenden Schätzung (Gruber 1993).

$$\hat{\beta}_{i}^{s} = \frac{\hat{\beta} \cdot s_{i}}{s_{v}} = \frac{typischer \ Effekt \ des \ i - ten \ \text{Re} \ gressors}{typische \ \ddot{A}nderung \ des \ \text{Re} \ gressanden}$$
Gleichung 25

 $\hat{\beta}$ = Schätzwert des Regressionskoeffizienten

s_i = empirische Standardabweichung des i-ten Regressors x_i

s_y = empirische Standardabweichung des Regressanden y

Elastizitätskoeffizienten

Der Elastizitätskoeffizient ermöglicht es, den Effekt eines Regressors auf den Regressanden beurteilen zu können, ohne auf die Maßeinheit achten zu müssen.

$$\hat{\varepsilon}_{i} = \frac{\partial y_{t}}{\partial x_{ti}} \cdot \frac{x_{i}'}{y'}$$
(allgemeine Formel) Gleichung 26
$$\hat{\varepsilon}_{i} = \hat{\beta}_{i} \cdot \frac{x_{i}'}{y'}$$
(für lineare Regressionsgleichungen)

Ändert sich certeris paribus der i-te Regressor um ein Prozent, dann ändert sich der Regressand nach dieser Schätzung um $\hat{\varepsilon}_i$ Prozent.

Der t-Wert

Sowohl der t- als auch der F-Test setzen das klassische Regressionsmodell mit Normalverteilungsannahme voraus, d.h. u.a.

- 1. der Störvariablenvektor ist T-dimensional normalverteilt mit Erwartungswert 0 und Kovarianzmatrix $\sigma_u^2 \cdot I$
- 2. die Regressormatrix X ist fest vorgegeben und besitzt vollen Rang K

Eine vorsichtige Interpretation der Ergebnisse dieser Tests ist bei Verletzung dieser Annahme (z.B. stochastische Regressormatrix X) zu empfehlen (Gruber, 1993).

Mit dem t-Test wird überprüft, ob der k-te Regressor in der Grundgesamtheit einen Effekt auf den Regressanden hat. Es wird also statistisch geprüft, ob der wahre, aber unbekannte Wert des Regressionskoeffizienten von Null verschieden ist. Der t-Test ist sowohl bei einseitigen Fragestellungen (ist der k-te Regressor z.B. größer oder kleiner als x) als auch bei zweiseitigen Fragestellungen (ist der Regressor von x verschieden) durchführbar. Eine Linearkombination von k-Regressoren kann ebenfalls getestet werden. Der t-Test wird durch die Anzahl der gegebenen Freiheitsgrade beeinflusst. Als Faustregel gilt: Bei einem zweiseitigen Test für eine nicht zu kleine Anzahl von Freiheitsgeraden (FG =50) genügt ein kalkulierter t-Wert betragsmäßig größer als 2, um die Nullhypothese β_k = 0 zum Signifikanzniveau α =0,05 abzulehnen (Gruber 1993).

Der F-Wert

Mit dem F-Test können zweiseitige Hypothesen über den Wert mehrerer Koeffizienten bzw. mehrerer Linearkombinationen als Gruppe in einem Zuge statistisch geprüft werden. Es wird also die Nullhypothese geprüft, dass alle Koeffizienten gleich Null sind.

Der Wert der F-Statistik lässt sich unter Verwendung des Bestimmtheitsmaßes berechnen:

$$F = \frac{R^2}{1 - R^2} \times \frac{T - K}{K - 1}$$
 Gleichung 27

Aus dieser Formel (Gruber, 1993) wird ersichtlich, dass mit zunehmendem Freiheitsgrad der F-Wert ansteigt, solange das Bestimmtheitsmaß sich nicht ändert.

Zur weiteren Beurteilung der Ergebnisse der verschiedenen Regressionsmodelle werden noch die verwendeten Freiheitsgrade angegeben. Es sind folgende Freiheitsgrade zu unterscheiden:

Freiheitsgrad der Modellgleichung

Der Freiheitsgrad der Gleichung entspricht der Anzahl der Regressoren in der jeweiligen Regressionsgleichung und damit der Anzahl der zu schätzenden Regressionskoeffizienten (ohne Intercept).

Freiheitsgrad des Fehlers der Modellgleichung

Der Freiheitsgrad des Fehlers berechnet sich aus der Differenz der zur Regression verwendeten Datensätze (T) minus der Anzahl der Regressoren (K) (ohne Intercept) in der jeweiligen Gleichung. Dieser Freiheitsgrad wird als Divisor zur Schätzung des RMSE verwendet.

Freiheitsgrad Total

Die Addition des Freiheitsgrades des Modells und des Fehlers ergibt in der Summe den Freiheitsgrad Total.

7.2 Kennwerte der ex-ante-Prognose (Validierung)

Da für den Testdatensatz alle Laborwerte (Zielvariablen) vorliegen, lässt sich wie bei der expost-Prognose für die ex-ante-Prognose der Fehler $\hat{e} = y - \hat{y}$ für die jeweilige Vorhersage berechnen. Zur Unterscheidung zur ex-post-Prognose sei der Fehler mit "e" gekennzeichnet. Die mathematischen Berechnungen entsprechen dabei weitgehend den Kennwerten zur Beschreibung der Regression mit dem Unterschied, dass anstatt dem ex-post-Fehler u der ex-ante-Fehler e über alle Vorhersagen verwendet wird. Das tiefgestellte "p" bzw. die Abkürzung "prog" zeigt dabei immer Kennzahlen der ex-ante-Vorhersagen an.

Mittlerer Fehler (Mean Error = ME_P)

$$\hat{\overline{e}}_p = \frac{1}{n} \sum (y - \hat{y})$$
 Gleichung 28

Im Gegensatz zur ex-post-Prognose wird der mittlere Fehler der ex-ante-Prognosen nicht gleich Null sein. Er sollte jedoch bei einer konsistenten Schätzung nahe bei Null liegen. Anhand dieser Kennzahl lässt sich feststellen, ob das jeweilige Modell im Mittel (über alle ex-ante-Vorhersagen) die Laborwerte richtig geschätzt hat. Des weiteren kann beurteilt werden, ob im Mittel die Laborwerte über – oder unterschätzt worden sind.

Gleichung 29

Absoluter Fehler (Absolut Error = AE_P)

$$AE_p = \begin{vmatrix} y_t - \hat{y}_t \end{vmatrix} = \begin{vmatrix} \hat{e}_t \end{vmatrix}$$

Absoluter prozentualer Prognosefehler (Absolut Percent Error = APE_P)

$$APE_{p} = \frac{|y_{t} - \hat{y}_{t}|}{y_{t}} \times 100\% = \frac{|\hat{e}_{t}|}{y_{t}} \times 100\%$$
 Gleichung 30

Mittlere absoluter prozentualer Prognosefehler (Mean Absolut Percent Error = MAPE_P)

$$MAPE_{p} = \frac{\sum |\hat{e}_{t} / y_{t}|}{n} \times 100\% \qquad y_{t} \neq 0 \qquad \qquad \text{Gleichung 31}$$

Mittlere absolute Abweichung (Mean Absolut Deviation = MAD_P)

$$MAD_{P} = \frac{\sum \left| forecast \ error \right|}{number \ of \ forecast} = \frac{\sum \left| y_{t} - \hat{y}_{t} \right|}{n}$$
 Gleichung 32

Mittlerer quadratischer Fehler der Vorhersage (MSEp bzw. RMSEp)

$$MSE_{p} = \frac{\sum_{i=1}^{T} (y_{i} - \hat{y}_{i})^{2}}{n-1} \qquad RMSE_{p} = \sqrt{MSE} \qquad \text{Gleichung 33}$$

Entsprechend dem weiter oben für den Beobachtungsraum beschriebenen Vorgehen, wird an dieser Stelle der Quotient aus der empirischen Standardabweichung der Testdaten und dem RMSE_p gebildet. Diese Kennzahl ermöglicht den Vergleich der Schätzgenauigkeit im Vorhersageraum für die verschieden Gleichungen.

Mittlerer Variationskoeffizient des Vorhersagefehlers (RMSPE_p)

$$RMSPE_{p} = \sqrt{\frac{100}{n-1} \sum_{t=1}^{n} ((y_{t} - \hat{y}_{t}) / y_{t})^{2}}$$
 Gleichung 34

Korrelationskoeffizient r_p

$$Korrel_{p} = r_{p} = \frac{\operatorname{cov}(y; \hat{y})}{\sigma_{y} \times \sigma_{\hat{y}}}$$

$$\operatorname{cov} = \frac{1}{n} \sum_{i=1}^{n} (y_{i} - \overline{y}) \times (\hat{y}_{i} - \overline{\hat{y}})$$

$$\sigma_{y} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{i} - \overline{y})} \qquad \sigma_{\hat{y}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_{i} - \overline{\hat{y}})}$$

Zur Quantifizierung des linearen Zusammenhangs zwischen den Vorhersagen und den Laborwerten lässt sich der Korrelationskoeffizient verwenden. Bei gleichen Korrelationskoeffizienten, ist das Modell zu verwenden, bei dem sich die lineare Regressionsgerade am besten der 45°-Linie anpasst. Die Signifikanz des Korrelationskoeffizienten ist von dem Freiheitsgrad abhängig (siehe Anhang A2).

Gleichung 35

7.3 Beurteilung der Modellgüte

Im obigen Kapitel sind die Kennwerte der Validierung definiert worden. Zur Beurteilung der ex-ante-Vorhersagen sind diese Kennwerte für sich genommen nicht aussagefähig. Sie müssen immer im Vergleich zu den Testdaten selbst bzw. zu den ex-post-Kennzahlen oder auch zu eigenen Anforderungen (Welche maximale durchschnittliche Abweichung ist tolerabel?) ins Verhältnis gesetzt werden. Angelehnt an Janssen (1996) wird für ein stabiles Vorhersagemodell gefordert, dass der RMSE_p kleiner sein muss als die Standardabweichung der Testdaten selbst. (In der Arbeit von Janssen wird der RMSE_P als Abweichungsmaß a bezeichnet). Im Gegensatz zur mittleren prozentualen Abweichung (MAPE) geht in der Formel für den RMSE das Quadrat des Vorhersagefehlers ein. Große Abweichungen der Vorhersagen von den tatsächlichen Laborwerten werden damit stärker bestraft, als kleinere Abweichungen. Dies ist insofern wichtig, da Modelle mit gleichmäßig moderaten Fehlern, Modellen mit kleineren, aber gelegentlich sehr große Fehlern, zu bevorzugen sind (Farnum und Stanton, 1998). Des weiteren sollten die Korrelationen zwischen den ex-ante-Vorhersagen und den tatsächlich gemessenen Laborwerten auf dem Niveau von 95% statistisch signifikant sein. Ein statistisch gesicherter Korrelationskoeffizient zwischen Vorhersagen und Laborwerten ist von den verwendeten Freiheitsgraden abhängig. In der Tabelle A2 im Anhang finden sich die Zufallshöchstwerte des Korrelationskoeffizienten in Abhängigkeit vom Freiheitsgrad. Als drittes sehr leicht nachprüfbares Kriterium sollte die mittlere Abweichung des ex-ante-Vorhersagefehlers (ME Prog) annährend Null sein.

Tabelle 2 gibt einen Überblick zu den im Ergebnisteil für die Regressionsmodelle standardmäßig angegebenen Kennziffern. Die Tabellen enthalten im obersten Teil die Kennzahlen des Regressionsmodells und der ex-post-Prognose (Beobachtungsraum) und im unteren Teil die Kennzahlen der ex-ante-Vorhersagen (Vorhersageraum). Wenn es der besseren Übersicht dient, werden noch zusätzlich die empirischen Standardabweichungen der Lern- und Testdaten in den Tabellen mit angegeben.

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD		
R-Square		Gleichung 2	1/22			
Adj R-Square		Gleichung	23			
F-Value		Gleichung	27			
DF_Model		Anzahl der Variable	n im Modell			
MAD		Gleichung	15			
MAPE		Gleichung	16			
RMSE (T-K)	Gleichung 19					
RMSE (n) Gleichung 19 (mit Freiheitsgrad n)						
RMS Pct Error	Error Gleichung 20					
Korrel_Prog		Gleichung	35			
ME_Prog		Gleichung	28			
MAD_Prog		Gleichung	32			
MAPE_Prog		Gleichung	31			
RMSE_Prog		Gleichung	33			
RMS Pct Error Prog		Gleichuna	34			

Taballa Or	I lharaiaht	dam dargaatalltan	Kann-iffarn in	. Erachaiotail
rapelle Z	UDErsicht Zu	oen oaroesiellien	Kennzmern m	i Froeonisieii
	0001010111 20	aon aargootonton		- Ligobinoton

Die Kennzahlen der ex-post- und der ex-ante-Vorhersagen stellen dabei Mittelwerte über den gesamten Vorhersageraum dar. Der RMSE des Modells ist dabei im Ergebnisteil der Regressionsanalyse mit und ohne Freiheitsgradkorrektur dargestellt. Ohne Freiheitsgradkorrektur steigen die Kennwerte mit zusätzlich aufgenommenen Variablen.

In der PLS-Analyse weichen die Kennzahlen des Beobachtungsraumes geringfügig von der Darstellung in Tabelle 2 ab. So werden dort sowohl das Bestimmtheitsmaß für das y-Modell (R-Square R_y^2) als auch für das x-Modell (R-Square R_x^2) angegeben. Die Angabe des adjustierten Bestimmtheitsmaßes und des RMSE (T-K) entfallen. Die um den Freiheitsgrad korrigierten Kennzahlen weichen in den PLS-Modellen aufgrund der geringen Anzahl an latenten Variablen nur unwesentlich von den nicht korrigierten Kennzahlen ab. Als weitere Kennzahl entfällt der F-Wert der Gleichungen. Anstatt der Anzahl an Variablen im Modell wird die Anzahl an PLS-Faktoren in den Tabellen angegeben.

Die Formeln zur Berechnung der Kennzahlen der ex-ante-Vorhersagen (Vorhersageraum) sind mit denen aus der Regressionsanalyse identisch. In beiden Fällen werden die Residuen der ex-ante-Vorhersagen zur Berechnung dieser Kennzahlen herangezogen.

8 Datenerfassung

Ein ausführlicher Datensatz einer kontinuierlichen HDF-Produktion wurde aus der Industrie zur Verfügung gestellt. Eine Darstellung der Variablen mit den statistischen Kennwerten Mittelwert, Standardabweichung und Variationskoeffizient ist im Anhang A1 gegeben. Einzelwerte der Laborwerte werden im Ergebnisteil normiert dargestellt. Dabei wird der Mittelwert des Datensatzes gleich 100% gesetzt.

Die Daten entstammen einer kontinuierlichen HDF-Produktion. Bei dem Produkt handelt es sich um eine 8mm (brutto) HDF Laminat-Trägerplatte.

Der Datensatz wurde direkt dem Datenbankrechner des SPOC-Systems als kommaseparierte Textdatei aus der Produktion entnommen. Die firmeneigenen Laborprüfungen sind im Datensatz enthalten. Zur Auswertung stehen die Querzugfestigkeit, 24-Stunden-Dickenquellung, Abhebefestigkeit und Rohdichte zur Verfügung.

Aufgrund von Schwierigkeiten der Datenübertragung zwischen der Dickenmessung und der SPS, erfolgte zum Teil keine Übertragung oder Aktualisierung der Daten des Dickenmesssystems in die SPS. Um trotzdem den Messwert "Dicke" mit in der Modellierung verwenden zu können, wurden alle Beobachtungen gelöscht, in der die Dicke nicht korrekt übertragen wurde. Ebenfalls sind die Messwerte der Plattenwaage nach der Heißpresse nicht in die SPS eingebunden. Die Daten des Dichtemesssystems Stenograph wurden in das Prozessleitsystem übertragen, konnten aber aufgrund der hohen Geschwindigkeiten in der HDF-Produktion noch nicht zuverlässig aufgenommen werden. Dabei kam es zu Verschiebungen der absoluten Rohdichte. Für die Auswertung des Datensatzes bedeutet dies, dass keine online Messwerte hinter der Presse (außer Dicke) zur Verfügung standen.

Der Lerndatensatz zur Erstellung der Modelle umfasste letztendlich 132 Beobachtungen über einen Zeitraum von 9 Monaten mit insgesamt 321 Variablen. Eine weitere Variablenreduzierung erfolgte in der Daten- und Plausibilitätsanalyse.

Im direkten zeitlichen Anschluss an den Lerndatensatz wurde ein weiterer Datensatz, der Testdatensatz, aufgenommen. Dieser Testdatensatz umfasste einen Zeitraum von 3 Monaten und beinhaltete nach der Daten- und Plausibilitätsanalyse 34 Beobachtungen. Anhand dieses Testdatensatzes wurden die mit den Lerndaten erzeugten Modelle und deren ex-ante-Vorhersagen validiert.

Neun Monate später wurde ein weiterer Datensatz des gleichen Plattentyps gewonnen. Nach der Daten- und Plausibilitätsanalyse enthielt dieser Datensatz 82 Beobachtungen. Anhand dieses zweiten Testdatensatzes konnten die erstellten Modelle auf ihre langfristige Stabilität und Vorhersagegenauigkeit überprüft werden. Der Testdatenraum umfasste gemeinsam mit dem ersten Testdatensatz einen Zeitraum von einem Jahr.

9 Statistische Prozessmodellierung

9.1 Daten- und Prozessanalyse

9.1.1 Ziele der Untersuchungen

Das vorrangige Ziel der Daten- und Prozessanalyse es, die Strukturen des Datensatzes aufzudecken und einfache Zusammenhänge sowohl zwischen den Prozessvariablen als auch zwischen Prozess- und Zielvariablen zu erkennen. Insbesondere sind die Daten auf Plausibilität und auf Ausreißer zu untersuchen. Alle Untersuchungen dienen der Vorbereitung zur nachfolgenden Modellbildung. Eine ausführliche Daten- und Prozessanalyse ist die Vorraussetzung einer erfolgreichen Modellierung.

9.1.2 Statistische Kennzahlen

Tabelle 3 zeigt die Aufteilung der Variablen in die jeweiligen Prozessunterbereiche. Die Dominanz der Presse und die Unterrepräsentanz der Bereiche Holzplatz, Endfertigung und Beleimung ist sehr deutlich.

Die statistischen Kennzahlen Mittelwert, Standardabweichung und Variationskoeffizient der wichtigsten Variablen für den Lerndatensatz sind im Anhang A1 zu finden. Des weitern sind dort die Einheiten und die Zuordnung der Variablen zu den einzelnen Prozessabschnitten, wie in Tabelle 3 angegeben, zu finden.

Bei der Variable "Remark" handelt es sich um eine Handeingabe, mit der unterschiedliche Klebstofflieferanten des gleichen Harztyps codiert werden. In diesem Datensatz werden 2 Harnstoff-Formaldehydharze unterschieden (1 und 4). Ein dritter Harnstoff-Formaldehydharz (Remark=2) wurde in der Vergangenheit kurzzeitig verwendet.

Die Variationskoeffizienten der Labordaten liegen im normalen branchentypischen Durchschnitt, wobei die Rohdichte mit einem CV=1% sehr geringe Schwankungen aufweist.

Bereich	Anzahl der Prozessvariablen
Umgebung	4
Endfertigung	13
Presse	196
Vorpresse	20
Formstrang	10
Beleimung	9
Trockner/Sichter	15
Refiner/Kocher	29
Holzplatz	5
Labor	10
Sonstiges	10
Gesamt	321

Tabelle 3: Aufteilung Prozessvariablen in ihre Unterbereiche

9.1.3 Prüfung auf Stationärität

Der Begriff der Stationärität kommt aus dem Bereich der Ökonometrie bzw. der Zeitreihenanalyse. Stationärität bedeutet, dass jeder Ausschnitt von Prozessvariablen eine gleiche gemeinsame Verteilung hat (Schlittgen, 2001).

Nach Lobenhoffer (1990) ist die Stationärität des Prozesses eine strenge Vorraussetzung für eine konsistente Schätzung der Zielvariablen y aus den Prozessvariablen x. Für eine sehr schnelle und einfache Überprüfung der Stationärität lassen sich die ersten und zweiten Momente (Mittelwerte und Varianzen) heranziehen. Dazu wird die Beobachtungsreihe in nicht allzu kleine Segmente eingeteilt und die dazugehörigen Mittelwerte und Varianzen berechnet. Bei Stationärität des Prozesses sollten die Momente in den Segmenten nicht stark voneinander abweichen.

Innerhalb der Prozessdaten sind verschiedene instationäre Zustände festzustellen. Diese Instationäritäten werden durch jahreszeitliche Einflüsse (Hallenfeuchte, Hallentemperatur, Fasertemperaturen) als auch durch eine ständige Optimierung bzw. Anpassung des Prozesses (Beleimung, Pressengeschwindigkeit, Flächenwaage) verursacht (siehe Abbildung 3). Prozessdaten mit jahreszeitlichen Schwankungen zeigen bei Berücksichtigung eines Jahresdatensatzes obere und untere Grenzen, womit die Stationärität bei Berücksichtigung eines Jahres erneut unterstellt werden kann. In der Regel ist die Forderung nach Stationärität durch Berücksichtigung längerer Zeiträume einzuhalten. Drifts aufgrund von Optimierungen müssen dagegen durch Modellnachführungen bzw. -aktualisierungen berücksichtigt werden. Es bleibt zu prüfen, ob durch eine Modellaktualisierung mit den jeweils neuesten Datensätzen eine konsistente Schätzung der Zieleigenschaften bei instationären Verhalten möglich ist.



Abbildung 3: Instationäre Prozessvariablen aufgrund technologischer Anpassungen (links) bzw. jahreszeitlichen Einflüssen (rechts)

In den Labordaten zeigen sich über den Zeitraum des Lern- und Testdatenraumes von insgesamt 21 Monaten ebenfalls instationäre Zustände. Die Querzugfestigkeit und die Rohdichte zeigen einen fallenden und die Dickenquellung einen steigenden Trend. Dieser Trend wird jedoch erst im Zeitraum des zweiten Testdatensatzes signifikant. Die Abhebefestigkeit zeigt dagegen keinen signifikanten Trend, jedoch im Zeitraum des

Lerndatensatzes einen kurzeitigen Abfall. Abbildung 4 zeigt den Verlauf der Laborwerte über den gesamten Zeitraum von 21 Monaten.

Es bleibt festzuhalten, dass sowohl im Raum der Prozessvariablen als auch in den Labordaten mit instationären Zuständen gerechnet werden muss. Diese Trends treten jedoch nur zeitweilig auf und schwanken in längeren Zeiträumen, als die betrachteten, in oberen und unteren Grenzen. Diese Grenzen sind entweder natürlicher Art oder durch maschinenbauliche Zwänge bestimmt.

9.1.4 Grafische Darstellungen (Line Plots)

Der erste Schritt jeder Datenanalyse sollte die grafische Veranschaulichung der Daten sein. Dabei ergibt die Darstellung der linearen Verbindungslinien meist einen besseren Eindruck der Gesamtstruktur der Reihe als eine Darstellung einzelner Punkte (Schlittgen 2001). Aus diesen Plots lassen sich sehr schnell Trends und die Existenz von Werten herausfinden, die stark vom Mittelwert abweichen. Abbildung 4 zeigt den Verlauf der Laborwerte über 21 Monate. Die Werte sind dabei auf den Mittelwert des gesamten Zeitraums normiert. Der linke Abschnitt in den Grafiken zeigt den Lerndatensatz, der mittlere Abschnitt stellt den ersten Testdatensatz und der rechte Abschnitt den zweiten Testdatensatz dar. Die oben beschriebenen Trends in den Labordaten sind in der grafischen Darstellung sehr gut nachzuvollziehen. Die Schwankungen der Einzelwerte um den Mittelwert sind bei der Rohdichte mit ± 3% sehr gering. Die übrigen Laborwerte zeigen Schwankungen von bis zu 25% um den langfristigen Mittelwert.



Abbildung 4: Verlauf der Laborwerte (normiert auf Mittelwert) über 21 Monate

9.1.5 Plausibilitätsprüfung

In einer Plausibilitätsprüfung werden die minimalen bzw. maximalen Werte der einzelnen Variablen mit den technologischen bzw. maschinenbaulichen Grenzen verglichen. Für diese Prüfung ist es notwendig die technischen Grenzen der Prozessvariablen zu kennen. Alle Werte, die außerhalb dieser Grenzen liegen, sollten aus der weiteren Analyse ausgeschlossen werden. Ausgang der Analyse bildete die Datei mit 321 Variablen.

Die Hallentemperatur und die Außentemperatur konnten vorerst als Prozessvariablen nicht verwendet werden, da im Frühjahr 2002 ein Blitzeinschlag die Kalibrierung zerstörte und die folgenden Temperaturaufzeichnungen bis in den Sommer unbrauchbar machte. Eine Löschung der Datenbeobachtungen (60 Datensätze) mit falschen Temperaturen hätte einen zu großen Verlust an Beobachtungsdaten bedeutet. Daher wurde auf die Hallen- und Außentemperatur als Einflussgrößen verzichtet. (Mit weiterer Datensammlung können zu einem späteren Zeitpunkt die Temperaturen wieder berücksichtigt werden.)

Wie bereits weiter oben erwähnt, konnten die Messgrößen des online-Dichtemessgerätes (Stenograph) nicht verwendet werden. Damit konnte der Einfluss von Rohdichteprofilkennwerten nicht untersucht werden.

Die Variablen der Dickenmessung der Spuren 1 und 2 mussten ebenfalls aus der weiteren Analyse genommen werden. Diese beiden Spuren zeigten zu viele Fehlwerte an, um die Beobachtungen zu löschen. Daher wurde entschieden, die Variablen aus dem Datensatz zu nehmen. Es verbleiben damit noch die Messspuren 3 und 4 und 5 im Datensatz.

Die gemessenen Drücke des Vorverdichters zeigten sprunghafte Anstiege im Verlauf. Sie wurden zunächst im Datensatz beibehalten, jedoch als kritische Variablen gekennzeichnet. Die Multipots zeigten im drucklosen Zustand negative Distanzen. Anstatt der Distanzen wurden in der Analyse die Drücke verwendet. Abbildung 5 zeigt den Druckverlauf des Vorverdichters und die Distanz des Multipots 2.



Nach der Plausibilitätsanalyse verblieben 281 Prozessvariablen im Datensatz.

Abbildung 5: Verlauf Druck Vorverdichter (links) und Distanz Multipot 2 (rechts)

9.1.6 Schwankungsanalyse

In Verbindung mit den Line-Plots lassen sich Schwankungen in den Variablen sehr gut veranschaulichen. Ein objektives Maß für gering und stark schwankende Variablen ist der Variationskoeffizient (CV). Mit dieser Kennzahl lassen sich die Schwankungen von Variablen mit verschiedenen physikalischen Einheiten vergleichen. Gering schwankende Variablen zeichnen sich durch einen sehr kleinen Variationskoeffizienten aus. Nach Lobenhoffer (1990) werden Schätzungen schwierig, wenn die Streuweite unter 2% liegt.

Tabelle 4 zeigt die Variablen des Lerndatensatzes mit einem Variationskoeffizienten größer 20% und Tabelle 5 kleiner 2%.

Es ließen sich 49 Variablen mit einem Variationskoeffizienten größer 20% feststellen. Es zeigt sich, dass bei Datensätzen, die über die Sommer- und Wintermonate eines Jahres gesammelt worden sind, jahrzeitlich schwankende temperatur- und feuchteabhängige Variablen entsprechend Variationskoeffizienten größer 20% aufweisen. Weitere stark schwankende Variablen sind die Variablen Harnstoff und Härter. Da diese nur bei Bedarf in diskreten Mengen zugeführt werden, ist die Streuung entsprechend hoch. Die Füllstände von Bunkern zeigen ebenso hohe Schwankungen im laufenden Betrieb.

Messfehler mit Temperaturen größer 60°C zeigen sich in der Außen- und Hallentemperatur (siehe Plausibilitätsprüfung). Zudem zeigt die Luftfeuchtigkeit mit Werten oberhalb von 100% Ungenauigkeiten an.

Alle weiteren Variablen mit einem Variationskoeffizienten größer 20% entstammen größtenteils der Presse und der Vorpresse und einige dem Bereich Refiner und Hackschnitzelaufbereitung.

Die hohen Streuungen der Multipots lassen sich durch die zeitweise Herausnahme dieser erklären. Messfehler zeigen sich in den Distanzen der Multipots und der Mahlscheiben (negative Werte). Als kritisch zu beurteilen sind die meisten Messwerte der Vorpresse, die Bandverläufe und die Vibrationen des Refiners. Technologisch erklärbar hohe Streuungen finden sich in der Dampfzugabe im Hackschnitzelbunker und in den Austragsgeschwindigkeiten der Schnecken in den Silos.

Es bleibt festzuhalten, dass Variablen mit sehr hohen Schwankungen individuell mit Technologen vor Ort begutachtet werden sollten, um deren Plausibilität festzustellen. Oft finden sich gerade in diesen Variablen Messfehler. Diese Variablen sollten möglichst alle identifiziert und eliminiert werden. Andererseits sollten technologisch erklärbare stark streuende Variablen nicht automatisch durch einen Filter ausgeschlossen werden. Eine Entscheidungshilfe bietet zudem die Beurteilung der Line-Plots.

Nr	Variable	CV									
1	u_Halle	25,2	15	SDR_R18	21,1	29	MPOS_25	184	43	VIB_REF	30,3
2	u_AUS	24,7	16	MKDR_2	56,8	30	B_AL_V_O	63,1	44	DDR_KOCH	55,1
3	TEMP_AUS	67,3	17	MKDR_3	50,1	31	B_AL_V_U	86,3	45	DA_HKS	45,5
4	Temp_Ha	33,3	18	MKDR_4	48,1	32	KVA_DR_R	27,8	46	VEN_HKS	37,9
5	FUEL_FAB	24,9	19	MKDR_5	55,4	33	KVA_DR_L	33,6	47	V_SCHN_1	48,7
6	TEMP_SI	39,1	20	MKDR_6	56,3	34	KVA_SDR	33,2	48	V_SCHN_2	50,1
7	u_SI	24,0	21	MKDR_21	40,4	35	KVE_DR_L	26,3	49	DIST_Mah	-117
8	TEMP_L_G	38,0	22	MKDR_23	39,1	36	KVE_DR_R	26,1			
9	TEMP_RE2	39,1	23	MKDR_25	39,1	37	STR_CBVo	30,3			
10	Haert	86,0	24	MPOS_2	84,8	38	CBV_O_L	39,4			
11	Harnst	21,9	25	MPOS_4	54,0	39	CBV_O_R	30,2			
12	FUEL_S1	33,0	26	MPOS_5	61,4	40	CBV_U_L	28,6			
13	FUEL_S2	30,1	27	MPOS_6	52,4	41	AL_CBV_L	47,1			
14	FUEL_S3	35,8	28	MPOS_23	-319	42	AL_CBV_R	49,6			

Tabelle 4: Variablen mit Variationskoeffizienten größer 20% im Lerndatensatz

Die häufigsten sehr gering schwankenden Variablen sind ebenfalls in der Presse zu finden. Die meisten Distanzen der Pressrahmen 5 bis 15, die Distanzen des Bandes am Auslauf und sämtliche Presstemperaturen zeigten einen kleineren Variationskoeffizienten als 2%. Im Gegensatz zu den Pressdistanzen weisen die Pressdrücke i.d.R. mit Variationskoeffizienten zwischen 7 und 16% eine höhere Schwankungsbreite auf. Dies ist auf die isochore Pressensteuerung zurückzuführen.

Die im Labor gemessene mittlere Rohdichte zeigt trotz des langen Beobachtungszeitraumes mit einem Variationskoeffizienten von 1,1% eine sehr geringe Schwankungsbreite. Dies spiegelt sich ebenfalls in den geringen Schwankungen der Flächengewichts (CV=1,5%) wider. Weitere kaum streuende Variablen sind die Plattendicke, Plattenbreite und die Temperatur im Mahlraum des Refiners.

Nr	Variable	CV	Nr	Variable	CV	Nr	Variable	CV	Nr	Variable	CV
1	Lab_MD	1,1	10	B_AL_D4	1,4	19	DIST_R15	2,0	28	DIST_L10	1,8
2	Lab_Dicke	0,3	11	TEMP_H1	0,7	20	DIST_L14	2,0	29	DIST_L09	1,8
3	Dicke_S3	0,9	12	TEMP_H2	0,6	21	DIST_R14	2,0	30	DIST_L8	1,7
4	Dicke_S4	0,7	13	TEMP_H3	0,9	22	DIST_L13	1,9	31	DIST_L7	1,7
5	Dicke_S5	0,9	14	Temp_H4	1,5	23	DIST_R13	2,0	32	DIST_L6	1,9
6	FG_Gew_2	1,5	15	TEMP_H5	1,0	24	DIST_L12	1,8	33	DIST_L5	2,0
7	B_AL_D1	0,9	16	TEMP_H6	0,9	25	DIST_R12	1,8	34	Breite_I	1,8
8	B_AL_D2	0,7	17	TEMP_ROL	1,5	26	DIST_L11	1,8	35	TEMP_MR	1,3
9	B_AL_D3	1,3	18	DIST_L15	2,0	27	DIST_R11	1,8			

Tabelle 5: Variablen mit Variationskoeffizienten kleiner 2% im Lerndatensatz

9.1.7 Ausreißeranalyse

Zur Detektierung von Ausreißern gibt es verschiedene Untersuchungsmöglichkeiten. Keine dieser Möglichkeiten sollte allein als Grundlage zur Entscheidung der Entfernung dieses Beobachtungspunktes dienen. Ein Vorschlag zur Entfernung von Ausreißern beruht darauf, die Analyse jeweils mit und ohne Ausreißer durchzuführen. Hierbei zeigte sich aber in den Voruntersuchungen, dass je nach Modellauswahl bzw. Modellgüte eine Beobachtung ein Ausreißer sein kann und mit einem anderem Modell diese Beobachtung vollständig erklärt werden kann. Von einer automatischen Entfernung von Ausreißern wird abgeraten.

Ausreißer können Informationen besitzen, die andere Daten nicht liefern, da sie zum Teil durch ungewöhnliche Kombinationen und Umständen entstehen und eher nähere Untersuchungen verdienen, als entfernt zu werden. Draper und Smith (1998) schlagen als generelle Regel vor, Beobachtungen per Hand nur dann zu entfernen, wenn die Ursache auf Fehler in der Aufzeichnung oder aber im Messgerät zurückverfolgt werden können.

Das Problem der Detektierung von Ausreißern verschärft sich zunehmend im Fall der multivariaten Betrachtung. Beobachtungen, die außerhalb des Datenraumes liegen, führen allgemein zu "extrapolierten" Vorhersagen, die um so kritischer zu bewerten sind, je weiter diese Beobachtungen von diesem Datenraum entfernt sind (Draper und Smith, 1998). Abbildung 6 verdeutlicht das Problem der Detektierung von Ausreißern im zweidimensionalem Raum. Obwohl der Punkt P jeweils im Datenbereich von x₁ und x₂ liegt, ist der



Punkt P selbst außerhalb des Datenraumes. Je mehr Dimensionen zu berücksichtigen sind, desto schwieriger ist es, solche Ausreißer zu entdecken.

Abbildung 6: Ausreißer im zweidimensionalem Raum, nach Draper und Smith (1998)

Aus Darstellungsgründen (möglichst geringe Abweichungen) werden oft Extremwerte subjektiv entfernt. Entscheidungshilfen mit Hilfe statistischer Tests überfordert in der Regel die meisten Nicht-Statistiker. Zudem besteht oft das Problem über zuverlässige Angaben über die Verteilungsform der Grundgesamtheit. Aus diesem Grunde wird in der Literatur oft zu praktischen objektiven Entscheidungsregeln geraten. Einen Kompromiss aus der Materialprüfung schlägt Dengel (1996) vor: Es werden Mittelwert x und Standardabweichung s der Beobachtungsreihe ohne den vermuteten Ausreißer berechnet und mit dem Faktor k=3,7+12/n der Bereich Mittelwert $x \pm k \times s$ gebildet. Liegt der "verdächtige" Wert nicht in diesem Bereich, so ist er als Ausreißer einzustufen. Nach Dengel hat sich diese Regel in der Praxis als technisch recht brauchbar erwiesen. "Selbst bei deutlicheren Abweichungen der Messwertverteilung von der Normalverteilung ist dieses Vorgehen noch immer weit sinnvoller

als die dann fälschliche Anwendung der normalverteilungsgebundenen Tests oder die rein subjektive Entscheidung." Diese Ausreißerbehandlung hat den entscheidenden Nachteil, dass die "verdächtigen" Werte ohne objektive Entscheidungsregel gefunden werden müssen und in Abhängigkeit von der Anzahl der verdächtigen Werte sich der Wert n ändert. Damit ist dieses Verfahren nicht automatisch auszuführen und für große Datenmengen äußerst unpraktikabel. Daher wurde neben dem exakten Verfahren nach Dengel, das Intervall zusätzlich ohne Berücksichtigung eventueller Ausreißer berechnet.

Um einen Überblick über die mögliche Anzahl der Ausreißer in den Daten zu bekommen, wurden folgende 4 Ausreißerintervalle gebildet:

- 1. Mittelwert ± 3-fache Standardabweichung (entspricht bei einer Normalverteilung 99,7% aller Werte) =out1
- 2. Mittelwert ± 4-fache Standardabweichung (entspricht bei einer Normalverteilung 99,994% aller Werte) =out2
- 3. Mittelwert ± (3,7+12/n) × Standardabweichung, ohne Berücksichtigung eventueller Ausreißer (abgewandelte Berechnung nach Dengel) =out3
- 4. Mittelwert ± (3,7+12/n) × Standardabweichung, mit Berücksichtigung eventueller Ausreißer (exakte Berechnung nach Dengel) =out4

Die Anzahl an Variablen, die diese Intervalle mit maximalen oder minimalen Werten überschreiten, zeigt die Tabelle 6 in einer Übersicht:

	out1_unten	out1_oben	out2_unten	out2_oben	out3_unten	out3_oben	out4_unten	out4_oben
Anzahl Ausreißer	62	64	18	6	22	11	22	14
gesamt	1:	24	2	4	3	33	53	6

Tabelle 6: Anzahl Ausreißer nach verschiedenen Verfahren

Die Bildung eines Intervalls um die Mittelwerte der einzelnen Variablen ergab bei 3-facher Standardabweichung 62 Unterschreitungen der unteren Grenze und 64 Überschreitungen der oberen Grenze. Bei einem Intervall mit 4-facher Standardabweichung um den Mittelwert reduzieren sich die Unterschreitungen auf 18 und die Überschreitungen auf 6 Variablen. Das abgewandelte und das exakte Verfahren nach Dengel unterscheidet sich in einer Differenz von 3 Variablen bei der Überschreitung des oberen Intervalls.

Keines dieser Verfahren ist für eine automatische Entfernung von "Ausreißern" geeignet. Dabei würden zu viele wertvolle Informationen verloren gehen. Die Entscheidung ob eine Messung sicher oder unsicher ist, kann kein automatisches Verfahren treffen. Es scheint aber praktikabel einen Filter mit einem Intervall (z.B. 4-fache Standardabweichung oder Intervall nach Dengel) zu setzen und sich die Werte, die außerhalb dieses Intervalls liegen, anzeigen zu lassen (z.B. als markierte Punkte eines Lineplots). Der Technologe bekommt damit die Möglichkeit an die Hand, die Messtechnik zu überwachen und verdächtigen Werten nachzugehen. Damit können auch sehr schnell Eingabefehler im Labor erkannt werden. Als Messfehler festgestellte Werte sollten entfernt werden, damit diese Werte nicht in die Regressionsanalyse aufgenommen werden.

Abbildung 7 zeigt zwei Beispiele von detektierten "Ausreißern", die außerhalb der 4-fachen Standardabweichung, als auch außerhalb des Dengel-Intervalls lagen. So werden z.B. die Mutipots nicht permanent genutzt, beim "rausnehmen" dieser fällt der Druck entsprechend ab. Diese Werte wurden bei den Multipots von allen angewendeten Verfahren zur Bestimmung von Ausreißern als solche identifiziert, obwohl sie technologisch korrekt sind. Die rechte Abbildung zeigt die Stromaufnahme des unteren Bandantriebes der Vorpresse. Der markierte Wert wurde ebenfalls als Ausreißer gekennzeichnet. Zeitgleich an dieser Stelle fiel die Stromaufnahme des oberen Bandabtriebes rapide ab, so dass diese scheinbaren Ausreißer keine Messfehler darstellten.



Abbildung 7: Detektierte scheinbare Ausreißer in der Datenaufzeichnung

Die Entscheidung, ob Werte als Ausreißer zu definieren sind, lässt sich oft erst bei der Analyse der Residuen nach der Modellbildung feststellen. In der Praxis erscheint diese Möglichkeit jedoch als nicht durchführbar. Es wurden alle Werte beibehalten, die nicht mit Sicherheit auf Mess- oder- Kalibrierfehler zurückzuführen waren (wie z.B. die Distanz der Mahlscheiben des Refiners bzw. die Distanzen der Multipots).

9.1.8 Test auf Normalverteilung

Eine Annahme der Regressionsanalyse besteht in der Annahme der Normalverteilung der Störvariablen. Diese Annahme ist jedoch nicht für die Regressionsanalyse selbst entscheidend, sondern für die Anwendung der t- und F-Tests. Da die Störvariablen sich aus nicht messbaren Größen zusammensetzen und damit unbekannt sind, wird diese Annahme stattdessen an den Residuen getestet. Ein Rückschluss auf Nichtnormalverteilung der Störvariablen auf Grund von nichtnormalverteilten Regressoren ist nicht möglich. Es zeigte sich in den Untersuchungen, dass auch bei nichtnormalverteilten Regressoren die Residuen i.d.R. normalverteilt sind.

Nach Gruber (1994) kann dann von einer Normalverteilung der Störvariablen ausgegangen werden, wenn der zentrale Grenzwertsatz anwendbar ist. "Dies ist dann der Fall, wenn u.a. die Störvariable das Ergebnis des Zusammenwirkens einer sehr großen Anzahl von zufälligen und unabhängigen Einzeleinflüssen ist, die, jeder für sich genommen, keinen dominierenden Einfluss haben." Von dieser Annahme kann bei der Herstellung von Holzwerkstoffen ausgegangen werden. Dies konnte durch die Untersuchungen der Residuen bestätigt werden.
Unter den in der Literatur dargestellten Tests auf Normalverteilung sollen die Ergebnisse von zwei grafischen und zwei rechnerischen Verfahren vorgestellt werden. Vor allem die grafischen Darstellungen bieten die Möglichkeit sich einen schnellen Überblick über die tatsächliche Verteilung der Daten zu verschaffen und identifizieren zugleich extreme Werte in dem Datenraum.

Zu den grafischen Möglichkeiten der Überprüfung der Normalverteilung zählt das Normalverteilungs-Diagramm (normal plot) und das Histogramm mit eingezeichneter theoretischer Normalverteilungskurve auf Grundlage der empirischen Daten.

Abbildung 8 zeigt die Diagramme für die Querzugfestigkeit (links) und für die Feuchte der Platten im Labor (rechts). Die Einzelwerte sind auf den Mittelwert normiert worden. Eine perfekte Normalverteilung zeigt sich im normal plot dadurch, dass alle Werte auf der durchgezogenen Linie liegen, im Histogramm durch eine Anpassung der Balken an der Glockenkurve. Tabelle 7 zeigt die rechnerischen Ergebnisse des Shapiro-Wilk und des Kolmogoroff-Smirnov-Tests auf Normalverteilung für die beiden in der Abbildung 8 dargestellten Szenarien. P-Werte kleiner als 0,10 weisen dabei auf keine normalverteilten Daten hin. Die unterschiedlichen Ergebnisse dieser beiden Tests verdeutlichen die Schwierigkeit, eine Normalverteilung rechnerisch zu identifizieren. Die Ergebnisse sind je nach Testmethode unterschiedlich.

Es lässt sich feststellen, dass der Shapiro-Wilk-Test leicht konservativer testet als der Kolmogoroff-Smirnov-Test. Von 243 getesteten Variablen sind nach dem Shapiro-Wilk-Test 108 nach dem Kolmogoroff-Smirnov-Test 124 Variablen normalverteilt.

Eigenschaft	Shapiro-Wilk	KolmSmirnoff
Querzug ⁽¹⁾	0,002	0,150
Feuchte ⁽¹⁾	0,000	0,010
Querzug ⁽²⁾	0,106	0,150
Feuchte ⁽²⁾	0,229	0,030

Tabelle 7: p-Werte für den Test auf Normalverteilung

(1) alle Beobachtungen in der Analyse

(2) nach Entfernen der markierten Punkte in Abbildung 8

Es sei an dieser Stelle nochmals darauf hingewiesen, dass für die Regressionsmodelle die Variablen selbst keine Normalverteilung aufweisen müssen. Jedoch können insbesondere durch die grafischen Darstellungen Rückschlüsse auf die Streuungsbreite und auf extreme Einzelwerte der Variablen gewonnen werden. Im normal plot präsentieren sich Ausreißer in der rechten Hälfte der Grafik oberhalb der diagonalen Linie, in der linken Hälfte dagegen unterhalb der Linie. In der unteren Hälfte der Abbildung 8 sind jeweils die Diagramme nach Löschen der vermeintlichen Ausreißer abgebildet.



Diagramme nach löschen der markierten Punkte



Abbildung 8: Histogramm- und Normalverteilungs-Diagramme für die Querzugfestigkeit und die Plattenfeuchte (normiert, Mittelwert entspricht 100%)

9.1.9 Korrelationsanalyse und Bewertungszahlen

Eine weitere Voraussetzung der Regressionsmodelle liegt in der Annahme, dass die Regressoren frei von extremer Kollinearität sind. In erster Näherung kann man "Kollinearität der Regressoren" mit "Korrelation zwischen Regressoren" übersetzen (Gruber, 1994). Die drei bedeutendsten Folgen von extremer Kollinearität sind:

- (1) Große Varianzen und Kovarianzen der Schätzer für die Regressionskoeffizienten
- (2) Mangel an Schätzbarkeit, Regressionskoeffizienten können mit üblichen Methoden nicht errechnet werden (Kreuzproduktmatrix X'X ist nicht invertierbar)
- (3) Instabile Regressionskoeffizienten (z.B. Vorzeichenwechsel)

Der dritte Punkt ist der Grund dafür, warum die Vorzeichen einzelner Variablen in den Modellen technologisch falsch sein können. Modelle mit Kollinearität lassen sich technologisch nicht mehr interpretieren. Nach Lobenhoffer (1990) können die Vorhersagen trotz Multikollinearität jedoch brauchbar genau bleiben, da sich die negativen Einflüsse aufheben, solange man im Datenraum von X bleibt. Problematisch bleiben diese Modelle in jedem Fall, da technologische Ableitungen zur Optimierung des Prozesses auf Grundlage kollinearer Modelle nicht mehr seriös möglich sind.

Kennzahlen zur Detektierung von Kollinearitäten sind:

- (1) Determinante der Kreuzproduktmatrix
- (2) Determinante der Korrelationsmatrix
- (3) Einzelne Korrelationskoeffizienten
- (4) Varianzinflationsfaktor
- (5) Konditionsindex
- (6) Konditionszahl

Nach Gruber (1994) sind nahe bei 1,0 liegende Korrelationskoeffizienten bei drei oder mehr Regressoren eine hinreichende aber nicht notwendige Bedingung für Kollinearität. Das heißt, es kann auch Kollinearität vorliegen, ohne dass auch nur ein Korrelationswert nahe bei 1,0 liegt. Die Korrelationsmatrix R vermittelt aber trotzdem einen ersten Eindruck über Kollinearität und wird in der Praxis daher auch oft verwendet. Des weiteren lassen sich mit der Korrelationsmatrix sehr leicht direkte Linearkombinationen (r=1) von Variablen aufdecken. Linearkombinationen verhindern eine Matrizeninversion und damit die Schätzung der Regressionskoeffizienten. Nach Judge et al. (1988) wird bei Korrelationskoeffizienten größer 0,80 bzw. 0,90 die Multikollinearität ein ernsthaftes Problem. Eine ausführliche Darstellung der Kollinearität und deren Folgen findet sich in Gruber (1994), Judge et al. (1988) und Fomby et al. (1988).

In den Untersuchungen dieser Arbeit wurde der Korrelationskoeffizient und der Varianzinflationsfaktor zur Aufdeckung von Kollinearitäten herangezogen.

Um einen ersten Eindruck der Korrelationen der Prozessdaten zu den Laboreigenschaften zu erhalten, zeigt die Abbildung 9 grafisch die einzelnen Korrelationskoeffizienten aller Prozessdaten zu den Laboreigenschaften. Die horizontalen Linien verdeutlichen die Grenzen der statistischen Signifikanz (S=95%) der Korrelationskoeffizienten. Die vertikalen Linien

dagegen kennzeichnen die jeweiligen Prozessabschnitte. Die Zuordnung der Zahlen zu den Prozessabschnitten ist in Tabelle 8 zu finden. Dort sind des weiteren die Anzahl der statistisch signifikanten Variablen zu den einzelnen Laboreigenschaften in Abhängigkeit des Prozessabschnittes aufgeführt. Im Unterschied zu der Grafik sind in der Tabelle nur die Variablen aufgeführt, die nach der Plausibilitätsprüfung noch im Datensatz enthalten waren und eine statistisch signifikante Korrelation vorwiesen. So zeigten z.B. verschiedene Variablen eine statistische Signifikanz, obwohl sie nachweislich mit Messfehlern behaftet waren (Temperaturen der Umgebung, Kennzahlen des Dichtemesssystems Stenograph, Einlauftemperatur der Presse). Dies verdeutlicht, die Notwendigkeit einer technologischen Überprüfung der einzelnen Variablen, bevor statistische Analysen angewendet werden.



Abbildung 9: Korrelationskoeffizienten zwischen den Prozessdaten und den Laboreigenschaften, Korrelationskoeffizienten außerhalb der horizontalen Linien statistisch signifikant (S=95%), vertikale Linien kennzeichnen Prozessabschnitte

Sowohl der Abbildung 9 als auch der Tabelle 8 ist zu entnehmen, dass die mittlere Rohdichte und die Querzugfestigkeit die meisten statistisch signifikanten Korrelationen zu den Prozessdaten aufweisen. Dies wird in erster Linie durch die Variablen der Presse verursacht. So weisen mehr als 100 Pressenvariablen eine statistisch signifikante Korrelation zur mittleren Rohdichte und zur Querzugfestigkeit auf.

Codierung	1	2	3	4	5	6	
Bereich	Endfertigung/	Presse	Vorpresse	Formstrang/	Trockner	Kocher,	Gesamt
	Umgebung			Beleimung	Sichter	Refiner/	
						Holzplatz	
Querzug	0/1	110	6	3/2	5	3/1	131
Dickenquell.	0/0	34	7	6/6	2	9/0	74
Abhebefest.	0/1	21	13	5/5	8	5/1	67
Rohdichte	1/0	128	9	1/6	7	5/0	159

Tabelle 8: Anzahl signifikanter Korrelationen der Prozessdaten nach der Plausibilitätsprüfung zu den Laboreigenschaften (Lerndatensatz, S=95%)

Die Vorpresse und der Trocknungs- und Sichtungsprozess zeigt die meisten gesicherten Korrelationen zur Abhebefestigkeit. Der Kocher bzw. Refiner wiederum weist dagegen die meisten gesicherten Korrelationen zur Dickenquellung auf. In der Beleimung fällt auf, dass nur zwei Variablen (Härtermenge und Leimlieferant) eine statistische Signifikanz zur Querzugfestigkeit aufweisen. Alle anderen Eigenschaften zeigen dagegen signifikante Korrelationen zu fünf bzw. sechs Variablen aus der Beleimung.

Grundsätzlich lässt sich jedoch aus der Höhe des einfachen Korrelationskoeffizienten einer Prozessvariable nicht auf die Bedeutung zur Laboreigenschaft schließen. Dazu müssen die Einflüsse der restlichen Prozessvariablen berücksichtigt werden.

Tabelle 9 zeigt die Korrelationsmatrix der Laborwerte. Diese Matrix beruht auf dem Lerndatensatz mit 132 Beobachtungen. Alle Korrelationen sind auf dem 99% Niveau statistisch signifikant. Die höchste Korrelation ist zwischen der Querzugfestigkeit und der Abhebefestigkeit gegeben. Die Dickenquellung zeigt eine negative Korrelation zu den übrigen Laborwerten.

	Querzug-	Rohdichte	Dicken-	Feuchte
	festigkeit		quellung	
Rohdichte	0,22			
Dickenquellung	-0,29	-0,35		
Feuchte	0,38	0,26	-0,34	
Abhebefestigkeit	0,53	0,42	-0,35	0,32

Tabelle 9: Korrelationsmatrix der Laborwerte (N=132)

Tabelle 10 zeigt die zehn höchsten korrelierten Prozessdaten zu den Laboreigenschaften. Die Temperaturen des Streubunkers, des Faserbunkers und des Trocknerausgangs sind alle (höchst signifikant) negativ zur Querzugfestigkeit korreliert. Das bedeutet, je höher diese Temperaturen, desto geringer war innerhalb dieses Zeitraumes die Querzugfestigkeit. Mit zunehmender Härtermenge fällt ebenfalls die Querzugfestigkeit. Die Remark-Variable ist eine Codierung zur Unterscheidung der Klebstofflieferanten. Es besteht ein starker Zusammenhang zwischen Höhe der Querzugfestigkeit und des Klebstofflieferanten (siehe auch Kapitel 9.1.12). Weiterhin zeigen die Drücke der Multipots 5 und 6 eine positive und die Drücke der Rahmen 9 und 17 eine negative Korrelation zur Querzugfestigkeit. Die Verwendung beider Multipots bzw. beider Drücke führte in den Regressionsmodellen auf grund der hohen Korrelationen zwischen den Variablen stets zu verschiedenen Vorzeichen.

-			-				
LAB_IB		Lab_MD		LAB_TS		Lab_SS	
TEMP_STB	-45,9	DR_L22	37,3	DZ_SCHN	-41,4	B_AL_D3	41,4
REMARK	43,5	DR_R22	36,4	Faser	-40,8	B_AL_D4	40,7
TEMP_FAB	-41,4	SDR_R22	35,5	STR_CBVu	-40,6	STR_CBVo	37,6
MKDR_5	37,8	DR_L21	34,9	Leim	-39,3	TEMP_RE1	-36,0
TEMP_TRA	-37,6	SDR_R21	34,7	Wasser	-38,0	KVE_DI_R	-35,0
SDR_R9	-36,9	DR_L24	34,4	Härter	37,7	KVE_DI_L	-34,7
Härter	-36,7	SDR_R13	34,2	REMARK	-36,5	STR_CBVu	34,6
SDR_R17	-36,2	DR_L25	33,6	KVK_DI_R	29,2	TEMP_L_G	-34,3
DR_R9	-35,3	SDR_R23	33,6	u_Matte	-28,5	EL_CBV	-34,0
MKDR_6	35,3	DR_L26	33,6	DAMPF_GE	-28,2	AL_CBV_R	-33,6

Die 10 maximalen Korrelationen zur Rohdichte bestehen alle zu den Variablen der Presse. Die Drücke, vorrangig aus dem letzten Drittel der Presse (Kalibrierzone), zeigen alle positive Korrelationen zur Rohdichte. Diese Variablen würden innerhalb eines Prozessmodells starke Multikollinearitätsprobleme aufwerfen, da sie alle miteinander sehr hoch korreliert sind.

Die am höchsten korrelierten Variablen zur Dickenquellung entstammen vorrangig aus der Beleimung. Die Drehzahl der Austragsschnecke des Refiners (DZ_Schn) ist dabei hoch korreliert mit den Variablen aus der Beleimung, da die Durchflussmengen der Beleimung nach der Menge an Fasern geregelt wird. Die Korrelationskoeffizienten zwischen der Drehzahl der Austragsschnecke des Refiners und der Leimmenge und des Paraffins betragen 0,92 bzw. 0,95.

Eine Erhöhung der Härtermenge zeigt wie bei der Querzugfestigkeit einen negativen Effekt, mit zunehmender Härtermenge stieg die Dickenquellung in diesem Zeitraum an.

Eine höhere Stromaufnahme des Bandantriebes der Vorpresse (STR_CBVu) korreliert mit einer niedrigeren Dickenquellung. Eine höhere Stromaufnahme kann technologisch auf eine erhöhte Verdichtungsarbeit der Fasermatte zurückgeführt werden. Diese wird u.a. durch die Mattenhöhe (Verdichtungshöhe), Holzart, Bandgeschwindigkeit, Fasertemperatur und Faserfeuchte beeinflusst. Nicht nachweisbar, aber zu vermuten ist auch ein Zusammenhang zwischen Fasereigenschaften und der Stromaufnahme. Dieses Beispiel verdeutlicht, wie bedeutend eine auf den ersten Blick unwichtige Messgröße sein kann und deren komplexen Zusammensetzung aus den verschiedensten Einflüssen. Die technologische Interpretation der Stromaufnahme in den Modellen ist daher nicht eindeutig möglich, da sie selbst durch andere Prozessvariablen stark beeinflusst wird.

Die Mattenfeuchte (u_Matte) und die Dampfmenge des Refiners sind negativ mit der Dickenquellung korreliert.

Das höchsten Korrelationen der Prozessvariablen zu der Abhebefestigkeit zeigen die Distanzen der Vorpresse, des Keilverdichters und des Bandauslaufes der Hauptpresse, die Stromaufnahmen der Bandantriebe der Vorpresse sowie die Temperaturen im Sichtungsprozess. Ein anderes Bild in der Reihenfolge der ersten zehn Eingangsvariablen zeigt sich, wenn die Korrelationen der Prozessparameter untereinander und die Anzahl der vollständigen Datensätze berücksichtigt werden (siehe Kapitel 6.1.7.4 optimale Eingänge). Tabelle 11 zeigt die Korrelationskoeffizienten in % der Prozessparameter zu den Laboreigenschaften (AusCorr), die maximalen Korrelationen zu in der Tabelle weiter oben stehenden Prozessparametern (EinCorr), die Anzahl der vollständigen Datensätze dieser Variablen (Anzahl) und letztendlich die daraus berechneten Bewertungszahlen (BWZ).

Die Bewertungszahl berechnet sich nach folgender Gleichung:

BWZ= (vollständige Datensätze / vorhandene Datensätze)×100% + AusCorr - EinCorr

Die jeweils erste dargestellte Variable in Tabelle 11 weist die maximale Korrelation zur Zieleigenschaft auf. Im Beispiel der Querzugfestigkeit liegt die höchste absolute Korrelation mit 0,459 zur Temperatur im Streubunker (Temp_STB) vor. Die höchste theoretische Bewertungszahl liegt bei vollständigen Datensätzen bei 200. In diesem Fall würde der Korrelationskoeffizient zwischen Ziel- und Eingangsvariable bei eins liegen. In der Praxis bestimmt die höchste vorhandene Korrelation zur Zieleigenschaft die maximale Bewertungszahl. Erst bei der Bewertungszahl zur Berechnung der zweiten Eingangsvariablen wird die (maximale) Eingangskorrelation zur ersten Variablen subtrahiert. Möglich ist auch die Subtraktion der mittleren Eingangskorrelation von allen Variablen mit höheren Bewertungszahlen. In der Validierung der Modelle zeigte diese Auswahl jedoch stets geringere Schätzgenauigkeiten im Vorhersageraum. Die kleinste mögliche Bewertungszahl liegt bei vollständigen Datensätzen bei Null. In diesem Fall wäre die Ausgangskorrelation null und die Eingangskorrelation eins. Sind dagegen die Datensätze unvollständig ist auch eine Bewertungszahl kleiner Null als untere Grenze möglich.

Die Bewertungszahlen sind umso größer, je größer die Differenz zwischen Ausgangs- und Eingangskorrelation ist. Damit wird sicher gestellt, dass Variablen mit maximalen "eigenen Informationsgehalten" vorrangig berücksichtigt werden.

Eine statistische Signifikanz einer Variablen in einem Prozessmodell kann aus einer hohen Bewertungszahl nicht zwangsläufig abgeleitet werden. Diese Methode der Variablenauswahl berücksichtigt jedoch, im Gegensatz zu den klassischen statistischen Auswahlverfahren, die Korrelationen zwischen den Eingangsparametern. Auftretende Multikollinearitäten in den Modellen sollten, bei einer Auswahl der Variablen mit maximalen Bewertungszahlen, weitgehend ausgeschlossen sein.

Querzug	AusCorr	EinCorr	Anzahl	BWZ
TEMP_STB	45,9	0	130	145,9
REMARK	43,5	15,3	127	125,9
TEMP_TRA	37,6	17,3	130	120,3
SDR_R9	36,9	29,3	130	107,6
MKDR_5	37,8	31,4	130	106,5
Haert	36,7	35,5	130	101,1
STR_CBVu	25,7	26,2	130	99,6
FUEL_FAB	12,0	17,8	130	94,2
u_Matte	23,74	31,2	130	91,8
STR_BSCH	25,31	34,2	130	91,1
Rohdichte				
DR_L22	37,3	0	130	137,3
TEMP_H6	31,4	21,2	130	110,2
TEMP_L_G	31,8	24,6	130	107,2
TEMP_STB	32,9	25,8	130	107,1
DA_HKS	23,3	27,6	130	95,7
TEMP_H5	16,7	27,6	130	89,1
TEMP_H3	18,5	32,4	130	86,1
STR_CBVu	28,6	42,9	130	85,7
FUEL_FAB	7,0	25,3	130	82,7
FUEL_STB	11,0	30,9	130	80,1
Dickenquellung				
DZ_SCHN	41,5	0	130	141,4
STR_CBVu	40,6	31,4	130	109,2
REMARK	36,5	35,5	127	98,6
Breite_I	18,3	26,9	130	91,4
FUEL_STB	22,2	30,9	130	91,3
FUEL_FAB	14,6	24,3	130	90,4
DAMPF_GE	28,2	38,2	130	90,1
Haert	37,7	49,5	130	88,2
TEMP_STB	23,5	35,5	130	88,1
TEMP_H6	14,6	26,8	130	87,8
Abhebefestigkei	t			
B_AL_D3	41,4	0	130	141,4
REMARK	29,6	16,6	127	110,7
TEMP_STB	30,9	21,7	130	109,2
FUEL_FAB	14,1	16,7	130	97,4
CBV_O_R	24,1	28,6	130	95,4
TEMP_H5	16,7	27,6	130	89,1
Haert	23,4	35,5	130	87,8
u_Matte	23,6	36,2	130	86,6
STR_BSCH	31,7	45,8	130	85,9
TEMP_TRA	27,8	44,1	130	83,7

Tabelle 11: Herleitung der Bewertungszahlen für die Eingangsparameter (Regressoren)

9.1.10 Analyse der Varianzen

Die Anforderungen für verschiedene Einsatzzwecke der Holzwerkstoffe werden in EN-Normen beschrieben. Die Anforderungen an Faserplatten sind in Verbindung mit den Prüfnormen in den Normen EN 622 (1-6) geregelt. Dabei müssen die beschriebenen Anforderungen von den 5% Quantil-Werten (95% Quantil-Werten im Fall der Dickenquellung) erfüllt werden, die aus den Mittelwerten einzelner Platten berechnet wurden. Die Standardabweichung einer Platteneigenschaft zwischen den Plattenmittelwerten ist direkt proportional zu dem 5%-Grenzwert. Je größer die Streuung einer Platteneigenschaft zwischen den Platten, desto höher muss der Eigenschaftsmittelwert der Produktion gefahren werden (Noack, 2003). Somit ist jede Verringerung der Streuung zwischen den Platten mit einer Absenkung des langfristigen Eigenschaftsmittelwertes und damit mit Kostenvorteilen verbunden. Die Streuungen zwischen den Platten werden durch Rohstoff- bzw.-Prozessänderungen hervorgerufen. Diese Streuungen können durch ein ausreichend genaues Prozessmodell erklärt werden.

Dagegen können Streuungen innerhalb der Platten nicht durch die Modelle erklärt werden, da zur Modellierung die Plattenmittelwerte herangezogen werden. Dementsprechend stellen Streuungen innerhalb der Platten eine in den Modellen nicht berücksichtigte Störgröße dar.

Zur Untersuchung der Streuung zwischen und innerhalb der Platten wurden innerhalb von 14 Tagen 33 Platten aus dem Produktionsprozess entnommen, klimatisiert und geprüft. Es wurden je Prüfstreifen 8 Proben für die Abhebefestigkeit und 10 Proben für Dickenquellung, Querzugfestigkeit und Rohdichte über die Produktionsbreite hinweg entnommen. Die Streuungen dieser Untersuchungen sind im oberen Teil der Tabelle 12 in Form des Variationskoeffizienten angegeben. Zum Vergleich wurden Aufzeichnungen der Eigenüberwachung der Firma von 164 Platten über einen Zeitraum von 11 Monaten ausgewertet. Es wurden die Eigenschaften der Platten eines Klebstofflieferanten ausgewertet, um nicht Streuungen zwischen den Lieferanten mit einzubeziehen. Die Streuungen innerhalb der Platten konnten nur für die Eigenschaften Rohdichte und Querzugfestigkeit ermittelt werden, da für die Abhebefestigkeit und Dickenquellung keine Standardabweichungen der Einzelwerte innerhalb der Platten vorlagen.

Prüfung	CV [%]	Rohdichte	Querzug	Abhebefestigkeit	Dickenquellung
klimatisiert	zwischen Platten ¹	1,3	7,4	9,2	6,9
klimatisiert	innerhalb Platten ²	1,6	7,4	13,9	6,6
Industrie	zwischen Platten ¹	1,1	7,4	11,4	10,4
Industrie	innerhalb Platten ²	1,6	6,8	k. A.	k. A.
Industrie	innerhalb Kommission ³	0,6	5,6	6,0	7,0

Tabelle 12: Analyse der Varianzen

1 Variationskoeffizient der Mittelwerte zwischen den Platten

klimatisiert: n = 33 Platten je 8 bzw. 10 Proben

Industrie: n = 164 Platten je 8 Proben

2 durchschnittlicher Variationskoeffizient der Einzelwerte innerhalb der Platten

3 durchschnittlicher Variationskoeffizient zwischen Plattenmittelwerten innerhalb einer Kommission (n=25)

k. A. keine Angabe

Mit Ausnahme der Abhebefestigkeit sind die Streuungen innerhalb und zwischen den Platten in den einzelnen Platteneigenschaften annährend gleich groß. Die Abhebefestigkeit wies

aufgrund eines ausgeprägten Profils über die Plattenbreite eine höhere Streuung innerhalb der Platten auf als zwischen den Platten. Die Streuungen innerhalb einer Fertigung (Kommission) sind deutlich geringer. Ein Teil der Streuungen basiert damit auf Unterschieden zwischen den Kommissionen. Jede Unterbrechung einer Kommission durch einen Plattentypwechsel verursacht zusätzliche Streuungen. Abbildung 10 zeigt den Verlauf der Platteigenschaften über die Plattenbreite. Eingezeichnet sind die normierten Mittelwerte und die Variationskoeffizienten der 33 geprüften Platten. Die Abhebfestigkeit zeigt von außen nach innen einen deutlichen Abfall in der Festigkeit. Die Streuungen der Einzelwerte sind außen höher als in der Plattenmitte. Der maximale Unterschied beträgt zwischen Plattenmitte und Plattenrand 27%. Die Dickenquellung weist in Plattenmitte eine um ca. 12% höhere Quellung als am Plattenrand auf. Die Querzugfestigkeit und die Rohdichte zeigen keinen ausgeprägten symmetrischen Verlauf über die Plattenbreite. Der Verlauf der Querzugfestigkeit über die Plattenbreite korreliert jedoch signifikant mit dem Rohdichteverlauf. Der Grund für die geringere Abhebefestigkeit und die höhere Dickenquellung in Plattenmitte ist vermutlich auf eine höhere Mattentemperatur in der Plattenmitte der Presse zurückzuführen. In den Deckschichten führen diese höheren Temperaturen zu verstärkten Hydrolyseeffekten. Die Querzugfestigkeit bleibt davon unberührt.



Abbildung 10: normierte Mittelwerte (%) und Variationskoeffizienten der Laboreigenschaften über die Produktionsbreite (HDF 8mm, n=33, 100% entspricht minimalen Mittelwert)

Die geringeren Schwankungen der Rohdichte und der Querzugfestigkeit über die Plattenbreite sind u.a. auf eine höhere Kontrolldichte dieser Eigenschaften durch sogenannte Schnelltests zurückzuführen. Die Schnelltests werden i.d.R. zu Beginn einer neuen Kommissionsproduktion durchgeführt, um die Rohdichte und die Querzugfestigkeit über die Plattenbreite zu kontrollieren. Bei Unregelmäßigkeiten wird die Streuung entsprechend ausgeregelt.

9.1.11 Pressprogramm

Abbildung 11 veranschaulicht das Pressprogramm zur Herstellung der Faserplatten. Zu sehen sind die Distanzen und die Pressdrücke über die einzelnen Pressrahmen.



Abbildung 11: Pressprogramm

Deutlich zu sehen sind die einzelnen Abschnitte im Pressprogramm. Nach der ersten Verdichtung (Rahmen 0-3) wird die Presse wieder leicht geöffnet und bis Rahmen 16 relativ konstant gehalten, anschließend erfolgt der zweite Verdichtungsschritt auf Nenndicke (Rahmen 16-20) und letztendlich die Kalibrierzone (Rahmen 20-28).

Da jeder Rahmen nur in festen Grenzen zum vorherigen bzw. nachfolgenden Rahmen geändert werden kann, ist es verständlich, dass die Variablen der Presse besonders hoch miteinander korreliert sind. Die Distanzen und Drücke der linken und rechten Seite sind alle mit einem Korrelationskoeffizienten r>0,95 verbunden.

9.1.12 Hilfsvariablen

Eine Forderung des klassischen linearen Regressionsmodells (Annahme 9) besagt, dass die Regressionskoeffizienten

- 1. feste Größen und
- 2. in der Zeit bzw. im Querschnitt konstante Größen sind.

Teil 1 der Annahme besagt, dass die Regressionskoeffizienten annahmegemäß einen festen Wert und nicht wie eine Zufallsvariable eine Wahrscheinlichkeitsfunktion haben. Dieser feste Wert ist unbekannt und wird als Ziel der Regressionsanalyse als Parameterschätzwert ermittelt. Im verallgemeinerten Regressionsmodell wird diese Annahme aufgehoben.

Teil 2 der Annahme 9 besagt, dass der Regressionskoeffizient als eine in der Zeit konstante Größe betrachtet wird. In Analysen, in denen Daten mit großen Zeiträumen verwendet werden, ist diese Annahme oft verletzt. Diese Verletzungen können im Fall der Holzwerkstoffherstellung auf Wechsel oder Wartung von Maschinenteilen oder Messsystemen, Änderungen in den Anforderungen der Platten oder durch unterschiedliche Lieferanten von Rohstoffen beruhen.

Zur Lösung des Problems können sogenannte Hilfsvariablen (Synonyme: Scheinvariable, dummy variable, künstliche Variable, Indikatorvariable oder Sprungvariable) eingeführt werden.

Der Lern- und Testdatensatz beinhaltet die Variable "Remark". Damit wurde per Handeingabe der Klebstofflieferant gekennzeichnet. Im vorliegenden Datensatz sind zwei Klebstofflieferanten eines UF-Harzes codiert. Mit dieser Variable im Modell wird der Datensatz faktisch in zwei Blöcke eingeteilt. Es wird der Umstand berücksichtigt, dass die beiden Klebstoffe einen unterschiedlich hohen Effekt auf die Zielvariablen ausüben, aber in die gleiche Richtung wirken. Der Regressionskoeffizient einer Dummy-Variable mit zwei Levels erklärt den Unterschied in der Höhe der Zielvariablen der Gruppe "Klebstoff 1" verglichen mit der Gruppe "Klebstoff 2".

Eine Dummy-Variable ist z.B. auch vorstellbar, um den Einfluss verschiedener Holzlieferanten zu bestimmen, insofern diese Unterscheidung konsequent auf dem Holzplatz eingehalten wird. Auch denkbar ist eine Dummy-Variable zur Unterscheidung des verantwortlichen Personals in den einzelnen Schichten. Dummy-Variablen sollten möglichst immer dann verwendet werden, sobald eine Unterscheidung in zwei oder mehrere kategorische Klassen möglich ist. Damit können Varianzen des Fehlers weiter verkleinert werden, insofern diese Unterscheidung einen signifikanten Einfluss auf die jeweilige Eigenschaft besitzt.

Der Vorteil der Verwendung von Dummy-Variablen liegt in der Möglichkeit, Strukturänderungen im Regressionsmodell zuzulassen. Das wiederum bietet den Vorteil den Datensatz nicht aufteilen und verschiedene Modelle bilden zu müssen.

Tabelle 13 zeigt die prozentualen Unterschiede der Laboreigenschaften zwischen den beiden Klebstofflieferanten. Es zeigt sich, dass bei gleicher Rohdichte, die Eigenschaften der Platten des Klebstofflieferanten 1 alle deutlich besser sind als die Platteneigenschaften des Lieferanten 2. Gleichzeitig ist die Produktionsgeschwindigkeit bei Lieferant 1 um ein 1% höher. Die Unterschiede der Mittelwerte der ersten drei aufgeführten Laborwerte sind auf dem 99% Niveau statistisch signifikant.

	Lieferant 1	Lieferant 2
Querzugfestigkeit	100%	108%
Abhebefestigkeit	100%	108%
Dickenquellung	100%	89%
Rohdichte	100%	100%
Pressengeschwindigkeit	100%	101%

Tabelle 13: Qualitätsunterschied in Abhängigkeit vom Klebstofflieferanten (Lerndatensatz)

Würde die Variable zur Unterscheidung des Klebstofflieferanten nicht zur Verfügung stehen, so können diese Unterschiede im Modell nicht erklärt werden. Das Bestimmtheitsmaß des Modells sinkt und die Varianz des Fehlers steigt. Als weitere Folge der Nichtunterscheidung zwischen den Klebstofflieferanten könnte in diesem Fall die Schlussfolgerung gezogen werden, dass mit zunehmender Pressengeschwindigkeit sich die Laborwerte verbessern (siehe Tabelle 13).

Dieses Beispiel verdeutlicht, wie bedeutsam es ist, alle wichtigen Informationen ins Modell einfließen zu lassen. Annahme 8, die Regressormatrix enthält alle wichtigen Bestimmungsfaktoren des Regressanden, wäre bei Nichtverwendung der Dummy-Variable "Remark", verletzt. Die Verletzung dieser Annahme kann beträchtliche Folgen für die Schätzung der Regressionskoeffizienten haben. Diese können verzerrt und mit falschen Vorzeichen geschätzt werden.

Erkenntnisse

Neben der Berechnung einfacher statistischer Kennzahlen, ist die grafische Darstellung der Variablen in Abhängigkeit der Zeit unbedingt zu empfehlen. Diese Darstellung erlaubt es auf sehr effizienter Weise, "verdächtige" Variablen bzw. Beobachtungen zu identifizieren. Des weiteren sind Trends in den Daten sehr schnell erkennbar. Die Untersuchungen zeigten, dass, instationäre Zustände einzelner Variablen, aufgrund von ständigen Weiterentwicklungen bzw. Optimierungen, vorhanden sind. Je größer jedoch die betrachteten Zeiträume sind, desto eher können stationäre Prozessbedingungen unterstellt werden.

Die Korrelationsanalyse und die Berechnung der Bewertungszahlen ermöglicht es, einfache Zusammenhänge zwischen den Variablen zu finden. Die Bewertungszahlen reihen die Prozessvariablen nach ihren "merkmalseigenen Informationen" zu den Platteneigenschaften. Sie stellen eine unverzichtbare Hilfe zur Variablenauswahl als auch zur späteren Interpretation der Modellergebnisse dar.

Die Identifizierung von möglichen Dummy-Variablen sollte möglichst vor Beginn der Datensammlung erfolgen. Mit diesen Variablen können kategoriale Unterschiede in der Modellierung berücksichtigt werden. Im vorhandenen Datensatz konnten signifikante Unterschiede in den Platteneigenschaften, in Abhängigkeit des Leimlieferanten (Variable "Remark"), festgestellt werden.

9.2 Regressionsanalyse

9.2.1 Ziele der Untersuchungen

Zunächst soll der Einfluss verschiedener Variablenauswahlverfahren auf die Schätzgenauigkeit im Vorhersageraum untersucht und verglichen werden. Gesucht wird eine allgemein gültige Methode zur Erstellung empirischer Regressionsmodelle mit maximalen Vorhersagegenauigkeiten für die Echtzeitqualitätskontrolle, bei der die Variablenauswahl auf korrelierte Prozessvariablen beruht.

Des weiteren soll festgestellt werden, inwiefern endogene Regressoren und eine Modellaktualisierung die Schätzgenauigkeit im Vorhersageraum beeinflussen. Durch einen zweiten ausführlicheren Testdatensatz wird die langfristige Stabilität der Modellvorhersagen überprüft und mit den Ergebnissen der Validierung des ersten Testdatensatzes verglichen.

Die Diskussion signifikanter Zusammenhänge zwischen Prozessparametern und den Platteneigenschaften schließt die Untersuchungen zur Regressionsanalyse ab.

9.2.2 Modelle nach der Methode der optimalen Eingänge

9.2.2.1 Einfluss der Anzahl der Regressoren

Abbildung 12 zeigt einen typischen Verlauf der Bewertungszahlen für die Prozessvariablen aus einer kontinuierlichen Holzwerkstoffproduktion. Variablen mit sehr kleinen Bewertungszahlen weisen entweder einen sehr kleinen Korrelationskoeffizienten zum Regressanden und/oder einen hohen Korrelationskoeffizienten zu in der Matrix weiter oben stehenden Variablen auf. Im letzteren Fall besitzen sie ähnliche Informationen wie Variablen mit hohen Bewertungszahlen und stellen im Falle eines Ausfalls dieser Variablen Ersatzgrößen dar.



Abbildung 12: Typischer Verlauf der Bewertungszahlen, dargestellt am Beispiel für die Querzugfestigkeit

Tabelle 14 zeigt die ersten 50 Regressoren mit den höchsten Bewertungszahlen nach der Methode der optimalen Eingänge (siehe auch Kapitel 6.1.7.4 und 9.1.9). Technologisch

offensichtlich nicht mit den Platteneigenschaften in Zusammenhang stehende Variablen, wie z.B. die Silofüllstände und deren Austragsmengen auf dem Holzplatz sind vor der Berechnung der Bewertungsmatrix passiv gesetzt worden und somit nicht mit in die Bewertung eingegangen. Eine Berücksichtigung dieser Variablen führte aufgrund der sehr geringen Korrelationen zu den übrigen Prozessvariablen zu einer hohen Bewertung.

	IB	BWZ	Roh	BWZ	TS	BWZ	Abhe	BWZ
1	TEMP_STB	145,9	DR_L22	137,3	DZ_SCHN	141,4	B_AL_D3	141,4
2	REMARK	125,9	TEMP_H6	110,2	STR_CBVu	109,2	REMARK	110,7
3	TEMP_TRA	120,3	TEMP_L_G	107,2	TEMP_STB	98,7	TEMP_STB	109,2
4	SDR_R9	107,6	TEMP_STB	107,1	REMARK	98,6	CBV_O_R	98,3
5	MKDR_5	106,5	DA_HKS	95,7	TEMP_H6	93,5	FUEL_FAB	97,4
6	Haert	101,1	TEMP_H5	89,1	Breite_I	91,4	TEMP_H5	89,1
7	STR_CBVu	99,6	TEMP_H3	86,1	FUEL_STB	91,3	Haert	87,8
8	FUEL_FAB	94,2	STR_CBVu	85,7	FUEL_FAB	90,3	u_Matte	86,6
9	u_Matte	91,8	FUEL_FAB	82,7	DAMPF_GE	90,1	STR_BSCH	85,9
10	STR_BSCH	91,1	FUEL_STB	80,1	Haert	88,2	TEMP_TRA	83,7
11	POS_ABV	87,2	Dampfm	79,9	u_Matte	87,7	FUEL_KO	83,1
12	TEMP_H1	85,6	Dicke_S4	79,6	DR_R2	87,1	TEMP_H6	82,8
13	TEMP_H6	83,7	CBV_O_R	76,5	MKDR_3	85,0	FUEL_STB	77,8
14	STR_SCH	81,7	DDR_MAHL	74,6	DIST_L6	84,5	Dicke_S4	77,7
15	CBV_O_R	81,5	DR_L4	74,5	STR_SCH	83,5	STR_CBVu	77,5
16	FUEL_STB	81,0	TEMP_ROL	74,4	TEMP_RE3	82,8	TEMP_RE3	77,0
17	u_TRO	80,5	TEMP_RE3	73,0	FL_Waage	81,2	MKDR_6	75,0
18	u_AUS	80,1	FUEL_KO	72,6	TEMP_H5	80,1	DDR_MAHL	73,1
19	Breite_I	80,1	FG_Gew_2	72,3	u_TRO	79,6	TEMP_ROL	73,0
20	B_AL_DR3	79,8	REMARK	72,2	FG_Gew_2	78,6	DR_E_R	70,2
21	TEMP_RE3	79,0	Haert	71,5	B_AL_DR1	77,9	PZF	70,1
22	DDR_MAHL	77,4	Wasser	71,1	CBV_O_R	77,6	DR_L19	69,9
23	TEMP_H5	75,9	DIST_L4	70,8	DR_AL_L	75,7	u_TRO	69,7
24	Dicke_S4	74,9	Breite_I	70,8	TEMP_ROL	74,5	Breite_I	69,2
25	FL_Waage	69,1	MDR_R4	69,8	B_AL_D2	74,1	MDR_R4	69,1
26	TEMP_ROL	68,9	TEMP_H1	69,1	MDR_R4	73,3	TEMP_H4	68,1
27	DA_HKS	68,1	FL_Waage	67,8	DDR_MAHL	73,2	POS_ABV	67,6
28	B_AL_DR1	68,0	POS_ABV	66,4	TEMP_H1	71,3	MKDR_21	67,1
29	FUEL_KO	66,9	DR_E_R	66,0	B_AL_D3	71,3	DR_AL_L	66,7
30	Dampfm	65,8	STR_SH_1	64,0	B_AL_DR4	71,3	FL_Waage	66,3
31	SDR_R22	64,9	STR_BSCH	63,7	PZF	70,9	B_AL_DR1	66,0
32	DZ_DB2	64,3	MKDR_4	63,4	STR_SH_1	70,5	B_AL_D4	65,5
33	FG_Gew_2	63,9	B_AL_DR1	62,9	POS_ABV	70,2	MDR_R23	64,1
34	TEMP_HKS	63,0	B_AL_DR3	62,7	FUEL_KO	67,7	Skalper_2	64,1
35	SDR_R17	61,5	TEMP_H4	62,6	Dicke_S5	66,8	B_AL_DR4	63,7
36	DR_E_R	60,8	DR_L20	62,4	STR_BSCH	65,3	u_Halle	63,5
37	DDR_KOCH	60,2	u_Matte	62,2	DA_HKS	63,2	DDR_KOCH	63,4
38	MDR_R4	59,8	DR_AL_R	61,9	TEMP_TRE	63,1	Dampfm	63,2
39	u_Halle	59,1	u_AUS	61,4	DZ_SCH_2	63,0	STR_SCH	61,8
40	B_AL_D4	58,4	SDR_R13	59,6	Wasser	62,2	FG_Gew_2	60,6
41	MKDR_21	57,6	u_TRO	59,4	DR_E_R	61,2	TEMP_H3	60,5
42	TEMP_FAB	57,6	DZ_SCHN	59,2	MKDR_23	60,8	STR_REF	59,6
43	TEMP_H4	57,5	SDR_R3	58,3	VEN_HKS	59,7	DAMPF_GE	59,1
44	SDR_R1	57,0	TEMP_TRA	58,1	STR_REF	57,9	TEMP_RE1	59,1
45	B_AL_D2	53,7	PZF	58,1	DIST_R19	57,4	MDR_R6	58,7
46	SDR_R3	52,5	STR_RUEW	56,5	MDR_R6	53,9	STR_SH_1	58,5
47	MDR_R21	52,0	MKDR_21	56,2	SDR_R1	53,2	DA_HKS	56,4
48	STR_REF	51,9	VEN_HKS	55,8	MDR_R21	53,0	Wasser	56,2
49	MKDR_3	51,2	B_AL_D2	55,3	STR_RUEW	52,7	TEMP_H2	55,7
50	DR_Mahls	49,9	DIST_L22	53,7	u_Halle	52,7	TEMP_MR	54,1

Tabelle 14: Bewertungsmatrix der Regressoren

Die Rangfolge der optimalen Eingänge ist zwangsläufig von der vorhergehenden Datenanalyse und –bereinigung abhängig. Die erste Variable in der Bewertungsmatrix weist die höchste Korrelation zum jeweiligen Regressanden auf. Bei allen folgenden Variablen wird neben der Korrelation zum Regressanden die maximale Korrelation zu den Variablen mit höheren Bewertungszahlen berücksichtigt. Aufgrund der zum Teil hohen Korrelationen zwischen den Eingangsparametern und den relativ geringen Korrelationen zu den Regressanden fällt die Bewertungszahl im oberen Bereich sehr schnell ab. In Tabelle 15 sind die Standardabweichungen und die Variationskoeffizienten für die Laborqualitätsvariablen der Lern- und Testdaten zu sehen. Aufgrund des längeren Zeitraums der Lerndaten zeigen diese geringfügig höhere Variationskoeffizienten als die Testdaten. Im Mittel zeigt die Dickenquellung mit einem Variationskoeffizienten von ca. 10% die höchsten Schwankungen, gefolgt von der Abhebefestigkeit mit ca. 9% und der Querzugfestigkeit mit ca. 7%. Die Rohdichte weist mit einem Variationskoeffizienten von 1% eine sehr geringe Schwankungsbreite auf. Die Kennzahlen der Tabelle 15 werden zum Vergleich zu den Kennzahlen des Modells im Beobachtungsraum (Lerndaten) und im Vorhersageraum (Testdaten) herangezogen. Erst durch diesen Vergleich lassen sich die Kennzahlen des Modells und der Vorhersage beurteilen. Wie bereits weiter oben beschrieben, sollte u.a. für ein stabiles Vorhersagemodell die Kennzahl RMSE_Prog kleiner sein als die Standardabweichung der Testdaten selbst.

Kennzahlen	N	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	34	0,128	0,846	0,200	0,009
CV_Testdaten	34	6,7	9,2	8,3	1,0
STD_Lerndaten	132	0,142	0,950	0,234	0,010
CV Lerndaten	132	7.4	10.9	10.2	1.1

Tabelle 15: Standardabweichung und Variationskoeffizient der Lern- und Testdaten

Um den Einfluss der Anzahl der Regressoren auf die Schätzgenauigkeit zu bestimmen, wurden zunächst Modelle mit unterschiedlicher Regressorenanzahl berechnet. Das kleinste Modell beinhaltet dabei alle Regressoren bis zu einer Bewertungszahl von 80% und das größte Modell bis zu 50%. In Abhängigkeit der Höhe der Bewertungszahl enthalten die Gleichungen der Modelle unterschiedlich viele Regressoren. Je kleiner die Bewertungszahl desto mehr Variablen werden in die Modellgleichungen aufgenommen (siehe Tabelle 14).

Die Tabelle 16 bis 22 zeigen die beschreibenden Kennzahlen der Modelle im Beobachtungsraum und die ex-ante-Kennwerte im Vorhersageraum in Abhängigkeit der Höhe der Bewertungszahl. Es ist zu beachten, dass die Zahlen in den Tabellen mittlere Kennwerte über den Prognosezeitraum des Testdatensatzes darstellen.

Tabelle 16: Gütekennzahlen Modell 1 (BWZ=50%)

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,71	0,77	0,63	0,74
Adj R-Square	0,54	0,62	0,38	0,53
F-Value	4,38	5,12	2,53	3,55
DF_Model	50	55	57	63
MAD	0,06	0,36	0,12	0,00
MAPE	3,28	4,18	5,11	0,47
RMSE (T-K)	0,096	0,585	0,188	0,0096
RMSE (n)	0,077	0,454	0,145	0,0051
RMS Pct Error	4,08	5,33	6,40	0,57
Korrel Prog	0,54	0,49	0,40	0,28
ME Prog	0,003	-0,069	0,023	0,001
MAD Prog	0,096	0,663	0,173	0,009
MAPE Prog	5,05	7,20	7,27	1,03
RMSE Prog	0,119	0,824	0,220	0,0108
RMS Pct Error_Prog	6,22	8,94	9,25	1,21

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,68	0,71	0,60	0,68
Adj R-Square	0,53	0,57	0,38	0,50
F-Value	4,69	5,26	2,80	3,93
DF_Model	44	45	49	49
MAD	0,06	0,41	0,12	0,00
MAPE	3,40	4,74	5,30	0,51
RMSE (T-K)	0,098	0,615	0,187	0,0070
RMSE (n)	0,080	0,504	0,151	0,0057
RMS Pct Error	4,28	5,95	6,70	0,63
Korrel_Prog	0,56	0,54	0,35	0,31
ME_Prog	-0,005	-0,024	0,018	0,000
MAD_Prog	0,092	0,573	0,188	0,009
MAPE_Prog	4,82	6,23	7,87	1,00
RMSE_Prog	0,113	0,756	0,227	0,0106
RMS Pct Error_Prog	5,94	8,24	9,53	1,19

Tabelle 17: Gütekennzahlen Modell 2 (BWZ=55%)

Tabelle 18: Gütekennzahlen Modell 3 (BWZ=60%)

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,67	0,71	0,52	0,65
Adj R-Square	0,54	0,58	0,32	0,51
F-Value	5,40	5,59	2,59	4,72
DF_Model	38	43	42	40
MAD	0,06	0,41	0,13	0,00
MAPE	3,45	4,75	5,84	0,53
RMSE (T-K)	0,097	0,610	0,196	0,0070
RMSE (n)	0,082	0,506	0,165	0,0059
RMS Pct Error	4,38	5,97	7,34	0,66
Korrel_Prog	0,57	0,55	0,19	0,42
ME Prog	-0,004	-0,028	0,014	-0,001
MAD_Prog	0,085	0,565	0,180	0,008
MAPE Prog	4,52	6,15	7,62	0,89
RMSE_Prog	0,110	0,738	0,227	0,0094
RMS Pct Error_Prog	5,88	8,03	9,90	1,06

Tabelle 19: Gütekennzahlen Modell 4 (BWZ=65%)

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,60	0,67	0,50	0,59
Adj R-Square	0,49	0,56	0,35	0,48
F-Value	5,43	6,02	3,38	5,51
DF_Model	31	36	32	29
MAD	0,07	0,43	0,13	0,01
MAPE	3,79	5,06	5,92	0,60
RMSE (T-K)	0,102	0,626	0,192	0,0072
RMSE (n)	0,089	0,537	0,169	0,0067
RMS Pct Error	4,72	6,41	7,51	0,76
Korrel_Prog	0,55	0,50	0,30	0,48
ME_Prog	-0,004	0,017	0,017	0,000
MAD_Prog	0,084	0,629	0,165	0,007
MAPE_Prog	4,46	6,77	6,92	0,77
RMSE_Prog	0,110	0,784	0,203	0,0086
RMS Pct Error_Prog	5,93	8,34	8,69	0,97

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,58	0,67	0,45	0,57
Adj R-Square	0,49	0,57	0,34	0,47
F-Value	6,75	6,68	4,23	6,05
DF_Model	24	33	23	25
MAD	0,07	0,43	0,14	0,01
MAPE	3,83	5,08	6,16	0,60
RMSE (T-K)	0,101	0,619	0,193	0,0072
RMSE (n)	0,092	0,539	0,176	0,0068
RMS Pct Error	4,84	6,43	7,72	0,77
Korrel_Prog	0,64	0,51	0,34	0,48
ME_Prog	-0,008	0,016	0,041	-0,001
MAD_Prog	0,074	0,635	0,164	0,007
MAPE_Prog	3,91	6,86	6,80	0,74
RMSE_Prog	0,099	0,785	0,197	0,0086
RMS Pct Error_Prog	5,36	8,40	8,12	0,97

Tabelle 20: Gütekennzahlen Modell 5 (BWZ=70%)

Tabelle 21: Gütekennzahlen Modell 6 (BWZ=75%)

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,58	0,64	0,43	0,52
Adj R-Square	0,49	0,56	0,36	0,46
F-Value	6,75	8,61	5,60	8,55
DF_Model	24	24	17	16
MAD	0,07	0,44	0,14	0,01
MAPE	3,85	5,21	6,20	0,62
RMSE (T-K)	0,100	0,650	0,188	0,0073
RMSE (n)	0,092	0,563	0,178	0,0071
RMS Pct Error	4,84	6,80	7,76	0,80
Korrel_Prog	0,64	0,47	0,39	0,47
ME_Prog	-0,009	0,184	0,033	0,000
MAD_Prog	0,074	0,654	0,153	0,006
MAPE_Prog	3,94	7,04	6,37	0,73
RMSE_Prog	0,099	0,826	0,189	0,0083
RMS Pct Error_Prog	5,35	8,83	7,83	0,93

Tabelle 22: Gütekennzahlen Modell 7 (BWZ=80%)

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,57	0,56	0,42	0,49
Adj R-Square	0,50	0,49	0,37	0,44
F-Value	8,01	8,18	7,79	10,23
DF_Model	20	19	12	12
MAD	0,07	0,49	0,14	0,01
MAPE	3,94	5,77	6,22	0,64
RMSE (T-K)	0,101	0,672	0,189	0,0075
RMSE (n)	0,093	0,621	0,180	0,0074
RMS Pct Error	4,92	7,56	7,87	0,83
Korrel_Prog	0,62	0,37	0,39	0,46
ME_Prog	-0,011	-0,013	0,054	-0,001
MAD_Prog	0,078	0,697	0,154	0,006
MAPE_Prog	4,13	7,56	6,37	0,71
RMSE_Prog	0,102	0,844	0,193	0,0083
RMS Pct Error_Prog	5,47	9,15	7,91	0,94

Je kleiner die Bewertungszahl desto mehr Variablen werden in die Modellgleichungen aufgenommen. Wie zu erwarten, steigt mit zunehmender Variablenanzahl in den Gleichungen das Bestimmtheitsmaß und sinkt die geschätzte Standardabweichung des Fehlers im Modell. In Abbildung 13 ist grafisch der Zusammenhang zwischen der Höhe der Bewertungszahl und damit der Anzahl an Variablen im Modell und dem Bestimmtheitsmaß dargestellt. Alle Eigenschaften weisen ca. den gleichen Anstieg des Bestimmtheitsmaßes mit abnehmender Bewertungszahl bei gleichzeitig zunehmender Variablenanzahl auf. Der Anteil

der erklärten Varianz durch die Regressionsgleichung liegt bei der Dickenquellung am höchsten, bei der Abhebefestigkeit am geringsten. Dieser Umstand lässt sich einerseits mit einer höheren Ungenauigkeit der Prüfung und andererseits mit einer stärkeren Schwankung der Abhebefestigkeit über die Produktionsbreite erklären. So zeigten sich über die Produktionsbreite bei der Abhebefestigkeit ein maximaler Unterschied von 27% und bei der Dickenguellung von 12% zwischen Plattenmitte und Plattenrand (siehe Abbildung 10). Da in den Modellgleichungen die Mittelwerte der Eigenschaften über die Produktionsbreite herangezogen werden, stellen starke Schwankungen über die Produktionsbreite eine zusätzliche, nicht berücksichtigte, Fehlerquelle dar. Der Variationskoeffizient der Abhebefestigkeit bzw. der Dickenquellung innerhalb der Platten lag bei einer Stichprobe von 33 geprüften Industrieplatten bei 13,9% bzw. 6,6%. Die maximalen Schwankungen über die Produktionsbreite fallen bei der Querzugfestigkeit und der Rohdichte mit 8% und 3% wesentlich geringer aus als bei den oben aufgezählten Eigenschaften. Der Grund hierfür dürfte in der höheren Kontrolldichte der Querzugfestigkeit und der Rohdichte durch sogenannte Schnelltests und der damit verbundenen Ausregelung dieser beiden Eigenschaften liegen.



Abbildung 13: Zusammenhang zwischen Anzahl der Regressoren in den Gleichungen (bestimmt durch die Bewertungszahl) und dem Bestimmtheitsmaß

Da das Bestimmtheitsmaß mit zunehmender Anzahl an Regressoren zunimmt, ist es nicht für die Auswahl eines Modells geeignet. Auch das adjustierte Bestimmtheitsmaß, dass bei Anfügen von Leervariablen wieder abfällt, ist bei einer Variablenauswahl nach der Methode der optimalen Eingänge für die Bestimmung des optimalen Vorhersagemodells nicht geeignet, da die Auswahl nicht nach der Signifikanz der Variablen erfolgt. Der Aussagewert des Bestimmtheitsmaßes für die Güte eines Vorhersagemodells ist begrenzt. Eine Maximierung des Bestimmtheitsmaßes ist in der industriellen Prozessmodellierung nur dann sinnvoll, solange sich die Schätzgenauigkeit im Vorhersageraum erhöhen lässt. Abbildung 14 zeigt in Abhängigkeit der Höhe der Bewertungszahl die Vorhersagegenauigkeit, als Mittel aller vier Gleichungen, für den Beobachtungs- und Vorhersageraum. Die Schätzgenauigkeit ist als Verhältnis der Standardabweichung der Lern- bzw. Testdaten und des entsprechenden RMSE definiert. Es zeigt sich, das die maximale Schätzgenauigkeit im

Vorhersageraum bei den Modellen zu finden ist, die Variablen bis zu einer Bewertungszahl zwischen 70% und 75% enthalten. Bei Aufnahme weiterer Variablen fällt die Schätzgenauigkeit im Vorhersageraum im Mittel aller vier Gleichungen wieder ab. Einzige Ausnahme davon zeigt die Gleichung der Dickenquellung, deren Schätzgenauigkeit bis zu einer Bewertungszahl von 60% ansteigt. Im Gegensatz zum Vorhersageraum, steigt die Schätzgenauigkeit im Beobachtungsraum mit zunehmender Variablenanzahl in den Gleichungen kontinuierlich an.



Abbildung 14: Einfluss der Anzahl an Regressoren, bestimmt durch die Bewertungszahl, auf die durchschnittliche normierte Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum

Der durchschnittliche prozentuale Fehler der ex-ante-Vorhersagen über den Testzeitraum von 3 Monaten liegt bei der Rohdichte mit ca. 0,7% am geringsten, gefolgt von der Querzugfestigkeit mit ca. 4% und der Dickenquellung und Abhebefestigkeit mit 7% bzw. 6%.

Werden die Schätzgenauigkeiten im Vorhersageraum auf die tatsächlichen Schwankungen bezogen, ergeben sich mit einem Verhältnis von 1,29 für die Querzugfestigkeit die genauesten Vorhersagen. Mit einem maximalen Verhältnis von 1,06 und 1,10 weisen die Eigenschaften Abhebefestigkeit und die Rohdichte die geringsten Schätzgenauigkeiten im Vorhersageraum auf. Tabelle 23 zeigt die Verhältniszahlen für den Beobachtungs- und den Vorhersageraum für die erstellten Modelle.

Die Kennzahlen (Kz) der zweiten Spalte in der Tabelle 23 bedeuten dabei:

- 1 Anzahl der Regressoren im Modell
- 2 Quotient aus Standardabweichung Beobachtungsraum und RMSE_Modell (T-K)
- 3 Quotient aus Standardabweichung Testdatenraum und RMSE_Prog.

Modell	Kz	LAB_IB	LAB_TS	LAB_SS	LAB_MD	Mittelwert
Modell 1 (50%)	1	50	55	57	63	
	2	1,50	1,62	1,21	1,52	1,46
	3	1,07	1,03	0,91	0,85	0,96
Modell 2 (55%)	1	44	45	49	49	
	2	1,49	1,54	1,21	1,48	1,43
	3	1,13	1,12	0,88	0,86	1,00
Modell 3 (60%)	1	38	43	42	40	
	2	1,50	1,55	1,16	1,49	1,42
	3	1,16	1,15	0,88	0,97	1,04
Modell 4 (65%)	1	31	36	32	29	
	2	1,42	1,51	1,18	1,44	1,39
	3	1,16	1,08	0,98	1,06	1,07
Modell 5 (70%)	1	24	33	23	25	
	2	1,41	1,54	1,23	1,41	1,40
	3	1,29	1,08	1,01	1,06	1,11
Modell 6 (75%)	1	24	24	17	16	
	2	1,41	1,53	1,24	1,40	1,39
	3	1,29	1,02	1,06	1,10	1,12
Modell 7 (80%)	1	20	19	12	12	
	2	1,43	1,41	1,20	1,39	1,36
	3	1,25	1,00	1,03	1,10	1,10

Tabelle 23: Schätzgenauigkeiten im Vergleich

Die Ziffer 2 in der Tabelle 23 beschreibt die Anpassung der Daten im Beobachtungsraum, die Ziffer 3 im Vorhersageraum. Da der "Root Mean Square Error" für eine ausreichende Schätzgenauigkeit kleiner sein sollte als die Standardabweichung der Beobachtungs- bzw. der Testdaten, sollte der Quotient entsprechend größer 1 sein. Je größer der Quotient, desto geringer sind die Abweichungen (genauer die Summe der Abweichungsquadrate) zwischen den Vorhersagen und den Laborwerten und desto höher ist die Schätzgenauigkeit einzustufen. Die letzte Spalte (Gesamt) gibt den Mittelwert der Quotienten für das gesamte Gleichungssystem an. Für den Beobachtungsraum ist in der Regel ein Quotient größer 1 immer zu erreichen. Für den Vorhersageraum dagegen ist diese ohne weiteres nicht zu erwarten. Wie in der Tabelle 23 zu sehen, ist dies bei Modellen nach der Methode der optimalen Eingänge tendenziell eher bei kleineren Modellen mit weniger Variablen in den Gleichungen der Fall.

Erst durch die Bildung der oben beschriebenen Quotienten lassen sich die Schätzgenauigkeiten verschiedener Gleichungen bzw. die Ergebnisse verschiedener Anlagen vergleichen.

Die höchsten Korrelationskoeffizienten von 0,64 und 0,54 zwischen ex-ante-Vorhersagen und gemessenen Laborwerten zeigen ebenfalls die Eigenschaften Querzugfestigkeit und Dickenquellung. In den Modellen 5 bis 7 sind die Korrelationskoeffizienten in allen Gleichungen des Modells auf dem 95% Niveau signifikant.

Um den Einfluss des einzelnen Regressors auf die Vorhersagegenauigkeit zu untersuchen, wurden 50 Modelle, beginnend mit einem Ein-Variablen-Modell, berechnet. Entsprechend der Reihenfolge der Bewertungsmatrix der Tabelle 14 wurden die Modelle um je eine





Abbildung 15: Einfluss der Anzahl der Regressoren auf das Bestimmtheitsmaß, F-Wert und die Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum für alle vier Gleichungen

Der höchste Anstieg des Bestimmtheitsmaßes und der Schätzgenauigkeit im Beobachtungsraum erfolgt bei allen vier Gleichungen innerhalb der ersten 10 aufgenommenen Regressoren. Gleichermaßen fällt der F-Wert der Gleichungen in diesem Bereich am stärksten. Die Schätzgenauigkeit im Beobachtungsraum steigt ab 10 Variablen in den Gleichungen für die Dickenquellung, Querzugfestigkeit und Rohdichte stetig, aber für die beiden letztgenannten, nur sehr geringfügig, an.

Die Ergebnisse für die Schätzgenauigkeit im Vorhersageraum fallen je nach Gleichung unterschiedlich aus. Allen Gleichungen gemein ist der deutliche Abfall der Schätzgenauigkeit im Bereich von 40 bis 50 Variablen. Dort zeigt sich das sogenannte "Overfitting" der Modelle sehr deutlich. Die Vorhersagekraft verliert sich mit zunehmender Variablenaufnahme. Die Eigenschaft Querzugfestigkeit zeigt sich über den fast gesamten Bereich der Variablenaufnahme als am genauesten vorherzusagen. Dagegen zeigt die Abhebefestigkeit nur im Bereich zwischen 10 und 24 Variablen eine ausreichende Schätzgenauigkeit, in der die geschätzte Standardabweichung der Prognosen kleiner ist als die Standardabweichung der Testdaten. Bei der Dickenquellung fällt auf, dass die Schätzgenauigkeit im Vorhersageraum sich mit der Aufnahme des 14., 31. und 42 (siehe Pfeile in Abbildung 15 und Tabelle 14). Regressors sprunghaft erhöht. Verantwortlich dafür sind die Variablen "Distanz Rahmen 6", "Presszeitfaktor" und der "Druck Mutipot Rahmen 23". In den dazwischenliegenden Bereichen lässt sich keine signifikante Änderung der Schätzgenauigkeit feststellen. Die Schätzgenauigkeit der Rohdichte fällt ab der Aufnahme des siebten Regressors kontinuierlich und fällt mit der Aufnahme des 38. Regressors unter die geforderte Schätzgenauigkeit. Werden die Ergebnisse der Einzelgleichungen als Gesamtsystem betrachtet, zeigt sich die maximale Vorhersagegenauigkeit im Bereich zwischen 15 und ca. 25 Regressoren. Abbildung 16 zeigt die durchschnittliche Schätzgenauigkeit des Gesamt-gleichungssystems in Abhängigkeit der aufgenommenen Regressoren im Beobachtungsund Vorhersageraum.



Abbildung 16: Einfluss der aufgenommenen Regressoren im Modell auf die durchschnittliche normierte Vorhersagegenauigkeit im Beobachtungs- und im Vorhersageraum

Es bleibt festzuhalten, dass die nach der Methode der optimalen Eingänge entwickelten Mehrgleichungsmodelle, die aufgestellten Kriterien eines Vorhersagemodells erfüllen können. Eine maximale Vorhersagegenauigkeit im Vorhersageraum erreichen diejenigen Modelle, die Regressoren bis zu einer Bewertungszahl von 70 bis 75% aufnehmen. Die Eigenschaften Querzugfestigkeit und Dickenquellung lassen sich von den untersuchten Eigenschaften im Vergleich zu den tatsächlichen Schwankungen im Testdatenraum am genauesten vorhersagen. Die kleinsten prozentualen Abweichungen der ex-ante-Vorhersagen von den Laborwerten weist mit ca. 0,7% dagegen die Rohdichte auf.

9.2.2.2 Einfluss der Signifikanz der Regressoren

Mit dem t-Test wird überprüft, ob der k-te Regressor in der Grundgesamtheit einen Effekt auf den Regressanden hat. Nur statistisch signifikante Regressoren zeigen einen nachweisbaren Effekt auf den Regressanden. Wie hoch der typische Effekt des Regressors zum Regressanden ist, kann durch den standardisierten Regressionskoeffizienten zum Ausdruck gebracht werden. Da die Annahmen des t-Tests u.a. durch die Anwesenheit von stochastischen Regressoren und einer Nichtnormalverteilung der Störvariablen verletzt werden können, sind die Ergebnisse vorsichtig zu interpretieren (Gruber, 1993).

Während der Bildung der weiter oben beschriebenen 50 Modelle, angefangen vom Ein-Variablen bis zum 50-Variablen-Modell (entsprechend der Bewertungsmatrix), wurde die Signifikanz der neu aufgenommenen Variablen jeder Gleichung und deren Einfluss auf die Schätzgenauigkeit im Vorhersageraum untersucht. Als Kennzahl für die Schätzgenauigkeit wurde erneut das Verhältnis der Standardabweichung der Testdaten zur geschätzten Standardabweichung der Vorhersage (RMSE_Prog) herangezogen. Änderte sich bei einer Aufnahme eines Regressors diese Kennzahl bis zur zweiten Kommastelle nicht, wurde dieses als keine Änderung der Schätzung gewertet (Schätzung konstant). Zur Beurteilung der Veränderung der Schätzgenauigkeit bei der Aufnahme des zweiten Regressors wurde die Schätzgenauigkeit des Ein-Variablen-Modells, bei der Aufnahme des dritten Regressors die des Zwei-Variablenmodells, usw. herangezogen. Insgesamt konnten somit 196 Fälle beurteilt werden (49 Modelle mal 4 Gleichungen). Abbildung 17 zeigt die Ergebnisse dieser Untersuchung.





Von den 196 aufgenommenen Regressoren waren insgesamt 46 Regressoren signifikant. Getestet wurde dabei auf einem Signifikanzniveau von 0,90. Von diesen 46 aufgenommenen signifikanten Regressoren erhöhten 23 und verringerten 19 die Schätzgenauigkeit im Vorhersageraum. Bei vier Regressoren zeigte sich keine Änderung. Je größer die Modelle wurden, desto geringer wurde die Wahrscheinlichkeit, dass die aufgenommene Variable eine Signifikanz auf dem getesteten Niveau im Modell aufwies. Das heißt, je kleiner die Bewertungszahl einer Variable ist, desto kleiner ist die Wahrscheinlichkeit einer Signifikanz dieser Variable im Modell. Bis zum 5-Variablen-Modell zeigten sich ausnahmslos alle Regressoren signifikant.

Um zu verstehen, warum signifikante aufgenommene Variablen im Startmodell einmal zu einer Verbesserung und einmal zu einer Verschlechterung der Vorhersagegenauigkeit führen können, wurde neben dem t-Wert zum Zeitpunkt der ersten Vorhersage (Startmodell) auch der t-Wert zum Ende der Vorhersage (hier nach 34 Beobachtungen) untersucht. Von den 19

Fällen, in denen signifikante Regressoren zu einer Verschlechterung der Vorhersagegenauigkeit führten, zeigten 17 Regressoren über den Verlauf der Vorhersagen einen fallenden t-Wert. Die Hälfte davon zeigten am Ende der Vorhersage keine Signifikanz mehr an. Von den 23 signifikanten Regressoren, die zu einer Verbesserung der Schätzgenauigkeit führten, zeigten 20 am Ende der Vorhersage einen größeren t-Wert als zu Beginn der Vorhersagen.

Von den insgesamt 196 untersuchten Fällen war die große Mehrheit der aufgenommenen Variablen (150) nicht signifikant. Davon führten 75 Variablen zu einer Verringerung und 30 zu einer Erhöhung der Schätzgenauigkeit im Vorhersageraum. Bei allen 30 Fällen, die zur Erhöhung der Schätzgenauigkeit führten, waren die t-Werte am Ende der Schätzung größer als zu Beginn der Schätzgenauigkeit im Vorhersageraum führten, der aller größte Anteil am Ende der Vorhersage einen kleinern t-Wert als zu Beginn der Vorhersage. Bei 45 aufgenommenen nichtsignifikanten Variablen war keine Änderung in der Vorhersage-genauigkeit festzustellen.

Nach Annahme 9 des klassischen linearen Regressionsmodells sind die Regressionskoeffizienten feste und in der Zeit konstante Größen. Diese Annahme kann nur gültig sein, solange sich die Variablen über den Zeitraum stationär verhalten. Dies ist im vorliegenden Datensatz nicht für alle Variablen der Fall.

Instationäre Variablen bedeuten für die Regressionsrechnung insofern ein Problem, dass deren Bedeutung für die Regression bzw. die Signifikanz im Modell sich im Laufe des Vorhersagebetriebes ändern können. Zum Beginn der Vorhersage bedeutungslose Variablen können während des Vorhersagebetriebes einen bedeutenden Einfluss erhalten. Sind diese Variablen nicht im Modell enthalten, fehlen diese Informationen für die Vorhersage. Die Vorhersagegenauigkeit nimmt ab. So zeigte z.B. im Startmodell der Presszeitfaktor in der Gleichung der Dickenquellung keinen signifikanten Einfluss auf die Dickenquellung. Dieser Umstand lässt sich durch die sehr geringe Schwankungsbreite dieser Variablen im Lerndatenraum erklären. Während des Vorhersagebetriebes, zeigte der Presszeitfaktor, aufgrund einer starken Änderung im Testdatenraum, einen signifikanten Einfluss auf die Dickenquellung. Eine Aufnahme dieser Variablen ins Modell (trotz ihrer Nichtsignifikanz im Startmodell) führte zu einer Verbesserung der Schätzgenauigkeit im Vorhersageraum. Im Fall der Dickenquellung zeigt sich diese Verbesserung in der Abbildung 15 am sprunghaften Anstieg der Vorhersagegenauigkeit mit der Aufnahme des Presszeitfaktors als 31. Variable im Modell.

Der andere Fall ist genauso möglich. Zu Beginn der Vorhersage für die Regression bedeutende Variablen nehmen während des Vorhersagebetriebes, aufgrund von Veränderungen im System, in ihrer Bedeutung ab. Diese Variablen verursachen dann eine Verringerung der Vorhersagegenauigkeit.

In den oben dargestellten Betrachtungen wurden nur die aufgenommenen Variablen berücksichtigt und aufgrund ihrer Veränderung in der Signifikanz auf die Veränderung in der Vorhersagegenauigkeit geschlossen. Dies ist insofern eine vereinfachende Darstellung, da sich durch die Aufnahme einer Variable ins Modell auch die Signifikanz der schon vorhandenen Variablen ändern kann. Es ist immer das Gesamtsystem für eine Änderung in der Vorhersagegenauigkeit verantwortlich und nicht ausschließlich die aufgenommene Variable. Trotz dieser Vereinfachung lassen sich verschiedene Schlussfolgerungen aus dieser Untersuchung ziehen.

Signifikante Regressoren haben auf die Schätzgenauigkeit im Vorhersageraum einen wesentlich höheren absoluten Einfluss als nichtsignifikante Regressoren. Nichtsignifikante Regressoren sollten, insofern sie keine technologisch bedeutende Kontrollvariablen sind, nicht in das Modell aufgenommen werden, da sie in der großen Mehrheit der Fälle zu einer Verringerung der Vorhersagegenauigkeit führen. Des weiteren besteht die zunehmende Gefahr der Kolliniarität durch im Modell verbleibende nichtsignifikante Regressoren.

Entscheidend ist bei instationären Verhältnissen jedoch nicht die Bedeutung bzw. die Signifikanz der Variablen zum Beginn der Vorhersage, sondern deren Verlauf während der Vorhersage. Dieser Umstand lässt sich durch keine automatisierte Methode der Variablenauswahl berücksichtigen.

Prinzipiell sollte ein Vorhersagemodell alle relevanten Variablen im Modell haben, die sich über die Zeit ändern können. Eine Variablenauswahl, die allein auf die Signifikanz der Variablen im Lerndatenraum beruht, ist für Vorhersagemodelle nicht zielführend.

Eine leicht zu überwachende Kennzahl während des Vorhersagebetriebes ist der F-Wert der einzelnen Gleichungen. Fällt der F-Wert zunehmend ab, so ist das Modell zunehmend weniger gut bestimmt, einzelne Variablen verlieren an Bedeutung für die Regressionsgleichung. Um langfristig stabile Modelle bei instationären Verhältnissen (die in der Industrie kurzfristig immer vorhanden sein werden) zu erreichen, ist neben der Aktualisierung der Koeffizientenmatrix durch Berücksichtigung der neuesten Datensätze eine zusätzliche Auswahl von technologisch relevanten Variablen erforderlich.

Tabelle 24 zeigt die Ergebnisse eines Modells, in dem nur Variablen aufgenommen wurden, deren t-Werte am Ende der Vorhersage größer waren als zu Beginn der Vorhersage. Dieses Vorgehen ist in der industriellen Umgebung in dieser Weise nicht möglich, da der zukünftige Verlauf des t-Wertes im Vorhersagebetrieb nicht bekannt ist. Aus diesem Grund stellt das Ergebnis dieses Modells eine theoretisch erreichbare Vorhersagegenauigkeit dar, die erzielbar wäre, wenn genau diese Variablenauswahl getroffen würde. Eine technologische Interpretation dieses Modells ist jedoch aufgrund der rein statistischen Auswahl nicht möglich. Im Vordergrund der Untersuchung steht die Überprüfung, ob aus dem Verlauf des F-Wertes (und damit aus dem mittleren Verlauf der t-Werte der Modellvariablen) als übergeordnete Modellkennziffer auf die Modellgültigkeit bzw. einer ausreichenden Schätzgenauigkeit geschlossen werden kann.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
R-Square	0,59	0,68	0,50	0,58
Adj R-Square	0,54	0,63	0,43	0,53
F-Value	12,0	13,1	7,4	11,0
DF_Model	15	20	17	16
MAD	0,07	0,42	0,13	0,01
MAPE	3,84	4,90	5,89	0,58
RMSE (T-K)	0,097	0,573	0,179	0,0068
RMSE (n)	0,091	0,551	0,167	0,0067
RMS Pct Error	4,81	6,57	7,34	0,75
Korrel_Prog	0,79	0,79	0,75	0,70
ME Prog	0,003	-0,069	0,024	0,000
MAD_Prog	0,057	0,368	0,105	0,005
MAPE_Prog	3,02	3,99	4,40	0,58
RMSE_Prog	0,078	0,520	0,135	0,0065
RMS Pct Error_Prog	4,20	5,56	5,56	0,73

Tabelle 24: Gütekennzahlen Modells 8

Anhand dieser Variablenauswahl verringert sich der prozentuale Vorhersagefehler für die Querzugfestigkeit auf ca. 3%, für die Dickenquellung und Abhebefestigkeit auf ca. 4% und für die Rohdichte auf rund 0,6%. Auch die Korrelationskoeffizienten zwischen Vorhersage und Laborwerten liegen mit Werten größer/gleich 0,70 sehr hoch. Das Verhältnis zwischen Standardabweichung der Testdaten und RMSE_Prog liegt zwischen 1,40 für die Rohdichte und 1,64 für die Querzugfestigkeit.

Abbildung 18 zeigt den Verlauf der F-Werte der einzelnen Gleichungen über den Vorhersagebetrieb des Modells. Der Anstieg der F-Werte über den Vorhersageraum korreliert hierbei mit der Schätzgenauigkeit im Vorhersageraum. Diese Trendgrafik der F-Werte im Vorhersageraum kann genutzt werden, um die Modellgültigkeit zu überprüfen. Erste Warnhinweise zur Gültigkeit des Modells sind dann gegeben, wenn der F-Wert bei fortlaufender Vorhersage konstant bleibt. Eine Überprüfung des Modells ist dann notwendig, wenn der F-Wert abfällt. In diesem Fall ist davon auszugehen, dass das Modell die Prozessgegebenheiten nicht mehr ausreichend beschreibt.



Abbildung 18: Entwicklung der F-Werte der vier Modellgleichungen über den Zeitraum der ex-ante-Vorhersagen (Modell 8)

Zur weiteren Untersuchung des Einflusses der Signifikanz der Regressoren auf die Vorhersagegenauigkeit wurde, ausgehend von dem Modell mit je 50 Variablen in den Gleichungen (entsprechend der Bewertungsmatrix), sukzessive jeweils die Variable mit dem kleinsten absolutem t-Wert aus jeder Gleichung entfernt. Die Ergebnisse sind grafisch in Abbildung 19 dargestellt.



Abbildung 19: Einfluss der Herausnahme von Regressoren mit dem jeweils kleinstem t-Wert auf das Bestimmtheitsmaß, F-Wert und die Vorhersagegenauigkeit im Beobachtungsund Vorhersageraum

Durch das konsequente Entfernen von Leervariablen bis zu einer Variablenanzahl von ca. 20 verringert sich das Bestimmtheitsmaß nur geringfügig und die Schätzgenauigkeit im Beobachtungsraum steigt an. Der Fehler des Modells wird kleiner. Das adjustierte Bestimmtheitsmaß zeigt das gleiche Verhalten wie der Fehler des Modells. Unterhalb dieser Grenze fällt das Bestimmtheitsmaß als auch die Schätzgenauigkeit im Beobachtungsraum stark ab. Der F-Wert steigt kontinuierlich mit der Verringerung der Variablen an. Die Schätzgenauigkeit im Vorhersageraum (Abbildung 19 rechts unten) steigt bei den Gleichungen der Rohdichte und der Abhebefestigkeit durch das Entfernen der Leervariablen konsequent an. Bei der Querzugfestigkeit und der Dickenquellung ist dieser Effekt nicht sehr stark ausgeprägt. Beide Eigenschaften sind im ex-ante-Vorhersageraum erneut am genauesten vorherzusagen. Durchschnittlich fallen die Schätzgenauigkeiten der einzelnen Eigenschaften bei der sukzessiven Entfernung von Variablen mit den kleinsten t-Werten geringer aus, als bei der schrittweisen Integration der Variablen nach den Bewertungszahlen (Vergleiche Abbildung 15 und Abbildung 19).

Entscheidend für die Schätzgenauigkeit im Vorhersageraum ist auch in dieser Betrachtung nicht die Signifikanz der Variable zum Beginn der ex-ante-Vorhersagen, sondern dessen Verlauf über die Vorhersagezeit. Unabhängig von der Größe des t-Wertes, erhöhte sich die Schätzgenauigkeit im Vorhersageraum beim Entfernen einer Variable aus dem Modell, wenn

Erkenntnisse

größer aus, je höher der t-Wert über diesen Zeitraum ansteigt.

Die Variablenauswahl nach der Methode der optimalen Eingänge ermöglicht es, Modelle zu bilden, die die aufgestellten Kriterien eines Vorhersagemodells erfüllen können. Jede Gleichung des Modells weist dabei eine optimale Anzahl von Variablen auf, bei der die Schätzgenauigkeiten im Vorhersageraum maximal sind. Die höchste durchschnittliche Schätzgenauigkeit wurde bei einer Variablenauswahl bis zu einer Bewertungszahl von ca. 75% erreicht. Dies entspricht im vorliegenden Fall einer Anzahl von 16 bis 24 Variablen für die einzelnen Modellgleichungen. Eine höhere Anzahl an Variablen in den Modell-gleichungen führte überwiegend zu einer Abnahme der ex-ante-Schätzgenauigkeit. Dies wird i.d.R. durch die Aufnahme nichtsignifikanter Leervariablen verursacht. Wie insbesondere die Modellvorhersagen der Dickenquellung zeigten, kann sich jedoch, durch die Aufnahme einzelner Variablen mit kleineren Bewertungszahlen, die Schätzgenauigkeit im Vorhersageraum erhöhen.

Da die Berechnung der Bewertungszahlen und damit die Variablenauswahl auf Grundlage der Lerndaten erfolgt, beschreibt die Variablenauswahl die Verhältnisse in diesem Zeitraum. Werden Prozesseinstellungen während des Vorhersagezeitraums geändert, so sind diese durch die Variablenauswahl auf Basis der Lerndaten nicht berücksichtigt. Je stärker die Änderungen im Vorhersageraum gegenüber dem Lerndatenraum sind, desto mehr verschieben sich die Einflüsse der Prozessvariablen zur Erklärung der Platteneigenschaften. Damit verbunden ist eine Änderung der Signifikanzen der im Modell enthaltenen Variablen. Eine statistische Auswahl nach der Signifikanz der Variablen ist damit für ein Vorhersagemodell nicht zielführend.

Eine Variablenauswahl sollte bei dynamischen (instationären) Prozessen zusätzlich technologisch bedeutsame Variablen enthalten, die nicht durch die Variablenauswahl nach der Methode der optimalen Eingänge berücksichtigt wurden. Die zusätzliche Aufnahme von technologisch bedeutsamen, jedoch statistisch nicht signifikanten Variablen, kann kurzfristig die Schätzgenauigkeit im Vorhersageraum verringern, sollte jedoch über längere Vorhersagezeiträume zu einer genaueren Abbildung des Prozesses führen. Dies ist insbesondere dann der Fall, wenn die Variablen zukünftigen Optimierungen unterliegen. Die Validierungsergebnisse eines Modells sind somit entscheidend vom Zeitraum und Ausmaß der Änderungen im Vorhersageraum abhängig.

9.2.3 Modelle unter Berücksichtigung des Varianzinflationsfaktors VIF

Die nachteiligen Folgen von Kollinearität der Regressoren sind große Varianzen und Kovarianzen der Schätzer für die Regressionskoeffizienten, instabile Regressionskoeffizienten und im Extremfall der Mangel an Schätzbarkeit. Ein bekanntes Kollinearitätsmaß ist der Varianzinflationsfaktor VIF_k. Je größer der Varianzinflationsfaktor, um so stärker ist die Kollinearität zwischen den k-ten Regressor x_k und den anderen Regressoren in X. Nachteilig am VIF ist, das es kein formales Kriterium gibt, ab welcher Höhe des VIF ein Einfluss auf den Regressanden zu erwarten ist. Die vorliegenden Untersuchungen zeigten jedoch, dass Varianzinflationsfaktoren größer vier Vorzeichenwechsel in den Regressionskoeffizienten verursachen können. Die Folgen von starker Kollinearität auf die Schätzung der Regressionskoeffizienten ist anhand eines Beispiels in der Tabelle 25 dargestellt. In dem abschnittsweise gezeigten Modell zur Schätzung der Querzugfestigkeit sind unter anderem die Pressendistanzen der Rahmen 24 und 26 und die Temperaturen im Sichter, im Register 2 des Trockners, am Trocknerausgang als auch im Faser- und Streubunker enthalten. Tabelle 26 zeigt die Korrelationsmatrix der Temperaturen untereinander und zu der Querzugfestigkeit.

Regressor	Regkoeff	t-Wert	stand. Reg	VIF
DIST_L26	1,92	1,91	2,93	785
DIST_L24	-1,97	-1,94	-2,98	789
TEMP_SI	0,38	1,99	15,21	19469
TEMP_RE2	-0,27	-1,95	-14,84	19372
STR_BSCH	0,01	2,85	0,17	1
TEMP_TRA	-0,02	-5,20	-0,31	1
TEMP_FAB	-0,01	-2,98	-0,38	5
Haert	-0,02	-1,73	-0,11	1
TEMP_STB	-0,02	-3,73	-0,39	4
	•			
Regressor	Regkoeff	t-Wert	stand. Reg	VIF
DIST_L24	-0,02	-0,53	-0,03	1
TEMP_RE2	0,01	2,71	0,34	5
STR_BSCH	0,01	2,56	0,15	1
TEMP_TRA	-0,02	-5,29	-0,32	1
TEMP_FAB	-0,01	-2,57	-0,32	5
Haert	-0,02	-1,56	-0,10	1
TEMP STB	-0.02	-3.53	-0.38	4

Tabelle 25: Einfluss von stark korrelierten Variablen in der Regressionsanalyse

Anhand der standardisierten Regressionskoeffizienten zeigt sich in Tabelle 25 eine überragende Bedeutung der Temperaturen im Register 2 und im Sichter, gefolgt von den Distanzen. Beide Paare zeigen jedoch einen entgegengesetzten Einfluss auf die Querzugfestigkeit an. Vom technologischem Standpunkt ist ein unterschiedlicher Einfluss der Rahmen 24 und 26 und der Temperaturen nicht zu erwarten. In der Korrelationsmatrix zeigt jede Temperatur für sich einen negativen Korrelationskoeffizienten zur Querzugfestigkeit. Das Phänomen der wechselnden Vorzeichen im Regressionsmodell ist hier eindeutig auf die hohen Korrelation zwischen den beiden Distanzen und zwischen den Temperaturen zurückzuführen. Die beiden Distanzen weisen untereinander einen Korrelationskoeffizienten von 0,99 (n=132) auf.

	Lab_IB	Temp_Tra	Temp_FAB	Temp_STB	Temp_Si
Lab_IB	1				
Temp_Tra	-0,38	1			
Temp_FAB	-0,42	0,17	1		
Temp_STB	-0,46	0,10	0,83	1	
Temp_Si	-0,30	0,18	0,88	0,84	1
Temp_RE2	-0,30	0,18	0,88	0,84	0,99

Tabelle 26: Korrelationsmatrix (n=132)

Anhand des Variationsinflationsfaktors lassen sich Kollinearitätsprobleme aufdecken. Der VIF zeigt für die Distanzen jeweils einen Wert von 785 bzw. 789 und für die Temperaturen im Register und im Sichter Werte oberhalb von 19000. Es ist zwingend erforderlich, Variablen mit hohen VIF-Werten aus dem Modell zu entfernen bzw. nicht ins Modell aufzunehmen. Die Ergebnisse der Regressionsanalyse nach dem Löschen der Variablen Dist_L26 und Temp_Si zeigt die untere Hälfte der Tabelle 25. Die Varianzinflationsfaktoren weisen nun noch maximale Werte von 5 auf. Die Temperatur im Register 2 zeigt hier ein entgegengesetztes Vorzeichen gegenüber den Temp_RE2 und Temp_FAB zeigen alle Variationsinflationsfaktoren einen Wert von eins an.

Es lässt sich festhalten, dass bei einer Berücksichtigung des Varianzinflationsfaktors, extreme Kollinearitäten im Modell aufgedeckt werden können und damit grobe Fehlspezifikationen des Modells vermieden werden können. Es ist jedoch nicht ausgeschlossen, dass, gleichwohl sehr kleiner VIF-Werte, Kollinaritätsprobleme auftreten können. Zur Auswahl der Variablen ist die Berücksichtigung der Bewertungsmatrix bzw. der Eingangskorrelationen daher unbedingt empfehlenswert. Die Bewertung der Vorzeichen der Regressionskoeffizienten technologisch bedeutsamer Variablen sollte bei der Variablen-auswahl mit berücksichtigt werden. Dieser Bewertung sind jedoch bei Modellen, die auf einer passiv erhobenen Datenmatrix beruhen, Grenzen gesetzt.

In Tabelle 27 ist die Variablenauswahl eines einfachen technologischen Modell dargestellt. Einzige statistische Vorgabe bei der Erstellung dieses Modells war, dass kein Regressor mit einem Varianzinflationsfaktor größer vier in das Modell aufgenommen wurde. Unter dieser Maßgabe war es z.B. nicht möglich, die Leim-, Wasser und die Harnstoffmenge innerhalb eines Modells zu berücksichtigen. Diese Variablen zeigten zu hohe Korrelationen untereinander und verursachten sehr hohe VIF-Werte. Alle vier Gleichungen enthalten 15 Variablen. Berücksichtigt wurde in allen Gleichungen die Leimmenge, die Härtermenge, die Temperatur im Streubunker und am Trocknerausgang, die Mattenfeuchte, die Flächenwaage nach der Streuung, der spezifische Druck am Rahmen 2, 16, 20 bzw. 26 und die Temperaturen des Heizfeldes 6 und der Rollstangen. Weiterhin wurde in den Gleichungen der Querzugfestigkeit, Dickenguellung und Abhebefestigkeit der Leimlieferant, die Stromaufnahme der Beschickschnecke, die Temperatur des Heizfeldes 1 und die Pressengeschwindigkeit berücksichtigt. In der Gleichung der Rohdichte wurde anstatt der vier letztgenannten Variablen die Dampfmenge im Refiner und im Vorkocher, die Skalperhöhe und die Distanz am Pressenauslauf einbezogen.

		IB	MD	TS	SS
1	Remark	✓		✓	✓
2	Leim1	\checkmark	\checkmark	\checkmark	\checkmark
3	V Presse	\checkmark		\checkmark	\checkmark
4	Haert	\checkmark	\checkmark	\checkmark	\checkmark
5	Temp STB	\checkmark	\checkmark	\checkmark	\checkmark
6	Temp Tra	\checkmark	\checkmark	\checkmark	\checkmark
7	u Matte	\checkmark	\checkmark	\checkmark	\checkmark
8	FI Waage	\checkmark	\checkmark	\checkmark	\checkmark
9	SDR RŽ	\checkmark	\checkmark	\checkmark	\checkmark
10	SDR R16	\checkmark	\checkmark	\checkmark	\checkmark
11	SDR_R20	\checkmark		\checkmark	
12	SDR_R26		\checkmark	\checkmark	\checkmark
13	Temp H1	\checkmark		\checkmark	\checkmark
14	Temp_H6	\checkmark	\checkmark	\checkmark	\checkmark
15	Temp_ROL	\checkmark	\checkmark	\checkmark	\checkmark
16	STR_BSCH	\checkmark		\checkmark	\checkmark
17	Dampf		\checkmark		
18	Da_HKS		\checkmark		
19	Dist_A_R		\checkmark		
20	Skalper		\checkmark		

Tabelle 27: Variablenauswahl technologisches Modell

Tabelle 28 zeigt die Kennzahlen des Modells und der Validierung. Die Gleichungen der Querzugfestigkeit und der Dickenquellung weisen mit dieser Variablenauswahl schon eine ausreichende Vorhersagegenauigkeit auf. In den beiden anderen Gleichungen dagegen überschreitet die Schätzung der Standardabweichung in der Vorhersage (RMSE_Prog) noch geringfügig die Standardabweichung der Testdaten.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD Testdaten	0,128	0,846	0,200	0,0091
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,51	0,44	0,33	0,50
Adj R-Square	0,45	0,37	0,25	0,44
F-Value	8,67	6,67	4,10	8,48
DF_Model	15	15	15	15
MAD	0,08	0,57	0,15	0,01
MAPE	4,13	6,56	6,64	0,63
RMSE (T-K)	0,106	0,743	0,205	0,0074
RMSE (n)	0,099	0,697	0,193	0,0072
RMS Pct Error	5,22	8,28	8,54	0,81
Korrel_Prog	0,63	0,50	0,40	0,33
ME_Prog	0,020	0,042	0,077	0,001
MAD_Prog	0,075	0,628	0,162	0,007
MAPE_Prog	3,91	6,84	6,61	0,83
RMSE_Prog	0,101	0,746	0,201	0,0094
RMS Pct Error Prog	5,23	8,12	8,02	1,05

Tabelle 28: Gütekennzahlen Modell 9 (technologisches Modell)

Nachfolgend wurde die Variablenauswahl nach technologischen Gesichtspunkten mit der Auswahl nach der Methode der optimalen Eingänge kombiniert. Dazu wurden die Variablen bis zu einer Bewertungszahl von 75% mit in das technologische Modell integriert, die bis dahin noch nicht berücksichtigt wurden. Multikollinearitätsprobleme wurden derart vermieden, dass alle Variablen mit einem VIF-Wert größer vier aus dem Modell ausgeschlossen wurden. Tabelle 29 zeigt die Ergebnisse des kombinierten Modells.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	0,1280	0,846	0,200	0,0091
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,60	0,68	0,46	0,56
Adj R-Square	0,49	0,60	0,35	0,49
F-Value	5,70	8,36	4,23	7,38
DF_Model	29	29	24	21
MAD	0,07	0,42	0,14	0,01
MAPE	3,78	4,95	6,08	0,58
RMSE (T-K)	0,102	0,595	0,190	0,0071
RMSE (n)	0,090	0,528	0,173	0,0066
RMS Pct Error	4,74	6,29	7,56	0,74
Korrel_Prog	0,65	0,58	0,46	0,43
ME_Prog	0,007	0,036	0,028	0,000
MAD_Prog	0,075	0,551	0,154	0,007
MAPE_Prog	3,91	5,92	6,40	0,77
RMSE_Prog	0,098	0,693	0,180	0,0088
RMS Pct Error_Prog	5,15	7,28	7,51	0,99

Tabelle 29: Gütekennzahlen Modell 10

Die Schätzgenauigkeit im Vorhersageraum konnte durch die ins technologische Modell zusätzlich aufgenommenen Variablen in jeder Gleichung verbessert werden. Das Modell erfüllt alle geforderten Bedingungen: die mittleren Abweichungen der ex-ante-Vorhersagen von den Laborwerten sind nahe Null, die Korrelationen zwischen den Vorhersagen und den Laborwerten sind auf dem 95% Niveau signifikant und die geschätzte Standardabweichung der Prognosen ist kleiner als die Standardabweichung der Testdaten (Quotient ist größer eins). Ein Vergleich des kombinierten Modells zu dem Modell 6, in der nur die Variablen bis zu einer Bewertungszahl von 75% ins Modell aufgenommen wurden, zeigt für das kombinierte Modell eine höhere Schätzgenauigkeit für die Gleichungen der Dickenquellung und Abhebefestigkeit. Die Gleichung der Querzugfestigkeit zeigt im kombinierten Modell keine wesentlichen Verbesserungen, die Schätzgenauigkeit der Rohdichte ist im kombinierten Modell etwas geringer.

Insgesamt ist das kombinierte Modell den anderen Modellen vorzuziehen. Die zusätzliche Auswahl bedeutender technologischer Variablen, die zum großen Teil auch Kontrollvariablen darstellen, lässt das Modell besser interpretieren und wird langfristig eine bessere Stabilität aufweisen. Eine Verbesserung der Schätzgenauigkeit durch zusätzlich technologisch ausgewählte Variablen ist allerdings nicht immer gegeben.

Abbildung 20 veranschaulicht grafisch den Verlauf der Vorhersagen der Querzugfestigkeit, Dickenquellung, Abhebefestigkeit und der Rohdichte über den Bebachtungsraum und den Vorhersageraum. Dabei wurden die Eigenschaften auf den Mittelwert der jeweiligen Datenreihe (=100%) normiert. Es ergeben sich dabei keine Veränderungen bzw. Verzerrungen zur Darstellung der Originalwerte. Die unterste Kurve in den Grafiken zeigt jeweils den Verlauf der Residuen in prozentualer Abweichung vom Laborwert und sind auf der sekundären y-Achse abgetragen.





Abbildung 20: Qualitätsvorhersagen für die Platteneigenschaften (Modell 10)

Die Abbildung 20 zeigt, dass der generelle Trend sowohl in der ex-post- als auch in der exante-Vorhersage durch das Modell nachvollzogen wird. In der Abbildung zur Abhebefestigkeit wird deutlich, warum das Bestimmtheitsmaß so gering ausfällt. Die gesamten Minimas und Maximas ("Spitzen") in den Laborwerten werden durch das Modell in diesen Extremen nicht modelliert. Es ist anzunehmen, dass diese Spitzen nicht die "wahren" Werte sind, sondern vor allem durch Fehler in der Materialprüfung verursacht werden. Es ist in der Holzwerkstoffqualitätskontrolle bekannt, dass die Prüfung der Abhebefestigkeit den höchsten Fehler in der Prüfung aufweist.

Erkenntnisse

Der Varianzinflationsfaktor VIF ist ein wichtiges Hilfsmittel in der Variablenauswahl zur Aufdeckung von Multikollinearitäten in den Regressionsmodellen. Im Modell nicht entdeckte Multikollinearitäten führen u.a., durch Vorzeichenwechsel in den Regressionskoeffizienten (die rein mathematisch bedingt sind), zu einer falschen Interpretation der Modellzusammenhänge. In den Modellen auf Basis des vorliegenden Datensatzes waren Auswirkungen der Multikollinearität in Form von Vorzeichenwechsel der Regressionskoeffizienten bis zu einem Varianzinflationsfaktor von 5 nachweisbar.

Eine Modellauswahl nach rein technologischen Gesichtspunkten unter Berücksichtigung eines Varianzinflationsfaktors kleiner 5 führte in den Gleichungen der Abhebefestigkeit und der mittleren Rohdichte noch zu einer ungenügenden Schätzgenauigkeit im Vorhersageraum. Die Integration technologisch ausgewählter Variablen in die Variablenauswahl nach der Methode der optimalen Eingängen bis zu 75% führte unter der Berücksichtigung des Varianzinflationsfaktors zu einem Modell, das alle aufgestellten Kriterien eines Vorhersagemodells erfüllte. Im kombinierten Modell wurden nur Variablen berücksichtigt, deren Varianzinflationsfaktoren kleiner 5 waren.

9.2.4 Modelle nach klassischer Variablenauswahl

Die klassischen Variablenauswahlverfahren beruhen auf eine Maximierung des Bestimmtheitsmaßes (oder des adjustierten Bestimmtheitsmaßes) bzw. auf die Auswahl der Variablen nach deren Signifikanzniveau. Die am meisten genutzten Verfahren sind dabei die rückwärtige (backward) und die schrittweise (stepwise) Auswahl der Variablen. Draper und Smith (1998) empfehlen für praktische Anwendungen die schrittweise Auswahl.

Modelle nach diesen Auswahlverfahren zeigten jeweils im Beobachtungsraum eine sehr hohe Anpassung, konnten aber im Vorhersageraum nicht die Schätzgenauigkeiten der nach der Methode der optimalen Eingänge gebildeten Modelle erreichen (Hasener, 2003). Dabei konnte beobachtet werden, dass die Bestimmtheitsmaße mit den ersten ex-ante-Vorhersagen sehr schnell abfielen. Dies wird durch eine Überanpassung der Lerndaten verursacht. Die Modelle wiesen keine Stabilität über den Vorhersageraum auf. Die Auswahl der Variablen wurde dabei wesentlich durch das Signifikanzniveau bestimmt. Die Stepwise-Modelle zeigten dabei mit weniger Variablen in den Gleichungen eine höhere Schätzgenauigkeit im Vorhersageraum als die Backward-Modelle. Die aufgestellten Kriterien für ein Vorhersagemodell konnten mit beiden Methoden nicht erfüllt werden.

Beide Methoden eignen sich nicht für stark korrelierte Datenmatrizen. Bei der Auswahl werden Korrelationen zwischen den Variablen nicht berücksichtigt. So wurden bevorzugt Variablen aus der Heißpresse in den Modellen aufgenommen, die stark miteinander korreliert sind. Die Varianzinflationsfaktoren zeigten schon bei einer geringen Variablenanzahl deutlich Multikollinearitäten an. Eine technologische Interpretation dieser Modelle ist nicht möglich.

Kruse (1997) berichtet, dass mit der abgewandelten schrittweisen Variablenauswahl (MaxR), in der das Bestimmtheitsmaß der Modelle maximiert wird, die besten Vorhersagen der Querzugfestigkeit erreicht wurden. Das Gütekriterium war dabei ein minimales RMSE, dessen Berechnung auf die durch Cross-Validierung (leave-one-out) gewonnenen Residuen basierte. Bei der Modellbildung wurde dabei auf den vorherzusagenden Datensatz verzichtet. Diese Validierungsmethode ist verwendet worden, da nicht ausreichend externe Datensätze zur Validierung vorlagen. Die dargestellten univariaten Modelle zeigen sehr hohe Bestimmtheitsmaße (0,75-0,95).

Eigene Untersuchungen zeigen, dass eine Beurteilung der Schätzgenauigkeit für den exante-Vorhersageraum auf Basis der leave-one-out-Cross-Validierung stark überschätzt wird. Es wurden mittels der Variablenauswahlprozedur "MaxR" Modelle mit 8 bis 35 Variablen gebildet. Die Schätzgenauigkeit der Modelle wurde zum einem anhand des Testdatensatzes und zum anderen durch die Cross-Validierung der Lerndaten ermittelt. Zur Variablenauswahl wurde dabei jeweils der gesamte Lerndatensatz herangezogen. Die Variablenauswahl blieb für das jeweilige Modell über die Cross-Validierung hinweg konstant. Tabelle 30 zeigt die Ergebnisse exemplarisch für die Modelle der Querzugfestigkeit.

Der Fehler des Modells (Beobachtungsdaten) sinkt mit zunehmender Variablenanzahl. Der Fehler der Cross-Validierung liegt größenmäßig zwischen dem Fehler des Modells und der Testdaten. Er sinkt bis zu einer Variablenanzahl von 20 und bleibt dann annährend konstant. Der Fehler der ex-ante-Vorhersagen fällt dagegen bis zu einer Variablenanzahl von 14 und
steigt anschließend mit weiteren Variablen im Modell wieder an. Hier zeigt sich deutlich, dass die Schätzgenauigkeit für den Vorhersageraum auch durch Beurteilung des Fehlers aus der Cross-Validierung noch überschätzt wird. Die Standardabweichung der Beobachtungsdaten lag bei 0,142. Nach der Beurteilung der Schätzgenauigkeit durch eine Cross-Validierung würde das Modell mit 30 Variablen und einem Bestimmtheitsmaß von R²=0,83 zur Vorhersage ausgewählt werden. Dieses Modell erreicht im Testdatensatz jedoch nicht die gestellten Anforderungen. Der RMSE_P ist mit 0,150 größer als die Standardabweichung der Testdaten.

Anzahl Variablen	RMSE	RMSE	RMSE _P	Bestimmt-
Im Modell	Modell ⁽¹⁾	Cross-Validierung ^(1, 2)	ex-ante-Vorhersage ⁽¹⁾	heitsmaß
35	0,079	0,089	0,153	0,85
30	0,081	0,089	0,150	0,83
20	0,085	0,090	0,142	0,76
18	0,086	0,092	0,130	0,74
16	0,087	0,094	0,117	0,72
14	0,092	0,101	0,110	0,66
12	0,095	0,101	0,113	0,64
10	0,097	0,102	0,113	0,61
8	0,101	0,105	0,122	0,57

Tabelle 30: Vergleich der Validierungsergebnisse der MaxR-Modelle

(1) allen Berechnungen liegt der tatsächliche Freiheitsgrad T-K zugrunde

(2) Summe der Fehlerquadrate ist der PRESS-Statistik der Prozedur Prog Reg entnommen (OLS-Residuen)

Nach der Validierung der Modelle durch einen externen Datensatz zeigt das Modell mit 14 Variablen und einem Bestimmtheitsmaß von R²=0,66 die höchste Schätzgenauigkeit im Vorhersageraum.

Im Vergleich zu den Modellen nach der Methode der optimalen Eingänge zeigen die MaxR-Modelle bei gleicher Regressorenanzahl eine wesentlich höhere Anpassung im Beobachtungsraum. Dieser Effekt beruht auf der Auswahl der Variablen nach der maximalen Signifikanz bzw. Bestimmtheitsmaßes in der MaxR-Methode. Im Vorhersageraum fällt die durchschnittliche Schätzgenauigkeit geringer aus als bei den Modellen nach der Methode der optimalen Eingänge. Die Schere in der Schätzgenauigkeit zwischen dem Beobachtungsraum und Vorhersageraum öffnet sich mit fortschreitender Variablenanzahl zunehmend. Abbildung 21 zeigt für das gesamte Gleichungssystem die auf die entsprechenden Standardabweichungen bezogenen durchschnittlichen Vorhersagegenauigkeiten im Beobachtungs- und Vorhersageraum.

Eine deutliche Abnahme der Schätzgenauigkeit im Vorhersageraum beginnt im Vergleich zu den Modellen nach der Methode der optimalen Eingänge bei den MaxR-Modellen bei einer geringeren Variablenanzahl. Dieses Ergebnis wird durch die unterschiedlichen Methoden in der Variablenauswahl verursacht. Bei der MaxR-Methode werden Variablen ausgewählt, die zu einer Maximierung des Bestimmtheitsmaßes führen. Bei dieser Auswahl werden Korrelationen zwischen den Eingangsvariablen nicht berücksichtigt. Dies führt zur Aufnahme von Variablen, die eine scheinbare Signifikanz aufweisen. Dieses Phänomen konnte ab einer Anzahl von 9 Variablen beobachtet werden. Ab dieser Variablenanzahl zeigten erste Variablen Varianzinflationsfaktoren größer 5. Mit einer weiteren Variablenaufnahme erhöhten sich zunehmend die Varianzinflationsfaktoren der Modellvariablen. Die Schätzgenauigkeit im Vorhersageraum nimmt zunehmend ab.



Abbildung 21: Einfluss der Regressorenanzahl im Modell auf die durchschnittliche Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum (Methode MaxR)

Das Verhalten der MaxR-Modelle ist denen der Stepwise-Modelle im Beobachtungs- und Vorhersageraum ähnlich. Die aufgeführten Aussagen über die Stepwise- und Backward-Modelle gelten uneingeschränkt auch für die MaxR-Modelle. Eine Variablenauswahl für die industrielle Prozessmodellierung nach der Methode der Regressionsanalyse muss die Korrelationen zwischen den Eingangsparametern berücksichtigen. Andernfalls können keine Modelle entwickelt werden, die ausreichend hohe und stabile Schätzgenauigkeiten im Vorhersageraum liefern. Eine Modellinterpretation ist zudem nicht möglich.

Tabelle 31 zeigt die Kennzahlen für das MaxR-Modell mit 35 Variablen in jeder Gleichung für den Beobachtungs- und Vorhersageraum. Die Kennzahl RMSE_Prog ist in jeder Gleichung höher als die Standardabweichung der Testdaten. Die ungenügende Schätzgenauigkeit dieses Modells lässt sich auch am Verlauf der F-Werte über den Vorhersagezeitraum nachvollziehen. Abbildung 22 zeigt den Verlauf der F-Werte. Mit Beginn der ex-ante-Vorhersagen ist eine Abnahme der F-Werte zu sehen. Diese stetige Abnahme der F-Werte weist auf eine Überanpassung der Daten im Beobachtungsraum hin, die immer mit einer ungenügenden Vorhersagegenauigkeit einhergeht.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	0,128	0,846	0,200	0,009
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,77	0,81	0,68	0,76
Adj R-Square	0,69	0,75	0,57	0,68
F-Value	10,52	13,76	6,50	9,87
DF_Model	35	35	35	35
MAD	0,06	0,32	0,11	0,00
MAPE	2,95	3,67	4,72	0,45
RMSE (T-K)	0,079	0,470	0,155	0,0057
RMSE (n)	0,068	0,406	0,134	0,0054
RMS Pct Error	3,62	4,77	5,93	0,60
Korrel Prog	0,46	0,26	0,39	0,32
ME_Prog	0,019	-0,037	0,013	0,001
MAD Prog	0,128	0,821	0,204	0,008
MAPE Prog	6,65	8,89	8,58	0,94
RMSE Prog	0,153	0,998	0,284	0,0108
RMS Pct Error Prog	7,87	10,87	12,05	1,21



Abbildung 22: Entwicklung der F-Werte der vier Modellgleichungen über den Zeitraum der ex-ante-Vorhersagen (Modell 11)

Die Folgen einer Überanpassung der Lerndaten durch ein Modell sind in Abbildung 23 grafisch am deutlichsten anhand der Residuen zu sehen. Innerhalb des Lerndatenraums schwankt der Fehler sehr dicht um die Nulllinie. Die ex-post-Prognosen stimmen sehr gut mit den Labordaten überein. Im Testdatenraum dagegen schwankt der Fehler in wesentlich größeren Wellen um die Nulllinie. Der Vorhersageraum wird mit der Variablenauswahl ungenügend beschrieben.



Abbildung 23: Vorhersagen der Querzugfestigkeit (Modell 11)

Erkenntnisse

Die Fähigkeit zur Beurteilung eines Modells zur zukünftigen Vorhersage durch die Methode der Cross-Validierung wird stark überschätzt. Diese Überschätzung der Vorhersagegenauigkeit ist um so größer einzustufen, je größer die Unterschiede der Prozesseinstellungen im Beobachtungs- und im Vorhersageraum sind.

Die klassischen statistischen Variablenauswahlverfahren sind für korrelierte Datenmatrizen zur Erstellung von Vorhersagemodellen nicht geeignet. Die Auswahlverfahren berücksichtigen nicht die Korrelationen zwischen den Prozessparametern. Dabei kommt es zu einer Variablenauswahl, deren Signifikanzen insbesondere durch die Korrelationen der im Modell enthaltenen Variablen verursacht werden. Es sind "Scheinsiginifikanzen", die bei Entfernung einer einzigen Variable kollabieren. Diese Auswahl führt insgesamt zu einer Überanpassung ("overfitting") der Beobachtungsdaten. Da diese Verfahren voraussetzen, dass die Regressoren frei von Fehlern sind, werden die realen Mess- und Zuordnungsfehler durch die Variablenkombination mit angepasst. Im Vorhersageraum wird der im Beobachtungsraum mitmodellierte Fehler übertragen. Die Vorhersagen sind entsprechend ungenügend.

9.2.5 Untersuchung der Residuen

Anhand der Residuen lassen sich die Annahmen über die Störvariablen am zweckmäßigsten überprüfen. Da die Residuen den nichtmodellierten Anteil darstellen, sind durch die Prüfung der Residuen Fehlspezifikationen des Modells leicht aufzudecken. Sind in den Residuen noch Strukturen erkennbar, so kann davon ausgegangen werden, dass das Modell nicht richtig spezifiziert ist.

Das klassische lineare Regressionsmodell setzt verschiedene Annahmen über die Störvariablen voraus. Zum einen wird vorausgesetzt, dass der Erwartungswert der Störvariablen Null ist, die Störvariablen gleiche Varianzen aufweisen, unabhängig voneinander (keine Autokorrelation) und mehrdimensional normalverteilt sind. Letztere Annahme ist für das Modell selbst nicht Voraussetzung, sondern für die Anwendung des tund F-Tests. Neben verschiedenen numerischen Tests, eignen sich insbesondere grafische Darstellungen der Residuen zur Überprüfung der Annahmen.

Zur Überprüfung der Verteilung der Residuen eignen sich die Darstellungen des Histogramms und des normal plots (siehe Kapitel 9.1.8). Abbildung 24 zeigt die Residuen des Modells 10 in diesen Darstellungen. Alle Residuen der einzelnen Gleichungen zeigen eine sehr gute Annäherung an die Normalverteilung. Gleichwohl zeigen sich an den Rändern gewöhnlich einige Abweichungen.



Abbildung 24: Residuen des Modells 10 (n=132) als Histogramm und normal plot

Zur Überprüfung von Zeiteffekten (Autokorrelationen) und der Varianzhomogenität eignet sich die Darstellung der Residuen gegenüber der Zeit bzw. der ex-post bestimmten Vorhersagen. In diesen Darstellungen sollten die Residuen kein zeitliches Muster bzw. ein möglichst rechteckiges Band bilden. In Abbildung 25 wird auf der linken Seite das geforderte ideale Verhalten der Residuen gezeigt, die rechte Abbildung dagegen zeigt eine nichtkonstante Varianz (Heteroskedastizität) in den Residuen an.



Abbildung 25: Homoskedastizität (links) und Heteroskedastizität (rechts) der Residuen (vereinfachende Darstellung)

Abbildung 26 zeigt den tatsächlichen Verlauf der Residuen des Modells 10 über den zeitlichen Verlauf der Lerndaten. In allen Grafiken kann eine Varianzhomogenität der Störvariablen unterstellt werden. Es zeigen sich bis auf kleinere "Ausreißer" keine besonderen Auffälligkeiten.



Abbildung 26: Verlauf der Residuen über die Zeit

In den obigen Grafiken ist keine eindeutige Autokorrelation der Residuen ersichtlich. Es ist jedoch zweckmäßig, auch den numerischen Durbin-Watson-Test zur Überprüfung von Autokorrelationen der Residuen zu berechnen. Nach dem Test liegt keine Autokorrelation der Residuen vor, wenn der Durbin-Watson-Koeffizient gleich zwei ist (Draper und Smith, 1998). Tabelle 32 zeigt die berechneten Durbin-Watson-Koeffizienten zu Beginn und Ende der exante-Vorhersagen des Modells 10. Anhand dieser Koeffizienten lässt sich auf dem 95% Signifikanzniveau keine serielle Korrelation der Residuen über den Vorhersageraum feststellen.

	DW-Koeffizient	DW-Koeffizient
	Beginn Vorhersage	Ende Vorhersage
Querzugfestigkeit	1,96	1,91
Dickenquellung	1,81	1,85
Abhebefestigkeit	2,34	2,38
Rohdichte	2,40	2,18

Tabelle 32: Durbin-Watson- (DW) Koeffizienten des Modells 10

Eine weitere elegante grafische Methode zur Überprüfung von seriellen Korrelationen besteht in der Darstellung der Residuen gegenüber den Residuen der vorhergehenden Periode (lag=1). Eine positive Autokorrelation zeigt sich in einer positiv ansteigenden Punktwolke, eine negative Autokorrelation dagegen in eine negativ fallende Punktwolke. Abbildung 27 zeigt beispielhaft die unkorrelierte Darstellung der Residuen gegenüber den um eine Periode vorhergehenden Residuen für die Gleichung der Querzugfestigkeit des Modells 10.



Abbildung 27: Eine unkorrelierte Serie von Residuen (IB) des Modells 10

Das Auftragen der Residuen gegenüber den Vorhersagen des Modells ließ ebenfalls keine Unregelmäßigkeiten in den Residuen erkennen. Auf Grundlage der vorhandenen Daten sind die Annahmen über die Störvariablen für das Modell 10 nicht abzulehnen. Es lassen sich anhand der Residuen keine Fehlspezifikationen des Modells feststellen. Es kann von einer Normalverteilung und einer homogenen Varianz der Residuen ausgegangen werden. Serielle Korrelationen der Residuen können in Abhängigkeit des Daten auftreten, sollten aber vernachlässigbar klein sein. Die Annahme, ob der Erwartungswert der Störvariablen gleich Null ist, lässt sich leicht anhand des Mittelwertes der Residuen überprüfen.

Erkenntnisse

Die Überprüfung der Residuen stellt eine der wichtigsten Methoden zur Aufdeckung von Fehlspezifikationen eines Modells dar. Methoden zur Untersuchung der Residuen sind in der Literatur ausführlich beschrieben. (siehe Draper und Smith, 1998). Neben ausführlichen numerischen Tests eignen sich jedoch insbesondere grafische Methoden zur Überprüfung der Residuen. Jegliche erkennbare systematische Strukturen in den Residuen offenbaren Fehler in der Modellspezifikation.

Grundsätzlich sollten die Residuen auf Zeiteffekte, Normalverteilung, Varianzinhomogenitäten und Krümmungen untersucht werden. Dazu können die Residuen auf der vertikalen Achse gegen die Zeit (wenn bekannt), den ex-post-Vorhersagen und gegen die verschiedenen Regressoren aufgetragen werden. Zur Überprüfung der Normalverteilung der Residuen wird die Darstellung des Histogramms bzw. des normal plots verwendet.

Die Residuen des kombinierten Modells zeigten keine Zeiteffekte (Autokorrelationen), keine Varianzinhomogenitäten und keine Krümmungen, die auf Modellglieder zweiter oder höherer Ordnung hinweisen. Die Residuen aller Modellgleichungen waren normal verteilt.

9.2.6 Endogene Regressoren

Die Verwendung endogener Variablen als Regressoren ist in der Ökonometrie weit verbreitet. In wirtschaftstheoretischen Modellen sind sowohl Dependenzen als auch Interdependenzen, also wechselseitige Beziehungen, oft zu beobachten. Es ergeben sich z.B. wechselseitige Beziehungen zwischen dem Zinsniveau und der Investitionstätigkeit. Beide Größen werden in eigenen Gleichungen modelliert und aufgrund der Interdependenz gegenseitig als Regressor in der anderen Gleichung verwendet. Dieses Vorgehen ist für den Prozess der Holzwerkstoffherstellung übertragbar und wurde erstmals von Lobenhoffer (1982) vorgeschlagen.

Abhängigkeiten zur mittleren Rohdichte zeigen prinzipiell alle technologischen Eigenschaften der Holzwerkstoffe. Diese Abhängigkeit ist jedoch nur einseitig vorhanden. Die mittlere Rohdichte wird nicht von den technologischen Eigenschaften selbst beeinflusst. Dieser Zusammenhang wird als dependent bezeichnet.

Echte Inderdependenzen im eigentlichen Sinne sind zwischen den technologischen Eigenschaften der Holzwerkstoffplatten nicht vorhanden. Jedoch zeigen die Korrelationen zwischen den Laboreigenschaften einen ausgeprägten Zusammenhang (Tabelle 9). Am stärksten ausgebildet ist dabei die Korrelation zwischen der Abhebefestigkeit und der Querzugfestigkeit. Dieser Zusammenhang wird in der Industrie insofern ausgenutzt, dass auf

Grundlage sogenannter Querzug-Schnelltests auf die Einhaltung bzw. Verletzung der Zielvorgaben für die Abhebefestigkeit geschlossen wird. Wird ein bestimmter Grenzwert in der Querzugprüfung nicht unterschritten, so wird in aller Regel auch die Vorgabe für die Abhebefestigkeit eingehalten. Die Korrelation ist aber ein von der Kausalität unabhängiger, d.h. ungerichteter Zusammenhang zwischen verschiedenen Merkmalen (Rasch und Verdorren, 2003). Auf Basis eines signifikanten Korrelationskoeffizienten kann nicht auf eine wechselseitige Abhängigkeit geschlossen werden. Technologisch ist dies auch einsichtig. Trotz der hohen Korrelation zwischen Querzug- und Abhebefestigkeit, beeinflusst weder die Querzugfestigkeit die Abhebefestigkeit noch besteht der kausale Zusammenhang in der anderen Richtung. Der Zusammenhang zwischen diesen beiden Eigenschaften ist darauf zurückzuführen, dass beide Eigenschaften durch die Prozessvariablen in ähnlicher Weise beeinflusst werden. Da die technologischen Eigenschaften an dem gleichen Materialstrang getestet werden, wirken die Prozessvariablen in gleicher Weise auf den Probenkörper.

Die Verwendung endogener Variablen in den Vorhersagemodellen und damit die Nutzung der direkten Korrelationen zwischen den technologischen Eigenschaften, ist im Einzelfall zu prüfen. Im vorliegenden Fall konnte durch die Verwendung endogener Variablen keine Verbesserung in den ex-ante-Vorhersagen erzielt werden. Dazu wurden die endogenen Variablen in den verschiedenen, weiter oben dargestellten Modellen, eingesetzt. Dies ist dadurch zu erklären, dass durch die Aufnahme der endogenen Regressoren keine wesentlich neuen Informationen ins Modell gelangten. Die Informationen waren durch die exogenen Regressoren schon jeweils im Modell enthalten. So verursachte z.B. die Aufnahme der Querzugfestigkeit als endogener Regressor in die Gleichung der Abhebefestigkeit des Modells 10 eine geringe Abnahme der Schätzgenauigkeit im Vorhersageraum. Die Querzugfestigkeit zeigte keine Signifikanz. Wurden dagegen nach der Aufnahme der Querzugfestigkeit in die Gleichung der Abhebefestigkeit alle Variablen in dieser Gleichung gelöscht, die schon in der Gleichung der Querzugfestigkeit enthalten waren, zeigte die Querzugfestigkeit mit einem t-Wert von 5,49 eine hohe Signifikanz im Modell der Abhebefestigkeit. Mit der Querzugfestigkeit waren insgesamt nur noch 5 Variablen (B_AL_D3, CBV_O_R, FUEL_KO, MKDR_6, Lab_IB) in der Gleichung der Abhebefestigkeit enthalten. Dieses Modell (siehe Tabelle 33) zeigte die gleiche Schätzgenauigkeit im Vorhersageraum für die Abhebefestigkeit wie das Modell 10. Dieses Ergebnis bedeutet, dass durch endogene Regressoren die Anzahl an exogenen Regressoren im Modell stark verringert werden kann, ohne die Schätzgenauigkeit im Vorhersageraum zu verringern. Dies setzt sowohl eine hohe Korrelation der endogenen Variablen voraus als auch eine ausreichende Schätzgenauigkeit des endogenen Regressors. Diese sind selbst im ersten Schritt der 3SLS-Prozedur durch die Instrumentenvariabeln zu schätzen. Ein möglicher praktischer Nutzen endogener Variablen liegt in der Möglichkeit, schon bei geringer Stichprobenzahl ein kleineres Modell mit ausreichender Vorhersagegenauigkeit zu erstellen. Sobald genügend Datensätze vorliegen, können die endogenen Regressoren vernachlässigt werden. Modelle mit ausschließlich exogenen Regressoren sind zudem wesentlich einfacher zu interpretieren.

Die Nutzung endogener Variablen der Vorperiode (Lag-Variable) erbrachte keine Vorteile für die Schätzgenauigkeit.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	0,128	0,846	0,200	0,009
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,60	0,68	0,38	0,56
Adj R-Square	0,49	0,60	0,36	0,49
F-Value	5,70	8,36	16,83	7,38
DF_Model	29	29	5	21
MAD	0,07	0,43	0,15	0,01
MAPE	3,77	4,96	6,43	0,58
RMSE (T-K)	0,102	0,595	0,183	0,0071
RMSE (n)	0,090	0,528	0,183	0,0067
RMS Pct Error	4,74	6,29	8,28	0,75
Korrel_Prog	0,64	0,58	0,47	0,41
ME_Prog	0,003	0,037	0,029	0,000
MAD_Prog	0,077	0,553	0,143	0,007
MAPE_Prog	4,04	5,95	5,95	0,78
RMSE_Prog	0,099	0,695	0,179	0,0088
RMS Pct Error_Prog	5,23	7,29	7,38	0,99

Tabelle 33: Gütekennzahlen Modell 11 mit der Querzugfestigkeit als endogenen Regressor

9.2.7 Einfluss der Modellaktualisierung in den Parametern

Da im Vorhersagebetrieb mit Veränderungen im Produktionsprozess zu rechnen ist, sollte das Modell den neuen Bedingungen angepasst werden. Greubel (1999) berichtet von einer notwendigen ständigen Modellanpassung, da mit unvollständigen Wissen gearbeitet wird. Zu unterscheiden ist in der Modellaktualisierung die Aktualisierung der Parametermatrix durch hinzufügen der neuesten Datensätze und eine Aktualisierung des Modells in der Variablenauswahl. Im folgenden wird auf den Einfluss der Modellaktualisierung der Regressionskoeffizientenmatrix auf die Schätzgenauigkeit des Modells eingegangen.

Es sei an dieser Stelle wiederholt, dass bei den bisher dargestellten Modellen, die Datenmatrix nach jeder ex-ante-Vorhersage durch den letzten Datensatz aktualisiert wurde. Die Parametermatrix ist entsprechend nach jeder Vorhersage neu berechnet worden. Um den Einfluss der Parameteraktualisierung auf die Schätzgenauigkeit quantifizieren zu können, wurden zwei weitere Varianten berechnet:

Variante 1: keine Aktualisierung der Parameter über dem Vorhersageraum (worst case)

Variante 2: drei Aktualisierungen über dem Vorhersageraum, d.h. jeweils nach vorliegen acht neuer Beobachtungen erfolgte eine Neuberechnung der Parametermatrix

Als Vergleichsmodelle werden das technologische und das kombinierte Modell herangezogen (Modelle 9 und 10). Da in der industriellen Praxis nicht davon ausgegangen werden kann, dass eine ständige Aktualisierung der Parametermatrix erfolgt, zeigen diese Berechnungen, inwieweit die Schätzgenauigkeiten durch den Aktualisierungsmodus beeinflusst werden.

Tabelle 34 und Tabelle 35 zeigen die Ergebnisse des Modells 10 und 9 ohne jegliche Aktualisierung des Modells im Vorhersageraum. Der Freiheitsgrad des Fehlers bleibt über die Zeit der Vorhersagen konstant.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD Testdaten	0,1280	0,846	0,200	0,009
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,60	0,70	0,47	0,60
Adj R-Square	0,48	0,61	0,35	0,51
F-Value	5,06	7,73	3,77	7,31
DF_Model	29	29	24	21
MAD	0,07	0,42	0,14	0,01
MAPE	3,76	4,93	6,04	0,56
RMSE (T-K)	0,104	0,597	0,193	0,0070
RMSE (n)	0,090	0,521	0,173	0,0065
RMS Pct Error	4,74	6,27	7,61	0,73
Korrel_Prog	0,69	0,61	0,44	0,47
ME_Prog	-0,011	0,218	0,048	-0,001
MAD_Prog	0,072	0,558	0,154	0,007
MAPE_Prog	3,80	5,87	6,38	0,78
RMSE_Prog	0,093	0,720	0,186	0,0088
RMS Pct Error_Prog	5,00	7,31	7,68	0,99

Tabelle 34	: Gütekennzahlen	Modell 10	ohne A	ktualisierung	Variante 1)
------------	------------------	-----------	--------	---------------	------------	---

Im Vergleich zu dem Modell 10 mit ständiger Aktualisierung der Parameter zeigt sich im Modell 10 ohne Aktualisierung nur eine geringfügige Abnahme der Schätzgenauigkeit in den Gleichungen der Dickenquellung und der Abhebefestigkeit. Diese Abnahme in der Schätzgenauigkeit fällt jedoch sehr gering aus und offenbart damit die Stabilität des kombinierten Vorhersagemodells. Die geschätzte Standardabweichung der Vorhersage (RMSE_Prog) bleibt in allen Gleichungen kleiner als die Standardabweichung der Testdaten. Die Korrelationen zwischen den Vorhersagen und den Laborwerten sind weiterhin für alle Gleichungen auf dem 95% Niveau statistisch signifikant.

Im Gegensatz zu dem kombinierten Modell hat die Nichtaktualisierung des technologischen Modells eine höhere Abnahme der Schätzgenauigkeit der Eigenschaften Dickenquellung und Abhebefestigkeit im Vorhersageraum zur Folge (siehe Tabelle 35). Die Schätzgenauigkeit der Rohdichte und die der Querzugfestigkeit werden dagegen von der Aktualisierung der Parametermatrix nur unwesentlich beeinflusst.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD Testdaten	0,1280	0,846	0,200	0,009
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,51	0,45	0,33	0,53
Adj R-Square	0,44	0,38	0,24	0,47
F-Value	7,66	6,11	3,66	8,36
DF_Model	15	15	15	15
MAD	0,08	0,56	0,15	0,01
MAPE	4,22	6,54	6,63	0,62
RMSE (T-K)	0,107	0,750	0,207	0,0073
RMSE (n)	0,100	0,698	0,194	0,0071
RMS Pct Error	5,24	8,31	8,63	0,80
Korrel_Prog	0,64	0,46	0,34	0,37
ME_Prog	0,033	0,235	0,149	0,002
MAD_Prog	0,078	0,666	0,202	0,007
MAPE_Prog	4,05	7,11	8,18	0,83
RMSE_Prog	0,104	0,806	0,247	0,0094
RMS Pct Error_Prog	5,33	8,46	9,77	1,06

Tabelle 35: Gütekennzahlen Modell 9 ohne Aktualisierung (Variante 1)

Die Eigenschaften Dickenquellung und Abhebefestigkeit reagieren auf eine Nichtaktualisierung wesentlich empfindlicher als die Rohdichte und Querzugfestigkeit. Je präziser diese Eigenschaften jedoch modelliert werden können (Modell 10 im Vergleich zu Modell 9), desto geringer ist die Wirkung einer Nichtaktualisierung der Parametermatrix auf die Abnahme der Schätzgenauigkeiten. Durch eine gelegentliche Aktualisierung (Variante 2) des Modells kann die Schätzgenauigkeit des Modells schon wesentlich verbessert werden. Tabelle 36 zeigt die Ergebnisse des Modells 9, bei dem die Aktualisierung nach jeweils 8 neuen Beobachtungen, insgesamt drei mal im Vorhersageraum, vorgenommen wurde. Es sind keine wesentlichen Veränderungen zu den Ergebnissen der ständigen Aktualisierung des Modells festzustellen.

				-
Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	0,1280	0,846	0,200	0,009
CV_Testdaten	6,7	9,2	8,3	1,0
R-Square	0,51	0,44	0,33	0,51
Adj R-Square	0,45	0,37	0,25	0,45
F-Value	8,41	6,47	4,09	8,48
DF_Model	15	15	15	15
MAD	0,08	0,57	0,15	0,01
MAPE	4,17	6,58	6,64	0,63
RMSE (T-K)	0,106	0,745	0,206	0,0074
RMSE (n)	0,100	0,698	0,193	0,0072
RMS Pct Error	5,25	8,30	8,56	0,80
Korrel_Prog	0,62	0,52	0,35	0,33
ME_Prog	0,022	0,054	0,085	0,001
MAD_Prog	0,076	0,616	0,167	0,008
MAPE_Prog	3,93	6,70	6,78	0,84
RMSE_Prog	0,103	0,741	0,209	0,0096
RMS Pct Error Prog	5,31	8,06	8,30	1,07

Tabelle 36: Gütekennzahlen Modell 9 mit Teilaktualisierung (Variante 2)

Tabelle 37 fasst den Einfluss der Modellaktualisierung auf die Schätzgenauigkeit der Modelle zusammen. Die Kennziffern (Kz) 2 und 3 zeigen die auf die Standardabweichung normierte Vorhersagegenauigkeit im Beobachtungs- und im Vorhersageraum. Je höher die Kennziffern sind, desto größer ist die Schätzgenauigkeit im jeweiligen Datenraum.

	Kz	Lab_IB	Lab_TS	Lab_SS	Lab_MD	Gesamt
Modell 9	1	15	15	15	15	
	2	1,37	1,27	1,10	1,39	1,28
	3	1,26	1,13	0,99	0,97	1,09
Modell 10	1	29	29	24	21	
	2	1,42	1,59	1,18	1,45	1,41
	3	1,30	1,22	1,11	1,04	1,17
Modell 9 (Variante 1)	1	15	15	15	15	
	2	1,35	1,26	1,08	1,41	1,28
	3	1,23	1,05	0,81	0,97	1,01
Modell 9 (Variante 2)	1	15	15	15	15	
	2	1,36	1,27	1,09	1,40	1,28
	3	1,24	1,14	0,96	0,96	1,07
Modell 10 (Variante 1)	1	29	29	24	21	
	2	1,40	1,58	1,17	1,48	1,41
	3	1,37	1,18	1,07	1,04	1,17
1: Anzahl Variablen in den Gleichung						

2:

Tabelle 37: Genormte Schätzgenauigkeiten in Abhängigkeit der Parameteraktualisierung

Anzahl Variablen in den Gleichungen

Standardabweichung Lerndaten/RMSE Modell

3: Standardabweichung Testdaten/RMSE_Prog Das gleichzeitige Löschen des ältesten Datensatzes bei der Aktualisierung mit dem neuesten Datensatz führte zu einer Abnahme der Schätzgenauigkeit im Vorhersageraum. Bei diesem Vorgehen wurden dem Lerndatenraum vermutlich Prozesskonstellationen entzogen, die für die Vorhersage bedeutsam waren. Solange im Modell keine Strukturbrüche (Maschinenumstellung, neue Anforderungen in den Plattenqualitäten o.ä.) vorhanden sind, sollte der Lerndatenraum nicht eingeschränkt werden. In der Praxis wird der Lerndatenraum durch die Kapazität der Datenbank bzw. des Computers bestimmt.

Erkenntnisse

Eine regelmäßige Aktualisierung der Regressionskoeffizientenmatrix durch die neuesten Datensätze sorgt für die ständige Erweiterung des Lerndatenraums. Prozessänderungen durch natürliche Drifts (Verschleiß) oder Optimierungen werden dem Modell zunehmend angelernt. Die Modelle, die nach jeder Vorhersage aktualisiert worden sind, zeigten folgerichtig die höchsten Schätzgenauigkeiten. Am deutlichsten profitierten im vorliegenden Fall die Dickenquellung und die Abhebefestigkeit von einer Parameteraktualisierung. Auch ohne eine kontinuierliche Aktualisierung der Modelle können noch ausreichende Schätzgenauigkeiten (siehe Tabelle 37) erzielt werden. Dieses Ergebnis kann dann erwartet werden, wenn alle wichtigen Bestimmungsfaktoren im Modell enthalten sind und keine gravierenden Veränderungen im Prozess vorliegen.

9.2.8 Langfristige Vorhersagen

Anhand des zweiten Testdatensatzes, der sich zeitlich nahtlos an den ersten Testdatensatz anfügt, wurde das Modell 5 (Variablenauswahl nach Methode der optimalen Eingänge bis BWZ=70%) und das kombinierte Vorhersagemodell eingesetzt, um die langfristige Stabilität der Modelle zu untersuchen. In dieser Zeit erfolgte eine zunehmende Anhebung der Pressengeschwindigkeit. Diese Optimierung hatte u.a. Anpassungen im Pressenbereich zur Folge. Dabei wurde u.a. zusätzlich ein Multipot am Ende der Presse installiert. Aufgrund dieser strukturellen Änderungen, die dem Modell nicht gesamt zur Verfügung stehen, ist von einer geringeren Schätzgenauigkeit in diesem Zeitraum auszugehen.

Abbildung 28 zeigt exemplarisch für die Querzugfestigkeit die Vorhersagen des Modells 10 im Beobachtungsraum sowie über den gesamten Testdatenraum von einem Jahr. An der Stelle des Datensatzes 216 ist, in den Residuen sehr gut ersichtlich, ein offensichtlicher Ausreißer in der Vorhersage zu sehen. An dieser Beobachtung kam es zu einem Ausfall des Temperatursensors im Pressenheizfeld 1. Sehr starke Ausreißer in der Vorhersagen weisen i.d.R. auf eine große Abweichung vom Durchschnitt der Variablen im Datenraum hin. Diese Abweichungen beruhen sehr oft auf Ausfälle der Messtechnik. Damit lassen sich die Vorhersagen ferner für die Kontrolle der Messtechnik nutzen. Entsprechend fehlerhafte Datensätze sollten nicht für die Aktualisierung des Modells herangezogen werden. Bei längerfristigen Ausfall der Messtechnik, ist es dagegen vorzuziehen, die entsprechende Variable aus dem Modell zu nehmen. Im vorliegenden Fall wurde dies für die Temperaturmessung nach dem Ventilator in der Sichtung (ab Datensatz 172) für beide Modelle und für die Stromaufnahme des oberen Bandantriebes der Vorpresse (ab Datensatz 209) für Modell 10 praktiziert.



Abbildung 28: Vorhersagen der Querzugfestigkeit des Modells 10 mit Ausreißer

Tabelle 38 zeigt die entsprechenden Kennzahlen des Modells 5 und Tabelle 39 die des Modells 10 über den einjährigen Vorhersagebetrieb.

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD Testdaten	0,140	0,887	0,186	0,010
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square	0,55	0,62	0,42	0,53
Adj R-Square	0,48	0,53	0,33	0,46
F-Value	8,11	7,16	4,93	7,28
DF_Model	24	33	23	25
MAD	0,07	0,48	0,14	0,01
MAPE	3,98	5,49	6,08	0,62
RMSE (T-K)	0,101	0,671	0,187	0,0075
RMSE (n)	0,094	0,603	0,174	0,0071
RMS Pct Error	4,98	6,92	7,64	0,80
Korrel_Prog	0,61	0,38	0,34	0,49
ME_Prog	-0,041	0,033	-0,004	-0,002
MAD_Prog	0,093	0,709	0,143	0,007
MAPE_Prog	5,26	7,39	6,10	0,81
RMSE_Prog	0,119	0,899	0,180	0,0093
RMS Pct Error_Prog	6,87	9,21	7,73	1,05

Tabelle 38: Gütekennzahlen Modell 5 über einen Zeitraum von 1 Jahr

Kennzahlen	LAB_IB	LAB_TS	Lab_SS	Lab_MD
STD_Testdaten	0,140	0,887	0,186	0,010
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square	0,58	0,64	0,43	0,51
Adj R-Square	0,50	0,57	0,34	0,44
F-Value	7,55	9,43	4,94	8,17
DF_Model	29	29	24	21
MAD	0,07	0,47	0,14	0,01
MAPE	3,84	5,29	6,02	0,62
RMSE (T-K)	0,100	0,640	0,185	0,0076
RMSE (n)	0,091	0,583	0,172	0,0072
RMS Pct Error	4,80	6,73	7,51	0,80
Korrel_Prog	0,68	0,46	0,37	0,45
ME_Prog	-0,022	0,013	-0,001	-0,001
MAD_Prog	0,085	0,646	0,139	0,007
MAPE_Prog	4,72	6,78	5,96	0,81
RMSE_Prog	0,104	0,823	0,178	0,0092
RMS Pct Error_Prog	5,92	8,60	7,65	1,03

Die beiden obersten Zeilen in den Tabellen geben die neu berechnete Standardabweichungen und Variationskoeffizienten für den gesamten einjährigen Testdatensatz an. Aufgrund des verlängerten Testzeitraums änderten sich die Streuungen geringfügig. Diese Kennzahlen bilden nun die Vergleichsbasis zu den Fehlern der ex-ante-Vorhersagen.

Tabelle 40 fasst die Schätzgenauigkeiten der Modelle 5 und 10 im Vergleich über 3 und 12 Monate zusammen. Die Schätzgenauigkeit im Vorhersageraum des Modells 5 hat für die Eigenschaften Querzugfestigkeit und Dickenquellung über das Jahr abgenommen. Die Schätzgenauigkeit der Abhebefestigkeit und der Rohdichte zeigten sich in diesem Modell auch nach einem Jahr unverändert. Die Rohdichte und die Abhebefestigkeit weisen in beiden Modellen (5 und 10) über das gesamte Jahr annährend die gleichen Schätzgenauigkeiten im Vorhersageraum auf. Die Querzugfestigkeit und Dickenquellung werden dagegen mit dem kombinierten Modell im Jahresvergleich genauer geschätzt. Der ursächliche Grund hierfür lieqt zusätzlichen Berücksichtigung der in der Pressengeschwindigkeit und der Leimmenge im kombinierten Modell. Beide Prozessvariablen sind innerhalb der ersten 3 Monate der Validierung in der Gleichung der Querzugfestigkeit nicht relevant. Beide Modelle zeigen noch die gleiche Schätzgenauigkeit im Vorhersageraum (siehe Tabelle 40, Kennziffer 3: 1,29 bzw. 1,30). Erst mit einer Anhebung der Pressengeschwindigkeit und einer größeren Schwankungsbreite der Leimmenge im Zeitraum des zweiten Testdatensatzes werden beide Variablen im Modell 10 für die Querzugfestigkeit statistisch signifikant. Das Modell 10 weist durch die Berücksichtigung dieser technologisch bedeutenden Variablen im Jahresmittel eine wesentlich höhere Schätzgenauigkeit für die Querzugfestigkeit auf (Kennziffer 3: 1,18 zu 1,34).

	Kz	Lab_IB	Lab_TS	Lab_SS	Lab_MD	Gesamt
Modell 5	1	24	33	23	25	
3 Monate	2	1,41	1,54	1,23	1,41	1,40
	3	1,29	1,08	1,01	1,06	1,11
Modell 5	1	24	33	23	25	
12 Monate	2	1,43	1,41	1,21	1,39	1,36
	3	1,18	0,99	1,03	1,06	1,06
Modell 10	1	29	29	24	21	
3 Monate	2	1,42	1,59	1,18	1,45	1,41
	3	1,30	1,22	1,11	1,04	1,17
Modell 10	1	29	29	24	21	
12 Monate	2	1,45	1,48	1,22	1,37	1,38
	3	1,34	1,08	1,04	1,07	1,13
1: Anzahl der Variablen im Modell						

Tabelle 40: Genormte Schätzgenauigkeiten des Modells 5 und 10 über 3 und 12 Monate im Vergleich

> Anzahl der Variablen im Modell Standardabweichung Lerndaten/RMSE_Modell

Standardabweichung Testdaten/RMSE_Prog

In beiden Modellen ist im Beobachtungsraum eine geringe Abnahme des Bestimmtheitsmaßes in den Gleichungen der Dickenquellung, Abhebefestigkeit und Rohdichte während des einjährigen Vorhersagebetriebes festzustellen. Abbildung 29 zeigt die Entwicklung der Bestimmtheitsmaße der vier Gleichungen während des Vorhersagebetriebes für das Modell 10. Eine Stabilisierung des Bestimmtheitsmaßes ist ab der 80. Vorhersage festzustellen.

2: 3[.]



Abbildung 29: Entwicklung der Bestimmtheitsmaße während der ex-ante-Vorhersagen über ein Jahr (Modell 10)

Eine mit der Zeit überproportionale Abnahme des Bestimmtheitsmaßes würde sich in einem fallenden F-Wert zeigen. Sobald dieser über einen längeren Zeitraum im Vorhersagebetrieb fällt, kann dies als Hinweis zur Modellaktualisierung in der Variablenauswahl verstanden werden. Abbildung 30 zeigt für Model 10 den Verlauf der F-Werte im Vorhersagebetrieb.



Abbildung 30: Entwicklung der F-Werte während der ex-ante-Vorhersagen über ein Jahr (Modell 10)

Alle F-Werte steigen im Vorhersagebetrieb kontinuierlich an, wobei die Gleichungen der Abhebefestigkeit den geringsten und die Gleichung der Querzugfestigkeit den höchsten Anstieg aufweisen. Die Gleichung der Rohdichte zeigt im Bereich zwischen der 70. und 90. Vorhersage einen kurzzeitigen Abfall im Verlauf des F-Wertes. Dieser Bereich fällt mit einer Änderung des Pressenprogramms zusammen, dessen Auswirkung in erster Linie im Modell der Rohdichte zu beobachten ist.

Die Schätzgenauigkeit im Vorhersageraum ist bezogen auf die tatsächlichen Streuungen von allen Eigenschaften für die Querzugfestigkeit am höchsten. Dieses Ergebnis korrespondiert mit dem annährend konstanten Verlauf des Bestimmtheitsmaßes über den Vorhersageraum. Das Modell der Querzugfestigkeit erfasst die wesentlichen Bestimmungsfaktoren über die gesamte Zeit. Die Dickenquellung scheint dagegen durch nicht im Modell enthaltene Variablen beeinflusst. Die Schätzgenauigkeit der Dickenquellung zeigte in den 12 Monaten der Validierung den höchsten Verlust. Dennoch ist die Dickenquellung ausreichend genau vorherzusagen. Ausreichend genau heißt, dass der Fehler des Modells im einjährigen Vorhersageraum kleiner ist als die Standardabweichung der Testdaten. Die systematischen Schwankungen in den Platteneigenschaften werden durch das Modell vorhergesagt. Die systematischen Schwankungen der Abhebefestigkeit und der Rohdichte lassen sich im Verhältnis zur Standardabweichung der Testdaten nicht so präzise vorhersagen, wie die der Querzugfestigkeit. Im Fall der Rohdichte liegt der Grund in der sehr geringen Schwankungsbreite der Daten. Die Abhebefestigkeit dagegen weist in den Laborwerten und über die Produktionsbreite einen größeren Fehler auf als die Querzugfestigkeit.

Erkenntnisse

Der prozentuale Fehler der einjährigen Vorhersage liegt durchschnittlich höher als innerhalb der ersten 3 Monate. Über ein gesamtes Jahr konnte mit dem kombinierten Modell die Querzugfestigkeit mit einem durchschnittlichen absoluten Fehler von 4,7%, die Dickenquellung mit 6,8%, die Abhebefestigkeit mit 6,0% und die Rohdichte mit 0,8% vorhergesagt werden.

Werden die Vorhersagefehler der einzelnen Modellgleichungen auf die empirischen Standardabweichungen des Jahres bezogen, so bleibt die normierte Vorhersagegenauigkeit der Querzugfestigkeit und der Rohdichte über das Jahr konstant. Die Abhebefestigkeit weist nur eine sehr geringe Abnahme in der Schätzgenauigkeit innerhalb des einjährigen Vorhersagebetriebs des Modells auf. Die Schätzgenauigkeit der Dickenquellung dagegen zeigt während des Jahres die höchsten Verluste. Damit wird die Dickenquellung am stärksten von Variablen außerhalb des Modells beeinflusst.

Das auf Basis der Bewertungszahlen erstellte Modell 5 weist für die Querzugfestigkeit und die Dickenquellung eine geringere Vorhersagegenauigkeit auf, als das kombinierte Modell 10 mit zusätzlich technologisch ausgesuchten Variablen. Dennoch kann das Modell 5 bis auf die Dickenquellung die geforderten Genauigkeiten erreichen.

Anhand des F-Wertes der Modellgleichungen kann die Modellgültigkeit im Vorhersageraum überprüft werden. Sobald diese Kennzahl mit fortschreitender Vorhersage über einen längeren Zeitabschnitt abnimmt, sollte die Variablenauswahl überprüft werden.

9.2.9 Technologische Betrachtungen

Neben den Kriterien zur Beurteilung der Modellgüte, spielt die Güte der theoretischen Fundierung der Regressionsgleichungen, eine bedeutende Rolle. Im folgenden werden die wichtigsten technologischen Zusammenhänge zwischen den Prozessvariablen und den Laboreigenschaften in den Modellen diskutiert. Da die Signifikanzen einzelner Variablen sowohl in der Zeit als auch zwischen den Modellen sehr stark variieren können (je nach Variablenauswahl), werden an dieser Stelle nur Zusammenhänge dargestellt, die sich im Durchschnitt über alle validierten Modelle hinweg als signifikant herausstellten. Da die einzelnen Prozessvariablen jedoch nicht unabhängig voneinander in verschiedenen Stufen eingestellt worden sind (Experiment), sondern auf passive Beobachtungen basieren, sind Interpretationen kausaler Zusammenhänge zwischen einzelnen Prozessvariablen und den Laboreigenschaften immer der Gefahr ausgesetzt, dass diese auch durch andere, nicht berücksichtigte Effekte beeinflusst worden sind.

Kocher und Refiner

Dieser Prozessabschnitt beeinflusst besonders stark die Fasergeometrie und Faserqualität. Innerhalb des Prozessabschnittes liegen jedoch keine Messungen vor, die direkt an den Fasern gemessen worden sind. Die gemessenen Prozessvariablen (Temperaturen, Drücke, Drehzahlen und Stromaufnahmen) stellen damit indirekte Variablen der Faserqualität dar. Dies führt in der Regel zu einer schwierigen Interpretation der Variablen im Bezug auf die direkten Platteneigenschaften. Nach Groom et al. (2002) sind die individuellen Fasereigenschaften für die mechanischen Eigenschaften von MDF-Platten sehr bedeutsam.

Als bedeutend stellten sich in den Modellen besonders die Stromaufnahme der Dosierschnecke (auch Speise- bzw. Beschickschnecke genannt) und die Dampfmengen im Kocher, Refiner bzw. im Hackschnitzelbunker heraus.

Nach den Modellen führte dabei eine erhöhte Stromaufnahme der Dosierschnecke zu einer Erhöhung der Querzugfestigkeit, Abhebefestigkeit und Dickenquellung. Eine direkte Interpretation dieses Zusammenhangs ist nicht möglich, da die Stromaufnahme der Dosierschnecke selbst von vielen Parametern abhängig ist. So wird die Stromaufnahme insbesondere durch die Holzart, die Holzfeuchte, die Kochbedingungen (Zeit, Temperatur, Druck) und den Austragsmengen beeinflusst.

Eine Erhöhung der Dampfmenge im Mahlraum des Refiners bzw. im Kocher führte zu einer Erhöhung der Rohdichte bzw. zu einer Verringerung in der Dickenquellung. Signifikant positiv korreliert zu der Variable Dampfmenge im Kocher ist die Variable Dampfmenge gesamt. Sie zeigte die gleiche Wirkung auf die Dickenquellung, konnte jedoch nicht gleichzeitig mit der Variable Kocherdampfmenge im Modell berücksichtigt werden.

Krug und Kehr (2001) berichten von einer signifikanten Abnahme der Dickenquellung von MUF und PMDI gebundenen MDF mit einer Erhöhung des Druckes im Kocher von 8 auf 12 bzw. 16 bar. Mit der Druckerhöhung erhöhte sich zugleich der Feinanteil an Fasern mit geringeren Faserlängen. Die kürzeren Faserlängen waren mit einer Abnahme der Biegefestigkeit und des Biege-Elastizitätsmoduls der MDF-Platten verbunden. Die Abnahme

der Dickenquellung mit zunehmenden Aufschlussdruck im Kocher wird nach den Autoren dagegen durch eine dauerhafte Quellungsvergütung erzielt. Die Quellungsvergütung beruht auf eine Abnahme des Hemizellulosenanteils mit zunehmenden Dampfdruck (bei gleichzeitiger Temperaturerhöhung) im Kocher (Krug und Tobisch, 2002). Die Hemicellulosen (Polyosen) sind die temperaturempfindlichsten Holzkomponenten, gefolgt von der Cellulose und dem Lignin (Burmester, 1970). Eine Verringerung von hydrophilen Bestandteilen in der Zellwand ist mit einer Abnahme in der Dickenquellung verbunden. Nach Faix (2003) sind die Hemicellulosen in isolierter Form ca. um den Faktor 1,6 hydrophiler als die Cellulose. Versuche von Goroyias und Hale (2002) zeigten, dass die Dickenquellung bei 24 Stunden Wasserlagerung von OSB-Strands bei einer Temperaturbehandlung von 200 bis 260°C um mehr als 90% verbessert werden konnte.

Im Datensatz waren die Druckschwankungen im Kocher mit einem Variationskoeffizienten von 2,8% zu klein, um einen statistisch signifikanten Effekt nachzuweisen. Es ist jedoch davon auszugehen, dass eine Erhöhung der Dampfmenge im Kocher den gleichen Effekt beschreibt, wie eine Druckerhöhung im Kocher.

Ein Einfluss der Druckschwankungen im Refiner auf die Platteneigenschaften war ebenfalls nicht signifikant nachweisbar. In Untersuchungen von Labosky et al. (1993) führte eine Druckerhöhung im Refiner von 50 psi (3,48 bar) auf 100 psi (6,89 bar) zu keinen signifikanten Einfluss auf die mechanisch-technologischen Eigenschaften von MDF-Platten. Dagegen deuteten Voruntersuchungen auf einen hohen Einfluss der Hackschnitzelfeuchte auf die Platteneigenschaften hin.

Eine Erhöhung der Dampfmenge im Hackschnitzelbunker führte nach den Modellen zu einer Erhöhung der Rohdichte. Der Füllstand im Kocher zeigte dagegen einen signifikant negativen Zusammenhang zur Rohdichte. Beide Zusammenhänge sind technologisch nicht eindeutig zu interpretieren. Es kann vermutet werden, dass die eigentliche Beziehung zwischen der Holzartenzusammensetzung und der mittleren Rohdichte besteht.

Beleimung

Die Beleimung stellt einen zentralen Punkt der Prozesssteuerung sowohl hinsichtlich der Qualität als auch der Kosten des Fertigproduktes dar. Variablen zur Beschreibung der Beleimungsqualität stehen online i.d.R. nicht zur Verfügung. In der Literatur beschriebene Größen zur Charakterisierung der Beleimung sind der Kontaktwinkel, die Klebstoffverteilung, Tröpfchengröße und der Bedeckungsgrad. Bekannt ist, dass die Klebstoffverteilung (und damit auch der übrigen Größen) einen Einfluss auf die Platteneigenschaften ausübt, jedoch nicht, wie ein Optimum der Klebstoffverteilung für spezifische Eigenschaften der Platte auszusehen und zu erreichen ist (Stahl, 2002). Stahl untersuchte anhand von 47 Datensätzen (4 Monate) aus der MDF-Herstellung den Einfluss des Bedeckungsgrades des Klebstoffes auf die Platteneigenschaften. Während der Datenerhebung wurden nach Stahl sowohl der Beleimungsgrad als auch die Refinerbedingungen variiert. Der mittlere Bedeckungsgrad der Fasern mit Klebstoff lag bei den untersuchten Proben bei 45% und schwankten zwischen 39% und 62%. In dieser Untersuchung konnte zwischen dem Bedeckungsgrad der Fasern mit Klebstoff und den Standardeigenschaften Querzugfestigkeit, 24-Stunden-Dickenquellung und Biegefestigkeit kein statistisch signifikanter Zusammenhang

gefunden werden. Zu der Prüfung der Querzugfestigkeit und der Dickenquellung nach dem Standard 313V zeigte sich jedoch eine signifikante positive bzw. negative Korrelation zum mittleren Bedeckungsgrad. Der Bedeckungsgrad der Fasern mit Klebstoff selbst wurde anhand eines statistischen Modells beschrieben. Dazu wurden 20 Variablen aus der Klebstoff- und Fasercharakteristik und aus dem Refiner und der Blowline herangezogen. Das Bestimmtheitsmaß dieses Modells lag bei R²=0,67.

In den Regressionsmodellen dieser Arbeit konnten neben der Indikatorvariable "Leimhersteller", nur die Variablen Leimmenge und Härtermenge berücksichtigt werden. Die zusätzliche Verwendung der Variablen Fasermenge, Harnstoff, Wasser und Paraffin führte zu sehr hohen Varianzinflationsfaktoren und zu Vorzeichen in den Regressionskoeffizienten, die dem technologischen Zusammenhang widersprachen. Die Faser-, Wasser- und Paraffinmenge sind signifikant mit der Leimmenge korreliert (Tabelle 41). Dieser Zusammenhang ist durch die Rezeptur vorgegeben, in der sich die zugegebenen Mengen jeweils auf die Faser- bzw. Leimmengen beziehen.

	Leim	Härter	Wasser	Paraffin	Faser
Harnstoff	-0.01	0.33	-0.75	0.38	0.44
	0.84	<.0001	<.0001	<.0001	<.0001
Leim	1	0.01	0.64	0.76	0.83
		0.92	<.0001	<.0001	<.0001
Härter		1	-0.37	0.05	0.11
			<.0001	0.47	0.097
Wasser			1	0.19	0.19
				0.002	0.003
Paraffin				1	0.86
					<.0001

Tabelle 41: Korrelationen zwischen den Beleimungsdaten (N=248)

Die Härter- und Harnstoffmengen weisen keine signifikanten Korrelationen zu der eingesetzten Leimmenge auf. Diese Zugaben wurden je nach Klebstofflieferant, technologischen Gegebenheiten (Holzmischung, Temperaturen) und Kundenanforderungen (Formaldehydgehalt) in unterschiedlichen Niveaus eingesetzt. Die signifikante Korrelation zwischen Härter- und Harnstoffmenge erscheint dagegen nicht kausaler, sondern statistischer Natur. Tabelle 42 stellt die Korrelationen der Beleimungsdaten zu den Laboreigenschaften dar. In den Klammern sind jeweils die Signifikanzniveaus (p-Werte) angegeben.

Tabelle 42: Korrelationen zwischen Beleimungsdaten und Laboreigenschaften

	Querzug-	Rohdichte	Dicken-	Abhebe-	Formalde-
	festigkeit		quellung	festigkeit	hydgehalt
Anzahl	248	248	248	248	110
Harnstoff	-0,25	-0,34	0,33	0,02	-0,28
	(<0,0001)	(<0,0001)	(<0,0001)	(0,78)	(0,003)
Härter	-0,32	-0,17	0,24	-0,13	-0,38
	(<0,0001)	(0,006)	(0,0002)	(0,04)	(<0,001)
Leim	-0,15	0,09	-0,35	-0,04	0,12
	(0,02)	(0,18)	(<0,0001)	(0,58)	(0,21)
Wasser	0,12	0,31	-0,49	-0,04	0,29
	(0,07)	(<0,0001)	(<0,0001)	(0,51)	(0,002)
Paraffin	-0,20	-0,05	0,09	0,01	-0,04
	(0,001)	(0,46)	(0,18)	(0,82)	(0,67)

Die Härtermenge ist als einzige Beleimungsvariable zu allen Laboreigenschaften signifikant korreliert. Die Leimmenge dagegen zeigt mit ihrer geringen Schwankungsbreite nur eine signifikante Korrelation zur Querzugfestigkeit und Dickenquellung. Dabei ist die negative Korrelation zwischen Leimmenge und Querzugfestigkeit technologisch falsch. Werden die Korrelationen zwischen diesen beiden Variablen in den Gruppen der Leimhersteller berechnet, sind die Korrelationen nicht mehr signifikant. Die signifikante Korrelation zwischen Leimmenge und Dickenquellung bleibt jedoch erhalten.

In den Modellen zeigte sich die Leimmenge in der Gleichung der Dickenquellung über das gesamte Jahr der Validierung hinweg als hochsignifikant (p-Wert: <0,0001) und mit dem höchsten standardisierten Regressionskoeffizienten als die bedeutendste Variable zur Erklärung der Schwankungen in der Dickenquellung. Die gleiche Wirkung ist auch die Variable Drehzahl Austragsschnecke im Refiner nachzuweisen. Sie kann als Ersatzgröße für die Leimmenge herangezogen werden, da diese beiden Größen signifikant positiv korreliert sind. In der Gleichung der Querzugfestigkeit wird die Leimmenge dagegen erst gegen Ende der Jahresvorhersage signifikant, jedoch nicht als die bedeutendste Variable. Dies könnte auf einer leicht höhere Schwankungsbreite der Leimmenge am Ende der Vorhersage zurück geführt werden. Der Variationskoeffizient der Leimmenge lag mit den Beobachtungen des zweiten Testdatensatzes bei 4,3%. Kruse (1997) konnte bei einem Variationskoeffizienten von 3,2% keine Signifikanz der Leimmenge der Mittelschicht im Modell zur Beschreibung der Querzugfestigkeit einer Spanplattenfertigung feststellen. In den MDF-Modellen dieser Arbeit zur Vorhersage der Abhebefestigkeit und Rohdichte wies die Leimmenge über den gesamten Zeitraum keine Signifikanz auf. Es kann aus dieser Beobachtung die Hypothese aufgestellt werden, dass die Dickenquellung wesentlich sensibler auf geringfügige Schwankungen in der Beleimungsmenge reagiert als die anderen Platteneigenschaften.

Hague et al. (1999) untersuchten u.a. den Einfluss der Leimmenge auf die Dickenquellung und Querzugfestigkeit an UF-gebundenen MDF-Platten aus Fichte. Eine Erhöhung des Leimgehaltes von 8% auf 14% führte zu einer Erhöhung der Querzugfestigkeit um ca. 30% und zu einer Abnahme in der Dickenquellung um ca. 40%.

Niemz (1982) kam bei Untersuchungen zum Einfluss des Harzmenge auf die technologischen Eigenschaften von UF gebundenen Laborspanplatten aus Kiefer zu ähnlichen Ergebnissen. Eine Erhöhung des Festharzanteils von 6% auf 18% führte zu einem Festigkeitszuwachs von etwa 40%. Die Dickenquellung (2h) verringerte sich dagegen um annährend 50%.

Labosky et al. (1993) untersuchten u.a. den Einfluss der Harzmenge auf die technologischen Eigenschaften von UF gebundenen MDF-Platten aus Ahorn (Acer rubrum L.). Eine Erhöhung der Harzmenge von 6% auf 12% führte in diesen Untersuchungen zu einen Anstieg in der Querzugfestigkeit um 174% und zu einer Verringerung der Dickenquellung um 53%. Die Autoren geben eine Verringerung der Dickenquellung um 113% an, tatsächlich ist damit jedoch eine Erhöhung der Dickenquellung um 113% gemeint, wenn die Harzmenge von 12% auf 6% reduziert wird.

Es kann festgehalten werden, dass der Einfluss der Harzmenge auf die technologischen Eigenschaften von den jeweiligen Prozessbedingungen abhängig ist. Ein allgemeingültiger

Zusammenhang des Einflusses der Harzmenge auf die technologischen Eigenschaften kann nicht angegeben werden. Die in der Literatur untersuchten Spannweiten der Leimmenge sind jedoch wesentlich größer als im vorliegenden Fall.

Der Leimtyp bzw. der Leimhersteller zeigte in allen Gleichungen, bis auf die Rohdichte, einen signifikanten Einfluss auf die Platteneigenschaften. Neben den Einfluss der Härtermenge und der Temperatur im Streubunker war der Leimtyp die bedeutendste Variable zur Erklärung der Schwankungen in den Platteneigenschaften.

Die eingesetzte Paraffinmenge zeigt in Tabelle 42 nur zur Querzugfestigkeit eine negative statistisch signifikante Korrelation. In den Modellen konnte die Paraffinmenge, auch bei Entfernung der korrelierten Leim- und Härtermenge, nicht als signifikante Variable identifiziert werden.

Der Einfluss der Wasser- und Harnstoffmenge konnte aufgrund der hohen Korrelationen in den Modellen nicht eindeutig nachgewiesen werden. Als einzelne Variablen der Beleimung in den Modellen zeigten sie keine Signifikanz. Wurden sie zusätzlich zu den Variablen Leimund Härtermenge in die Modelle mit aufgenommen zeigten diese Variablen zum Teil auch eine Signifikanz an. Diese Signifikanz ging jedoch mit einem sehr hohen Varianzinflationsfaktor einher. Sie wurden daher nicht in die Modelle mit aufgenommen. Es kann jedoch davon ausgegangen werden, dass beide Variablen einen eigenen Erklärungsanteil an den Schwankungen in den Platteneigenschaften besitzen. Um diesen Einfluss nachzuweisen, müssten beide Variablen unabhängig von den übrigen Beleimungsvariablen stufenweise variiert werden. Ein Indiz für den Einfluss des Harnstoffes auf die Platteneigenschaften ergibt sich aus den Daten selbst. Harnstoff wurde im wesentlichen in zwei Niveaus zugegeben (kleiner und größer 9 kg/min). Wird der Datensatz nach diesen beiden Niveaus aufgeteilt und der Mittelwertunterschied der Platteneigenschaften mit dem entsprechenden t-Test getestet, so ergeben sich folgende signifikante Mittelwertsunterschiede in den Eigenschaften:

Einfluss Harnstoff auf	Formaldehyd	Rohdichte	Querzug	Quellung
Mittelwertunterschied	1,12 [mg/100g]	8,0 [kg/m³]	0,06 [N/mm²]	0,7 [%]
Anzahl Datensätze	110	248	248	248

Tabelle 43: Einfluss von Harnstoff auf die Platteneigenschaften

Das höhere Harnstoffniveau führt danach zu einer Abnahme des Formaldehydgehalts, der Querzugfestigkeit und Rohdichte und zu einer höheren Dickenquellung. Die Abhebefestigkeit zeigte keinen signifikanten Unterschied in den Mittelwerten. In dieser Darstellung bleiben alle anderen Einflüsse unberücksichtigt. Es kann davon ausgegangen werden, dass die Unterschiede in dieser Betrachtung überschätzt sind. Technologisch ist der Zusammenhang jedoch korrekt.

Trockner und Sichter

In diesen Prozessabschnitten liegen in erster Linie Feuchte- und Temperaturmessungen vor. Die Temperaturmessungen korrelieren dabei entsprechend den Messstellen zum Teil sehr stark. Höhere Temperaturen in diesen Bereichen führten in allen Modellen zu einer Erniedrigung der Querzugfestigkeit, Abhebefestigkeit und Rohdichte und zu einer Erhöhung der Dickenquellung. Dies lässt sich technologisch durch eine zunehmende Aushärtung des Klebstoffes mit zunehmenden Temperaturen erklären. Den gleichen Effekt beschreibt Hague et al. (1999). Eine Temperaturerhöhung am Trocknereingang von 150°C auf 200°C bewirkte eine signifikante Verringerung der Querzugfestigkeit und eine Erhöhung der Dickenquellung.

Eine höhere Faserfeuchte nach dem Trockner führte zu einer geringeren Dickenquellung. Der Einfluss der Faserfeuchte wird im Abschnitt Formband diskutiert, da dort auch die Mattenfeuchte einen signifikanten Effekt auf die Querzugfestigkeit und Rohdichte ausübte.

Ein höherer Füllstand des Faserbunkers führte in den Modellen zu einer Abnahme der Querzug- und Abhebefestigkeit und zu einer erhöhten Rohdichte. Ein Einfluss auf die Dickenquellung war nicht nachzuweisen. Dieser Zusammenhang ist technologisch nicht untersucht. Signifikante Korrelationen zwischen dem Füllstand und den technologischen Eigenschaften lagen nicht vor. Aus den einzelnen Korrelationskoeffizienten zwischen Prozessgrößen und technologischen Eigenschaften lässt sich i.d.R. nicht auf eine Bedeutung dieser Variablen in den Modellen schließen. Variablen, die nicht signifikante Korrelationen zu den technologischen Eigenschaften zeigen, können in den Modellen, im Zusammenspiel mit anderen Einflussgrößen, signifikante Einflussgrößen werden. Eine Ursache für die Abnahme der Festigkeiten könnte darin vermutet werden, dass ein höherer Füllstand des Faserbunkers mit einer längeren Verweilzeit der Fasern und damit mit einer höheren Voraushärtung des Klebstoffes verbunden ist.

Formband

Die Formbandgeschwindigkeit wurde in diesem Abschnitt nicht berücksichtigt. Diese ist durch die Pressengeschwindigkeit ersetzbar und wird im Abschnitt der Heißpresse diskutiert. In diesem Abschnitt zeigte sich die Temperatur des Streubunkers als eine für alle Gleichungen sehr bedeutende Variable. Die Wirkung ist dabei die gleiche, wie der Einfluss der Temperatur im Trocknungs- bzw. Sichtungsabschnitt. Je höher die Temperatur im Streubunker, desto geringer die untersuchten mechanischen Eigenschaften und um so höher die Dickenquellung. Der Effekt zeigte sich ebenfalls bei der Faserbunkertemperatur, konnte aber aufgrund der hohen Korrelation zur Streubunkertemperatur in den Modellen nicht berücksichtigt werden.

Weniger deutlich als die Temperatur war der Einfluss der Mattenfeuchte und der Flächenwaage zu sehen. Die Mattenfeuchte nach der Streuung lag ca. 2% (absolut) unter der gemessenen Feuchte nach dem Trockner, zeigte jedoch mit einem Variationskoeffizienten innerhalb von 21 Monaten von ca. 4% wesentlich geringere Schwankungen als die Feuchtemessung nach dem Trockner mit ca. 12%. Die Mattenfeuchte korreliert zudem mit der im Labor gemessenen Plattenfeuchte wesentlich besser als die Trocknerausgangs-feuchte. Beide Messungen sind jedoch statistisch signifikant zu der im Labor gemessenen Feuchte korreliert. Es ist anzunehmen, dass die Feuchtemessung nach dem Trockner, aufgrund der höheren Umgebungstemperaturen, eine größere Unsicherheit aufweist. Eine Feuchtemessung ist nach Schäckel (2001) im Bereich zwischen 0°C und 50°C Umgebungs-temperatur zufriedenstellend. Außerhalb dieses Temperaturbereiches sind entsprechende Kühl- bzw. Heizeinrichtungen notwendig, um ausreichende Messgenauigkeiten zu erreichen. Eine erhöhte Faserfeuchte zeigte in den Modellen einen signifikanten positiven Einfluss auf die Querzugfestigkeit und Rohdichte. Einen signifikanten Einfluss der Feuchte auf die Dickenquellung konnte durch die Feuchte nach dem Trockner nachgewiesen werden. Lediglich ein Einfluss der Feuchte auf die Abhebefestigkeit war nicht festzustellen.

Lee und Maloney (1995) berichten, dass in den Untersuchungen eine Mattenfeuchte von 13% ein Optimum zur Maximierung der technologischen Eigenschaften von MDF darstellte. Gleichzeitig geben die Autoren an, dass die Mattenfeuchte bei der MDF-Herstellung zwischen 8-11% liegt, um die Produktqualität und Produktivität auszubalancieren.

Hague et al. (1999) stellten fest, dass selbst kleine Änderungen der Mattenfeuchte einen signifikanten Einfluss auf die Platteneigenschaften von MDF ausüben. Ein Anstieg der Mattenfeuchte von ca. 8,5% auf 10,5% führte zu einer Erhöhung der Querzugfestigkeit um 70% und zu einer Verringerung der Dickenquellung um 40%. Die Mattenfeuchte zeigte in den Untersuchungen einen stärkeren Einfluss als die Leimmenge auf die technologischen Eigenschaften der MDF-Platten.

Der Einfluss des Flächengewichts [kg/m²] war mit einem Variationskoeffizienten von kleiner als 2% nicht nachweisbar. Der Einfluss der Flächenwaage nach der Streuung (CV=3,6%) konnte dagegen in einigen Modellen für die Rohdichte aufgezeigt werden. Dieser war jedoch im Vergleich zu anderen Einflüssen relativ gering. Die höheren Schwankungen der Flächenwaage im Vergleich zur Flächengewichtsmessung (BWQ) dürften vor allem auf die Feuchteschwankungen zurückzuführen sein. Die tatsächlichen Schwankungen im Fasergewicht (atro) waren demnach zu klein, um einen signifikanten Einfluss auf die mechanisch-technologischen Platteneigenschaften auszuüben.

Die Mattenbreite, gemessen nach der Streuung, zeigte sich im Modell der Dickenquellung als hochsignifikante Variable. Innerhalb der 8 mm Plattenproduktion wurden im wesentlichen zwei Plattenbreiten für unterschiedliche Kunden hergestellt. Die Unterschiede in den Plattenqualitäten dieser Kunden liegen in verschiedenen Anforderungen zur Dickenquellung und des Formaldehydgehaltes. Entsprechend den Kundenanforderungen werden u.a. verschiedene Harnstoffmengen eingesetzt. Alle anderen Eigenschaften zeigten in Abhängigkeit der Plattenbreite keine signifikanten Unterschiede. Mit der Variable Plattenbreite konnte der systematische Unterschied in der Dickenquellung zwischen den beiden Plattentypen berücksichtigt werden, der durch die restlichen im Modell enthaltenen Variablen nicht vollständig erklärt werden konnte. Die Plattenbreite ist damit nicht technologisch zu interpretieren, sondern als eine Indikatorvariable, die kategorische Unterschiede zwischen zwei Plattenbreiten beschreibt.

Vorpresse

Die Variablen in der Vorpresse umfassen vor allem Druck- und Distanzmessungen und Stromaufnahmen der Bandantriebe. Insgesamt sind die Variationen der Variablen in der Vorpresse wesentlich höher als in der Heißpresse. Die Variationskoeffizienten der Variablen der Vorpresse lagen zwischen 8% und 38%.

In den Modellen zeigten sich insbesondere die Variablen Stromverbrauch des unteren Bandantriebes und der Druck oben rechts als signifikante Einflussfaktoren. Ein höherer Druck in der Vorpresse führte dabei in den Modellen zu einer Erhöhung der Querzugfestigkeit, Abhebefestigkeit und Rohdichte. Ein höherer Druck ist mit einer höheren Vorverdichtung der Matte verbunden und führt in der Heißpresse zu einer schnelleren Durchwärmung der Matte. Damit ist bei konstanter Pressengeschwindigkeit ein verbesserter Vernetzungsgrad des Klebstoffes zu erwarten.

Eine höhere Stromaufnahme des unteren Bandantriebes der Vorpresse ist nach dem Modellen mit einer Erhöhung der Querzugfestigkeit und der mittleren Rohdichte sowie einer Verringerung der Dickenquellung verbunden. Der Einfluss dieser Variable ist technologisch nicht eindeutig zu beschreiben, da diese von sehr vielen Parametern abhängig ist. Die Stromaufnahme des Bandantriebes wird von dem Verdichtungsverhalten der Matte beeinflusst. Das Verdichtungsverhalten der Matte ist nach Haas (1998) von der Rohdichte, der Partikelgeometrie, der Verdichtungsgeschwindigkeit (hier Pressengeschwindigkeit) und von der Temperatur und Feuchte der Matte abhängig. Als zusätzlicher Einflussfaktor ist die Holzart zu nennen.

<u>Heißpresse</u>

Die kontinuierliche Presse umfasst als Prozessabschnitt die meisten Variablen. Zugleich zeigen die Variablen sowohl quer als auch längs der Pressenrahmen signifikante Korrelationen. Aufgrund dieser hohen Korrelationen der Pressenvariablen ist ein kausaler Zusammenhang zwischen einem einzelnen Pressenrahmen und den jeweiligen Platteneigenschaften nicht herzustellen. Obwohl die Presse die meisten Variablen liefert, konnten in den Modellen nur sehr wenige davon berücksichtigt werden, ohne ernsthafte Multikolliniaritätsprobleme zu verursachen. Es konnten nur sehr wenige Variablen identifiziert werden, die sich über die Modelle hinweg als gleich bedeutsam herausstellten. In der Regel sind die Druck- und Distanzvariablen in den Modellen austauschbar.

Die Pressengeschwindigkeit war nur im Modell der Dickenquellung über das gesamte Jahr der Vorhersagen als signifikante Einflussgröße zu identifizieren. Mit der kontinuierlichen Erhöhung der Pressengeschwindigkeit im Laufe des Vorhersagejahres, zeigte sich diese auch im Modell der Querzugfestigkeit als signifikant. Für die Abhebefestigkeit und die Rohdichte konnte selbst nach der Erhöhung der Pressengeschwindigkeit kein signifikanter Einfluss nachgewiesen werden. Mit Beginn der Vorhersagen schwankte die Pressengeschwindigkeit mit einem Variationskoeffizienten von 3%, durch die Optimierung der Geschwindigkeit erhöhte sich dieser auf 4,7%. Ähnlich dem Leimgehalt zeigt sich die Dickenquellung auch hier bei schon geringeren Schwankungen der Pressengeschwindigkeit als beeinflussbar. Die Querzugfestigkeit reagierte dagegen erst bei einem höheren Geschwindigkeitsniveau mit einer Abnahme der Festigkeit.

Die Temperaturen in den sechs Heizfeldern der Presse schwankten mit Variationskoeffizienten von 0,8 bis 2,5% über den gesamten Zeitraum von 21 Monaten nur sehr gering. Mit einer ausreichenden Sicherheit konnte deren Einfluss nicht nachgewiesen werden. Eine geringere Temperatur des letzten Heizfeldes 6 war jedoch in verschiedenen Modellen mit einer Erhöhung der Querzugfestigkeit und der mittleren Rohdichte und einer Verringerung der Dickenquellung verbunden. Dieser Effekt wäre durch einen verringerten Dampfdruck in der Platte zu erklären, der nach Austritt der Platte aus der Heißpresse zu einer geringeren Rückfederung führt. Eine geringere Rückfederung wirkt sich positiv auf die Platteneigenschaften aus (Heinemann, 2003).

Eine erhöhte Rollstangentemperatur ist in den Modellen mit einer verringerten Rohdichte verbunden.

Obwohl die Querzugfestigkeit neben der Rohdichte die meisten signifikanten Korrelationen zu den Pressenvariablen aufwies (siehe Abbildung 9), konnten in den Modellen nur sehr wenige signifikante Zusammenhänge festgestellt werden. Die direkten Pressenvariablen (ohne Pressengeschwindigkeit) spielten in den Vorhersagemodellen der Querzugfestigkeit eine eher untergeordnete Rolle. Neben dem Multipotdruck am Rahmen 5 zeigte sich i.d.R. nur noch eine Druckvariable aus den mittleren Segmenten der Heißpresse als signifikant.

Die Dickenquellung wurde nach den Modellen dagegen von mehreren Pressenvariablen signifikant beeinflusst. Hier zeigten sich insbesondere Distanzen und Drücke aus der Schließ-, Lüftungs- und Kalibrierphase als bedeutsam. Eine Erhöhung des Multipotdrucks am Rahmen 23 ist in den Modellen mit einer verringerten Dickenquellung verbunden. Im Gegensatz zur Querzugfestigkeit war kein Einfluss der Multipots im vorderen Bereich der Presse nachweisbar.

Für die Abhebefestigkeit und für die Rohdichte zeigte sich über alle Modelle hinweg, die Distanz am Pressenauslauf (Band) bzw. der Druck am Rahmen 22 als sehr bedeutsam. Beide Variablen stehen in der Bewertungsmatrix der optimalen Eingänge an erster Stelle. Im Modell der Rohdichte war die Variable Druck_R22 durch jede andere Druckvariable aus der Kalibrierzone ersetzbar, jedoch war der Zusammenhang zwischen der Rohdichte und dem Druck am Rahmen 22 am stärksten ausgeprägt. Sie ist im Regressionsmodell als "Stellvertretervariable" für die Kalibrierzone zu interpretieren. Danach ist die mittlere Rohdichte um so größer, je höher der Druck bzw. je geringer die Distanz in der Kalibrierzone ausgeprägt war.

Plattenkontrolle

Als Variablen der Plattenkontrolle stand nur die Dickenmessung zur Verfügung. In den Modellen zeigten sie keine Signifikanz. Die Plattendicke ist eine sehr wichtige Kontrollgröße und wird manuell permanent ausgeregelt. Platten, die stark von der Solldicke abweichen, werden automatisch aussortiert und stehen damit der Prozessmodellierung nicht zur Schwankung in der Plattendicke ist damit sehr Verfügung. Die gering. Der Variationskoeffizient betrug innerhalb von 21 Monaten ca. 1%. Diese geringen Schwankungen übten statistisch keinen signifikanten Einfluss aus. In den Modellen der Rohdichte und der Querzugfestigkeit wies die Plattendicke jedoch einen, über alle Vorhersagen hinweg, negativen Regressionskoeffizienten auf. Die Hypothese des negativen Einflusses einer zunehmenden Rückfederung nach der Presse und daraus folgend einer höheren Plattendicke auf die technologischen Eigenschaften wird damit zumindest gestützt.

Tabelle 44 fasst die signifikanten Einflussgrößen in den verschiedenen Modellen nochmals zusammen. Dort sind die Variablen aufgeführt, die sich in der Mehrheit der validierten Modelle als signifikant (S=95%) herausstellten. Die Vorzeichen der Regressionskoeffizienten der Prozessvariablen sind jeweils in den Klammern angegeben. In den Einzelmodellen zeigten sich weitere signifikante Variablen. Diese sind i.d.R. zu denen in der Tabelle 44 aufgeführten Variablen stark korreliert und weisen ähnliche technologische Zusammenhänge auf.

	Querzugfestigkeit	Dickenquellung	Abhebefestigkeit	Rohdichte
Beleimung	(+) Leimhersteller	(-) Leimhersteller	(+) Leimhersteller	(-) Härtermenge
	(+) Leimmenge	(-) Leimmenge	(-) Härtermenge	
	(-) Härtermenge	(+) Härtermenge		
Kocher/	(+) Strom Dosierschnecke	(+) Strom Dosierschnecke	(+) Strom Dosierschnecke	(+) Dampf Mahlraum
Refiner		(-) Dampf Kocher (Gesamt)		(+) Dampf HKS
				(-) Füllstand Kocher
Trockner/	(-) Temp. Trocknerausgang	(-) Feuchte nach Trockner	(-) Temp. Trocknerausgang	(-) Temp. Register 1
Sichter	(-) Füllstand Faserbunker	(+) Temp. Register 3	(-) Füllstand Faserbunker	(+) Füllstand Faserbunker
Formung	(-) Temp. Streubunker	(+) Temp. Streubunker	(-) Temp. Streubunker	(-) Temp. Streubunker
	(+) Mattenfeuchte	(+) Mattenbreite		(+) Mattenfeuchte
Vorpresse	(+) Strom Bandantrieb unten	(-) Strom Bandantrieb unten		(+) Strom Bandantrieb unten
	(+) Druck Vorpresse oben		(+) Druck Vorpresse oben	(+) Druck Vorpresse oben
Presse	(-) Pressengeschwindigkeit	(+) Pressengeschwindigkeit	(+) Distanz Band Auslauf	(-) (Temp. Heizfeld 6)
	(-) (Temp. Heizfeld 6)	(+) (Temp. Heizfeld 6)	(+) Druck Rahmen 26	(-) Temp. Rollstangen
	(-) Druck Rahmen 9	(-) Druck Band Auslauf		(+) Druck Rahmen 22
	(+) Multipot Rahmen 5	(-) Druck Auslauf		
		(+) Distanz Rahmen 6		
		(+) Druck Rahmen 2		
		(-) Druck Multipot 23		

Tabelle 44: Signifikante Einflussgrößen in den Modellen

Erkenntnisse

Eine technologische Interpretation von Regressionsmodellen, die auf eine Auswahl korrelierter Prozessvariablen beruht, ist im Einzelnen nicht immer direkt möglich. Die Angabe des Einflusses (in Form des standardisierten Regressionskoeffizienten bzw. des Elastizitätskoeffizienten) einer einzelnen Variable auf die Zieleigenschaft ist, bei vorhandenen Korrelationen zu anderen Prozessvariablen, oftmals irreführend. Jedoch können durch die Modelle Effekte nachgewiesen werden, bei der eine Variable nur eine Realisation des Effektes darstellt. Die Interpretation bzw. Diskussion einer Variable im Modell ist damit "stellvertretend" für diesen Effekt zu führen.

Bei vorliegenden Fachkenntnissen, kann dieser Effekt beschrieben werden. So ist z.B. nicht direkt eine Temperaturerhöhung im Streubunker für die Verringerung der Festigkeiten und für die Erhöhung der Dickenquellung verantwortlich, sondern der Effekt der zunehmenden Klebstoffaushärtung. Die verschiedenen Temperaturmessungen im Trocknungs,- Sichtungsbzw. Streuungsprozess sind messbare Realisationen dieses Effektes. Die Temperatur im Streubunker und am Trocknerausgang konnten diesen Einfluss im vorliegenden Fall am genauesten beschreiben.

Insbesondere die Variablenauswahl der Heißpresse in die Regressionsmodelle und deren Interpretation kann nur stellvertretend für einen Pressenabschnitt bzw. für die gesamte Presse geführt werden. Eine unabhängige Einstellung benachbarter Pressenrahmen ist durch die Konstruktionsbauweise der Heißpresse stark eingegrenzt und damit verantwortlich für die hohen Korrelationskoeffizienten zwischen den einzelnen Pressenrahmen. Generell zeigte sich die Auswahl der Pressenvariablen durch die Bewertungsmatrix der optimalen Eingänge für die Vorhersage als geeigneter als eine technologische Auswahl.

Schwierig wird eine Interpretation von Variablen, bei der kein Erfahrungswissen vorliegt bzw. deren Einfluss durch Effekte anderer (messbar und nicht messbar) Variablen verstärkt oder aufgehoben werden kann. Zu nennen sind hier die Stromaufnahmen der Beschickschnecke im Kocher und des Bandantriebes der Vorpresse. Beide Parameter zeigten signifikante Zusammenhänge zu den Schwankungen in den Platteneigenschaften.

Die absoluten Einflüsse der Variablen im Regressionsmodell sind entscheidend von der Variablenauswahl abhängig. Erschwerend kommt hinzu, dass, in Abhängigkeit der Veränderungen in den Prozesseinstellungen im Vorhersageraum, sich die Einflüsse und damit die Signifikanzen der Variablen im Vorhersagebetrieb ändern. Es ist daher nicht zielführend, als alleiniges Kriterium zur Modellerstellung und Interpretation die Signifikanz der Variablen im Beobachtungsraum heranzuziehen. Um so stärker die Variablen einer Drift ausgesetzt sind, desto eher ändern sich die Einflüsse und Signifikanzen im Vorhersageraum.

9.3.1 Ziele der Untersuchungen

Mit der Faktorenanalyse ist es möglich, korrelierte Daten zu übergeordneten (latenten) Variablen (Faktoren) zusammenzufassen. Diese Faktoren sind untereinander nicht korreliert, ihre Kovarianzen sind Null.

In der Regressionsanalyse bereiten korrelierte Variablen große Schwierigkeiten bei der Berechnung der Regressionskoeffizienten. Idealerweise sind die Regressoren unabhängig voneinander. In diesem Fall kann der Einfluss einer Variablen unabhängig von den anderen Variablen auf die Zielgröße bestimmt werden. Dieser Fall ist jedoch nur in Form eines Versuchsplans (Experiment) einzuhalten.

Die vorliegenden Prozessvariablen sind dagegen miteinander korreliert und nicht unabhängig voneinander eingestellt. Die kontinuierliche Heißpresse ist aufgrund ihrer Rahmenbauweise davon am meisten betroffen. Jeder einzelne Pressrahmen wird innerhalb eines Pressprogramms mit Sollwertvorgaben eingestellt. Aus den vorliegenden Pressenvariablen können nur vereinzelte Variablen in den Regressionsmodellen untersucht werden. Eine Beschreibung des Einflusses des gesamten Pressprogramms auf die technologischen Eigenschaften der Platten ist innerhalb von Regressionsmodellen nicht möglich.

Mit der Faktorenanalyse soll im ersten Schritt aufgezeigt werden, inwiefern sich das Pressprogramm durch Faktoren bzw. Komponenten beschreiben lässt. Die Interpretation dieser Faktoren ist durch entsprechende Rotationsverfahren darzustellen.

Im zweiten Schritt werden die Faktoren auf ihre statistische Signifikanz innerhalb der bestehenden Regressionsmodelle untersucht. Dazu werden die einzelnen Pressenvariablen des Distanz- und Druckprofils durch die extrahierten Faktoren des Pressprogramms aus der Faktorenanalyse ersetzt. Als Basismodelle werden das technologische und das kombinierte Modell herangezogen. In diesen Modellen werden die ausgewählten Pressenvariablen durch die extrahierten Faktoren Faktoren Faktoren ersetzt.

Der Einfluss der signifikanten Pressenfaktoren auf die technologischen Eigenschaften wird diskutiert und mit den bestehenden Regressionsmodellen verglichen.

9.3.2 Faktoranalytische Untersuchungen des Pressprogramms

Die Variablen der Heißpresse sind aufgrund eines festen Pressprogramms stark miteinander korreliert. Die Höhe der absoluten Korrelationen nehmen mit zunehmender Entfernung zwischen den Rahmen ab. Abbildung 31 zeigt den Verlauf der Korrelationskoeffizienten ausgewählter Distanzen zu allen anderen Distanzen der Heißpresse. Die Nullhypothese, das der Korrelationskoeffizient gleich Null ist, wird bei einem Signifikanzniveau von 95%, bei einem Korrelationskoeffizienten r>0,17 abgelehnt (n=132). Die Distanzen der Rahmen 8 bis 16 weisen einen Korrelationskoeffizienten größer 0,9 untereinander auf. Entsprechend flach fällt der Korrelationskoeffizient des Rahmen 13 zu den benachbarten Pressenrahmen ab. Dieser Bereich der Presse zeichnet sich durch einen sehr gleichmäßigen Verlauf der Distanzen nach dem Schließen und Lüften und vor dem zweiten Verdichtungsschritt ab Rahmen 17 aus. In diesem Bereich lässt sich die Information der Distanz aus jeder anderen Distanz fast vollständig ableiten. In der Faktoren bzw. Komponentenanalyse kann dieser Pressenbereich durch einen Faktor bzw. Komponente erklärt werden. (Zur Vereinfachung wird im folgenden immer von Faktoren gesprochen, auch bei einer Hauptkomponentenanalyse. Zur eindeutigen Unterscheidung der Hauptfaktorenanalyse (PFA) von der Hauptkomponentenanalyse (PCA) werden die Abkürzungen mit angegeben).



Abbildung 31: Grafische Darstellung der Korrelationen ausgewählter Distanzen

Prinzipiell lässt sich im jeweiligen Pressenabschnitt die Information eines Rahmens durch die Information eines benachbarten Rahmens sehr gut ersetzen. Eine Ausnahme davon gilt für den Rahmen 20. In der Abbildung 31 kann für diesen Rahmen ein starker Abfall der Korrelationen zu den benachbarten Rahmen festgestellt werden. So weisen die benachbarten Rahmen einen Korrelationskoeffizienten von 0,60 (Rahmen 21) bzw. 0,62 (Rahmen 19) auf. Der Rahmen 20 bildet im Pressprogramm den Abschluss des zweiten Verdichtungsschrittes und weist sowohl in der Distanz- als auch in der Druckkurve den schärfsten Knick zu den benachbarten Rahmen auf. Dieser Rahmen hat die höchste

merkmalseigene Varianz und kann in der Faktorenanalyse am ungenauesten durch die Faktoren erklärt werden.

Zunächst sind für die Distanzen und Drücke getrennt voneinander jeweils für die rechte und linke Seite eine PCA- und PFA-Analyse berechnet worden. Anschließend erfolgt eine gemeinsame faktoranalytische Betrachtung der linken und der rechten Seite. In Tabelle 45 sind die Ergebnisse der PCA-Analyse und in Tabelle 46 die der PFA-Analyse zusammengefasst. Alle Berechnungen basieren auf den Lerndatensatz der Faserplattenproduktion. Die Variablen werden in der Faktorenanalyse als Merkmale bezeichnet.

Tabelle 45: Eigenwerte der ersten 6 Hauptfaktoren der PCA-Analyse für die Korrelationsmatrix der Distanzen und Drücke getrennt nach rechter und linker Seite

Eige	Eigenwerte der Korrelationsmatrix Dist. rechts:					
Tota	al = 29 Durch	schnitt = 1,	M=29			
Eigenwert Differenz Anteil Summe						
1	17,38	13,71	0,599	0,599		
2	3,67	0,67	0,127	0,726		
3	2,99	1,05	0,103	0,829		
4	1,95	0,89	0,067	0,896		
5	1,06	0,46	0,037	0,933		
6	0,60	0,15	0,021	0,953		

Eige	Eigenwerte der Korrelationsmatrix Druck rechts:						
Tota	I = 27 Durch	schnitt = 1,	M=27				
	Eigenwert Differenz Anteil Summe						
1	17,03	14,47	0,631	0,631			
2	2,55	0,57	0,095	0,725			
3	1,98	0,45	0,073	0,799			
4	1,53	0,63	0,057	0,855			
5	0,90	0,25	0,033	0,889			
6	0,65	0,11	0,024	0,913			

Eige	Eigenwerte der Korrelationsmatrix Dist. links:					
Tota	al = 29 Durch	schnitt = 1,	M=29			
	Eigenwert Differenz Anteil Summe					
1	19,93	16,60	0,687	0,687		
2	3,33	1,31	0,115	0,802		
3	2,02	0,91	0,070	0,871		
4	1,11	0,01	0,038	0,910		
5	1,09	0,62	0,038	0,947		
6	0,47	0,14	0,016	0,964		

Eige	Eigenwerte der Korrelationsmatrix Druck links:					
Tota	al = 27 Durch	schnitt = 1,	M=27			
	Eigenwert	Differenz	Anteil	Summe		
1	17,94	15,51	0,664	0,664		
2	2,43	0,60	0,090	0,754		
3	1,83	0,54	0,068	0,822		
4	1,29	0,41	0,048	0,870		
5	0,88	0,26	0,033	0,903		
6	0,61	0,20	0,023	0,925		

Die Spalte "Eigenwert" gibt den absoluten Anteil des Faktors an der Gesamtvarianz der Ausgangsdatenmatrix an. Da diese standardisiert ist und damit jedes Merkmal eine Varianz von eins besitzt, ist in der PCA-Analyse die Gesamtvarianz gleich der Anzahl der Merkmale (M). Der prozentuale Anteil der erklärten Varianz durch den jeweiligen Faktor findet sich in der Spalte "Anteil". Die Spalte "Summe" gibt den aufsummierten Anteil der erklärten Varianz durch die Faktoren an. Es wird ersichtlich, dass die ersten 5 Faktoren 93,3% der gesamten Varianz des Distanzenprofils der rechten Seite und 94,7% der linken Seite erklären. Im Distanzprofil wurden neben den Rahmen 1 bis 27 noch der Einlauf (Rahmen 0) und der Auslauf (Rahmen 28) berücksichtigt. Im Druckprofil dagegen blieb der Einlauf- und der Auslaufrahmen unberücksichtigt. Das Druckprofil wird insgesamt geringfügig schlechter durch die ersten 5 Faktoren zusammengefasst. Die ersten 5 Faktoren fassen 88,9% der gesamten Varianz des Druckprofils der rechten Seite und 90,3% der linken Seite zusammen. Der Faktor 6 kann vernachlässigt werden, da der Zuwachs an erklärter Varianz mit 1,6% bis

2,4% gering ausfällt. Eine Berücksichtigung eines sechsten Faktors führt zu einem Einzelrestfaktor, der mit dem Rahmen 20 übereinstimmt. Die geringere erklärte Varianz des Druckprofils wird im Vergleich zu dem Distanzprofil durch die geringeren Korrelationen der Drücke untereinander verursacht. Die höheren Korrelationen zwischen den Distanzen erklärt sich durch die isochore Pressensteuerung. Die Drücke werden demnach im Gegensatz zu den Distanzen durch Schwankungen in den Mattenfeuchte und durch Streuungsungenauigkeiten beeinflusst.

In der PFA-Analyse entspricht die Gesamtvarianz nicht der Anzahl der Merkmale. Da bei jedem Merkmal die merkmalseigenen Varianzen abgezogen werden, ist die Gesamtvarianz stets kleiner als die Anzahl der Merkmale. Die Gesamtvarianz und die durchschnittliche Varianz je Faktor ist jeweils im Tabellenkopf angegeben. Aufgrund des Abzuges der merkmalseigenen Varianzen ist die erklärte Varianz durch die ersten 5 Faktoren in der PFA-Analyse höher als in der PCA-Analyse (siehe Tabelle 46). Dies macht sich insbesondere bei den weniger stark korrelierten Pressdrücken bemerkbar. So erklären die ersten 5 Faktoren der PFA-Analyse insgesamt 94% der gesamten Varianz des Druckprofils der rechten Seite. In der PCA-Analyse sind es dagegen nur 88,9%.

Tabelle 46: Eigenwerte der ersten 6 Hauptfaktoren der PFA-Analyse für die Korrelations	3-
matrix der Distanzen und Drücke getrennt nach rechter und linker Seite	

Eige	Eigenwerte der red. Korrelmatrix Dist. rechts:						
Tota	ıl = 28,44 Du	rchschnitt =	= 0,98; M=2	29			
	Eigenwert Differenz Anteil Summe						
1	17,37	13,72	0,611	0,611			
2	3,64	0,66	0,128	0,739			
3	2,98	1,05	0,105	0,843			
4	1,93	0,89	0,068	0,911			
5	1,04	0,47	0,037	0,948			
6	0,57	0,17	0,020	0,968			

Eigenwerte der red. Korrelmatrix Druck rechts:								
Total = 25,13 Durchschnitt = 0,93; M=27								
	Eigenwert Differenz Anteil Summe							
1	16,97	14,50	0,675	0,675				
2	2,47	0,55	0,098	0,774				
3	1,92	0,47	0,077	0,850				
4	1,45	0,65	0,058	0,908				
5	0,80	0,27	0,032	0,940				
6	0,53	0,15	0,021	0,961				

Eigenwerte der red. Korrelmatrix Dist. links:								
Tota	Total = 28,53 Durchschnitt = 0,98; M=29							
	Eigenwert Differenz Anteil Summe							
1	19,91	16,61	0,698	0,698				
2	3,31	1,31	0,116	0,814				
3	2,00	0,90	0,070	0,884				
4	1,09	0,02	0,038	0,922				
5	1,07	0,64	0,038	0,960				
6	0,43	0,12	0,015	0,975				

Eigenwerte der red. Korrelmatrix Druck links:								
Tota	Total = 25,16 Durchschnitt = 0,93; M=27							
	Eigenwert Differenz Anteil Summe							
1	17,88	15,53	0,711	0,711				
2	2,35	0,59	0,093	0,804				
3	1,76	0,56	0,070	0,874				
4	1,20	0,40	0,048	0,922				
5	0,79	0,30	0,032	0,953				
6	0,50	0,22	0,020	0,973				

Zur Interpretation der Faktoren ist das rotierte Faktorenmuster heranzuziehen. Unter Interpretation der Faktoren wird die Zuordnung der Variablen (Merkmale) zu den Faktoren verstanden. Tabelle 47 zeigt das rotierte Faktorenmuster (Ladungsmatrix) der 5 PCA-Hauptfaktoren zur Erklärung des Distanzenprofils rechts. Die Ladungsmatrizen der Distanzen links und der Drücke rechts und links finden sich im Anhang 3. Tabelle 47 zeigt exemplarisch die Zuordnung der Faktoren zu den einzelnen Distanzen. Eine Verbindung zwischen Faktoren und Merkmalen ergibt sich durch die "Hochladungen". Die höchsten Faktorladungen eines Merkmals zeigen eine Zugehörigkeit zum jeweiligen Faktor an. Je höher die Faktorladung (maximal=1), desto besser wird das Merkmal durch den Faktor beschrieben. Die Zugehörigkeit der Merkmale zu den Faktoren ist durch Fettdruck in den Tabellen markiert.

Tabelle 47: Rotiertes F	aktorenmuster der	· Methode `	Varimax für	5 Faktoren ((Distanzen
rechts)					

Faktorenmuster Methode Varimax							
	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5		
DIST A R	0,76	0,25	0,21	0,13	0,34		
DIST R27	0,89	0,21	0,07	0,13	0,21		
DIST R26	0,91	0,18	0,29	0,19	0,05		
DIST R25	0,91	0,18	0,29	0,19	0,04		
DIST R24	0,91	0,18	0,30	0,19	0,05		
DIST R23	0,91	0,18	0,29	0,19	0,05		
DIST R22	0,88	0,23	0,31	0,20	0,06		
DIST R21	0,85	0,26	0,29	0,17	0,07		
DIST R20	0,32	0,19	0,15	0,36	0,61		
DIST R19	0,02	0,07	0,15	0,03	0,96		
DIST R18	0,07	0,13	0,32	-0,02	0,88		
DIST R17	0,28	0,17	0,66	0,09	0,63		
DIST R16	0,39	0,26	0,65	0,08	0,52		
DIST R15	0,49	0,25	0,71	0,11	0,38		
DIST R14	0,51	0,26	0,71	0,14	0,30		
DIST R13	0,53	0,26	0,70	0,16	0,29		
DIST R12	0,51	0,21	0,77	0,19	0,18		
DIST R11	0,51	0,22	0,77	0,18	0,18		
DIST R10	0,23	0,71	0,60	0,09	0,16		
DIST R9	0,23	0,71	0,59	0,09	0,16		
DIST R8	0,21	0,87	0,33	0,10	0,16		
DIST R7	0,23	0,90	0,22	0,16	0,17		
DIST R6	0,22	0,91	0,14	0,20	0,09		
DIST R5	0,24	0,88	0,08	0,27	0,10		
DIST R4	0,25	0,76	0,00	0,51	0,02		
DIST R3	0,20	0,54	0,07	0,77	-0,01		
DIST R2	0,14	0,47	0,02	0,84	0,01		
DIST R1	0,26	0,12	0,18	0,91	0,12		
DIST E R	0,27	0,10	0,18	0,91	0,13		

Nach der Analyse des rotierten Faktorenmusters kann das Pressprogramm in fünf Abschnitte untergliedert werden. Diese Abschnitte sind in der Tabelle 48 dargestellt. Im Distanzenprofil der rechten und linken Seite ergeben sich leichte Verschiebungen der Rahmen in den Abschnitten "Lüften" und "Halten". Diese Verschiebungen sind durch Distanzunterschiede in der linken und rechten Seite in den Rahmen 4 bis 10 der Presse zu erklären. Dieser Unterschied wurde durch einer leichten "Kippstellung" der Presse zwischen der linken und rechten Seite Kippstellung ist nach ca. 85 Beobachtungen beseitigt

worden. Abbildung 32 zeigt den Verlauf der Distanzen und Drücke der rechten und linken Seite des Rahmens 5. Im Distanzenprofil sind die unterschiedlichen Distanzen zwischen der rechten und linken Seite bis zum Datensatz 85 zu sehen. Im Druckprofil sind diese Unterschiede nicht zu finden. Als Folge der Unterschiede in den Streuungen der rechten und linken Seite im Distanzprofil, werden die einzelnen Pressenabschnitte durch unterschiedliche Faktoren beschrieben. Das Druckprofil dagegen wird sowohl auf der linken als auch auf der rechten Seite durch die gleichen Faktoren gekennzeichnet (siehe Tabelle 48).

Tabelle 48: Interpretation und Zuordnung der Faktoren (PCA) für das Distanzen- ur	nd
Druckprofil links und rechts (Rotation Varimax)	

Rahmennummer		Pressenabschnitt	beschrieben durch Faktor.	
Dist. rechts	Dist. links		rechts	links
0 - 3	0 - 3	"Schließen"	4	3
4 - 10	4 - 7	"Lüften"	2	4
11 - 17	8 - 17	"Halten"	3	1
18 - 20	18 - 20	"zweiter Verdichtungsschritt"	5	5
21 - 28	21 - 28	"Kalibrierzone"	1	2

Rahmennummer		Pressenabschnitt	beschrieben durch Faktor	
Druck rechts	Druck links		rechts	links
0 - 3	0 - 3	"Schließen"	5	5
4 - 8	4 - 8	"Lüften"	3	3
9 – 15	9 – 15	"Halten"	2	2
16 – 19	16 – 19	"zweiter Verdichtungsschritt"	4	4
20 - 27	20 - 27	"Kalibrierzone"	1	1

(Die Rahmen 16 und 17 werden zu gleichen Anteilen von Faktor 2 und 4 beschrieben)

Die Ergebnisse der Hauptfaktorenanalyse (PFA) stimmen mit den Resultaten der Tabelle 48 überein.



Abbildung 32: Distanzen (links) und Drücke (rechts) des Rahmens 5

Die Abbildung 33 verdeutlicht grafisch die Zuordnung der Merkmale (Distanzen) zu den ersten 4 Faktoren der PCA-Analyse nach der Varimax-Rotation. Faktor 5 bzw. der zweite Verdichtungsschritt ist nicht dargestellt. Je höher die Faktorladungen der Variablen, desto besser werden die Variablen durch die jeweiligen Faktoren repräsentiert.



Abbildung 33: Grafische Repräsentation des Distanzenprofils bzw. der PCA-Faktoren nach der Varimax-Rotation

Tabelle 49 stellt die erklärten Varianzen der Ausgangsdatenmatrix durch die jeweiligen Faktoren vor und nach der Rotation dar. Die Varianzen entsprechen dabei genau den Eigenwerten (siehe Tabelle 45). Die Summe der absoluten erklärten Varianz wird durch eine Rotation nicht beeinflusst. In der Zeile "Prozent" ist nochmals die erklärte Varianz in Prozent durch die 5 Faktoren an der Gesamtvarianz der Ausgangsdatenmatrix angegeben. Abbildung 34 veranschaulicht grafisch die Verteilung der Varianzen für das Beispiel Distanzen rechts.

	Distanz I	inks	Distanz r	echts	Druck	links	Druck I	rechts
Rotation	vor	nach	vor	nach	vor	nach	vor	nach
Faktor1	19,9	8,0	17,4	8,4	17,9	7,2	17,0	7,2
Faktor2	3,3	7,5	3,7	6,1	2,4	6,4	2,6	6,1
Faktor3	2,0	4,9	3,0	5,2	1,8	4,4	2,0	4,8
Faktor4	1,1	3,7	1,9	3,9	1,3	3,3	1,5	3,1
Faktor5	1,1	3,4	1,1	3,4	0,9	3,1	0,9	2,8
Summe	27,5	27,5	27,1	27,1	24,4	24,4	24,0	24,0
Prozent	94,7	94,7	93,3	93,3	90,3	90,3	88,9	88,9

Tabelle 49: Erklärte Varianzen durch die PCA-Faktoren vor und nach der Rotation Varimax



Abbildung 34: Erklärte Varianz durch die Faktoren vor und nach der Varimax-Rotation (Distanzen rechts)

Vor einer Rotation wird immer durch den ersten Faktor ein Maximum der Varianz der Ausgangsmerkmale erklärt. Jeder weitere Faktor nimmt einen abnehmenden Varianzanteil auf. Diese mathematische Vorgehensweise hat den Nachteil, dass die Faktoren nicht interpretierfähig sind. Eine Rotation verteilt die Gesamtvarianz auf die Faktoren so, dass eine Zuordnung der Merkmale auf die Faktoren möglich ist. Diese Zuordnung ist je nach Rotationsverfahren unterschiedlich. Da im vorliegenden Fall die Merkmale selbst bekannt sind, ist eine interpretierbare Zuordnung zu den Faktoren sehr einfach.

Die Verteilung der Varianz auf die einzelnen Merkmale ist am Beispiel der Distanzen rechts in der Abbildung 35 dargestellt. Dort wird ersichtlich, zu welchen Anteil die Varianz der standardisierten einzelnen Merkmale durch die Faktoren erklärt werden. Dies entspricht genau der Kommunalität (Anteil der erklärten Varianz eines standardisierten Merkmals durch die Faktoren) der einzelnen Merkmale. Die einzelnen Kommunalitäten ergeben aufsummiert die dargestellte Summe in der Tabelle 49.


Abbildung 35: Kommunalitäten der Distanzen rechts der Presse; PCA mit 5 Faktoren

Die Varianzen der einzelnen Distanzen werden überwiegend zu über 90 % durch die fünf Hauptfaktoren beschrieben. Eine einzige, stärkere Ausnahme davon bildet der Pressenrahmen 20, der die höchste merkmalseigene Varianz und damit die geringste Kommunalität aufweist. Dieser Rahmen wird durch die Faktoren am geringsten erklärt, da diese Distanz die geringsten Korrelationen zu den anderen Distanzen aufweist. Im Pressprogramm zeigt sich dort das Ende des zweiten Verdichtungsschrittes und zugleich der "schärfste" Übergang von einem Rahmen zum nächsten.

Eine gemeinsame faktoranalytische Betrachtung der Distanzen der rechten und linken Seite, führte aufgrund der Kippstellung der Presse, zu einer nicht eindeutigen Zuordnung des Bereiches "Lüften" zu einem gemeinsamen Faktor. Alle anderen Bereiche der Presse konnten durch gemeinsame Faktoren beschrieben werden. Das Druckprofil der rechten und linken Seite konnte dagegen in der Faktorenanalyse gemeinsam betrachtet werden. Tabelle 50 fasst die Ergebnisse der PCA- und der PFA-Analyse des Druckprofils, für beide Seiten gemeinsam, zusammen. Insgesamt werden durch die ersten 5 Faktoren ca. 89% der Gesamtvarianz der 54 Druckvariablen der Presse erklärt. Die Zuordnung der Faktoren zu dem Merkmalen stimmt mit den Angaben der Tabelle 48 überein.

Eigenwerte der Korrelationsmatrix Druck								
Tota	Total = 54 Durchschnitt = 1, M=54							
Eigenwert Differenz Anteil Summe								
1	34,70	29,81	0,643	0,643				
2	4,88	1,08	0,090	0,733				
3	3,80	0,99	0,070	0,803				
4	2,81	1,04	0,052	0,856				
5	1,78	0,51	0,033	0,888				
6	1,27	0,34	0,024	0,912				

Eigenwerte der reduzierten Korrel.-matrix Druck Total = 53.67 Durchschnitt = 0.99; M=54 Differenz Anteil Eigenwert Summe 34,69 29,81 0,646 0,646 1 4,88 1,08 0,091 0,737 2 3,79 0,99 0,071 0,808 3 1,04 0.052 0.860 2,81 4 1,77 0,50 0,033 0,893 5 1,27 0,35 0,024 0,917 6

Tabelle 50: Eigenwerte der ersten 6 Hauptfaktoren der PCA-Analyse (links) und der PFA-Analyse (rechts) für die Korrelationsmatrix Drücke der rechten und linken Seite In der Faktorenanalyse gibt es keine eindeutige Regel, wie viel Faktoren zur Erklärung der Merkmale zu verwenden sind. Die Anzahl der Faktoren wird durch das Ziel der Untersuchung und den Genauigkeitsanforderungen mit beeinflusst. Steht eine Interpretation der Faktoren im Vordergrund, sollte diese die Auswahl der Faktoren mit bestimmen. Bisher ist das Pressprogramm mit fünf Faktoren erfasst worden. Eine Orientierung zur Auswahl der Faktoren bietet die Höhe des Eigenwertes. Danach werden alle Faktoren mit Eigenwerten größer eins verwendet. Nach dieser Regel liegt je nach Modell und Variable die Anzahl zwischen vier und fünf Faktoren. Der Zuwachs an erklärter Varianz von ca. 3% durch den fünften Faktor fällt im Vergleich zu den ersten vier Faktoren relativ gering aus. Eine Einteilung des Pressprogramms in vier Abschnitte ist ebenso möglich. Dabei werden die Abschnitte "Lüften" und "Halten" zu einem gemeinsamen Faktor "Aufwärmen" zusammengefasst. Die Ladungsmatrizen bzw. die rotierten Faktorenmuster der 4 PCA-Hauptfaktoren sind beispielhaft im Anhang 4 zur Erklärung der Distanzen links und der spezifischen Drücke dargestellt.

Tabelle 51 zeigt die Zuordnung der einzelnen Rahmen zu den Faktoren und deren Interpretation, die sich durch die Auswertung der rotierten Faktorenmustermatrix ergeben. Die Equamax-Rotation ergab im Vergleich zur Varimax-Rotation eine eindeutigere Zuordnung der Variablen zu den Faktoren.

Tabelle 51: Interpretation und Zuordnung der Faktoren (PCA) für das Distanzenprofil links
und spezifisches Druckprofil (Rotation Equamax)

Rahmennummer Distanz links	Pressenabschnitt	beschrieben durch Faktor
0 - 4	"Schließen"	3
5 - 15	"Erwärmen"	2
16 – 20	"zweiter Verdichtungsschritt"	4
21 – 28	"Kalibrierzone"	1

Rahmennummer spez. Druck	Pressenabschnitt	beschrieben durch Faktor
1 - 4	"Schließen"	4
5 - 15	"Lüften"	2
16 – 19	"zweiter Verdichtungsschritt"	3
20 - 27	"Kalibrierzone"	1

Erkenntnisse

Das Pressprogramm kann mit vier bzw. fünf Faktoren mit einer durchschnittlich erklärten Varianz von 90-95% beschrieben werden. Die fünf Faktoren definieren die Abschnitte des Pressprogramms "Schließen", "Lüften", "Halten", "zweiter Verdichtungsschritt" und "Kalibrierzone". Bei einer Vier-Faktorenlösung werden die Faktoren "Schließen" und "Lüften" zu einem gemeinsamen Faktor "Aufwärmen" zusammengeführt (siehe Anhang 3 und 4).

Für eine eindeutige Zuordnung der Variablen zu den Faktoren ist die Faktorenanalyse für das Distanz- und Druckprofil getrennt voneinander zu berechnen. Die Drücke der rechten und linken Seite der Presse konnten im Druckprofil gemeinsam berücksichtigt werden. Die Verwendung der spezifischen Drücke in der Faktorenanalyse ergibt gegenüber den

absoluten Drücken keinen Unterschied. Das Distanzprofil musste aufgrund einer Kippstellung der Presse und einer damit verbundenen unterschiedlichen Distanzmessung auf beiden Seiten getrennt für die linke bzw. rechte Seite berechnet werden.

Die Ergebnisse der Faktorenanalyse (PFA) unterscheiden sich nur unwesentlich von denen der Komponentenanalyse (PCA). Der Grund dafür liegt in den sehr geringen merkmalseigenen Varianzen der Pressenvariablen.

9.3.3 Einfluss der Fortschreibung des Datensatzes

Die berechneten Faktoren bilden in einer Regressionsanalyse die Regressoren der Heißpresse. Bei einer erwünschten Interpretation der Faktoren, muss zu jeder Zeit sichergestellt sein, welche Abschnitte der Presse durch welche Faktoren beschrieben werden. Da nach jeder ex-ante-Vorhersage der entsprechende Datensatz zur Aktualisierung der Parametermatrix an den Lerndatensatz angehängt wird, kann es zu einer Verschiebung in der Zuordnung der einzelnen Merkmale zu den Faktoren kommen. Es musste daher geprüft werden, ob sich während der Vorhersagen ein Wechsel in der Zuordnung der Faktoren zu den Variablen ergab. Dazu wurde nach jeder ex-ante-Vorhersage das rotierte Faktorenmuster untersucht, ob sich dort Änderungen in der Zuordnungsmatrix ergaben. Dargestellt werden die Ergebnisse der Equamax-Rotation anhand der spezifischen Drücke und der Distanzen links. Ähnliche Ergebnisse wurden auch mit der Varimax-Rotation erzielt.

Die Equamax-Rotation versucht im Vergleich zur Varimax-Rotation die Gesamtvarianz gleichmäßig über die Faktoren zu verteilen. Dies entspricht weitgehend dem Verhalten der einzelnen Ausgangsvariablen. Abbildung 36 zeigt die Verteilung der erklärten Varianz vor und nach der Rotation sowie bei Beginn am Ende der Vorhersagen für das Distanzprofil links. Die erklärte Gesamtvarianz ändert sich über die ex-ante-Vorhersagen nicht. Die Verteilung der Varianzen auf die einzelnen Faktoren verschiebt sich geringfügig während des Vorhersagebetriebs. Am Ende der Vorhersagen erklärt der erste Faktor vor der Rotation 22,5% der Gesamtvarianz im Vergleich zu 19,9% zu Beginn der Vorhersagen.



Abbildung 36: Erklärte Varianz durch die Faktoren vor und nach der Rotation (Distanzen. links) Methode Equamax; links: Beginn Vorhersage, rechts: Ende Vorhersage

Tabelle 52 zeigt die Interpretation und Zuordnung der Faktoren zu Beginn und Ende des Vorhersagebetriebs. Im Distanzenprofil links kommt es zu einem Wechsel der Faktorenzuordnung in den Abschnitten "Schließen" und Lüften". Wird zu Beginn der

Vorhersage der Abschnitt "Schließen" durch Faktor vier beschrieben, ist es am Ende der Vorhersage der Faktor drei. In den Regressionsmodellen ist für die Interpretation der Ergebnisse dieser Wechsel in der Zuordnung zu beachten.

In der Zuordnung der spezifischen Drücke zu den Faktoren erfolgte während des Vorhersagebetriebes keine Änderung.

Tabelle 52: Interpretation und Zuordnung der Faktoren (PCA) für das spezifische Druck- und Distanzenprofil links zu Beginn und Ende der Vorhersage

Methode Equamax						
Rahmennummer Distanz links		Pressenabschnitt	beschrieben durch Faktor			
Beginn	Ende		Beginn	Ende		
Vorhersage	Vorhersage		Vorhersage	Vorhersage		
0 - 3	0 - 3	"Schließen"	4	3		
4 - 8	4 - 8	"Lüften"	3	4		
9 – 15	9 – 15	"Halten"	2	2		
16 – 20	16 – 20	"zweiter Verdichtungsschritt"	5	5		
21 - 28	20 - 28	"Kalibrierzone"	1	1		

Methode Equamax						
Rahmennummer spez. Druck		Pressenabschnitt	beschrieben durch Faktor			
Beginn	Ende		Beginn	Ende		
Vorhersage	Vorhersage		Vorhersage	Vorhersage		
1 - 3	1 - 3	"Schließen"	5	5		
4 - 8	4 - 8	"Lüften"	3	3		
9 – 15	9 – 15	"Halten"	2	2		
16 – 19	16 – 19	"zweiter Verdichtungsschritt"	4	4		
20 - 27	20 - 27	"Kalibrierzone"	1	1		

Erkenntnisse

Die Zuordnung der Faktoren zu den Merkmalen kann sich durch die Aktualisierung der Datenmatrix ändern. Die Equamax-Rotation der Faktoren führte zu einer eindeutigen Identifizierung der Merkmale und wies im 4-Faktorenmodell während der Datenaktualisierung im fortlaufenden Vorhersagebetrieb keine Änderung in der Zuordnung auf. Im 5-Faktorenmodell der Distanzen änderte sich die Zuordnung der Abschnitte "Schließen" und "Lüften". Diese Änderung der Zuordnung ist in der Interpretation der Ergebnisse bzw. in den Regressionsmodellen zu berücksichtigen.

9.3.4 Regressionsmodelle mit Pressenfaktoren der PCA-Analyse

Durch Verwendung der unkorrelierten Faktoren der PCA- bzw. PFA-Analyse der Heißpresse in den Regressionsmodellen können Hinweise gewonnen werden, welche Abschnitte der Heißpresse einen statistisch signifikanten Einfluss auf die jeweiligen Platteneigenschaften ausüben. Die Vorzeichen der Regressionskoeffizienten der Pressenfaktoren zeigen den Einfluss des jeweiligen durch den Faktor repräsentierten Pressenabschnittes auf die technologischen Eigenschaften.

In den Regressionsmodellen ergaben sich für die Schätzgenauigkeiten im Vorhersageraum keine Unterschiede in Abhängigkeit der Verwendung von PFA- bzw. PCA-Faktoren der Heißpresse. Es wurden die Faktoren des Distanzenprofils und die der spezifischen Drücke unabhängig voneinander in das technologische Modell bzw. in das kombinierte Modell integriert. Dabei wurden die einzelnen Pressenvariablen der Distanzen und Drücke durch die entsprechenden Faktoren ersetzt. Die Regressionsmodelle wurden jeweils mit den 4- und 5-Faktorenmodellen der Heißpresse berechnet.

Die Schätzgenauigkeiten der Regressionsmodelle mit den Pressenfaktoren der PCA- bzw. PFA-Analyse zeigten im Vergleich zu den Modellen 9 und 10 keine wesentlichen Unterschiede. Einzig die Schätzgenauigkeit der Dickenquellung im Beobachtungs- und Vorhersageraum im technologischen Modell konnte durch die Berücksichtigung des gesamten spezifischen Druckprofils durch die extrahierten Faktoren verbessert werden. In den kombinierten Modellen zeigte sich keine signifikante Änderung in den Modellkennzahlen bei Verwendung der Faktoren der spezifischen Drücke. Wurden dagegen die Faktoren der Distanzen in die Modelle integriert, ergaben sich geringfügige Abnahmen in den Schätzgenauigkeiten gegenüber den Modellen mit den Faktoren der spezifischen Drücke. Der Grund hierfür ist in der höheren Schwankungsbreite der Druckvariablen gegenüber den Distanzvariablen zu sehen.

Ein Unterschied zwischen den Vorhersagegenauigkeiten bei der Integration der 4 bzw. 5 Faktoren des Pressenprogramms in die Regressionsmodelle war nicht festzustellen.

Die Integration der Heißpresse in Form von unkorrelierten Faktoren der PCA- bzw. PFA-Analyse in die multivariaten Regressionsmodelle führte nicht zu einer Verbesserung der Schätzgenauigkeit im Vorhersageraum. Die direkte Auswahl der Pressenvariablen in die Regressionsmodelle ist damit nicht mit einer Abnahme der Vorhersagegenauigkeit verbunden.

Alle 4 bzw. 5 Faktoren des Distanzen- und Druckprofils zeigten auf die Rohdichte einen signifikanten Einfluss. Bei Verwendung der Originalvariablen konnte dagegen nur bei einer Druckvariable aus der Kalibrierzone (Druck Rahmen 22) eine statistische Signifikanz nachgewiesen werden. Trotz dieser unterschiedlichen Gewichtung der Presse in den Modellen ist keine höhere Schätzgenauigkeit durch die Integration aller Pressenabschnitte durch die Faktoren in die Modelle erzielt worden. Die Ursache hierfür dürfte darin liegen, dass die Pressenabschnitte bzw. Pressenrahmen nicht unabhängig bzw. nur in den Toleranzen der einzelnen Heizplatten voneinander geändert werden. Je nach Kundenauftrag werden verschiedene Pressenprogramme genutzt, die bei gleicher Plattendicke nur

geringfügig unterschiedliche Distanzen bzw. Drücke über die Pressenrahmen aufweisen. Für die Vorhersagegenauigkeit der Rohdichte ist die Information der Distanz bzw. des Druckes eines Rahmens aus der Kalibrierzone ausreichend.

In Tabelle 53 sind die signifikanten Pressenfaktoren der PCA-Analyse in den Regressionsmodellen für die jeweiligen Platteneigenschaften mit Angabe des Vorzeichens der Regressionskoeffizienten zusammengefasst.

Tabelle 53: signifikanter Einfluss der Pressenfaktoren (PCA) auf die Platteneigenschafter
innerhalb der Regressionsmodelle mit Vorzeichen der Regressionskoeffizienten

	4 Faktoren		4 Faktoren		5 Faktoren		5 Faktoren	
	Distanz		spez. Druck		spez. Druck	spez. Druck		
IB	"Erwärmen"	(+)	"Erwärmen"	(-)	"Halten"	(-)	"Halten"	(+)
TS	"Schließen" "Erwärmen" " 2. Verdichtungsschritt" "Kalibrierung"	(-) (+) (+) (+)	"Erwärmen" "Kalibrierung"	(-) (-)	"Schließen" "Lüften" " 2. Verdichtungsschritt" "Kalibrierung"	(+) (-) (-) (-)	"Schließen" "Lüften" " 2. Verdichtungsschritt" "Kalibrierung"	(-) (+) (+) (+)
SS	" 2. Verdichtungsschritt"	(-)	" 2. Verdichtungsschritt"	(+)	"Kalibrierung"	(+)	" 2. Verdichtungsschritt"	(-)
MD	"Schließen" "Erwärmen" " 2. Verdichtungsschritt" "Kalibrierung"	(-) (-) (-)	"Schließen" "Erwärmen" " 2. Verdichtungsschritt" "Kalibrierung"	(+) (+) (+) (+)	"Schließen" "Lüften" "Halten" " 2. Verdichtungsschritt" "Kalibrierung"	(+) (+) (+) (+) (+)	"Schließen" "Lüften" "Halten" " 2. Verdichtungsschritt" "Kalibrierung"	(-) (-) (-) (-)

Insgesamt zeigten sich durch die Verwendung der PCA-Pressenfaktoren mehr Pressenvariablen in den Regressionsmodellen signifikant, als bei einzelhafter Variablenauswahl. Durch Ausschaltung der Korrelationen zwischen den Pressenfaktoren, kann der Einfluss der Pressenabschnitte auf die technologischen Eigenschaften sichtbar gemacht werden. Alle in Tabelle 53 aufgeführten Pressenfaktoren weisen in dem Vorhersagezeitraum des einjährigen Testdatensatzes auf dem Niveau von mindestens α =0,10 eine statistische Signifikanz auf. Diese Signifikanz trat insbesondere in den Gleichungen der Querzugfestigkeit, Dickenquellung und Abhebefestigkeit nie durchgängig auf.

Die Vorzeichen der Regressionskoeffizienten der Distanzfaktoren sind durchweg den Vorzeichen der Druckvariablen entgegengesetzt. Technologisch ist dieser Zusammenhang richtig, da mit zunehmender Distanz bei gleicher Mattendicke der Pressdruck abnimmt. Das physikalische Verhalten der Pressenvariablen wird damit durch die Modelle korrekt wiedergegeben.

Ähnlich der Auswahl der Einzelvariablen der Presse konnte nur bei einem Pressenfaktor in den Modellen zur Querzugfestigkeit und Abhebefestigkeit eine statistische Signifikanz nachgewiesen werden. Eine Erhöhung der Distanz bzw. eine Verringerung des Drucks im Abschnitt "Halten" bzw. "Erwärmen" war in allen Modellen mit einer höheren Querzugfestigkeit verbunden. Eine größere Distanz in diesem Bereich der Presse führt zu einer

besseren Entgasung der Matte und damit am Ausgang der Presse zu geringeren Rückstellkräften in der Matte. Die geringeren Rückstellkräfte führen zu einer geringeren Belastung der noch nicht vollständig ausgehärteten Klebverbindungen zwischen den Fasern und damit zu einer höheren Querzugfestigkeit. In der direkten Variablenauswahl für die Gleichung der Querzugfestigkeit wurde der Druck des Rahmens 9 ausgewählt.

Ein höherer Druck während des zweiten Verdichtungsschrittes bzw. in der Kalibrierzone steht mit einer höheren Abhebefestigkeit in Verbindung. In den Regressionsmodellen mit direkter Variablenauswahl wurde stellvertretend für die Kalibrierzone der Rahmen 26 ausgewählt.

Die Eigenschaften Dickenquellung und Rohdichte werden nach den Modellen wesentlich stärker durch Änderungen im Pressprogramm beeinflusst als die Querzug- und Abhebefestigkeit. Höhere Drücke bzw. geringere Distanzen in allen Pressenabschnitten sind mit höheren Rohdichten der Platten verbunden. Ob dieser Zusammenhang tatsächlich unabhängig in jedem Pressenabschnitt zutrifft, ist nicht nachweisbar. Technologisch ist der Zusammenhang – geringere Pressdistanzen bei gleichem Mattengewicht führen zu höheren Rohdichten – unstrittig. Von allen Pressenabschnitten zeigte die Kalibrierzone jedoch den stärksten Effekt auf die Rohdichte. In der einzelhaften Auswahl der Pressenvariablen wurde dies durch die Auswahl des Rahmens 22 berücksichtigt.

Eine höhere Dickenquellung ist dagegen mit einem höheren Druck in der Schließphase und mit geringeren Drücken in den Abschnitten der Entlüftung bzw. Erwärmung und im zweiten Verdichtungsschritt und Kalibrierzone verbunden. Bis auf den zweiten Verdichtungsschritt wurden diese signifikanten Abschnitte bei der einzelhaften Auswahl der Pressenvariablen berücksichtigt. Die Vorzeichen der Einzelvariablen stimmen mit denen der PCA-Faktoren überein.

Erkenntnisse

Die Zusammenfassung der korrelierten Pressenvariablen zu einzelnen Pressenabschnitten in Form von unkorrelierten Pressenfaktoren ermöglicht eine tiefergehende und sichere Interpretation des Einflusses der Heißpresse auf die technologischen Eigenschaften als die einzelne Betrachtung der Pressenvariablen in den Regressionsmodellen. Die Integration der PCA- bzw. PFA-Faktoren der Heißpresse in die Regressionsmodelle führte jedoch nicht zu einer Erhöhung der Schätzgenauigkeit im Vorhersageraum. Dieser Umstand liegt in der nur begrenzten unabhängigen Änderung der einzelnen Pressenabschnitte. Die Informationen der Einflüsse der einzelnen PCA- bzw. PFA-Pressenfaktoren auf die technologischen Platteneigenschaften können für die Auswahl einzelner direkter Pressenvariablen genutzt werden.

9.4 PLS-Analyse

9.4.1 Ziele der Untersuchungen

In den Regressionsmodellen kann nur der Einfluss nichtkorrelierter Prozessvariablen auf die technologischen Eigenschaften der Platten nachgewiesen werden. Die notwendige bzw.optimale Anzahl an Variablen in den Modellen ist u.a. von den Veränderungen der Prozesseinstellungen im Beobachtungs- bzw. Vorhersageraum abhängig. Mit einer Anzahl von 15 -30 Variablen je Gleichung erreichten die Regressionsmodelle maximale Schätzgenauigkeiten im langfristigen Vorhersageraum.

Um den Einfluss korrelierter Variablen innerhalb eines Modells auf die technologischen Eigenschaften untersuchen zu können, ist eine Reduktion der Datenmatrix zu latenten nicht Variablen notwendig. miteinander korrelierten Unter latenten Variablen (auch Linearkombinationen bzw. Scores genannt) werden übergeordnete nicht messbare Merkmale verstanden. Im Prozess der Holzwerkstoffherstellung kann unterstellt werden, dass dieser u.a. durch Qualitätsschwankungen der Ausgangsmaterialien, durch die Beleimungsgüte, die Reaktivität zwischen Klebstoff und Holzpartikel, eventuellen Verunreinigungen und durch das Pressprogramm beeinflusst wird. Diese Einflüsse sind jedoch nicht direkt als ein Merkmal messbar, sondern werden durch eine Vielzahl von Prozessvariablen bestimmt. Die Qualität des Fasermaterials wird z.B. wesentlich von der Holzart und durch den Koch- bzw. Refinerprozess bestimmt.

Welche Variablen zu den latenten Variablen zusammengefasst werden, kann durch die Höhe der Gewichtungen w beurteilt werden. Um die direkte Wirkung einer einzelnen Variable auf die technologischen Eigenschaften zu bemessen, lassen sich PLS-Regressionskoeffizienten für ein lineares Modell berechnen (Manne, 1987). Diese Modelle werden PLSR-Modelle genannt und sind in die Klasse der LVR-Modelle (Latente Variablen Regression) einzuordnen.

Es soll untersucht werden, ob durch die zusätzliche Berücksichtigung von korrelierten Variablen innerhalb der PLS-Modelle, die Schätzgenauigkeit im Vorhersageraum erhöht werden kann.

Die Validierung der PLS-Modelle erfolgt methodisch exakt nach dem Vorgehen aus der Regressionsanalyse. Es werden die gleichen Testdaten zur Validierung herangezogen, um die Vorhersagegenauigkeiten beider Modellarten miteinander vergleichen zu können.

9.4.2 Modellwahl (Einzel- oder Gesamtmodell)

Mit der PLSR-Analyse können mehrere Zieleigenschaften (Plattenmerkmale) gemeinsam (mulivariat) in einem Modell modelliert werden. Dies hat den Vorteil, den Einfluss von Prozessvariablen gemeinsam auf die Plattenmerkmale beschreiben zu können. Die Auswirkungen von Betriebsstörungen bzw. von abweichenden Prozesseinstellungen können in einem gemeinsamen Modell auf die technologischen Eigenschaften besser nachvollzogen werden. Wenn die Zieleigenschaften korreliert sind, sollten sie nach Wold et al. (2001) auch gemeinsam modelliert werden. Messen die Zieleigenschaften dagegen unterschiedliche

Dinge, tendiert ein gemeinsames Modell zu zu vielen Komponenten und damit zur Unüberschaubarkeit. In diesem Fall haben einzelne Modelle den Vorteil geringerer Dimensionen und damit der besseren Interpretationsmöglichkeit. Nach der Korrelationsmatrix der Labordaten (Tabelle 9) sind alle Laboreigenschaften auf dem 99% Niveau statistisch signifikant miteinander korreliert. Auf Basis der Korrelationsmatrix würde ein gemeinsames Modell adaquat sein. Wold et al. (2001) dagegen empfiehlt eine PCA-Analyse der Zieleigenschaften, um den praktischen Rang der zu modellierenden Eigenschaften zu erhalten. Danach ist ein gemeinsames Modell zu wählen, wenn die Anzahl der Komponenten klein im Vergleich zu der Anzahl der Eigenschaften ist. Sind dagegen Gruppen im PCA-Loading-Plot zu sehen, sollten diese Gruppen für sich modelliert werden. Tabelle 54 zeigt die Eigenwerte und die erklärte Varianz durch die Komponenten F1 bis F4. Danach erklärt die erste Komponente ca. 54%, die Zweite 20% und die Dritte ca. 16% der gesamten Varianz der Laboreigenschaften. Drei Faktoren sind notwendig um 90% der Varianz der vier Laboreigenschaften zu erklären. Die Anzahl der Komponenten im PCA-Modell mit 90% erklärter Varianz ist damit nur um eins kleiner als die Anzahl der zu modellierenden Eigenschaften.

	F1	F2	F3	F4
Eigenwerte	2,2	0,8	0,7	0,4
Varianz %	53,9	20,0	16,4	9,8
Summe %	53,9	73,9	90,2	100,0

Tabelle 54: PCA-Analyse, Eigenwerte und erklärte Varianz der 4 Laboreigenschaften

Abbildung 37 zeigt die Faktorladungen der ersten beiden Komponenten für die Laboreigenschaften eines 2-Komponenten PCA-Modells nach einer Varimax-Rotation. Die höchste Ähnlichkeit weisen dort die Eigenschaften Abhebefestigkeit und Querzugfestigkeit auf. Beide Eigenschaften werden mit Faktorladungen von 0,73 bzw. 0,93 ausreichend gut von der ersten Komponente beschrieben. Die Dickenquellung dagegen wird mit einer Faktorladung von -0,63 durch die zweite Komponente nur befriedigend beschrieben. Die Dickenquellung befindet sich auf Grund der negativen Korrelation zu den anderen Eigenschaften in der linken Hälfte der Grafik. Die Dichte wiederum wird durch die zweite Komponente mit einer Faktorladung von 0,91 ausreichend genau erklärt. Mit 2 Komponenten werden insgesamt 74% der Varianz der Laboreigenschaften beschrieben.

Zur genaueren Beschreibung der Dickenquellung ist eine dritte Komponente notwendig. Abbildung 38 zeigt die Faktorladungen des 3-Komponenten PCA-Modells. Es zeigt sich, dass die Querzugfestigkeit und die Abhebefestigkeit durch die erste, die Rohdichte durch die zweite und die Dickenquellung durch die dritte Komponente sehr gut beschrieben werden. Die Ergebnisse der PCA-Analyse legen nahe, die Dickenquellung und die Rohdichte in eigenen Modellen zu modellieren. Dieses Ergebnis spiegelt die physikalischen Messprinzipien der Eigenschaften wider. Ein gemeinsames Modell ist in der praktischen Anwendung jedoch nicht auszuschließen, da mit 2 PCA-Komponenten insgesamt 74% der Varianz beschrieben werden können. Die Ergebnisse aus den Regressionsmodellen zeigten, dass die Bestimmtheitsmaße kleiner als 70% sein müssen, um das Hintergrundrauschen des Prozesses nicht überproportional auf die ex-ante-Vorhersagen zu übertragen.



Abbildung 37: PCA Faktorladungen eines 2- Komponentenmodells für die Laboreigenschaften nach der Varimax-Rotation (Lerndatensatz)



Abbildung 38: PCA Faktorladungen eines 3- Komponentenmodells für die Laboreigenschaften nach der Varimax-Rotation (Lerndatensatz)

9.4.3 Anzahl der PLS-Faktoren

Die Anzahl der PLS-Faktoren (entspricht Anzahl latenter Variablen) bestimmt sowohl das Anpassungsverhalten des Modells an die Lerndaten und damit die Vorhersagekraft als auch die Komplexität des Modells. Mit sehr vielen und korrelierten Daten besteht die Gefahr eines "overfittings" des Modells. Diese Modelle besitzen eine geringe Vorhersagekraft. Es ist daher unbedingt notwendig die Vorhersagesignifikanz jeder Komponente zu testen. Die Methode der Cross-Validierung hat sich dabei zur Überprüfung der Vorhersagesignifikanz der PLS-Faktoren durchgesetzt (Wakeling und Morris, 1993). Abbildung 39 zeigt den Zusammenhang zwischen der Anzahl an PLS-Faktoren und der Vorhersagefähigkeit des Modells, dargestellt am Root Mean PRESS. Zusammengefasst in der PRESS-Statistik sind hierbei die Vorhersagen der vier Standardzieleigenschaften mittels der "Leave-one-out" Cross-Validierung. Die Grafik zeigt, dass anfangs mit zunehmender Anzahl an Komponenten die Vorhersagegenauigkeit des Modells steigt (Press-Statistik wird kleiner) und ca. ab der Komponente 10 wieder abfällt. Der Unterschied der PRESS-Statistik zwischen den Komponenten 3 bis 14 fällt dabei mit 0,90 bis 0,92 sehr gering aus. Der Test von van der Voet (1994) wählt dabei das Modell mit der geringsten Anzahl an Komponenten aus, deren Residuen insignifikant größer sind als das Modell mit der minimalen PRESS-Statistik. Nach der Hotteling's T² Statistik des van der Voet Tests wird das Modell mit 5 Komponenten gewählt. Nach Shao (1993) ist das Leave-one-out Verfahren asymptotisch inkonsistent und zu konservativ in dem Sinn, dass zu viele Komponenten gewählt werden.



Abbildung 39: Zusammenhang zwischen Anzahl der PLS-Faktoren und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS, mittels "Leave-one-out" Cross-Validierung (Modell mit 4 Zieleigenschaften, 281 Prozessvariablen)

Abbildung 40 zeigt die durch Block-Validierung bestimmte PRESS-Statistik, zur Darstellung der Vorhersagegenauigkeit, in Abhängigkeit der Anzahl an Komponenten. In der Block-Validierung werden Blöcke an Beobachtungen aus dem Lerndatensatz herausgenommen und vorhergesagt. Die PRESS-Statistiken werden am Ende der Validierung aufsummiert und über die Anzahl der Blöcke gemittelt. Die Anzahl der zusammenhängenden Beobachtungen innerhalb eines Blockes betragen in der Abbildung 40 n=5, 10 und 15. Die PRESS-Statistik fällt in der Block-Validierung allgemein höher aus als in der "Leave-one-out"-Validierung. Innerhalb der Block-Validierung zeigen sich, in Abhängigkeit der Blockgröße, die minimalen PRESS-Statistiken bei 2 bis 4 PLS-Faktoren. Damit weisen nach den Ergebnissen der Block-Validierung die multivariaten Modelle (alle 4 Zieleigenschaften) mit 2 bis 4 Komponenten die höchsten Vorhersagegenauigkeiten auf.

Da die Modelle anhand eines externen Testdatensatzes validiert werden, wird die optimale Komponentenanzahl zur maximalen Vorhersagegenauigkeit der Laboreigenschaften anhand dieses Datensatzes erneut überprüft. Mit diesem Vorgehen kann die Bestimmung der Komponentenanzahl aus der Cross-Validierung überprüft werden. Die weiteren Ergebnisse der Cross-Validierungen werden im weiteren Text mit der Darstellung der einzelnen PLS-Modelle aufgezeigt.



Abbildung 40: Zusammenhang zwischen Anzahl der PLS-Faktoren und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS, mittels Block-Cross-Validierung (Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)

Erkenntnisse

Die PCA-Analyse der Laboreigenschaften lässt erkennen, dass eine multivariate Betrachtung der Platteneigenschaften in einem PLS-Modell zu einer hohen Modelldimension führen kann. Dies ist immer dann zu erwarten, wenn die betrachteten Zieleigenschaften stark unterschiedliche Phänomene messen. Die 4 betrachteten Laboreigenschaften Querzugfestigkeit, Abhebefestigkeit, Dickenquellung und mittlere Rohdichte können mit zwei PCA-Komponenten und einer erklärten Varianz von 74% und mit 3 PCA-Komponenten und einer erklärten Varianz von 90% zusammengefasst werden. Eine starke Reduzierung der Modellkomponenten im Vergleich zu der betrachteten Anzahl an Ausgangsvariablen spricht für eine Betrachtung der Eigenschaften in einem gemeinsamen PLS-Modell. Im vorliegenden Fall kann die Querzugfestigkeit und die Abhebefestigkeit sehr gut und die Dickenquellung und die mittlere Rohdichte ausreichend durch eine übergeordnete (latente) Variable zusammengefasst werden.

Es kann aus den Voranalysen nicht eindeutig abgeleitet werden, ob eine effiziente Schätzung der 4 Laboreigenschaften in einem gemeinsamen PLS-Modell möglich ist. Als Konsequenz daraus, werden nachfolgend sowohl multivariate PLS-Modelle als auch Einzelmodelle für jede Eigenschaft berechnet und deren Schätzgenauigkeiten im Vorhersageraum verglichen.

In Abhängigkeit des Validierungsverfahren erreichen die multivariaten PLS-Modelle mit 2 bis 5 Komponenten die höchsten Schätzgenauigkeiten, insofern alle Prozessvariablen herangezogen werden. Sowohl eine Unterschreitung als auch eine Überschreitung dieser Komponentenanzahl führt zu deutlichen Verlusten in der Vorhersagegenauigkeit.

9.4.4 Diagnose und Interpretation

Dieses Kapitel stellt im folgenden die wichtigsten Zusammenhänge, Begriffe und Interpretationen der PLS-Analyse vor. Dazu werden insbesondere die grafischen Darstellungen dieser Methode genutzt, um in die PLS-Analyse einzuführen.

Die Beziehungen zwischen den Beobachtungen im X-Raum (Prozessvariablen) sind am anschaulichsten durch einen X-Score-Plot darstellen. Die X-Scores stellen die neuen und zusammenfassenden Variablen des Prozesses dar. Anhand dieser Darstellung lassen sich Ähnlichkeiten bzw. Unähnlichkeiten in den Beobachtungen aufzeigen. Abbildung 41 zeigt die ersten beiden X-Scores t1 und t2 eines gemeinsamen multivariaten Modells für alle Laboreigenschaften.



Abbildung 41: PLS Scores t(1) und t(2) eines gemeinsamen multivariaten PLS-Modells (Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)

Jeder Punkt in der Abbildung stellt eine Beobachtung (Datensatz) als Zusammenfassung des Prozessdatenraums dar. Die ersten beiden Komponenten (Scores) t1 und t2 erklären ca. 38% der Variation des gesamten x-Raumes mit insgesamt 281 Prozessvariablen. Die Höhe an erklärter Varianz im x-Raum ist erheblich von der Variablenauswahl und der Anzahl an Komponenten abhängig. Die Nummerierung stellt die fortlaufende Aufzeichnung der Datensätze dar. Es ist in der Abbildung 41 zu sehen, dass der Prozess stabil um einen bekannten Betriebspunkt (0,0) geregelt wird. Die meisten Beobachtungen befinden sich innerhalb des Bereiches von –10 bis +10 im Score t1 und t2.

Beobachtungen, die weit außerhalb des Nullpunktes im X-Score-Diagramm liegen, können außergewöhnliche Betriebspunkte im Prozessdatenraum darstellen. Beobachtung 81 (rechts außen) zeigt im Score 1 geringe Pressdrücke in der Heißpresse an. Zusätzlich weist dieser Datensatz die geringsten Beleimungsmengen und Pressengeschwindigkeit im ganzen Lerndatensatz auf. Die Beobachtung 16 (links außen) dagegen weist ungewöhnlich hohe Drücke in der Heißpresse auf.

Des weiteren ist in der Abbildung 41 zu sehen, dass sich die älteren Beobachtungen im oberen (Score t(2) positiv) und die jüngeren Beobachtungen im unteren (Score t(2) negativ) Bereich der Grafik befinden. Dies spiegelt eine Drift des Prozesses im X-Raum wider. Diese Verschiebung wird insbesondere durch eine Drift in den gemessenen Temperaturen verursacht. Auch die Änderungen im Stromverbrauch des Bandantriebes der Vorpresse und der Beschickschnecke als auch in der zugegebenen Wassermenge werden durch die zweite PLS-Komponente erfasst und tragen zur Verschiebung der Punktwolke bei. Die erste PLS-Komponente erfasst dagegen fast hauptsächlich die Variablen der Heißpresse, bei denen keine gravierenden Änderungen über die Zeit zu sehen sind.

Die Beziehungen zwischen den Beobachtungen im Y-Raum (Labordaten) können durch einen Y-Score-Plot dargestellt werden. Jeder Punkt stellt eine Beobachtung als Zusammenfassung der Platteneigenschaften dar. Je weiter die Punkte voneinander entfernt sind, desto unähnlicher sind sich die Platteneigenschaftsprofile. Abbildung 42 zeigt die ersten beiden Y-Scores u(1) und u(2) eines multivariaten Modells für alle 4 Platteneigenschaften.



Abbildung 42: PLS Scores u(1) und u(2) eines gemeinsamen multivariaten PLS-Modells (Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)

Die Interpretation des Y-Score-Plots ist gegenüber dem X-Score-Plot schwieriger, da die u-Vektoren des Y-Prozesses nicht orthogonal aufeinander stehen. In Abbildung 42 ist jedoch sehr gut zu sehen, dass die Eigenschaftsprofile der Platten der einzelnen Beobachtungen in der Mehrheit um den Nullpunkt schwanken. Je weiter die Datenpunkte vom Nullpunkt abweichen, desto mehr weicht das von den Score-Vektoren u(1) und u(2) beschriebene Platteneigenschaftsprofil vom Mittelwert ab. So weist der Datensatz 107 (rechts außen) z.B. die minimalste und der Datensatz 74 (links außen) die maximalste Rohdichte im Lerndatensatz auf. Beide Datenpunkte liegen auf dem u(1)-Score am weitesten voneinander entfernt. Weiter vom Nullpunkt entfernte Datenpunkte weisen eine Besonderheit bzw. überdurchschnittliche Abweichung vom Mittelwert in den Labordaten auf. Der Datensatz 25 mit einem maximalen u(2)-Score weist die maximale Querzugfestigkeit im Lerndatensatz auf. Zur vollständigeren Beschreibung des Eigenschaftsprofils sind weitere Komponenten heranzuziehen. Die beiden ersten Komponenten erklären 22% der Variation des Y-Raumes. Der erklärte Anteil der Varianz des Y-Raums durch die ersten beiden Komponenten liegt bei gleicher Variablenanzahl bei den Einzelmodellen höher. In Tabelle 55 ist die erklärte aufsummierte Varianz für das Gesamtmodell mit den 4 Laboreigenschaften als auch für die jeweiligen Einzelmodelle in Abhängigkeit von den PLS-Faktoren dargestellt. Für alle Modelle wurden zunächst alle Prozessvariablen (nach der Plausibilitätsanalyse) verwendet.

Komponente	Gesamt-Modell		IB-Modell		TS-Modell		SS-Modell		MD-Modell	
	R _X ^²	R_y^2	R _X ^²	Ry ²	R _X ²	Ry ²	R _X ²	Ry ^²	R _X ²	Ry ²
1	27,1	7,7	27,6	17,6	18,6	22,8	14,9	20,0	26,4	21,4
2	38,2	21,7	33,7	45,4	34,3	42,1	38,1	26,8	38,2	36,7
3	42,8	36,9	39,7	57,4	42,0	58,1	42,8	46,4	42,6	57,2
4	47,0	43,6	47,6	62,3	46,3	70,0	47,0	55,3	46,8	66,1
5	52,4	48,9	50,3	69,8	49,2	75,9	50,1	62,6	49,9	73,0
6	54,7	53,8	53,6	73,9	53,7	78,6	53,7	67,1	53,3	77,4

Tabelle 55: Anteil der erklärten aufsummierten Varianz in % im X- und Y-Raum durch die ersten 6 PLS-Faktoren für ein Gesamt- und den 4 Einzelmodellen

Unabhängig vom Modell wird der X-Raum in Abhängigkeit der PLS-Faktoren annährend gleich modelliert. Mit 6 PLS-Faktoren werden zwischen 53 und 54% der Varianz der 281 Prozessvariablen beschrieben. Dieser Wert stellt den Mittelwert der erklärten Varianzen für alle 281 Variablen dar. Die Einzelwerte schwanken dabei sehr deutlich zwischen 2% und 90%. Prozessvariablen mit hohen merkmalseigenen Varianzen weisen keine deutlich ausgeprägte Korrelationen zu anderen Prozessvariablen auf. Diese können nicht ausreichend durch die latenten Variablen erfasst werden, da diese nur stark miteinander korrelierte Variablen zusammenfassen, die in Beziehung zu Y stehen. Prozessvariablen mit hohen merkmalseigenen Varianzen bilden u.a. die Silo- bzw. Bunkerfüllstände, verschiedene Ventilstellungen, die Dickenmessungen und die Luftfeuchte. Die Druckvariablen des Vorverdichters und der Vorpresse weisen ebenfalls sehr geringe Werte in den erklärten Varianzen auf.

Im Y-Raum liegen die erklärten Varianzen der Einzelmodelle bei 6 PLS-Faktoren mit ca. 67% - 77% höher als beim Gesamtmodell mit annährend 54%. Bei den Einzelmodellen wird die Abhebefestigkeit bei gleicher Komponentenanzahl am ungenauesten und ab der dritten Komponente die Dickenquellung im Lerndatenraum am genauesten erklärt. Dieses Ergebnis stimmt mit den Bestimmtheitsmaßen der multivariaten Regressionsmodelle in Abhängigkeit der Variablenanzahl überein.

Aus den bisherigen Grafiken lassen sich die Abweichungen vom Mittelwert in den jeweiligen X- bzw. Y-Räumen feststellen. Erst wenn beide Räume in Beziehung zueinander gesetzt werden, kann nachvollzogen werden, ob Abweichungen im X-Raum zu entsprechenden Folgen im Y-Raum geführt haben. Die Beziehungen zwischen dem X- und dem Y-Raum lassen sich mit dem Score-Plot u (y) gegen t (x) aufzeigen. Nichtlineare Zusammenhänge können in diesem Plot aufgedeckt werden. Abbildung 43 zeigt jeweils die ersten X- und Y-Scores t1 und u1 und Abbildung 44 die Beziehung zwischen X und Y in der zweiten Dimension des multivariaten Modells für alle Laboreigenschaften. Mit zunehmender Modelldimension nehmen i.d.R. die linearen Zusammenhänge zwischen dem x- und y-Raum

ab. So ist der lineare Zusammenhang der ersten Dimension in Abbildung 43 straffer als in der zweiten Dimension in Abbildung 44.



Abbildung 43: PLS Scores t(1) und u(1) eines multivariaten PLS-Modells (4 Zieleigenschaften mit 218 Prozessvariablen)



Abbildung 44: PLS Scores t(2) und u(2) eines multivariaten PLS-Modells (4 Zieleigenschaften mit 218 Prozessvariablen)

Ein linearer Zusammenhang zwischen dem X- und Y-Raum lässt sich in Abbildung 43 gut erkennen. Die extremen Beobachtungen 16 (unten links) und 107 (oben rechts) stellen Änderungen im Pressenprogramm dar, die entsprechende Folgen für die Rohdichte aufwiesen. So weist die Beobachtung 107 bei durchschnittlich geringeren Pressdrücken (höhere Pressdistanzen) die geringste Rohdichte im Lerndatensatz und Beobachtung 16 mit hohen Pressdrücken eine sehr hohe Rohdichte auf.

Die Beobachtung 81 zeigt nach dem Modell in der ersten Dimension einen zu geringen y-Score auf. In den gemessenen Platteneigenschaften sind dagegen keine ungewöhnlichen Abweichungen zu sehen. (Der Wert der Dickenquellung ist in diesem Datensatz jedoch nicht vorhanden, so dass nicht festgestellt werden kann, ob diese durch die Prozessstörung beeinflusst worden ist.) Es kann unterstellt werden, dass mit der geringeren Pressengeschwindigkeit auf die Störung in der Beleimung reagiert worden ist und somit die Plattenanforderungen weitgehend eingehalten werden konnten. Die Dickenquellung kann auf diese Konstellation jedoch empfindlicher reagiert haben und aufgrund fehlender Informationen aus dem Prozess im Labor als "Fehlmessung" aussortiert worden sein. In der zweiten Dimension weist der Datensatz 81 keine weiteren Unregelmäßigkeiten auf.

Um festzustellen, inwieweit eine Beobachtung im x- bzw. y-Raum einen Ausreißer darstellt, sind die entsprechenden Residuen heranzuziehen. Diese spielen ein bedeutende Rolle in der Beurteilung eines Modells. Für die y-Residuen eignet sich die Darstellung des normal probability plots. In dieser Darstellung lassen sich Ausreißer als Punkte wiederfinden, die stark von der "Normal-Linie" abweichen. Als starke Ausreißer bezeichnet Wold et al. (2001) Werte, die ober- bzw. unterhalb der 4-fachen Standardabweichung liegen. Bei entsprechender Standardisierung der Residuen auf der x-Achse (Residuen dividiert durch Standardabweichung der Residuen) können Ausreißer sofort erkannt werden. Abbildung 45 zeigt für alle 4 Platteneigenschaften die Residuen des multivariaten PLS-Modells mit 5 Komponenten in den jeweiligen normal plots.



Abbildung 45: Residuen des multivariaten PLS-Modells mit 5 Komponenten als normal plot (4 Zieleigenschaften und 281 Prozessvariablen)

Die Mehrheit der Residuen liegen innerhalb der zweifachen Standardabweichung. Ein Wert im Modell der Rohdichte übersteigt die 4-fache Standardabweichung geringfügig. Die im Labor gemessene Rohdichte weist an dieser Stelle einen stark unterdurchschnittlichen Wert auf. Auch die Regressionsmodelle konnten diesen Wert nicht modellieren und wiesen an dieser Stelle ebenfalls maximale Residuen im Lerndatensatz auf.

Für die X-Residuen ist die Darstellung der normal plots nicht geeignet, da zu viele Residuen existieren (N*K). Im vorliegenden Fall müssten 281 normal plots auf Ausreißer überprüft werden. Eine zusammenfassende Darstellung für jede Beobachtung liefert die Standardabweichung der X-Residuen (RSD) der entsprechenden Zeile in der Residuen-Matrix E. Da diese Standardabweichung proportional zur euklidischen Distanz der Datenpunkte und der Modellebene im X-Raum ist, wird diese in der Fachliteratur auch oft DModX genannt (Distance to the Model in X-space). Nach Wold et al. (2001) weisen DModX-Werte größer als 2,5 mal der durchschnittlichen Standardabweichung aller Residuen auf einen Ausreißer für diese Beobachtung hin. Die Darstellung der Standardabweichung der Residuen ist ebenso für die Zieleigenschaften (Y-Raum) geeignet. Abbildung 46 und Abbildung 47 zeigen die entsprechenden Grafiken für den X- und für den Y-Raum. Die durchschnittlichen Standardabweichung der X-Residuen beträgt 11,4. Im X-Raum ist somit kein außergewöhnlicher Ausreißer aufzufinden. In der Abbildung 46 findet sich im Y-Raum an der Beobachtung 26 der Ausreißer in der Rohdichte wieder.



Abbildung 46: DModX für jede Beobachtung des Lerndatensatzes eines gemeinsamen 5-Komponenten PLS-Modells (4 Zieleigenschaften, 281 Prozessvariablen)



Abbildung 47: DModY für jede Beobachtung des Lerndatensatzes eines gemeinsamen 5-Komponenten PLS-Modells (4 Zieleigenschaften, 281 Prozessvariablen)

Die Darstellung der DMod-Grafiken ermöglicht eine sehr schnelle Überprüfung der Modelle.

Zur Interpretation der PLSR-Modelle bzw. deren latenten Strukturen wird standardmäßig der Gewichts-Plot herangezogen. Die X-Werte mit den größten Gewichten auf der jeweiligen Achsen-Komponente werden durch die jeweilige Komponente am besten repräsentiert. X-Werte mit geringen absoluten Gewichten sind für das Modell weniger bedeutend als die mit hohen Gewichten. Variablen mit Gewichten nahe Null in allen Komponenten können zur Verbesserung der Vorhersagefähigkeit des Modells entfernt werden. Abbildung 48 zeigt die ersten beiden PLS-Gewichte eines 3-Komponenten Modells für die Dickenquellung. Grundlage des Modells ist der Lerndatensatz mit 132 Beobachtungen und 60 Variablen. In der Abbildung 48 sind zur besseren Übersichtlichkeit nur eine Auswahl an Variablen dargestellt. Die vollständige Liste ist im Anhang 5 angegeben. Dort finden sich auch die Gewichte der dritten Komponente sowie weitere Kennzahlen des PLS-Modells für die Dickenquellung.

Die Reduzierung von 281 auf 60 Variablen erfolgte anhand der Höhe der standardisierten Regressionskoeffizienten und der VIP-Werte. Nähere Erläuterungen dazu befinden sich im Text weiter unten.



Abbildung 48: PLS-Gewichte für die ersten beiden PLS-Faktoren eines Modells für die Dickenquellung (60 Prozessvariablen)

Abbildung 48 zeigt eine Momentaufnahme, im weiteren Verlauf der Vorhersagen ändern sich die Gewichte entsprechend den veränderten Prozesseinstellungen. Die Variablen der Beleimung (Leim, Faser, Wasser, Härter) werden gut durch die erste Komponente repräsentiert. Dabei hat die Härtermenge den entgegengesetzten Einfluss auf die Dickenquellung als die restlichen Variablen der Beleimung. Die Harnstoffmenge dagegen weist ein sehr geringes positives Gewicht auf. Im Lerndatenraum hat die Harnstoffmenge keine signifikante Bedeutung auf die Dickenquellung. Bei 60 Variablen erklärt das Modell mit 3 Komponenten 25% der Varianz im X-Raum und 72% im y-Raum. Bei Verwendung aller Prozessvariablen dagegen wurden mit 3 Komponenten 42% der Varianz des x-Raums und nur 58% des y-Raums erklärt (Tabelle 55). Durch die Reduzierung der Variablenanzahl in der PLS-Analyse kann die Vorhersagegenauigkeit im y-Raum eines Modells verbessert werden. Die erklärten Varianzen im x-Raum nehmen dagegen ab, da zunehmend korrelierte Variablen mit geringer Bedeutung für die Zielvariable aus der Analyse herausgenommen werden. Je weniger die übrigen Variablen jedoch miteinander korreliert sind, desto weniger lassen sich deren Schwankungen durch wenige Komponenten erklären. Die erklärte Varianz der einzelnen Prozessvariablen findet sich im Anhang 5. Im Fall der Dickenquellung können mit 25% erklärter Varianz im x-Raum knapp ³/₄ der Varianz der Dickenquellung erklärt werden. Neben ein hohes "Rauschen" im x-Raum, ist dieser Umstand damit zu erklären, dass nur wenige Prozessvariablen einen tatsächlich bedeutenden Einfluss auf die Dickenquellung ausüben.

Bei Modellen mit mehreren Komponenten kann nicht ohne weiteres aus dem Gewichtsplot der ersten beiden Dimensionen auf den Einfluss der einzelnen Variablen auf die Zieleigenschaften geschlossen werden. Dazu ist die Berechnung der PLS-Regressionskoeffizienten notwendig. Um so größer diese sind, desto bedeutender ist die Variable für die Zieleigenschaft. Die Regressionskoeffizienten fassen den Einfluss einer Variablen über alle Komponenten hinweg auf die Zieleigenschaft zusammen. Eine weitere Statistik, die die Bedeutung der Variable für das gesamte Modell angibt, ist die Kennziffer VIP (Variable Importance for Projection) von Wold (1994). Im Gegensatz zum Gewicht w bzw. dem Regressionskoeffizienten, die die Bedeutung der Prozessvariablen zur Erklärung der Zieleigenschaften angeben, repräsentiert die VIP-Kennziffer einen Wert für jede Variable zur Anpassung des Modells sowohl in den Zieleigenschaften als auch in den Prozessvariablen. Die Kennziffer VIP berechnet sich als eine gewichtete Summe der Quadrate der PLS-Gewichte w^{*}_{ak} mit den erklärten Anteilen der Varianz von Y durch jede einzelne Komponente.

Hat eine Variable sowohl einen kleinen Regressionskoeffizienten als auch einen kleinen VIP-Wert, dann kann diese Variable zur Erhöhung der Vorhersagefähigkeit des Modells aus dem Modell herausgenommen werden. Nach Wold (1994) sind VIP-Werte kleiner 0,8 als "klein" einzustufen.

Abbildung 49 zeigt die standardisierten Regressionskoeffizienten und die VIP's für das PLS-Dickenquellungs-Modell mit 3 Komponenten. Dargestellt ist zur besseren Übersicht erneut nur eine Auswahl an Variablen mit VIP-Werten größer 0,8 (siehe Anhang 5 für die vollständige Liste). Die bedeutendsten Variablen des Modells der Dickenquellung sind im Lerndatenraum demnach der Leimlieferant (Remark), die Beleimungsdaten (ohne Paraffin und Harnstoff), die mit den Beleimungsmengen korrelierte Drehzahl der Austragsschnecke im Refiner und der Stromverbrauch des unteren Vorpressenbandes. Aus dem Vorzeichen der PLS-Regressionskoeffizienten kann die Wirkungsrichtung der Variablen abgelesen werden.



Abbildung 49: VIP und standardisierte Regressionskoeffizienten eines 3 Komponenten Modells für die Dickenquellung (Lerndatensatz, 60 Prozessvariablen)

Eine Erhöhung der Leim-, Faser-, bzw. Wassermenge führt innerhalb dieses Modells zu einer Verringerung und eine Erhöhung der Härtermenge dagegen zu einer Erhöhung der Dickenquellung. Eine Erhöhung der Drücke der Multipots am Ende der Presse (MKDR 21, 23 und 25) ist mit einer geringeren Dickenquellung verbunden. Diese Informationen sind insoweit besonders wertvoll, da diese Zusammenhänge in der Regressionsanalyse nicht aufgedeckt werden können. Bei Verwendung der drei Multipots als Regressoren im Regressionsmodell ergaben sich aufgrund der hohen Korrelationen untereinander sowohl positive als auch negative Vorzeichen in diesen Variablen.

Es bleibt zu beachten, dass die Abbildung 49 die Zusammenhänge aus dem Lerndatenraum widerspiegelt. Sowohl die Höhe der VIP-Kennwerte als auch die der standardisierten Regressionskoeffizienten ändern sich im Vorhersageraum entsprechend den neuen Prozesseinstellungen.

Erkenntnisse

Durch die Zusammenfassung der einzelnen Variablen zur latenten Variablen (Scores) lassen sich in der multivariaten PLS-Analyse Phänomene aufdecken, die durch klassische Verfahren sehr schwer aufzuzeigen sind. Der gesamte x-Raum kann durch ein Score-Plot der ersten beiden Komponenten überwacht werden. Sobald eine oder mehrere Variablen außerhalb der üblichen Betriebspunkte liegen, fällt diese Beobachtung durch einen hohen Score-Wert auf. In der univariaten Prozesskontrolle müsste jede einzelne Variable durch ein eigenes Diagramm überwacht werden.

Welche Variablen zu den einzelnen Komponenten zusammengefasst werden lässt sich durch die Höhe der Gewichte erkennen. Je höher das Gewicht einer Variablen, desto besser wird diese durch die Komponente erfasst. Diese Komprimierung korrelierter Variablen zu latenten Variablen ermöglicht es, den Einfluss aller Variablen auf die Platteneigenschaften nachzuweisen. Hierin ist der größte Vorteil der PLS-Analyse gegenüber der Regressionsanalyse zu sehen.

Die Berechnung der PLS-Regressionskoeffizienten erlaubt es, die absolute Bedeutung der Variablen auf die Zieleigenschaften aufzuzeigen.

9.4.5 Multivariate PLS-Modelle

Im Gegensatz zur multivariaten Regressionsanalyse, in der jede Zieleigenschaft mit individuellen Prozessvariablen beschrieben werden kann, können in den PLS-Modellen mehrere Zieleigenschaften nur mit einem gemeinsamen "X-Block" an Prozessvariablen modelliert werden. Dies ist dann in einem multivariaten PLS-Modell von Nachteil, wenn die Zieleigenschaften stark unterschiedliche Eigenschaftsmerkmale darstellen.

Um zu prüfen, inwiefern sich die 4 Platteneigenschaften (Querzugfestigkeit, Dickenquellung, Abhebefestigkeit und Rohdichte) durch ein gemeinsames PLS-Modell beschreiben und vorhersagen lassen, wurden alle 281 (plausiblen) Prozessvariablen herangezogen, um die Kennzahlen des Modells im Beobachtungs- und Vorhersageraum zu beschreiben. Um den Einfluss der Anzahl an Komponenten auf die Schätzgenauigkeit im Vorhersageraum zu bestimmen, wurden die Modelle mit jeweils 1 bis 12 Komponenten mit dem Testdatensatz validiert. Nach den Ergebnissen der Cross-Validierung (siehe Kapitel 9.4.3) ist in Abhängigkeit der Methode die maximale Vorhersagegenauigkeit zwischen 2 und 5 Komponenten zu erwarten.

Abbildung 50 zeigt den Einfluss der Komponentenanzahl auf die Bestimmtheitsmaße der einzelnen Platteneigenschaften im Lerndatenraum. Mit zunehmender Komponentenanzahl steigt erwartungsgemäß der Anteil der erklärten Varianz der Platteneigenschaften durch das PLS-Modell.





Im PLS-Modell kann wie in den Regressionsmodellen die Abhebefestigkeit im Lerndatenraum am ungenauesten beschrieben werden. Um die Schätzgenauigkeiten zwischen den Platteneigenschaften zu vergleichen, wurden die Vorhersagefehler (RMSE und RMSE_P) erneut auf die entsprechenden Standardabweichungen aus den Beobachtungsraum und Vorhersageraum (3 Monate) bezogen. In Abbildung 51 sind die Vorhersagegenauigkeiten des gemeinsamen Modells für die entsprechenden Datenräume

dargestellt. Der Fehler im Lerndatenraum wird durch eine zunehmende Komponentenanzahl verringert. Erst durch einen externen, nicht zur Modellierung verwendeten Testdatensatz, kann der Einfluss der Komponenten auf die Vorhersagegenauigkeit des Modells beurteilt werden. Danach können die Eigenschaften Dickenquellung und Querzugfestigkeit über den gesamten Bereich am genauesten vorhergesagt werden. Die maximale Schätzgenauigkeit wird bei 8 bzw. 9 Komponenten erreicht. Im Gegensatz dazu nimmt die Schätzgenauigkeit der Abhebefestigkeit und der Rohdichte mit der 5. PLS-Komponente ab. Insgesamt sind die Schätzgenauigkeiten dieser beiden Eigenschaften geringer als die der Querzugfestigkeit und Dickenquellung. Die gleichen Trends wurden durch die Regressionsmodelle aufgezeigt.



Abbildung 51: Einfluss der Anzahl der PLS-Faktoren auf die genormte Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum (281 Prozessvariablen)

Im Durchschnitt über alle vorherzusagenden Eigenschaften weist das 9-Komponentenmodell die höchste Schätzgenauigkeit auf (Abbildung 52). Dieser Durchschnittswert wird jedoch in erster Linie durch die Dickenquellung und Querzugfestigkeit bestimmt. Die Vorhersagegenauigkeit der Rohdichte ereicht bei 9 Komponenten gerade noch die geforderte Grenze.



Abbildung 52: Einfluss der PLS-Faktoren auf die durchschnittliche normierte Vorhersage genauigkeit im Beobachtungs- und Vorhersageraum (281 Prozessvariablen)

Da die Platteneigenschaften ein unterschiedliches Anpassungsniveau im Lerndatenraum besitzen, um die maximale Schätzgenauigkeit der einzelnen Eigenschaften im Vorhersage-

raum zu erreichen, tendiert das gemeinsame Modell mit den 281 Prozessvariablen zu einer Überanpassung der Rohdichte und der Abhebefestigkeit. Beide Eigenschaften dürfen im Lerndatenraum nicht so hoch angepasst werden, wie es bei der Dickenquellung und der Querzugfestigkeit möglich ist. Die Ursachen dafür liegen jedoch in verschiedenen Mechanismen. Die Rohdichte weist im Lerndatenraum eine sehr geringe Schwankungsbreite auf. Um so stärker die Rohdichte mit zunehmender Faktoren- bzw. Variablenanzahl angepasst wird, desto stärker und schneller übertragen sich die Fehler aus den Prozessdaten in die Zieleigenschaft. Bei der Abhebefestigkeit dagegen ist der Fehler in der Prüfung als auch die Schwankungen über die Plattenbreite wesentlich höher als bei den übrigen Eigenschaften. Hier führt eine erhöhte Komponentenanzahl und damit eine vermehrte Berücksichtigung an Prozessvariablen zu einer zunehmenden "Modellierung der Fehler" in der Abhebefestigkeit.

Einen Kompromiss aus dem unterschiedlichen Anpassungsniveau der Platteneigenschaften stellt das gemeinsame Modell mit 5 PLS-Faktoren dar. In der rechten Grafik der Abbildung 52 zeigt sich dort für alle Eigenschaften eine ausreichende Schätzgenauigkeit bezogen auf die Standardabweichung der Testdaten. Tabelle 56 und Tabelle 57 fassen detailliert die Kennzahlen des multivariaten Modells im Beobachtungs- und Vorhersageraum mit 5 bzw. 9 PLS-Faktoren zusammen.

Tabelle 56: Gütekennzahlen	gemeinsames Mode	ell mit 5 PLS-Fal	ctoren (281 Prozess-
variablen)			

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
R-Square Rx ²	0,52	0,52	0,52	0,52
R-Square Ry ²	0,53	0,55	0,36	0,47
PLS Faktoren	5	5	5	5
MAD	0,078	0,517	0,148	0,006
MAPE	4,13	6,04	6,53	0,63
RMSE (n)	0,097	0,631	0,187	0,0071
RMS Pct Error	5,17	7,56	8,34	0,80
Korrel_Prog	0,54	0,56	0,45	0,43
ME_Prog	-0,045	0,290	0,026	0,001
MAD_Prog	0,095	0,615	0,153	0,008
MAPE_Prog	5,11	6,60	6,37	0,87
RMSE_Prog	0,115	0,743	0,180	0,0085
RMS Pct Error_Prog	6,27	7,78	7,46	0,95

Tabelle 57: Gütekennzahlen gemeinsames Modell mit 9 PLS-Faktoren (281 Prozessvariablen)

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
R-Square Rx ²	0,61	0,61	0,61	0,61
R-Square Ry ²	0,61	0,71	0,55	0,64
PLS Faktoren	9	9	9	9
MAD	0,072	0,415	0,119	0,005
MAPE	3,81	4,80	5,24	0,54
RMSE (n)	0,089	0,506	0,156	0,0060
RMS Pct Error	4,73	5,95	6,93	0,67
Korrel_Prog	0,60	0,56	0,41	0,39
ME_Prog	-0,008	-0,066	-0,002	-0,002
MAD_Prog	0,081	0,571	0,155	0,008
MAPE_Prog	4,27	6,36	6,62	0,87
RMSE_Prog	0,103	0,678	0,186	0,0091
RMS Pct Error_Prog	5,45	7,75	8,21	1,02

Eine Reduzierung der Prozessvariablen, durch Entfernung aller Variablem mit VIP-Kennziffern kleiner als 0,8, führt zu einem Modell mit einer geringeren Anpassung im Beobachtungsraum und einer durchschnittlich höheren Schätzgenauigkeit im Vorhersageraum. Abbildung 53 stellt den Verlauf der Schätzgenauigkeit in Abhängigkeit der PLS-Faktoren in beiden Datenräumen für jede einzelne Platteneigenschaft und Abbildung 54 für alle Platteneigenschaften gemeinsam dar.



Abbildung 53: Einfluss der Anzahl der PLS-Faktoren auf die genormte Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum (200 Prozessvariablen)



Abbildung 54: Einfluss der PLS-Faktoren auf die durchschnittliche normierte Vorhersage genauigkeit im Beobachtungs- und Vorhersageraum (200 Prozessvariablen)

Das PLS-Modell mit 6 Faktoren bei reduzierter Variablenanzahl weist im Durchschnitt für alle Platteneigenschaften die maximale Schätzgenauigkeit im Vorhersageraum auf. Die Querzugfestigkeit erreicht jedoch erst bei 9 PLS-Faktoren die höchste Schätzgenauigkeit. Tabelle 58 zeigt die Kennzahlen des PLS-Modells mit 6 Komponenten im Detail.

Tabelle 58: Gütekennzahlen gemeinsames Modell mit 6 PLS-Faktoren (200 Prozessvariablen)

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
R-Square Rx ²	0,63	0,63	0,63	0,63
R-Square Ry ²	0,58	0,59	0,44	0,45
PLS Faktoren	6	6	6	6
MAD	0,075	0,496	0,135	0,006
MAPE	3,96	5,79	5,94	0,63
RMSE (n)	0,092	0,601	0,174	0,0072
RMS Pct Error	4,87	7,19	7,69	0,81
Korrel_Prog	0,57	0,57	0,48	0,39
ME Prog	-0,042	0,132	0,030	0,001
MAD Prog	0,091	0,578	0,151	0,008
MAPE Prog	4,88	6,26	6,29	0,85
RMSE Prog	0,112	0,701	0,178	0,0086
RMS Pct Error_Prog	6,11	7,48	7,47	0,97

9.4.6 PLS-Einzelmodelle

Um zu prüfen, ob durch eine individuelle Modellierung der Eigenschaften mit geringerer Komponentenzahl eine Erhöhung der Schätzgenauigkeit zu erreichen ist, werden im folgenden die Ergebnisse der Einzelmodelle dargestellt. Methodisch wurden im ersten Schritt jeweils alle Prozessvariablen zur PLS-Analyse herangezogen. Anschließend wurde eine Variablenreduktion vorgenommen, in der für die Modelle unbedeutende Variablen aus der Analyse herausgenommen worden sind. Die Bedeutung der Variablen für die Modelle wurden dabei durch die VIP-Kennzahl und den standardisierten Regressionskoeffizienten bestimmt. Es verblieben i.d.R. alle Variablen mit VIP-Kennzahlen größer 0,8 und einen standardisierten Regressionskoeffizienten größer 0,01 bzw. 0,02 in der Analyse. Technologisch bedeutsame Variablen, die diese Grenzen im Lerndatensatz unterschritten, blieben ebenfalls in der Datenmatrix.

9.4.6.1 Querzugfestigkeit

Für die Querzugfestigkeit ist nach der Cross-Validierung eine maximale Schätzgenauigkeit zwischen 3 und 4 Komponenten zu erwarten. Abbildung 55 stellt die Ergebnisse der "Leaveone-out"- und der Block-Validierung grafisch dar.



Abbildung 55: Zusammenhang zwischen Anzahl der PLS Komponenten und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-Validierung, Modell für die Querzugfestigkeit mit 281 Prozessvariablen

In der Abbildung 56 ist der Einfluss der Komponentenanzahl auf bedeutende Kennzahlen im Beobachtungs- und Vorhersageraum dargestellt. Die Bestimmtheitsmaße R_x^2 und R_y^2 steigen dabei mit zunehmender Komponentenanzahl und der Fehler des Modells (RMSE Modell) sinkt entsprechend. Die maximale Vorhersagegenauigkeit im Vorhersageraum wird mit 4 Komponenten erreicht. Der Fehler des Modells in der ex-ante-Vorhersage der Querzugfestigkeit liegt in diesem Modell bei 0,104 N/mm² bzw. bei 4,5%. Diese Werte bilden die Durchschnittswerte aus den ex-ante-Vorhersagen der 34 Testdaten innerhalb von 3 Monaten. Abbildung 57 gibt die auf die Standardabweichung der jeweiligen Datenräume bezogenen Schätzgenauigkeiten in Abhängigkeit der PLS-Faktoren wieder.



Abbildung 56: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im Beobachtungs- und Vorhersageraum (IB-Modell mit 281 Prozessvariablen)



Abbildung 57: Einfluss der PLS Faktoren auf die normierte Schätzgenauigkeit im Beobachtungs- und Vorhersageraum (IB-Modell mit 281 Prozessvariablen)

In der Abbildung 57 ist deutlich zu sehen, dass die Schätzgenauigkeiten im Beobachtungsund Vorhersageraum ab der 4. bzw. 5. Komponente zunehmend auseinander driften. Das bedeutet, dass ab dieser Komponentenanzahl ein "overfitting" des Beobachtungsraumes stattfindet. Das Bestimmtheitsmaß bzw. der Fehler des Modells für die Querzugfestigkeit bei maximaler Vorhersagegenauigkeit im Vorhersageraum (4 Komponenten) liegt mit R_y^2 =0,62 bzw. mit RMSE=0,088 N/mm² in einem Bereich, in dem auch die Regressionsmodelle die maximalen Schätzgenauigkeiten im Vorhersageraum aufweisen. Tabelle 59 zeigt im Detail die Kennzahlen im Beobachtungs- und Vorhersageraum für die Modelle der Querzugfestigkeit mit 2 bis 5 PLS Komponenten.

Kennzahlen	LAB_IB	LAB_IB	LAB_IB	LAB_IB
R-Square Rx ²	0,34	0,40	0,48	0,50
R-Square Ry ²	0,43	0,57	0,61	0,69
PLS Faktoren	2	3	4	5
MAD	0,086	0,077	0,070	0,063
MAPE	4,55	4,08	3,72	3,33
RMSE (n)	0,106	0,093	0,088	0,079
RMS Pct Error	5,61	4,92	4,65	4,21
Korrel_Prog	0,48	0,56	0,60	0,59
ME_Prog	0,015	-0,010	-0,018	-0,018
MAD_Prog	0,091	0,090	0,084	0,085
MAPE_Prog	4,74	4,75	4,48	4,52
RMSE_Prog	0,114	0,109	0,104	0,104
RMS Pct Error_Prog	5,88	5,77	5,58	5,54

Tabelle 59: Gütekennzahlen IB-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)

Durch die individuelle Modellierung der Querzugfestigkeit in einem Modell kann die Modellkomplexizität stark verringert werden. Im Einzelmodell sind bei gleicher Schätzgenauigkeit im Vorhersageraum nur 4 PLS-Faktoren notwendig. Im multivariaten Modell wurden für die maximale Schätzgenauigkeit der Querzugfestigkeit dagegen 9 PLS-Komponenten benötigt. Eine höhere Schätzgenauigkeit des Einzelmodells gegenüber dem multivariaten Modell, bei gleicher Variablenanzahl, konnte dagegen nicht festgestellt werden. Eine höhere Schätzgenauigkeit kann jedoch im Einzelmodell durch eine gezielte Variablenreduktion wesentlich einfacher realisiert werden als im multivariaten Fall. Tabelle 60 zeigt die Modellkennzahlen für die Querzugfestigkeit für 2 bis 5 Komponenten bei Verwendung von nur insgesamt 44 Prozessvariablen.

Kennzahlen	LAB_IB	LAB_IB	LAB_IB	LAB_IB
R-Square Rx ²	0,22	0,31	0,37	0,43
R-Square Ry ²	0,59	0,63	0,65	0,67
PLS Faktoren	2	3	4	5
MAD	0,073	0,069	0,067	0,066
MAPE	3,87	3,69	3,58	3,51
RMSE (n)	0,090	0,086	0,083	0,081
RMS Pct Error	4,79	4,57	4,45	4,35
Korrel_Prog	0,69	0,67	0,63	0,63
ME_Prog	-0,018	-0,035	-0,017	-0,029
MAD_Prog	0,072	0,082	0,080	0,086
MAPE_Prog	3,83	4,36	4,28	4,61
RMSE_Prog	0,093	0,100	0,100	0,104
RMS Pct Error_Prog	5,02	5,48	5,40	5,69

Tabelle 60: Gütekennzahlen IB-Modelle mit 2-5 PLS-Faktoren (44 Prozessvariablen)

Die maximale Schätzgenauigkeit im Vorhersageraum weist das Modell mit 2 Komponenten auf. Der ex-ante-Vorhersagefehler beträgt in diesem Modell für die Querzugfestigkeit 0,093 N/mm² bzw. 3,8%. Die Höhe des ex-ante-Fehlers entspricht dabei dem des ex-post-Fehlers. Durch die Eliminierung unbedeutender Variablen kann die Schätzgenauigkeit der Querzugfestigkeit verbessert werden. Die Kennzahlen dieses PLS-Modells entsprechen dabei weitgehend den Kennzahlen des kombinierten Modells aus der Regressionsanalyse.

Abbildung 58 stellt den Einfluss der PLS-Komponenten auf die wichtigsten Kennzahlen des Modells im Beobachtungs- und Vorhersageraum grafisch dar.



Abbildung 58: Einfluss der PLS-Faktoren auf Kennzahlen des Modells (IB) im Beobachtungsund Vorhersageraum (44 Prozessvariablen)

Der ex-post-Fehler wird im Lerndatenraum mit zunehmender Komponentenanzahl nur geringfügig verkleinert. Das Bestimmtheitsmaß R_y^2 weist ab der 2. Komponente nur noch einen sehr flachen Anstieg auf, obwohl die erklärte Varianz des x-Raums R_x^2 mit zunehmender Komponentenzahl weiter ansteigt. Das bedeutet, dass eine höhere Anpassung des x-Raums mit 44 Variablen keine wesentliche Verbesserung in der Vorhersage im Y-Raum nach sich zieht. Im Vorhersageraum wird bei 2 Komponenten der geringste Fehler erreicht, steigt jedoch mit weitern Komponenten nur noch geringfügig an. Abbildung 59 zeigt die normierte Vorhersagegenauigkeit der beiden Datenräume. Beide Kurven verlaufen aufgrund der starken Variablenreduktion annährend parallel.

Die Reduktion von 281 auf 44 Variablen, ohne Einbußen in der Schätzgenauigkeit hinzunehmen, offenbart deutlich, dass nur sehr wenige Prozessvariablen tatsächlich einen bedeutenden Einfluss auf die Querzugfestigkeit ausüben. Der Einfluss der Prozessvariablen auf die Platteneigenschaften ist jedoch ein dynamischer Prozess, die Gewichte der Einflussgrößen verschieben sich mit der Zeitachse. Aus diesem Grund sind für längere Vorhersageräume mehr Variablen in den Modellen zu berücksichtigen, als zum Zeitpunkt to zur Vorhersage notwendig wären. Die Pressengeschwindigkeit zeigte z.B. mit einem standardisierten Regressionskoeffizienten von 0 zu Beginn der Vorhersage keinen Einfluss des Testdatensatzes jedoch auf die Querzugfestigkeit. Am Ende wies der Regressionskoeffizient einen Wert von -0,8 auf. Dieses Beispiel verdeutlicht, dass eine Variablenauswahl auf Basis eines Lerndatensatzes und statistischer Kennzahlen für zukünftige Vorhersagen nicht ausreichend ist. Zusätzlich sollten technologisch bedeutsame (und bekannte) Prozessvariablen in die Modelle aufgenommen werden, auch wenn sie

aufgrund geringer Schwankungen bis zu diesem Zeitpunkt keinen Einfluss auf die Platteneigenschaften aufweisen.

Das Konfliktfeld der statistischen Prozessmodellierung liegt damit in einer maximal möglichen Berücksichtigung von Prozessvariablen und einer ausreichend hohen Schätzgenauigkeit in der Zukunft. Bei der Berücksichtigung aller Prozessvariablen in der PLS-Analyse ist sichergestellt, dass zukünftige Prozessänderungen und deren Auswirkungen auf die Platteneigenschaften erfasst werden können. Der Nachteil dieser umfassenden Variablenüberwachung innerhalb eines Modells liegt in der durchschnittlich geringeren Vorhersagegenauigkeit. Dieser Nachteil beziffert sich im vorliegenden Fall auf eine prozentuale durchschnittliche Abweichung der Vorhersagen von den Laborwerten von 3,8% (44 Variablen) bzw. 4,5% (281 Variablen).



Abbildung 59: Einfluss der PLS-Faktoren auf die normierte Schätzgenauigkeit im Beobachtungs- und Vorhersageraum (IB-Modell mit 44 Prozessvariablen)

9.4.6.2 Dickenquellung

Abbildung 60 zeigt die Ergebnisse der Cross-Validierung für das PLS-Modell der Dickenquellung bei Verwendung aller 281 Prozessvariablen. Die maximale Vorhersagegenauigkeit ist danach zwischen 3 und 4 PLS-Komponenten zu erwarten.



Abbildung 60: Zusammenhang zwischen Anzahl der PLS-Faktoren und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-Validierung, Modell für die Dickenquellung mit 281 Prozessvariablen

Die Ergebnisse der Validierung anhand des Testdatensatzes bestätigen die Ergebnisse der Cross-Validierung. In Abbildung 61 unten ist der minimale ex-ante-Vorhersagefehler bei 4 PLS Komponenten ersichtlich. Der ex-ante-Fehler der Vorhersage der Dickenquellung liegt bei Verwendung aller Prozessvariablen bei durchschnittlich 6,2% (MAPE_Prog) bzw. bei 0,66% (RMSE_Prog) und ist damit geringfügig kleiner als im multivariaten PLS-Modell. Jedoch sind im Einzelmodell für die annährend gleiche Modellperfomance nur 4 PLS-Komponenten notwendig.

Im Vergleich zu dem Regressionsmodell mit der höchsten Vorhersagegenauigkeit (Modell10) sind die Kennwerte beider Modelltypen nahezu identisch. Im Regressionsmodell wurden 29 Prozessvariablen berücksichtigt. Eine mögliche Berücksichtigung aller Prozessvariablen in dem PLS-Modell führt damit nicht zu einer Erhöhung der Vorhersagegenauigkeit. Andererseits können durch das PLS-Modell alle Prozessvariablen in einem Modell überwacht werden, ohne Verluste in der Vorhersagegenauigkeit hinnehmen zu müssen. Einen bedeutenden Einfluss auf die Dickenquellung üben jedoch nur wenige Prozessvariablen aus. Eine Verringerung der Prozessvariablen von 281 zu 60 Variablen führt im PLS-Modell entsprechend auch zu einer höheren ex-ante-Vorhersagegenauigkeit. Tabelle 61 stellt die Kennzahlen des Modells mit Berücksichtigung aller Prozessvariablen den Kennzahlen des PLS-Modells mit reduzierter Variablenauswahl in Tabelle 62 gegenüber. Die maximale ex-ante-Vorhersagegenauigkeit wird mit reduzierter Variablenanzahl bei 3 Komponenten und einem prozentualen Fehler von 5,95% erreicht.



Abbildung 61: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im Beobachtungs- und Vorhersageraum (TS-Modell mit 281 Prozessvariablen)

In Abbildung 62 ist die normierte Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum für die Dickenquellung in Abhängigkeit der PLS-Komponenten dargestellt. Wie schon bei der Querzugfestigkeit beobachtet werden konnte, öffnet sich die Schere der Schätzgenauigkeit in beiden Datenräumen mit zunehmender Komponentenanzahl.



Abbildung 62: Einfluss der PLS-Faktoren auf die normierte Schätzgenauigkeit im Beobachtungs- und Vorhersageraum (TS-Modell mit 281 Prozessvariablen)

Kennzahlen	LAB_TS	LAB_TS	LAB_TS	LAB_TS
R-Square Rx ²	0,34	0,42	0,46	0,49
R-Square Ry ²	0,41	0,57	0,69	0,75
PLS Faktoren	2	3	4	5
MAD	0,566	0,502	0,432	0,374
MAPE	6,65	5,86	5,00	4,31
RMSE (n)	0,719	0,614	0,522	0,470
RMS Pct Error	8,87	7,36	6,16	5,47
Korrel_Prog	0,47	0,55	0,58	0,54
ME_Prog	0,091	0,085	-0,028	0,099
MAD_Prog	0,605	0,616	0,558	0,586
MAPE_Prog	6,63	6,76	6,19	6,39
RMSE_Prog	0,730	0,700	0,661	0,702
RMS Pct Error_Prog	8,03	7,71	7,45	7,65

Tabelle 61: Gütekennzahlen TS-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)

Tabelle 62: Gütekennzahlen TS-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen)

Kennzahlen	LAB_TS	LAB_TS	LAB_TS	LAB_TS
R-Square Rx ²	0,20	0,25	0,31	0,51
R-Square Ry ²	0,62	0,67	0,71	0,72
PLS Faktoren	2	3	4	5
MAD	0,471	0,436	0,403	0,387
MAPE	5,49	5,04	4,69	4,48
RMSE (n)	0,578	0,534	0,507	0,495
RMS Pct Error	6,95	6,31	6,02	5,85
Korrel_Prog	0,64	0,66	0,60	0,56
ME_Prog	0,291	0,069	0,089	0,185
MAD_Prog	0,600	0,541	0,558	0,579
MAPE_Prog	6,46	5,95	6,14	6,33
RMSE_Prog	0,712	0,632	0,680	0,721
RMS Pct Error_Prog	7,54	7,01	7,61	7,96

9.4.6.3 Abhebefestigkeit

Abbildung 63 zeigt die Ergebnisse der Cross-Validierung für das PLS-Modell der Abhebefestigkeit bei Verwendung aller 281 Prozessvariablen. Im Gegensatz zu den Ergebnissen der Querzugfestigkeit und der Dickenquellung ist in der PRESS-Statistik der Abhebefestigkeit in Abhängigkeit der PLS-Komponenten kein ausgeprägtes Minimum zu sehen. Die maximale Schätzgenauigkeit ist je nach Verfahren der Cross-Validierung zwischen 2 und 5 PLS Komponenten zu erwarten.

Abbildung 64 zeigt die Bestimmtheitsmaße R_x² und R_y², die absoluten und prozentualen Fehler im Beobachtungs- bzw. Vorhersageraum für die PLS-Modelle der Abhebefestigkeit in Abhängigkeit der Komponenten. Die höchste Schätzgenauigkeit im Vorhersageraum wird bei 4 bzw. 5 PLS-Komponenten erzielt und weist keine Verbesserung gegenüber dem multivariaten Modell auf. In der normierten Schätzgenauigkeit in Abbildung 65 ist das geringe ex-ante-Vorhersageniveau der Abhebefestigkeit zu sehen. Auch eine Verringerung der Variablenanzahl in der PLS-Analyse führt bei der Modellierung der Abhebefestigkeit nicht zu einer Verbesserung der Vorhersagegenauigkeit. Der ursächliche Grund dürfte in der höheren Fehlerquote der Messung der Abhebefestigkeit und den daraus resultierenden geringeren Korrelationen zu den Prozessvariablen liegen. Tabelle 63 und Tabelle 64 geben die Kennzahlen für die PLS-Modelle mit 2 bis 5 PLS-Komponenten auf Basis der gesamten und reduzierten Variablenmatrix im Detail wieder.



Abbildung 63: Zusammenhang zwischen Anzahl der PLS-Faktoren und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-Validierung, Modell für die Abhebefestigkeit (SS) mit 281 Prozessvariablen



Abbildung 64: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im Beobachtungs- und Vorhersageraum (SS-Modell mit 281 Prozessvariablen)




Kennzahlen	LAB_SS	LAB_SS	LAB_SS	LAB_SS
R-Square Rx ²	0,38	0,43	0,47	0,50
R-Square Ry ²	0,27	0,44	0,54	0,62
PLS Faktoren	2	3	4	5
MAD	0,158	0,137	0,122	0,112
MAPE	6,86	6,02	5,36	4,92
RMSE (n)	0,201	0,175	0,158	0,145
RMS Pct Error	8,69	7,70	6,97	6,41
Korrel_Prog	0,16	0,34	0,45	0,44
ME_Prog	0,055	-0,028	-0,019	0,001
MAD_Prog	0,174	0,165	0,151	0,153
MAPE_Prog	7,14	7,01	6,47	6,54
RMSE_Prog	0,208	0,198	0,186	0,184
RMS Pct Error_Prog	8,39	8,68	8,24	8,13

Tabelle 64: Gütekennzahlen SS-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen)

Kennzahlen	LAB_SS	LAB_SS	LAB_SS	LAB_SS
R-Square Rx ²	0,31	0,39	0,45	0,53
R-Square Ry ²	0,44	0,50	0,54	0,57
PLS Faktoren	2	3	4	5
MAD	0,137	0,131	0,124	0,120
MAPE	6,01	5,73	5,43	5,25
RMSE (n)	0,175	0,166	0,159	0,154
RMS Pct Error	7,70	7,31	7,01	6,79
Korrel Prog	0,33	0,36	0,35	0,27
ME_Prog	-0,085	-0,017	0,004	-0,017
MAD Prog	0,183	0,169	0,176	0,180
MAPE Prog	7,96	7,19	7,45	7,71
RMSE_Prog	0,215	0,195	0,200	0,216
RMS Pct Error_Prog	9,69	8,52	8,65	9,61

9.4.6.4 Rohdichte

In der Abbildung 66 sind die Ergebnisse der Cross-Validierung für das PLS-Modell der Rohdichte bei Verwendung aller Prozessvariablen zu sehen. Ähnlich den Modellen der Abhebefestigkeit weist die Vorhersagefehler kein deutliches Minimum in der PRESS-Statistik auf. Die maximale Schätzgenauigkeit ist jedoch zwischen 2 und 5 PLS-Komponenten zu erwarten.



Abbildung 66: Zusammenhang zwischen Anzahl der PLS-Faktoren und der Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-Validierung, Modell für die Rohdichte (MD) mit 281 Prozessvariablen

Abbildung 67 stellt den Einfluss der Anzahl der PLS Komponenten für die PLS-Modelle der Rohdichte auf die wichtigsten Kennzahlen des Beobachtungs- und Vorhersageraums dar. Die höchste Schätzgenauigkeit im Vorhersageraum wird mit 2 PLS-Komponenten erreicht. Im Beobachtungsraum findet dagegen mit jeder weiteren Komponente eine höhere Anpassung der ex-post-Vorhersagen an die Labordaten statt. Mit der 4 PLS-Komponente wird der Vorhersagefehler (RMSE_Prog) größer als die Standardabweichung der Testdaten. Das Modell ist bei 4 Komponenten mit einem Bestimmtheitsmaß von Ry²=0,64 "overfitted".

Abbildung 68 zeigt die genormte Schätzgenauigkeit der Rohdichte in Abhängigkeit der PLS Komponenten in beiden Datenräumen. Die Schere der Vorhersagegenauigkeiten in den beiden Datenräumen laufen ab der 2 Komponente zunehmend auseinander. Die Ursache für das geringe ex-ante-Vorhersageniveau liegt in der sehr geringen Schwankungsbreite der Rohdichte sowohl im Beobachtungs- als auch im Vorhersageraum. Die prozentuale ex-ante-Vorhersagefehler ist jedoch mit 0,82% äußerst gering. Das bedeutet, dass die Rohdichte im Mittel mit einer Genauigkeit von über 99% vorhergesagt werden kann. Schwierig dagegen sind die sehr geringen Abweichungen vom Mittelwert exakt vorherzusagen.

Eine Reduktion der Prozessvariablen erhöht die Schätzgenauigkeit im Vorhersageraum weiter. Mit 60 Variablen und 2 Komponenten liegt der ex-ante-Vorhersagefehler der Rohdichte bei 8,3 kg/m³ bzw. 0,79%. Tabelle 65 und Tabelle 66 zeigen für die gesamte und reduzierte Datenmatrix die Ergebnisse der Modelle im Detail.



Abbildung 67: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im Beobachtungs- und Vorhersageraum (MD-Modell mit 281 Prozessvariablen)



Abbildung 68: Einfluss der PLS Faktoren auf die normierte Schätzgenauigkeit im Beobachtungs- und Vorhersageraum (MD-Modell mit 281 Prozessvariablen)

Kennzahlen	LAB_MD	LAB_MD	LAB_MD	LAB_MD
R-Square Rx ²	0,38	0,43	0,47	0,50
R-Square Ry ²	0,36	0,55	0,64	0,71
PLS Faktoren	2	3	4	5
MAD	0,006	0,005	0,005	0,004
MAPE	0,72	0,59	0,54	0,49
RMSE (n)	0,0079	0,0066	0,0060	0,0054
RMS Pct Error	0,89	0,74	0,67	0,60
Korrel_Prog	0,36	0,39	0,34	0,33
ME_Prog	0,001	-0,002	-0,002	0,000
MAD_Prog	0,007	0,008	0,008	0,008
MAPE Prog	0,82	0,85	0,90	0,89
RMSE_Prog	0,0086	0,0090	0,0094	0,0095
RMS Pct Error_Prog	0,96	1,01	1,06	1,06

Tabelle 65: Gütekennzahlen MD-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)

Tabelle 66: Gütekennzahlen MD-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen)

Kennzahlen	LAB_MD	LAB_MD	LAB_MD	LAB_MD
R-Square Rx ²	0,40	0,51	0,55	0,59
R-Square Ry ²	0,41	0,47	0,51	0,54
PLS Faktoren	2	3	4	5
MAD	0,006	0,006	0,005	0,005
MAPE	0,69	0,66	0,61	0,60
RMSE (n)	0,0077	0,0073	0,0070	0,0068
RMS Pct Error	0,87	0,82	0,79	0,77
Korrel_Prog	0,41	0,40	0,29	0,30
ME_Prog	0,000	-0,002	-0,004	-0,002
MAD Prog	0,007	0,007	0,008	0,008
MAPE_Prog	0,79	0,82	0,94	0,85
RMSE Prog	0,0083	0,0088	0,0103	0,0097
RMS Pct Error_Prog	0,93	0,99	1,16	1,10

Erkenntnisse

Das multivariate PLS-Modell tendiert aufgrund der unterschiedlichen Platteneigenschaften zu einer hohen Modelldimension. Bei Berücksichtigung aller plausiblen Prozessvariablen erreichen im gemeinsamen Modell die Dickenquellung und die Querzugfestigkeit erst mit 8 bzw. 9 und die Abhebefestigkeit und Rohdichte mit 5 PLS-Faktoren ihre maximale Schätzgenauigkeit im Vorhersageraum. Mit 8 bzw. 9 Faktoren werden jedoch die beiden letztgenannten Eigenschaften zu stark im Beobachtungsraum angepasst. Die ex-ante-Vorhersagegenauigkeit dieser Eigenschaften ist ungenügend. In einem Modell mit 5 PLS-Faktoren dagegen muss eine geringere Schätzgenauigkeit der Dickenquellung und der Querzugfestigkeit in Kauf genommen werden. Eine Reduktion auf 200 Prozessvariablen führt im Durchschnitt aller Platteneigenschaften zu einer geringfügig höheren Schätzgenauigkeit im Vorhersageraum. Das Problem des unterschiedlichen Anpassungsniveaus der Eigenschaftsmerkmale bleibt jedoch in einem gemeinsamen PLS-Modell bestehen.

In den Einzelmodellen werden zur Erreichung der maximalen Schätzgenauigkeiten im Vorhersageraum bei Berücksichtigung aller Prozessvariablen für die Rohdichte nur 2 und für die restlichen Platteneigenschaften 4 PLS-Faktoren benötigt. Durch eine individuelle Variablenreduzierung konnten bis auf die Abhebefestigkeit die Schätzgenauigkeiten der Einzelmodelle im Vorhersageraum verbessert werden. Die berücksichtigten 44 bis 60 Variablen in den Einzelmodellen wurden dabei auf 2 bzw. 3 latente Variablen reduziert.

Die prozentualen ex-ante-Vorhersagefehler lagen über den validierten Zeitraum von 3 Monaten für die Rohdichte bei ca. 0,8%, für die Querzugfestigkeit bei kleiner 4%, für die Dickenquellung bei ca. 6% und schließlich für die Abhebefestigkeit bei ca. 7%.

9.4.7 Langfristige Vorhersagen

Um zu überprüfen, wie stabil die PLS-Modelle die Schwankungen der Platteneigenschaften längerfristig erklären können, wurden die ex-ante-Vorhersagen der multivariaten als auch der Einzelmodelle über weitere 9 Monate und damit insgesamt über 12 Monate validiert. Tabelle 67 und Tabelle 68 zeigen die Ergebnisse der einjährigen Validierung für die multivariaten PLS-Modelle unter Berücksichtigung von 281 bzw. 200 Prozessvariablen.

Bei Berücksichtigung aller Variablen verfehlen die Vorhersagen zur Dickenquellung und Abhebefestigkeit knapp die aufgestellte Anforderung, dass die geschätzte Standardabweichung der Vorhersagen (RMSE_Prog) kleiner sein sollte als die Standardabweichung der Testdaten selbst. Mit der reduzierten Variablenmatrix kann das multivariate Modelle dagegen alle gestellten Anforderungen an die Vorhersagegenauigkeiten erfüllen.

Tabelle 67: Gütekennzahlen gemeinsames Modell mit 5 PLS-Faktoren über einen Zeittraum von 1 Jahr (281 Prozessvariablen)

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
STD_Testdaten	0,140	0,887	0,186	0,010
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square Rx ²	0,52	0,52	0,52	0,52
R-Square Ry ²	0,51	0,52	0,33	0,44
PLS Faktoren	5	5	5	5
MAD	0,080	0,547	0,146	0,006
MAPE	4,27	6,24	6,37	0,66
RMSE (n)	0,098	0,672	0,184	0,0075
RMS Pct Error	5,23	7,86	8,15	0,84
Korrel_Prog	0,64	0,43	0,29	0,48
ME_Prog	-0,018	0,270	0,031	-0,001
MAD_Prog	0,092	0,726	0,151	0,007
MAPE_Prog	5,13	7,45	6,39	0,82
RMSE_Prog	0,114	0,895	0,190	0,0089
RMS Pct Error_Prog	6,39	8,95	7,99	1,00

Tabelle 68: Gütekennzahlen gemeinsames Modell mit 6 PLS-Faktoren über einen Zeittraum von 1 Jahr (200 Prozessvariablen)

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
STD_Testdaten	0,140	0,887	0,186	0,0098
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square Rx ²	0,63	0,63	0,63	0,63
R-Square Ry ²	0,56	0,57	0,39	0,44
PLS Faktoren	6	6	6	6
MAD	0,077	0,518	0,138	0,006
MAPE	4,08	5,90	6,01	0,65
RMSE (n)	0,093	0,635	0,175	0,0075
RMS Pct Error	4,98	7,37	7,71	0,84
Korrel_Prog	0,63	0,47	0,33	0,48
ME_Prog	-0,030	0,230	0,015	-0,001
MAD_Prog	0,093	0,674	0,148	0,007
MAPE_Prog	5,21	6,95	6,30	0,79
RMSE_Prog	0,117	0,856	0,185	0,0088
RMS Pct Error_Prog	6,65	8,63	7,93	0,99

Die Genauigkeit der ex-ante-Vorhersagen der Dickenquellung und der Abhebefestigkeit nimmt im einjährigen Mittel gegenüber den ersten 3 Monaten ab. Die Genauigkeiten der exante-Vorhersagen der Querzugfestigkeit und der mittleren Rohdichte sind dagegen im Jahresmittel leicht zunehmend. Tabelle 69 fasst die auf die Standardabweichungen bezogenen Schätzgenauigkeiten der multivariaten Modelle im Beobachtungs- und Vorhersageraum über 3 und 12 Monate zusammen.

Tabelle 69: Genormte Schätzgenauigkeiten der Gesamtmodelle über 3 und 12 Monate im Vergleich

Gesamtmodell	Kz	Lab_IB	Lab_TS	Lab_SS	Lab_MD	Gesamt
281 Variablen	1	5	5	5	5	
3 Monate	2	1,48	1,49	1,20	1,45	1,41
	3	1,12	1,14	1,11	1,08	1,11
281 Variablen	1	5	5	5	5	
12 Monate	2	1,47	1,40	1,22	1,38	1,37
	3	1,23	0,99	0,98	1,11	1,08
200 Variablen	1	6	6	6	6	
3 Monate	2	1,57	1,57	1,29	1,43	1,46
	3	1,14	1,21	1,12	1,06	1,13
200 Variablen	1	6	6	6	6	
12 Monate	2	1,54	1,48	1,28	1,39	1,42
	3	1,19	1,04	1,00	1,11	1,09
1 Anzahl der PLS-Faktoren im Modell						

Quotient aus STD_Beobachtungsraum und RMSE_Modell

Quotient aus STD_Testdatenraum und RMSE_Prog

Tabelle 70 und Tabelle 71 fassen die Kennzahlen der einjährigen Validierung für die PLS-Einzelmodelle bei Berücksichtigung der gesamten und einer reduzierten Variablenmatrix zusammen. Die Querzugfestigkeit und die Dickenquellung können in den Einzelmodellen genauer geschätzt werden als in den multivariaten Modellen. Die Vorhersagegenauigkeit der mittleren Rohdichte ist im multivariaten Modell und im Einzelmodell nahezu identisch. Einzig die Abhebefestigkeit ist in den multivariten Modellen genauer zu schätzen als in den Einzelmodellen.

2

3

Tabelle 70: Gütekennzahlen Einzelmodelle mit unterschiedlichen PLS-Faktoren über einen Zeittraum von 1 Jahr (281 Prozessvariablen)

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
STD_Testdaten	0,140	0,887	0,186	0,010
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square Rx ²	0,48	0,46	0,47	0,38
R-Square Ry ²	0,60	0,65	0,51	0,38
PLS Faktoren	4	4	4	2
MAD	0,072	0,467	0,124	0,006
MAPE	3,85	5,28	5,41	0,72
RMSE (n)	0,089	0,576	0,158	0,0080
RMS Pct Error	4,72	6,58	6,95	0,89
Korrel_Prog	0,66	0,45	0,28	0,45
ME_Prog	-0,018	-0,046	-0,025	0,000
MAD_Prog	0,088	0,676	0,161	0,007
MAPE_Prog	4,91	7,10	6,96	0,81
RMSE_Prog	0,110	0,850	0,201	0,0088
RMS Pct Error_Prog	6,20	8,87	8,86	1,00

Kennzahlen	LAB_IB	LAB_TS	LAB_SS	LAB_MD
STD_Testdaten	0,140	0,887	0,186	0,010
CV_Testdaten	7,6	9,3	7,9	1,1
R-Square Rx ²	0,22	0,25	0,39	0,40
R-Square Ry ²	0,55	0,63	0,45	0,41
PLS Faktoren	2	3	3	2
MAD	0,076	0,474	0,134	0,006
MAPE	4,02	5,35	5,82	0,71
RMSE (n)	0,093	0,589	0,167	0,0079
RMS Pct Error	4,97	6,74	7,30	0,88
Korrel_Prog	0,65	0,47	0,31	0,47
ME_Prog	-0,020	0,113	-0,029	-0,001
MAD_Prog	0,089	0,666	0,158	0,007
MAPE_Prog	4,94	6,92	6,81	0,80
RMSE_Prog	0,111	0,834	0,193	0,0088
RMS Pct Error_Prog	6,30	8,51	8,39	0,99

Tabelle 71: Gütekennzahlen Einzelmodelle mit unterschiedlichen PLS-Faktoren über einen Zeittraum von 1 Jahr (44-60 Prozessvariablen)

Die höhere Schätzgenauigkeit der Abhebefestigkeit im multivariaten Modell ist u.a. auf die geringere Anpassung dieses Modells im Beobachtungsdatenraum zurückzuführen. Damit werden die Messfehler der Abhebefestigkeit weniger stark auf die ex-ante-Vorhersagen übertragen. Zudem profitiert im gemeinsamen Modell die Abhebefestigkeit von der höheren Messgenauigkeit der anderen Platteneigenschaften, da die Bildung der y-Scores durch alle Platteneigenschaften bestimmt wird. Tabelle 72 fasst die genormten Schätzgenauigkeiten der PLS-Einzelmodelle für die unterschiedlichen Validierungszeiträume zusammen.

Tabelle 72: Genormte Schätzgenauigkeiten der PLS-Einzelmodelle über 3 und 12 Monate im Vergleich

Einzelmodelle	Kz	Lab_IB	Lab_TS	Lab_SS	Lab_MD	Gesamt
281 Variablen	1	4	4	4	2	
3 Monate	2	1,58	1,74	1,38	1,27	1,49
	3	1,23	1,28	1,07	1,06	1,16
281 Variablen	1	4	4	4	2	
12 Monate	2	1,62	1,64	1,42	1,30	1,50
	3	1,27	1,04	0,92	1,11	1,09
red. Variablen	1	2	3	3	2	
3 Monate	2	1,56	1,71	1,32	1,31	1,48
	3	1,38	1,34	1,03	1,10	1,21
red. Variablen	1	2	3	3	2	
12 Monate	2	1,55	1,60	1,35	1,32	1,46
	3	1,26	1,06	0,96	1,12	1,10
1 Anzahl dar DLS Eaktoron im Madall						

Anzahl der PLS-Faktoren im Modell

Quotient aus STD_Beobachtungsraum und RMSE_Modell

Quotient aus STD_Testdatenraum und RMSE_Prog

Die Abnahme der Vorhersagegenauigkeit der Dickenquellung und Abhebefestigkeit ist verbunden mit einer Abnahme der Bestimmtheitsmaße dieser Eigenschaften über den gesamten einjährigen Zeitraum der Validierung. Die Bestimmtheitsmaße der Querzug-festigkeit und der mittleren Rohdichte weisen nach einer anfänglichen Abnahme dagegen eine Stabilisierung dieser Kennzahl ab ca. der 60. Vorhersage auf. In Abbildung 69 ist der Verlauf der Bestimmtheitsmaße für die PLS-Einzelmodelle während des einjährigen Vorhersagezeitraums zu sehen. Eine Verringerung des Bestimmtheitsmaßes weist auf einen abnehmenden Erklärungsanteil der Variablen im Modell hin. Es können verschiedene Ursachen hierfür als Erklärung dienen:

2

3

- (1) Verschiebung der Einflussanteile (Signifikanzen) zwischen den Variablen
- (2) zunehmende Bedeutung von nicht im Modell enthaltenen Variablen
- (3) Umbaumaßnahmen in der Anlage



Abbildung 69: Entwicklung der Bestimmtheitsmaße der PLS-Einzelmodelle während der exante-Vorhersagen über 1 Jahr (reduzierte Variablenanzahl)

Erkenntnisse

Die Eigenschaften Querzugfestigkeit und die mittlere Rohdichte konnten mit allen aufgestellten PLS-Modellen auch langfristig über ein Jahr mit einer hohen Genauigkeit vorhersagt werden. Alle Beurteilungskriterien für ein stabiles Vorhersagemodell wurden von den Modellen erfüllt. Die Fehler der Echtzeitvorhersagen lagen im Jahresdurchschnitt in allen Modellen der Querzugfestigkeit bei ca. 0,11 N/mm² bzw. 5% und bei der mittleren Rohdichte bei 8,8 kg/m³ bzw. 0,8%.

Die Genauigkeit der ex-ante-Vorhersagen der Dickenquellung liegt im Jahresdurchschnitt bei absolut ca. 0,85% und relativ bei 7%. Die Schätzgenauigkeit der Dickenquellung nimmt im Jahresdurchschnitt von allen Eigenschaften am stärksten ab. Insgesamt werden jedoch auch im Jahresmittel die Beurteilungskriterien für ein stabiles Vorhersagemodell von den PLS-Einzelmodellen und dem multivariaten Modell mit reduzierter Variablenanzahl erreicht. Der absolute Fehler der ex-ante-Vorhersage (RMSE_Prog.) liegt mit 0,85% jedoch im validierten Zeitraum von insgesamt einem Jahr nur geringfügig unter der Standardabweichung der Testdaten mit 0,89%.

Die Abhebefestigkeit ist langfristig die am ungenauesten zu schätzende Eigenschaft. Obwohl der langfristige prozentuale Vorhersagefehler mit 6,3-6,9% unterhalb der Dickenquellung liegt, ist der langfristige absolute ex-ante-Vorhersagefehler i.d.R. geringfügig höher als die Standardabweichung der Testdaten. Einzig im multivariaten Modell mit reduzierter Variablenanzahl ist der Fehler der ex-ante-Vorhersage gleich der Standardabweichung der Testdaten. Das bedeutet, dass das langfristige Niveau der Abhebefestigkeit mit den PLS-Modellen im Mittel gut geschätzt, die kurzfristigen Schwankungen um den Mittelwert jedoch nicht ausreichend vorhersagt werden können.

9.4.8 Technologische Zusammenhänge

Die aus den Regressionsmodellen entwickelten technologischen Zusammenhänge konnten durch die PLS-Modelle grundsätzlich bestätigt werden. Durch die Berücksichtigung von korrelierten Prozessvariablen können die PLS-Modelle mehr signifikante technologische Beziehungen aufzeigen, als die Regressionsmodelle. Jedoch führte die zusätzliche Berücksichtigung dieser Prozessvariablen bzw. Zusammenhänge in den PLS-Modellen zu keiner höheren Schätzgenauigkeit im Vorhersageraum. Der Grund dafür liegt darin, dass eine Gruppe von technologischen Zusammenhängen i.d.R. auf eine grundsätzliche Beziehung reduziert werden kann. Dies ist vor allem dann der Fall, wenn Prozessvariablen nicht unabhängig voneinander eingestellt werden und miteinander korreliert sind.

Im Folgenden werden an exemplarischen Beispielen aus dem Trocknungs- und Sichtungsprozess, der Beleimung und der Heißpresse anhand der PLS-Regressionskoeffizienten die Beziehungen korrelierter Variablen auf die Platteneigenschaften dargestellt. Die PLS-Regressionskoeffizienten geben den gemeinsamen Einfluss der Prozessvariablen auf die Zieleigenschaften wieder. Eine isolierte Betrachtung der Koeffizienten ist prinzipiell nicht möglich, da Korrelationen zwischen den Prozessvariablen im Modell zugelassen werden.

Trocknung und Sichtung

Durch die PLS-Modelle kann nachgewiesen werden, dass eine Erhöhung der Temperatur im Trocknungs- bzw. Sichtungsprozess mit einer verringerten Querzugfestigkeit, Abhebefestigkeit und mittleren Rohdichte und einer höheren Dickenquellung verbunden ist. Dieser Nachweis kann für jede Temperaturmessstelle gemeinsam innerhalb eines PLS-Modells erbracht werden. In den Regressionsmodellen kann dagegen, aufgrund der Korrelationen zwischen den Temperaturmessstellen, dieser Zusammenhang nur exemplarisch an einer Messstelle aufgezeigt werden. Da jede Messstelle jedoch grundsätzlich den gleichen Effekt nur an unterschiedlichen Stellen im Trocknungs- und Sichtungsprozess darstellt - die zunehmende Leimaushärtung mit höheren Temperaturen - ist prinzipiell nur eine Temperaturvariable aus diesem Prozessbereich im Modell notwendig. Aufgrund unterschiedlicher Störeinflüsse an den Messstellen, kann der tatsächliche Erklärungsanteil einer Temperaturvariable jedoch höher sein als der einer anderen Temperaturvariable. In den PLS-Modellen werden alle diese Temperaturmessungen zu einer idealerweise übergeordneten (latenten) Temperaturvariable zusammengefasst. Die Berechnung der PLS-Regressionskoeffizienten erfordert einer Rücktransformation der übergeordneten Temperaturvariable zu den einzelnen Temperaturvariablen. Diese PLS-Regressionskoeffizienten können nicht unabhängig voneinander diskutiert werden. Sie geben den gemeinsamen Einfluss der Prozessvariablen auf die Platteneigenschaften an.

Abbildung 70 stellt die PLS-Regressionskoeffizienten der Temperaturvariablen aus dem Trocknungs- und Sichtungsprozess sowie der Streubunkertemperatur (Formung) für die einzelnen Platteneigenschaften dar. Für jede Temperatur sind mit Ausnahme der Lufttemperatur nach dem Ventilator im Sichtungsprozess jeweils die PLS-Regressionskoeffizienten zu Beginn und am Ende der Vorhersage des einjährigen Testdatensatzes angegeben. Die PLS-Regressionskoeffizienten beziehen sich auf die skalierten und zentrierten Variablen. Da die Schwankungen der Rohdichte um den langfristigen Mittelwert

sehr gering sind, weisen die PLS-Regressionskoeffizienten der Rohdichte absolut gesehen die geringsten Werte auf.

Es ist in Abbildung 70 zu erkennen, dass die Temperaturen zu Beginn und am Ende der Vorhersage einen unterschiedlichen absoluten Einfluss auf die Platteneigenschaften ausüben. Die Unterschiede werden durch Veränderungen im Prozessdatenraum mit fortwährender Zeit verursacht. Die Einflüsse der einzelnen Prozessvariablen können sich entsprechend mit der Zeit verändern. Insbesondere die Temperatur am Trocknerausgang hat am Ende der Vorhersagen eine höhere Bedeutung für die Querzugfestigkeit und Abhebefestigkeit als zu Beginn der Vorhersagen. Die Temperatur am Trocknereingang, mit einen höheren Variationskoeffizienten als die Temperatur am Trocknerausgang, zeigt dagegen auf keine Platteneigenschaft einen signifikanten Einfluss.

Die Rohdichte wird, im Gegensatz zu den anderen Platteneigenschaften, von den meisten Temperaturen signifikant beeinflusst. Der absolute Einfluss ist jedoch sehr gering. Technologisch ist die Rohdichteabnahme mit höheren Trocknungstemperaturen auf einen zunehmenden Springback zurückzuführen.



Abbildung 70: Einfluss der Temperaturen im Trocknungs- und Sichtungsprozess sowie im Streubunker (PLS-Einzelmodelle, 281 Prozessvariablen, n.s. = nicht signifikant bei S=95%)

Beleimung

In den Regressionsmodellen konnte aus dem Bereich der Beleimung neben dem Leimlieferanten nur die Härter- und die Leimmenge als signifikante Einflussgrößen identifiziert werden. In den PLS-Modellen sind weitere signifikante Beziehungen zwischen den Beleimungsvariablen und den Platteneigenschaften nachweisbar. Abbildung 71 stellt die Beziehungen zwischen den Beleimungsvariablen und den Platteneigenschaften durch die standardisierten PLS-Regressionskoeffizienten dar. Zusätzlich zu den eigentlichen Beleimungsvariablen ist die Drehzahl der Austragsschnecke (DZ_SCHN) dargestellt, nach der die Durchflussmenge der Leimflotte geregelt wird.



Abbildung 71: Einfluss der Beleimungsvariablen auf die Platteneigenschaften (PLS-Einzelmodelle, 281 Prozessvariablen)

Die Querzug- und Abhebefestigkeit wurden in den PLS-Modellen nur durch die Härtermenge und den Leimlieferanten signifikant beeinflusst. Die Schwankungen der Leim-, Wasser-, Harnstoff- und Paraffinmenge beeinflussten dagegen die Eigenschaften nicht signifikant.

Die Dickenquellung wird, wie die Querzug- und Abhebefestigkeit, am stärksten durch die Härtermenge und den Leimlieferanten beeinflusst. Zusätzlich besteht bis auf die Harnstoffmenge ein signifikanter Zusammenhang zu allen anderen Variablen der Beleimung. Die Höhe der Paraffinzugabe hat von den Beleimungsparametern den geringsten Effekt auf die Dickenquellung in diesem Zeitraum. Die mittlere Rohdichte ist in den PLS-Modellen neben der Härtermenge auch von der Wasser- und Harnstoffmenge signifikant beeinflusst worden. Eine zunehmende Harnstoffmenge ist dabei mit einer geringeren Rohdichte verbunden. Die mittlere Rohdichte ist die einzige der untersuchten Eigenschaften, die von der Menge des eingesetzten Harnstoffs beeinflusst wurde. (Hier nicht betrachtet ist der Formaldehydgehalt, der durch die Zugabe von Harnstoff verringert werden soll.) Eine höhere Wasserzugabe, in Verbindung mit einer höheren Leimmenge, ist mit einer höheren mittleren Rohdichte verbunden.

Heißpresse

Der Einfluss der stark korrelierten Pressendistanzen und -drücke ist innerhalb der Regressionsmodelle nicht eindeutig bestimmbar. Signifikant auftretende Pressenrahmen sind als "Stellvertretervariable" für ein Pressenabschnitt bzw. für das gesamte Pressenprogramm aufzufassen. In der Faktorenanalyse wurden einzelne Pressenrahmen zu Pressabschnitten zusammengefasst. Innerhalb der Regressionsmodelle konnte anschließend der Einfluss verschiedener Pressenabschnitte auf die Platteneigenschaften nachgewiesen werden (siehe Tabelle 53). In den PLS-Modellen ist der Nachweis des gesamten Pressenprogramms auf die Platteneigenschaften berechenbar. Abbildung 72 zeigt den Einfluss des Pressprogramms anhand der einzelnen Pressendistanzen auf die Dickenquellung. Die Schwankungen innerhalb des Pressprogramms beeinflussen diese Eigenschaft am stärksten. Jedoch ist die absolute Bedeutung der Pressenvariablen mit ihren geringen Variationskoeffizienten innerhalb der Modelle für die Vorhersage weit weniger stark als z.B. die Schwankungen der Härtermenge oder der Temperaturen innerhalb des Streubunkers. Die Bedeutung der Variablen kann anhand der Höhe des standardisierten PLS-Regressionskoeffizienten beurteilt werden.



Abbildung 72: Einfluss der Pressendistanzen auf die Dickenquellung (PLS-Einzelmodell, 281 Prozessvariablen)

In der Abbildung 72 weisen die Rahmen 0, 1, 4, 5, 6, 19, 27 und 28 die höchsten PLS-Regressionskoeffizienten auf. Damit sind geringere Distanzen in der Schließphase und höhere Distanzen in der Lüftungs-, zweiten Verdichtungs- und in der Kalibrierzone mit höheren Werten in der Dickenquellung verbunden. Diese Zusammenhänge konnten ebenfalls mit den PCA-Pressenabschnitten innerhalb der Regressionsmodelle nachgewiesen werden (Tabelle 53). Die PLS-Analyse ermöglicht jedoch die direkte Einflussbestimmung der einzelnen Pressenrahmen.

Abbildung 73 zeigt zum Vergleich, den scheinbaren Einfluss der Pressendistanzen, der sich ergibt, wenn alle Pressendistanzen innerhalb eines Regressionsmodells berechnet werden. Dazu wurden alle Pressendistanzen in das kombinierte Regressionsmodell (Modell 10) eingesetzt und die standardisierten Regressionskoeffizienten zu Beginn und am Ende des Vorhersagezeitraums (1 Jahr) berechnet. Die Regressionskoeffizienten sind, augrund der Korrelationen zwischen den Pressendistanzen, irreführend und nicht interpretierbar. Sehr gut zu sehen ist der regelmäßige Wechsel in den Vorzeichen der Regressionskoeffizienten zwischen Pressenrahmen. Dieses Phänomen wird durch die Multikollinearität im Regressionsmodell verursacht.



Abbildung 73: Scheinbarer Einfluss der Pressendistanzen auf die Dickenquellung im Regressionsmodell

Das oben aufgeführte Beispiel zeigt deutlich, wie problematisch korrelierte Prozessvariablen für die Interpretation von Regressionsmodellen sind. Die Variablenauswahl ist für die Interpretierbarkeit von Regressionsmodellen entscheidend.

Abbildung 74 stellt den Einfluss der Multipotdrücke auf die Querzugfestigkeit und die Dickenquellung dar. Für die Eigenschaften Abhebefestigkeit und mittlere Rohdichte war kein einheitlicher und signifikanter Einfluss der Multipotdrücke über die Vorhersagezeit festzustellen. Nach den PLS-Modellen sind die Multipotdrücke im vorderen Bereich der Heißpresse (Rahmen 2 bis 6) für die Querzugfestigkeit und die Multipotdrücke in der Kalibrierzone (Rahmen 21, 23 und 25) für die Dickenquellung bedeutsam.



Abbildung 74: PLS-Regressionskoeffizienten der Multipotdrücke für die Querzugfestigkeit und Abhebefestigkeit (PLS-Einzelmodelle, 281 Prozessvariablen)

Erkenntnisse

Anhand der PLS-Modelle kann der gemeinsame Einfluss korrelierter Prozessvariablen auf die Platteneigenschaften dargestellt werden. Diese Möglichkeit ist insbesondere für die technologische Interpretation der Zusammenhänge zwischen Prozessvariablen und den Platteneigenschaften bedeutsam. Die durch die PLS-Modelle gewonnenen technologischen Zusammenhänge sind bei Prozessen mit stark korrelierten Variablen deutlich besser erklärbar als aus den Regressionsmodellen.

10 Diskussion und Schlussfolgerungen

10.1 Modellannahmen und -nutzen

Jedes Modell ist ein vereinfachtes Abbild eines realen Prozesses. "Ein Modell wird erstellt, um das System (Prozess) besser zu verstehen und entspricht nie genau der Realität. Modellieren bedeutet Hervorheben des Wesentlichen und Weglassen von unwichtigen Details. Wichtig bei der Erstellung eines Modells sind die Ziele, die mit dem Modell verfolgt werden, die Betrachter und Benutzer des Modells." (aus Grässle et al., 2004). Um reale Prozesse zu verstehen bzw. berechenbar zu machen, müssen jedoch Systemgrenzen bzw. Modellannahmen definiert werden. Die Bedeutung der Modellannahmen ist je nach Zielstellung unterschiedlich. Myers (1986) gibt 4 Ziele der Regressionsanalyse an, die sich jedoch überschneiden können:

- (1) Vorhersage,
- (2) Variablenauswahl,
- (3) Modellspezifizierung, (Systemerklärung)
- (4) Parameterschätzung.

In der vorliegenden Arbeit bestand das Hauptziel in der online-Vorhersage der Platteneigenschaften. Jedoch kann dieses Ziel nicht losgelöst von den anderen Zielen betrachtet werden.

Die Variablenauswahl war weniger ein Ziel der Untersuchungen, als viel mehr eine Notwendigkeit, um die grundlegenden Voraussetzungen der Regressionsanalyse einzuhalten und um interpretierfähige Modellzusammenhänge aufzufinden. Entsprechend war der Einfluss der Variablenauswahl auf die Vorhersagegenauigkeiten und auf die Interpretierfähigkeit der Modelle zu untersuchen.

In der Arbeit sind ausschließlich lineare Modelle mit Regressoren erster Ordnung (Modellspezifizierung) betrachtet worden, da der zugrundeliegende Datenraum der Zielvariablen sehr engen Grenzen unterworfen ist. Die Modelle umfassen jeweils einen Plattentyp, der durch Plattendicke und mittlere Rohdichte definiert wird. Die zu berücksichtigenden Schwankungen sind entsprechend gering. Die Einbeziehung quadratischer Anteile oder Wechselwirkungen zwischen Regressoren sind nur bei Versuchsplänen sinnvoll, die einen weiten Datenraum umfassen. Aufgrund der korrelierten Prozessdaten, deren gemeinsamer Einfluss in Regressionsmodellen nicht berechenbar ist, sind Regressionsmodelle zur Systemerklärung nur eingeschränkt verwendbar. Wenn ein großer Teil des Systems durch wenige Haupteinflüsse bestimmt wird ("system of reduced rank"), dann sollte dies nach Burmham et al. (2001) ein bedeutender Teil der Modellspezifizierung sein. Die PLSR-Modelle setzten voraus, dass der Prozess nur durch wenige latente Strukturen beeinflusst wird. Zur Systemerklärung von Prozessen, die Korrelationen zwischen Einflussfaktoren aufweisen, sind PLSR-Modelle entsprechend besser geeignet als Regressionsmodelle.

Das vierte Ziel, die Parameterschätzung, wird sowohl durch die Modellspezifikation als auch durch die Variablenauswahl beeinflusst. Für die fachspezifische Interpretation ist das

Vorzeichen und die Höhe der Regressionskoeffizienten bedeutsam. Die Parameterschätzung wurde in den Regressionsmodellen wesentlich stärker durch die Variablenauswahl beeinflusst als die PLSR-Modelle. Sobald erste Multikollinearitäten (lineare Abhängigkeiten zwischen den Regressoren) in den Regressionsmodellen auftreten, ist die Parameterschätzung, insbesondere die Vorzeichen der Regressionskoeffizienten, irreführend. Nach Lobenhoffer (1990) können Vorhersagen trotz Multikollinearität brauchbar genau bleiben, solange man im Datenraum x bleibt, d.h. solange der Prozess stationär verläuft. Die Stationäritätsannahme kann jedoch, wie in dieser Arbeit festgestellt wurde, nicht für alle Variablen unterstellt werden. Die Schätzgenauigkeit der Regressionsmodelle nahm entsprechend in den vorliegenden Untersuchungen mit zunehmender Kollinearität ab. Die PLSR-Modelle vermeiden das Problem der Multikollinearität durch die Reduzierung der Datenmatrix zu wenigen, unkorrelierten, latenten Variablen. Eine Variablenauswahl (siehe Kapitel 10.2) muss in der Regressionsanalyse entsprechend wesentlich sorgfältiger durchgeführt werden als in der PLS-Analyse.

Ein Vorteil der PLSR-Modelle kann darin gesehen werden, dass sie hinsichtlich der X-Variablen weniger bzw. für industrielle Prozesse realere Annahmen definieren als die Regressionsmodelle. Tabelle 73 fasst die generellen Unterschiede in den Annahmen zusammen.

	Regressionsanalyse		PLS-Analyse
٠	Regressoren sind frei von extremer	٠	X-Variablen sind korreliert
	Kollinearität		
٠	Die Datenreihenlänge ist größer als die	٠	Die Datenreihenlänge kann kleiner sein,
	Anzahl der Regressoren		als die Anzahl an X-Variablen
•	X ist frei von Fehlern in den Variablen	•	X kann fehlerbehaftet sein

Tabelle 73: Unterschiede in den Modellannahmen

Die Annahme, dass die Datenreihenlänge (T) in der Regressionsanalyse größer sein muss als die Anzahl der Regressoren (K) ist konstitutiv. Es müssen genügend Freiheitsgrade FG=T-K zur Verfügung stehen. Eine statistisch fundierte Regressionsanalyse ist nur möglich, wenn FG \geq 1 ist. Wie groß der Freiheitsgrad sein soll, ist im einzelnen nicht bekannt (Gruber 1994). Nach Gruber (1994) ist es jedoch ideal, möglichst viele, mindestens aber 15 bis 20 Freiheitsgrade, zu haben. Je kleiner der Freiheitsgrad, desto größer ist die Gefahr der reinen Anpassung des Modells an die Prozessdaten. Die Prognosefähigkeit dieser Modelle ist unzureichend, da das Modell sämtliche Fehler mit anpasst und keinen generellen Trend der Daten modelliert. Das Modell ist "overfitted". Schweitzer (1992), Léon-Méndez und Thömen (1996), Janssen (1997) und Kruse (1997) verwenden zur Bestimmung der Regressorenanzahl eine von Mager (1982) aufgestellte Faustformel. Danach soll die Anzahl der Regressoren maximal ¼ des Stichprobenumfanges betragen. Der notwendige Freiheitsgrad einer Regressionsanalyse ist jedoch allgemein nicht bestimmbar und sowohl von der Art des zu modellierenden Prozesses als auch von der Zielstellung abhängig.

Der Einfluss des Freiheitsgrades der Regressionsmodelle auf die ex-ante-Vorhersagegenauigkeit wurde in der vorliegenden Arbeit untersucht. Die Ergebnisse sind für die einzelnen Platteigenschaften in Abhängigkeit des Variablenauswahlverfahrens unterschiedlich. Insgesamt konnten die Eigenschaften Querzugfestigkeit und Dickenquellung, mit mehr Variablen (ca. 30) in den Gleichungen, genauer im Beobachtungsraum angepasst werden, ohne im Vorhersageraum an Schätzgenauigkeit zu verlieren. Die Bestimmtheitsmaße dieser Eigenschaften waren entsprechend höher als bei der Abhebefestigkeit und der mittleren Dichte, die ihre maximalen Schätzgenauigkeiten mit ca. 15-20 Variablen zeigten. Bei 132 Beobachtungen des Lerndatensatzes betrug das Verhältnis zwischen Regressoren und Beobachtungen entsprechend ca. 1/5 bis 1/6.

In der PLS-Analyse kann die Datenreihenlänge dagegen kleiner sein, als die Anzahl an X-Variablen, da zur Berechnung der Modellparameter nur die Anzahl latenter Variablen herangezogen wird. Nach Burnham et al. (1999) sind bei kontinuierlichen Prozessen mit mehreren hundert Variablen selten mehr als 7 bis 8 latente Variablen zu finden. In der PLS-Analyse bezieht sich der notwendige Freiheitsgrad auf die Anzahl der latenten Variablen und nicht auf die Anzahl an Prozessvariablen. Die Prozessvariablen werden als Realisationen der latenten Variablen aufgefasst. Als notwendige Bedingung zur Berechung der Modellparameter muss die Anzahl an Beobachtungen größer sein, als die Anzahl an latenten Variablen. Wie hoch der Freiheitsgrad im einzelnen sein muss, ist ebenfalls nicht bekannt. Jedoch gilt auch in der PLS-Analyse, je höher der Freiheitsgrad, desto besser.

Die Annahme der Regressionsanalyse, dass die X-Variablen frei von Fehlern sind, ist in der industriellen Prozessmodellierung unrealistisch. Diese Annahme hat u.a. zur Folge, dass die Modelle nicht anhand der Beobachtungsdaten bzw. daraus abgeleiteten Kennzahlen validiert werden können. Es wurden Regressionsmodelle entwickelt, die bei der Parameterschätzung Messfehler berücksichtigen (Gleser, 1981, Schneeweiss und Mittag, 1987). Dabei wird jedoch vorausgesetzt, dass die Verteilung der einzelnen Messfehler bekannt bzw. messbar ist. Für die industrielle Prozessmodellierung ist dieser Modelltyp nicht geeignet, da keine vollständigen Informationen über die Messerfehlerverteilungen vorliegen.

Die PLS-Modelle modellieren sowohl den Y-Raum als auch den X-Raum. Die Residuen des X-Modells umfassen den Teil von Messfehlern, der nicht zur Modellierung der Zieleigenschaften genutzt wird. Die X-Residuen sind für die Indentifizierung von Ausreißern in den Prozessdaten nützlich, z.B. von Abweichungen von den "normalen" Prozessbedingungen (Wold et al., 2001).

10.2 Variablenauswahl

Eine Variablenauswahl kann je nach Modelltyp zwingend notwendig sein, um die mathematische Berechnung der Modellparameter sicherzustellen bzw. interpretierfähige Ergebnisse zu erhalten oder kann als Werkzeug verstanden werden, um die Vorhersagegenauigkeit der Modelle zu erhöhen. Regressionsmodelle stellen zwei Anforderungen, die bei einer industriellen Prozessmodellierung i.d.R. zu einer zwingenden Variablenauswahl führen. So muss die Anzahl der Variablen kleiner sein als die Anzahl der Beobachtungen (Annahme 5) und stark korrelierte Regressoren können nicht gemeinsam in das Modell aufgenommen werden (Annahme 4). Während die Annahme 5 explizit erfüllt sein muss, um eine Parameterschätzung durchführen zu können, beinflussen korrelierte Variablen die Parameterschätzung nachteilig. Erst bei extremer Kollinearität der Daten ist eine Parameterschätzung nicht mehr möglich. Für die prinzipielle Anwendung der PLS-Modelle muss dagegen grundsätzlich keine Variablenauswahl erfolgen. Eine Variablenauswahl in der PLS-Analyse kann jedoch dazu beitragen, die Schätzgenauigkeit der Modelle zu verbessern, indem für die Vorhersage nicht relevante Variablen entfernt werden.

In heutigen modernen kontinuierlichen Anlagen zur Herstellung von Holzwerkstoffen sind bis zu 2500 Prozessvariablen (Young, et al. 2004) abrufbar. Eine Variablenauswahl zur Aufstellung der Modellgleichungen wird generell durch zwei entgegengesetzte Kriterien beeinflusst (Draper und Smith, 1998):

- 1. Um möglichst genaue Vorhersagen zu erhalten, müssen möglichst viele (bedeutende) Variablen im Modell aufgenommen werden, um den Biasfehler der Vorhersage so gering wie möglich zu halten.
- 2. Um die Vorhersagevarianz so gering wie möglich zu halten, die proportional zur Anzahl der Parameter im Modell und umgekehrt proportional zur Anzahl an Beobachtungen steigt, sollten so wenig wie möglich Variablen im Modell enthalten sein.

Das zweite Kriterium legt nahe, nur Variablen in das Modell aufzunehmen, die einen signifikanten Einfluss auf die Zieleigenschaften ausüben. Alle weiteren sogenannten Leervariablen erhöhen den Vorhersagefehler und damit die Vorhersagevarianz. Die Auswahl der signifikanten Variablen beruht jedoch auf den Lerndatenraum und ist insbesondere bei instationären Prozessdaten nicht auf den Vorhersageraum übertragbar. Aus diesem Grund müssen alle Variablen ins Modell aufgenommen werden, die zukünftig einen bedeutenden Einfluss auf die Platteneigenschaften ausüben könnten. Da diese nicht bekannt sind, sollten, unabhängig von der Signifikanz, technologisch bedeutsame (und bekannte) Variablen ins Modell aufgenommen werden. Dieses Vorgehen erhöht durchschnittlich den aktuellen Vorhersagefehler, verringert jedoch den zukünftigen Vorhersagefehler, wenn die zusätzlich aufgenommenen Variablen einen signifikanten Einfluss auf die Zielvariablen ausüben, jedoch nicht im Modell enthalten sind, führen zu einem systematischen Vorhersagefehler (Bias) der Modelle.

Die Schätzgenauigkeiten der Regressionsmodelle und der PLSR-Modelle werden gleichermaßen von den genannten Kriterien beeinflusst. Zu diesen Überlegungen muss die Variablenauswahl zur Regressionsanalyse jedoch zusätzlich die Annahmen 4 und 5 (siehe oben) berücksichtigen. Die Variablenauswahl zur Erstellung der Regressionnsmodelle wird mit zunehmender Variablenanzahl und Korrelation zwischen den Regressoren erschwert. Klassische Variablenauswahlverfahren, die auf eine Minimierung des ex-post-Vorhersagefehlers basieren, führen bei kollinearen Datenmatrizen zu ungenügenden ex-ante-Vorhersagegenauigkeiten (Hasener, 2003). Die Ursache hierfür liegt u.a. in der Auswahl von Variablen, die Scheinsignifikanzen aufweisen. Scheinsignifikanzen können entstehen, wenn zwei oder mehrere stark korrelierte Variablen im Modell enthalten sind. Eine Berücksichtigung der Korrelationsstruktur bei der Variablenauswahl nach der Methode der optimalen Eingänge verhindert, die durch Multikollinearität verursachten Probleme und führt zu höheren Schätzgenauigkeiten im Vorhersageraum.

Nach Box (1966) und Draper und Smith (1998) ist jede automatische Variablenauswahl zur Erstellung von Regressionsmodellen bei ungeplanten Versuchen (wie z.B. kontinuierliche Industrieprozesse) mit einigen Gefahren verbunden. Die Unwägbarkeiten der automatischen

Variablenauswahl können bei ungeplanten Versuchen zu ungenügenden Vorhersagegenauigkeiten bzw. zu falschen Schlussfolgerungen führen. Draper und Smith (1998) nennen dafür drei verschiedene Ursachen. Zum einen kann der Fehler des Modells nicht zufällig sein (wie angenommen), sondern kann durch einen gemeinsamen Effekt verschiedener Variablen beeinflusst werden, die nicht im Modell enthalten bzw. gar nicht messbar sind. Box (1966) nennt sie versteckte (lurking) bzw. latente Variablen. Die nicht messbaren Variablen können einen systematischen Fehler (Bias) in der Schätzung verursachen. Da die latenten bzw. versteckten Variablen nicht messbar sind, können Veränderungen nicht erkannt und nicht in den Modellen berücksichtigt werden. Dies kann, in Abhängigkeit der Stärke des Einflusses, zu unzuverlässigen Vorhersagen und zu falschen Schlussfolgerungen in den Effekten führen.

In den betrachteten Modellen umfassen die nicht berücksichtigten Effekte z.B. Änderungen in der Holzartenzusammensetzung und daraus folgenden Veränderungen der Faserqualität. Verschiedene indirekte Variablen, wie z.B. die Änderungen in den Stromstärken der Beschickschnecke im Refiner oder des Bandantriebes der Vorpresse, die u.a. durch die Holzart bzw. Holzartenzusammensetzung beeinflusst werden, waren für die Vorhersage bedeutsam. Eine sichere Interpretation dieser Variablen war jedoch nicht möglich, da diese auch durch prozessbedingte Variablen beeinflusst werden. Sie lassen jedoch annehmen, dass die Schwankungen in der Holzartenzusammensetzung zur Erklärung von Änderungen in den Platteneigenschaften bedeutsam sind. Weitherhin konnten Schwankungen der Leimqualität (Viskosität, pH-Wert u.a.) und der Beleimungsqualität (Leimbedeckung der Fasern, Tröpfchgröße u.a.) nicht in den Modellen berücksichtigt werden.

Ein weiterers Manko in der automatischen Variablenauswahl von ungeplanten Versuchen liegt in der Tatsache, dass die meisten technologisch bedeutsamen Variablen in sehr engen Bereichen ausgesteuert werden, um die Zieleigenschaften in den spezifizierten Gütegrenzen zu halten. Die geringen Schwankungen in diesen Variablen führen oftmals zu "nicht-signifikanten" Regressionskoeffizienten. Eine automatische Variablenauswahl würde diese Variablen nicht in das Modell aufnehmen. Bei zukünftigen Anpassungen bzw. Optimierungen in der Fertigung erhöhen sich jedoch oftmals die Schwankungsbreiten der technologisch bedeutsamen Variablen (Pressengeschwindigkeit, Leimmenge) und weisen in Folge dessen auch statistisch signifikante Effekte auf. Sind diese Variablen nicht im Modell enthalten, führt dies erneut zu systematischen Vorhersagefehlern (Bias).

Ein drittes Problem der automatischen Variablenauswahl von ungeplanten Versuchen wird durch die Korrelationen zwischen den Prozessvariablen verursacht. Korrelationen zwischen den Prozessvariablen ergeben sich in industriellen Prozessen aus maschinentechnischen Gründen (z.B. Heispresse) bzw. aus sogenannten Handlungsanweisungen (z.B. wenn Mattenfeuchte zu hoch, dann reduziere die Pressengeschwindigkeit). Diese Korrelationen erlauben es in Regressionsmodellen nicht, festzustellen, ob eine Änderung in Y mit X₁ oder X₂ oder mit beiden Variablen assoziiert ist. Eine eindeutige Zurückverfolgung von Effekten einer <u>einzelnen</u> Variable auf die Zieleigenschaften ist nur durch geplante Versuche, in denen die zu untersuchenden Effekte unabhängig voneinander eingestellt werden, möglich. Die PLS-Analyse erlaubt es dagegen, den gemeinsamen Einfluss korrelierender Variablen auf die Zieleigenschaften.

Die ersten beiden aufgeführten Problemfelder bei ungeplanten Versuchen sind auch für die PLSR-Modelle zutreffend. Jedoch ist die Gefahr eines systematischen Fehlers, durch nicht in das Modell aufgenommene, aber messbare Variablen, geringer als bei den Regressionsmodellen. Die Ursache liegt in der prinzipiellen Möglichkeit, alle Prozessvariablen in den PLSR-Modellen zu berücksichtigen. Versteckte, nicht messbare und kleine Schwankunsbreiten von bedeutenden technologischen Variablen sind für die Schätzung und Interpretation der Modellparamter für beide Modelltypen gleich problematisch.

10.3 Vorhersagegenauigkeit

Der Vorhersagefehler bzw. die Vorhersagegenauigkeit eines Prozessmodells kann durch statistische Kennzahlen aus dem Beobachtungsraum nicht abgeleitet werden (Léon-Méndez und Thömen, 1996; Janssen, 1997). Nach Myers (1986) beschreiben die Kennzahlen aus dem Beobachtungsraum (Bestimmtheitsmaß, Fehlervarianz) die Fähigkeit zur Anpassung der aktuellen Daten durch das Modell, lassen aber keine Aussage über die Vorhersagefähigkeit des Modells zu. Ein weiterer Grund, dass die Vorhersagegenauigkeit nicht aus den Beobachtungsdaten abgeleitet werden kann, liegt in den Daten selbst. Sie sind fehlerbehaftet. Die Fehler in den Prozessdaten setzen sich aus zufälligen und systematischen Messungenauigkeiten und durch Fehler in der zeitlichen Zuordnung zu den Platteneigenschaften zusammen. Die Laborwerte (Platteneigenschaften) beziehen neben den zufälligen und systematischen Messfehler auch Fehler (Schwankungen) über die Plattenbreite mit ein. Je größer die Schwankungen über die Plattenbreite, desto ungenauer werden die Vorhersagen, da diese sich auf den Mittelwert einer Platte beziehen. In Voruntersuchungen konnte gezeigt werden, dass die Abhebefestigkeit, gefolgt von der Querzugfestigkeit und der Dickenquellung, die größten Schwankungen über die Plattenbreite aufwiesen. Die Schwankungen der Rohdichte über die Plattenbreite waren am geringsten (siehe Tabelle 12). Aus den Kennzahlen des Beobachtungsraumes kann nicht beurteilt werden, ab welchem Grad der Anpassung zunehmend auch die Fehler in den Platteneigenschaften angepasst werden. Die Untersuchungen zeigten, dass eine zunehmende Überanpassung der Lerndaten einen starken Anstieg des ex-ante-Vorhersagefehlers zur Folge hatte. Der Bereich, ab dem eine Überanpassung der Lerndaten (overfitting) beginnt, kann jedoch nur durch einen externen Testdatensatz ermittelt werden.

Die PLSR-Modelle erreichten ungeachtet der Berücksichtigung aller Prozessvariablen keine höheren ex-ante-Vorhersagegenauigkeiten als die Regressionsmodelle mit ca. 20-30 Regressoren. Der Grund dafür liegt in der Tatsache, dass tatsächlich nur wenige Prozessvariablen unabhängig voneinander die Platteneigenschaften beeinflussten. Insofern diese Prozessvariablen in den Modellen berücksichtigt sind, kann die ex-ante-Vorhersagegenauigkeit durch Hinzunahme weiterer Variablen nicht erhöht werden. Eine Reduzierung der Variablenanzahl in den PLSR-Modellen führte entsprechend auch zu höheren Schätzgenauigkeiten im Vorhersageraum. Die Möglichkeit, den gemeinsamen Einfluss korrelierter Variablen in PLSR-Modellen zu berechnen, ist damit für die technologische Interpretation der Zusammenhänge zwischen Prozessvariablen und den Platteneigenschaften und weniger für die Vorhersagegenauigkeit bedeutsam. Tabelle 75 und Tabelle 76 fasst die typische Spannweite der absoluten und realtiven exante-Vorhersagegenauigkeiten der Regressions- und der PLSR-Modelle in Abhängigkeit des validerten Zeitraumes (3 und 12 Monate) zusammen. Zum Vergleich fasst Tabelle 74 die Standardabweichungen (STD) und Variationskoeffizienten (CV) der einzelnen Platteneigenschaften in den verschiedenen Datenräumen zusammen. Angegeben sind die Kennzahlen für den Lerndatenraum und für den 3 und 12 Monate umfassenden Testdatenraum.

Tabelle 74: Standardabweichungen und Variationskoeffizienten der Platteneigenschaften in den einzelnen Datenräumen

	Ν	LAB_IB	LAB_TS	Lab_SS	Lab_MD
Einheit STD		[N/mm²]	[%]	[N/mm²]	[kg/m³]*
STD Lerndaten (9 Monate)	132	0,14	0,95	0,23	10,2
CV Lerndaten (9 Monate)	132	7,4	10,9	10,2	1,1
STD Testdaten (3 Monate)	34	0,13	0,85	0,20	9,1
CV Testdaten (3 Monate)	34	6,7	9,2	8,3	1,0
STD Testdaten (12 Monate)	116	0,14	0,89	0,19	9,8
CV Testdaten (12 Monate)	116	7,6	9,3	7,9	1,1

* im Ergebnisteil ist die Einheit [g/cm³], Abweichungen beruhen auf Rundungsfehler

Tabelle 75: Vergleich der e	ex-ante-Vorhersagegenauigkeiten	über 3 Monate
-----------------------------	---------------------------------	---------------

Modellart	Querzug	festigkeit	Dickenquellung		Abhebefestigkeit		mittlere Rohdichte	
	MAPE _P	RMSE _P	MAPE _P	RMSE _P	MAPE _P	RMSE _P	MAPE _P	RMSE _P
	[%]	[N/mm ²]	[%]	[%]	[%]	[N/mm²]	[%]	[kg/m³]
mult. Reg.	3,9-4,5	0,10-0,11	5,9-6,9	0,69-0,79	6,4-6,9	0,18-0,20	0,7-0,8	8,3-8,8
mult. PLS	4,3-4,9	0,10-0,12	6,3-6,6	0,68-0,74	6,3-6,6	0,18-0,19	0,9	8,5-9,1
single PLS	3,8-4,8	0,09-0,11	6,0-6,8	0,63-0,73	6,5-8,0	0,19-0,22	0,8-0,9	8,3-10,3

Tabelle 76: Vergleich der	ex-ante-Vorhersagegenauigkeiten über 12 Monate

Modellart	Querzug	festigkeit	Dickenquellung		Abhebefestigkeit		mittlere Rohdichte	
	MAPE _P	RMSE _P	MAPE _P	RMSE _P	MAPE _P	RMSE _P	MAPE _P	RMSE _P
	[%]	[N/mm²]	[%]	[%]	[%]	[N/mm²]	[%]	[kg/m³]
mult. Reg.	4,7-5,3	0,10-0,12	6,8-7,4	0,82-0,90	6,0-6,1	0,18	0,8	9,2-9,3
mult. PLS	5,1-5,2	0,11-0,12	7,0-7,5	0,86-0,90	6,3-6,4	0,19	0,8	8,8-8,9
single PLS	4,9	0,11	6,9-7,1	0,83-0,85	6,8-7,0	0,19-0,20	0,8	8,8

Zur Abschätzung der Entwicklung der Vorhersagegenauigkeiten sind die Änderungen in den Schwankungen der Platteneigenschaften mit fortlaufender Vorhersage zu berücksichtigen. Der prozentuale Vorhersagefehler (MAPE_P) der Querzugfestigkeit und der Dickenquellung erhöhte sich mit zunehmender Einsatzzeit der Modelle (Vergleich zwischen 3 und 12 Monate). Der prozentuale Vorhersagefehler der Abhebefestigkeit und der mittleren Rohdichte verbesserte sich dagegen geringfügig bzw. blieb weitgehend unverändert. Werden jedoch die Schätzgenauigkeiten auf die veränderten empirischen Standardabweichungen in den Testdatenräumen bezogen, verbesserte sich die Schätzgenauigkeit der Querzugfestigkeit und der mittleren Rohdichte über die Zeit geringfügig. Die auf gleicher Weise normierten Schätzgenauigkeiten der Dickenquellung und der Abhebefestigkeit nahmen dagegen ab. Die Dickenquellung wurde von allen untersuchten Eigenschaften am

stärksten von Effekten beeinflusst, die nicht im Modell berücksichtigt worden sind. Als mögliche nicht berücksichtigte Einflussfaktoren sind Änderungen in der Holzartenzusammensetzung und ein Einbau eines zusätzlichen Multipots am Ende der Heißpresse zu nennen. Dieser nachträgliche Einbau erfolgte in der Zeit des zweiten Testdatensatzes.

10.4 Industrielle Relevanz und Bewertung der Ergebnisse

Die Qualitätskontrolle in der Holzwerkstoffindustrie basiert auf einer Kontrolle von Stichproben, die der laufenden Produktion entnommen werden. Dabei hat der Hersteller die Platteneigenschaften, die nach den entsprechenden EN-Anforderungen gefordert werden, für jede Fertigungslinie zu kontrollieren (EN 326-2). Stichproben sind dabei von jedem Produkttyp und jedem Dickenbereich zu entnehmen. Der genaue Umfang der Stichproben ist jeweils in den Gütenormen festgelegt. Die mechanisch-technologischen Eigenschaften sind i.d.R. im Abstand von 8 Stunden zu überprüfen. In der werkseigenen Produktionskontrolle (Eigenüberwachung) sind die Plattenmittelwerte jeder Einzelplatte, die Standardabweichung innerhalb der Platten, das gleitende Gesamtmittel der letzten 30 Platten-Mittelwerte und die gleitende Standardabweichung zwischen den letzten 30 Platten-Mittelwerte zu berechnen und unter Verwendung von Kontrollkarten aufzuzeichnen. Neben der Eigenüberwachung erfolgt halbjährlich eine Fremdüberwachung, die durch ein neutrales Prüfinstitut durchgeführt wird.

Die gesamte Qualitätskontrolle in der Holzwerkstoffindustrie stützt sich auf die Überwachung der in den Anforderungsnormen aufgeführten Platteneigenschaften. Die Prüfung der genormten mechanisch-technologischen Platteneigenschaften ist heute noch nicht zerstörungsfrei möglich. Der Nachteil der auf Stichproben basierenden Qualitätskontrolle ist die stark zeitlich verzögerte Feststellung von signifikanten Abweichungen von den Güteanforderungsgrenzen ("Außer-Kontrolle-Geraten") und die daraus entstehenden Folgekosten. Das Erkennen von signifikanten Abweichungen vom gleitenden Gesamtmittelwert durch Verwendung von Kontrollkarten ist je nach Kontrollkartentyp unterschiedlich, benötigt jedoch mehrere Stichprobenmittelwerte (Mittag, 1993, Noack 2003). Um eine Unterschreitung der Güteanforderung durch den gleitenden Mittelwert zu verhindern, wird industriell mit hohen Sicherheitsbeiwerten gearbeitet.

Eine auf die Prozessvariablen gestützte Echtzeitqualitätskontrolle dagegen ermöglicht es, ohne Zeitverluste Trends in den Prozessvariablen als auch in den Platteneigenschaften zu erkennen und entsprechende Gegenmaßnahmen einzuleiten. Eine sinnvolle Gütekontrolle muss demnach nicht nur die Qualitätskontrolle der Platteneigenschaften, sondern auch die Kontrolle der Prozessvariablen beinhalten. Da ein Hauptteil der Streuungen in den Platteneigenschaften durch Schwankungen in den Prozessvariablen verursacht wird, ist es notwendig und ratsam, die Prozesskontrolle in die Gütekontrolle (SPC) unterstellt werden, dass die Prozessdaten (X) von größerer Bedeutung sind als die Qualitätsdaten (Y). Die Prozessvariablen können im Gegensatz zu den Qualitätseigenschaften ohne zusätzliche Kosten mit einer sehr hohen zeitlichen Auflösung überwacht werden. Die multivariate statistische Prozesskontrolle ermöglicht es, innerhalb einer zusammenfassenden online-Kontrollkarte (Hotteling T²) den Prozess zu überwachen.

Durch die Speicherung der Prozess- und Qualitätsdaten in einer Datenbank ist es auch zukünftig (z.B. bei späteren Kundenreklamationen) möglich, den Nachweis einer norm- oder kundengerechten Herstellung der Holzwerkstoffplatten zu führen. Die Eigenschaften markierter Liefermengen sind rückführbar und statistisch gesichert. Es ist bei Haftungsfragen ein entscheidender Vorteil, die normgerechte Produktion aus den **Prozessdaten** nachweisen zu können. Das Verfahren der validierten Echtzeitkontrolle ist zur Qualitätskontrolle nach ISO 9001:2000 geeignet.

Sowohl die Methode der multivariaten Regressionsanalyse als auch die PLS-Analyse erlauben es, die Platteneigenschaften in Echtzeit, bei vergleichbaren Schätzgenauigkeiten, vorherzusagen. Die erreichbaren Schätzgenauigkeiten im Vorhersageraum sind dabei von verschiedenen Parametern, wie die Verfügbarkeit und Genauigkeit der Messtechnik, insbesondere aber von der Genauigkeit der Datenanalyse und -bereinigung und der Variablenauswahl abhängig.

Zur Beantwortung der Frage, welcher Modelltyp prozessangemessen ist, wird oftmals nur die quantitative Vorhersagegenauigkeit der Zieleigenschaften herangezogen. Nach Burmham, et al. (1999) sind jedoch auch andere Aspekte der Nutzungsmöglichkeit von Modellen zu berücksichtigen. Dazu zählen nach den Autoren z.B. die Analyse bzw. die Interpretation von Ereignissen in den Daten, die Prozessüberwachung oder die Fähigkeit der Modelle, fehlende Werte in den Daten zu handhaben. Auch zu beurteilen ist der notwendige Aufwand zur Erstellung der Modelle bzw. die Benutzerfreundlichkeit für industrielle Anwendungen.

In der vorliegenden Arbeit unterscheiden sich die beiden Modelltypen nicht signifikant in der langfristigen Vorhersagegenauigkeit. Wie bereits oben erwähnt, ist dies, insbesondere bei Regressionsmodellen, jedoch stark von der Vorgehensweise während den der Modellerstellung abhängig. Der Aufwand der Modellerstellung bzw. der Variablenauswahl ist bei der Regressionsanalyse höher einzuschätzen als bei den PLSR-Modellen. Grundsätzlich können in den PLSR-Modellen alle Prozessvariablen, ohne Berücksichtigung der Korrelationen zwischen den Variablen, einbezogen werden. In den Regressionsmodellen dagegen ist die Entfernung aller korrelierten Prozessvariablen notwendig, um technologisch interpretierbare Modelle zu erhalten. Eine weitgehend automatische Auswahl der Prozessvariablen für die Regressionsmodelle auf Grundlage der Korrelations-Bewertungsmatrix kann jedoch den zeitlichen Aufwand zur Modellerstellung beschränken und stellt unter Berücksichtigung weiterer technologischer Variablen eine maximale Schätzgenauigkeit im Vorhersageraum sicher. Dies konnte in zwei weiteren Anlagen zur Herstellung von Holzwerkstoffen bestätigt werden.

Aufgrund der notwendigen Reduzierung der Variablen in den Regressionsmodellen, können technologische Zusammenhänge zwischen den Prozessvariablen und den Zielvariablen nur beschränkt an ausgewählten nicht miteinander korrelierenden Variablen aufgezeigt werden. Anhand der PLSR-Modelle kann der gemeinsame Einfluss korrelierender Variablen auf die Platteneigenschaften nachgewiesen werden. Dies ist als einer der entscheidenden Vorteile der PLSR-Modelle gegenüber den Regressionsmodellen zu nennen.

Ein weiterer Vorteil besteht darin, dass die PLSR-Modelle Fehler in den Prozessvariablen zulassen, da auch der X-Prozess modelliert wird und Messungenauigkeiten bzw. Fehler

aufgrund von ungenauen zeitlichen Zuordnungen zu den Plattenmerkmalen als Residuen des X-Modells nicht auf den Y-Prozess übertragen werden. Bei Regressionsmodellen wird dagegen angenommen, dass die Regressoren fehlerfrei sind. Höhere Schätzgenauigkeiten der PLSR-Modelle konnten, trotz Berücksichtigung aller Prozessvariablen, dennoch nicht erreicht werden. Die Ursache liegt darin, dass tatsächlich nur sehr wenige Variablen einen bedeutenden und unabhängigen Einfluss auf die Platteneigenschaften ausübten. Werden alle vorhandenen unabhängigen Einflüsse in den Regressionsmodellen berücksichtigt, so müssen deren Vorhersagegenauigkeiten mit denen der PLSR-Modelle übereinstimmen.

Die PLSR-Modelle werden mit der Integration weiterer, insbesondere hochauflösender Messtechnik in die Holzwerkstoffproduktion und der damit verbundenen Erhöhung von korrelierten Messdaten den Regressionsmodellen in bezug auf Erstellung, Handhabung und in den Interpretationsmöglichkeiten zunehmend überlegen sein. Auch können die PLSR-Modelle durch die Berücksichtigung der Korrelationsstruktur der Prozessvariablen und deren Reduzierung auf wenige latente Variablen eine umfassendere Prozesskontrolle gewährleisten als die Regressionsmodelle.

Industrielle Anwendungen der statistischen Prozesskontrolle, insbesondere in der chemischen Industrie, durch multivariate Projektionsmethoden sind in Kresta et al. (1991), Wold (1993), Nomikos und MacGregor (1995), Kourti und McGregor (1995), MacGregor (1997), Burnham et al. (1999) und Wold et al. (2001) beschrieben. Erste industrielle Anwendungen der multivariaten Projektionsmethoden in Verbindung mit der NIR-Spektros-kopie in der MDF- bzw. Spanplattenindustrie zur Überwachung von spezifischen Material-eigenschaften (Feuchte, Partikelgröße, Holzqualität) als auch zur Vorhersage der Platten-eigenschaften sind in Engström et al. (1998), Johnsson et al. (2000) und Carlsson et al. (2002) aufgezeigt.

10.5 Weiterer Forschungs- und Entwicklungsbedarf

Ein vorrangiges Ziel sollte die Entwicklung von Messtechnik sein, die sowohl direkte Aussagen über die Materialeigenschaften als auch der Plattengualitäten zulassen. Es ist unbestritten, dass die (zerstörenden) Standardprüfungen, wie die Querzugfestigkeit, die Dickenquellung, die Biegefestigkeit oder die Abhebefestigkeit nur einen Teil des Eigenschaftsprofils der Werkstoffe beschreiben. Zur vollständigeren Beschreibung und einer sinnvollen Optimierung der Plattengualitäten müssten z.B. Kennzahlen zur Langzeitstabilität, Witterungsbeständigkeit oder des Emissionspotenzials in die Modelle mit aufgenommen werden. Dazu müssen jedoch sowohl umfassende Kennzahlen als auch Messmethoden entwickelt werden, die auch in der Praxis bezahlbar und innerhalb eines angemessenen Zeitraums durchführbar sind. Eine Optimierung des Prozesses kann nur unter Einbeziehung aller Eigenschaften erfolgen. Als Beispiel kann der Einsatz von Harnstoff genannt werden. Dieser wird als Formaldehydfänger eingesetzt, um die Anforderungen zur Formaldehydabgabe einzuhalten. Für die mechanisch-technologischen Eigenschaften ist der Einsatz von Harnstoff nachteilig. Wenn die Formaldehydabgabe nicht im Prozessmodell enthalten ist, würde jede Optimierungsrechnung den Harnstoffeinsatz minimieren. Erst mit Berücksichtigung der Formaldehydabgabe kann die optimale Harnstoffmenge berechnet werden.

Die Entwicklung von zerstörungsfreien und online einsetzbaren Messmethoden zur Beschreibung der Platteneigenschaften würde es neben der Kosteneinsparung erlauben, die

Vorhersagen der Modelle ohne Zeitverzug validieren zu können. Eine Optimierung des Prozesses kann damit vereinfacht, erweitert und beschleunigt werden.

Auf der Prozessseite werden in der Mehrheit Anlagen- und Umweltparameter (Druck, Distanz, Mengen, Drehzahlen, Temperaturen usw.) überwacht. Direkte von Holz und Klebstoff abhängige Parameter, die die Schwankungen der Holz- und Leimqualitäten beschreiben, sind i.d.R. online nicht vorhanden. Die NIR-Spektroskopie bietet hier eine Möglichkeit, diese zu erfassen. Der Einfluss dieser Messsignale, durch zusätzliche Integration in die Modelle, auf die Vorhersagegenauigkeiten sollte wissenschaftlich untersucht werden. Gleiches gilt für die online zur Verfügung stehenden Ultraschallsignale bzw. Rohdichteprofilkennzahlen.

11 Zusammenfassung

In der industriellen Fertigung von Holzwerkstoffen werden vom Rohstoff bis zum Endprodukt zahlreiche Teilprozesse durchlaufen. Die Prozesskette umfasst: Holzaufbereitung, Trocknung und Sichtung, Beleimung, Streuung, Vorverdichtung, Heißpresse und End-fertigung. Die Qualitätskontrolle basiert auf einer Kontrolle von Proben, die der laufenden Produktion entnommen werden und im Labor analysiert werden. Die Informationen der Qualitätskontrolle stehen damit jedoch erst zeitlich stark verzögert für eine Regelung des Prozesses zur Verfügung. Eine systematische Verknüpfung der Informationen der einzelnen Teilprozesse, der Prozessparameter und der Plattenqualitäten zur statistischen Prozesskontrolle erfolgt bisher nicht.

Zusammenhänge zwischen den Prozessparametern und den Platteneigenschaften aus der Qualitätskontrolle können durch ein statistisches Prozessmodell beschrieben werden Lobenhoffer (1990). Besitzt das Modell eine ausreichende Genauigkeit, kann zukünftig aus den Daten des laufenden Prozesses die Produktqualität zuverlässig vorhergesagt werden. Dieses Verfahren ermöglicht eine Echtzeitqualitätskontrolle und eine Regelung des Gesamtprozesses.

Weiterhin kann das Modell dazu verwendet werden, den Prozess innerhalb des vorliegenden Datenraums zu simulieren. Dazu werden Prozesseinstellungen variiert, um den Einfluss auf die Produktqualität zu bestimmen. Zur vollständigen Interpretation der Simulationen ist jedoch die Korrelationsstruktur der Prozessdaten zu berücksichtigen. Eine umfassende Prozessoptimierung erfordert i.d.R. Daten aus einem statistischen Versuchsplan, mit dem Ziel der Prozessoptimierung (Wold et al., 2001).

Die mathematischen Grundlagen zur regressionsanalytischen Prozessmodellierung wurden von Lobenhoffer (1990) beschrieben. Seit dem ist mit der Möglichkeit zur automatischen Prozessdatenerfassung durch Prozessleitsysteme sowohl die Prozessdatenanzahl als auch die Anzahl der aufgezeichneten Beobachtungen stark angestiegen. Ein weiterer technologischer Innovationsschritt ist mit der Umstellung von Mehretagenpressen auf kontinuierliche Pressen einher gegangen. Diese Entwicklungen und der zunehmende Einsatz von moderner Messtechnik führten dazu, dass heute bis zu 2500 Prozessvariablen (Young, 2004) vorliegen können. Diese Daten sind jedoch stark untereinander korreliert.

In dieser Arbeit werden die multivariate Regressionsanalyse und die Partial-Least-Square-Analyse (PLS) angewendet, um Modelle zur Echtzeitqualitätskontrolle von Holzwerkstoffen zu entwickeln. Da innerhalb von Regressionsmodellen jedoch keine stark korrelierten Variablen berücksichtigt werden können, lag ein Schwerpunkt dieser Arbeit in der Entwicklung und Beschreibung einer Prozedur zur Variablenauswahl, nach deren Anwendung eine erfolgreiche regressionsanalytische Prozessmodellierung erfolgen kann. Weiterhin sollte der Einfluss der Modellaktualisierung und der Einfluss endogener Regressoren auf die Schätzgenauigkeit im Vorhersageraum untersucht werden.

Ein weiterer Teil der Arbeit beschäftigt sich mit der Faktorenanalyse zur Beschreibung des Pressprogramms der Heißpresse durch eine geringe Anzahl unkorrelierter Faktoren. Der

Einfluss der extrahierten Pressenfaktoren auf die Platteneigenschaften wird innerhalb der Regressionsmodelle beschrieben.

Um den gemeinsamen Einfluss korrelierter Variablen auf die Platteneigenschaften und auf die Vorhersagegenauigkeit zu untersuchen, wurde die PLS-Analyse angewendet. Sowohl die Vorhersagegenauigkeiten als auch die technologischen Interpretationsmöglichkeiten der beiden Modelltypen sollten miteinander verglichen werden. An beiden Modelltypen sollte weiterhin die Langzeitstabilität bzw. die Entwicklung der Vorhersagegenauigkeit untersucht werden.

Zur Erstellung der Modelle wurde ein umfangreicher **Datensatz** aus einer Faserplattenproduktion mit kontinuierlicher Presse (8 mm HDF-Laminat-Trägerplatte) verwendet. Dieser Datensatz umfasst einen Produktionszeitraum von 9 Monaten und wird als Lerndatensatz bzw. als Beobachtungsraum bezeichnet. Um eine realistische Beurteilung der Modellgüte und der Vorhersagegenauigkeit zu erhalten, wird von León-Méndez und Thömen (1996) empfohlen, die Modelle anhand zusätzlich erhobener Datensätze zu überprüfen.

Die **Validierung** der in dieser Arbeit gebildeten Modelle erfolgte anhand eines Testdatensatzes (Vorhersageraum), der nicht zur Modellerstellung herangezogen wurde und sich zeitlich dem Lerndatensatz anschließt. Der erste Testdatensatz umfasste einen Zeitraum von 3 Monaten und wurde zur grundsätzlichen Überprüfung der Modelle verwendet. Ein zweiter, 9 Monate umfassender, Testdatensatz wurde zusätzlich herangezogen, um in Verbindung mit dem ersten Testdatensatz die langfristige Vorhersagestabilität der Modelle zu prüfen. Um die Vorhersagen im Lerndatensatz von denen im Testdatensatz zu unterscheiden, werden erstere als ex-post-Vorhersagen und letztere als ex-ante-Vorhersagen bezeichnet.

Zur Abschätzung der **Modellgüte** werden verschiedene Kennzahlen definiert. Die Kennzahlen werden sowohl für den Beobachtungsraum als auch für den Vorhersageraum berechnet. Von einem gesicherten Vorhersagemodell wird u.a. gefordert, dass der Vorhersagefehler im Vorhersageraum (RMSE_p) kleiner sein muss als die empirische Standardabweichung der Testdaten. Diese Anforderung stellt sicher, dass das Modell die systematischen Schwankungen, also Schwankungen die durch Prozesseinflüsse verursacht werden, berücksichtigt und deren Einfluss auf die Platteneigenschaften modelliert. Dabei wird unterstellt, dass die Schwankungen der Platteneigenschaften durch zufällige, jedoch hauptsächlich durch systematische Einflüsse aus dem Prozessdatenraum verursacht werden.

Vor der eigentlichen Modellerstellung wurden die Daten mittels einfacher Methoden der **Daten- und Prozessanalyse** untersucht. Analysen zur Stationärität, zur Schwankungsbreite, Verteilung und Korrelationsstruktur der Daten, zur Streuung der Varianzen innerhalb und zwischen den Platteneigenschaften, zum Pressenprogramm und zur Identifizierung von Hilfsvariablen (Dummy-Variablen) wurden angewendet, um erste Strukturen im Datensatz zu erkennen und fehlerbehaftete und nicht plausible Variablen bzw. Beobachtungen zu entfernen.

Nach Lobenhoffer (1990) ist die Stationärität eine strenge Voraussetzung für eine konsistente Schätzung der Zielvariablen y aus den Prozessvariablen x. Diese Bedingung schließt z.B. aus, dass der Produktionsprozess einen Trend bzw. einen zyklischen Verlauf aufweist. Mittag (1993) weist darauf hin, dass diese Annahme jedoch oftmals in der

industriellen Praxis verletzt wird. Im vorliegenden Datensatz der Faserplattenproduktion konnten sowohl in den Zielvariablen als auch in den Prozessvariablen Trends nachgewiesen werden. Insbesondere sind Trends bei Variablen vorzufinden, die natürlichen jahreszeitlichen Schwankungen unterliegen (Lufttemperatur bzw. -feuchte) bzw. von diesen beeinflusst werden und bei Variablen, die einer ständigen Optimierung bzw. Anpassung unterliegen (Beleimung, Pressengeschwindigkeit). Niveauänderungen in den Zieleigenschaften sind während des zweiten Testdatensatzes aufgetreten.

Eine Plausibilitätsüberprüfung der Daten erfolgt anhand von maschinenbaulich oder technologisch vorgegebenen Grenzwerten. Alle Werte, die außerhalb dieser Grenzen liegen, sind aus der weiteren Analyse ausgeschlossen worden. Ausreißer dagegen sollten nur dann aus dem Datensatz entfernt werden, wenn diese eindeutig auf Messfehler bzw. Übertragungsfehler zurückzuführen sind. Technologisch erklärbare Werte, die durch statistische Analysemethoden als "Ausreißer" erkannt werden, sollten im Datensatz verbleiben. Oftmals besitzen "Ausreißer" Informationen, die andere Daten nicht liefern können, da sie zum Teil durch ungewöhnliche Kombinationen und Umständen entstehen. Automatisierte Verfahren der Ausreißerentfernung können nicht empfohlen werden. Zweckmäßig ist die grafische Darstellung der Variablen über die Zeit (Line-plots) und eine optische Markierung von Werten, die z.B. außerhalb des der 4-fachen Standardabweichung oder außerhalb des Dengel-Intervalls (Dengel, 1996) liegen. Der Technologe bekommt damit die Möglichkeit, die Messtechnik zu überwachen und verdächtigen Werten nachzugehen.

Durch die Berechnung der Variationskoeffizienten und eine Reihung der Variablen nach deren Höhe lässt sich erkennen, welchen Schwankungen die Variablen unterworfen sind. Variationskoeffizienten kleiner 2% zeigten, aufgrund der isochoren Steuerung der Presse, insbesondere die Distanzen der Presse aber auch die Pressentemperaturen der einzelnen Heizkreise. Von den Platteneigenschaften wies die mittlere Rohdichte mit einem Variations-koeffizienten von ca. 1% eine äußerst geringe Schwankungsbreite auf. Variablen mit Variationskoeffizienten größer 20% traten in verschiedenen Bereichen des Herstellungsprozesses auf. Innerhalb dieser Gruppe waren Variablen zu identifizieren, die Messfehler beinhalteten.

Neben metrischen können auch kategoriale Variablen für die Modellierung bedeutsam sein. Diese Variablen besitzen eine bestimmte Anzahl an Ausprägungen, die nicht metrisch messbar sind. Im vorliegenden Datensatz wurde für die Plattenproduktion Harnstoff-Formaldehydharze (UF) verwendet, die von unterschiedlichen Herstellern bezogen wurden. Die Leimlieferanten A und B wurden in die Modelle als kategoriale Variable mit zwei Ausprägungen integriert. In den Voruntersuchungen konnte ein signifikanter Unterschied in den Platteneigenschaften, in Abhängigkeit des Leimlieferanten, festgestellt werden. Die kategorialen Variablen (z.B. Lieferanten, Holzarten, verantwortliche Schichtführer für die Pressensteuerung, Mattenbesprühung usw.) müssen vor Beginn der Datensammlung festgelegt werden. Problematisch ist in der Industrie jedoch oft die automatische bzw. konsequente Erfassung dieser Ausprägungen.

Die Variablenauswahl zur Erstellung der multivariaten **Regressionsmodelle** beeinflusst entscheidend die Vorhersagegenauigkeit und die Interpretation der Modellzusammenhänge. Es konnte gezeigt werden, dass die klassischen statistischen Methoden zur Variablenselektion (stepwise, backward u.a.) zu Modellen führen, die den Beobachtungsraum gut beschreiben, aber in der ex-ante-Vorhersage geringe Schätzgenauigkeiten aufweisen. Dieses Phänomen wird durch eine i.d.R. zu hohen Anpassung des Beobachtungsraumes versucht. Die Variablenauswahl dieser Verfahren berücksichtigen insbesondere nicht die Korrelationen zwischen den Prozessgrößen. Dies führt in Abhängigkeit der gewählten Signifikanzen zu sehr unterschiedlichen Auswahlergebnissen. Die Modelle sind sehr stark durch Multikollinearitäten geprägt und als Folge dessen technologisch nicht zu interpretieren.

Eine Modellauswahl nach statistischen Kriterien aus dem Beobachtungsraum, wie z.B. das Bestimmtheitsmaß, führte nicht zu stabilen Vorhersagemodellen. Diese Beurteilungskriterien beruhen u.a. auf der Annahme, dass die Regressoren frei von Fehlern in den Variablen sind. Davon darf sowohl in industriellen Prozessen als auch in der zerstörenden Materialprüfung nicht ausgegangen werden. Eine Beurteilung von Vorhersagemodellen muss daher an den ex-ante-Vorhersagen erfolgen. Dies hat jedoch den methodischen Nachteil, dass die Modellgüte erst nach erfolgten ex-ante-Vorhersagen beurteilt werden kann. Daher war ein Verfahren zu finden, das die Kriterien eines stabilen Vorhersagemodells implizit erfüllen kann. Die Methode der optimalen Eingänge (Bärmann, 2002) beruht auf einer Gewichtung der Korrelationen der Eingangsvariablen zu den Zielvariablen als auch auf den Korrelationen zu den schon aufgenommenen Eingangsvariablen. Durch die Berücksichtigung der Eingangskorrelationen können die durch Multikollinearität verursachten Probleme vermieden werden. Mit der Methode der optimalen Eingänge wurden Vorhersagemodelle erstellt, die die aufgestellten Kriterien zur Schätzgenauigkeit im Vorhersageraum erfüllten.

Diese Methodik der Modellerstellung wurde bisher in zwei weiteren Unternehmen der Holzwerkstoffindustrie angewendet. Dadurch konnte der durchschnittliche Fehler der Vorhersagen im Vergleich zu den vorher bestehenden Modellen stark reduziert werden.

Die Interpretation dieser Modelle erweist sich jedoch als sehr schwierig, da die Auswahl allein auf der Korrelationsmatrix beruht. Durch die zusätzliche Berücksichtigung technologisch relevanter Variablen konnte neben einer besseren Interpretation der Modelle vor allem die langfristige Stabilität bzw. Vorhersagegenauigkeit in der Dickenquellung und Querzugfestigkeit der Modelle verbessert werden. Dazu wurden Variablen, deren technologische Zusammenhänge bekannt sind, unabhängig von ihrer momentanen Schwankungsbreite bzw. Signifikanz in die Modelle aufgenommen. Bei der Integration dieser technologischen Variablen in die bestehenden Modelle nach der Methode der optimalen Eingänge sind erneut die Korrelationen zwischen den neu aufzunehmenden und im Modell vorhandenen Variablen zu berücksichtigen. Als Maß der Kollinearität wurde dabei der Varianzinflationsfaktor herangezogen. Es wurden nur die Variablen im Modell belassen, deren Varianzinflationsfaktoren einen Wert kleiner 5 aufwiesen.

Eine technologische Auswahl der Variablen der Heißpresse anhand des Pressprogramms war für die Modellerstellung nicht zielführend. Eine Berücksichtigung der Korrelationen zwischen den Pressenvariablen ist unbedingt notwendig. Eine technologische Auswahl (insbesondere bei einer Auswahl der Variablen durch Technologen vor Ort), führte zu starken Kollinearitäten in den Modellen. Höhere Schätzgenauigkeiten und keine Schwierigkeiten mit Kollinearitäten ergaben sich bei der Verwendung der Pressenvariablen nach der Methode der optimalen Eingänge.

Ob eine Variable einen statistisch signifikanten Einfluss auf die jeweilige Zielvariable ausübt, wird in den Regressionsmodellen anhand des t-Wertes (bzw. des p-Wertes) des Regressionskoeffizienten beurteilt. Eine technologische Interpretation von Modellzusammenhängen ist entsprechend nur bei signifikanten Variablen sinnvoll. Problematisch dabei ist jedoch, dass die Annahmen zur Berechnung der t-Werte verletzt sein können (z.B. durch

stochastische Variablen) und sich die Signifikanzen durch nichtstationäre Variablen im Laufe der ex-ante-Vorhersagen ändern. Aus diesen Gründen ist eine alleinige Auswahl der Variablen anhand des Signifikanzniveaus nicht erfolgversprechend.

Das durchschnittliche Verhältnis zwischen Regressoren und Datensätzen, lag bei den Modellen mit maximaler Vorhersagegenauigkeit bei ca. 1:5 bis 1:6. Vom ursprünglichen Datensatz mit 321 Variablen zeigten sich nur zwischen 20 und 30 Variablen für die Vorhersage der Platteneigenschaften als bedeutsam. Der Einfluss der meisten Variablen kann aus methodischen Gründen in den Regressionsmodellen jedoch nicht nachgewiesen werden, da sie untereinander zu hohe Korrelationen aufweisen.

Ein Abweichen des Modells im Vorhersagemodus mit gegebener Variablenauswahl von den tatsächlichen Gegebenheiten des Prozesses konnte anhand des Verlaufes der F-Statistik einer Modellgleichung nachvollzogen werden. Bei biasfreien ⁽¹⁾ Vorhersagen sollte als notwendige Bedingung der F-Wert mit zunehmendem Freiheitsgrad, also mit fortschreitender Vorhersage, ansteigen. Bei konstanten oder fallenden F-Werten über dem Vorhersageraum sollte eine neue Variablenauswahl auf Basis des fortgeschriebenen Datensatzes überprüft werden.

Bei Modellen, die bei einem ausreichenden Freiheitsgrad alle bedeutenden Prozessvariablen berücksichtigen, konnte durch die zusätzliche Berücksichtigung von **Interdependenzen** der endogenen Variablen keine Erhöhung der Schätzgenauigkeit erreicht werden. Mit der Querzugfestigkeit als endogenen Regressor lässt sich jedoch mit geringem Aufwand die Abhebefestigkeit abschätzen.

Die Schätzgenauigkeit der Modelle wird durch eine sofortige **Aktualisierung** der Parametermatrix, nach Vorliegen eines neuen Datensatzes, erhöht. Versuche zeigten aber, dass eine regelmäßige, nicht permanente Aktualisierung der Parameter, zu keinen wesentlichen Verlusten in der Schätzgenauigkeit führen muss. Dieses Ergebnis kann dann erwartet werden, wenn alle wichtigen Bestimmungsfaktoren im Modell enthalten und im Vorhersageraum keine gravierenden Änderungen im Prozess vorgefallen sind.

Zwei ausgewählte Regressionsmodelle wurden herangezogen, um die **langfristige Entwicklung der Schätzgenauigkeiten** im Vorhersageraum zu untersuchen. Zur Abschätzung der Entwicklung der Vorhersagegenauigkeiten sind die Änderungen in den Schwankungen der Platteneigenschaften mit fortlaufender Vorhersage zu berücksichtigen. Der prozentuale Vorhersagefehler (MAPE_P) der Querzugfestigkeit und der Dickenquellung erhöhte sich mit zunehmender Einsatzzeit der Modelle. Der prozentuale Vorhersagefehler der Abhebefestigkeit und der mittleren Rohdichte verbesserte sich dagegen geringfügig bzw. blieb weitgehend unverändert. Werden jedoch die Schätzgenauigkeiten auf die veränderten empirischen Standardabweichungen in den Testdatenräumen bezogen, verbesserte sich die Schätzgenauigkeit der Querzugfestigkeit und der mittleren Rohdichte über die Zeit geringfügig. Die auf gleicher Weise normierten Schätzgenauigkeiten der Dickenquellung und der Abhebefestigkeit nahmen dagegen ab. Die Dickenquellung wurde von allen untersuchten Eigenschaften am stärksten von Effekten beeinflusst, die nicht im Modell berücksichtigt worden sind (z.B. nachträglich eingebauter Multipot, Änderungen in der Holzartenzusammensetzung).

²⁰⁰

⁽¹⁾ Bias = Abweichung durch systematischen Fehler

Die **bedeutendsten Prozessvariablen** zur Erklärung der Schwankungen der Platteneigenschaften im Vorhersageraum waren aus der Gruppe der Kontrollvariablen (steuerbare Größen) der Leimtyp, die Leim- und die Härtermenge, die Dampfmenge im Kocher bzw. im Refiner, die Temperatur am Trocknerausgang, die Feuchte der Fasern, der Druck in der Vorpresse, die Pressengeschwindigkeit, die Multipots und Distanzen bzw. Drücke verschiedener Pressenrahmen. Aus der Gruppe der nicht steuerbaren Einflussgrößen waren die Variablen Stromaufnahme der Dosierschnecke, die Stromaufnahme des Bandantriebes der Vorpresse, die Temperatur im Streubunker und der Füllstand des Faserbunkers für die Vorhersage bedeutsam.

Die ausgewählten Pressenrahmen in den Regressionsmodellen sind als "Stellvertretervariablen" für die jeweiligen Pressenabschnitte zu interpretieren. Die Einflüsse konnten auch an anderen Pressenrahmen nachgewiesen werden, als in den letztendlich ausgewählten. Die Einbeziehung aller Pressenrahmen war jedoch aufgrund der hohen Korrelationen zwischen den Pressenrahmen nicht möglich.

Um den Einfluss des Pressprogramms einer kontinuierlichen Presse in Form der einzelnen Pressenabschnitte auf die technologischen Eigenschaften untersuchen zu können, wurden die Variablen zunächst durch eine **Faktorenanalyse** zusammengefasst und anschließend in die gebildeten Regressionsmodelle integriert. Dabei konnte sowohl das Druckprofil als auch das Distanzprofil durch 4 bzw. 5 Faktoren mit einer erklärten Varianz von 86-95% ausreichend genau beschrieben werden. Bei einem 5 Faktoren-Modell bildeten die extrahierten Faktoren die Pressenabschnitte "Schließen", "Lüften", "Halten", "2. Verdichtungsschritt" und die "Kalibrierzone". Die neu gebildeten Variablen konnten dabei eindeutig den einzelnen Pressenrahmen zugeordnet werden. Bei einem 4 Faktoren-Modell der Heißpresse wurden aufgrund der höchsten Ähnlichkeit die Abschnitte "Lüften" und "Halten" zu einem gemeinsamen Faktor "Erwärmen" zusammengefasst.

Der Einfluss der einzelnen Pressenfaktoren auf die untersuchten Platteneigenschaften konnte innerhalb der Regressionsmodelle nachgewiesen werden. Es konnte festgestellt werden, dass insbesondere die Dickenquellung und die mittlere Rohdichte durch die Schwankungen aller Pressenabschnitte (Distanzen und Drücke) beeinflusst wurden. Die Querzugfestigkeit dagegen wurde nur durch den Abschnitt "Halten" und die Abhebefestigkeit durch den "2. Verdichtungsschritt" bzw. der "Kalibrierzone" signifikant beeinflusst.

Die Schätzgenauigkeit der Regressionsmodelle konnte jedoch durch die Integration der Pressenfaktoren aus der PCA-Analyse nicht erhöht werden.

Der Einfluss aller korrelierten Prozessvariablen auf die Platteneigenschaften wurde innerhalb der **PLSR-Modelle** untersucht. Die Platteneigenschaften werden im Gegensatz zu den multivariaten Regressionsmodellen in einem PLSR-Modell durch einen gemeinsamen X-Variablen-Block erklärt. Grundsätzlich kann jede Eigenschaft auch individuell modelliert werden. Je unterschiedlicher die zu betrachtenden Eigenschaften sind, desto eher sollten diese auch individuell modelliert werden. Gleichwohl die vier zu untersuchenden Platteneigenschaften Querzugfestigkeit, Dickenquellung, Abhebefestigkeit und mittlere Rohdichte signifikant miteinander korreliert sind, werden diese Eigenschaften durch unterschiedliche Phänomene geprägt.

Das multivariate (gemeinsame) PLSR-Modell tendierte zu einer hohen Modelldimension (bis zu 9 PLS-Komponenten). Die Ursache dafür liegt in der unterschiedlichen Anpassungsfähigkeit und Charakteristika der Eigenschaften. Eine höhere Schätzgenauigkeit wurde für die Eigenschaften Querzugfestigkeit, Dickenquellung und mittlere Rohdichte durch eine individuelle Modellierung und einer gleichzeitigen Variablenreduzierung erreicht. Die Berücksichtigung aller Prozessvariablen in den Modellen führte stets zu einem höheren Vorhersagefehler, da tatsächlich nur wenige Variablen einen bedeutenden Einfluss ausüben. Die Einbeziehung aller Prozessvariablen stellt jedoch sicher, dass zukünftige Prozessänderungen und deren Auswirkungen auf die Platteneigenschaften erfasst werden.

Die Untersuchungen zur Langzeitstabilität der Vorhersagen durch die PLSR-Modelle bestätigten die Ergebnisse der Regressionsmodelle. Die Schätzgenauigkeiten der beiden Modelltypen unterschieden sich nicht signifikant voneinander. Die Abhebefestigkeit konnte jedoch mit den Regressionsmodellen etwas genauer vorhergesagt werden. Im Vorhersagezeitraum von 12 Monaten wurde die Querzugfestigkeit mit einem prozentualen Fehler von ca. 5%, die Dickenquellung von 7%, die Abhebefestigkeit von 6,5% und die mittlere Rohdichte mit 0,8% vorhergesagt.

Die technologischen Interpretationsmöglichkeiten der PLSR-Modelle sind aufgrund der Berücksichtigung der Korrelationsstrukturen gegenüber den Regressionsmodellen weitergehend. Es zeigte sich jedoch, dass alle unabhängigen Haupteinflussfaktoren in den Regressionsmodellen berücksichtigt wurden. Anderenfalls hätten sich signifikante Unterschiede in den Vorhersagegenauigkeiten zwischen den Modelltypen ergeben müssen. Anhand exemplarischer Beispiele aus dem Trocknungs- und Sichtungsprozess, der Beleimung und der Heißpresse wurden die Einflüsse der korrelierten Variablen auf die Platteneigenschaften dargestellt.

Durch die Arbeit wurde nachgewiesen, dass eine Echtzeitqualitätskontrolle von kontinuierlich hergestellten Faserplatten auf Basis der Prozessdaten möglich ist. Der Nutzen der Prozessmodelle, insbesondere der PLSR-Modelle, geht jedoch weit über die Echtzeitqualitätskontrolle hinaus. Sie sind ein optimales Tool für Wissenschaft und Praxis, um den Einfluss von Prozessvariablen auf die Platteneigenschaften zu untersuchen und zu quantifizieren. Die Modelle ermöglichen es, den Sicherheitsabstand zwischen dem gleitenden Produktionsmittelwert und der Güteschränke zu verringern, da auf Qualitätsabweichungen ohne zeitliche Verzögerung reagiert werden kann. Eine fundierte Optimierung des Prozesses auf Grundlage des Prozessmodells ermöglicht eine Senkung der Herstellbzw. der Rohstoffkosten.

12 Tabellenverzeichnis

Tabelle 1: Vor- und Nachteile von analytischen und empirischen Modellen	4
Tabelle 2: Übersicht zu den dargestellten Kennziffern im Ergebnisteil	51
Tabelle 3: Aufteilung Prozessvariablen in ihre Unterbereiche	54
Tabelle 4: Variablen mit Variationskoeffizienten größer 20% im Lerndatensatz	59
Tabelle 5: Variablen mit Variationskoeffizienten kleiner 2% im Lerndatensatz	59
Tabelle 6: Anzahl Ausreißer nach verschiedenen Verfahren	61
Tabelle 7: p-Werte für den Test auf Normalverteilung	63
Tabelle 8: Anzahl signifikanter Korrelationen der Prozessdaten nach der	
Plausibilitätsprüfung zu den Laboreigenschaften (Lerndatensatz, S=95%)	67
Tabelle 9: Korrelationsmatrix der Laborwerte (N=132)	67
Tabelle 10: Maximale Korrelationen der Prozessdaten zu den Labordaten (N=132)	68
Tabelle 11: Herleitung der Bewertungszahlen für die Eingangsparameter	
(Regressoren)	70
Tabelle 12: Analyse der Varianzen	71
Tabelle 13: Qualitätsunterschied in Abhängigkeit vom Klebstofflieferanten	
(Lerndatensatz)	74
Tabelle 14: Bewertungsmatrix der Regressoren	77
Tabelle 15: Standardabweichung und Variationskoeffizient der Lern- und Testdaten	78
Tabelle 16: Gütekennzahlen Modell 1 (BWZ=50%)	78
Tabelle 17: Gütekennzahlen Modell 2 (BWZ=55%)	79
Tabelle 18: Gütekennzahlen Modell 3 (BWZ=60%)	79
Tabelle 19: Gütekennzahlen Modell 4 (BWZ=65%)	79
Tabelle 20: Gütekennzahlen Modell 5 (BWZ=70%)	80
Tabelle 21: Gütekennzahlen Modell 6 (BWZ=75%)	80
Tabelle 22: Gütekennzahlen Modell 7 (BWZ=80%)	80
Tabelle 23: Schätzgenauigkeiten im Vergleich	83
Tabelle 24: Gütekennzahlen Modells 8	89
Tabelle 25: Einfluss von stark korrelierten Variablen in der Regressionsanalyse	92
Tabelle 26: Korrelationsmatrix (n=132)	93
Tabelle 27: Variablenauswahl technologisches Modell	94
Tabelle 28: Gütekennzahlen Modell 9 (technologisches Modell)	94
Tabelle 29: Gütekennzahlen Modell 10	95
Tabelle 30: Vergleich der Validierungsergebnisse der MaxR-Modelle	99
Tabelle 31: Gütekennzahlen Modell 11	100
Tabelle 32: Durbin-Watson- (DW) Koeffizienten des Modells 10	105
Tabelle 33: Gütekennzahlen Modell 11 mit der Querzugfestigkeit als endogenen	
Regressor	108
Tabelle 34: Gütekennzahlen Modell 10 ohne Aktualisierung (Variante 1)	109
Tabelle 35: Gütekennzahlen Modell 9 ohne Aktualisierung (Variante 1)	109
Tabelle 36: Gütekennzahlen Modell 9 mit Teilaktualisierung (Variante 2)	_ 110
Tabelle 37: Genormte Schätzgenauigkeiten in Abhängigkeit der	
Parameteraktualisierung	_ 110
Tabelle 38: Gütekennzahlen Modell 5 über einen Zeitraum von 1 Jahr	_ 112
Tabelle 39: Gütekennzahlen Modell 10 über einen Zeitraum von 1 Jahr	112

Tabelle 40: Genormte Schätzgenauigkeiten des Modells 5 und 10 über 3 und 12	
Monate im Vergleich	113
Tabelle 41: Korrelationen zwischen den Beleimungsdaten (N=248)	118
Tabelle 42: Korrelationen zwischen Beleimungsdaten und Laboreigenschaften	118
Tabelle 43: Einfluss von Harnstoff auf die Platteneigenschaften	120
Tabelle 44: Signifikante Einflussgrößen in den Modellen	125
Tabelle 45: Eigenwerte der ersten 6 Hauptfaktoren der PCA-Analyse für die	
Korrelationsmatrix der Distanzen und Drücke getrennt nach rechter und linker	
Seite	_129
Tabelle 46: Eigenwerte der ersten 6 Hauptfaktoren der PFA-Analyse für die	
Korrelationsmatrix der Distanzen und Drücke getrennt nach rechter und linker	
Seite	_130
Tabelle 47: Rotiertes Faktorenmuster der Methode Varimax für 5 Faktoren (Distanzen	
rechts)	_131
Tabelle 48: Interpretation und Zuordnung der Faktoren (PCA) für das Distanzen- und	
Druckprofil links und rechts (Rotation Varimax)	_132
Tabelle 49: Erklärte Varianzen durch die PCA-Faktoren vor und nach der Rotation	
Varimax	_134
Tabelle 50: Eigenwerte der ersten 6 Hauptfaktoren der PCA-Analyse (links) und der	
PFA-Analyse (rechts) für die Korrelationsmatrix Drücke der rechten und linken	
Seite	_135
Tabelle 51: Interpretation und Zuordnung der Faktoren (PCA) für das Distanzenprofil	
links und spezifisches Druckprofil (Rotation Equamax)	_136
Tabelle 52: Interpretation und Zuordnung der Faktoren (PCA) für das spezifische	
Druck- und Distanzenprofil links zu Beginn und Ende der Vorhersage	_138
Tabelle 53: signifikanter Einfluss der Pressenfaktoren (PCA) auf die	
Platteneigenschaften innerhalb der Regressionsmodelle mit Vorzeichen der	
Regressionskoeffizienten	_140
Tabelle 54: PCA-Analyse, Eigenwerte und erklärte Varianz der 4 Laboreigenschaften	_143
Tabelle 55: Anteil der erklärten aufsummierten Varianz in % im X- und Y-Raum durch	
die ersten 6 PLS-Faktoren für ein Gesamt- und den 4 Einzelmodellen	_149
Tabelle 56: Gütekennzahlen gemeinsames Modell mit 5 PLS-Faktoren (281	
Prozessvariablen)	_159
Tabelle 57: Gütekennzahlen gemeinsames Modell mit 9 PLS-Faktoren (281	
Prozessvariablen)	_159
Tabelle 58: Gütekennzahlen gemeinsames Modell mit 6 PLS-Faktoren (200	
Prozessvariablen)	_161
Tabelle 59: Gütekennzahlen IB-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen) _	_164
Tabelle 60: Gütekennzahlen IB-Modelle mit 2-5 PLS-Faktoren (44 Prozessvariablen)	_164
Tabelle 61: Gütekennzahlen TS-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)	_169
Tabelle 62: Gütekennzahlen TS-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen)	_169
Tabelle 63: Gütekennzahlen SS-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)	_171
Tabelle 64: Gütekennzahlen SS-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen) _	_171
Tabelle 65: Gütekennzahlen MD-Modelle mit 2-5 PLS-Faktoren (281 Prozessvariablen)	_174
Tabelle 66: Gütekennzahlen MD-Modelle mit 2-5 PLS-Faktoren (60 Prozessvariablen)	_174
Tabelle 67: Gütekennzahlen gemeinsames Modell mit 5 PLS-Faktoren über einen	
Zeittraum von 1 Jahr (281 Prozessvariablen)	175

Tabelle 68: Gütekennzahlen gemeinsames Modell mit 6 PLS-Faktoren über einen	
Zeittraum von 1 Jahr (200 Prozessvariablen)	175
Tabelle 69: Genormte Schätzgenauigkeiten der Gesamtmodelle über 3 und 12 Monate	
im Vergleich	176
Tabelle 70: Gütekennzahlen Einzelmodelle mit unterschiedlichen PLS-Faktoren über	
einen Zeittraum von 1 Jahr (281 Prozessvariablen)	176
Tabelle 71: Gütekennzahlen Einzelmodelle mit unterschiedlichen PLS-Faktoren über	
einen Zeittraum von 1 Jahr (44-60 Prozessvariablen)	177
Tabelle 72: Genormte Schätzgenauigkeiten der PLS-Einzelmodelle über 3 und 12	
Monate im Vergleich	177
Tabelle 73: Unterschiede in den Modellannahmen	186
Tabelle 74: Standardabweichungen und Variationskoeffizienten der	
Platteneigenschaften in den einzelnen Datenräumen	191
Tabelle 75: Vergleich der ex-ante-Vorhersagegenauigkeiten über 3 Monate	191
Tabelle 76: Vergleich der ex-ante-Vorhersagegenauigkeiten über 12 Monate	191

13 Abbildungsverzeichnis

Abbildung 1: Produktion von Holzwerkstoffen in Europa (nach EPF 2003)	1
Abbildung 2: Annahmen des klassischen linearen Modells der Normalregression und	
einige Verallgemeinerungen (Gruber, 1994)	31
Abbildung 3: Instationäre Prozessvariablen aufgrund technologischer Anpassungen	
(links) bzw. jahreszeitlichen Einflüssen (rechts)	55
Abbildung 4: Verlauf der Laborwerte (normiert auf Mittelwert) über 21 Monate	56
Abbildung 5: Verlauf Druck Vorverdichter (links) und Distanz Multipot 2 (rechts)	57
Abbildung 6: Ausreißer im zweidimensionalem Raum, nach Draper und Smith (1998)	60
Abbildung 7: Detektierte scheinbare Ausreißer in der Datenaufzeichnung	62
Abbildung 8: Histogramm- und Normalverteilungs-Diagramme für die Querzugfestigkeit	
und die Plattenfeuchte (normiert, Mittelwert entspricht 100%)	64
Abbildung 9: Korrelationskoeffizienten zwischen den Prozessdaten und den	
Laboreigenschaften, Korrelationskoeffizienten außerhalb der horizontalen	
Linien statistisch signifikant (S=95%), vertikale Linien kennzeichnen	
Prozessabschnitte	66
Abbildung 10: normierte Mittelwerte (%) und Variationskoeffizienten der	
Laboreigenschaften über die Produktionsbreite (HDF 8mm, n=33, 100%	
entspricht minimalen Mittelwert)	_72
Abbildung 11: Pressprogramm	_73
Abbildung 12: Typischer Verlauf der Bewertungszahlen, dargestellt am Beispiel für die	
Querzugfestigkeit	_76
Abbildung 13: Zusammenhang zwischen Anzahl der Regressoren in den Gleichungen	
(bestimmt durch die Bewertungszahl) und dem Bestimmtheitsmaß	_81
Abbildung 14: Einfluss der Anzahl an Regressoren, bestimmt durch die	
Bewertungszahl, auf die durchschnittliche normierte Vorhersagegenauigkeit im	
Beobachtungs- und Vorhersageraum	_82
Abbildung 15: Einfluss der Anzahl der Regressoren auf das Bestimmtheitsmaß, F-Wert	
und die Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum für alle	
vier Gleichungen	84
Abbildung 16: Einfluss der aufgenommenen Regressoren im Modell auf die	
durchschnittliche normierte Vorhersagegenauigkeit im Beobachtungs- und im	
Vorhersageraum	85
Abbildung 17: Einfluss der aufgenommenen Regressoren auf die Schätzgenauigkeit im	~~
Vorhersageraum	_86
Abbildung 18: Entwicklung der F-Werte der vier Modeligleichungen über den Zeitraum	~~~
der ex-ante-Vorhersagen (Modell 8)	_89
Abbildung 19: Einfluss der Herausnahme von Regressoren mit dem jeweils kleinstem t-	
Wert auf das Bestimmtneitsmals, F-Wert und die Vornersagegenauigkeit im	~~~
Beobachtungs- und Vornersageraum	_90
Abbildung 20: Qualitatsvornersagen für die Platteneigenschaften (Modell 10)	_97
Abbildung 21: Einfluss der Regressorenanzani im Modell auf die durchschnittliche	
vornersagegenauigkeit im Beobachtungs- und Vornersageraum (Methode	100
Maxx)	_100
Abbildung 22: Entwicklung der F-vverte der vier Modeligielchungen über den Zeitraum	101
der ex-ante-vornersagen (Modell 11)	_ 101
Abbildung 23: Vorhersagen der Querzugfestigkeit (Modell 11)	101
---	-------------------
Abbildung 24: Residuen des Modells 10 (n=132) als Histogramm und normal plot	103
Abbildung 25: Homoskedastizität (links) und Heteroskedastizität (rechts) der Residuen	
(vereinfachende Darstellung)	104
Abbildung 26: Verlauf der Residuen über die Zeit	104
Abbildung 27: Eine unkorrelierte Serie von Residuen (IB) des Modells 10	105
Abbildung 28: Vorhersagen der Querzugfestigkeit des Modells 10 mit Ausreißer	112
Abbildung 29: Entwicklung der Bestimmtheitsmaße während der ex-ante-Vorhersagen	
über ein Jahr (Modell 10)	114
Abbildung 30: Entwicklung der F-Werte während der ex-ante-Vorhersagen über ein	
Jahr (Modell 10)	114
Abbildung 31: Grafische Darstellung der Korrelationen ausgewählter Distanzen	128
Abbildung 32: Distanzen (links) und Drücke (rechts) des Rahmens 5	132
Abbildung 33: Grafische Repräsentation des Distanzenprofils bzw. der PCA-Faktoren	
nach der Varimax-Rotation	133
Abbildung 34: Erklärte Varianz durch die Faktoren vor und nach der Varimax-Rotation	
(Distanzen rechts)	134
Abbildung 35: Kommunalitäten der Distanzen rechts der Presse; PCA mit 5 Faktoren	135
Abbildung 36: Erklärte Varianz durch die Faktoren vor und nach der Rotation	
(Distanzen. links) Methode Equamax; links: Beginn Vorhersage, rechts: Ende	
Vorhersage	137
Abbildung 37: PCA Faktorladungen eines 2- Komponentenmodells für die	
Laboreigenschaften nach der Varimax-Rotation (Lerndatensatz)	144
Abbildung 38: PCA Faktorladungen eines 3- Komponentenmodells für die	
Laboreigenschaften nach der Varimax-Rotation (Lerndatensatz)	144
Abbildung 39: Zusammenhang zwischen Anzahl der PLS-Faktoren und der	
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS, mittels "Leave-one-	
out" Cross-Validierung (Modell mit 4 Zieleigenschaften, 281 Prozessvariablen)	14
Abbildung 40: Zusammenhang zwischen Anzahl der PLS-Faktoren und der	
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS, mittels Block-Cross-	
Validierung (Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)	146
Abbildung 41: PLS Scores t(1) und t(2) eines gemeinsamen multivariaten PLS-Modells	
(Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)	147
Abbildung 42: PLS Scores u(1) und u(2) eines gemeinsamen multivariaten PLS-	
Modells (Modell mit 4 Zieleigenschaften und 281 Prozessvariablen)	148
Abbildung 43: PLS Scores t(1) und u(1) eines multivariaten PLS-Modells (4	
Zieleigenschaften mit 218 Prozessvariablen)	150
Abbildung 44: PLS Scores t(2) und u(2) eines multivariaten PLS-Modells (4	
Zieleigenschaften mit 218 Prozessvariablen)	150
Abbildung 45: Residuen des multivariaten PLS-Modells mit 5 Komponenten als normal	
plot (4 Zieleigenschaften und 281 Prozessvariablen)	. 15 ⁻
Abbildung 46: DModX für jede Beobachtung des Lerndatensatzes eines gemeinsamen	
5- Komponenten PLS-Modells (4 Zieleigenschaften, 281 Prozessvariablen)	15
Abbildung 47: DModY für jede Beobachtung des Lerndatensatzes eines gemeinsamen	
5- Komponenten PLS-Modells (4 Zieleigenschaften, 281 Prozessvariablen)	15
Abbildung 48: PLS-Gewichte für die ersten beiden PLS-Faktoren eines Modells für die	
Dickenquellung (60 Prozessvariablen)	154

Abbildung 49: VIP und standardisierte Regressionskoeffizienten eines 3 Komponenten Medelle für die Diekonguellung (Lerndatensetz, 60 Prozessystiablen)	155
Abbildung 50: Entwicklung der Bestimmtheitsmaße Ry ² in Abhängigkeit der Anzahl der	_ 155
PLS-Faktoren (mit 281 Prozessvariablen)	157
Abbildung 51: Einfluss der Anzahl der PLS-Faktoren auf die genormte	-
Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum (281	
Prozessvariablen)	158
Abbildung 52: Finfluss der PI S-Faktoren auf die durchschnittliche normierte	
Vorhersage genauigkeit im Beobachtungs- und Vorhersageraum (281	
Prozessvariablen)	158
Abbildung 53: Einfluss der Anzahl der PLS-Faktoren auf die genormte	-
Vorhersagegenauigkeit im Beobachtungs- und Vorhersageraum (200	
Prozessvariablen)	160
Abbildung 54: Einfluss der PLS-Faktoren auf die durchschnittliche normierte	_
Vorhersage genauigkeit im Beobachtungs- und Vorhersageraum (200	
Prozessvariablen)	160
Abbildung 55: Zusammenhang zwischen Anzahl der PLS Komponenten und der	_
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-	
Validierung, Modell für die Querzugfestigkeit mit 281 Prozessvariablen	162
Abbildung 56: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im	
Beobachtungs- und Vorhersageraum (IB-Modell mit 281 Prozessvariablen)	163
Abbildung 57: Einfluss der PLS Faktoren auf die normierte Schätzgenauigkeit im	
Beobachtungs- und Vorhersageraum (IB-Modell mit 281 Prozessvariablen)	163
Abbildung 58: Einfluss der PLS-Faktoren auf Kennzahlen des Modells (IB) im	_
Beobachtungs- und Vorhersageraum (44 Prozessvariablen)	165
Abbildung 59: Einfluss der PLS-Faktoren auf die normierte Schätzgenauigkeit im	_
Beobachtungs- und Vorhersageraum (IB-Modell mit 44 Prozessvariablen)	166
Abbildung 60: Zusammenhang zwischen Anzahl der PLS-Faktoren und der	_
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-	
Validierung, Modell für die Dickenquellung mit 281 Prozessvariablen	167
Abbildung 61: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im	_
Beobachtungs- und Vorhersageraum (TS-Modell mit 281 Prozessvariablen)	168
Abbildung 62: Einfluss der PLS-Faktoren auf die normierte Schätzgenauigkeit im	
Beobachtungs- und Vorhersageraum (TS-Modell mit 281 Prozessvariablen)	_168
Abbildung 63: Zusammenhang zwischen Anzahl der PLS-Faktoren und der	
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-	
Validierung, Modell für die Abhebefestigkeit (SS) mit 281 Prozessvariablen	170
Abbildung 64: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im	
Beobachtungs- und Vorhersageraum (SS-Modell mit 281 Prozessvariablen)	170
Abbildung 65: Einfluss der PLS-Faktoren auf die normierte Schätzgenauigkeit im	
Beobachtungs- und Vorhersageraum (SS-Modell mit 281 Prozessvariablen)	171
Abbildung 66: Zusammenhang zwischen Anzahl der PLS-Faktoren und der	_
Vorhersagegenauigkeit, dargestellt am Root Mean PRESS mittels Cross-	
Validierung, Modell für die Rohdichte (MD) mit 281 Prozessvariablen	_172
Abbildung 67: Einfluss der PLS-Faktoren auf Kennzahlen des Modells im	-
Beobachtungs- und Vorhersageraum (MD-Modell mit 281 Prozessvariablen)	_173

Abbildung 68: Einfluss der PLS Faktoren auf die normierte Schätzgenauigkeit im	
Beobachtungs- und Vorhersageraum (MD-Modell mit 281 Prozessvariablen)	173
Abbildung 69: Entwicklung der Bestimmtheitsmaße der PLS-Einzelmodelle während	
der ex-ante-Vorhersagen über 1 Jahr (reduzierte Variablenanzahl)	178
Abbildung 70: Einfluss der Temperaturen im Trocknungs- und Sichtungsprozess sowie	
im Streubunker (PLS-Einzelmodelle, 281 Prozessvariablen, n.s. = nicht	
signifikant bei S=95%)	180
Abbildung 71: Einfluss der Beleimungsvariablen auf die Platteneigenschaften (PLS-	
Einzelmodelle, 281 Prozessvariablen)	181
Abbildung 72: Einfluss der Pressendistanzen auf die Dickenquellung (PLS-	
Einzelmodell, 281 Prozessvariablen)	182
Abbildung 73: Scheinbarer Einfluss der Pressendistanzen auf die Dickenquellung im	
Regressionsmodell	183
Abbildung 74: PLS-Regressionskoeffizienten der Multipotdrücke für die	
Querzugfestigkeit und Abhebefestigkeit (PLS-Einzelmodelle, 281	
Prozessvariablen)	184

14 Literatur

Amemiyar, T. (1980) Selection of regressors. International Economic Review 21, 331-354.

- Andres, F. (1989) Untersuchungen zur Modellierung von Teilprozessen der
 Spanplattenfertigung Teilprozess Pressen. Ingenieurbeleg, TU Dresden, Sektion
 Verarbeitungs- und Verfahrenstechnik, Bereich Holz- und Faserwerkstofftechnik.
- Bärmann, F. (2001) Modellierung von Kontianlagen mit neuronalen Netzen. www.baermann.de.
- Bärmann, F. (2002) NN-Tool 2000. Handbuch, Eigenverlag.
- Belimow, D. (1978) Ein Beitrag zur Optimierung der Spanplattenfertigung unter besonderer Beachtung des Heißpressvorganges. Dissertation, TU Dresden.
- Benker, B., Erbreich, M., Plinke, B., Wollmann, A. (2003) Characterisation of wood fibres for MDF – off-line and on-line measurements for production monitoring. Proceedings 7th European Panel Products Symposium, Llandudno, Wales, S. 1-13.
- Bernardy, G. et al. (1999) Cost reduction by the integration of process control technology and statistical process optimisation. Proceedings 3rd European Panel Products Symposium, Llandudno, Wales, S. 235 246.
- Bernardy, G., Scherff, B. (1997) Prozessmodellierung führt zur on-line-Qualitätskontrolle und Prozessoptimierung bei der Span- und Faserplattenproduktion. Holz Roh- Werkstoff 55, S.133-140.
- Box, G. E. P. (1966) The Use and Abuse of Regression. Technometrics 8 (4), 625 -629.
- Box, G. E. P., Hunter, W. G. (1978): Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley series in probability and mathematical statistics, New York.
- Brade, F. (1987) Untersuchungen zur Modellierung des Zusammenhangs zwischen Rohdichteprofil und strukturellen sowie prozesstechnischen Parametern. Diplomarbeit, TU Dresden.
- Burnham, A. J., MacGregor, J. F., Viveris, R. (1999) Latent variable regression modeling. Chemometrics and Intelligent Laboratory Systems 48, 167-180.
- Burnham, A. J., MacGregor, J. F., Viveros, R. (2001) Interpretation of regression coefficients under a latent variable regression model. J. Chemometr. 15, 265-284.
- Burmester, A. (1970) Formbeständigkeit von Holz gegenüber Feuchtigkeit. BAM-Bericht Nr. 4, Berlin, 179 S.
- Carlsson, J., Mathiasson, M., Goeminne, J. (2002) On-line monitoring of fiber quality in the MDF process. Proceedings 6th European Panel Products Symposium, Llandudno, Wales, 167-177.

- Carvalho, L. M. H., Costa, C. A. V. (1998) Modeling and simulation of the hot-pressing process in the production of medium density fibreboard (MDF). Gordon & Breach Science Publishing Inc. Chemical Engineering Communications170, 1-21.
- Dai C., Yu, C., Hubert, P. (2000) Modelling vertical density profile in wood composites during hot pressing. 5th Pacific Rim Bio-Based Composites Symposium, Canberra, 10-13.
- Dengel, D. (1996) Ausreißer einfach bewerten. Materialprüfung 38, 3 S. 99-101.
- Deppe, H. J., Ernst, K. (2000) Taschenbuch der Spanplattentechnik. Leinfelden-Echterdingen, DRW-Verlag, 467 S.
- Deppe, H. J., Ernst, K. (1996) MDF Mitteldichte Faserplatten. Leinfelden-Echterdingen, DRW-Verlag, 200 S.
- Draper, N. R., Smith, H. (1998) Applied regression analysis. New York, Wiley, 706 p.
- Engström, B., et al. (1998) Process modelling system for particleboard manufacturing, incorporating near infrared spectroscopy on dried wood particles. Proceedings of the 2nd European Panel Products Symposium, Llandudno, Wales, 107-114.
- EPF (2003) European Panel Federation. Annual Report 2002-2003. 251 S.
- Faix, O. (2003) Grundlagen der Holzchemie. Teil 1. Vorlesungsskript Studiengang Holzwirtschaft, 445 S.
- Farnum; N. R., Stanton, L. W. (1992) Quantitative forecasting methods. Boston, PWS-Kent, 573 p.
- Fomby, T. B., Hill, R. C., Johnson, S. R. (1988) Advanced econometric methods. New York, Springer Verlag, 623 S.
- Friedrich, B. (1987) Identifikation und Beschreibung des Zusammenhanges zwischen technisch-technologischen Einflussfaktoren und den Produkteigenschaften von Spanplatten. Dissertation, TU Dresden.
- Fuchs, M. (2003) Ultrasonic camera minimizes press factors. 37th International Wood Composite Materials, University of Washington, Pullman, 89-94.
- Garcia, P. J. (2002) Three-dimensional heat and mass transfer oriented strandboard hotpressing. University of British Columbia, CA. Ph.D. Thesis, 254 p.
- Geladi P. and Kowalski B.R. (1986) Partial Least Squares Regression: A Tutorial. Analytica Chimica Acta 185, 1-17.
- Geyer, S., Tietze, D. (1984) Untersuchungen zur Optimierung des stofflich-strukturellen Aufbaus von Spanplatten. Ingenieurbeleg, TU Dresden, Sektion Verarbeitungs- und Verfahrenstechnik, Bereich Holz- und Faserwerkstofftechnik.
- Gleser, L. J. (1981) Estimation in a multivariate ,Errors-in-variables' regression model: large sample results. The Annals of Statistics 9, 24-44.

- Goroyias, G. J., Hale, M. D. (2002) Heat treatment of wood strands for OSB prduction: Effect on the mechanical properties, water absorption and dimensional stability. 33rd Annual meeting of the international research group on wood preservation. Section 4., Cardiff, Wales, 1-18.
- Goroyias, G. J., McCarthy, E. T. (2002) A new method for the assessment of fibre properties and optimisation of refining parameters in MDF manufacture. Proceedings 6th European Panel Products Symposium, Llandudno, Wales, 186-197.
- Gotsman, V. (2004) Latest Development in MDF Thin-Board Production. 4th European Wood-Based Panel Symposium, Hannover.
- Grässle, P., Baumann, H., Baumann, P. (2004) Geschäftsprozessmodellierung, IT-System-Spezifikation und Systemintegration. Galileo Computing, 288 S.
- Greubel, D. (1999) Untersuchungen von Methoden zur Qualitätssicherung durch Prozessmodelle. Abschlußbericht AIF Vorhaben 10508 N, Braunschweig, 63 S.
- Greubel, D., Lobenhoffer, H., Ballüer, L. (1999) On-Line Qualitätskontrolle in der Holzwerkstoffindustrie. Holz-Zentralblatt 52, S. 767-770.
- Groom, L., Mott, L., Shaler, S. (1999) Relationship between Fiber Furnish Properties and the Structural Performance of MDF. 33rd International Particleboard Composite Materials Symposium, Washington State University, Pullman, S. 89-100.
- Gruber, J. (1981) Ökonometrische Schätz- und Testverfahren. Ökonometrie 3, Hochschule Hagen, Fachbereich Wirtschaftswissenschaft.
- Gruber, J. (1993) Einführung in die multiple Regression und Ökonometrie, Ökonometrie 1. Hochschule Hagen, Fachbereich Wirtschaftswissenschaft.
- Gruber, J. (1994) Ergänzungen und Erweiterungen der klassischen Regressionsanalyse. Ökonometrie 1, Teil 2. Hochschule Hagen, Fachbereich Wirtschaftswissenschaft.
- Gruber, J. (1995) Ökonomische Prognose- und Optimierungsmodelle. Strukturelle Form und Typen ökonometrischer Modelle, Ökonometrie 2, Hochschule Hagen, Fachbereich Wirtschaftswissenschaft
- Hague, J., Robson, D., Riepen, M. (1999) MDF Process Variables An Overview of their Relative Importance. 33rd International Particleboard Composite Materials Symposium, Washington State University, Pullman, S. 79-87.
- Hanvongjirawat, W. (2003) Permeabilität von Holzwerkstoffmatten. Universität Hamburg, Dissertation, 205 S.
- Harless, T. et al. (1987) A model to predict the density profile of particleboard. Wood Fiber Science 19 (1), 81-92.
- Hartung, J., Elpelt, B. (1984) Multivariate Statistik Lehr- und Handbuch der angewandten Statistik. München, Oldenbourg.

- Haselein, C. R. (1998) Numerical simulation of pressing wood-fiber composites. Ph.D. Thesis, Oregon State University, 244 p.
- Hasener, J. (2003) Prozess- und Qualitätskontrolle bei der Herstellung von Holzwerkstoffen mit statistischen Prozessmodellen. Abschlussbericht AiF 12977N, 117 S.
- Haslbeck, O. (1995) Anwendung der qualitätsbedingten optimalen Prozesskontrolle in einem Spanplattenwerk. Diplomarbeit, FH Rosenheim, 91 S.
- Heinemann, C. (2003): Charakterisierung der Aushärtung von Harnstoffharzen in einer Holzpartikelmatrix, Dissertation, Universität Hamburg, 219 S.
- Humphrey, P. E. (1982) Physical aspects of wood particleboard manufacture. Ph.D. Thesis, University of Wales.
- Humphrey, P. E., Bolton, A. J. (1989) The hot pressing of dry-formed wood-based composites. Part II. A simulation model for heat and moisture transfer. Holzforschung 43 (3), 199-206.
- Janssen, A. (1997) Erfassung und detaillierte Auswertung von Prozessdaten aus der MDF-Produktion mit dem Ziel einer Prozessoptimierung. Diplomarbeit, Universität Hamburg. 137 S.
- Johnsson, B., Renmarker, P., Wallstén, P. (2000) BoardModel A new NIR-based process modeling system for wood panels. Proc. Int. Particleboard/Composite Materials Symposium, Forest Products Soc. Madison.
- Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., Lee, T.-S. (1988) Introduction for the theory and practice of econometrics. New York, Wiley, 1024 S.
- Kessler, R. W., Reinhardt, T., Kessler, W., Zimmer, H. (2000) Adaptiv processing of wood for fibreboards by high level spectroscopic on-line control. 4th European Panel Products Symposium, Llandudno, 227-235.
- Kessler, R. W., Kessler, W., Ganz, A., Zimmer, H. (2002) Quality monitoring for smart processing of fibre boards through on line spectroscopy. International Symposium, Vienna, Wood composites and chemistry, COST Action E13 Final Workshop "Wood Adhesion and Glued products", 257-265.
- Kresta, J., MacGregor, J. F., Marlin, T. E. (1991) Multivariate statistical monitoring of process operating performance. The Canadian Journal of Chemical Engineering 69, 35-47.
- Kourti, T., MacGregor, J. F. (1995) Tutorial: process analysis, monitoring and diagnosis, using multivariate projection methods. Chemometrics Intell. Lab. Syst. 28, 3-21.
- Kourti, T., Lee, J., MacGregor, J. F. (1996) Experiences with industrial applications of projection methods for multivariate statistical process control. Computers in Chemical Engineering 20, 2745-2750, Special supplement.

- Krug, D., Kehr, E. (2001) Einfluss des Aufschlussdruckes bei der Faserstoffherstellung auf die Quellungsvergütung von MDF. Holz als Roh- und Werkstoff 59, 342-343.
- Krug, D., Tobisch, S. (2002) Influence of the fibre pulping conditions and the resin type on the weathering resistance of MDF in exterior use. Proceedings 6th European Panel Products Symposium, Llandudno, Wales, S. 38-50.
- Kruse, K. (1997) Entwicklung eines Verfahrens der berührungslosen Ermittlung von Schallgeschwindigkeiten zur zerstörungsfreien Bestimmung mechanischer Eigenschaften an Holzwerkstoffplatten und dessen Integration in die Prozesskontrolle. Dissertation, Universität Hamburg, 195 S.
- Kruse, K.; Thömen, H.; Maurer, H.; Steffen, A.; León-Méndez, R. (1997) Optimierung des Spanplattenproduktionsprozesses mit Hilfe der Prozessmodellierung. Holz Roh-Werkstoff 55 (1), 17-21.
- Krüzner, M. (2004) The "Dry Blow Line". 4th European Wood-Based Panel Symposium, Hannover.
- Labosky, P., Yobp, R. D., Janowiak, J. J., Blankenhorn, P.R. (1993): Effect of steam pressure refining and resin levels on the properties of UF-bonded red maple MDF. Forest Product Journal 43 (11/12), 82-88.
- Landmesser, W. (1990) Rechnergestützte Qualitätssicherung in der Spanplattenfertigung. Unveröffentlichter Forschungsbericht, WTZ Holz GmbH Dresden.
- Landmesser, W., Ritter, C., Niemz, P. (1988) Ergebnisse und Probleme der Modellierung von Teilprozessen der Spanplattenfertigung. Holztechnologie 6, 316-319.
- Lee, H. H., Maloney, T. M. (1995) The effect of final moisture conntent of mat on the physical and mechanical properties of UF-bonded MDF. Journal of Korean Wood Science and Technology 23 (4), 85-90.
- Léon-Méndez, R., Thömen, H. (1996) Untersuchungen zur Optimierung eines Prozessleitsystems in einem Spanplattenwerk. Diplomarbeit, Universität Hamburg. 240 S.
- Liiri, O. (1977) Der Einfluss Holzart, Spangröße und Bindemittel auf die Festigkeit und die Quellung von Spanplatten mit höheren elastomechanischen Eigenschaften. Holzforschung und -verwertung 29 (6), 117-121.
- Lobenhoffer, H. (1982) Auswertung von Prozessdaten mittels statistischer Methoden. Holz als Roh- und Werkstoff 40, 395-401.
- Lobenhoffer, H. (1990) Qualitätsbedingte Regelung eines Spanplattenformstrangs. Dissertation, Georg-August-Universität Göttingen. 190 S.
- Lobenhoffer, H. (1991) Qualitätsbestimmte Regelung in der Spanplattenindustrie. Holz als Roh- und Werkstoff 49, 7-12.

- Lobenhoffer, H. et al. (1997) Regelung von Plattenqualität und -kosten mit Prozessdaten. Holz-Zentralblatt 68, 1038.
- Lobenhoffer, H., Ballüer, L., Greubel, D. (1999) Optimale Zustandsregelung von Plattenstraßen. Holz-Zentralblatt 103, 1354-1355.
- Loxton, C., Thumm, A., Grigsby, W. J., Adams, T. A., Ede, R. M. (2003) Resin distribution in MDF. Quantification of UF resin distribution on blowline and dry-blended MDF fiber and panels. Wood and Fiber Science 35 (3), 370-380.
- MacGregor, J. F. (1997) Using on-line process data to improve quality: Challenges for Statisticians. International Statistical Review 65 (3), 309-323.
- Mager, H. (1982): Moderne Regressionsanalyse. Frankfurt/M, Otto Saale Verlag.
- Maloney, T. (1993) Modern particleboard and dry-process fiberboard manufacturing. San Francisco, Miller Freeman.
- Manne, R. (1987) Analysis of two partial-least-squares algorithms for multivariate calibration. Chemometrics Intell. Lab. Syst. 2, 187-197.
- Mathiasson, M., Carlsson, J. (2003) NIR Technology for on-line property prediction. 37th International Wood Composite Materials Symposium, Washington State University, Pullman, 99-108.
- Mays, S. L., Kleinschmidt, H.-P. (2003) Increased panel production using delamination inspection systems. 37th International Wood Composite Materials, University of Washington, Pullman, 79-82.
- Mehlhorn, L., Landmesser, W. (1991) Prozessdatenerfassung und -auswertung in der Holzwerkstoffindustrie. WKI-Bericht Nr. 27.
- Mehlhorn, L., Thole, V. (1992) Datenerfassung und Datenauswertung zur Qualitätsüberwachung in einem Spanplattenwerk. Abschlussbericht 03 18800 A9.
- Mittag, H. J. (1993) Qualitätsregelkarten. München, Carls Hanser Verlag, 250 S.
- Myers, R., H. (1986) Classical and Modern Regression with Applications. Boston, Duxbury Press, 359 S.
- Nimz, P. (1982) Untersuchungen zum Einfluss der Struktur auf die Eigenschaften von Spanplatten; Teil 1: Einfluss von Partikelform, Rohdichte, Festharzanteil und Festparaffinanteil. Holztechnologie 23 (4), 206-213.
- Noack D. (2003) CE-Kennzeichnung von Holzwerkstoffen. 5. Holzwerkstoff-Kolloquium, Nischen für Holzwerkstoffe, Dresden.
- Nomikos, P., MacGregor, J. F. (1995) Multivariate SPC charts for monitoring batch processes. Technometrics 37, 41-59.

- Rackwitz, G. (1963) Der Einfluss der Spanabmessung auf einige Eigenschaften von Holzspanplatten. Holz Roh- Werkstoff 21, 200-209.
- Rasch, D., Verdorren, R. (2003) Einführung in die Biometrie. Grundlagen der Korrelations analyse und der Regressionsanalyse. Senat der Bundesforschungsanstalten des Bundesministeriums für Verbraucherschutz, Ernährung und Landwirtschaft.
- Reine, F. (1998) Konkurrierende neuronale Netze und ihre Anwendungen. Technische Universität Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, Dissertation, 131 S.
- Rials, T. G., Kelley, S. S., So, C.-L. (2002) Use of advanced spectroscopic techniques for predicting the mechanical properties of wood composites. Wood and Fiber Science 34 (3), 398-407.
- Ritter, C., Schweitzer, F. (1992) Neue Wege der Prozessmodellierung für die rechnergestützte Qualitätsvorhersage und -sicherung in der Spanplattenfertigung. Datenerfassung und multiple Regressionsanalyse. Holz als Roh- u. Werkstoff 50, 62-66.
- Sachs L. (1968) Statistische Auswertungsmethoden. Berlin, Heidelberg, New York, Springer Verlag, 671 S.
- Sahling, A. (1989) Untersuchungen zur Modellierung ausgewählter Teilprozesse in der Spanplattenindustrie. Teilprozess Vliesbilden. Ingenieurbeleg, TU Dresden, Sektion Verarbeitungs- und Verfahrenstechnik, Bereich Holz- und Faserwerkstofftechnik.
- Schäckel, R. (2001) New systems für online moisture measurement, one of the most important process parameters during the wood-based panel production. 3rd European Wood-Based Panel Symposium, Hanover.
- Scherff, B. et al. (1999) Produktqualität und Herstellkosten im Visier. Leitsystem bestimmt Plattenqualität und zeigt Optimierungspotentiale auf. Holz- und Möbelind. 11, 71-74.
- Schneeweiss, H., Mittag, H.-J. (1987) Lineare Modelle mit Fehler-behafteten Daten. Heidelberg, Physika-Verlag, 50 S.
- Schönfeld, P. (1971) Methoden der Ökonometrie. Band I und II, Verlag Franz Vahlen GmbH.
- Schubert, K. (1990) Untersuchungen zur Modellierung des Teilprozesses Heißpressens innerhalb der Spanplattenherstellung. Ingenieurbeleg, TU Dresden, Sektion Verarbeitungs- und Verfahrenstechnik, Bereich Holz- und Faserwerkstofftechnik.
- Schütte, B. (1998) Untersuchungen zur Vorhersagegenauigkeit der statistischen Prozesskontrolle in einem Spanplattenwerk. Diplomarbeit, FH Rosenheim, 110 S.
- Schweitzer, F. (1992) Modellierung des Heißpressenvorgangs zur Herstellung von Spanplatten in diskontinuierlich arbeitenden Pressen. Dissertation, TU Dresden.

- Schweitzer, F., Ritter, C. (1992) Neue Wege der Prozessmodellierung für die rechnergestützte Qualitätsvorhersage und –sicherung in der Spanplattenfertigung. Holz Roh-Werkstoff 50, 101-105.
- Suo, S., Bowyer, J. L. (1994) Simulation modelling of particleboard density profile. Wood Fiber Science 26 (3), 397-411.
- Schlittgen, R. (2001) Angewandte Zeitreihenanalyse. München, Oldenburg.
- Shao, J. (1993) Linear Model Selection by Cross-Validation. Journal of American Statistical Association 88, 486-494.
- Stahl, W. (1993) Machbarkeitsstudie über die Einführung von QCOPC in einem MDF-Werk. Diplomarbeit, FH Rosenheim.
- Stahl, W. (2002) Blending Factors Influencing the Quality of Medium Density Fibreboard. Ph.D. Thesis, University of Melborne, 212 p.
- Steffen, A., Janssen, A., Kruse, K. (2001) Analyse der Herstellung von MDF mit Hilfe der statistischen Prozessmodellierung. Holz als Roh- und Werkstoff 58, 419-431.
- Steffen, A. (1999) Möglichkeiten der Optimierung von Herstellungsprozess und Produkt qualität in der MDF-, Spanplatten- und OSB-Industrie. Habilitationsschrift, Universität Hamburg, 311 S.
- Suzuki, S. (2000) Recycled wood as a raw material for current PB production and a potential use for J-OSB. 5th Pacific Rim Bio-based Composites.
- Theil, H. (1971) Principles of Econometrics. Wiley, New York.
- Thole V. (2002) Strands für OSB aus großformatigen Holzteilen. WKI-Kurzbericht Nr. 16.
- Thömen, H. (2001) Modelling the physical Process in natural fiber composites during batch and continuous pressing. Ph.D. Thesis, Oregon State University, 188 p.
- Thömen, H., Humphrey, P. E. (2003) Modeling the continoius pressing process for wood based composites. Wood and Fibre Science 35 (3), 456-468.
- Trinks, R.-H. (1989) Untersuchungen zur Modellierung ausgewählter Teilprozesse in der Spanplattenindustrie. Teilprozess Leimaufbereitung/Beleimung. Diplomarbeit, TU Dresden.
- Thurstone, L. L. (1947) Multiple factor analysis. Chicago, University of Chicago Press,
- Van der Voet, H. (1994) Comparing the Predictive Accuracy of Models Using a Simple Ramdomiziation Test. Chemometrics Intell. Lab. Syst. 25, 313-323.

- von Haas, G. (1998) Untersuchungen zur Heißpressung von Holzwerkstoffmatten unter besondere Berücksichtigung des Verdichtungsverhaltens, der Permeabilität, der Temperaturleitfähigkeit und der Sorptionsgeschwindigkeit. Universität Hamburg, Dissertation, 264 S.
- Wakeling, I. N., Morris, J. J. (1993) A test of significance for partial least squares regression. J. Chemometrics 7, 291-304.
- Wold, H. (1966) Estimation of Principal Components and Related Models by Iterative Least Squares. In Multivariate Analysis, New York, Academic Press, 391-420.
- Wold, S., Ruhe, A., Wold, H., Dunn, W. J. (1984) The collinearity problem in linear regression. The partial least square approach to generalized inverses. SIAM J. Sci. Stat. Comput. 5, 735-743.
- Wold, S. (1993) PLS in chemical practice, discussion of a statistical view of some chemometrics regression tools. Technometrics 35, 136-139.
- Wold, S. (1994) PLS for Multivariate Linerar Modeling. QSAR: Chemometric Methods in Molecular Designs. Methods and Principles in Medicinical Chemistry, ed. H. van de Waterbeemd, Weinheim, Verlag-Chemie.
- Wold, S., Sjöström, M., Eriksson, L. (2001) PLS-Regression: A Basic Tool of Chemometrics. Chemomtrics and Intelligent Laboratory Systems 58, 109-130.
- Wolter, B. et al. (2002) Kernresonanz in Aufsatztechnik. Nuclear Magnetic Resonance in One-Sided Access Technique. Technisches Messen 69 (1), 43-48.
- Young, T. M., Andre, N., Huber, C. W. (2004) Predictive modeling of the internal bond of MDF using genetic algorithms with distributed data fusion. 8th European Panel Products Symposium, Llandudno, 1-13.
- Younger, M. S., (1979) A Handbook for Linear Regression. Duxbury Press, North Scituate, Massachusetts, 569 S.
- Zombori, B. G. (2001) Modeling the transient effects during the hot-pressing of wood based composites. Ph.D. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, 212 p.

15 Anhang

A 1: Lerndatensatz HDF (n=132)

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
Date						
Time						
Remark	Leimhersteller		-	-	-	-
LAB_IB	Querzugfestigkeit	Labor	N/mm²	1,90	0,14	7,4
LAB_Abhe	Abhebefestigkeit	Labor	N/mm²	2,29	0,23	10,2
LAB_Rohd	Mittlere Rohdichte	Labor	kg/m³	894,6	10,23	1,1
LAB_TS	Dickenquellung (24 h)	Labor	%	8,74	0,95	10,9
LAB_u	Plattenfeuchte	Labor	%	4,49	0,59	13,1
Dicke_S3	Platendicke Spur 3	Endfertigung	mm	8,03	0,07	0,9
Dicke_S4	Plattendicke Spur 4	Endfertigung	mm	8,00	0,06	0,7
Dicke_S5	Plattendicke Spur 5	Endfertigung	mm	8,03	0,07	0,9
u_Halle	Feuchte Produktionshalle	Endfertigung	%	24,49	6,18	25,2
B_AL_D1	Band Wegwert Auslauf oben links	Presse	mm	239,53	2,07	0,9
B_AL_D2	Band Wegwert Auslauf oben rechts	Presse	mm	243,09	1,71	0,7
B_AL_D3	Band Wegwert Auslauf unten links	Presse	mm	274,71	3,46	1,3
B_AL_D4	Band Wegwert Auslauf unten rechts	Presse	mm	277,75	3,92	1,4
B_AL_DR1	Band Druckwert Auslauf oben links	Presse	bar	103,45	7,38	7,1
B_AL_DR2	Band Druckwert Auslauf oben rechts	Presse	bar	122,01	5,31	4,3
B_AL_DR3	Band Druckwert Auslauf unten links	Presse	bar	109,50	6,78	6,2
B_AL_DR4	Band Druckwert Auslauf unten rechts	Presse	bar	117,68	5,71	4,9
B_AL_V_O	Bandverlauf Auslauf oben	Presse	mm	4,88	3,08	63,1
B_AL_V_U	Bandverlauf Auslauf unten	Presse	mm	5,22	4,51	86,3
DIST_A_L	Auslauf Wegistwert links	Presse	mm	7,65	0,19	2,5
DIST_A_R	Auslauf Wegistwert rechts	Presse	mm	7,64	0,19	2,5
DIST_E_L	Einlauf Wegistwert links	Presse	mm	9,76	0,35	3,5
DIST_E_R	Einlauf Wegistwert rechts	Presse	mm	9,94	0,34	3,5
DIST_L1	Distanz Rahmen 1 links	Presse	mm	9,80	0,34	3,5
DIST_L2	Distanz Rahmen 2 links	Presse	mm	9,74	0,32	3,3
DIST_L3	Distanz Rahmen 3 links	Presse	mm	9,77	0,29	3,0
DIST_L4	Distanz Rahmen 4 links	Presse	mm	9,95	0,24	2,5
DIST_L5	Distanz Rahmen 5 links	Presse	mm	10,15	0,21	2,0
DIST_L6	Distanz Rahmen 6 links	Presse	mm	10,21	0,19	1,9
DIST_L7	Distanz Rahmen 7 links	Presse	mm	10,25	0,18	1,7
DIST_L8	Distanz Rahmen 8 links	Presse	mm	10,24	0,18	1,7
DIST_L9	Distanz Rahmen 9 links	Presse	mm	10,24	0,18	1,8
DIST_L10	Distanz Rahmen 10 links	Presse	mm	10,25	0,18	1,8
DIST_L11	Distanz Rahmen 11 links	Presse	mm	10,24	0,19	1,8
DIST_L12	Distanz Rahmen 12 links	Presse	mm	10,24	0,19	1,8
DIST_L13	Distanz Rahmen 13 links	Presse	mm	10,21	0,20	1,9
DIST_L14	Distanz Rahmen 14 links	Presse	mm	10,22	0,21	2,0
DIST_L15	Distanz Rahmen 15 links	Presse	mm	10,19	0,20	2,0
DIST_L16	Distanz Rahmen 16 links	Presse	mm	10,18	0,23	2,3
DIST_L17	Distanz Rahmen 17 links	Presse	mm	9,95	0,23	2,4
DIST_L18	Distanz Rahmen 18 links	Presse	mm	9,55	0,36	3,7

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
DIST_L19	Distanz Rahmen 19 links	Presse	mm	8,54	0,37	4,3
DIST_L20	Distanz Rahmen 20 links	Presse	mm	7,68	0,24	3,1
DIST_L21	Distanz Rahmen 21 links	Presse	mm	7,60	0,18	2,4
DIST_L22	Distanz Rahmen 22 links	Presse	mm	7,59	0,17	2,3
DIST_L23	Distanz Rahmen 23 links	Presse	mm	7,59	0,18	2,3
DIST_L24	Distanz Rahmen 24 links	Presse	mm	7,59	0,17	2,3
DIST_L25	Distanz Rahmen 25 links	Presse	mm	7,59	0,18	2,3
DIST_L26	Distanz Rahmen 26 links	Presse	mm	7,59	0,18	2,3
DIST_L27	Distanz Rahmen 27 links	Presse	mm	7,67	0,19	2,5
DIST_R1	Distanz Rahmen 1 rechts	Presse	mm	9,80	0,34	3,5
DIST_R2	Distanz Rahmen 2 rechts	Presse	mm	9,81	0,35	3,6
DIST_R3	Distanz Rahmen 3 rechts	Presse	mm	9,85	0,33	3,4
DIST_R4	Distanz Rahmen 4 rechts	Presse	mm	10,03	0,30	3,0
DIST_R5	Distanz Rahmen 5 rechts	Presse	mm	10,23	0,26	2,6
DIST_R6	Distanz Rahmen 6 rechts	Presse	mm	10,31	0,26	2,5
DIST_R7	Distanz Rahmen 7 rechts	Presse	mm	10,35	0,25	2,5
DIST_R8	Distanz Rahmen 8 rechts	Presse	mm	10,34	0,26	2,5
DIST_R9	Distanz Rahmen 9 rechts	Presse	mm	10,32	0,24	2,3
DIST_R10	Distanz Rahmen 10 rechts	Presse	mm	10,32	0,24	2,3
DIST_R11	Distanz Rahmen 11 rechts	Presse	mm	10,24	0,19	1,8
DIST_R12	Distanz Rahmen 12 rechts	Presse	mm	10,20	0,18	1,8
DIST_R13	Distanz Rahmen 13 rechts	Presse	mm	10,23	0,20	2,0
DIST_R14	Distanz Rahmen 14 rechts	Presse	mm	10,24	0,21	2,0
DIST_R15	Distanz Rahmen 15 rechts	Presse	mm	10,23	0,21	2,0
DIST_R16	Distanz Rahmen 16 rechts	Presse	mm	10,19	0,24	2,3
DIST_R17	Distanz Rahmen 17 rechts	Presse	mm	9,90	0,23	2,3
DIST_R18	Distanz Rahmen 18 rechts	Presse	mm	9,55	0,36	3,7
DIST_R19	Distanz Rahmen 19 rechts	Presse	mm	8,52	0,36	4,3
DIST_R20	Distanz Rahmen 20 rechts	Presse	mm	7,68	0,24	3,1
DIST_R21	Distanz Rahmen 21 rechts	Presse	mm	7,81	0,18	2,3
DIST_R22	Distanz Rahmen 22 rechts	Presse	mm	7,81	0,17	2,2
DIST_R23	Distanz Rahmen 23 rechts	Presse	mm	7,80	0,17	2,2
DIST_R24	Distanz Rahmen 24 rechts	Presse	mm	7,81	0,17	2,2
DIST_R25	Distanz Rahmen 25 rechts	Presse	mm	7,81	0,17	2,2
DIST_R26	Distanz Rahmen 26 rechts	Presse	mm	7,81	0,17	2,2
DIST_R27	Distanz Rahmen 27 rechts	Presse	mm	7,87	0,19	2,5
DR_AL_L	Druck Auslauf links	Presse	bar	105,89	9,02	8,5
DR_AL_R	Druck Auslauf rechts	Presse	bar	111,48	11,12	10,0
DR_E_L	Druck Einlauf links	Presse	bar	167,13	14,63	8,8
DR_E_R	Druck Einlauf rechts	Presse	bar	168,78	11,16	6,6
DR_L1	Druck Rahmen 1 links	Presse	bar	164,16	13,32	8,1
DR_L2	Druck Rahmen 2 links	Presse	bar	158,89	10,49	6,6
DR_L3	Druck Rahmen 3 links	Presse	bar	143,46	11,62	8,1
DR_L4	Druck Rahmen 4 links	Presse	bar	124,38	11,24	9,0
DR_L5	Druck Rahmen 5 links	Presse	bar	106,71	10,58	9,9
DR_L6	Druck Rahmen 6 links	Presse	bar	93,25	9,20	9,9

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
DR_L7	Druck Rahmen 7 links	Presse	bar	90,42	7,96	8,8
DR_L8	Druck Rahmen 8 links	Presse	bar	79,78	8,03	10,1
DR_L9	Druck Rahmen 9 links	Presse	bar	89,76	8,89	9,9
DR_L10	Druck Rahmen 10 links	Presse	bar	88,07	8,23	9,3
DR_L11	Druck Rahmen 11 links	Presse	bar	87,88	9,01	10,3
DR_L12	Druck Rahmen 12 links	Presse	bar	84,65	9,16	10,8
DR_L13	Druck Rahmen 13 links	Presse	bar	81,69	8,30	10,2
DR_L14	Druck Rahmen 14 links	Presse	bar	76,36	8,60	11,3
DR_L15	Druck Rahmen 15 links	Presse	bar	77,36	7,99	10,3
DR_L16	Druck Rahmen 16 links	Presse	bar	73,17	8,54	11,7
DR_L17	Druck Rahmen 17 links	Presse	bar	75,85	8,02	10,6
DR_L18	Druck Rahmen 18 links	Presse	bar	88,91	14,50	16,3
DR_L19	Druck Rahmen 19 links	Presse	bar	121,15	16,15	13,3
DR_L20	Druck Rahmen 20 links	Presse	bar	152,50	12,90	8,5
DR_L21	Druck Rahmen 21 links	Presse	bar	140,05	11,62	8,3
DR_L22	Druck Rahmen 22 links	Presse	bar	140,10	10,22	7,3
DR_L23	Druck Rahmen 23 links	Presse	bar	139,18	10,64	7,6
DR_L24	Druck Rahmen 24 links	Presse	bar	140,41	10,46	7,4
DR_L25	Druck Rahmen 25 links	Presse	bar	136,26	10,42	7,6
DR_L26	Druck Rahmen 26 links	Presse	bar	134,37	9,73	7,2
DR_L27	Druck Rahmen 27 links	Presse	bar	140,55	13,96	9,9
DR_R1	Druck Rahmen 1 rechts	Presse	bar	164,33	12,49	7,6
DR_R2	Druck Rahmen 2 rechts	Presse	bar	158,16	11,34	7,2
DR_R3	Druck Rahmen 3 rechts	Presse	bar	143,18	11,92	8,3
DR_R4	Druck Rahmen 4 rechts	Presse	bar	122,63	11,34	9,2
DR_R5	Druck Rahmen 5 rechts	Presse	bar	107,05	10,50	9,8
DR_R6	Druck Rahmen 6 rechts	Presse	bar	94,36	9,13	9,7
DR_R7	Druck Rahmen 7 rechts	Presse	bar	89,00	8,06	9,1
DR_R8	Druck Rahmen 8 rechts	Presse	bar	79,49	7,75	9,7
DR_R09	Druck Rahmen 9 rechts	Presse	bar	87,20	8,58	9,8
DR_R10	Druck Rahmen 10 rechts	Presse	bar	86,82	8,02	9,2
DR_R11	Druck Rahmen 11 rechts	Presse	bar	88,83	9,22	10,4
DR_R12	Druck Rahmen 12 rechts	Presse	bar	81,48	8,56	10,5
DR_R13	Druck Rahmen 13 rechts	Presse	bar	79,82	7,96	10,0
DR_R14	Druck Rahmen 14 rechts	Presse	bar	76,39	7,93	10,4
DR_R15	Druck Rahmen 15 rechts	Presse	bar	72,00	7,27	10,1
DR_R16	Druck Rahmen 16 rechts	Presse	bar	71,01	7,93	11,2
DR_R17	Druck Rahmen 17 rechts	Presse	bar	74,99	7,75	10,3
DR_R18	Druck Rahmen 18 rechts	Presse	bar	86,98	14,11	16,2
DR_R19	Druck Rahmen 19 rechts	Presse	bar	121,21	15,94	13,2
DR_R20	Druck Rahmen 20 rechts	Presse	bar	151,69	12,98	8,6
DR_R21	Druck Rahmen 21 rechts	Presse	bar	136,79	11,24	8,2
DR_R22	Druck Rahmen 22 rechts	Presse	bar	136,20	10,12	7,4
DR_R23	Druck Rahmen 23 rechts	Presse	bar	137,77	10,42	7,6
DR_R24	Druck Rahmen 24 rechts	Presse	bar	136,23	10,40	7,6
DR_R25	Druck Ranmen 25 rechts	Presse	bar	135,89	10,28	7,6

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
DR_R26	Druck Rahmen 26 rechts	Presse	bar	131,39	9,40	7,2
DR_R27	Druck Rahmen 27 rechts	Presse	bar	127,74	13,42	10,5
KVA_DI_L	Keilverdichter Auslauf Distanz links	Presse	mm	24,32	3,70	15,2
KVA_DI_R	Keilverdichter Auslauf Distanz rechts	Presse	mm	24,35	3,72	15,3
KVA_DR_L	Keilverdichter Auslauf Druck links	Presse	bar	98,87	33,27	33,6
KVA_DR_R	Keilverdichter Auslauf Druck rechts	Presse	bar	108,11	30,09	27,8
KVA_SDR	Keilverdichter Auslauf spez. Druck	Presse	N/mm²	1,10	0,36	33,2
KVE_DI_L	Keilverdichter Einlauf Distanz links	Presse	mm	52,41	6,28	12,0
KVE_DI_R	Keilverdichter Einlauf Distanz rechts	Presse	mm	52,34	6,27	12,0
KVE_DR_L	Keilverdichter Einlauf Druck links	Presse	bar	121,98	32,03	26,3
KVE_DR_R	Keilverdichter Einlauf Druck rechts	Presse	bar	108,99	28,48	26,1
KVK_DI_R	Winkel Keilverdichter rechts	Presse	٥	2,23	0,17	7,4
MDR_R2	Druck Multipot Rahmen 2	Presse	bar	97,16	15,82	16,3
MDR_R3	Druck Multipot Rahmen 3	Presse	bar	88,53	13,48	15,2
MDR_R4	Druck Multipot Rahmen 4	Presse	bar	73,41	11,32	15,4
MDR_R5	Druck Multipot Rahmen 5	Presse	bar	65,16	8,54	13,1
MDR_R6	Druck Multipot Rahmen 6	Presse	bar	57,44	7,32	12,7
MDR_R21	Druck Multipot Rahmen 21	Presse	bar	113,80	11,43	10,0
MDR_R23	Druck Multipot Rahmen 23	Presse	bar	115,97	12,12	10,5
MDR_R25	Druck Multipot Rahmen 25	Presse	bar	112,54	11,86	10,5
TEMP_H1	Temperatur Heizkreis 1	Presse	°C	257,08	1,75	0,7
TEMP_H2	Temperatur Heizkreis 2	Presse	°C	255,38	1,62	0,6
TEMP_H3	Temperatur Heizkreis 3	Presse	°C	253,25	2,34	0,9
TEMP_H4	Temperatur Heizkreis 4	Presse	°C	238,85	3,60	1,5
TEMP_H5	Temperatur Heizkreis 5	Presse	°C	219,40	2,22	1,0
TEMP_H6	Temperatur Heizkreis 6	Presse	°C	195,07	1,83	0,9
TEMP_ROL	Temperatur Rollstangen	Presse	°C	251,56	3,65	1,5
V_PRESSE	Pressengeschwindigkeit	Presse	mm/s	516,50	15,63	3,0
PZF	Presszeitfaktor	Presse	s/mm	8,69	0,29	3,3
STR_CBVu	Stromaufnahme Bandantrieb unten	Vorpresse	А	72,64	8,72	12,0
STR_CBVo	Stromaufnahme Bandantrieb oben	Vorpresse	А	65,41	19,82	30,3
DR_HZ_L	Dickenregelung Auslauf links Hauptzylinder	Vorpresse	bar	123,17	10,24	8,3
DR_RUE_L	Dickenregelung Auslauf links Rückzug	Vorpresse	bar	96,61	7,61	7,9
DR_HZ_R	Dickenregelung Auslauf rechts Hauptzylinder	Vorpresse	bar	129,09	8,44	6,5
DR_RUE_R	Dickenregelung Auslauf rechts Rückzug	Vorpresse	bar	88,62	10,68	12,1
DR_EL_HZ	Dickenregelung Einlauf Rückzug	Vorpresse	bar	118,21	6,68	5,7
DR_EL_RU	Dickenregelung Einlauf Hauptzylinder	Vorpresse	bar	49,33	6,02	12,2
EL_VV	Distanz Einlauf Vorverdichter	Vorpresse	mm	323,87	25,30	7,8
VV_O_L	Druck Vorverdichter oben links	Vorpresse	bar	35,78	4,28	12,0
VV_O_R	Druck Vorverdichter oben rechts	Vorpresse	bar	37,12	4,83	13,0
VV_U_L	Druck Vorverdichter unten links	Vorpresse	bar	25,26	2,79	11,0
VV_U_R	Druck Vorverdichter unten rechts	Vorpresse	bar	34,97	2,47	7,0
CBV_O_L	Druck Vorpresse (CBV) oben links	Vorpresse	bar	67,08	26,40	39,4
CBV_O_R	Druck Vorpresse (CBV) oben rechts	Vorpresse	bar	45,08	13,61	30,2
CBV_U_L	Druck Vorpresse (CBV) unten links	Vorpresse	bar	67,47	19,27	28,6

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
CBV_U_R	Druck Vorpresse (CBV) unten rechts	Vorpresse	bar	73,49	13,63	18,5
V_FORM	Geschwindigkeit Formband	Formstrang	mm/s	515,83	15,20	2,9
u_Matte	Feuchte Fasermatte	Formstrang	%	9,69	0,40	4,2
Breite_I	Breite Fasermatte	Formstrang	mm	2798,93	49,89	1,8
Skalper_2	Höhe Abkämmwalze	Formstrang	mm	146,48	11,64	7,9
FL_Waage	Flächenwaage	Formstrang	kg	8,28	0,17	2,1
FG_Gew_2	Flächengewichtsmessung BWQ	Formstrang	kg/m²	7,37	0,11	1,5
DZ_SB2	Drehzahl Siebband	Formstrang	Hz	26,40	0,93	3,5
DZ_DB2	Drehzahl Dosierband	Formstrang	Hz	54,34	4,68	8,6
TEMP_STB	Temperatur Streubunker	Formstrang	°C	44,05	2,78	6,3
FUEL_STB	Füllstand Streubunker	Formstrang	bit	82,14	7,23	8,8
Faser	Durchsatz Faser	Beleimung	kg/min	484,48	18,52	3,8
Leim1	Durchsatz Leim	Beleimung	kg/min	129,29	5,07	3,9
Wasser	Durchsatz Waser	Beleimung	kg/min	22,80	1,53	6,7
Haert	Durchsatz Härter	Beleimung	kg/min	0,84	0,72	86,0
Harnst	Durchsatz Harnstoff	Beleimung	kg/min	7,99	1,75	21,9
Paraf	Durchsatz Paraffin	Beleimung	kg/min	7,31	0,46	6,3
TEMP_FAB	Temperatur Faserbunker	Trockner/Sichter	°C	44,61	3,72	8,3
TEMP_SI	Temperatur Sichter	Trockner/Sichter	°C	14,93	5,83	39,1
TEMP_TRA	Temperatur Trocknerausgang	Trockner/Sichter	°C	68,43	2,42	3,5
TEMP_TRE	Temperatur Trocknereingang	Trockner/Sichter	°C	189,24	15,89	8,4
TEMP_L_G	Temperatur Luft nach Ventilator	Trockner/Sichter	°C	20,53	7,80	38,0
TEMP_RE1	Temperatur Register 1	Trockner/Sichter	°C	34,14	4,64	13,6
TEMP_RE2	Temperatur Register 2	Trockner/Sichter	°C	20,72	8,11	39,1
TEMP_RE3	Temperatur Register 3	Trockner/Sichter	°C	88,62	9,41	10,6
u_SI	Luftfeuchte Sichter	Trockner/Sichter	%	23,21	5,56	24,0
u_TRO	Faserfeuchte Trockner	Trockner/Sichter	%	12,18	0,88	7,2
POS_KI	Position Anfahrklappe	Trockner/Sichter	%	86,20	8,60	10,0
FUEL_FAB	Füllstand Faserbunker	Trockner/Sichter	bit	39,87	9,93	24,9
STR_TR_V	Strom Trocknerventilator	Trockner/Sichter	А	121,62	2,95	2,4
V1_D_ABS	Dampfregelventil 1 absolut	Trockner/Sichter	bar	98,05	4,39	4,5
V2_D_ABS	Dampfregelventil 2 absolut	Trockner/Sichter	bar	17,82	0,80	4,5
POS_ABV	Position Ausblasventil	Refiner	%	94,50	8,31	8,8
VEN_KO	Ventilstellung Kocher	Refiner	%	62,19	3,88	6,2
VIB_REF	Vibration Refiner	Refiner	%	18,94	5,74	30,3
TEMP_MR	Temperatur Mahlraum	Refiner	°C	166,4	2,16	1,3
DR_Mahls	Kolbendruck Mahlspalt	Refiner	bar	114,6	6,45	5,6
DR_Mahlr	Dampfdruck Mahlraum	Refiner	bar	7,74	0,50	6,5
DDR_KOCH	Differentialdruck Kocher	Refiner	bar	0,84	0,46	55,1
SP_DD	Sperrdampfdruck	Refiner	bar	9,66	0,51	5,2
DDR_MAHL	Differentialdruck Mahlraum	Refiner	bar	1,95	0,05	2,6
DD_KO	Dampfdruck	Refiner	bar	8,59	0,24	2,8
Dampfm	Dampfmenge Mahlraum	Refiner	kg/h	298,3	11,42	3,8
DAMPF_KO	Dampfmenge Kocher	Refiner	kg/h	7986,7	1251	15,7
SP_DAM	Sperrdampfmenge	Refiner	kg/h	149,0	6,55	4,4
DAMPF_GE	Dampfmenge gesamt	Refiner	kg/h	13051,6	1181	9,1

Codierung	Variable	Bereich	Einheit	Mean	STD	CV
STR_REF	Stromaufnahme Refiner	Refiner	А	474,48	47,51	10,0
STR_BSCH	Stromaufnahme Beschickschnecke	Refiner	А	35,50	2,60	7,3
STR_SCH	Stromaufnahme Austragsschnecke	Refiner	А	96,80	6,94	7,2
STR_RUEW	Stromaufnahme Rührwerk	Refiner	А	67,37	4,90	7,3
STR_SH_1	Stromaufnahme Stopfschnecke M1	Refiner	А	362,82	26,15	7,2
STR_SH_2	Stromaufnahme Stopfschnecke M2	Refiner	А	363,72	25,92	7,1
DZ_SCHN	Drehzahl Austragsschnecke	Refiner	1/min	23,38	1,39	6,0
DZ_SCH_1	Drehzahl Stopfschnecke M1	Refiner	1/min	74,15	4,08	5,5
DZ_SCH_2	Drehzahl Stopfschnecke M2	Refiner	1/min	73,63	4,08	5,5
FUEL_KO	Füllstand Kocher	Refiner	bit	60,08	3,81	6,3
FUEL_HKS	Füllstand Hackschnitzelbunker	Refiner	bit	45,62	7,80	17,1
DA_HKS	Dampfmenge Hackschnitzelbunker	Refiner	kg/h	4186,1	1903	45,5
VEN_HKS	Ventilstellung Hackschnitzelbunker	Refiner	%	31,29	11,86	37,9
TEMP_HKS	Temperatur Hackschnitzelbunker	Refiner	°C	89,40	14,05	15,7
V_SCHN_1	Austragsgeschwindigkeit Schnecke 1	Holzplatz	%	61,74	30,10	48,7
V_SCHN_2	Austragsgeschwindigkeit Schnecke 2	Holzplatz	%	61,86	30,98	50,1
FUEL_S1	Füllstand Silo 1	Holzplatz	bit	60,38	19,93	33,0
FUEL_S2	Füllstand Silo 2	Holzplatz	bit	69,17	20,84	30,1
FUEL_S3	Füllstand Silo 3	Holzplatz	bit	59,73	21,38	35,8
u_AUS	Luftfeuchte außen	Umgebung	%	75,16	18,58	24,7

	Signifikanz des Korrelationskoeffizienten				
Freiheitsgrade	95 %	99 %	99,9 %		
	Zuf	allshöchstwerte v	on r		
1	1,00	1,00	1,00		
2	0,95	0,99	1,00		
3	0,88	0,96	0,99		
4	0,81	0,92	0,97		
5	0,75	0,87	0,95		
6	0,71	0,83	0,93		
7	0,67	0,80	0,90		
8	0,63	0,77	0,87		
9	0,60	0,74	0,85		
10	0,58	0,71	0,82		
11	0,55	0,68	0,80		
12	0,53	0,66	0,78		
13	0,51	0,64	0,76		
14	0,50	0,62	0,74		
15	0,48	0,61	0,73		
16	0,47	0,59	0,71		
17	0,46	0,58	0,69		
18	0,44	0,56	0,68		
19	0,43	0,55	0,67		
20	0,42	0,54	0,65		
25	0,38	0,49	0,60		
30	0,35	0,45	0,55		
35	0,33	0,42	0,52		
40	0,30	0,39	0,49		
45	0,29	0,37	0,47		
50	0,27	0,35	0,44		
60	0,25	0,33	0,41		
70	0,23	0,30	0,38		
80	0,22	0,28	0,36		
90	0,21	0,27	0,34		
100	0,20	0,25	0,32		
120	0,18	0,23	0,30		
150	0,16	0,21	0,26		
200	0,14	0,18	0,23		
300	0,11	0,15	0,19		

A 2: Zufallshöchstwerte des Korrelationskoeffizienten

Zahl der Freiheitsgrade und Zufallshöchstwerte des Korrelationskoeffizienten nach Sachs (1968)

Faktorenmuster M	Faktorenmuster Methode Varimax Druck rechts							
	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5			
DIST_A_L	0,26	0,79	0,17	0,34	0,25			
DIST_L27	0,27	0,79	0,17	0,33	0,25			
DIST_L26	0,48	0,80	0,25	0,19	0,11			
DIST_L25	0,49	0,80	0,24	0,19	0,11			
DIST_L24	0,48	0,81	0,24	0,19	0,11			
DIST_L23	0,48	0,80	0,25	0,20	0,12			
DIST_L22	0,50	0,77	0,24	0,24	0,13			
DIST_L21	0,47	0,76	0,21	0,25	0,13			
DIST_L20	0,12	0,37	0,32	0,36	0,59			
DIST_L19	0,14	0,09	0,02	0,01	0,96			
DIST_L18	0,35	0,12	0,00	-0,03	0,89			
DIST_L17	0,71	0,24	0,11	0,07	0,63			
DIST_L16	0,70	0,33	0,11	0,24	0,50			
DIST_L15	0,74	0,45	0,14	0,27	0,35			
DIST_L14	0,74	0,47	0,16	0,29	0,27			
DIST_L13	0,73	0,49	0,18	0,29	0,26			
DIST_L12	0,82	0,45	0,22	0,17	0,15			
DIST_L11	0,83	0,44	0,22	0,17	0,16			
DIST_L10	0,82	0,40	0,23	0,26	0,15			
DIST_L9	0,82	0,40	0,23	0,27	0,16			
DIST_L8	0,62	0,44	0,24	0,55	0,18			
DIST_L7	0,48	0,42	0,29	0,66	0,18			
DIST_L6	0,39	0,36	0,28	0,77	0,06			
DIST_L5	0,28	0,34	0,33	0,81	0,04			
DIST_L4	0,18	0,27	0,62	0,65	-0,03			
DIST_L3	0,20	0,20	0,87	0,30	-0,03			
DIST_L2	0,13	0,15	0,94	0,20	0,00			
DIST_L1	0,18	0,20	0,93	0,11	0,13			
DIST_E_L	0,17	0,21	0,92	0,10	0,16			

Faktorenmuster Methode Varimax Druck links							
	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5		
DR_L27	0,76	0,22	0,23	0,14	0,10		
DR_L26	0,82	0,34	0,20	0,23	0,28		
DR_L25	0,83	0,33	0,20	0,24	0,25		
DR_L24	0,83	0,35	0,20	0,22	0,28		
DR_L23	0,83	0,38	0,19	0,21	0,25		
DR_L22	0,82	0,36	0,24	0,19	0,24		
DR_L21	0,77	0,41	0,27	0,07	0,25		
DR_L20	0,61	0,21	0,21	0,20	0,20		
DR_L19	0,22	0,10	0,09	0,93	0,02		
DR_L18	0,19	0,21	0,06	0,93	0,03		
DR_L17	0,33	0,61	0,06	0,63	0,20		
DR_L16	0,39	0,54	0,26	0,57	0,14		
DR_L15	0,50	0,69	0,29	0,36	0,11		
DR_L14	0,49	0,69	0,35	0,29	0,07		
DR_L13	0,47	0,74	0,35	0,14	0,06		
DR_L12	0,42	0,79	0,29	0,16	0,23		
DR_L11	0,39	0,81	0,26	0,14	0,21		

DR_L10	0,36	0,70	0,40	0,25	0,28
DR_L09	0,31	0,79	0,34	0,15	0,21
DR_L8	0,34	0,55	0,65	0,18	0,23
DR_L7	0,31	0,44	0,74	0,22	0,19
DR_L6	0,22	0,43	0,80	0,11	0,19
DR_L5	0,20	0,27	0,89	0,08	0,17
DR_L4	0,29	0,12	0,75	-0,04	0,43
DR_L3	0,25	0,27	0,35	0,03	0,80
DR_L2	0,25	0,09	0,26	0,02	0,87
DR_L1	0,40	0,21	0,12	0,16	0,76

Faktorenmuster Methode Varimax Druck rechts							
	Faktor1	Faktor2	Faktor3	Faktor4	Faktor5		
DR_R27	0,74	0,16	0,26	0,23	0,01		
DR_R26	0,84	0,32	0,19	0,20	0,27		
DR_R25	0,85	0,27	0,18	0,23	0,28		
DR_R24	0,85	0,33	0,18	0,20	0,24		
DR_R23	0,85	0,37	0,18	0,15	0,24		
DR_R22	0,83	0,38	0,23	0,12	0,23		
DR_R21	0,80	0,37	0,29	0,02	0,25		
DR_R20	0,56	0,32	0,16	0,17	0,15		
DR_R19	0,20	0,08	0,09	0,93	0,05		
DR_R18	0,18	0,23	0,07	0,92	0,05		
DR_R17	0,32	0,62	0,07	0,63	0,22		
DR_R16	0,39	0,54	0,29	0,55	0,14		
DR_R15	0,48	0,70	0,32	0,35	0,07		
DR_R14	0,48	0,68	0,37	0,29	0,04		
DR_R13	0,46	0,75	0,34	0,15	0,00		
DR_R12	0,42	0,80	0,29	0,13	0,17		
DR_R11	0,42	0,82	0,17	0,10	0,11		
DR_R10	0,30	0,64	0,52	0,22	0,27		
DR_R9	0,28	0,75	0,38	0,15	0,20		
DR_R8	0,30	0,49	0,72	0,18	0,20		
DR_R7	0,28	0,36	0,80	0,22	0,17		
DR_R6	0,19	0,36	0,84	0,05	0,17		
DR_R5	0,17	0,21	0,89	0,06	0,19		
DR_R4	0,29	0,05	0,74	-0,01	0,44		
DR_R3	0,25	0,24	0,38	0,02	0,80		
DR_R2	0,23	-0,04	0,32	0,08	0,86		
DR_R1	0,39	0,30	0,12	0,15	0,71		

Faktorenmuster Methode Equamax Distanz links							
	Faktor1	Faktor2	Faktor3	Faktor4			
DIST_A_L	0,72	0,34	0,29	0,29			
DIST_L27	0,73	0,34	0,29	0,29			
DIST_L26	0,83	0,37	0,31	0,26			
DIST_L25	0,83	0,38	0,30	0,26			
DIST_L24	0,83	0,37	0,30	0,26			
DIST_L23	0,83	0,38	0,30	0,27			
DIST_L22	0,78	0,42	0,30	0,28			
DIST_L21	0,77	0,41	0,28	0,28			
DIST_L20	0,24	0,22	0,43	0,54			
DIST_L19	0,01	-0,06	0,07	0,93			
DIST_L18	0,10	0,05	0,01	0,95			
DIST_L17	0,31	0,36	0,10	0,84			
DIST_L16	0,38	0,51	0,13	0,71			
DIST_L15	0,51	0,57	0,16	0,57			
DIST_L14	0,54	0,60	0,19	0,50			
DIST_L13	0,55	0,59	0,21	0,48			
DIST_L12	0,58	0,57	0,20	0,44			
DIST_L11	0,56	0,57	0,21	0,45			
DIST_L10	0,51	0,64	0,22	0,43			
DIST_L9	0,51	0,64	0,23	0,43			
DIST_L8	0,44	0,74	0,33	0,34			
DIST_L7	0,38	0,74	0,40	0,28			
DIST_L6	0,30	0,79	0,41	0,12			
DIST_L5	0,25	0,76	0,48	0,05			
DIST_L4	0,18	0,56	0,73	-0,03			
DIST_L3	0,17	0,28	0,90	0,02			
DIST_L2	0,12	0,16	0,95	0,04			
DIST_L1	0,18	0,09	0,93	0,19			
DIST_E_L	0,18	0,08	0,92	0,21			

A 4: Rotierte Faktorenmuster der Methode Equamax für 4 PCA-Faktoren

Faktorenmuster Methode Equamax spez. Druck							
	Faktor1	Faktor2	Faktor3	Faktor4			
SDR_R27	0,64	0,27	0,26	0,25			
SDR_R26	0,79	0,27	0,33	0,37			
SDR_R25	0,80	0,21	0,33	0,38			
SDR_R24	0,82	0,24	0,33	0,35			
SDR_R23	0,83	0,26	0,30	0,34			
SDR_R22	0,81	0,30	0,30	0,34			
SDR_R21	0,79	0,35	0,16	0,35			
SDR_R20	0,55	0,20	0,26	0,29			
SDR_R19	0,05	0,03	0,94	0,11			
SDR_R18	0,09	0,08	0,96	0,05			
SDR_R17	0,40	0,34	0,78	0,12			
SDR_R16	0,36	0,48	0,67	0,21			
SDR_R15	0,52	0,61	0,50	0,14			
SDR_R14	0,51	0,68	0,44	0,13			
SDR_R13	0,54	0,70	0,30	0,08			
SDR_R12	0,54	0,69	0,31	0,19			
SDR_R11	0,53	0,67	0,30	0,13			
SDR_R10	0,33	0,68	0,40	0,37			
SDR_R9	0,37	0,72	0,33	0,23			
SDR_R8	0,23	0,77	0,31	0,41			
SDR_R7	0,17	0,74	0,28	0,48			
SDR_R6	0,10	0,79	0,15	0,46			
SDR_R5	0,03	0,72	0,09	0,55			
SDR_R4	0,12	0,48	-0,01	0,74			
SDR_R3	0,24	0,24	0,11	0,82			
SDR_R2	0,17	0,02	0,09	0,90			
SDR_R1	0,45	-0,01	0,24	0,63			

A 5: Kennzahlen des PLS-Modells mit 3 Komponenten für die Dickenquellung

	Gewicht	Gewicht	Gewicht	erklärte	stand.	VIP
	w1	w2	w3	Varianz	Regkoeff.	
Haert	0,29	0,19	0,18	26,2	0,19	2,07
Faser	-0,27	0,18	0,00	70,4	-0,06	1,98
DZ_SCHN	-0,27	0,19	-0,03	69,3	-0,06	1,93
	-0,27	0,18	-0,04	69,3	-0,07	1,93
STR_CBVu	-0,28	0,01	-0,16	39,3	-0,13	1,89
Wasser	-0,27	0,13	-0,13	67,0	-0,09	1,88
REMARK	-0,25	-0,18	0,00	22,5	-0,14	1,84
V_PRESSE	-0,16	0,28	0,19	58,8	0,03	1,51
DAMPF_GE	-0,19	0,14	0,24	44,9	-0,01	1,41
STR_CBVo	-0,19	0,14	-0,18	50,3	-0,07	1,41
B_AL_D3	-0,18	0,17	-0,12	52,1	-0,05	1,41
KVK_DI_R	0,20	0,05	-0,16	18,8	0,07	1,38
FG_Gew_2	-0,11	-0,29	0,00	33,2	-0,12	1,36
u_Matte	-0,20	0,07	0,28	38,9	-0,02	1,35
Paraf	-0,17	0,17	0,16	33,9	0,00	1,32
DR_RUE_L	0,17	-0,18	0,16	54,0	0,05	1,31
B_AL_DR1	-0,17	-0,13	-0,01	10,0	-0,10	1,22
EL_CBV	0,17	-0,10	0,09	28,8	0,05	1,21
Skalper_2	-0,06	-0,30	-0,01	34,6	-0,10	1,19
Breite_I	0,12	0,22	0,20	7,0	0,13	1,19
TEMP_H4	0,08	0,26	0,12	20,3	0,11	1,14
TEMP_STB	0,16	0,10	-0,12	13,0	0,07	1,11
FUEL_STB	0,15	0,09	-0,02	11,4	0,08	1,05
DR_AL_L	-0,15	0,06	-0,05	15,7	-0,05	1,04
STR_BSCH	-0,10	0,20	0,22	30,4	0,04	1,02
u_TRO	-0,14	-0,06	0,11	10,2	-0,05	0,99
DR_R2	0,13	0,10	-0,14	19,8	0,06	0,98
DIST_L6	0,14	-0,02	0,15	19,1	0,07	0,96
DIST_A_L	0,14	0,02	0,10	10,7	0,07	0,95
SDR_R1	0,13	0,10	-0,06	16,9	0,07	0,94
B_AL_DR2	0,13	0,08	0,00	5,9	0,07	0,92
MDR_25	-0,10	-0,15	0,04	11,1	-0,07	0,89
TEMP_RE3	0,07	0,20	0,43	26,7	0,14	0,88
MKDR_23	-0,07	-0,20	0,23	48,0	-0,04	0,88
MDR_R21	-0,11	-0,12	0,06	8,7	-0,06	0,86
DD_KO	-0,12	0,06	0,03	11,1	-0,03	0,86
MKDR_21	-0,05	-0,21	0,23	52,6	-0,04	0,85
VEN_KO	-0,12	-0,03	0,15	30,3	-0,03	0,85
TEMP_H6	0,11	0,11	0,15	16,0	0,09	0,85
MKDR_25	-0,06	-0,19	0,24	48,3	-0,04	0,84
MDR_R23	-0,09	-0,15	0,06	10,8	-0,06	0,84
DAMPF_KO	-0,12	-0,03	0,21	31,7	-0,02	0,82
VV_U_R	0,08	0,15	0,25	5,4	0,11	0,81
SDR_R2	0,11	0,08	-0,18	21,8	0,04	0,80
DIST_L5	0,12	-0,04	0,08	16,4	0,05	0,80
FUEL_FAB	0,09	0,11	0,08	5,3	0,07	0,73
Harnst	0,09	-0,11	0,24	40,4	0,04	0,70
TEMP_ROL	-0,09	-0,09	0,10	6,9	-0,04	0,66
TEMP_H1	-0,09	0,07	0,24	8,6	0,02	0,66
FL_Waage	0,09	0,04	0,05	3,0	0,05	0,59
DR_L14	-0,04	0,14	-0,06	17,4	0,01	0,59
TEMP_H2	-0,09	0,00	-0,01	5,7	-0,04	0,58
DR_R16	0,05	0,12	-0,07	9,3	0,04	0,55
DR_R14	-0,01	0,14	-0,09	16,6	0,02	0,54
DR_L16	0,03	0,12	-0,04	8,1	0,03	0,50
TEMP_FAB	0,06	0,07	-0,05	3,4	0,03	0,45
DR_R15	-0,02	0,11	-0,13	14,1	0,00	0,42
TEMP_SI	0,06	0,04	-0,05	2,7	0,02	0,41
TEMP_RE2	0,06	0,04	-0,05	2,6	0,02	0,40
SDR_R16	0,01	0,07	-0,14	8,2	0,00	0,28