Setzkasten Vergangenheit – ein kombinatorischer Ansatz für regionale Klimasimulationen

Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften im Department Geowissenschaften der Universität Hamburg

vorgelegt von

Boris Orlowsky

aus Bonn

Hamburg

2007

Als Dissertation angenommen vom Department Geowissenschaften der Universität Hamburg

Auf Grund der Gutachten von Prof. Dr. Klaus Fraedrich

und Prof. Dr. Friedrich-Wilhelm Gerstengarbe

Hamburg, den 11. April 2007

Prof. Dr. Kay-Christian Emeis Leiter des Department Geowissenschaften

Inhaltsverzeichnis

1	Ein	Einleitung: Zugänge zur Klimamodellierung 1						
	1.1	Der Königsweg – die direkte Modellierung der Physik						
		1.1.1	Allgemeine Zirkulationsmodell	3				
		1.1.2	Regionale Klimamodelle	4				
	1.2	Statistische Klimamodelle						
		1.2.1	Einige Ansätze der statistischen Klimamodellierung	8				
2	Das	Verfahren 13						
	2.1	Übersicht						
		2.1.1	Vorüberlegungen zu den Randbedingungen	13				
		2.1.2	Die Fragestellung	15				
		2.1.3	Die Herangehensweise	16				
		2.1.4	Die Herangehensweise für regionale Simulationen	24				
	2.2	Der Einstationen-Fall						
		2.2.1	Die Fragestellung	27				
		2.2.2	Der Ansatz	30				
		2.2.3	Der Algorithmus	32				
	2.3	Die Erweiterung auf mehrere Stationen						
		2.3.1	Die erweiterte Fragestellung	43				
		2.3.2	Der erweiterte Ansatz	44				
		2.3.3	Der erweiterte Algorithmus	45				
3	Kali	ibrierung und Begrenzung des Verfahrens						
	3.1	Kalibı	rierung	51				
		3.1.1	Wahl der Blocklänge	51				
		3.1.2	$\epsilon_{\rm jd}$ – die Toleranz beim julianischen Datum $\ldots \ldots \ldots$	54				
		3.1.3	Sonstige Parameter	58				
	3.2	Begre	nzungen	60				
4	Vali	dierun	ıg des Verfahrens	63				
	4.1	Aufbau des Experiments						
	4.2	Vorbe	ereitung des Experiments	66				

	4.3	3 Auswertungen des Ensembles				
		4.3.1	Klassische Statistiken	. 69		
		4.3.2	Persistenz	. 73		
		4.3.3	Extreme	. 76		
		4.3.4	Die simulierten Jahresgänge	. 80		
		4.3.5	Zusammenfassung	. 81		
	4.4	4.4 Auswertung einer ausgewählten Simulation				
		4.4.1	Auswertungen für die Temperatur	. 84		
		4.4.2	Auswertung für den Niederschlag	. 92		
		4.4.3	Zusammenfassung	. 100		
	4.5 Untersuchung der geforderten Eigenschaften					
		4.5.1	Regressionsparameter	. 103		
		4.5.2	Jahresgang	. 103		
		4.5.3	Persistenz	. 108		
	4.6	Vergle	eich mit REMO	. 109		
		4.6.1	Einige Vergleiche für die Temperatur	. 110		
		4.6.2	Einige Vergleiche für die trockenen Tage	. 110		
		4.6.3	Zusammenfassung	. 112		
5	Die klimatische Zukunft des Elbe-Einzugsgebiets 11					
	5.1	Vorbe	reitung der Simulationen	. 115		
	5.2	Die Si	mulationen zur ECHAM4-Vorgabe	. 117		
	5.3	Auswe	ertung der ausgewählten Simulationen	. 119		
		5.3.1	Die Zukunft der Temperatur	. 119		
		5.3.2	Die Zukunft des Niederschlags	. 122		
		5.3.3	Die Zukunft der klimatischen Wasserbilanz	. 132		
	5.4	Vergle	eich der EH4- und EH5-Simulationen	. 135		
	5.5	Zusan	nmenfassung	. 143		
G	7.1.0		nforoung und Aushlick	145		
0	Zus	amme	massung und Ausbrick	145		
A	Mat	hemat	ische Begriffe und Notationen	149		
	A.1	Allger	neine mathematische Konzepte	. 149		
	A.2	Statist	tische Konzepte	. 151		
В	Stat	Statistische Tests				
	B.1 Der <i>t</i> -Test für den Vergleich zweier Mittelwerte		Test für den Vergleich zweier Mittelwerte	. 155		
	B.2	Tests	für lineare Trends	. 157		
	B.3	Der K	olmogorov-Smirnov-Test	. 158		
	B.4	B.4 Der KS-Test für den Vergleich zweier Spektren				
	B.5	Der W	/ilcoxon- oder Mann-Whitney-Test	. 160		

Clus	steranalyse 163			
C.1	Verfahren der Clusteranalyse			
	C.1.1 Hierarchische Verfahren	164		
	C.1.2 Nicht-hierarchische Verfahren	164		
C.2	Bemerkungen zu den Verfahren			
	C.2.1 Wahl des Distanzmaßes	166		
	C.2.2 Repräsentation der Klassen	168		
	C.2.3 Abhängigkeit von der Starteinteilung	169		
	C.2.4 Abhängigkeit von der Anordnung	169		
C.3	Eine Kombination der beiden Ansätze	170		
STL		73		
011	-			
Der	: Datensatz des Elbe-Einzugsgebiets			
E.1	Probleme mit meteorologischen Daten			
E.2	Fehlererkennung und Vervollständigung			
	E.2.1 Fehlerhafte Messungen	178		
	E.2.2 Datenergänzung	179		
E.3	Homogenitätsanalyse und Homogenisierung	180		
	Clus C.1 C.2 C.3 STL Der E.1 E.2 E.3	Clusteranalyse I C.1 Verfahren der Clusteranalyse I C.1.1 Hierarchische Verfahren I C.1.2 Nicht-hierarchische Verfahren I C.2 Bemerkungen zu den Verfahren I C.2.1 Wahl des Distanzmaßes I C.2.2 Repräsentation der Klassen I C.2.3 Abhängigkeit von der Starteinteilung I C.2.4 Abhängigkeit von der Anordnung I C.3 Eine Kombination der beiden Ansätze I Der Datensatz des Elbe-Einzugsgebiets I E.1 Probleme mit meteorologischen Daten I E.2 Fehlererkennung und Vervollständigung I E.2.1 Fehlerhafte Messungen I E.2.2 Datenergänzung I E.3 Homogenitätsanalyse und Homogenisierung I		

Kapitel 1

Einleitung: Zugänge zur Klimamodellierung

Der Begriff "Klima" ist nicht leicht zu fassen. Schönwiese (2003) gibt ihn als "Statistische Beschreibung der Wetterelemente über eine relativ lange Zeit" an – grob gesprochen versteht man also darunter die über einen längeren Zeitraum typische Charakteristik des Wetters für einen bestimmten Ort oder eine Region. Eine solche Charakteristik beinhaltet mittlere Zustände wie Jahresmittel oder Jahresgang der Lufttemperatur, außerdem die Schwankungsbreiten um diese mittleren Zustände oder die Neigung zu bestimmten extremen Witterungsereignissen. Klima und Leben auf der Erde hängen wesentlich miteinander zusammen – das Klima bestimmt, wo und wann welche Lebensformen Bedingungen vorfinden, unter denen sie bestehen können, das Leben beeinflusst seinerseits bis zu einem gewissen Grad das Klima, verändert es zu seinem Vor- oder Nachteil.

Insgesamt für den Menschen nachteilige Auswirkungen zeigen seine eigenen Eingriffe in das System Erde seit der im 19. Jahrhundert einsetzenden Industrialisierung. Mit der Veränderung der chemischen Zusammensetzung der Atmosphäre geht eine Verstärkung des Treibhauseffekts einher. In der Folge steigt die globale Temperatur mit davor nicht beobachteter Geschwindigkeit, Wüsten wachsen mit zum Teil verheerenden Konsequenzen für ihre Anwohner und immer mehr spricht dafür, dass in diesem Zusammenhang extreme Ereignisse wie Stürme, Hitzewellen oder Überschwemmungen zunehmen.

Neben den auf der ganzen Erde zu beobachtenden Folgen wie der zunehmenden Erwärmung und einer Intensivierung des Wasserkreislaufs unterscheiden sich die konkreten Auswirkungen von Region zu Region erheblich. Zusätzlich zu einer einheitlichen Politik, die auf eine Gesamtentlastung des Klimas hinarbeitet, müssen daher Anpassungs- und Vermeidungsstrategien individuell für jede betroffene Gegend entwickelt werden. Dieses Anliegen wird beispielsweise im GLOWA-Elbe-Projekt (**GLO**baler Wandel des **WA**sserkreislaufs)¹ verfolgt, in dem diese Arbeit angesiedelt ist.

Untersuchungsgegenstand ist hierbei das gesamte Elbe-Einzugsgebiet, eine Region, in der sich die Wasserverfügbarkeit für mehrere Bereiche wie Land- und Forstwirtschaft oder Schiffsverkehr nahe der unteren Toleranzgrenze bewegt. Die Zielsetzung des Projekts ist daher die

Entwicklung integrierter Strategien zur nachhaltigen Bewältigung von durch den globalen Wandel bedingten Wasserverfügbarkeitsproblemen und Wassernutzungskonflikten und den daraus resultierenden Umwelt- und sozio-ökonomischen Problemen im Elbe-Einzugsgebiet.

Da der Klimawandel als eine der wesentlichen Triebkräfte für die Entwicklungen des Wasserhaushalts im Elbe-Einzugsgebiet der nächsten Jahrzehnte anzusehen ist, ist für die Untersuchung dieser Fragestellungen zunächst eine Abschätzung der zu erwartenden klimatischen Entwicklungen notwendig. Gegenstand meiner Arbeit ist daher ein Verfahren, das eben solche Abschätzungen ermöglicht.

Herkömmlicherweise werden solche Abschätzungen mit Klimamodellen angegangen, die ganz allgemein als vereinfachende Abbildungen des Wetter- und Klimageschehens auf physikalische Gleichungen oder statistische Zusammenhänge und deren Implementierung auf Computern verstanden werden können.

Klimamodelle liefern keine Prognosen wie etwa die Modelle der Wettervorhersage – viele externe Antriebskräfte, die auf der Zeitskala der Wettervorhersage als konstant angenommen werden können, folgen auf längeren Zeitskalen Entwicklungen, die heute nicht bekannt, für die Dynamik des Klimas jedoch relevant sind (beispielsweise die zukünftigen CO₂-Emissionen). Simulationsergebnisse beziehen sich daher notwendigerweise auf Annahmen, die für solche Antriebskräfte zugrunde gelegt werden. Die Annahmen, über die weltweit der größte Konsens besteht, sind sicherlich die vom Intergovernmental **P**anel of **C**limatic **C**hange (IPCC) entwickelten SRES-Szenarien², die plausible Entwicklungen für die wichtigsten anthropogenen Klimaeinflussfaktoren für das 21. Jahrhundert beschreiben (Houghton et al., 2001).

Zu den nicht einheitlich zu lösenden Problemen bei der Klimamodellierung gehört die enorme Bandbreite der raumzeitlichen Skalen im Klimageschehen.³

¹ Siehehttp://www.glowa.org/de/elbe_uebersicht.htm.

 $^{^2}$ Special Report on Emissions Scenarios, ein Bericht, den das IPCC 1996 in Auftrag gegeben hat.

³ Allein in der Atmosphäre reichen diese Skalen von wenigen Stunden und Quadratkilometern, auf denen man beispielsweise Gewitter beobachtet, über mehrere Tage der Lebensdauer eines Tiefdruckgebiets, das sich über hunderte Kilometer erstreckt zu stabilen Hochdruckgebieten, wie sie über Wochen hinweg über Sibirien stehen.

Inzwischen gibt es daher eine ganze Reihe von zum Teil sehr unterschiedlichen Modellansätzen, die in der Regel für einen bestimmten Skalenbereich entwickelt wurden. Einige wichtige von ihnen werden in den folgenden Absätzen vorgestellt.

1.1 Der Königsweg – die direkte Modellierung der Physik

Dynamische Ansätze versuchen, alle bekannten Prozesse des Klimageschehens zu erfassen, in Gleichungen zu beschreiben und diese Beschreibung in Computerprogramme zu übertragen. Das untersuchte Gebiet wird dabei diskretisiert, d. h. in Gitterzellen zerlegt – ebenso wie die zeitlichen Entwicklungen in einzelne Zeitschritte zerlegt werden.

1.1.1 Allgemeine Zirkulationsmodell

Sozusagen ganz oben in der Hierarchie der dynamischen Modelle stehen die allgemeinen Zirkulationsmodelle oder General Circulation Models (GCM), die in der Lage sind, das Geschehen auf dem gesamten Planeten zu simulieren.

Angetrieben werden diese Modelle durch vorgegebene äußere Randbedingungen für den Strahlungshaushalt der Erde, etwa die schwankende Energiemenge, die von der Sonne die Erde erreicht, außerdem Vulkanausbrüche und, was die Zeit seit der Industrialisierung betrifft, die vom Menschen verursachte Veränderung der Zusammensetzung der Atmosphäre.

Der Preis für diesen globalen Anspruch ist eine relativ grobe Auflösung – üblicherweise wie beim globalen Zirkulationsmodell ECHAM5 (Roeckner et al., 2003), das am *European Centre for Medium-Range Weather Forecasts* (ECMWF, Europäisches Zentrum für mittelreichweitige Wettervorhersage) und am Max-Planck-Institut für Meteorologie in Hamburg entwickelt wurde,⁴ in der Größenordnung von 100 km horizontal. Die typische Länge eines Zeitschritts beträgt ca. eine halbe Stunde, für feinere Auflösungen und kürzere Zeitschritte werden die Anforderungen an die Rechenleistung zu hoch. Durch diese grobe Rasterung der Prozesse in Raum und Zeit ist es unmöglich, kleinskalige Vorgänge direkt im Modell abzubilden. Sie werden indirekt durch eine sogenannte Parameterisierung

Neben dem atmosphärischen Geschehen bestimmen zahlreiche weitere Systeme das Klima mit, etwa die Vegetation, die für den Kohlenstoffkreislauf des Planeten und seine Albedo von großer Bedeutung ist und die sich teilweise auf Zeitskalen von Jahrhunderten entwickelt. Ferner der Ozean, der bereits nahe seiner Oberfläche Monate benötigen kann, um sich einer eventuell veränderten klimatischen Umwelt anzupassen, in der Tiefe auch Jahrhunderte bis Jahrtausende. Ähnlich träge ist die Kryosphäre, deren Kontinentaleisschilde Reaktions- und Wirkungszeiten bis zu 10000 Jahren haben können. Gebirgsgletscher und Seeeis sind hingegen deutlich schnelllebiger. Siehe hierzu Ruddiman (2001).

⁴ Siehe http://www.ecmwf.int bzw. http://www.mpimet.mpg.de/. Der Modellname ECHAM setzt sich aus dem EC des ECMWF und aus HAM von Hamburg zusammen.

berücksichtigt, z. B. für die Wolkenbildung. Hierfür können statistische Beziehungen genutzt werden wie Zusammenhänge zwischen Temperatur und Feuchtefeldern, die vom GCM brauchbar gut dargestellt werden können, und der Wolkenbildung, die sich auf viel kleineren Skalen abspielt. Insgesamt versucht man, durch die Parameterisierung die mittleren Verhältnisse in dem Raumgebiet, das durch eine Gitterzelle überdeckt wird, zu erfassen. Diese gehen dann mit in den Gesamtzustand der Zelle zu einem bestimmten Zeitschritt ein und beeinflussen ihre Entwicklung und Umgebung entsprechend.

Die Parameterisierungen führen ein erhebliches Unsicherheitspotential in die Ergebnisse eines Modelllaufs ein: Die genauen Parameterwerte sind in der Regel unbekannt und müssen aus Beobachtungsdaten geschätzt werden, die ihrerseits nicht in allen Regionen und nicht für alle Zeit verfügbar sind. Die Schätzungen müssen daher auch auf Regionen oder Zeiträume übertragen werden, für die sie nicht unbedingt Gültigkeit besitzen. Auch die Modellierung und Diskretisierung der Prozesse innerhalb der Gitterzellen ist mit Unsicherheit verbunden. Als ein Beispiel hierfür zeigt Abbildung 1.1 Abschätzungen für die Änderungen des mittleren Luftdrucks in Meereshöhe über Europa für die Wintermonate bei einem gegenüber der Gegenwart verdoppelten CO₂-Gehalt der Atmosphäre. Benutzt wurde das GCM HadAM3 (Gordon et al., 2000) des Hadley-Centres in Großbritannien⁵ mit 53 verschiedenen Einstellungen der Parameterisierungen, die z. B. Wolkenbildung, Niederschlag und Konvektionsvorgänge betreffen (Murphy et al., 2004). Die in der Abbildung dargestellte Differenz bezieht sich für jede Simulation auf 20-jährige Mittel, die aus Gleichgewichtsläufen mit gegenwärtiger CO₂-Konzentration bzw. mit der verdoppelten Konzentration geschätzt wurden. Die Ergebnisse sind extrem unterschiedlich, obwohl die physikalischen Antriebe für alle Simulationen die gleichen waren.

Einige Größen, wie etwa der für die Klimatologie so wichtige Niederschlag, können aufgrund der Skalenproblematik in GCMs für Teilregionen der Erde wie Europa, Deutschland oder europäische Flusseinzugsgebiete nur unbefriedigend bestimmt werden – die Ursachen für Niederschlag sind oft zu kleinräumig, als dass ein GCM sie erfassen könnte.

1.1.2 Regionale Klimamodelle

Niederschlag und Wolkenbildung, für die zum Teil Prozesse auf räumlich sehr viel feineren Skalen relevant sind, können erst in Modellen aus der Familie der Regionalen Klimamodelle oder Regional Climate Models (RCM) genauer simuliert werden, deren Auflösung sich horizontal zwischen 10 und 50km bewegt. Bei einem auf diese Weise sehr viel punktreicheren Gitter werden wegen des entsprechend erhöhten Bedarfs an Rechnerressourcen Vorgänge für einzelne Regionen,

⁵ Siehe http://www.metoffice.gov.uk/research/hadleycentre/index.html.



-4 -2 -1 0 1 2 4 Change in MSLP (hPa)

Abbildung 1.1: Änderungen des über 20 Jahre gemittelten Luftdrucks im Winter über Europa auf Meeresniveau aus 53 Läufen des HadAM3-Modells für eine verdoppelte CO₂-Konzentration mit unterschiedlichen Parameterisierungen. Nach Mitchell und Ewins (2003).

nicht mehr für die ganze Erde, simuliert. Um ins globale Klimageschehen eingebettet zu bleiben, werden die RCMs an den Rändern des Simulationsgebiets mit großskaligen Daten aus GCM-Simulationen oder, wenn etwa Entwicklungen der Vergangenheit nachvollzogen werden sollen, mit tatsächlich gemessenen und analysierten Beobachtungsdaten angetrieben. Hinzu kommen wie bei den GCMs die für die interessierende Entwicklung bedeutenden äußeren Einflussfaktoren. Die im englischen als *nesting* bezeichnete Einbettung in das gröbere Gitter eines GCMs kann als Einweg-nesting betrieben werden – d. h., dass die Simulationsergebnisse des RCMs nicht in das GCM zurückgehen – oder als gekoppelte Einbettung, bei der die Abläufe im RCM nach jedem Zeitschritt in die großskaligeren Abläufe im GCM eingehen, siehe Lorenz und Jacob (2005).

Auch wenn diese Modelle die lokalen Verhältnisse bedeutend genauer darstellen können als ihre grobskaligeren Verwandten, zeigt Abbildung 1.2 als ein repräsentatives Beispiel für Nachsimulationen der Jahre 1979 bis 1993 in Deutschland immer noch deutliche Abweichungen von den Beobachtungsdaten. Die Abbildung entstammt Kotlarski et al. (2005) mit einer Validierung der Modelle MM5 (Grell et al., 1995), REMO (Jacob, 2001) und CLM (Böhm et al., 2006):⁶ Für

⁶ Die Abkürzungen stehen für *Mesoscale Model* (MM) des nordamerikanischen *National Center*



Abbildung 1.2: Abweichung der mittleren Jahresdurchschnittstemperatur (links) und der mittleren Jahressumme des Niederschlags (rechts) aus verschiedenen RCMs von den Daten des Deutschen Wetterdienstes, 1979–1993. Nach Kotlarski et al. (2005).

die langjährigen Temperaturmittel ergeben sich Abweichungen von bis zu 1.5K von den Beobachtungsdaten, Jahresniederschlagssummen werden um bis zu 70% falsch eingeschätzt. Ähnliche Ergebnisse zeigt die Validierungsstudie von Giorgi et al. (2004), in der langjährige Mittel von Temperatur und Niederschlag, deren Trends und ihre Variabilität im Vergleich von Beobachtung und Simulationen des RegCM-Modells⁷ (Giorgi et al., 1993) untersucht werden.

Auch bei den RCMs sind nicht alle wichtigen Prozesse direkt in physikalischen Gleichungen erfasst und werden über Parametrisierungen berücksichtigt, beispielsweise bei Konvektionsvorgängen. Die dabei auftretenden Unsicherheiten können zu den gezeigten Abweichungen beitragen, ebenso wie noch immer zu grobe Auflösung, die Verletzung der Masse- oder Energieerhaltung (insbesondere auf Skalen zwischen regionalen und kontinentalen Bereichen) sowie Genauigkeits- und Stabilitätsbeschränkungen bei der numerischen Implementierung der verwendeten Differentialgleichungen.

Trotz ihrer Ungenauigkeiten in der exakten Bestimmung klimatischer Parameter haben die dynamischen Modelle den großen Vorzug, dass sie das physikalische Geschehen direkt simulieren. Ein unmittelbarer Zugang zu den auftretenden Prozessen ist damit geschaffen, der es unter anderem ermöglicht, die relativen Wichtigkeiten einzelner Teilaspekte zu analysieren. Außerdem bieten sie die Möglichkeit, wenn die entsprechenden Randbedingungen von einem GCM

for Atmospheric Research, REgionalMOdell (REMO), wie die ECHAM-Modelle vom MPI in Hamburg, sowie für *Climate version des LokalModells* des Deutschen Wetterdienstes.

⁷ Das **Reg**ional **C**limate **M**odel vom A. Salam International Centre for Theoretical Physics in Triest, Italien

verfügbar sind, Entwicklungen in Form von Szenarien zu simulieren, die weit entfernt von heutigen klimatischen Verhältnissen sind und für die keine Messdaten vorliegen.

Charakteristisch für dynamische Modelle ist der für ihren Betrieb erforderliche hohe Arbeitsaufwand: Sie sind sehr anspruchsvoll an Rechenleistung, Entwicklung und Bedienung, wirkliche Mannschaftsprojekte, die ganze Teams für Entwicklung und Betrieb erfordern sowie teure und seltene Großrechenanlagen.

Verzichtet man auf die Differentialgleichungs-Physik und verlässt sich statt dessen auf Beobachtungsdaten, kommt man zu den statistischen Klimamodellen. Sie sind der Gegenstand der nun folgenden Betrachtungen.

1.2 Statistische Klimamodelle

Statistische Klimamodelle verwenden anstelle von physikalischen Gleichungen Zusammenhänge zwischen gemessenen Daten. Dadurch verzichten sie in gewisser Weise auf ein Verständnis der Natur der zugrunde liegenden Prozesse, erreichen jedoch unter Umständen eine größere Konsistenz zwischen beobachteten und simulierten Daten. Bei Nachsimulationen der bereits von Messwerten bedeckten Vergangenheit ist die mit statistischen Modellen erreichte Übereinstimmung in vielen Fällen von dynamischen Modellen unerreicht. Besonders für Folgeanwendungen wie Untersuchungen des Wasserhaushalts oder der Vegetation, wie sie im GLOWA-Elbe-Projekt stattfinden, sind daher statistische Modelle noch immer unersetzbar. Sie setzen allerdings die Verfügbarkeit von Daten voraus, eine Bedingung, die für Europa im Allgemeinen recht brauchbar erfüllt ist, in weiten Teilen der Welt jedoch nur mangelhaft gegeben ist und damit die Verwendung solcher Modelle dort einschränkt.

Darüber hinaus stützt sich die Nutzung statistischer Zusammenhänge auf die Annahme, dass diese Zusammenhänge tatsächlich stationär sind, das heißt dass das Klima der Zukunft dieselben statistischen Beziehungen aufweist wie das beobachtete. Das verhindert den Einsatz dieser Verfahren für die Simulation völlig vom beobachteten Klima verschiedener Klimatologien.

Ein wichtiger Vorteil statistischer Verfahren ist ihr geringer Preis: Sie sind sehr viel einfacher aufgebaut als die dynamischen, können problemlos von einzelnen Anwendern benutzt werden und benötigen nur einen Bruchteil der Rechenleistung, die ein dynamisches Regionalmodell beansprucht.

In dieser Arbeit möchte ich ein solches Verfahren vorstellen, das Klimasimulationen auf der Basis von Stationsdaten und entsprechender räumlicher Auflösung realisiert, dabei einfach zu bedienen und schnell ist und sich zur Erzeugung ganzer Ensembles von Simulationen eignet.

Bevor ich dieses Verfahrens detailliert beschreibe, möchte ich einen Überblick über ihm verwandte Ansätze der statistischen Klimamodellierung und des Herunterskalierens⁸ geben, ergänzt durch Verweise auf Anwendungsbeispiele. Ziel all dieser Verfahren ist die Erzeugung klimatologischer Simulationsreihen, die je nach Fragestellung bestimmte Eigenschaften aufweisen. Für ihre ausführliche Diskussion möchte ich allerdings auf die Literatur verweisen, z. B. in Zorita und von Storch (1999), von Storch et al. (2000), Wilby und Wigley (1997) und Wilby et al. (1998).

1.2.1 Einige Ansätze der statistischen Klimamodellierung

Wettergeneratoren. Die Grundidee eines Wettergenerators besteht darin, ein stochastisches Modell für eine interessierende Variable an einer Station zu definieren und mit diesem eine Zeitreihe von Zufallszahlen zu erzeugen, deren Eigenschaften denen einer beobachteten Reihe entsprechen. Für Niederschlagsreihen können z. B. in einem ersten Schritt mit einer Markov-Kette Abfolgen von trockenen und feuchten Tagen generiert werden. Die Abfolge könnte auch aus einem GCM bestimmt werden. Als zweiter Schritt wird dann für die feuchten Tage die Niederschlagsmenge aus einer Verteilung simuliert, die aus beobachteten Niederschlägen geschätzt wird (siehe z. B. Wilks, 1999). Eine auf solche Weise erzeugte Niederschlagsreihe besitzt realistische statistische Eigenschaften, etwa im Hinblick auf die Verteilung der täglichen Niederschläge.

Kompliziertere Wettergeneratoren, wie etwa in Semenov und Barrow (1997) beschrieben, sind in der Lage, nicht nur einzelne Größen, sondern auch ganze Tupel – Niederschlag, Temperatur, Sonnenscheindauer etc. – zu simulieren. Erreicht wird dies durch eine entsprechend komplexere Hierarchie der Bedingtheiten, welche die Gesamtheiten, aus denen für die einzelnen Variablen ein beobachteter Wert gezogen wird, so einschränken, dass für jeden Tag realistische Kombinationen verschiedener Größen gezogen werden. Solche Bedingheiten können nicht nur für die verschiedenen Messgrößen, sondern auch für räumliche Zusammenhänge verwendet werden, wenn Wettergeneratoren zur Erzeugung von Klimasimulationen für mehrere Stationen gleichzeitig benutzt werden sollen. Auch die Prediktoren⁹ können sehr viel ausgefeilter sein als einfach eine Abfolge trockener und feuchter Tage.

Üblicherweise simulieren Wettergeneratoren die Beobachtungsvariablen einzeln. Werden mehrere Variablen simuliert, können sich daher unter Umständen physikalisch sinnlose Kombinationen ergeben.

⁸ Im Englischen *downscaling*: Verfahren, die zeitlich oder räumlich grob aufgelöste Simulationen auf eine feinere Auflösung übertragen.

⁹ Ein Prediktor ist eine Größe, mit der sich das Verhalten einer anderen Größe vorhersagen lässt. Diese wird als Prediktant bezeichnet. Da sich beispielsweise an Stationen des Elbe-Einzugsgebiets eine hohe Korrelation zwischen Sonnenscheindauer und Globalstrahlung feststellen lässt, wäre die Sonnenscheindauer ein guter Prediktor für die Globalstrahlung.

Übertragungsfunktionen. Besonders für Anwendungen des Herunterskalierens (*downscaling*) werden häufig Übertragungsfunktionen verwendet, d. h. empirisch bestimmte funktionale Abhängigkeiten zwischen Prediktoren und Prediktanten (von Storch und Zwiers, 1999). Das einfachste Beispiel einer solchen Übertragungsfunktion ist ein durch eine lineare Regression bestimmter linearer Zusammenhang zwischen erklärenden und abhängigen Variablen, beispielsweise Beobachtungen an der Oberfläche in Abhängigkeit geeigneter atmosphärischer Zustandsvariablen (siehe Murphy, 1999 für den Vergleich eines solchen Ansatzes mit dynamischem Herunterskalieren). Als Prediktoren werden häufig aggregierende Maße von simulierten Feldern verwendet, wie sie z. B. eine Hauptkomponentenanalyse liefert (siehe z. B. Kidson und Thompson, 1998). Neben den interessierenden Variablen selbst können Modelle für ausgewählte Statistiken der Variablen spezifiziert werden, beispielsweise für Wiederkehrintervalle.

Zu den Übertragungsfunktionen gehören auch die verschiedensten Methoden der Interpolation, die dazu verwendet werden können, die von auf grober Auflösung arbeitenden Modellen simulierten Felder auf eine feinere räumliche Skala zu transferieren.

Allgemein tendieren mit diesen Verfahren erzeugte Simulationsreihen zu einer zu geringen Variabilität (Zorita und von Storch, 1999).

Wetterlagenbasierte Ansätze. Andere Ansätze stützen sich auf großräumige Wetterlagen, die z. B. für einen Simulationszeitraum aus einem GCM-Lauf extrahiert werden können. Sie versuchen, Zusammenhänge zwischen großräumigen Mustern und lokalem Witterungsgeschehen zu nutzen, um Letzteres zu simulieren, wenn nur Informationen über Erstere verfügbar sind. Dazu werden typischerweise beobachtete großskalige Druckverteilungen in Ähnlichkeitsklassen einsortiert. Die simulierten Druckverteilungen können dann z. B. über Distanzmaße den ihnen entsprechenden Klassen zugewiesen werden. Da zu den beobachteten Druckverteilungen auch das lokale Witterungsgeschehen bekannt ist, kann damit über die Klassenzuweisung der simulierten Druckverteilungen auch deren lokales Wetter simuliert werden – das Verfahren kann sich für eine gegebene Druckverteilung aus den Witterungszuständen, die in deren Klasse fallen, bedienen. Wetterlagen können außerdem als Prediktoren für Wettergeneratoren dienen oder mit für wetterlagenabhängig definierten Übertragungsfunktionen kombiniert werden (z. B. Enke und Spekat, 1997).

Analogverfahren. Dem eng verwandt sind die Analogverfahren, die z. B. in der Wettervorhersage (van den Dool, 1989), für Hurrikan-Prognosen (Fraedrich et al., 2003) oder für statistisches Herunterskalieren (Zorita und von Storch, 1999) benutzt werden. Sie versuchen, Informationslücken für einen simulierten Zustand zu schließen, indem sie nach ähnlichen Zuständen in der Menge der Beobachtungen suchen und dem simulierten Zustand fehlende Information aus dem ähnlichen beobachteten beifügen. Dadurch können die entstehenden Simulationsreihen nur aus Elementen zusammengesetzt werden, die in der Menge der vorhandenen Beobachtungen enthalten sind.

Etwa könnte man sich im Kontext des Herunterskalierens in der Situation finden, für einen zukünftigen Tag zwar Informationen über die großräumige Druckverteilung zu besitzen – etwa aus einem GCM-Lauf –, jedoch keinerlei Wissen über die räumlich sehr viel feiner differenzierten Wetterverhältnisse am Boden. Beim klassischen "Nächster-Nachbar"-Ansatz würde aus der Menge von beobachteten Druckverteilungen diejenige ausgewählt, welche der Druckverteilung des zukünftigen Tages am ähnlichsten ist. Die zugehörigen beobachteten Wetterverhältnisse können dann als Annäherung an die des zukünftigen Tages angenommen werden. Dieser Ansatz ist deterministisch, da zu jeder Reihe großräumiger Druckverteilungen genau eine passende Neuanordnung der Tage des Beobachtungszeitraums gehört.

Diese klassische Strategie wurde vielfach erweitert (z. B. in Beersma und Buishand, 2003; Diez et al., 2005; Timbal und McAvaney, 2001), unter anderem, indem nicht nur der allernächste, sondern die k nächsten Nachbarn bestimmt werden. Unter diesen geschieht die Auswahl bedingt auf weitere Prediktoren, wie die unmittelbaren Vorgänger in der Reihe, was die Erhaltungsneigung der Atmosphäre mit berücksichtigt.

Durch solche Erweiterungen werden die Analogverfahren stochastisch und erlauben die Erzeugung von Ensembles, die eine Abschätzung der Bandbreite der Entwicklungen ermöglichen. Prinzipiell ist der analoge Ansatz nicht auf einzelne Stationen beschränkt und kann für regionale Klimasimulationen verwendet werden, indem die kompletten täglichen Datensätze der Region entsprechend den Prediktoren neu angeordnet werden. Die entstehenden Reihen bestehen dadurch aus physikalisch konsistenten Elementen.

Wie die Analogverfahren erzeugt mein Verfahren regionale Klimasimulationen durch eine Neuanordnung oder Rekombination von Beobachtungen einzelner Tage aus dem Beobachtungszeitraum. Die Simulationsreihen sind dabei allerdings einzig durch die Parameter einer linearen Regressionsgeraden beschränkt, durch die sich die simulierten Jahresmittel einer ausgewählten Bezugsgröße beschreiben lassen sollen. Mit den klassischen *nearest-neighbour(s)*-Ideen der Analogverfahren, die im Prinzip tägliche großräumige Information in lokales Wetter übersetzen – weshalb die Qualität ihrer Simulationsergebnisse von der Verfügbarkeit und Güte der täglichen großräumigen Information abhängig ist¹⁰ – hat es also nichts gemein.

¹⁰ Diesen Nachteil teilen sie übrigens mit allen dynamischen Regional-Modellen, die in die Gitter globaler Zirkulationsmodelle eingebettet sind.

Diese Sparsamkeit in den Randbedingungen birgt Chance und Risiko: Risiko, weil die Simulationen dadurch extrem unterdeterminiert sind; Chance, weil sich die Bezugsgröße frei und insbesondere unabhängig von einzelnen GCM-Simulationen wählen lässt (beispielsweise können die über ein Ensemble von GCM-Läufen gemittelten Parameter vorgegeben werden) und damit weiteren Plausibilitätsprüfungen zugänglich ist.

Das Verfahren basiert auf den Grundideen des **STA**tistischen **R**egional-, kurz STAR-Modells (Werner und Gerstengarbe, 1997), das am Potsdam Institut für Klimafolgenforschung von F.-W. Gerstengarbe und P.-C. Werner betrieben wird und in mehreren Klimafolgen-Untersuchungen Anwendung fand (siehe z. B. Wechsung et al., 2005; Stock et al., 2005).

Kapitel 2 stellt das Verfahren vor. Nach der Kalibrierung des Verfahrens in Kapitel 3 beschreibe ich in Kapitel 4 ein Kreuzvalidierungsexperiment mit Daten des Elbe-Einzugsgebiets, bei dem ein Teil des Beobachtungszeitraums simuliert wird. Hierdurch lässt sich im Vergleich der simulierten Jahre mit den beobachteten die Qualität des Verfahrens beurteilen. Zuletzt folgt in Kapitel 5 die Auswertung von Simulationen, die für die nächsten Jahrzehnte des Elbe-Gebiets gerechnet wurden.

Kapitel 2

Das Verfahren

Für ein besseres Verständnis des hier vorgestellten Verfahrens sollen zunächst in Abschnitt 2.1 einige Vorüberlegungen zur Fragestellung und zum gewählten Ansatz angestellt werden, denen sich eine übersichtsartige Beschreibung des Verfahrens und seiner Abläufe anschließt. Abschnitt 2.2 greift diese Ausführungen, in mathematischen Begriffen präzisiert, nochmals auf, zuerst für Simulationen für einzelne Stationen, anschließend in Abschnitt 2.3 für Simulationen für ganze Regionen.

2.1 Übersicht

2.1.1 Vorüberlegungen zu den Randbedingungen

Um regionale Auswirkungen zukünftiger Klimaänderungen zu untersuchen, benötigt man räumlich und zeitlich hoch aufgelöste Simulationen dieser Änderungen. Dafür müssen in irgendeiner Form großräumige bestehende Annahmen über die klimatische Entwicklung einer Region – häufig aus den grob aufgelösten Feldern, die ein GCM für den interessierenden Zeitraum bestimmt – in räumlich hoch aufgelöstes Witterungsgeschehen übersetzt werden. Dabei dienen die grob aufgelösten Felder entweder als Antrieb für die auf höherer Auflösung arbeitenden RCMs oder als Prediktoren für statistische Modelle.

Ein Verfahren, das ein solches *downscaling* vornimmt, ist Gegenstand dieser Arbeit. Die von ihm erzeugten Simulationsreihen sollen als Grundlage für hydrologische Modellierungen verwendbar sein, was einen hohen Anspruch an Genauigkeit und Konsistenz darstellt. Ein weiteres wichtiges Anliegen bei der Entwicklung war, dass das Verfahren möglichst genügsam sein sollte, was Rechenzeit und Aufwand in der Bedienung angeht. Damit schieden alle dynamischen Ansätze aus (siehe Abschnitt 1.1.2), und es entstand ein Verfahren, das mit Beobachtungsreihen und deren statistischen Beziehungen untereinander arbeitet.

Die Ergebnisse aller Versuche, die räumlich nur grob aufgelösten Aussagen

der GCMs auf die regionale Skala zu übertragen – sei es dynamisch oder statistisch – hängen zwangsläufig von der Qualität der GCMs ab. Auch wenn deren Aussagen auf der globalen Skala ein kohärentes Gesamtbild ergeben, sind sie auf der regionalen bis kontinentalen Skala oft alles andere als eindeutig: Wie Abbildung 1.1 zeigt, hängen zum Beispiel die Druckfelder, die ein solches GCM für Europa berechnet, auch bei gleichen Randbedingungen (wie z. B. in diesem Fall eine gegenüber der Gegenwart verdoppelte CO₂-Konzentration), wesentlich von den internen Parameterisierungen der Prozesse ab (siehe auch Abschnitt 1.1.1). Die Druckfelder als Prediktoren für regionale Simulationen zu verwenden, erscheint also nicht unbedingt als beste Wahl. Es stellt sich daher die Frage, ob eventuell auf dieser hier interessierenden Skala andere klimatische Charakteristika von GCMs zuverlässiger simuliert werden.

In der Literatur finden sich zahlreiche Arbeiten mit Vergleichen von GCM-Ergebnissen und Reanalyse-Daten, z. B. Benestad (2004) für Temperaturtrends in Nordeuropa, Fox-Rabinovitz et al. (2006) für eine Validierung von Stretched-Grid-GCMs oder Ruiz-Barradas und Nigam (2006) für die Überprüfung simulierter Niederschläge über den USA. Im Rahmen des PRUDENCE-Projekts¹ (Déqué, 2004) wurden u. a. die von vier GCMs reproduzierten 30-jährigen Mittel 1961– 1990 für Temperatur und Niederschlag in Sommer und Winter verglichen.

Insgesamt ergibt sich aus der Literatur kein einheitliches Bild, wie gut die Übereinstimmung verschiedener Variablen mit beobachtungsbasierten Daten ist. Differenzen ergeben sich sowohl zwischen verschiedenen Modellen als auch zwischen verschiedenen Simulationen gleicher Modelle (Murphy et al., 2004), die Variablen werden je nach Jahreszeit unterschiedlich gut wiedergegeben, je nach Region ist die Übereinstimmung zwischen Simulation und Reanalyse besser oder schlechter. Hinzu kommt die Unsicherheit in den Reanalysen selbst und die Tatsache, dass die Flächenmittel, die im GCM für die einzelnen Gitterzellen bestimmt werden, etwas physikalisch fundamental anderes sind als die Punktbeobachtungen einer Station.

Exemplarisch sind in Tabelle 2.1 einige Temperatur-Charakteristika für das Klima in Deutschland dargestellt, entsprechend den Beobachtungen bzw. den Simulation der GCMs ECHAM4 und ECHAM5 für 1959–1990, angetrieben durch beobachtete CO₂-Konzentrationen. Wirklich gute Übereinstimmung zeigt sich bei keiner der Größen.

Diese Absätze machen deutlich, dass es offensichtlich nicht unproblematisch ist, Simulationsergebnisse von GCMs auf der regionalen Skala als Antrieb für das Herunterskalieren zu verwenden. Es lag daher nahe, ein Verfahren zu entwickeln, das einen möglichst einfachen Antrieb erhält, der auch weiteren Plausibilitätsüberlegungen zugänglich ist.

¹ PRUDENCE steht für das *Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects*-Projekt des Dänischen Meteorologischen Instituts.

MWT	emp. [°C]	Trend Temp. [K]	
beobachtet	simuliert	beobachtet	simuliert
8.13	8.49 (ECHAM4)	0.50	0.12 (ECHAM4)
8.13	8.52 (ECAHM5)	0.50	0.73 (ECAHM5)

Tabelle 2.1: Verschiedene Statistiken aus Kontrollläufen von ECHAM4 und ECHAM5. Verglichen sind die Mittelwerte der Temperatur und die linearen Zunahmen in den Jahresmitteln der Temperatur. Die Temperaturreihen entstanden dabei durch Mittelung über die für Deutschland relevanten Gitterzellen für die GCMs bzw. über die Klimastationen in Deutschland. Für den Vergleich mit ECHAM5 standen Monatswerte 1959–1990 (Wild, 2005) zur Verfügung, für ECHAM4 Jahreswerte von 1959–1990 (Österle, 2005).

2.1.2 Die Fragestellung

Als eine der simpelsten Formen von Entwicklung bekommt mein Verfahren die Parameter einer linearen Regressionsgeraden, sprich eine lineare Zu- oder Abnahme sowie den Mittelwert für eine ausgewählte Bezugsgröße vorgegeben. Sie soll die *langfristige* Entwicklung beschreiben und bezieht sich daher auf eine Glättung der Bezugsgröße (hier: auf ihre Jahresmittel). Die Bezugsgröße wird dabei so gewählt, dass sie für das Klima der untersuchten Region besonders charakteristisch ist – für das Elbe-Einzugsgebiet beispielsweise die Temperatur, siehe Abschnitt 3.1.1 und Kapitel 4.

Eine solche Vorgabe ist selbstverständlich alles andere als vollständig bestimmend für die Simulation der klimatischen Entwicklung, insbesondere für die anderen Variablen außer der Bezugsgröße. Viele untereinander sehr verschiedene Simulationen, die alle einer solchen Vorgabe genügen, sind denkbar – auch physikalisch unsinnige. Das Verfahren muss daher so angelegt werden, dass die Simulationen trotz dieser Unterbestimmtheit physikalisch sinnvoll sind. Beispielsweise

- müssen die Kombinationen der einzelnen simulierten Beobachtungsvariablen physikalisch plausibel sein – ein Tag mit Niederschlag, aber ohne Bewölkung z. B. darf nicht auftauchen;
- müssen die simulierten Reihen realistische Jahresgänge aufweisen die Sommer müssen z. B. wärmer als die Winter simuliert werden;
- müssen die Persistenzen der einzelnen Variablen in den Simulationsreihen realistisch sein, einem heißen Tag sollte kein unmittelbarer Frosteinbruch folgen. Hierbei interessiert vor allem die Persistenz in den um langfristigen Trend- und Jahresgangkomponente bereinigten Reihen diese Komponenten können wesentlich zur Erhaltungsneigung in einer Reihe beitragen. Die interessierende Persistenz entsteht z. B. bei der Temperatur

aus der thermischen Trägheit der Atmosphäre oder der Verweildauer bestimmter Wetterlagen;

 müssen bei regionalen Simulationen die räumlichen Verteilungen der Beobachtungsvariablen physikalisch konsistent sein – ein simulierter lauer Frühlingstag in Berlin etwa sollte keinen Frosteinbruch in Potsdam begleiten.

Ein Verfahren wie dieses, das zwar nur die Regressionsparameter für eine Bezugsgröße vorgegeben bekommt, gleichzeitig jedoch auch die angesprochenen Eigenschaften für alle Beobachtungsvariablen realistisch simulieren soll, muss sich hierfür an den Eigenschaften der Beobachtungsreihen orientieren: realistisch ist, was aus der Kenntnis der Beobachtungsreihen plausibel erscheint. Letztlich kann man also ein solches Verfahren als den Versuch einer Antwort auf die Frage: "Wie hätte das Witterungsgeschehen der Beobachtungsperiode aussehen können, hätte es eine den Regressionsparametern entsprechende Entwicklung in der Bezugsgröße durchlaufen?" verstehen. Dadurch weisen simulierte und beobachtete Klimatologie zwangsläufig eine gewisse Mindestähnlichkeit auf – eine Eigenschaft des Verfahrens, die sich grundsätzlich bei statistischen Ansätzen findet.

2.1.3 Die Herangehensweise

Um die angesprochenen Konsistenzen sicherzustellen, bedient sich das Verfahren einer wichtigen Annahme: Zeitliche Abschnitte der Beobachtungsreihen können während der Simulationsperiode erneut oder zumindest in sehr ähnlicher Weise auftreten. Das Verfahren setzt daher die Beobachtungen der Simulationsreihen aus Beobachtungen aus der Beobachtungsreihen zusammen (vergleiche Abbildung 2.1).

Die Frage ist nun, wie eine solche Zuweisung gefunden werden kann, d. h. wie Abschnitte der Beobachtungsperiode so neu aneinandergefügt werden können, dass die entstehenden Reihen die in 2.1.2 genannten Eigenschaften aufweisen. Das Verfahren beschreitet dazu einen heuristischen Weg, der für Simulationen für eine einzelne Station in Abbildung 2.2 zusammengefasst ist und aus zwei Schritten besteht, die auf unterschiedlichen Zeitskalen arbeiten: der erste auf der Skala von Jahren (in der Abbildung grau hinterlegt), der zweite auf der Skala von 12-Tages-Blöcken (hellblau hinterlegt). Die Beobachtungsreihe wird daher zu Beginn einmal nach Jahren (vom 1. Januar bis 31. Dezember), einmal nach gleitenden 12-Tages-Blöcken unterteilt. Sowohl mit dem Begriff "Jahr" als auch mit dem "12-Tages-Block" verbindet sich also eine Liste von Kalenderdaten und ihren zugehörigen Beobachtungen. Die Blöcke² der Beobachtungspe-

² Im Grunde sind auch Jahre Blöcke der Länge 365 bzw. 366 Tage. In dieser Arbeit beziehe ich



Abbildung 2.1: Illustration der Fragestellung und des Ansatzes: Gegeben ist eine Beobachtungsreihe (dargestellt durch eine synthetische Temperaturreihe in schwarz) und eine vorgegebene Regressionsgerade (rot) für den Simulationszeitraum (links). Gesucht ist eine simulierte Reihe (ebenfalls rot), die der Vorgabe genügt und aus Abschnitten der Beobachtungsreihe zusammengesetzt ist (rechts). Die Punkte zeigen jeweils die Jahresmittel an, auf die sich die Regressionsvorgabe bezieht.

riode, charakterisiert durch ihre Werte der Bezugsgröße, werden in der Vorbereitung durch eine Cluster-Analyse klassifiziert, die einander ähnliche Blöcke in Klassen zusammenfasst. Diese vorbereitenden datenorganisierenden Maßnahmen sind im oberen "Input"-Teil der Abbildung mit angedeutet. Darunter findet sich im "Verfahren"-Teil ein Flussdiagramm, dass die beiden Schritte des Verfahrens darstellt, wobei die Zeitskala des jeweiligen Schritts mit der hinterlegten Farbe angezeigt wird (grau für Jahre, blau für Blöcke). Rechts finden sich in Wort und Bild die Grundideen zu den einzelnen Schritten. Die nächsten Absätze werden diese, Bezug nehmend auf die Abbildung, erläutern.

Schritt 1: Erste Näherung

Im ersten Schritt werden jahresweise Abschnitte der Beobachtungsreihe neu aneinandergereiht. Ziel dieses Schritts ist, durch die Neureihung bereits möglichst nah an die vorgegebenen Regressionsparameter zu gelangen. Dazu wird eine große Stichprobe von zufälligen Reihungen jahresweiser Abschnitte erzeugt, deren Länge der des Simulationszeitraums entspricht. Für jede dieser zufälligen Reihungen wird die Regressionsgerade der Bezugsgröße bestimmt und schließlich diejenige ausgewählt, die den vorgegebenen Parametern am nächsten liegt (siehe hierzu außerdem Abbildung 2.3). Diese "erste Näherung" weist garantiert realistische Jahresgänge und Persistenzen auf, da sie aus Abfolgen innerhalb der Jahre besteht, die bereits einmal beobachtet wurden. Allerdings, auch wenn ihre Regressionsgerade nicht allzu weit von der Vorgabe entfernt ist, wird es im Allgemeinen nur selten gelingen, bereits mit der ersten Näherung die vorgegebene

mich jedoch mit dem Begriff "Block" stets auf die 12-Tages-Blöcke.





Das Verfahren

18



Abbildung 2.3: Die erste Näherung: Neuanordnung ganzer Jahre aus dem Beobachtungszeitraum 1981–1990 als erste Approximation an die Beobachtungsreihe des Simulationszeitraums 1991–2000, dargestellt für die Temperatur als Bezugsgröße. Ausgewählt wird aus einer großen Stichprobe zufälliger Neuaneinanderreihungen so, dass die Regressionsgerade durch die umsortierten Jahresmittel (grün) möglichst nah an der vorgegebenen (rot) liegt. Unter der Kurve der Simulationsperiode finden sich die umsortierten Jahreszahlen aus der Beobachtungsperiode.

Regressionsgerade nachzuvollziehen.³

Um die Vorgabe tatsächlich zu erreichen, muss diese erste Näherung nachgebessert werden. Von der Zeitskala ganzer Jahre als neu anzuordnender Abschnitte begibt sich die Nachbesserung auf die kürzere Skala von 12-Tages-Blöcken.

Schritt 2: Nachbesserung

Die Nachbesserung besteht darin, die sich aus der ersten Näherung ergebende Reihe Block für Block durchzugehen und für jeden Block zu entscheiden, ob er aus der ersten Näherung übernommen wird. Dazu wird ein heuristisches Kriterium verwendet, das, stark verkürzt gesagt, bewertet, ob ein Block aus der ersten Näherung einer künstlichen Reihe der Bezugsgröße, die die Regressionsvorga-

³ Eine zufällige Neuanordnung (Ziehen mit Zurücklegen) einer Reihe von Jahresmitteln wird einen Erwartungswert vom Mittel der Jahresmittel aufweisen sowie eine Regressionsgerade mit einem verschwindenden Erwartungswert der Steigung. Je verschiedener das zu simulierende Mittel vom beobachteten und je weiter die vorgeschriebene Steigung von 0 abweicht, desto unwahrscheinlicher wird es, eine solche Neureihung mit passenden Parametern zufällig zu ziehen.

be erfüllt, zu unähnlich ist. In diesem Fall muss ein passender Ersatz aus den Blöcken des Beobachtungszeitraums gesucht werden. Der Ersatzblock wird so ausgewählt, dass sich mit seinen Beobachtungen einerseits physikalisch plausible Witterungsabfolgen ergeben; andererseits soll er dazu beitragen, dass die auf diese Weise schrittweise entstehende Reihe der Bezugsgröße der Regressionsvorgabe genügt.

Hinter diesem Vorgehen des zweiten Schritts stehen folgende Überlegungen:

- Durch die Verwendung von Blöcken als zeitlichen Abschnitten sind realistische Witterungsabfolgen innerhalb der Blöcke garantiert – sie entstammen ja einem tatsächlich beobachteten Wettergeschehen. Die Länge von 12 Tagen orientiert sich dabei an typischen Erhaltungsneigungen von Größen wie Temperatur, Luftdruck und Niederschlag an Stationen des Elbe-Einzugsgebiets, siehe Tabelle 3.1 und Abschnitt 3.1.1. Ein bei der Auswahl passender Ersatzblöcke verwendetes heuristisches Kriterium verlangt außerdem, dass die Blocklänge gerade sein muss.
- 2. Durch das Beibehalten einiger Blöcke aus der ersten Näherung bleibt ein "Grundgerüst" der beobachteten Jahre erhalten. Dies trägt zu einer physikalisch plausiblen Witterungsabfolge innerhalb der einzelnen Jahre bei.
- 3. Die Art der Ersetzung der Blöcke, die nicht beibehalten werden, geschieht anhand von Kriterien, die für eine physikalisch plausible Witterungsabfolge in der entstehenden Reihe sorgen.

Schritt 2 a: Identifikation der auszutauschenden Blöcke aus der 1. Näherung. Für die Identifikation wird zunächst eine künstliche Reihe der Bezugsgröße, die den vorgegebenen Regressionsparametern für die Jahresmittel entspricht, erzeugt: Dazu werden aus den Jahresmitteln der ersten Näherung "Zieljahresmittel" bestimmt, die die simulierte Reihe der Bezugsgröße am Ende aufweisen soll. Die Zieljahresmittel entstehen dabei aus den Jahresmitteln der ersten Näherung, indem diese um die Differenz an der Stelle des jeweiligen Jahres zwischen vorgegebener Regressionsgeraden und der Regressionsgeraden aus der ersten Näherung erhöht werden. Siehe hierzu die Abbildung 2.4 oben.

Für jedes Kalenderjahr des Simulationszeitraums lässt sich mit diesen Zieljahresmitteln die künstliche Reihe erzeugen, indem die Reihe der ersten Näherung jahresweise so erhöht bzw. abgesenkt wird, dass ihr Mittelwert dem jeweiligen Zieljahresmittel entspricht. Damit stehen zwei Reihen bereit: Die der ersten Näherung, nicht ganz der vorgegebenen Regressionsgeraden entsprechend, und die künstliche, die durch jahresweise Anpassung der ersten entsteht und die Vorgabe genau trifft. Für ein Beispieljahr sind diese beiden Reihen – diejenige aus Schritt 1 (grün) und die aus dieser künstlich generierte (blau) in Abbildung 2.2 dargestellt.



Abbildung 2.4: Entscheidung, welche Blöcke aus der ersten Näherung beibehalten werden, mit der Temperatur als Bezugsgröße. Oben: in grün die Temperaturreihe der ersten Näherung, in blau die künstliche Zieltemperaturreihe. Unten links: Zoom auf das erste Jahr. Die Boxen begrenzen zwei Blöcke der Länge 12 Tage. Darüber ist die Klassenzugehörigkeiten der Blöcke der beiden Temperaturreihen farbig codiert. Rechts unten sind schematisch die fünf Klassen der Beobachtungsblöcke dargestellt. Der rechte markierte Block fällt sowohl bezüglich der Temperaturen aus der ersten Näherung als auch bezüglich der Zieltemperaturen in die gleiche Klasse und wird beibehalten. Für den anderen (grau hinterlegt) gilt das nicht, er wird neu gesetzt.

Die Grundidee bei der Entscheidung darüber, welche Blöcke aus der ersten Näherung beibehalten werden, ist nun, die Blöcke bezüglich dieser beiden Reihen miteinander zu vergleichen. Sind sich die Blöcke beider Reihen ähnlich genug, wird der Block der ersten Näherung beibehalten, andernfalls wird er ersetzt. Als Ähnlichkeitskriterium wird die in der Vorbereitung aus der Clusteranalyse erhaltene Klassifikation benutzt: Fällt der Block sowohl bezüglich der ersten Näherung als auch der künstlichen Reihe in die gleiche Klasse, wird er aus der ersten Näherung übernommen. Bezüglich der künstlichen Reihe werden die Blöcke dabei der Klasse zugeordnet, zu deren Schwerpunkt sie am nächsten liegen. Am Kopf der Illustration zu 2 a finden sich in Abbildung 2.2 farbig codiert die Klassenzugehörigkeiten der beiden Reihen. Die Blöcke, bei denen sie sich unterscheiden, sind durch die grau hinterlegten Rechtecke markiert – diese Blöcke werden in Schritt 2 b ausgetauscht. Detailliert ist dieses Vorgehen in Abbildung 2.4 illustriert.



Abbildung 2.5: Die schrittweise Einschränkung der "Kandidatenmenge" zum Neusetzen eines Blocks. Ganz oben sind nummeriert die Blöcke bezüglich der Bezugsgröße der ersten Näherung dargestellt. Darunter die Blöcke bezüglich der Zielreihe der Bezugsgröße (mit apostrophierter Nummer). In Klammern für beide Reihen in rot die Klassenzugehörigkeiten der Blöcke. Für den grün markierten dritten Block unterscheiden sich diese, er wird also nicht aus der ersten Näherung übernommen. Für ihn wird zufällig aus der in vier Schritten eingeschränkten Kandidatenmenge Ersatz ausgewählt.

Schritt 2 b: Austausch der Blöcke. Für die Blöcke, die nicht aus der ersten Näherung übernommen werden, muss Ersatz beschafft werden. Hierfür wird für jeden neu zu besetzenden Block in mehreren Schritten die Menge potentieller Ersatzblöcke aufgrund von Kriterien reduziert, die nur Blöcke in der Auswahl lassen, die sich bruchlos in die schon bestehenden Teile der Reihe (das Grundgerüst) einfügen.

Dieses Verfahren der schrittweisen Eingrenzung ist andeutungsweise bereits in Abbildung 2.2 und detailliert in Abbildung 2.5 illustriert: Dort sind ganz oben nummeriert die Blöcke der ersten Näherung schematisch dargestellt; darunter die Blöcke der künstlichen Reihe (mit apostrophierter Nummer). In Klammern ist für beide Reihen die Klassenzugehörigkeiten der Blöcke angegeben. Für den grün markierten dritten Block unterscheiden sich diese, er wird also nicht aus der ersten Näherung übernommen.

Die Menge, aus der der Ersatzblock ausgewählt wird, wird schrittweise eingeschränkt:

- 1. Zunächst besteht sie aus sämtlichen Blöcken des Beobachtungszeitraums, die in die selbe Klasse gehören wie der zu besetzende Block bezüglich der künstlichen Reihe (in der Abbildung 2.5 die Klasse 3).
- 2. Um nicht zufällig einzelne Blöcke gehäuft zu verwenden, werden von ihnen nur diejenigen übrig behalten, die noch nicht in der Simulationsreihe auftauchen.⁴
- Das nächste Kriterium schränkt die Menge weiter auf diejenigen Beobachtungsblöcke ein, deren julianisches Anfangsdatum⁵ nicht zu weit von dem des zu besetzenden Block entfernt ist. Dadurch wird sichergestellt, dass nicht jahreszeitlich unpassende Blöcke ausgewählt werden.
- 4. Das letzte Kriterium behält daraus nur diejenigen Beobachtungsblöcke übrig, die bruchlos an ihre Vorgänger, und, falls bereits vorhanden, Nachfolger anschließen. Für seine genaue Formulierung siehe Abschnitt 2.2.3 und besonders die Abbildung $2.6.^6$

Aus der so gefundenen "Kandidatenmenge" wird schließlich zufällig ausgewählt.

Es kann vorkommen, dass trotz der Nachbesserung die Regressionsvorgabe nicht erreicht wird: Soll beispielsweise eine Regressionsgerade verwirklicht werden, die einen höheren Mittelwert aufweist als die Regressionsgerade aus der ersten Näherung, so wird die Mehrheit der im Grundgerüst beibehaltenen Blöcke systematisch zu einem zu niedrigen Mittelwert beitragen. Gegebenenfalls kann dieser Beitrag durch den Austausch der Blöcke, die nicht übernommen werden, nicht ausgeglichen werden. In diesem Fall geht das Verfahren in eine Iterationsschleife, in der, ausgehend von der ersten Näherung, so oft nachgebessert wird, bis die Vorgabe erfüllt ist. Dabei werden für den nächsten Iterationsschritt die Regressionsparameter intern übertrieben, was sich in einer künstli-

⁴ Dies bedeutet nicht, dass einzelne Tage nicht mehrfach gesetzt werden könnten: Einerseits können ganze Blöcke mehrfach vorkommen, wenn beispielsweise in der ersten Näherung ein Jahr zweimal auftritt und ein Block daraus an beiden Stellen in der ersten Näherung die Selektion aus Schritt 2 a überlebt. Außerdem können einzelne Tage mehrfach auftreten, wenn Blöcke ausgewählt werden, die in der Beobachtungsreihe weniger als die Blocklänge voneinander entfernt sind – in diesem Fall tauchen die "Überlapptage" doppelt auf. Das Kriterium verhindert damit einerseits, dass einzelne Tage des Beobachtungszeitraums zu häufig zugewiesen werden, was die Variabilität der Simulationsreihen einschränken würde, erlaubt andererseits eine maßvolle Mehrfachverwendung, die, nehmen in der Simulationsperiode gegenüber der Beobachtungsperiode bestimmte Zustände zu, auch angemessen sein kann. Bei den Experimenten aus Kapitel 4 beträgt der Anteil der mehrfach gesetzten Tage typischerweise 5%.

⁵ Das julianische Datum eines Tages ist der zeitliche Abstand (in Tagen) des Tages vom 31. Dezember des Vorjahres. Es gibt die Position des Tages innerhalb des Jahres an. Das julianische Datum eines 1. Januars ist also immer 1, das eines 3. Februars immer 34 usw.. Das julianische Datum nimmt in Schaltjahren Werte von 1..366 an, sonst 1..365.

⁶ Die genaue Formulierung des Kriteriums verlangt gerade Blocklängen, siehe Abschnitt 2.2.3.

chen Reihe niederschlägt, die etwas weiter als im vorangegangenen Schritt von der Reihe der ersten Näherung abweicht. Dadurch werden weniger Blöcke aus der ersten Näherung übernommen und das Verfahren kann mehr Blöcke so auswählen, dass eine eventuelle durch das Grundgerüst bedingte systematische Abweichung ausgeglichen werden kann.

Insgesamt entsteht dadurch eine Zuordnung, die jedem Block der Simulationsperiode die Beobachtungen eines Blocks aus der Beobachtungsreihe zuweist.

2.1.4 Die Herangehensweise für regionale Simulationen

Die beschriebene Herangehensweise lässt sich praktisch unverändert auf Simulationen für mehrere Stationen übertragen – die Abläufe finden lediglich in einem höherdimensionalen Raum statt. Anstatt eines einzelnen Werts der Bezugsgröße werden die Werte von allen betrachteten Stationen verwendet, anstatt einer einzelnen Regressionsvorgabe wird eine Vorgabe für jede betrachtete Station benötigt. Die Clusteranalyse klassifiziert Blöcke, die nicht nur durch ihre Werte der Bezugsgröße an einer Station, sondern durch ihre Werte an allen Stationen charakterisiert werden. Ein Jahr der Beobachtungsperiode enthält neben den Kalenderdaten die Beobachtungen aller Stationen der Untersuchungsregion, genauso wie die Blöcke.

Diese Unterschiede in der Dimensionalität sind die einzigen.

Je nach Stationendichte und klimatischer Heterogenität des Untersuchungsgebiets kann es überflüssig sein, wirklich alle Stationsreihen an der Konstruktion der Simulationsreihen zu beteiligen. In diesem Fall bietet sich wiederum die Cluster-Analyse an, diesmal, um die Stationen nach Ähnlichkeit zu klassifizieren. Für jede der Klassen wird nun eine Station als Repräsentantin gewählt und einzig diese repräsentativen Stationen bekommen Regressionsparameter vorgegeben und bringen ihre Beobachtungsreihen in das Verfahren ein.

Da die Blöcke der Simulationsreihe in diesem Fall die Beobachtungen aller untersuchten Stationen zu einem Block aus der Beobachtungsperiode zugewiesen bekommen, lassen sich die Simulationsreihen auch der nicht verwendeten Stationen konstruieren, obwohl diese zur Auswahl der verwendeten Blöcke nichts beitragen.

Im jetzt folgenden zweiten Teil dieses Kapitels wird die Beschreibung des Verfahrens in mathematische Begriffe gefasst. Dabei können bisher zwangsläufig unscharf verwendete Begriffe wie "realistischer simulierter Jahresgang" oder "plausibel im Sinne der Beobachtungsreihe" quantifizierbar gemacht werden. Die Erläuterung einiger bisher im Hintergrund gebliebener Details des Algorithmus wird ebenfalls nachgeholt.

Für diesen zweiten Teil verwendete mathematische Grundbegriffe werden

in Anhang A.1 rekapituliert, zusammen mit der hier verwendeten Standardnotation. Mengen werden beispielsweise mit Großbuchstaben, ihre Elemente mit Kleinbuchstaben bezeichnet, $M = \{m_1, m_2, \ldots\}$. Ein wichtiger Baustein besonders der algorithmischen Beschreibung des Verfahrens sind Listen, das sind geordnete Mengen, deren Elemente in eckige Klammern gefasst werden, z. B. $L = [l_1, l_2, \ldots]$. Die bei Mengen verwendete Notation der Mengenbildung ist auch für Listen gebräuchlich, z. B. für die Liste der ersten 10 Quadratzahlen $L_Q = [i^2 | i \leftarrow [1..10]]$. Statistische Konzepte wie Zufallsvariablen X und deren Realisierungen x, Prozesse $[X_i]$ und Zeitreihen $[x_i]$ als deren Realisierungen, Tests von Nullhypothesen etc. werden in Anhang A.2 eingeführt.

Damit komme ich zur Beschreibung der Fragestellung und des Ansatzes, mit dem ich eine Antwort unternehme.

2.2 Der Einstationen-Fall

Ganz grob gesprochen bestimmt das Verfahren für jede meteorologische Station eines Untersuchungsgebiets Zeitreihen, die zu jedem Kalenderdatum einer interessierenden Simulationsperiode meteorologische Beobachtungen enthalten. Die Entitäten, mit denen das Verfahren umgeht, sind also Beobachtungsreihen von meteorologischen Stationen. Ich werde allerdings, um die Darstellung sowohl der Fragestellung als auch des Lösungsansatzes einfach halten zu können, zunächst den Fall einer einzelnen meteorologischen Station betrachten, d. h. dass die Abhängigkeit von den Stationen verschwindet. Eine Übertragung auf den Kontext mehrerer Stationen schließt sich in Abschnitt 2.3 an.

Die Kalenderdaten (entsprechend dem gregorianischen Kalender), die zu den Beobachtungen gehören, möchte ich in einer Liste *C* organisieren⁷ (für die Listennotation [] siehe den Eintrag zu Listen in Anhang A.1). Sie besteht aus den Tupeln (Jahr (y), Monat (m), Tag (d)):

$$C = [(y, m, d) | y \leftarrow [0, 1, 2, ...], m \leftarrow [1..12], d \leftarrow [1..nd(y, m)]],$$
(2.1)

wobei

$$nd(y,m) = \begin{cases} m = 1: & 31\\ m = 2: & nf(y)\\ \vdots\\ m = 11: & 30\\ m = 12: & 31 \end{cases}$$

die Anzahl der Tage in Monat m und nf die Anzahl der Februartage für Schalt-

⁷ Die Symbole meiner Notation sind größtenteils von den entsprechenden englischen Begriffen motiviert, z. B. *C* für calendar.

und Nichtschaltjahre angibt,

nf(y) =
$$\begin{cases} 28, & \text{für } y \mod 4 \neq 0 \lor (y \mod 100 = 0 \land y \mod 400 \neq 0) \\ 29, & \text{sonst.} \end{cases}$$

mod bezeichnet dabei den Modulus-Operator.

Tägliche Beobachtungen einer Station lassen sich damit als Paar (c, b(c))darstellen, wobei c das Kalenderdatum und b(c) das Beobachtungstupel zu Datum c darstellen – z. B. (Maximum der Lufttemperatur, Tagesmittel der Lufttemperatur, Minimum der Lufttemperatur, Niederschlag,...). Allgemein wird daher $b(c) \in \mathbb{R}^n$. Ich werde für das Paar (c, b(c)) die abkürzende Schreibweise b_c verwenden.

Da Beobachtungs- und Zeitreihen die Basis der folgenden Ausführungen geben, soll an dieser Stelle die weiterhin verwendete Notation festgeschrieben werden (siehe auch den Eintrag zu stochastischen Prozessen in Anhang A.2). Die hier verwendeten Reihen liegen für Zeiträume vor, die aus vollständigen aufeinanderfolgenden Kalenderjahren bestehen. Reellwertige Zeitreihen (wiederum eine Liste aus Paaren $(c, x(c)), x : C \to \mathbb{R}$, abgekürzt x_c), die den Zeitraum der Jahre y_1 bis y_2 abdecken, werden dabei mit

$$x^{y_1, y_2} = [x_c \mid c \leftarrow C, \text{ fst } c \leftarrow [y_1 ... y_2]], y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1$$
(2.2)

bezeichnet. Der Operator fst greift dabei aus einem Tupel sein erstes Element heraus, z. B. fst (a, b) = a (analog dazu greift snd aus einem Tupel dessen zweites Element heraus, snd (a, b) = b). Ganz analog werden Reihen, die zu den Kalenderdaten meteorologische Beobachtungen $b(c) \in \mathbb{R}^n$ enthalten, notiert:

$$b^{y_1, y_2} = [b_c \mid c \leftarrow C, \text{ fst } c \leftarrow [y_1 .. y_2]], y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1.$$
(2.3)

Soll aus den Beobachtungstupeln eine bestimmte Variable v herausgegriffen werden, wird die entsprechende Reihe mit $(b^{y_1,y_2})_v$ bezeichnet.

Wie in der Einleitung beschrieben, wird aus den Beobachtungstupeln snd b^{y_1,y_2} eine reellwertige Bezugsgröße gebildet, die für die im folgenden vorgestellte Fragestellung von besonderer Bedeutung ist. Der Operator *t* projiziert dazu mittels der Projektion proj die meteorologischen Beobachtungstupel in der Reihe auf die Bezugsgröße,

$$t \ b^{y_1, y_2} = [(c, \operatorname{proj} b(c)) \mid c \leftarrow \operatorname{fst} b^{y_1, y_2}, \operatorname{proj} : \mathbb{R}^n \to \mathbb{R}].$$
(2.4)

 $t b^{y_1,y_2}$ liefert also die Reihe der reellwertigen Bezugsgröße für die Jahre von y_1 bis y_2 .

Besondere Bedeutung hat selbstverständlich die Reihe, die konkret während der Beobachungsperiode von y_1^{obs} bis y_2^{obs} registriert wurde. Werde die Beobachungsperiode mit

$$C^{\text{obs}} = \left[c \mid c \leftarrow C, \text{fst } c \leftarrow [y_1^{\text{obs}} ... y_2^{\text{obs}}] \right], \ y_1^{\text{obs}}, y_2^{\text{obs}} \in \mathbb{N}_0, \ y_2^{\text{obs}} > y_1^{\text{obs}}$$
(2.5)

bezeichnet, ist die Beobachtungsreihe

$$b^{\text{obs}} = [b_c \mid c \leftarrow C^{\text{obs}}]. \tag{2.6}$$

Ziel des Verfahrens ist, eine Simulationsreihe b^{sim} zu erzeugen, die entsprechend in der Simulationsperiode C^{sim} die Jahre von y_1^{sim} bis y_2^{sim} abdeckt und die eine im statistischen Sinne sinnvolle Approximation an die tatsächliche (allerdings häufig unbekannte) Beobachtungsreihe der Simulationsperiode darstellen soll. Für die Definition von "sinnvoll" siehe den folgenden Abschnitt 2.2.1.

Diese Approximation braucht natürlich Randbedingungen, die sie einschränken. Wie bereits beschrieben, ist in unserem Fall die einzige äußere Randbedingung eine vorgeschriebene lineare Regressionsgerade für eine ausgewählte Bezugsgröße⁸. Die Randbedingung besteht also aus den beiden Parametern (absolutes Glied und Steigung) einer Geraden, die mit tr $\in \mathbb{R}^2$ bezeichnet werden. Diese Gerade soll die langfristige Entwicklung der Bezugsgröße *t* vorschreiben und bezieht sich deshalb auf die zuvor geglättete Bezugsgröße, in diesem Fall schlicht ihre Jahresmittel.

2.2.1 Die Fragestellung

Ganz allgemein suche ich damit nach einer Funktion g, die aus einer gegebenen Beobachtungsreihe und einer Regressionsvorgabe tr $\in \mathbb{R}^2$ eine Simulationsreihe erzeugen soll:

$$g: \left\{ b^{y_1^{\text{obs}}, y_2^{\text{obs}}} \mid y_1^{\text{obs}}, y_2^{\text{obs}} \in N_0, \ y_2^{\text{obs}} > y_1^{\text{obs}} \right\} \times \mathbb{R}^2 \longrightarrow$$

$$\left\{b^{y_1^{\rm sim}, y_2^{\rm sim}} \mid y_1^{\rm sim}, y_2^{\rm sim} \in N_0, \ y_2^{\rm sim} > y_1^{\rm sim}\right\},\tag{2.7}$$

wobei ich mich auf den Fall beschränken werde, in dem die Anzahl der Jahre der Simulationsperiode die der Beobachtungsperiode nicht überschreitet, $y_2^{sim} - y_1^{sim} \le y_2^{obs} - y_1^{obs}$.

Wie gesagt soll b^{sim} eine sinnvolle Approximation an die tatsächliche Beobachtungsreihe der Simulationsperiode darstellen. Für die Charakterisierung einer Approximation als sinnvoll wird im folgenden eine Reihe von Eigenschaften wie realistische Jahresgänge und Persistenz sowie physikalisch konsistente Kombinationen verschiedener Beobachtungsvariablen beschrieben. Die Funktion *g* muss so beschaffen sein, dass die erzeugten b^{sim} diese Eigenschaften aufweisen.

Für die Spezifizierung der Eigenschaften benötige ich noch einige Funktionen.

⁸ Für das Elbe-Einzugsgebiet bietet sich beispielsweise die Tagesmitteltemperatur an.

Funktionen für die Spezifizierung

Erstens die lin_reg-Funktion, die aus einer reellwertigen Zeitreihe x^{y_1,y_2} , $y_2 > y_1$ die beiden Parameter einer linearen Regressionsgeraden schätzt,

$$\lim_{x \to 1} \operatorname{reg} : \{ x^{y_1, y_2} \mid y_1, y_2 \in \mathbb{N}_0, \ y_2 > y_1 \} \to \mathbb{R}^2,$$
(2.8)

die sich aus den Jahresmitteln der Beobachtungen in Abhängigkeit von den Jahren ergeben. Für die Definition von lin_reg siehe Gleichung (2.26).

Zweitens die acf-Funktion, die aus einer reellwertigen Zeitreihe x^{y_1,y_2} den Wert der Autokorrelationsfunktion ihres zugrundeliegenden Prozesses für eine gegebene Verschiebung $l \in \mathbb{N}_0$ schätzt,

acf:
$$\{x^{y_1, y_2} \mid y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1\} \times \mathbb{N}_0 \to \mathbb{R},$$
 (2.9)

und, dazu gleichwertig, die per-Funktion, die für diese Zeitreihe mittels des Periodogramms die Amplitude des Fourier-Spektrums an einer gegebenen Fourier-frequenz $\omega \in \mathbb{R}$ schätzt,

per:
$$\{x^{y_1, y_2} \mid y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1\} \times \mathbb{R} \to \mathbb{R}.$$
 (2.10)

Für die Definition dieser Funktionen siehe Gleichung (3.2) in Abschnitt 3.1.1 sowie Gleichung (B.1) in Abschnitt B.4.

Drittens die dist^{*e*}-Funktion, die den euklidischen Abstand zwischen zwei \mathbb{R}^{l} -Tupeln *x* und *y* bestimmt,

dist^e:
$$\mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}$$
, dist^e(x, y) = $\left(\sum_{i=1}^l (x_i - y_i)^2\right)^{1/2}$, (2.11)

und die dist m -Funktion, die zwei l-Tupeln reeller Zahlen ihren normierten Manhattan-Abstand zuordnet:

dist^{*m*}:
$$\mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}$$
, dist^{*m*}(*x*, *y*) = $\sum_{i=1}^l |x'_i - y'_i|$, (2.12)

wobei die x'_i, y'_i die in das (0, 1)-Intervall normierten Werte der x_i, y_i bezeichnen,

$$x'_{i} = \frac{x_{i} - \min(x)}{\max(x) - \min(x)}, \ \min_{\max}(x) = \min_{\max}\{x_{j} \mid j \leftarrow [1..l]\}, \ i \leftarrow [1..l].$$

 y'_i analog.

Viertens und letztens die nn-Funktion, die zu einem \mathbb{R}^l -Tupel *x* aus einer Liste von \mathbb{R}^l -Tupeln *L* das im Sinne der euklidischen Distanz nächste Tupel bestimmt:

nn:
$$\mathbb{R}^{l} \times \text{list } \mathbb{R}^{l} \to \mathbb{R}^{l}$$
, nn(x, L) = argmin dist^e(x, l), (2.13)
 $_{l \leftarrow L}$

wobei $\operatorname{argmin}_{m \in M} h(m)$ dasjenige $m \in M$ zurückliefert, an dem h(m) minimal wird.

Spezifizierung

Die durch g erzeugte Funktion soll die folgenden Eigenschaften aufweisen:

i) **Konsistenz zwischen Beobachtungsvariablen.** Die simulierten Tupel snd b^{sim} sollen im Sinne der in der Beobachtungsperiode beobachteten Tupel snd b^{obs} physikalisch plausibel sein. Konsistenz im Sinne der snd b^{obs} bedeutet hierbei, dass jedes simulierte Tupel $x \leftarrow \text{snd } b^{\text{sim}}$ einem beobachteten Tupel aus snd b^{obs} ähneln muss, wobei die erlaubte Verschiedenheit durch

 $\operatorname{dist}^{e}(x,\operatorname{nn}(x,\operatorname{snd} b^{\operatorname{obs}})) \leq \operatorname{dist}^{e}(\operatorname{nn}(x,\operatorname{snd} b^{\operatorname{obs}}),\operatorname{nn}(\operatorname{nn}(x,\operatorname{snd} b^{\operatorname{obs}}),\operatorname{snd} b^{\operatorname{obs}}))$ (2.14)

gegeben ist. D.h., dass jedes simulierte Tupel maximal so weit vom ihm ähnlichsten beobachteten Tupel entfernt sein darf, wie dieses ähnlichste seinerseits von seinem nächsten Nachbarn in snd $b^{\rm obs}$ entfernt ist.

ii) **Regressionsparameter.** Die Vorgaben für die Parameter der linearen Regressionsgeraden der Bezugsgröße sollen innerhalb einer gesetzten Toleranz ϵ_{tr} getroffen werden:⁹

$$\operatorname{dist}^{m}(\operatorname{lin}_{\operatorname{reg}}(t \ b^{\operatorname{sim}}), \operatorname{tr}) < \epsilon_{\operatorname{tr}}.$$
(2.15)

iii) **Jahresgang.** Werden die Reihen in b^{obs} und b^{sim} in die Anteile langzeitlicher Trend, Jahresgang und Rauschen zerlegt (z. B. mit dem STL-Verfahren, siehe Abschnitt D),

$$b^{\text{obs}} = b^{\text{obs}}_{\text{trend}} + b^{\text{obs}}_{\text{jahr}} + b^{\text{obs}}_{\text{rausch}},$$

$$b^{\text{sim}} = b^{\text{sim}}_{\text{trend}} + b^{\text{sim}}_{\text{iahr}} + b^{\text{sim}}_{\text{rausch}},$$
(2.16)

dann soll gelten:

- (a) Die beiden Jahresgänge b_{jahr}^{obs} und b_{jahr}^{sim} sollen bezüglich der Kalenderdaten (y^o , 1, 1) aus C^{obs} bzw. (y^s , 1, 1) aus C^{sim} in Phase sein.
- (b) Die Nullhypothese:

Für jede Variable v der Beobachtungstupel sind die beiden Periodogramme der Reihen $(b^{obs})_v$ bzw. $(b^{sim})_v$,

$$\begin{bmatrix} \operatorname{per}_{(b^{\operatorname{obs}})_{v}}(\omega) \mid \omega = j/m^{o}, \ j \leftarrow \left[-\left\lceil\frac{m^{o}-1}{2}..\left\lceil\frac{m^{o}}{2}\right\rceil\right] \end{bmatrix} \text{ und} \\ \begin{bmatrix} \operatorname{per}_{(b^{\operatorname{sim}})_{v}}(\omega) \mid \omega = j/m^{s}, \ j \leftarrow \left[-\left\lceil\frac{m^{s}-1}{2}..\left\lceil\frac{m^{s}}{2}\right\rceil\right] \end{bmatrix}, \ v \leftarrow [1..n], \end{cases}$$

 $^{^9}$ Für die Festlegung von $\epsilon_{
m tr}$ siehe Abschnitt 3.1.3

wobei m^o und m^s die Längen der beiden Reihen angeben und $\lceil x$ die kleinste ganze Zahl kleiner x bezeichnet, spektrale Realisierungen des gleichen Prozess.

soll auf einem gewählten Signifikanzniveau nicht abgelehnt werden. Dadurch soll sichergestellt werden, dass die Jahresperiode in der Frequenzzerlegung von b^{sim} ein zu b^{obs} vergleichbares Gewicht besitzt.

(c) Die Nullhypothese:

Die Werte der Rauschanteile von beobachteter und simulierter Reihe, snd $b_{\text{rausch}}^{\text{obs}}$ und snd $b_{\text{rausch}}^{\text{sim}}$, sind Stichproben der gleichen Grundgesamtheit.

soll auf einem gewählten Signifikanzniveau nicht abgelehnt werden. Diese Forderung soll sicherstellen, dass die simulierten Fluktuationen um Jahresgang- und Trendkomponente realistisch sind. Siehe auch Abschnitt 3.2.

iv) Persistenz. Die Nullhypothese:

Für jedes Element v der Beobachtungstupel charakterisieren die für alle sinnvollen Verschiebungen für die trend- und jahresgangbereinigten Reihen $(b_{rausch}^{obs})_v$ bzw. $(b_{rausch}^{sim})_v$ geschätzten Werte der Autokorrelationsfunktion,

$$\begin{bmatrix} \operatorname{acf}_{(b_{\operatorname{rausch}}^{\operatorname{obs}})_{v}}(l) \mid l \leftarrow [1..m_{v}'] \end{bmatrix} \quad \text{und} \\ \begin{bmatrix} \operatorname{acf}_{(b_{\operatorname{rausch}}^{\operatorname{sim}})_{v}}(l) \mid l \leftarrow [1..m_{v}'] \end{bmatrix}, \quad v \leftarrow [1..n] \\ \end{bmatrix}$$

wobe
i m_{ν}' eine jeweils sinnvolle maximale Verschiebung angibt, den selben Prozess.

ist auf einem gewählten Signifikanzniveau nicht abzulehnen. Diese Bedingung soll sicherstellen, dass b^{sim} auch über die Trend- und Jahresganganteile hinaus im Sinne von b^{obs} realistische Persistenzen aufweist.

2.2.2 Der Ansatz

Durch die geforderten Eigenschaften, die die durch *g* erzeugten Reihen besitzen sollen, ist *g* nicht vollständig definiert – bereits die statistische Formulierung der Eigenschaften *i*), *iii*) und *iv*) schließt eine eindeutige analytische Lösung aus. Ich möchte mich der Fragestellung daher heuristisch nähern, mit einem Verfahren, dass durch seine Konstruktion Reihen erzeugt, die die geforderten Eigenschaften aufweisen.
Das Verfahren erhebt nicht den Anspruch, die einzige oder die im Sinne irgendeiner Qualitätsprüfung beste Lösung des gestellten Problems zu sein. Es stellt vielmehr den Versuch dar, sich der Fragestellung durch eine Kette plausibler Überlegungen zu nähern. Wie die Kapitel 3 und 4 zeigen werden, bringt das Verfahren in den Fällen, in denen es angewandt und validiert wurde, sehr befriedigende Ergebnisse.

Eigenschaft *i*) enthält eine mathematische Formulierung der Forderung, dass die simulierten Tupel in b^{sim} physikalisch konsistent seien. Anschaulich gesprochen bedeutet dies beispielsweise, dass ein Tag mit Niederschlag nicht wolkenfrei simuliert werden sollte. Um diese Konsistenz sicherzustellen, liegt dem Verfahren der folgende **Ansatz** zugrunde: Alle simulierten Tupel in snd b^{sim} entstammen den beobachteten Tupeln aus snd b^{obs} , d. h. $\forall c \leftarrow C^{\text{sim}} : \exists c' \leftarrow C^{\text{obs}}$, so dass

$$\operatorname{snd} b_c = \operatorname{snd} b_{c'}.$$
 (2.17)

Für die simulierte Reihe werden damit Beobachtungstupel aus sn
d b^{obs} neu aneinandergereiht, siehe Abbildung 2.1.

Hinter diesem Ansatz steht die Annahme, dass Abschnitte aus b^{obs} während C^{sim} erneut oder ähnlich auftreten können und dass die in b^{obs} beobachtete Variabilität ausreicht, um b^{sim} zu erzeugen. Das wird sicherlich nicht immer der Fall sein – beispielsweise wäre es unrealistisch, eine zu den klimatischen Verhältnissen der letzten Eiszeit passende Simulationsreihe b^{sim} aus den Beobachtungen der letzten 50 Jahre erzeugen zu wollen. Der Ansatz garantiert zwar einerseits physikalisch konsistente Simulationsreihen, beschränkt aber andererseits den Anwendungsbereich des Verfahrens.

Mit dem Ansatz der Neuanordnung der Tupel aus sn
d $b^{\rm obs}$ ergibt sich die Menge möglicher Simulationsreihen zu

$$L_{i} = \left\{ \left[(c, f(c)) \mid c \leftarrow C^{\text{sim}} \right] \mid f : C^{\text{sim}} \to \text{snd } b^{\text{obs}} \right\}$$
(2.18)

Ein Element dieser Menge besteht also aus einer Simulationsreihe, deren Tupel aus der Liste der Tupel aus s
nd $b^{\rm obs}$ stammen.

Die Forderungen ii)-iv) schränken den Anwendungsbereich des Verfahrens ein, d. h. nicht alle Kombinationen von b^{obs} und Vorgaben tr $\in \mathbb{R}^2$ werden zu b^{sim} -Reihen führen können, die den Forderungen i)-iv) genügen.

Aus *ii*) ergibt sich die Einschränkung, dass sich die Regressionsvorgabe in einem der ϵ_{jd} -Intervalle um die Regressionsparameter der Simulationsreihen aus L_{i} bewegen muss:

$$\operatorname{tr} \in \left\{ \operatorname{tr}' \mid \operatorname{dist}^{m} \left(\operatorname{lin}_{\operatorname{reg}}(t \ l), \operatorname{tr}' \right) \le \epsilon_{\operatorname{jd}}, \ l \in L_{i} \right\}.$$

$$(2.19)$$

Erfüllt eine Regressionsvorgabe tr diese Bedingung, schränkt sie die Menge der mit ihr verträglichen Simulationsreihen ein auf

$$L_{ii} = \{l \mid l \in L_i, \operatorname{dist}^m(\operatorname{lin}_{\operatorname{reg}}(t \ l), \operatorname{tr}) \le \epsilon_{\operatorname{tr}}\}.$$
(2.20)

Ein Verfahren, das simulierte Reihen erzeugt, in denen z. B. der Sommer kälter als der Winter ist, ist offensichtlich nutzlos. Dem beugt Forderung *iii*) vor.

Forderung *iv*) verlangt realistische Persistenz in den Reihen. Allerdings wird der gewählte Ansatz dazu tendieren, Simulationsreihen mit geringerer Persistenz als in b^{obs} zu erzeugen: Durch das Umsortieren der Tupel aus snd b^{obs} und den damit möglicherweise entstehenden Brüchen¹⁰ ist dieser Effekt praktisch vorprogrammiert. Mit *iv*) muss daher ein Maß gefunden werden, mit dem diese Abweichung als zu groß oder als noch akzeptabel bewertet werden kann.

Im Gegensatz zur Einschränkung des Anwendungsbereichs durch Eigenschaft *ii*) können die durch *iii*) und *iv*) bedingten Einschränkungen nicht explizit angegeben werden und müssen für jedes erzeugte *b*^{sim} *a posteriori* überprüft werden.

Der folgende Abschnitt wird einen Algorithmus vorstellen, der eine Neuaneinanderreihung von Tupeln aus b^{obs} im Sinne der Eigenschaften *i*) bis *iv*) erzeugt. Es handelt sich hierbei um ein iteratives Vorgehen, bei dem eine erste Neuanordnung schrittweise verbessert wird, bis die Regressionsvorgabe aus *ii*) erreicht ist.

2.2.3 Der Algorithmus

Wie schon in Abschnitt 2.1.3 skizziert, spezifiziert der Algorithmus b^{sim} in zwei Phasen, der einer ersten Näherung und deren iterativer Verbesserung in der zweiten Phase. Diese beiden Phasen finden auf zwei verschiedenen Zeitskalen statt. Die der ersten ist ziemlich grob und versucht, unter Beibehaltung möglichst großer zusammenhängender Abschnitte des Beobachtungszeitraums bereits nahe an die Regressionsvorgabe *ii*) zu gelangen. In der zweiten Phase werden kleinere zusammenhängende Abschnitte dieser ersten Näherung ausgetauscht, bis die Vorgabe im Sinne von *ii*) getroffen wird.

Da die meisten meteorologischen Variablen in unseren Breiten einen mehr oder weniger ausgeprägten Jahresgang zeigen, bieten sich die Kalenderjahre als "große zusammenhängende Abschnitte" für die erste Phase an. Vielfache davon sind natürlich ebenso denkbar, etwa 2 oder 5 Jahre. Da jedoch für die meisten Regionen nur Beobachtungsdaten von etwa 50 Jahren zur Verfügung stehen und man daher z. B. für Abschnitte von 5 Jahren nur 10 solcher Abschnitte zur Verfügung hätte, fiel die Wahl für diesen Kontext auf ein Jahr. Die erste Näherung ist ziemlich wahrscheinlich mit den Eigenschaften *iii*) und *iv*) verträglich. "Nur" ziemlich wahrscheinlich, weil sich durch die eventuellen Brüche zwischen den einzelnen Jahren die Persistenz verringern kann, was *iv*) ggf. verletzt. Außerdem, falls nicht alle Beobachtungsjahre in der Neureihung vorkommen, wird

¹⁰ Die Verwendung von Blöcken reduziert diesen Effekt zwar erheblich – Brüche können nur noch zwischen den Blöcken auftreten – eliminiert ihn jedoch nicht vollständig.

die Menge der zueinander verschiedenen Jahresreihen kleiner sein als in der Beobachtungsperiode. Dadurch können sich Änderungen der Verteilungen des Rauschanteils ergeben, was ggf. *iii*) verletzt, auch wenn die Jahresgänge garantiert in Phase sind. In der Regel wird diese erste Näherung jedoch im Sinne der Eigenschaften *iii*) und *iv*) brauchbar sein, da die Beobachtungstupel-Abfolgen ganzer Jahre aus der Beobachtungsperiode unverändert übernommen werden.

Aus der ersten Näherung werden wie in 2.1.3 beschrieben während der zweiten Phase Blöcke¹¹ teils übernommen, teils ausgetauscht, wobei der Austausch so vorgenommen wird, dass die entstehenden Reihen den Forderungen *iii*) und *iv*) entsprechen können und gleichzeitig die Reihe näher an die Regressionsvorgabe bringen. Die Idee hinter diesem Vorgehen ist, möglichst viel von den Witterungsabfolgen der beobachteten Jahre beizubehalten, um die Simulationsreihe b^{sim} zu erzeugen. Dadurch wird es möglich, Simulationsreihen mit plausiblen Witterungsabfolgen zu generieren – plausibel auch innerhalb der einzelnen Jahre.

Die jetzt beginnende detaillierte algorithmische Beschreibung des Verfahrens beginnt mit der Einführung verwendeter Funktionen, gefolgt von der Beschreibung der ersten Näherung und ihrer iterativen Verbesserung.

Eine Erweiterung auf mehrere Stationen lässt sich durch eine einfache Übertragung der Ein-Stationen-Methode in einen höherdimensionalen Parameterraum erreichen, deren Erläuterung sich in Abschnitt 2.3 anschließt.

Verwendete Funktionen

Erstens die Klassifikation class, die jedem Element aus einer Menge M eine von n_c Klassen zuweist.

$$class: M \to \{1, 2, \dots, n_c\}.$$

$$(2.21)$$

Für den Kontext dieser Arbeit wird class mittels einer Clusteranalyse spezifiziert, siehe Anhang C.

Zweitens die *m*-Funktion, die einer endlichen Liste *L* reeller Zahlen das arithmetische Mittel zuordnet:

$$m: \text{list } \mathbb{R} \to \mathbb{R}, \ m(L) = \frac{1}{\text{length}(L)} \sum_{i=1}^{\text{length}(L)} l_i.$$
 (2.22)

Etwas spezieller erzeugt **drittens** die am-Funktion aus einer reellwertigen Zeitreihe x^{y_1,y_2} eine Liste von Paaren, die aus den Jahreszahlen von y_1 bis y_2

¹¹ Für die Anwendung des Verfahrens auf die Daten des Elbe-Einzugsgebiets werden die Blöcke eine Länge von 12 Tagen haben.

und den zugehörigen Jahresmittelwerten der snd x_i bestehen:

am:
$$\{x^{y_1, y_2} \mid y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1\} \rightarrow \\ \{[(y, f(y)) \mid y \leftarrow [y_1..y_2]], f: \text{list } \mathbb{R} \rightarrow \mathbb{R}, y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1\}, \\ am(x^{y_1, y_2}) = [(y, f(y)) \mid y \leftarrow [y_1..y_2]], f(y) = m([\text{snd } x_c \mid \text{fst } c = y]).$$

$$(2.23)$$

Viertens bestimmen die lr₁- und lr₂-Funktionen aus einer endlichen Liste *L* der Länge *k* von Paaren $l \in \mathbb{R}^2$ die beiden Parameter einer linearen Regression der Form snd $l = lr_1 + lr_2 \cdot (\text{fst } l)$,

$$\operatorname{lr}_1:\operatorname{list} \mathbb{R}^2 \to R, \, \operatorname{lr}_1(L) = m(\operatorname{snd} L) - \operatorname{lr}_2(L) \cdot m(\operatorname{fst} L), \quad (2.24)$$

wobei

$$\ln_2 : \text{list } \mathbb{R}^2 \to R, \ \ln_2(L) = \frac{\sum_{i=1}^k \left((\text{fst } L)_i - m(\text{fst } L) \right) \left((\text{snd } L)_i - m(\text{snd } L) \right)}{\sum_{i=1}^k \left((\text{fst } L)_i - m(\text{fst } L) \right)^2}.$$
 (2.25)

Damit lässt sich die lin_reg-Funktion aus (2.8) definieren als

$$\lim_{x \to 1} \operatorname{reg}(x^{y_1, y_2}) = \left(\ln_1(\operatorname{am}(x^{y_1, y_2})), \ln_2(\operatorname{am}(x^{y_1, y_2})) \right).$$
(2.26)

Die erste Näherung.

Die erste Näherung besteht aus einer Neuaneinanderreihung ganzer Jahre des Beobachtungszeitraums bzw. den zugehörigen Beobachtungen, die der Regressionsvorgabe *ii*) möglichst gut entspricht. Sie wird aus einer großen Stichprobe vom Umfang *L* zufälliger Aneinanderreihungen ausgewählt.¹²

Da es hier um Jahre geht, ist eine Liste hilfreich, die die Beobachtungen eines Jahres enthält:

$$A_{y} = [b_{c} \mid c \leftarrow C, \text{ fst}(c) = y], \ y \in \mathbb{N}_{0}.$$

$$(2.27)$$

 $A^{\rm obs}$ bezeichnet entsprechend die Liste all der Jahreslisten der Beobachtungsperiode,

$$A^{\rm obs} = [A_y \mid y \leftarrow [y_1^{\rm obs}..y_2^{\rm obs}]].$$
(2.28)

Bezeichne n_{sim} die Anzahl der Jahre in C^{sim} , dann lassen sich die L zufälligen Aneinanderreihungen der jahresweisen Abschnitte der Beobachtungsreihe – als versuchsweise erste Annäherung an die Simulationsreihe – darstellen als die Reihen R^l mit

fst
$$R^l = C^{\text{sim}}$$
, snd $R^l = \text{concat}(\text{snd sample}(A^{\text{obs}}, n_{\text{sim}}), \text{Rest}), l \leftarrow [1..L], (2.29)$

¹² Experimente mit Daten des Elbe-Einzugsgebiets haben gezeigt, dass mit L = 100000 die Stichprobe groß genug ist, d. h. dass die Chance auf eine wesentlich bessere Anordnung mit einer weiteren Vergrößerung der Stichprobe kaum zunimmt.

wobei sample $n_{\rm sim}$ -mal zufällig mit Zurücklegen aus der Liste der Jahresabschnitte $b^{\rm obs}$ zieht und die Züge in die Liste R^l legt. Die concat-Funktion fügt dabei mehrere Listen zu einer zusammen. Der zusätzlich zu den ganzen Jahren angehängte Rest stellt sicher, dass die sich ergebende Liste die gleiche Länge hat wie $C^{\rm sim}$ – Unterschiede können sich u. a. aufgrund der eventuell verschiedenen Anzahlen von Schaltjahren ergeben, siehe dazu außerdem den folgenden Abschnitt zur iterativen Verbesserung. Um diese auszugleichen, werden ggf. fehlende Beobachtungen einfach mit Tupeln vom Anfang des letzten Jahres $A_{R_{n_{\rm sim}}^l}$ aufgefüllt. Sollte die Liste zu lang sein, werden die überschüssigen letzten Einträge abgetrennt (in diesem Fall müsste man den Rest in (2.29) als einen zu entfernenden Teil der Liste auffassen).

Aus den *L* zufälligen Anordnungen wird diejenige ausgewählt, deren Regressionsparameter der Vorgabe *ii*) am nächsten sind. Sie wird die erste Näherung b^{start} an die zu erzeugende Simulationsreihe b^{sim} ergeben:

$$b^{\text{start}} = \operatorname*{argmin}_{R^l, \ l \leftarrow [1..L]} \operatorname{dist}^m \left(\left(\lim_{l \to \infty} \operatorname{reg}(t \ R^l), \operatorname{tr} \right) \right).$$
(2.30)

Die Verwendung der Manhattan-Distanz dist m wird sich als vorteilhaft erweisen, wenn in Abschnitt 2.3 das Verfahren auf regionale Simulationen übertragen wird.

Die Liste der Jahresmittel der Bezugsgröße aus dieser ersten Näherung wird mit am^{start} bezeichnet,

$$am^{\text{start}} = am(t \ b^{\text{start}}), \tag{2.31}$$

ihre zugehörigen Regressionsparameter mit

$$lr_i^{\text{start}} = lr_j(\text{am}^{\text{start}}), \ j \leftarrow [1..2].$$
(2.32)

Siehe Abbildung 2.3 für eine Illustration.

Iterative Verbesserung der ersten Näherung

Bei der sich anschließenden iterativen Verbesserung werden in den einzelnen Jahren Blöcke der Länge l_b so ausgetauscht, dass sich die entstehende Reihe der Bezugsgröße immer mehr der Regressionsvorgabe *ii*) annähert. Dabei wird zum einen durch das Belassen eines "Grundgerüsts" der einzelnen Jahre, zum anderen durch die Weise der Ersetzung dafür Sorge getragen, dass die entstehenden simulierten Jahresgänge und Persistenzen aller Variablen statistisch mit denen des Beobachtungszeitraums konsistent bleiben können (Vorgaben *iii*) und *iv*)), d. h. dass sich die Ersatzblöcke gut in das Grundgerüst einfügen.

Die Blöcke werden dabei mit den Listen

$$B_{c} = \left[b_{c} \mid c \leftarrow [c..(c+l_{b}-1)] \right]$$
(2.33)

bezeichnet, wobei analog zu den Listen ganzer Zahlen $[a..b]^{13}$ die Liste $[c_1..c_2]$, $c_1, c_2 \leftarrow C$, $c_2 \ge c_1$ alle Kalenderdaten von c_1 bis einschließlich c_2 enthält. $c + l_b - 1$ bezeichnet dabei das Datum, das um $l_b - 1$ Tage später ist als c. Blöcke entsprechend der ersten Näherung oder des *i*-ten Iterationsschritts (siehe unten) erhalten die Bezeichnung B_c^{start} bzw. B_c^i .

Für den nicht unwahrscheinlichen Fall, dass die Länge der Simulationsreihe length(b^{sim}) kein ganzzahliges Vielfaches der Blocklänge l_b ist, muss die Simulationsreihe verfahrensintern um so viele Beobachtungen verlängert werden, dass sich ein ganzzahliges Vielfaches von l_b ergibt. Intern werden damit maximal $l_b - 1$ Tagesbeobachtungen zu viel simuliert, die z. B. für die erste Näherung mit dem Restterm in Gleichung (2.29) aufgenommen werden müssen, allerdings am Ende nicht mit ausgegeben werden.

Das Verfahren arbeitet also die aufeinanderfolgenden Blöcke im (ggf. intern verlängerten) Simulationszeitraum ab, so dass die Liste mit den Startdaten dieser Blöcke hilfreich ist,

$$CB = [c_1^{sim}, c_1^{sim} + l_b, c_1^{sim} + 2l_b, \dots, c_{end}^{sim}], \qquad (2.34)$$

wobei c_{end}^{sim} dasjenige Datum bezeichnet, an dem der letzte vollständig im (ggf. intern verlängerten) C^{sim} enthaltene Block beginnt,

$$c_{\text{end}}^{\text{sim}} = \underset{c \leftarrow C^{\text{sim}}}{\operatorname{argmax}} \text{fst } B_c \subset C^{\text{sim}}$$
(2.35)

 $(\operatorname{argmax}_{m \in M} h(m) \text{ liefert dasjenige } m \in M \text{ zurück, an dem } h(m) \text{ maximal wird}).$

Für jeden dieser Blöcke muss entschieden werden, ob der Block aus der ersten Näherung beibehalten wird.

Der Prozess des Entscheidens, welche Blöcke verbleiben und das Ersetzen der übrigen wird iterativ wiederholt, bis die simulierte Reihe der Bezugsgröße eine Regressionsgerade aufweist, die innerhalb der gesetzten Toleranz ϵ_{tr} mit der Vorgabe *ii*) verträglich ist. Nach jedem Iterationsschritt, der dieses Ziel noch nicht erreicht hat, wird dabei indirekt das Entscheidungskriterium für den Verbleib von Blöcken verschärft, so dass der nächste Iterationsschritt weniger Blöcke aus der ersten Näherung übernimmt und mehr Freiheit erhält, die Regressionsgerade verbessernde Blöcke auszuwählen.

Die beiden Etappen eines Iterationsschritts – Entscheiden, welche Blöcke verbleiben, Blöcke ersetzen – und die Überleitung in den nächsten Iterationsschritt werden im folgenden detailliert dargestellt. Ein hochgestelltes i bezeichnet dabei den aktuellen Iterationsschritt.

Wie die erste Näherung eine b^{start} -Reihe liefert, definiert jeder Iterationsschritt *i* eine b^i -Simulationsreihe.

¹³ Siehe den Eintrag zu Listen in Anhang A.1.

Zu Beginn eines Iterationsschrittes i ist snd b^i noch undefiniert. Im Gegensatz zur ersten Näherung, bei der snd b^{start} einfach durch eine Aneinanderfügung von jahresweisen Abschnitten der Beobachtungsreihe gefunden wurde, lässt sich diese Spezifikation nicht mit einem Schritt angeben. snd b^i wird vielmehr in den beiden Stufen jedes Iterationsschritts aufgefüllt. snd b^i besteht daher zu Beginn des *i*-ten Iterationsschritts aus einer Liste von undef-Markierungen.

snd
$$b_c^i$$
 = undef, $c \leftarrow C^{\text{sim}}$. (2.36)

Blöcke beibehalten. Vorgabe *ii*) bezieht sich auf die Ausgleichsgerade der Jahresmittel der Bezugsgröße. Eine erste Näherung dahin zu verbessern, dass sie der Regressionsvorgabe genügt, kann daher auch formuliert werden als: Die erste Näherung dahin bringen, dass die Jahresmittel der simulierten Bezugsgröße eine Regressionsgerade aufweisen, die den vorgegebenen Parametern entspricht. Daher wird aus der Liste der Paare aus Jahreszahl und Jahresmittel der ersten Näherung am^{start} eine Liste \widetilde{am}^i von entsprechenden Paaren mit "Ziel"-Jahresmitteln bestimmt, die der aktuellen Regressionsvorgabe tr^{*i*} genügen:

$$\operatorname{snd} \widetilde{\operatorname{am}}_{y}^{i} = \operatorname{snd} \operatorname{am}_{y}^{\operatorname{start}} + (\operatorname{tr}_{1}^{i} - \operatorname{lr}_{1}^{\operatorname{start}}) + (\operatorname{tr}_{2}^{i} - \operatorname{lr}_{2}^{\operatorname{start}}) \cdot y, \ y \leftarrow \operatorname{fst} C^{\operatorname{sim}}.$$
(2.37)

Die Vorgabe des ersten Iterationsschrittes entspricht dabei derjenigen aus *ii*): $tr_1^1 = tr_1$, $tr_2^1 = tr_2$.

Diese Zielmittelwerte werden nun dazu verwendet, aus der Reihe der Bezugsgröße aus der ersten Näherung $t b^{\text{start}}$ eine künstliche "Zielreihe" $\tilde{t}^i b^{\text{start}}$ zu generieren, und zwar in dem $t b^{\text{start}}$ jahresweise so erhöht bzw. gesenkt wird, dass der Mittelwert für ein so verändertes Jahr dem entsprechenden Zielmittel gleich ist:

$$\tilde{t}^i b_c^{\text{start}} = t b_c^{\text{start}} + (\text{snd } \widetilde{am}_y^i - (\text{snd } am_y^{\text{start}})), \ c \leftarrow A_y, \ y \leftarrow \text{fst } C^{\text{sim}}.$$
 (2.38)

Es stehen damit zwei Reihen der Bezugsgröße bereit, die aus der ersten Näherung t b^{start} , und die daraus hervorgegangene Zielreihe \tilde{t}^i b^{start} . \tilde{t}^i b^{start} ist eine Reihe, die der aktuellen Regressionsvorgabe genügt, d. h. dass eine lineare Regression durch ihre Jahresmittel den für den aktuellen Iterationsschritt vorgeschriebenen Parametern entspricht. Aufgrund der Konstruktion der ersten Näherung sind \tilde{t}^i b^{start} und t b^{start} nicht allzu verschieden.

Die Grundidee bei der Entscheidung darüber, welche Blöcke aus der ersten Näherung beibehalten werden, ist nun, die Blöcke bezüglich dieser beiden Reihen der Bezugsgröße miteinander zu vergleichen. Sind sich die Blöcke beider Reihen ähnlich genug, wird der Block der ersten Näherung beibehalten, d. h. wenn für ein $c \leftarrow \text{CB}$ sich $t B_c^{\text{start}}$ der Zielreihe $\tilde{t}^i B_c^{\text{start}}$ anschmiegt, dann $B_c = B_c^{\text{start}}$. Andernfalls wird er neu gesetzt. Für die Entscheidung, ob sich die

Werte der Bezugsgröße eines Blocks aus der ersten Näherung denen der Zielreihe anschmiegen, wird die Klassifikationsfunktion class verwendet, die in Gleichung (2.21) eingeführt wurde. Sie weist beliebigen Blöcken eine von n_c Ähnlichkeitsklassen zu.¹⁴ Diese Ähnlichkeitsklassen sind bezüglich der Werte der Bezugsgröße der Blöcke definiert, d. h. jeder Block ist durch ein l_b -Tupel aus dem \mathbb{R}^{l_b} -Raum charakterisiert:

class:
$$\mathbb{R}^{l_b} \to \{1, 2, \dots, n_c\}.$$
 (2.39)

Die Funktion class wird durch eine Clusteranalyse spezifiziert, die mit den beobachteten Werten der Bezugsgröße an den Blöcken {snd $t B_c | B_c \subset b^{obs}$ } trainiert wurde (siehe Anhang C).

Für die erste Stufe der Spezifikation von b^i werden für alle aufeinanderfolgenden Blöcke des Simulationszeitraums ihre Klassenzugehörigkeiten, einmal bezüglich $t b^{\text{start}}$, einmal bezüglich $\tilde{t}^i b^{\text{start}}$, bestimmt – die Blöcke der Zielreihe werden dabei derjenigen Klasse zugewiesen, deren Schwerpunkt sie am nächsten sind, siehe Anhang C. Stimmen sie überein, wird der Block der ersten Näherung beibehalten:

$$c \leftarrow \text{CB}: \text{class(snd } t \ B_c^{\text{start}}) = \text{class(snd } \tilde{t}^i \ B_c^{\text{start}}) \implies B_c^i = B_c^{\text{start}}.$$
 (2.40)

Abbildung 2.4 illustriert diesen Schritt mit der Temperatur als Bezugsgröße.

Der nächste Abschnitt widmet sich der zweiten Stufe des *i*-ten Iterationsschritts, in der die Blöcke gesetzt werden, denen bislang nur undef-Markierungen zugeordnet werden.

Blöcke neu setzen. Wurde ein Block während der ersten Stufe nicht aus der ersten Näherung übernommen, so muss er neu gesetzt werden. Dazu wird die Menge der Blöcke aus dem Beobachtungszeitraum durch vier Kriterien schrittweise eingeschränkt, die Blöcke zur Auswahl lassen, die sich gut in die schon bestehenden Teile der Reihen einfügen und dazu beitragen, die entstehende Simulationsreihe näher an die Regressionsvorgabe zu bringen. Aus der so eingeschränkten Menge wird schließlich zufällig ausgewählt. Die vier Kriterien wurden bereits in Abschnitt 2.1.3 skizziert und in Abbildung 2.5 illustriert.

Für ihre jetzt folgende detailliertere Darstellung beginne an c' einer der aufeinanderfolgenden Blöcke aus der Simulationsperiode, für den der Block der ersten Näherung nicht beibehalten wurde, $c' \leftarrow CB$, $class(t B_{c'}^{start}) \neq class(\tilde{t}^i B_{c'}^{start})$. Dann wird aus der folgendermaßen eingeschränkten Menge Ersatz beschafft:

1. Einschränkung durch Klassenzugehörigkeit. Die Grundmenge $\Gamma_{c'}^1$, aus der ein Block für $B_{c'}$ ausgewählt wird, besteht aus den Blöcken der Beobachtungsreihe, die bezüglich ihrer Bezugsgröße unter der class-Funktion in die gleiche

¹⁴ Für die Festlegung von n_c siehe Kapitel C und 3.1.3.

Ähnlichkeitsklasse einsortiert werden wie der zu besetzende Block bezüglich der Zielreihe der Bezugsgröße, $\tilde{t}^i B_{c'}^{\text{start}}$:

$$\Gamma_{c'}^{1} = \{B_{\kappa} \mid \text{class(snd } t \ B_{\kappa}) = \text{class(snd } \tilde{t}^{i} \ B_{c'}^{\text{start}}\}, \ B_{\kappa} \subset b^{\text{obs}}\}.$$
(2.41)

Die Blöcke, die auf diese Weise in die Auswahl kommen, schmiegen sich im Sinne der Ähnlichkeitsklassifizierung an die Zielreihe an.

2. Einschränkung auf noch nicht gesetzte Blöcke. Um zu vermeiden, dass einzelne Blöcke zufällig gehäuft gesetzt werden, wird aus der Grundmenge nur diejenige Teilmenge $\Gamma_{c'}^2$ behalten, deren Tupel noch nicht in der Menge der bereits verwendeten Tupel snd b^i vorkommen:

$$\Gamma_{c'}^2 = \{B_{\kappa} \mid B_{\kappa} \in \Gamma_{c'}^1, \text{ snd } B_{\kappa} \not\subset \text{ snd } b^i\}.$$
(2.42)

Bislang werden die neu zu setzenden Blöcke lediglich anhand der Bezugsgröße ausgewählt, was für die weiteren Beobachtungsvariablen, die den Blöcken zugehören, unter Umständen nicht aussagekräftig genug ist. Mit der Temperatur beispielsweise lassen sich Sommer- und Winterblöcke im allgemeinen zuverlässig unterscheiden, für Frühlings- und Herbstblöcke gilt das nicht unbedingt. Daher die

3. Einschränkung anhand des julianischen Datums. Um die Jahreszeiten für alle Beobachtungsvariablen zu erhalten, schränkt das dritte Kriterium die Menge anhand des julianischen Datums ein: Lediglich Blöcke, für die sich das julianische Datum des Anfangs des Blocks nicht mehr als $\epsilon_{jd} \in \mathbb{N}_+$ von dem des Beginns des zu besetzenden Blocks $B_{c'}$ unterscheidet, bleiben in der Auswahl:¹⁵

$$\Gamma_{c'}^3 = \{B_{\kappa} \mid B_{\kappa} \in \Gamma_{c'}^2, \text{ jd}(\kappa) \ge \text{jd}(c') - \epsilon_{\text{jd}} \text{ und } \text{jd}(\kappa) \le \text{jd}(c') + \epsilon_{\text{jd}}\}.$$
(2.43)

Bislang sind also Blöcke in der Auswahl, die der Zielbezugsgröße für den zu besetzenden Blockes ähnlich sind, noch nicht gesetzt wurden und jahreszeitlich passend sind. Die letzte Einschränkung soll sicherstellen, dass nur Blöcke Kandidaten bleiben, die zu einem bereits gesetzten Vorgänger und, falls bereits gesetzt, auch zum nachfolgenden Block passen.

4. Einschränkung auf Blöcke, die sich gut in die schon bestehende Reihe einfügen. Dieses Kriterium verlangt gerade Blocklängen: Um Blöcke zu identifizieren, die zum in b^i gesetzten Vorgänger passen, bildet das Kriterium zunächst "Brückenblöcke" BB_{c'}, das sind Blöcke, die aus der zweiten Hälfte des Vorgängers und aus der ersten Hälfte eines Kandidatenblocks $B_{\kappa} \subset b^{\text{obs}}$ bestehen:

$$BB_{c'\kappa} = concat([b_c^{l} | c \leftarrow [(c'-l_b/2)..(c'-1)]], [b_c | c \leftarrow [\kappa..(\kappa+l_b/2-1)]]).$$
(2.44)

Die Idee für das Kriterium ist nun, diesen potentiellen Brückenblock mit dem Brückenblock zu vergleichen, der den schon gesetzten Vorgängerblock $B_{c'-l_{\nu}}^{i}$

¹⁵ Zur Wahl von ϵ_{id} siehe Abschnitt 3.1.2.



Abbildung 2.6: Das Brückenblock-Kriterium. Links die Blöcke eines Ausschnitts der Beobachtungsreihe mit den Brückenblöcken (schraffiert, Klassenzugehörigkeiten rot in Klammern). Rechts findet sich ein Ausschnitt in der Simulationsreihe wieder, wobei wie in Abbildung 2.5 der dritte Block nicht aus der ersten Näherung übernommen wird und also neu gesetzt werden muss. Kandidaten dafür sind in gelb dargestellt, schraffiert dazu die Brückenblöcke, die sie mit dem Block Nr. 2 bilden. In der Auswahl bleiben die Kandidaten, deren Brückenblöcke in die selbe Klasse (rot in Klammern) fallen wie der Brückenblock 2-3 in der ursprünglichen Reihe. Diese sind rot umrandet.

mit seinem ursprünglichen Nachfolgerblock in b^{obs} verbindet. Bezeichne dazu B_{c^*} den Block aus b^{obs} , dessen Tupel dem Vorgängerblock zugewiesen wurden, snd $B_{c'-l_b}^i$ = snd B_{c^*} . Dann ist der gesuchte Brückenblock aus der Beobachtungsreihe der Block $B_{c^*+l_b/2}$, der mittels der class-Funktion mit dem potentiellen Brückenblock BB $_{c'\kappa}$ verglichen wird: Fallen Brückenblock BB $_{c'\kappa}$ und der ursprüngliche Brückenblock des Beobachtungszeitraums $B_{c^*+l_b/2}$ bezüglich der Bezugsgröße in die gleiche Klasse,¹⁶ so wird der Kandidat B_{κ} als passend zu seinem Vorgänger angesehen (siehe Abbildung 2.6).

Für einen eventuell bereits gesetzten Nachfolger wird die Idee analog ange-

¹⁶ Hierzu werden die Blöcke wiederum derjenigen Klasse zugewiesen, deren Schwerpunkt sie am nächsten sind, siehe Anhang C.

wandt: Verglichen wird der Brückenblock $BB'_{c'\kappa}$,

$$BB'_{c'\kappa} = \operatorname{concat}([b_c \mid c \leftarrow [(\kappa + l_b/2)..(\kappa + l_b - 1)]], [b_c^i \mid c \leftarrow [(c' + l_b)..(c' + 3/2l_b - 1)]])$$
(2.45)

der aus der zweiten Hälfte des Kandidatenblocks und der ersten des Nachfolgers besteht, mit dem Block des Beobachtungszeitraums, der in b^{obs} eine halbe Blocklänge vor dem Block beginnt, dessen Tupel dem Nachfolgeblock zugewiesen wurden. Dieser ursprüngliche Nachfolgerblock aus b^{obs} werde mit $B_{c^{*\prime}}$ bezeichnet, der verglichene Brückenblock ist entsprechend $B_{c^{*\prime}-l_b/2}$. Ein Kandidat muss beide Vergleiche überstehen, falls auch der nachfolgende Brückenblock gebildet werden kann, ansonsten nur den Vergleich für den Vorgängerbrückenblock:

$$\Gamma_{c'}^4 = \left\{ B_{\kappa} \mid B_{\kappa} \in \Gamma_{c'}^3, \text{ class(snd } t \text{ BB}_{c',\kappa} \right\} = \text{class(snd } t B_{c^*+l_b/2}), \text{ und,} \qquad (2.46)$$

falls
$$B_{c'+l_b}^l \neq$$
 undef, class(snd $t BB_{c',\kappa}') = class(snd $t B_{c^{*'}-l_b/2})$ }.$

Damit ist die Einschränkung der Kandidatenmenge abgeschlossen, und der Ersatzblock wird zufällig aus $\Gamma_{c'}^4$ ausgewählt.

Eine wichtige Bemerkung: Natürlich gibt es keine Garantie, dass sich unter den beobachteten Blöcken immer solche finden, die alle vier Einschränkungen überstehen. In der Implementierung sind die Einschränkungen daher lediglich als Versuch formuliert – hinterlässt ein Kriterium eine leere Menge, so wird es nicht angewendet. Die Erfahrung aus Experimenten mit den Daten des Elbe-Einzugsgebiets (siehe Kapitel 4) zeigt, dass die Einschränkungen auf noch nicht gesetzte Blöcke und auf solche mit passendem julianischen Datum praktisch immer nicht-leere Mengen ergibt, wohingegen die vierte Einschränkung auf Blöcke mit passenden Brückenblöcken in etwa einem Drittel der Fälle nicht angewendet werden kann. In diesen Fällen kann es also zu Brüchen (im Sinne dieses Kriteriums) kommen.

Nach Neusetzung aller Blöcke, die nach der ersten Stufe noch undefiniert waren, ist damit b^i spezifiziert und die Jahresmittel der Bezugsgröße für diesen Schritt lauten

$$\operatorname{am}^{l} = \operatorname{am}(t \ b^{l}). \tag{2.47}$$

Falls die Regressionsparameter dieser Jahresmittel, noch weiter als ϵ_{tr} von der Vorgabe *ii*) entfernt sind,

dist^m(lin_reg(amⁱ), tr) >
$$\varepsilon_{\rm tr}$$
, (2.48)

so wird zum nächsten Iterationsschritt übergeleitet.

Überleitung zum nächsten Iterationsschritt. Dass am Ende eines Iterationsschritts die Regressionsvorgabe eventuell nicht erfüllt wird, kann vor allem daran liegen, dass die Ähnlichkeitsklassen, die die class-Funktion den Blöcken des Beobachtungszeitraums bezüglich der Bezugsgröße zuordnet, notwendigerweise eine gewisse Unschärfe für diese aufweist, "ähnlich" bedeutet naturgemäß nicht "gleich". Wie in Abschnitt 2.1.3 bei der Erläuterung zur Einleitung der nächsten Iteration beschrieben, kann diese Unschärfe zusammen mit dem Grundgerüst zu einer systematischen Verfehlung der Regressionsvorgabe führen.

Während des anstehenden Iterationsschritts wird versucht, diese systematische Verfehlung dadurch auszugleichen, dass eine Regressionsvorgabe gemacht wird, die um die Differenz zwischen der Vorgabe des letzten Schritts und den dabei tatsächlich erreichten Regressionsparametern verschärft wird:

$$\begin{aligned} \mathrm{tr}_{1}^{i+1} &= \mathrm{tr}_{1} + \mathrm{tr}_{1}^{i} - \mathrm{lr}_{1}^{i}, \\ \mathrm{tr}_{2}^{i+1} &= \mathrm{tr}_{2} + \mathrm{tr}_{2}^{i} - \mathrm{lr}_{2}^{i}. \end{aligned}$$
 (2.49)

Wenn sich etwa aus dem letzten Iterationsschritt eine Regressionsgerade ergeben hat, deren Mittel niedriger als das der Vorgabe liegt, so würde der nächste Iterationsschritt eine interne Vorgabe bekommen, deren Mittel noch höher liegt als das des letzten Iterationsschritts. Durch diese Verschärfung werden weniger Blöcke aus der ersten Näherung in der ersten Stufe des Iterationsschritts beibehalten, so dass das Verfahren mehr Blöcke neu setzt und diese auch noch an einer Zielreihe mit höherem Mittel orientiert. Auf diese Weise lässt sich für die Gesamtreihe die systematische Verfehlung durch das Grundgerüst ausgleichen.

Für den ersten Iterationsschritt sind interne Vorgabe tr^1 und die Vorgabe tr aus *ii*) identisch.

Führen auch häufige¹⁷ dieser Iterationen nicht zum Ziel, wird die Iteration abgebrochen und das Verfahren beginnt von vorn mit einer neuen (bedingt zufälligen) ersten Näherung.

Anwendungen dieses Einstationenfalls finden sich in Kapitel 3, in dem Experimente mit Daten der Station Potsdam beschrieben werden.

2.3 Die Erweiterung auf mehrere Stationen

Bisher waren sowohl die Fragestellung als auch der vorgeschlagene Lösungsweg für eine einzelne Station formuliert, wodurch keine Abhängigkeiten von der meteorologischen Station berücksichtigt werden mussten. Für Simulationen für ei-

¹⁷ Experimente mit Daten des Elbegebiets haben gezeigt, dass nach mehr als 20 Iterationen, in denen die Regressionsvorgabe nicht erreicht wurde, die Hoffnung auf weitere Iterationen unbegründet ist.

ne ganze Region, repräsentiert durch ihre Stationen, ist die Situation offensichtlich eine andere. Ich werde daher zunächst die im Einstationenfall festgelegte Notation so erweitern, dass die Abhängigkeit von den Stationen berücksichtigt wird.

Bezeichne \mathscr{S} die Liste aller Stationen, deren Teilliste $S \subset \mathscr{S}$ enthalte die Stationen der untersuchten Region (die Stationen könnten in diesen Listen z. B. anhand ihrer geografischen Koordinaten angeordnet sein). Die Beobachtungsund Zeitreihen bekommen daher noch ein Superskript *s* hinzu, der ihre Station indiziert, z.B. $b^{s,obs}$ für eine Beobachtungsreihe oder für eine einzelne Beobachtung b_c^s .

Auch die Regressionsvorgabe tr beschränkt sich nicht mehr auf eine Station und wird von den Stationen abhängig:

$$\mathrm{tr}:\mathscr{S}\to\mathbb{R}^2.\tag{2.50}$$

2.3.1 Die erweiterte Fragestellung

Gesucht ist also eine Funktion, die aus den Beobachtungsreihen $b^{s,obs}$, $s \leftarrow S$ und den Regressionsvorgaben Simulationsreihen $b^{s,sim}$, $s \leftarrow S$ erzeugt.

$$g: \{\{b^{s,\text{obs}} \mid s \leftarrow S\} \mid y_1^{\text{obs}}, y_2^{\text{obs}} \in \mathbb{N}_0, \ y_2^{\text{obs}} > y_1^{\text{obs}}\} \times (S \to \mathbb{R}^2) \longrightarrow \{\{b^{s,\text{sim}} \mid s \leftarrow S\} \mid y_1^{\text{sim}}, y_2^{\text{sim}} \in \mathbb{N}_0, \ y_2^{\text{sim}} > y_1^{\text{sim}}\}$$
(2.51)

Die Eigenschaften *i*) bis *iv*) der Konsistenz, Trendvorgabe, Jahresgänge und Persistenz haben für die Simulationsreihen aller Stationen $b^{s,sim}$, $s \leftarrow S$ zu gelten, hinzu kommt die Forderung nach räumlicher Konsistenz der simulierten Felder, d. h. der Werte einer Beobachtungsvariable zu einem Kalenderdatum an allen Stationen des Untersuchungsgebiets. Die Definition eines im Sinne der Beobachtungsreihen konsistenten simulierten Feldes ist analog zur Konsistenz zwischen einzelnen Beobachtungsvariablen *i*) aufgebaut:

v) Sei die Liste der beobachteten Felder der Beobachtungsvariable $v \leftarrow [1..n]$ gegeben durch

$$X_{v}^{\text{obs}} = \left[[\text{snd} (b_{c}^{s})_{v} \mid s \leftarrow S], \ c \leftarrow C^{\text{obs}} \right].$$

Dann muss jedes simulierte Feld dieser Variable,

$$X_{v}^{\text{sim}}(c) = [\text{snd} (b_{c}^{s})_{v} | s \leftarrow S], c \leftarrow C^{\text{sim}},$$

einem beobachteten Feld aus X_v^{obs} ähneln, wobei auch hier die erlaubte Verschiedenheit durch

$$\operatorname{dist}^{e}\left(X_{v}^{\operatorname{sim}}(c), \operatorname{nn}(X_{v}^{\operatorname{sim}}(c), X_{v}^{\operatorname{obs}})\right) \leq \operatorname{dist}^{e}\left(\operatorname{nn}(X_{v}^{\operatorname{sim}}(c), X_{v}^{\operatorname{obs}}), \operatorname{nn}\left(\operatorname{nn}(X_{v}^{\operatorname{sim}}(c), X_{v}^{\operatorname{obs}}), X_{v}^{\operatorname{obs}}\right)\right).$$
(2.52)

gegeben ist, d. h. dass ein simuliertes Feld $X_v^{\text{sim}}(c)$ sich nicht mehr von seinem nächsten Nachbarn aus der Menge der beobachteten Felder, nn($X_v^{\text{sim}}(c), X_v^{\text{obs}}$), unterscheiden darf als dieser sich von seinem nächsten Nachbarn unter den beobachteten Feldern.

Der im folgenden vorgestellte Ansatz und seine algorithmische Umsetzung sind eine natürliche Erweiterung des Einstationenverfahrens. Ich möchte allerdings nochmals betonen, dass weder im Einstationen- noch im Mehrstationenfall sich die hier vorgestellten Verfahren zwingend aus der Fragestellung ergeben, sie sind vielmehr heuristisch konstruierte Lösungen, die sich in mehreren Experimenten sehr befriedigend bewährt haben (siehe Kapitel 4).

2.3.2 Der erweiterte Ansatz

Um die räumliche Konsistenz v) sicherzustellen, liegt dem auf mehrere Stationen erweiterten Verfahren ein ähnlich radikaler Ansatz zugrunde wie derjenige für die Sicherstellung der Konsistenz zwischen den einzelnen Beobachtungsvariablen: Die Beobachtungstupel aller Stationen eines Datums der Simulationsperiode entsprechen den Beobachtungstupeln aller Stationen, die an einem Datum des Beobachtungszeitraums beobachtet wurden:

 $\forall c \leftarrow C^{\text{sim}} : \exists c' \leftarrow C^{\text{obs}}, \text{ so dass } \forall s \leftarrow S$

$$\operatorname{snd} b_c^s = \operatorname{snd} b_{c'}^s. \tag{2.53}$$

Durch diesen gewählten Ansatz ist es möglich und sinnvoll, die Tupel aus den Beobachtungsreihen aller Stationen zusammenzufassen und die entstehende Reihe der Einfachheit halber mit β^{y_1,y_2} zu bezeichnen (ich verwende die für den Einstationenfall eingeführten Bezeichner als griechische Buchstaben). β^{y_1,y_2} ist also eine Liste von Paaren, deren erstes Element ein Kalenderdatum enthält und deren zweites Element aus den Beobachtungen aller Stationen aus *S* zu diesem Datum besteht:

fst
$$\beta_c = c$$
,
snd $\beta_c = [\text{snd } b_c^s | s \leftarrow S].$ (2.54)

Dieser Ansatz garantiert zwar sowohl die Konsistenz der Kombinationen verschiedener Beobachtungsvariablen (i) an allen Stationen als auch die Konsistenz der räumlichen Felder aller Variablen (v)), bringt aber erhebliche Einschränkungen mit sich. Zum einen können auf diese Weise nur Felder simuliert werden, die im Beobachtungszeitraum bereits vorkamen, was insofern eine starke Vereinfachung ist, als sich das Klima bekanntlich räumlich differenziert ändert. Zum anderen heißt das, dass falls der Ansatz angemessen ist, die Simulationsreihen *aller* Stationen in *S* die Eigenschaften *i*) bis *iv*) bereits aufweisen, wenn nur die Simulationsreihe *einer* Station aus *S* diese Eigenschaften aufweist – eine Eigenschaft des Ansatzes, die sicherlich nur zutreffen kann, wenn die Verschiedenartigkeit des Witterungsgeschehens an den einzelnen Stationen nicht zu stark ist.

Diese beiden Einschränkungen begrenzen mit der klimatischen Heterogenität auch die räumliche Ausdehnung der Region, für die die Anwendung dieses Ansatzes Sinn macht. Wie für Jahresgang und Persistenz im Einstationenfall muss auch hier *a posteriori* überprüft werden, ob die erzeugten Simulationsreihen die gewünschten Eigenschaften aufweisen.

2.3.3 Der erweiterte Algorithmus

Bei nicht zu ausgedehnten Regionen wie dem Elbe-Einzugsgebiet, für welche die Informationen aus GCM-Läufen über die Trends der Bezugsgröße räumlich sehr viel gröber aufgelöst als die Stationen voneinander entfernt sind, würde die Verwendung aller Stationen Redundanz erzeugen würde. *S* kann daher geschrumpft werden. Die Idee ist dabei, klimatisch ähnliche Stationen in Klassen zusammenzufassen und jede dieser Klassen durch eine Station zu repräsentieren. Lediglich diese repräsentativen Stationen erhalten, stellvertretend für die Stationen ihrer Klasse, jeweils eine Regressionsvorgabe für die Bezugsgröße. Sie werden zusammengefasst in der Liste $\tilde{S} \subset S$, und die Regressionsvorgabe in *ii*) wird entspannt zu

ii) Für jede repräsentative Station sollen die für diese Station vorgegebenen Parameter einer linearen Regressionsgeraden durch die Jahresmittel der Bezugsgröße innerhalb einer gesetzten Toleranz ϵ_{tr} getroffen werden:

$$\forall s \leftarrow \tilde{S} : \operatorname{dist}^{m} \left(\operatorname{lin_reg} \left(t \ b^{s, \operatorname{sim}} \right), \operatorname{tr}(s) \right) < \epsilon_{\operatorname{tr}}$$
(2.55)

(Der gewählte Ansatz impliziert, dass die Bezugsgröße auch an den restlichen Stationen sinnvolle Regressionsparameter aufweist, siehe Abschnitt 2.3.2.)

Der erste Schritt bei der Simulation mehrerer Stationen wird es daher sein, die Menge der repräsentativen Stationen zu identifizieren. Hierzu wird wieder eine Clusteranalyse verwendet, die die Stationen aus *S* bezüglich klimatischer Parameter, die mit

$$p: \{b^{y_1, y_2} \mid y_1, y_2 \in \mathbb{N}_0, y_2 \ge y_1\} \to \mathbb{R}^l$$
(2.56)

aus den Beobachtungsreihen $b^{s,obs}$, $s \leftarrow S$ bestimmt werden, in Gruppen ähnlicher Stationen einteilt (siehe auch im Anhang das Kapitel C zur Clusteranalyse). Verwendete klimatische Parameter $p(b^{s,obs})$ für die Charakterisierung der Stationen sind z. B. Mittelwert und Varianz der Tagesmitteltemperatur und des Niederschlags und Differenzen zwischen den Mitteln der ersten und zweiten Hälfte des Beobachtungszeitraums.

Ergebe die Clusteranalyse eine Einteilung in n_{rep} Klassen¹⁸ von Stationen, die bezüglich dieser klimatischen Parameter einander ähnlich sind, dann lässt sich für jede dieser Klassen ihr Schwerpunkt cm_i (*center of mass*) bestimmen:

$$\operatorname{cm}_{i} = \frac{1}{n_{i}} \sum_{\operatorname{class}(p(s))=i} p(s), \ s \leftarrow S, \ i \leftarrow [1..n_{\operatorname{rep}}], \tag{2.57}$$

wobei n_i die Anzahl der Stationen in Klasse *i* bezeichne. Über die Schwerpunkte können Stationen identifiziert werden, die den Schwerpunkten möglichst ähnlich sind und damit die entsprechende Klasse bestmöglich repräsentieren. Diese Stationen werden in der Liste \tilde{S} zusammengefasst,

$$\tilde{S} = \left[\underset{s \leftarrow S}{\operatorname{argmin}} \operatorname{dist}^{e} \left(p(b^{s, \operatorname{obs}}), \operatorname{cm}_{i} \right) \mid i \leftarrow [1..n_{\operatorname{rep}}] \right].$$
(2.58)

Wie im Einstationenfall arbeitet das Verfahren auf den Zeitskalen der Jahre und der Blöcke, nur dass hier zu einem solchen Abschnitt nicht nur die Beobachtungen einer einzelnen Station gehören, sondern die Beobachtungen aller beteiligten Stationen. A_y und B_c erfahren deshalb wie $b_c^s \rightsquigarrow \beta_c$ eine entsprechende Erweiterung:

$$fst A_y = [c \mid c \leftarrow C, fst c = y],$$

$$snd A_y = [snd \beta_c \mid c \leftarrow C, fst c = y],$$

$$fst B_c = [c' \mid c' \leftarrow [c..(c + l_b - 1)]],$$

$$snd B_c = [snd \beta_c, c' \leftarrow [c..c + l_b - 1]].$$

(2.59)
(2.59)
(2.59)

Auch der *t*-Operator muss erweitert werden, bezeichnet mit θ . Da für die Konstruktion der Simulationsreihen nur die Reihen und Regressionsvorgaben der repräsentativen Stationen relevant sind, wirkt θ lediglich auf die Beobachtungstupel aus den Reihen der repräsentativen Stationen und erzeugt ein Paar aus dem Kalenderdatum und den zugehörigen Werten der Bezugsgröße an den repräsentativen Stationen:

$$\theta \ \beta_c = \left(c, [\text{snd } t \ b_c^s | \ s \leftarrow \tilde{S}]\right). \tag{2.61}$$

Die Bezeichnung $\tilde{\theta}^i$ wird wie im Einstationenfall beim Operator *t* verwendet.

Solcherart gerüstet, komme ich zur Beschreibung der beiden Phasen, der ersten Näherung durch bedingt zufällige Neureihung ganzer Jahre und deren iterativer Verbesserung.

¹⁸ Für die Festlegung von $n_{\rm rep}$ siehe den Abschnitt 3.1.3

Die erste Näherung für regionale Simulationen

Wie im Einstationenfall beginnt die Bestimmung der ersten Näherung mit einer großen Anzahl *L* von zufälligen Neuaneinanderreihungen ganzer Kalenderjahre, aus denen eine ausgewählt wird. Im Gegensatz zum Einstationenfall, bei dem die Auswahl anhand der Parameter lediglich der einen Regressionsgeraden durch die Jahresmittel der Bezugsgröße geschieht, die diesen Neureihungen entsprechen, müssen im Mehrstationenfall die Parameter mehrerer Regressionsgeraden berücksichtigt werden.

Wie im Einstationenfall bezeichne R^l eine zufällige Neuaneinanderreihung jahresweiser Abschnitte der zusammengefassten Beobachtungsreihe β^{obs} . Für jede der repräsentativen Stationen $s \leftarrow \tilde{S}$ werden dann mit lin_reg $\left(\text{am}\left((\theta R^l)_s\right)\right)$ die Parameter der linearen Regression bestimmt und mit der zugehörigen Vorgabe tr(*s*) verglichen. Die Summe der individuellen Abweichungen ist damit das Maß, anhand dessen eine der *L* Neureihungen ausgewählt wird:

$$\beta^{\text{start}} = \underset{R^l, \ l = [1..L]}{\operatorname{argmin}} \sum_{s \leftarrow \tilde{S}} \operatorname{dist}^m \left(\lim_{s \leftarrow \tilde{S}} \operatorname{reg}((\theta \ R^l)_s), \operatorname{tr}(s) \right).$$
(2.62)

Die iterative Verbesserung für regionale Simulationen

Bei der iterativen Verbesserung sind die Abläufe ebenfalls ganz analog zum Einstationenfall: Im ersten Schritt wird darüber entschieden, welche Blöcke aus der ersten Näherung beibehalten werden, im zweiten werden die benötigten Austauschblöcke bestimmt. Im Unterschied zum Einstationenfall, bei dem die Blöcke lediglich durch die Bezugsgröße charakterisiert werden, wie sie an ihren Tagen an der einen Station gemessen wurden, besteht die Charakterisierung eines Blockes hier aus der Bezugsgröße, wie sie an den Tagen des Blocks an allen repräsentativen Stationen $s \leftarrow \tilde{S}$ beobachtet wurden. Ein Block wird also durch $l_b n_{rep}$ Werte der Bezugsgröße charakterisiert. Die Klassifizierung, die in beiden Schritten eine wichtige Rolle spielt, wird entsprechend in den höherdimensionalen Raum übertragen,

class:
$$\mathbb{R}^{l_b \cdot n_{\text{rep}}} \rightarrow \{1, 2, \dots, n_c\},$$
 (2.63)

und durch eine Clusteranalyse spezifiziert, die mit den Charakterisierungen der " θ -Blöcke" des Beobachtungszeitraums {snd $\theta B_c | B_c \subset \beta^{obs}$ } trainiert wird.

Blöcke beibehalten. Wie im Einstationenfall wird die Entscheidung darüber, welche Blöcke der ersten Näherung beibehalten werden, durch den Vergleich der Klassenzugehörigkeiten eines Blockes getroffen, der einmal durch die Bezugsgröße entsprechend der ersten Näherung, einmal durch die Bezugsgrößen der "Zielreihen" charakterisiert wird. Die n_{rep} Zielreihen gehen dabei aus den

Reihen für die Bezugsgröße der ersten Näherung hervor, die individuell jahresweise so angehoben bzw. abgesenkt werden, dass die Regressionsvorgaben getroffen werden. Sie werden ganz analog zum Einstationenfall erzeugt,

$$\tilde{t}^i b_c^{s,\text{start}} = t \ b_c^{s,\text{start}} + (\widetilde{\text{am}}_{s,y}^i - \text{am}_{s,y}^{\text{start}}), \ c \leftarrow \text{fst } A_y, \ y \leftarrow \text{fst } C^{\text{sim}}, \ s \leftarrow \tilde{S}, \quad (2.64)$$

und unter dem Operator $\tilde{\theta}^i$ zusammengefasst entsprechend Gleichung (2.61):

$$\tilde{\theta}^{i} \beta_{c}^{\text{start}} = (c, [\tilde{t}^{i} \ b_{c}^{s, \text{start}} | s \leftarrow \tilde{S}]).$$
(2.65)

Damit kann der erste Teil von β^i spezifiziert werden:

$$c \leftarrow \text{CB: class(snd } \theta \ B_c^{\text{start})} = \text{class(snd } \tilde{\theta}^i \ B_c^{\text{start})} \Longrightarrow B_c^i = B_c^{\text{start}}.$$
 (2.66)

Die Blöcke aus β^i , die auf diese Weise noch nicht spezifiziert sind, enthalten wie im Einstationenfall die undef-Markierung.

Blöcke neu setzen. Auch hier wird für jeden $B_{c'}$ mit $c' \leftarrow CB$ | class(snd $\theta B_{c'}^{\text{start}}$) \neq class(snd $\tilde{\theta}^i B_{c'}^{\text{start}}$) aus einer in vier Schritten eingeschränkten Menge potentieller Blöcke ausgewählt. Die Kriterien, anhand derer diese Menge eingeschränkt wird, sind wie im Einstationenfall,

1. die Menge aller Blöcke B_{κ} der aggregierten Beobachtungsreihe β^{obs} , die bezüglich der von den repräsentativen Stationen zusammengefügten Bezugsgrößen snd θ B_{κ} in die gleiche Klasse gehören wie der zu besetzende Block $B_{c'}$ bezüglich der von den repräsentativen Stationen zusammengefassten Zielreihen der Bezugsgröße snd $\tilde{\theta}^i$ $B_{c'}^{\text{start}}$ (vergleiche Gleichung (2.41)),

$$\Gamma_{c'}^{1} = \{B_{\kappa} \mid \text{class(snd } \theta \mid B_{\kappa}) = \text{class(snd } \tilde{\theta}^{i} \mid B_{c'}^{\text{start}}), B_{\kappa} \subset \beta^{\text{obs}}\}; \qquad (2.67)$$

- 2. aus dieser Menge nur diejenigen Blöcke des Beobachtungszeitraums, deren Beobachtungen bislang keinem Block aus β^{sim} zugeordnet werden (vergleiche Gleichung (2.42));
- 3. daraus nur diejenigen Blöcke des Beobachtungszeitraums, deren julianisches Datum innerhalb der gesetzten Toleranz ϵ_{jd} mit dem julianischen Datum des zu besetzenden Blocks verträglich ist (vergleiche Gleichung (2.43));
- 4. daraus nur diejenigen Blöcke, die zu einem bereits gesetzten Vorgänger und, falls schon definiert, auch zu ihrem Nachfolger passen. Vergleiche

hierzu Gleichung (2.46) und die vorangehenden Definitionen der Brückenblöcke. Für den Mehrstationenfall werden die Klassenzugehörigkeiten der Brückenblöcke wiederum bezüglich θ bestimmt:

$$\Gamma_{c'}^4 = \{B_{\kappa} \mid B_{\kappa} \in \Gamma_{c'}^3, \text{ class}(\text{snd }\theta \text{ BB}_{c',\kappa}) = \text{class}(\text{snd }\theta B_{c^*+l_b/2}), \text{ und, } (2.68)$$

falls $B_{c'+l_b} \neq \text{undef, class}(\text{snd }\theta \text{ BB}_{c',\kappa}') = \text{class}(\text{snd }\theta B_{c^{*'}-l_b/2}).$

Aus $\Gamma_{c'}^4$ wird dann ein Block zufällig ausgewählt.

Nachdem alle Blöcke, die nach der ersten Stufe noch nicht "versorgt" waren, auf diese Weise abgearbeitet sind, ist die Konstruktion von β^i dieses Iterationsschritts *i* vollständig abgeschlossen und es bleibt zu überprüfen, ob die damit erzeugten Reihen der Bezugsgröße bereits mit der Regressionsvorgabe *ii*) verträglich sind. Dazu werden die Jahresmittel der Bezugsgröße für alle repräsentativen Stationen am $((\theta \ \beta^i)_s), \ s \leftarrow \tilde{S}$ und ihre Regressionsparameter bestimmt und für jede repräsentative Station geprüft, ob diese Gerade mit der jeweiligen Vorgabe tr(*s*) verträglich ist. Nur wenn diese Verträglichkeit für alle repräsentativen Stationen gegeben ist, wird β^i als β^{sim} akzeptiert und die Iteration beendet:

$$\forall s \leftarrow \tilde{S} : \operatorname{dist}^{m} \left(\operatorname{lin}_{\operatorname{reg}} \left((\theta \ \beta^{i})_{s} \right), \operatorname{tr}(s) \right) \leq \epsilon_{\operatorname{tr}}.$$

$$(2.69)$$

Überleitung zum nächsten Iterationsschritt. Ist das Abbruchkriterium (2.69) nicht erfüllt, werden intern verschärfte Regressionsparameter für den nächsten Iterationsschritt vorgegeben. Im Gegensatz zum Einstationenfall müssen hier die Vorgaben aller repräsentativer Stationen angepasst werden:

$$\begin{aligned} \mathrm{tr}_{1}^{i+1}(s) &= \mathrm{tr}_{1}(s) + \mathrm{tr}_{1}^{i}(s) - (\mathrm{lr}_{1}^{i})_{s}, \\ \mathrm{tr}_{2}^{i+1}(s) &= \mathrm{tr}_{2}(s) + \mathrm{tr}_{2}^{i}(s) - (\mathrm{lr}_{2}^{i})_{s}, \ s \leftarrow \tilde{S}. \end{aligned}$$
 (2.70)

Kapitel 3

Kalibrierung und Begrenzung des Verfahrens

In der Beschreibung meines Verfahrens in Kapitel 2 tauchen einige Parameter auf, die dem Verfahren subjektiv vorgegeben werden, z. B. die Blocklänge l_b oder die beiden Toleranzen ϵ_{jd} und ϵ_{tr} . Anhaltspunkte für ihre jeweilige Einstellung werde ich in diesem Kapitel erläutern und die Sensitivität des Verfahrens für diese Einstellungen untersuchen. Viele dieser Überlegungen gründen sich auf beobachtete und simulierte Daten der Station Potsdam, die eine ununterbrochene Reihe täglicher Beobachtungen seit 1893 weltweit einzigartiger Homogenität bietet. Diese Länge gestattet es, für Kreuzvalidierungsexperimente ausreichend lange Teilperioden herauszugreifen.

Sie dienen ebenfalls dem Versuch einer Abschätzung, wie klimatologisch verschieden Beobachtungs- und Simulationsperiode sein dürfen, d. h. welche Regressionsvorgaben mit einem gegebenen Datensatz umgesetzt werden können.

Allgemein wird es nicht offensichtlich sein, dass die aus solchen Einstationenexperimenten gezogenen Schlussfolgerungen direkt auf Mehrstationensimulationen übertragen werden können. In diesem Fall ist es aber aus den Überlegungen in Abschnitt 2.3.2 heraus zulässig: Dadurch, dass der Ansatz impliziert, dass die Simulationsreihen aller Stationen die geforderten Eigenschaften i)-iv) aus Abschnitt 2.2.2 aufweisen, wenn bereits eine Station dies tut, reicht die Betrachtung einer Station des Untersuchungsgebiets aus.

3.1 Kalibrierung

3.1.1 Wahl der Blocklänge

Die Blöcke als kleinste zeitliche Einheit zu verwenden, garantiert realistische Abfolgen und Erhaltungsneigung innerhalb der Blöcke. Je länger die Blöcke, desto längere realistische Abfolgen sind garantiert. Auf der anderen Seite schrumpft

Station	T _{mit}	RR	р
Grambeck	8.35	1.85	5.64
Potsdam	7.65	1.47	5.52
Gera	7.56	1.41	6.07
Churanov	6.13	1.89	6.54

Tabelle 3.1: Dekorrelationszeiten τ_d (in Tagen) für Tagesmittel der Lufttemperatur T_{mit} , Niederschlag *RR* und Luftdruck *p* an verschiedenen Stationen des Elbe-Einzugsgebiets.

mit zunehmender Blocklänge die Zahl der Bausteine, aus denen eine Simulationsreihe zusammengesetzt wird, was die Variabilität der verschiedenen Simulationen beschränkt. Grundsätzlich sollte es daher das Ziel sein, die kürzestmögliche Blocklänge zu verwenden, die noch für akzeptable Persistenz in den Simulationsreihen sorgt. Generell lässt der gewählte Ansatz erwarten, dass die Persistenz der Simulationsreihen geringer ausfällt als in der Trainingsreihe b^{obs} , da durch das Neuaneinandersetzen der Blöcke aus b^{obs} Brüche entstehen, die zu geringeren Autokorrelationen in b^{sim} führen.

Ein einfaches Maß für die Persistenz in einer Zeitreihe ist die Dekorrelationszeit τ_D . Für ihre Schätzung wird die acf-Funktion aus Gleichung (2.9) verwendet. Vorbereitend wird die Zeitreihe durch das STL-Verfahren (siehe Abschnitt D) in Trend-, Jahresgang- und Rauschkomponente zerlegt. Aus der Rauschkomponente [r_i] schätzt sich die Dekorrelationszeit zu (Taubenheim, 1969; von Storch und Zwiers, 1999)

$$\tau_d = 1 + 2\sum_{l=1}^{n'} \left(\frac{n-l}{n} \operatorname{acf}_{[r_i]}(l)\right), \tag{3.1}$$

wobei $acf_{[x_i]}(l)$ den Wert der Autokorrelationsfunktion der Verschiebung l für den Prozess $[X_i]$ an einer Realisierung, der Zeitreihe $[x_i]$ der Länge n, schätzt,

$$\operatorname{acf}_{[x_i]}(l) = \frac{1}{\operatorname{acf}_{[x_i]}(0)} \sum_{j=l}^n (x_j - m(x))(x_{j-l} - m(x)). \tag{3.2}$$

dabei sollte lediglich $l \leftarrow [0..n']$ bis zu einem sinnvollen maximalen n', z. B. dasjenige an dem der Wert der Autokokorrelationsfunktion das erste Mal nicht mehr signifikant von 0 verschieden ist (siehe z. B. Schönwiese, 2000). m(x) bezeichnet das arithmetische Mittel der Beobachtungen aus der Zeitreihe $[x_i]$.

 τ_D ist ein Schätzwert dafür, wie weit der Einfluss eines bestimmten Messwerts auf folgende Messungen reicht, bemisst also das Gedächtnis oder die Erhaltungsneigung des Prozesses, der der beobachteten Zeitreihe zugrunde liegt.

In Tabelle 3.1 sind einige Dekorrelationszeiten für die repräsentativen Stationen des Elbegebiets dargestellt (vergleiche Kapitel 4). Erwartungsgemäß besitzen unterschiedliche Größen unterschiedliche Erhaltungsneigungen, wobei der Niederschlag die geringste aufweist. Insgesamt sind die $\tau_D < 9$. Da die Blocklänge die Persistenz in den Simulationsreihen entscheidend kontrolliert und damit für die Einhaltung der Eigenschaft *iv*) vital ist (siehe Abschnitt 2.2.1), habe ich ihren Einfluss auf die simulierten Dekorrelationszeiten systematisch untersucht. Dafür habe ich Blocklängen zwischen 2 und 100 Tagen verwendet, um das Potsdamer Klima 1951–2003 aus den Daten von 1893–1950 zu simulieren. Da die Temperatur allgemein eine wichtige atmosphärische Zustandsgröße ist – sie ist ein Maß für die Energie eines Luftpakets und kontrolliert viele Vorgänge wie etwa Verdunstung entscheidend mit – und außerdem in diesen Zeiträumen ein wichtiger Indikator für die klimatische Entwicklung ist, bestand die Regressionsvorgabe aus den Parametern, die eine Regressionsanalyse der Jahresmitteltemperaturen 1951–2003 ergeben hatte. Für jede Blocklänge wurden 100 Simulationen erstellt.

Für die drei Beobachtungsvariablen Tagesmittel der Lufttemperatur, Niederschlag und Luftdruck wurde für jede Simulation τ_D mit (3.1) bestimmt. Die 100 Schätzungen zu jeder Blocklänge werden für jede der drei Variablen in einem Boxplot¹ dargestellt, siehe Abbildungen 3.1, 3.2 und 3.3. Die blaue waagrechte Linie zeigt das jeweilige τ_D der Trainingsreihe an, die rote dasjenige für die Referenzperiode.

Für die Temperatur (Abbildung 3.1) ergibt sich folgendes Bild: Eindeutige Auswirkungen im Sinne einer Steigerung der Persistenz mit zunehmender Blocklänge lassen sich erst für sehr lange Blocklängen ($l_b \ge 30$) ausmachen. Für kürzere Blocklängen ist das Bild diffus, jedoch werden die Persistenzen der Trainingsreihe stets unterschätzt. Erst die Blocklänge von 100 Tagen kommt wirklich in den Bereich der Trainingsreihe. Da in diesem Fall ein Simulationsjahr fast nur aus 3 Blöcken zusammengesetzt würde (was wenig Sinn macht), muss man diese Unterschätzung in Kauf nehmen. In der Kreuzvalidierung (Kapitel 4.3.2) zeigt sich allerdings, dass trotz dieser Unterschätzung das Andauernverhalten der untersuchten Ereignistage befriedigend simuliert wird.

Einfacher ist es für den Niederschlag (Abbildung 3.2): Durch die bereits in b^{obs} viel geringere Erhaltungsneigung dieser Variable passen die Simulationen schon ab $l_b = 10$ gut zur in b^{obs} festgestellten Persistenz. Die Erhaltungsneigung des Niederschlags ist für längere Blocklängen stabil.

Die Persistenz des Luftdrucks (Abbildung 3.3) wird ebenfalls mit wachsender Blocklänge der Persistenz in b^{obs} passender. Einem deutlichen Zuwachs der Persistenz bis etwa $l_b = 12$ folgt eine Sättigung für größere l_b , in der sich die simulierte Persistenz nur noch geringfügig verbessert.

Aufgrund dieser Auswertung (die Unterschätzung der Persistenz der Tempe-

¹ Ein Boxplot ist eine vereinfachende Darstellung einer empirischen Verteilung: Der fette waagrechte Strich bezeichnet den Median, die Box darum wird vom 25%- bzw. 75%-Quantil begrenzt, und der Rest einschließlich Ausreißer befindet sich zwischen den beiden Enden der gestrichelten Linien.



Dekorrelationszeit Tmit

Abbildung 3.1: In den Boxen τ_D für jeweils 100 Simulationen zu verschiedenen Blocklängen für die Tagesmitteltemperatur. Rot: Geschätztes τ_D für beobachtete Reihe der Referenzperiode, in blau die Schätzung für die Reihe der Trainingsperiode.

raturreihen verschwindet erst bei unpraktikablen großen l_b , die Erhaltungsneigung des Niederschlags scheint ab $l_b = 10$ befriedigend und der Luftdruck hat die gewünschte Persistenz im wesentlichen bei $l_b = 12$ erreicht), verwende ich $l_b = 12$ für die weiteren Simulationen.

3.1.2 ϵ_{jd} – die Toleranz beim julianischen Datum

Die dritte Einschränkung der Menge potentieller Ersatzblöcke, die nicht aus der ersten Näherung übernommen werden, geschieht anhand des julianischen Datums der Blöcke, siehe Abschnitt 2.2.3. Die julianischen Anfangsdaten der Ersatzblöcke, die dieses Kriterium überstehen, liegen maximal um ϵ_{jd} von dem des zu besetzenden Blocks entfernt. Dadurch wird sichergestellt, dass nur jahreszeitlich passende Blöcke in der Auswahl bleiben, was allein durch die Charakterisierung der Blöcke durch die Bezugsgröße eventuell nicht gewährleistet ist. ϵ_{jd} sollte dabei keinesfalls kleiner als nötig gewählt werden, um möglichst viele Blöcke übrig zu behalten, die noch das vierte Einschränkungskriterium überleben können, das dafür sorgt, dass sich der ausgewählte Block möglichst ohne Bruch in die Reihe der bereits gesetzten einfügt. Die nachfolgende Untersuchung stützt sich auf die Potsdamer Reihe von 1951–2003.



Abbildung 3.2: Wie Abbildung 3.1, jedoch für den Niederschlag.



Dekorrelationszeit Druck

Abbildung 3.3: Wie Abbildung 3.1, jedoch für den Luftdruck.

Einen Hinweis für die sinnvolle Festlegung der Toleranz versuche ich über die Frage zu erhalten: Gegeben sei ein beliebiger Block aus einem "prototypischen" Jahr, an welchen julianischen Daten in den Jahren der Beobachtungsperiode beginnen Blöcke, deren Beobachtungstupel zu dem aus dem prototypischen Jahr ähnlich sind und wie weit sind sie (bezüglich des julianischen Datums) von diesem entfernt?

Für die Beantwortung dieser Frage werden die Tage durch Tagesmittel und Amplitude der Temperatur ($A = T_{max} - T_{min}$) sowie den Niederschlag charakterisiert und aus diesen Blöcke der Länge 12 Tage gebildet. Die 36-Tupel dieser Blöcke ($T_1, ..., T_{12}, A_1, ..., A_{12}, ..., RR_1, ..., RR_{12}$) werden mit einer Clusteranalyse (siehe Abschnitt C) zu Klassen ähnlicher Tupel zusammengefasst.² Jedem Block des Beobachtungszeitraums entspricht auf diese Weise eine Ähnlichkeitsklasse, in die das an ihm beobachtete Tupel gehört.

Um aus der Beobachtungsreihe einer Station die Reihe eines "prototypischen Jahrs" zu bestimmen, wird für jede der drei Beobachtungsvariablen der mittlere Jahresgang berechnet, indem jeweils über die Beobachtungen zu Tagen mit gleichem julianischen Datum gemittelt wird. Die Beobachtungsreihen werden zuvor um den linearen Trend in den Jahresmitteln bereinigt. Aus den Reihen des prototypischen Jahrs werden wiederum Blöcke gebildet, wobei jedem Block des prototypischen Jahrs die Ähnlichkeitsklasse zugeordnet wird, in die sein 36-Tupel gehört.

Vergleicht man nun die julianischen Daten, an denen die Klassen im prototypischen Jahr auftreten, mit jenen, an denen die Klassen in den tatsächlich beobachteten Jahren auftreten, so erhält man eine Abschätzung für die Streubreite der Positionen im Jahr der einzelnen Klassen. Dies ist in Abbildung 3.4 illustriert: Dargestellt ist rechts eine synthetische Temperaturreihe von drei Jahren, links der mittlere Jahresgang aus diesen drei Jahren, das prototypische Jahr. Exemplarisch sind für eine Klasse die Tage, deren Beobachtungen in diese gehören, rot markiert. Außerdem zeigt eine senkrechte rote Linie für jedes Jahr den Median der julianischen Daten an, die zu dieser Klasse gehören.

In Abbildung 3.5 sind die Streubreiten bezüglich des julianischen Datums für alle Klassen, die im prototypischen Jahr auftauchen, dargestellt. Dazu wird für jedes Jahr der Beobachtungsreihe der Median der julianischen Daten bestimmt, denen eine bestimmte Klasse entspricht (siehe hierzu noch einmal Abbildung 3.4). Ebenso wird der entsprechende Median für das prototypische Jahr ermittelt. Die Boxplots geben die Verteilung der Differenzen zwischen den Medianen in den 53 beobachteten Jahren und dem Median des prototypischen Jahrs an. Ein Großteil der Abweichungen bewegt sich im Bereich von ± 20 (die roten waagrechten Linien).

 $^{^2}$ Als Cluster-Anzahl wurde hier 40 gewählt. Experimente mit verschiedenen Cluster-Anzahlen zwischen 10 und 100 ergaben stets vergleichbare Ergebnisse.



Abbildung 3.4: Abschätzung von ϵ_{jd} . Rechts eine synthetische Temperaturreihe der Länge drei Jahre, links der mittlere Jahresgang. In rot markiert die Kalenderdaten, deren Beobachtungen in eine Klasse fallen. Die senkrechten roten Linien markieren die Mediane dieser julianischen Daten der einzelnen Jahre.

Diese Abbildung wird so interpretiert, dass klimatisch ähnliche Blöcke einem "normalen" Jahresablauf ohne weiteres um bis zu 20 Tagen versetzt auftreten können. 20 Tage scheint damit für die Toleranz beim julianischen Datum ϵ_{jd} nicht zu hoch gewählt.

Eine weitere Überlegung ist in diesem Zusammenhang von Bedeutung: Dieses Toleranzkriterium lässt Blöcke zu, deren astronomisch mögliche Sonnenscheindauer um bis zu 10% die astronomisch mögliche Sonnenscheindauer des zu besetzenden Blocks übersteigt (die 10% wurden für die geografische Breite von Potsdam und $\epsilon_{id} = 20$ bestimmt).

Möchte man sich gegen diesen Fall absichern, ist eine Möglichkeit, das Toleranzfenster assymetrisch anzulegen: Für Blöcke vor der Sommersonnenwende kommen nur Blöcke in Frage, deren julianische Daten um bis zu ϵ_{jd} vor denen des zu besetzenden Blocks lagen, nach der Sommersonnenwende werden entsprechend nur Blöcke zugelassen, deren julianische Daten um bis zu ϵ_{jd} nach den zu besetztenden liegen.³

³ Alternativ können die simulierten Reihen nachkorrigiert werden, indem die simulierten Sonnenscheindauern, die die astronomisch möglichen überschreiten, entsprechend verringert werden. Eine solche Korrektur stellt einen Eingriff in die Beobachtungstupel dar und entspricht damit nicht dem hier gewählten Ansatz. Da diese Korrekturen jedoch sehr klein ausfallen werden, stel-



Bereich der julianischen Daten verschiedener Klassen

Abbildung 3.5: Zur Festlegung von ϵ_{jd} . Dargestellt sind die Abweichungen der Media-

Abbildung 3.5: Zur Festlegung von ϵ_{jd} . Dargestellt sind die Abweichungen der Mediane der julianischen Daten zu einer Klasse für jedes Jahr, bezogen auf den zugehörigen Median des prototypischen Jahrs.

Ich habe sowohl mit symmetrischen als auch mit asymmetrischen Fenstern mit den Daten des Elbe-Einzugsgebiets experimentiert, allerdings bei den simulierten Sonnenscheindauen keine Unterschiede in den Statistiken festgestellt. Für diese Arbeit verwende ich daher die symmetrische Variante.

3.1.3 Sonstige Parameter

Anzahl der Klassen für die Beobachtungsblöcke

Die Anzahl der Klassen, in die die beobachteten Blöcke eingeteilt werden, ist nicht objektiv definiert. Hinweise auf eine sinnvolle Klassenanzahl kann wie in Abschnitt C.3 beschrieben eine Kurve liefern, in der für verschiedene Klassenanzahlen die Gesamtsumme der klasseninternen Varianzen dargestellt ist (siehe Gleichung C.2). Für Daten des Elbegebiets legt diese Kurve eine Klassenanzahl > 10 nahe (siehe Abbildung 3.6), üblicherweise verwende ich 40 Klassen, um eine zu grobe Generalisierung auszuschließen. Bei Experimenten mit Klassenanzahlen zwischen 20 und 100 wurde kein Effekt dieser Wahl auf die Ergebnisse beobachtet.

len sie eine praktikable Option dar, die hier allerdings nicht verfolgt wird.



Abbildung 3.6: Zur Festlegung der Klassenanzahl für die durch die Bezugsgröße charakterisierten Blöcke des Beobachtungszeitraums. Dargestellt ist für verschiedene Klassenanzahlen die Summe aller klasseninternen Varianzen. Erkennbar ist ein Knick bei um die 10 Klassen, der anzeigt, dass mit weniger Klassen die Generalisierung durch die Klassen unbrauchbar wird.

Anzahl der Klassen für die Stationen

Die Anzahl der Cluster klimatisch ähnlicher Stationen bzw. der repräsentativen Stationen für regionale Simulationen ist vom Anwender festzulegen, gestützt auf die Kenntnis der beobachteten Klimatologie der Region. Hinweise auf eine sinnvolle Klassenanzahl kann darüber hinaus wie im vorangegangenen Abschnitt die Kurve liefern, in der die Summe aller klasseninternen Varianzen in Abhängigkeit von der Klassenanzahl aufgetragen sind.

Grundsätzlich sollte die geringste Clusteranzahl gewählt werden, die eine ausreichende räumliche Differenzierung bietet – je mehr repräsentative Stationen, in desto höherdimensionalem Parameterraum findet die Konstruktion statt, so dass die einzelnen Stationen relativ weniger Gewicht bei der Auswahl der Blöcke erhalten.

$\epsilon_{\rm tr}$ – Die Toleranz bei der Regressionsvorgabe

Die Toleranzen festzulegen, innerhalb derer die Regressionsvorgabe erfüllt werden muss, ist dem Anwender überlassen und allenfalls dadurch eingeschränkt, dass besonders bei der Erzeugung regionaler Simulationen mit mehreren Bezugsstationen zu geringe Toleranzen auch nach vielen Iterationen nicht unterschritten werden – ein Effekt, der sich der Abnahme des relativen Gewichts der einzelnen Stationen schuldet, die aus der mit der Anzahl der Bezugsstationen wachsenden Dimensionalität entsteht. Bei Experimenten mit Daten des Elbe-Einzugsgebiets erwiesen sich Toleranzen als realisierbar, die einem maximalen Abstand von vorgegebener zu simulierter Regressionsgerade von 0.1-0.2K im gesamten Simulationszeitraum entsprechen. Für Ein-Stationen-Rechnungen können auch geringere Toleranzen, etwa 0.05 - 0.1K entsprechend, umgesetzt werden.

3.2 Begrenzungen

Dieser Abschnitt widmet sich der Frage, wie weit sich – exemplarisch für die Temperatur als Bezugsgröße – die vorgeschriebene lineare Entwicklung vom Klima des Beobachtungszeitraums entfernen darf.

Beispielsweise zeigt die Potsdamer Temperaturreihe in den Jahresmitteln 1951–2003 eine lineare Zunahme von 1.2K. Aus diesem Datensatz eine Annäherung an ein um 10K wärmeres Klima gewinnen zu wollen, ist offensichtlich sinnlos. Wie steht es aber mit einer linearen Zunahme um weitere 1.2K?

Gesucht ist also eine Größe, deren Erhaltung von den Beobachtungen zur Simulation anzeigt, dass die Simulation aus den Beobachtungsdaten darstellbar ist. Eine statistische Trennung dieser Größe, bestimmt einmal für die Beobachtungs- und einmal für die simulierten Reihen, bedeutete dann, dass die Regressionsvorgabe für den Simulationszeitraum den Beobachtungsdatensatz überfordert.

Eine solche Größe liegt nicht auf der Hand, da Unterschiede zwischen simulierter und beobachteter Klimatologie mit meinem Verfahren gerade dargestellt und untersucht werden sollen. Die Unveränderlichkeit der mittleren Temperatur oder irgendeines anderen Mittels kommt daher nicht als Ziel in Frage. Auch Jahresgänge werden sich plausiblerweise ändern, so dass die Forderung ihrer Erhaltung physikalisch nicht sinnvoll ist. Die Idee ist daher, als Erhaltungsgröße das zu verwenden, was außer Jahresgängen und langfristigen Trends in den simulierten Zeitreihen fluktuiert, das Rauschen: Sind die Verteilungen der Rauschen zu b^{sim} und b^{obs} nicht unterscheidbar, so bietet b^{obs} offensichtlich genügend Variabilität, um daraus trotz unterschiedlicher Trend- und Jahresgangkomponenten b^{sim} zusammenzusetzen. Aus diesem Grund ist in der geforderten Eigenschaft *iii*), die die Forderung realistischer Jahresgänge charakterisiert, die statistische Ununterscheidbarkeit der Verteilungen der Rauschen aus beobachteter und simulierter Reihe verlangt.

Um zu untersuchen, welche Regressionsvorgaben sich mit den Daten der



Abbildung 3.7: p-Werte des KS-Tests in Abhängigkeit von der der Simulation vorgegebenen Erwärmung (Punkte). Die waagrechte Linie markiert eine Irrtumswahrscheinlichkeit von 5%.

Potsdamer Beobachtungsreihe simulieren lassen, habe ich einen Satz von Simulationen erstellt, die für die Jahre 2004–2055 unterschiedliche lineare Temperaturzunahmen vorgeschrieben bekamen. Insgesamt wurden mit den Beobachtungen von 1951–2003 als Trainingsdaten 10 Simulationen gerechnet, mit linearen Temperaturzunahmen von 0.2, 0.4, ..., 2.0K ab 2004.

Ich verwende für die Zerlegung von beobachteter und simulierter Temperaturreihe wiederum das in Anhang D beschriebene STL-Verfahren. Für den Vergleich der beiden Rauschverteilungen bietet sich der Kolmogorov-Smirnov-Test (KS-Test) für zwei Stichproben an, der die Nullhypothese prüft, ob zwei Stichproben aus der selben Grundgesamtheit stammen (siehe z. B. von Storch und Zwiers, 1999 und Abschnitt B.3).

Abbildung 3.7 zeigt die p-Werte des KS-Tests, also die Wahrscheinlichkeit, dass beide Rauschen aus der gleichen Grundgesamtheit stammen, für die Experimente mit der Potsdamer Reihe: Ab der Erwärmung um 1.2K identifiziert dieses Kriterium die Simulationen bei einer Irrtumswahrscheinlichkeit von 5% als unzulässig.

Kapitel 4

Validierung des Verfahrens

Mein Verfahren versucht, eine Antwort auf die Frage nach realistischen Klimaentwicklungen einer Region zu geben, die der vorgegebenen Entwicklung einer für das Klima der Region charakteristischen Beobachtungsvariable entsprechen. Da diese einzige äußere Randbedingung nur eine sehr vage Einschränkung darstellt, ist der Weg zur Antwort nicht eindeutig und muss heuristisch beschritten werden.

Allgemein ist es bei heuristischen Verfahren nicht trivial, ihre Verwendbarkeit zu zeigen. Pauschale Bewertungen im Stil von "Das Verfahren funktioniert." können nicht vergeben werden – im Gegensatz zu analytischen Problemen, deren Lösungen anhand des Problems eindeutig auf Korrektheit überprüft werden können.

Das "Funktionieren" heuristischer Verfahren lässt sich daher nur an Fallstudien untersuchen, für die zum einen die Robustheit der Antworten in Abhängigkeit von verfahrensinternen Annahmen, z. B. der Einstellung bestimmter Parameter, geprüft werden kann. Zum anderen können in Kreuzvalidierungsexperimenten, bei denen dem Verfahren gewissermaßen eine Frage mit bekannter Antwort gestellt wird, die Antwort des Verfahrens mit der "richtigen" Antwort verglichen werden.

Die Vorgaben und Parameter-Einstellungen sowie die Robustheit der Ergebnisse in Abhängigkeit von ihnen wurden bereits in Kapitel 3 behandelt. Dieses Kapitel stellt nun die Auswertung eines Kreuzvalidierungsexperiments mit Daten des Elbe-Einzugsgebiets vor, ergänzt durch einen kurzen Vergleich der Ergebnisse meines Verfahrens mit denen des REMO-Modells¹ aus einem vergleichbaren Experiment.

Die Wahl für die Bezugsgröße fiel auf die Temperatur – das verdankt sich dem klaren Signal, mit dem die Temperaturreihen in dieser Region den Klimawandel nachzeichnen. Darüber hinaus ist die Temperatur allgemein eine Größe, die

¹ Das **RE**gional-**MO**dell des Max-Planck-Instituts für Meteorologie in Hamburg.

eng mit dem Energiehaushalt der Atmosphäre zusammenhängt und wesentlichen Einfluss auf viele Prozesse in Boden und Luft nimmt, etwa auf die für den Wasserhaushalt so wichtige Verdunstung. Sie ist damit eine besonders wichtige Zustandsvariable der Atmosphäre.

Die meisten Abbildungen und Auswertungen wurden mit dem frei verfügbaren Statistikpaket R (R Development Core Team, 2004) erstellt, Karten mit der ebenfalls freien Software GMT (Wessel und Smith, 2006).

4.1 Aufbau des Experiments

Für das Elbe-Einzugsgebiet steht ein hochwertiger Datensatz täglicher Beobachtungen – bestehend aus Maximum, Mittel und Minimum der Lufttemperatur, Niederschlag, relativer Feuchte, Luftdruck, Dampfdruck, Sonnenscheindauer, Bedeckungsgrad, Globalstrahlung und mittlerer Windgeschwindigkeit – für die Jahre 1951–2003 zur Verfügung. Dieser wurde von Österle am PIK aufbereitet und homogenisiert, siehe Abschnitt E sowie Österle et al. (2006). Für das Kreuzvalidierungsexperiment werden daraus zwei 25-Jahresperioden herausgegriffen: 1951–1975 als Trainingsperiode und 1976–2000 als Referenzperiode. Mit der Vorgabe der beobachteten Temperaturtrends und -mittelwerte der Referenzperiode soll die Klimatologie der Referenzperiode simuliert werden, wobei ausschließlich die Daten der Trainingsperiode verwendet werden. Aus dem Vergleich von simulierter und beobachteter Klimatologie der Referenzperiode lässt sich auf diese Weise die Leistungsfähigkeit des Verfahrens bewerten.

Die Festlegung dieser beiden Abschnitte macht den Test sehr anspruchsvoll – ihre beiden Klimatologien sind einander ziemlich verschieden, d. h. dass der Ansatz, die zu simulierende Klimatologie aus einer Neuaneinanderreihung vergangener Beobachtungen zu gewinnen, sich erst als brauchbar erweisen muss.

Einen ersten Eindruck von der Verschiedenartigkeit der beiden Zeiträume gewinnt man beim einfachen Blick auf die Jahresmitteltemperaturen dieser Zeitspanne, dargestellt am Beispiel Potsdam in Abbildung 4.1. Während die Temperaturen in der ersten Hälfte mehr oder weniger um ein konstantes Niveau fluktuieren, findet in der zweiten Hälfte ein rapider Anstieg statt, der die Temperaturen insgesamt auf ein wärmeres Niveau hebt.

Auch in den großräumigen Zirkulationsmustern der Atmosphäre über Europa finden sich zwischen den beiden Zeiträumen gewaltige Unterschiede. Die Abbildung 4.2 zeigt die mittleren Andauern der zonalen, meridionalen und gemischten Großwetterlagen nach Hess und Brezowsky (1977) im Winter (10jährige Schätzintervalle). Besonders die Andauern der zonalen Lagen nehmen etwa ab den 70er Jahren sprunghaft zu. Da diese zu den häufig auftretenden Wetterlagen gehören, zeigt diese Änderung im Andauernverhalten eine wirkliche Änderung in der allgemeinen atmosphärischen Zirkulation über Europa an. Das Wet-



Jahresmittel der Temperatur Potsdam

Abbildung 4.1: Jahresmitteltemperaturen 1951–2000 der Station Potsdam. Die vertikale Linie trennt Trainings- und Referenzperiode für das Kreuzvalidierungsexperiment. In rot eingezeichnet eine Glättung mit gleitendem Median.

tergeschehen der beiden Perioden entsteht also aus verschiedenen atmosphärischen Dynamiken.

Zeigt sich, dass das Verfahren trotz der klimatischen Verschiedenheit der beiden Zeiträume befriedigende Ergebnisse produziert, gibt das Simulationen für zukünftige Zeiträume einiges Gewicht.

Insgesamt konzentriert sich die Auswertung auf drei Temperaturparameter – Tagesmaximum, -mittel und -minimum der Lufttemperatur – sowie den Niederschlag.

Für das Kreuzvalidierungsexperiment wurden 1000 Simulationen gerechnet, alle mit den selben Parametern.² Der erste Teil der Validierung wird von diesem Ensemble handeln, für das ausgewählte Statistiken für simulierte und beobachtete Daten verglichen werden. Die Statistiken beschreiben dabei "klassische" Größen wie Mittel, Varianz oder simulierte Jahresgänge, die Persistenz in den Reihen sowie Extreme. Ich möchte diesen Teil der Validierung vor allem visuell durchführen, um zunächst ein Gefühl für die Stärken und Schwächen des Verfahrens zu bekommen. Anschließend wird aus dem Ensemble eine einzelne Simulationen herausgegriffen, für die diese Statistiken detailliert dargestellt werden – dieser Teil wird im statistischen Sinne sehr viel rigoroser gestaltet sein und

² Die Rechenzeit hierfür beträgt etwa 40 Stunden auf einer 3GHz-CPU.



Abbildung 4.2: Andauern der Zirkulationsformen zonal (blau), meridional (grün) und gemischt (rot) nach Hess und Brezowsky (1977) in Gerstengarbe und Werner (2005).

die etablierten Testverfahren verwenden. Dem folgt ein kurzer aber wichtiger Abschnitt, in dem für dieses Experiment die Erfüllung der von den Simulationsreihen geforderten Eigenschaften aus Abschnitt 2.2.1 überprüft wird.

4.2 Vorbereitung des Experiments

Im Elbe-Einzugsgebiet befinden sich 63 Klimastationen (siehe Tabelle 4.1), deren Daten für das Experiment verwendet werden. Ganz grob gesagt nähern sich die Stationen mit steigender Nummer von der Mündung der Elbe her der Quelle an. Wie in Abschnitt 2.3 beschrieben, wird in der Vorbereitung mittels einer Clusteranalyse eine Einteilung der Stationen in Ähnlichkeitsklassen bezüglich klimatischer Parameter vorgenommen. Diese Parameter waren: Mittelwert der ersten Hälfte des Beobachtungszeitraum für die Tagesmitteltemperaturen und den Niederschlag, die Differenzen dieser Mittelwerte für die zweite Hälfte zu den Mittelwerten der ersten Hälfte so wie die Standardabweichungen der beiden Beobachtungsvariablen geschätzt für den gesamten Beobachtungszeitraum. Diese Parameter sollen die mittleren Verhältnisse, zeitliche Tendenzen (durch die Mittelwertdifferenzen) und Variabilität (durch die Standardabweichung) der beiden Beobachtungsvariablen erfassen. Die sich damit ergebende Klasseneinteilung und die repräsentativen Stationen jeder Klasse sind in Abbildung 4.3 dargestellt.
Nr.	Name	Lon. [°]	Lat. [°]	h [m]	Nr.	Name	Lon. [°]	Lat. [°]	h [m]
1	Neu	9.98	54.09	26	33	Halle	11.95	51.52	96
2	Jork	9.69	53.53	1	34	Weimar	11.32	50.98	264
3	Helse	9.01	53.96	2	35	Hof	11.88	50.32	567
4	Cuxhavn	8.70	53.86	5	36	Jena	11.58	50.93	155
5	Glückstadt	9.43	53.80	2	37	Bad	10.55	51.60	335
6	Boizenburg	10.69	53.40	45	38	Chemnitz	12.86	50.80	418
7	Gardelegen	11.24	52.52	47	39	Boberlug-Kirchh.	13.58	51.65	100
8	Grambek	10.69	53.57	27	40	Dresden	13.78	51.13	222
9	Hamburg/W	10.14	53.59	19	41	Fichtelberg	12.95	50.43	1213
10	Hamburg	10.00	53.63	13	42	Gera	12.13	50.89	311
11	Lüchow	11.14	52.97	17	43	Leipzig	12.42	51.32	141
12	Lüneburg	10.43	53.26	11	44	Marienberg	13.15	50.65	639
13	Magdeburg	11.58	52.12	79	45	Plauen	12.14	50.48	386
14	Marnitz	11.93	53.32	85	46	Selb	12.17	50.20	611
15	Seehausen	11.73	52.90	21	47	Torgau	13.00	51.59	80
16	Schwerin	11.38	53.65	59	48	Wittenberg	12.65	51.89	105
17	Ülzen	10.54	52.95	49	49	Cottbus	14.32	51.78	70
18	Berlin/Dahlem	13.30	52.47	51	50	Görlitz	14.95	51.17	238
19	Berlin/Tempelh.	13.40	52.46	48	51	Cheb	12.40	50.08	471
20	Berlin/Schönef.	13.53	52.38	46	52	Churanov	13.61	49.06	1126
21	Neuglobsow	13.04	53.15	62	53	Praha-Ruzyn	14.17	50.06	365
22	Neuruppin	12.82	52.90	38	54	Pribyslav	15.76	49.58	536
23	Potsdam	13.07	52.38	81	55	Primda	12.66	49.66	745
24	Waren	12.67	53.52	70	56	Doksany	14.16	50.46	159
25	Zehdenick	13.34	52.98	51	57	Ceske	14.45	48.93	424
26	Lindenberg	14.12	52.22	98	58	Praha-Kbely	14.53	50.11	287
27	Artern	11.30	51.39	164	59	Caslav	15.38	49.93	239
28	Bad	11.88	51.40	119	60	Usti	16.43	49.98	402
29	Braunlage	10.60	51.73	607	61	Karlovy	12.90	50.20	604
30	Brocken	10.61	51.80	1142	62	Pec	15.75	50.66	816
31	Eisleben	11.57	51.50	146	63	Pardubice	15.73	50.01	230
32	Erfurt	10.97	50.99	316					

Tabelle 4.1: Koordinaten (Lon. – Länge; Lat. – Breite; h – Höhe) der 63 untersuchten Klimastationen des Elbeeinzugsgebiets. Die repräsentativen Stationen sind **fett** gesetzt.

Unterschieden werden in der Klassifizierung die küstennahen (rot) von den kontinentaleren Stationen (gelb und blau) sowie die Gebirgsstationen (weiß).

An den vier repräsentativen Stationen Gera, Potsdam, Churanov und Grambeck wurden die Ausgleichsgeraden durch die Jahresmitteltemperaturen von 1976–2000 bestimmt, deren Parameter die Regressionsvorgabe aus *ii*) für die Simulationen darstellen. Sie beschreiben eine lineare Zunahme von 0.8–1.2K über diese 25 Jahre.

Die verfahrensinternen Parameter wurden anhand der in Abschnitt 3.1 ausgeführten Überlegungen gesetzt: die Blocklänge auf 12 Tage, $l_b = 12$, die Toleranzen für julianisches Datum auf 20 Tage, $\epsilon_{jd}=20$, die Toleranzen der Regressionsvorgabe auf 0.15K.



Abbildung 4.3: Die Ähnlichkeitsklassen der Klimastationen des Elbe-Einzugsgebiets farbig codiert. Schwarz umrandet die repräsentativen Stationen jeder Klasse.

4.3 Auswertungen des Ensembles

Für die Auswertung eines Ensembles von Simulationen bietet sich die Darstellung der Statistiken in Boxplots an, in denen für jede Station grob die empirische Verteilung der untersuchten Statistik aus den 1000 Simulationen dargestellt wird. Mit roten Punkten werden die Werte der Statistik dargestellt, die aus den Beobachtungsdaten des Referenzzeitraums 1976–2000 an den jeweiligen Stationen bestimmt wurden. Bei einigen Abbildungen habe ich auch mit blauen Punkten die Schätzungen aus den Trainingsdaten 1951–1975 eingezeichnet, wenn etwaige Diskrepanzen zwischen Simulation und Beobachtung dadurch verständlich werden.

Die untersuchten 63 Stationen des Elbeeinzugsgebiets werden in den Boxplots von 1 bis 63 durchnummeriert, entsprechend Tabelle 4.1.

Werden Statistiken für Sommer und Winter getrennt ausgewertet, so umfasst der Sommer dabei die Monate Juni, Juli und August. Der Winter besteht aus Dezember, Januar, Februar.

4.3.1 Klassische Statistiken

Langjähriges Mittel der Temperatur. Abbildung 4.4 zeigt in den Boxen die Spannbreite der 1000 Simulationen für die über die komplette Referenzperiode gemittelten Tagesmittel der Lufttemperatur an allen 63 Stationen. Die Spannbreiten zeigen sich allgemein als klein und liegen sehr nahe an den beobachteten Mitteln, die Abweichungen überschreiten die 0.2K nicht. Dass die Temperaturen an den repräsentativen Stationen im Mittel gut getroffen werden, ist zu erwarten, da mit der linearen Regressionsvorgabe natürlich auch das Mittel vorgegeben wird. Auch für die restlichen Stationen ist die Übereinstimmung jedoch ausgezeichnet, was die "Repräsentierbarkeit" der Stationen durch die als repräsentativ bestimmten Stationen unterstreicht.

Langjähriges Mittel der Niederschläge. Auch wenn die Übereinstimmung zwischen simulierten und beobachteten Jahresniederschlägen im Mittel nicht ganz so gut ist wie für die Mitteltemperaturen, fällt die beobachtete mittlere Summe für die meisten Stationen in den simulierten Bereich, der wiederum im allgemeinen sehr eng ist, siehe Abbildung 4.4. Die Abweichungen überschreiten relativ betrachtet eine Größenordnung von 10% nicht. Dieses Bild bestätigt sich beim Blick auf die mittleren Sommer- und Winterniederschläge (Abbildungen 4.6 und 4.7): Auch hier fallen für die meisten Stationen die beobachteten Mittel in die durch die 1000 Simulationen aufgespannten Bereich, der vor allem für die Sommerniederschläge etwas breiter ausfällt als für die Jahressummen.

Dass diese mittleren Niederschläge insgesamt so gut getroffen werden, ist in der regionalen Klimamodellierung keineswegs selbstverständlich, siehe z. B.



Abbildung 4.4: Langjähriges Mittel der Temperatur.



Mittlerer Jahresniederschlag 1976–2000

Abbildung 4.5: Langjähriges Mittel der Jahresniederschläge.



Mittlerer Sommerniederschlag

Abbildung 4.6: Langjähriges Mittel der Sommerniederschläge.



Mittlerer Winterniederschlag

Abbildung 4.7: Langjähriges Mittel der Winterniederschläge.



Standardabweichung Tmit

Abbildung 4.8: Standardabweichung der Tagesmittel der Lufttemperatur.

Frei et al. (2003) für einen Modellvergleich für die Alpenregion, Kotlarski et al. (2005) für einen Modellvergleich für Deutschland, Abbildung 1.2 und Abschnitt 4.6.

Standardabweichung der Tagesmittel der Lufttemperatur. In Abbildung 4.8 sind die simulierten und beobachteten Standardabweichungen der Tagesmitteltemperatur an den einzelnen Stationen dargestellt. Im Vergleich zu den Mitteln fällt eine größere Spannbreite in den Simulationen auf, im Mittel werden die beobachteten Standardabweichungen jedoch sehr gut getroffen.

Standardabweichung der Jahresmittel der Lufttemperatur. Ebenfalls relativ breit streuen die Standardabweichungen der Jahresmittel der Lufttemperatur. Hinzu kommt eine systematische Unterschätzung der beobachteten Standardabweichung, siehe Abbildung 4.9. Verständlich wird diese Unterschätzung, betrachtet man die interannuellen Standardabweichungen der Trainingsperiode (in blau): diese liegen noch unter den Simulationen und deutlich tiefer als die Standardabweichungen der Referenzperiode. An dieser Stelle wird also die Begrenzung des statistischen Ansatzes deutlich: Das Verfahren generiert Simulationsreihen, deren Eigenschaften statistisch ähnlich zu denen der Beobachtungsreihen sind – unterscheidet sich eine Eigenschaft für die beiden Zeiträume gravierend, wie es hier bei der interannuellen Standardabweichung der Fall ist,



Abbildung 4.9: Standardabweichung der Jahresmittel der Lufttemperatur.

kann das Verfahren diese Verschiedenheit in manchen Fällen nur ungenügend abbilden.

4.3.2 Persistenz

Eine realistische Persistenz ist bei statistischen Verfahren häufig ein Problem, insbesondere bei solchen kombinatorischen wie dem hier vorgestellten. Dem wird hier durch die Verwendung von Blöcken anstatt einzelner Tage und durch die ganze Anlage der Neuanordnung der Blöcke begegnet. Dieser Abschnitt widmet sich der Auswertung dieses Aspekts.

Dekorrelationszeiten. Bereits in Abschnitt 3.1.1 eingeführt, stellt die Dekorrelationszeit τ_D ein einfaches Maß für die Erhaltungsneigung in einer Zeitreihe dar, das in etwa aussagt, wieviele Zeitschritte ein System benötigt, um seinen aktuellen Zustand zu vergessen. Da es im wesentlichen die Autokorrelationsfunktion aufsummiert, drückt sich darin die (lineare) Erhaltungsneigung der kompletten Reihe aus, nicht nur spezielle Aspekte, wie etwa in der Andauernuntersuchung bestimmter Ereignistage. Von Interesse ist natürlich die Erhaltungsneigung jenseits der Trends und Jahresgänge in den Reihen. Diese Anteile werden mit dem STL-Verfahren aus Anhang D eliminiert und die Dekorrelationszeit aus dem übrig gebliebenen Rauschanteil geschätzt (vgl. Gleichung (3.1)).



Dekorrelationszeit Temperatur

Abbildung 4.10: Dekorrelationszeiten der trend- und jahresgangbereinigten Tagesmitteltemperaturen.

In Abbildung 4.10 ist τ_D dargestellt für das Tagesmittel der Lufttemperatur, in Abbildung 4.11 für die täglichen Niederschläge. Für beide Beobachtungsvariablen streuen die Dekorrelationszeiten zwischen den einzelnen Simulationen stark. Die aus den beobachteten Reihen der Referenzperiode geschätzten τ_D liegen bei der Temperatur für die meisten Stationen zwischen ersten und dritten Quartil der Dekorrelationszeiten aus den Simulationen, wobei an vielen Stationen tendenziell niedrigere Dekorrelationszeiten simuliert werden als beobachtet wurden. Dies ist mit dem bereits in Abschnitt 3.1.1 angesprochenen Effekt zu erklären, dass eine Umsortierung der Elemente einer Zeitreihe im allgemeinen zu einer Verringerung der Erhaltungsneigung führt. Für die Temperatur erweist sich dieser Effekt jedoch als nicht allzu gravierend, die Unterschätzungen bleiben im Mittel in einer Größenordnung von 5 – 10%.

Auch beim Niederschlag werden die beobachteten τ_D in der Regel unterschätzt, wobei diesmal die Beobachtungen klar außerhalb der sich zwischen ersten und dritten Quartil befindenen Simulationen liegen. Dies erklärt sich, neben dem eben angesprochenen Umsortierungs-Effekt, außerdem aus der Konstruktion des Verfahrens, die darauf abzielt, Simulationsreihen mit Persistenzen zu erzeugen, die mit den Persistenzen der Trainingsreihen statistisch vergleichbar sind (Eigenschaft *iv*)). Diese Persistenzen sind in blau der Grafik beigefügt: Die Simulationen entsprechen ihnen sehr viel besser als denen der Referenz-



Abbildung 4.11: Dekorrelationszeiten der trend- und jahresgangbereinigten täglichen Niederschläge.

periode. Die beobachtete Persistenz hat offensichtlich von der Beobachtungszur Simulationsperiode deutlich zugenommen, was zu der in Abbildung 4.2 gezeigten deutlichen Zunahme der Andauern wichtiger Großwetterlagen von der Trainings- zur Referenzperiode passt.

Beide Größen zeigen in Küstennähe (entsprechend den niedrigen Stationsnummern) höhere Persistenz als weiter im Landesinnern, was auf den dämpfenden Einfluss des Meeres zurückzuführen ist. Dieser Effekt lässt sich allgemein sowohl in Beobachtungen als auch in GCM-Simulationen feststellen, siehe z. B. Fraedrich und Blender (2003) und Blender und Fraedrich (2003).

Andauern von Ereignistagen. Neben dem allgemeinen Maß der Dekorrelationszeit sind Auswertungen von Andauern bestimmter Ereignistage ein intuitiver Zugang zur Persistenz in einer klimatologischen Zeitreihe. Typische Ereignistage sind Sommertage ($T_{\text{max}} \ge 25.0^{\circ}$ C) oder trockene Tage (RR = 0mm). Die mittleren Andauern dieser beiden Ereignistage sind in Abbildung 4.12 und Abbildung 4.13 dargestellt. Die in der Referenzperiode beobachteten Andauern finden sich zwar für praktisch alle Stationen innerhalb der simulierten Bereiche, werden jedoch insgesamt leicht unterschätzt. Die Übereinstimmung zwischen den beobachteten Andauern der Trainingsperiode (blau) und den Simulationen ist sehr viel besser. Für die Trockenperioden lässt sich an den küstennahen Stationen eine



Abbildung 4.12: Mittlere Andauern von Sommertagen.

systematische Unterschätzung feststellen. Diese erklärt sich, betrachtet man die Trockendauern der Trainingsreihen (in blau): Diese Andauern unterscheiden sich zwischen den beiden Zeiträumen beträchtlich und das Verfahren konnte in diesem Fall die Instationarität nicht abbilden (vergleiche hierzu die Auswertung der Dekorrelationszeiten).

4.3.3 Extreme

Extreme Witterungsereignisse sind gegenwärtig intensiv beforschte Erscheinungen. Dahinter steht die Erfahrung, dass in vielen Regionen zwar mit den allmählichen Klimaänderungen der letzten Jahrzehnte, etwa der Erwärmung in der Größenordnung von 1K während der letzten 50 Jahre im Elbgebiet, noch keine Einbußen an Lebensqualität einhergehen. Extreme Witterungsereignisse wie die außergewöhnlichen Starkniederschläge im Sommer 2002, die zu dem verheerenden Elbhochwasser führten, wirken hingegen instantan und stark. Die Frage, ob solche Ereignisse im Zuge der zu erwartenden Klimaentwicklungen zunehmen werden, ist daher von großer gesellschaftlicher Relevanz. Sollen die mit meinem Verfahren erhaltenen Abschätzungen verwertbar sein, muss seine Fähigkeit, solche Ereignisse abzubilden, nachgewiesen werden.

Eine gewisse Einschränkung dieser Fähigkeit ergibt sich dabei aus dem Ansatz des Verfahrens: Indem die simulierten Reihen aus Abschnitten der Beobach-



Abbildung 4.13: Mittlere Andauern von trockenen Tagen.

tungsreihen zusammengesetzt werden, kann der Wertebereich der einzelnen Beobachtungsvariablen nicht überschritten werden. An einer Station können daher beispielsweise nicht stärkere tägliche Niederschläge simuliert werden als in den Beobachtungsdaten dieser Station bereits vorhanden sind. Hingegen können sich extreme Andauerereignisse, etwa Hitzewellen oder Dürren durchaus in den Simulationen von den beobachteten Verhältnissen unterscheiden. Auch alle Extreme von aggregierten Größen wie monatlichen Niederschlägen können den Bereich der beobachteten Daten verlassen.

Speziell für das Elbe-Gebiet ist dabei zu bedenken, dass in den Beobachtungsdaten zwei extreme Extremereignisse dokumentiert sind: die bereits erwähnten Starkniederschläge von 2002 und der Hitzesommer 2003. Dem Verfahren stehen damit Daten wirklich extremer Verhältnisse zur Verfügung, aus denen es sich bedienen kann.

90%-Quantile des Niederschlags. Interessiert man sich für Starkniederschlags-Ereignisse und ihre Wiedergabe durch das Verfahren, so empfiehlt sich ein Blick auf den rechten Rand der Verteilung der täglichen Niederschlagsmengen. Abbildung 4.14 zeigt das 90%-Quantil der empirischen Verteilung der täglichen Niederschläge (ohne niederschlagsfreie Tage), d. h. die tägliche Niederschlagsmenge, jenseits der sich die stärksten 10% der Niederschlagsmengen befinden. Für viele Stationen fallen die in der Referenzperiode



Abbildung 4.14: q90-Quantile der täglichen Niederschläge.

beobachteten Quantile in den simulierten Bereich, die Abweichungen in den übrigen Fällen sind jedoch ebenfalls klein und betragen nicht mehr als wenige mm.

Anzahl von Starkniederschlagstagen. Eine anderer Zugang zu Starkniederschlagsereignissen besteht darin, einen Schwellenwert für den Niederschlag festzulegen und die Tage zu zählen, an denen der Schwellenwert überschritten wird. Ein üblicher Schwellenwert sind 10mm (z. B. im STARDEX-Projekt³, das sich mit der Simulation von extremen Witterungsereignissen beschäftigt). Abbildung 4.15 wertet die Anzahl der Tage mit RR > 10mm pro Jahr aus. Die in der Referenzperiode beobachteten Anzahlen an den meisten Stationen fallen in den simulierten Bereich, die Übereinstimmung ist sehr gut.

Anzahl von Sommer- und Frosttagen. Ein Analogon zu den Starkniederschlagstagen sind für die Temperatur Sommertage, deren Anzahl vom Verfahren sehr befriedigend wiedergegeben wird (siehe Abbildung 4.16). In die gleiche Kategorie fallen die Frosttage, also die Tage, bei denen die Temperatur unter 0°C fällt (Abbildung 4.17). Die Übereinstimmung zwischen Simulationen und Beobachtung ist wie für die Sommertage befriedigend.

³ Siehe http://www.cru.uea.ac.uk/projects/stardex/.



Abbildung 4.15: Anzahl der Tage mit Starkniederschlägen (RR > 10mm) 1976–2000.



Anzahl der Sommertage

Abbildung 4.16: Anzahl der Sommertage ($T_{\text{max}} \ge 25^{\circ}$ C) 1976–2000.



Anzahl der Frosttage

Abbildung 4.17: Anzahl der Frosttage ($T_{\min} \le 0^{\circ}$ C) 1976–2000.

4.3.4 Die simulierten Jahresgänge

Aus Platzgründen werte ich die simulierten Jahresgänge des kompletten Ensembles nur für zwei der repräsentativen Stationen aus, nämlich Potsdam und Grambeck, die für das kontinentale bzw. das küstennahe Klima des Elbgebiets stehen. Dargestellt sind jeweils die Monatsmittel für die Temperatur (Abbildung 4.18) bzw. die Monatssumme für den Niederschlag (Abbildung 4.19), in den Boxen die Werte, die sich aus den 1000 Simulationen ergeben haben, rot wie gewohnt die Schätzungen aus den Referenzdaten. Der Jahresgang der Temperatur wird sehr gut getroffen und die Spannbreite der Abweichungen beträgt für die meisten Simulationen weniger als 1K. Für den Niederschlag ist die Übereinstimmung nicht so gut: Zwar liegen die beobachteten Monatsniederschläge alle im Bereich der simulierten Monatssummen, sie werden jedoch für einzelne Monate um bis zu etwa 10mm bezüglich des Medians verfehlt, was einem relativen Fehler von ca. 20% entspricht. Die Streubreite der simulierten Niederschläge ist dabei sehr viel größer als für die Monatsmittel bei der Temperatur. Die Abweichungen erklären sich, betrachtet man die monatlichen Niederschläge, wie sie in der Trainingsperiode beobachtet wurden (in blau dargestellt). Sie befinden sich stets in der Nähe der Mediane der simulierten Summen. Dies ist ein Ergebnis sehr im Sinne des Ansatzes des Verfahrens, das Geschehen innerhalb der Jahre der Beobachtungsperiode als Annäherung an die Jahre des Simulationszeitraums zu



Abbildung 4.18: Monatsmittel der Temperatur für Grambeck und Potsdam.

verwenden. Es zeigt allerdings auch die Grenzen dieses Ansatzes auf.

4.3.5 Zusammenfassung

Die bisherige visuelle Auswertung des Ensembles ergab, dass das Verfahren wesentliche Statistiken der Referenzreihen befriedigend reproduziert.

Systematische, jedoch akzeptable, Abweichungen gibt es bei den interannuellen Standardabweichungen der Temperatur, bei den Persistenzen des Niederschlags sowie beim Jahresgang des Niederschlags, die sich letztlich auf die dem Verfahren implizite Stationaritätsannahme zurückführen lassen – wie alle statistischen Ansätze ist auch dieser nur begrenzt in der Lage, von den Trainingszeitreihen deutlich verschiedene Simulationsreihen zu erzeugen.

Berücksichtigt man die klimatologische Verschiedenheit der Trainings- und Simulationsperiode (siehe Abschnitt 4.1) in diesem Experiment, so sind die Ergebnisse insgesamt sehr zufriedenstellend, insbesondere für den Niederschlag, dessen Modellierung generell problematisch ist. Dies gilt umso mehr, als meinem Verfahren für den Niederschlag keinerlei Vorgaben gemacht werden. Da die Übereinstimmung der Statistiken aus den Simulationen und Beobachtungen an allen Stationen gleichmäßig gut ist, ist das Verfahren offensichtlich in der Lage, auch die räumliche Charakteristik des Klimas zu reproduzieren. Die gute Wiedergabe der für Extremereignisse relevanten Statistiken empfiehlt das Verfahren



Abbildung 4.19: Monatssummen des Niederschlags für Grambeck und Potsdam.

für Impact-Analysen dieser Ereignisse.

4.4 Auswertung einer ausgewählten Simulation

Um aus diesem Ensemble eine Simulation für eine detailliertere Untersuchung auszuwählen, muss zunächst entschieden werden, anhand welchen Kriteriums die Auswahl stattfinden soll.

Auswahl einer Simulation. Da der Wasserhaushalt im Elbe-Gebiet Untersuchungsgegenstand des GLOWA-Elbe-Projekts ist, suche ich eine Charakterisierung der 1000 Simulationen, die eine Einordnung auf einer Skala von trocken bis feucht ermöglicht. Dafür bietet sich die klimatische Wasserbilanz (KWB) an, die Differenz zwischen Niederschlag und potentieller Evapotranspiration. Die potentielle Evapotranspiration pE ist die Wassermenge, die die Atmosphäre theoretisch dem Boden entziehen könnte (ein komplett versiegelter Parkplatz wird da nicht so freigiebig sein). Die KWB ist somit ein Maß für die Wassermenge, die der Boden aus der Atmosphäre erhalten kann oder auch potentiell an sie abgeben muss. Die potentielle Evapotranspiration pE berechnet sich dabei mit der empirischen Formel nach Turc-Ivanov in DVWK (1996) zu

$$pE = \begin{cases} 0.0031\Omega(c) (Rad(c) + 209.4) \frac{t(c)}{t(c) + 15}, & t(c) \ge 5^{\circ}C \\ 0.000036 (25 + t(c))^{2} (100 - relH), & t(c) < 5^{\circ}C, \end{cases}$$
(4.1)

 $c \leftarrow C$ bezeichnet dabei wieder das Kalenderdatum, Rad(c) die Globalstrahlung dieses Datums in $[J/(\text{cm}^2 d)]$, relH die relative Feuchte in Prozent und t(c) die Tagesmitteltemperatur in [°C]. $\Omega(c)$ ist ein monatsabhängiger Korrekturfaktor, der sich zu

bestimmt.

Für jede der vier repräsentativen Station habe ich den Mittelwert der Wasserbilanz sowie die Differenz zwischen dem Mittelwert der 2. Hälfte des Simulationszeitraums und dem der ersten Hälfte bestimmt. Diese beiden Parameter beschreiben damit das mittlere Niveau und ganz grob die zeitliche Entwicklung dieser Größe. Insgesamt wird eine einzelne Simulation S_r also durch 8 Parameter beschrieben, zusammengefasst im 8-Tupel p_r , $r \leftarrow [1..1000]$. Bestimmt man den Schwerpunkt cm aller Simulationen in diesem Parameterraum,

$$\mathrm{cm} = \frac{1}{1000} \sum_{r=1}^{1000} p_r,$$

so kann die Simulation bestimmt werden, die ihm am nächsten liegt:

$$r' = \underset{r \leftarrow [1..1000]}{\operatorname{argmin}} \operatorname{dist}^{e}(p_r, \operatorname{cm})$$

r' repräsentiert in diesem Sinne die Mitte der von den Simulationen aufgespannten Entwicklungen bezüglich des Wasserhaushalts und wird im folgenden detailliert ausgewertet.

Darstellung der ausgewerteten Statistiken. Um die Differenzen zwischen Simulation und Beobachtung bezüglich einer Statistik darzustellen, verwende ich entweder Karten oder Punktdarstellungen ähnlich den Boxplots des vorangegangenen Abschnitts. Bei den Karten wird stets der Wert der Statistik für die Beobachtungsdaten der Referenzperiode dargestellt sowie die Abweichung in der Simulation. Je nach Statistik sind die Differenzen absolut oder relativ in Prozent dargestellt, siehe z. B. Abbildung 4.22 oder 4.32.

Die Punktdarstellungen (z. B. Abbildung 4.27) bestehen aus zwei Abbildungen: in der ersten ist für jede Station der Wert der jeweiligen Statistik bezüglich der simulierten abzüglich der in der Referenzperiode beobachteten Reihe eingetragen. Die zweite, untere Abbildung enthält die selbe Abweichung, allerdings relativ auf den Wert der Statistik für die Referenzreihe bezogen. In beiden Abbildungen ist als waagrechte Linie die mittlere Abweichung eingezogen. Zusätzlich fassen Tabellen die Ergebnisse gebräuchlicher Tests der Verschiedenheit von Beobachtung und Simulation zusammen (die Tests sind in Anhang B beschrieben). Diese Tabellen sind wie folgt aufgebaut (siehe beispielsweise Tabelle 4.2): Die ersten drei Spalten fassen die Vergleichstests an allen Stationen zusammen. Sie enthalten die über alle Stationen gemittelte Abweichung der Statistik, den Bereich der Abweichungen über alle Stationen, sowie den Bereich der p-Werte, die sich beim stationsweisen Vergleich ergeben. Der p-Wert gibt dabei die Wahrscheinlichkeit an, dass bezüglich der jeweiligen Statistik Simulationsreihe und beobachtete Reihe nicht unterscheidbar sind. Je geringer er ausfällt, desto wahrscheinlicher die Verschiedenheit der beiden Reihen bezüglich der jeweiligen Statistik.

Die nächsten beiden Spalten enthalten die Zusammenfassung des Vergleichs der über alle Stationen gemittelten Simulations- und Beobachtungsreihen. Auf diese Weise lässt sich das für das Gebiet mittlere Klima mit dem simulierten mittleren Klima vergleichen.

Die letzten drei Spalten enthalten die Anzahl von Stationen, für die für drei verschiedene Irrtumswahrscheinlichkeiten die Verschiedenheit von simulierter und beobachteter Reihe bezüglich der betrachteten Statistik gesichert werden kann.

Die Einheiten der Abweichungen sind in den Erläuterungen der Tabellen in eckigen Klammern angegeben.

Die Verwendung der Tests wird im Text zu den untersuchten Statistiken motiviert.

Als ersten Eindruck von der ausgewählten Simulation zeigt Abbildung 4.20 die simulierten und beobachteten Jahresmittel der Temperatur und die Jahresniederschläge 1976–2000, jeweils über alle 63 Stationen gemittelt, sowie die Regressionsgeraden für die Temperatur und die Mittelwerte für den Niederschlag (die lineare Regressionsanalyse ergibt mit dem Test aus B.2 im Gegensatz zur Temperatur keinen statistisch gesicherten Trend, daher zeige ich nur die Mittelwerte, deren Differenz mit dem Test in B.5 allerdings auch nicht gesichert ist). Diese liegen bei der Temperatur quasi exakt aufeinander, was wegen der Regressionsvorgabe für die Temperatur zu erwarten ist. Die beiden Mittelwerte des Niederschlags unterscheiden sich um knapp 20mm bzw. rund 2% (siehe Tabelle 4.5).

4.4.1 Auswertungen für die Temperatur

Klassische Statistiken der Temperatur

Mittelwerte (1. Moment) der Temperatur. In Karte 4.22 sind die beobachteten Mittelwerte der Temperatur und die Abweichungen in der Simulation dargestellt. Die Abweichungen bewegen sich im Bereich ± 0.2 K und lassen keine räum-



Abbildung 4.20: Jahresmitteltemperaturen und Jahresniederschläge 1976–2000, beobachtet (rot) und simuliert (schwarz).

liche Systematik erkennen, sie scheinen zufällig zu sein. Möchte man testen, ob diese Abweichungen statistisch unterscheidbar sind, bietet sich der *t*-Test aus Anhang B.1 für den Vergleich der Mittelwerte zweier Stichproben an. Anwendbar ist er für Stichproben aus normalverteilten Grundgesamtheiten, deren Elemente voneinander unabhängig sind.

Streng genommen ist für die Reihe der Tagesmitteltemperaturen die Normalverteiltheit nicht gegeben, das Histogramm in Abbildung 4.21 zeigt eine leichte Bimodalität an, die sich dem Jahresgang der Temperatur schuldet. Auch der qq-Plot, der die geschätzten Quantile gegen die Quantile der theoretischen Normalverteiltheit aufträgt, zeigt leichte Abweichungen um die Gerade, die eine tatsächlich normalverteilte Stichprobe ergäbe. Da diese Abweichungen jedoch nicht zu groß sind und der *t*-Test moderate Abweichungen von der Normalverteiltheit verträgt (von Storch und Zwiers, 1999), nehme ich diese Voraussetzung als erfüllt an.

Unverträglich sind hingegen Verletzungen der Unabhängigkeit der Stichprobenelemente (von Storch und Zwiers, 1999; Gerstengarbe, 2002), die für geophysikalische Zeitreihen allgemein nicht und für Temperaturreihen besonders wenig gegeben ist. Dieser reiheninternen Abhängigkeit muss daher durch die Bestimmung der effektiven Freiheitsgrade Rechnung getragen werden, siehe Abschnitt B.1.



Abbildung 4.21: Histogramm und qq-Plot für die Reihe der Tagesmittel der Temperatur Potsdam 1951–2003: Die Daten sind nicht normalverteilt, die Abweichungen sind jedoch nicht zu stark.

Für Tabelle 4.2 wurden beobachtete und simulierte Mittelwerte stationsweise mit dem *t*-Test verglichen – selbst für die Irrtumswahrscheinlichkeit von 10% lassen sich für keine der Stationen Abweichungen statistisch nachweisen, so dass die Mittelwerte als reproduziert angesehen werden können.

Nicht ganz so gut ist die Übereinstimmung, betrachtet man die mittleren Temperaturen im Sommer und im Winter (Abbildungen 4.23 und 4.24). Für den Sommer finden sich gleichermaßen Über- und Unterschätzungen in der Größenordnung bis zu 0.35K, die allerdings nur für wenige Stationen zu einer statistisch gesicherten Verschiedenheit von Simulation zu Beobachtung führen. Die Winter zeigen eine systematische Überschätzung von im Mittel 0.41K, die sogar auf dem 1%-Niveau bei 38 Stationen einer gesicherten Verschiedenheit entsprechen (siehe Tabelle 4.2).

Streuung (2. Moment) der Temperatur. Karte 4.25 zeigt den Vergleich von simulierter und beobachteter Standardabweichung für die Tagesmitteltemperatur. Er ergibt für einzelne Stationen Abweichungen bis zu 0.2K, die Abweichungen der meisten Stationen liegen unter ± 0.1 K.

Interessant ist außerdem die interannuelle Variabilität, gemessen mit der Standardabweichung der Jahresmittel der Temperatur. Sie ist in Karte 4.26 dar-



Abbildung 4.22: Gemittelte Tagesmittel der Lufttemperatur 1976–2000. Links die Mittel aus den Beobachtungen, rechts die absoluten Abweichungen der simulierten Temperaturmittel von den beobachteten.



Abbildung 4.23: Über die Sommer gemittelte Tagesmittel der Lufttemperatur 1976–2000. Links die Mittel aus den Beobachtungen, rechts die absoluten Abweichungen der simulierten Temperaturmittel von den beobachteten.



Abbildung 4.24: Über die Winter gemittelte Tagesmittel der Lufttemperatur 1976–2000. Links die Mittel aus den Beobachtungen, rechts die absoluten Abweichungen der simulierten Temperaturmittel von den beobachteten.



Abbildung 4.25: Standardabweichung der Tagesmittel der Lufttemperatur 1976–2000. Links die Standardabweichungen aus den Beobachtungen, rechts die absoluten Abweichungen der simulierten Standardabweichungen von den beobachteten.



Abbildung 4.26: Standardabweichung der Jahresmittel der Lufttemperatur 1976–2000. Links die Mittel aus den Beobachtungen, rechts die absoluten Abweichungen der simulierten interannuellen Standardabweichungen von den beobachteten.

gestellt, die eine leichte systematische Unterschätzung in der Simulation von maximal 0.25K (im Mittel rund 0.1K) ausweist. Siehe hierzu auch die Erläuterungen zu Abbildung 4.9.

Will man die durch die Standardabweichung gemessene Streubreite in den Temperaturreihen statistisch vergleichen, steht man vor einem Problem: Alle üblichen Tests wie der *F*- oder der χ^2 -Test für die Beurteilung der an Stichproben geschätzten Streuungen sind nur für normalverteilte Stichproben mit unabhängigen Elementen brauchbar. Im Gegensatz zum t-Test, der mit moderaten Verletzungen der Normalverteiltheit noch umgehen kann, reagieren diese Tests sehr sensibel (von Storch und Zwiers, 1999; Taubenheim, 1969). Parameterfreie Tests wie der Mann-Whitney-Test oder andere rangbasierte Tests, die dieses Problem nicht haben, bieten hingegen keine Möglichkeit, die reiheninterne Abhängigkeit zu berücksichtigen. Die Standardabweichung oder Varianz einer Zeitreihe bieten damit kein statistisch bewertbares Maß für deren zweites Moment. Dieses lässt sich besser durch die Autokorrelationsfunktion oder das Spektrum der Reihe erfassen. Für einen statistischen Vergleich ist hierbei das Spektrum der Autokorrelationsfunktion vorzuziehen, da es sich in eine Funktion transformieren lässt, die die Eigenschaften einer kummulativen Verteilungsfunktion aufweist. Auf diese Weise kann die Kolmogorov-Smirnov-Statistik zur Bewertung der Verschiedenheit zweier Spektren verwendet werden (siehe Abschnitt B.4).

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		♯ Stationen <i>p</i> <		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
MT (t)	-0.02	-0.18 / 0.17	0.10 / 0.40	-0.02	0.39	0	0	0
ST (t)	0.05	-0.23 / 0.35	0.00 / 0.40	0.05	0.34	18	8	2
WT (t)	0.41	0.11 / 0.63	0.00 / 0.29	0.41	0.00	57	54	38
SpT (KS)	-0.04	-0.20 / 0.09	1.00 / 1.00	-0.06	1.00	0	0	0
SpTR (KS)	-0.10	-0.24 / 0.01	1.00 / 1.00	-0.13	1.00	0	0	0
VT (KS)	0.02	0.01 / 0.03	0.00 / 0.54	0.02	0.06	35	26	3

Tabelle 4.2: Tests klassischer Statistiken der Temperatur. Beschreibung der Tabelle am

 Anfang von Abschnitt 4.4.

MT (t): Mittelwert der Temperatur, t-Test, [K].

ST (t): Sommermittel der Temperatur, t-Test, [K].

WT (t): Wintermittel der Temperatur, t-Test, [K].

SpT (K): Standardabweichung und Spektrum der Tagesmitteltemperatur (2. Moment), KS-Test, [K].

SpTR (K): Standardabweichung und Spektrum des Rauschens der Tagesmitteltemperatur (2. Moment), KS-Test, [K].

VT (KS): Verteilung der Tagesmitteltemperaturen, KS-Test, [].

In den Karten und Tabellen sind für die Abweichungen trotzdem die Differenzen in den Standardabweichungen dargestellt, da sie ein sehr viel intuitiveres Maß für die Verschiedenheit der Streuungen darstellt als die KS-Statistik. Die *p*-Werte zu den Tests hingegen beziehen sich auf den Vergleich der Spektren.

Zusätzlich zur Streuung bzw. zum Spektrum der kompletten Reihe wurde auch das Spektrum lediglich des Rauschanteils der beiden Reihen verglichen, um auf eine eventuelle Verschiedenheit jenseits der Trend- und Jahresgangkomponenten zu prüfen. Für die Zerlegung wurde wieder das in Abschnitt D beschriebene STL-Verfahren benutzt. An allen Stationen konnten sich weder für das Spektrum der kompletten Reihe noch für das der Rauschkomponente Unterschiede zwischen Simulation und Beobachtung sichern lassen.

Verteilung. Als letzte Statistik ist in Tabelle 4.2 die Kolmogorov-Smirnov-Statistik aufgeführt (siehe Abschnitt B.3), die die Verschiedenheit zweier Verteilungen bemisst. Der Test mit dieser Statistik ergab, dass sich die Verteilungen beobachteter und simulierter Temperaturen an 26 von den 63 Stationen auf dem 5%-Niveau unterscheiden, allerdings lässt sich nur für drei Stationen ein Unterschied auf dem 1%-Niveau feststellen.

Persistenz der Temperatur

Abbildung 4.27 zeigt die Differenzen der simulierten zu den beobachteten Dekorrelationszeiten für die Tagesmittel der Lufttemperatur. Im Mittel werden die Dekorrelationszeiten um einen knappen Tag unterschätzt, wobei die Abweichung für die küstennahen Stationen etwas stärker ausgeprägt erscheint. Diese



Abbildung 4.27: Differenzen in den Dekorrelationszeiten zwischen simulierten und beobachteten Temperaturreihen 1976–2000.

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
VSP (KS)	-0.15	-0.56 / 0.20	0.00 / 1.00	-0.33	0.09	5	2	1

Tabelle 4.3: Tests für die Persistenz der Temperatur. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

VSP (KS): Verteilung der Andauern von Sommertagen (Sommerperioden), KS-Test, [Ta-ge].

Unterschätzung kündigte sich bereits bei den Untersuchungen zur Blocklänge in Abschnitt 3.1.1 an: Durch das Neuaneinanderfügen einzelner Abschnitte der Beobachtungsreihen ist der Verlust an Persistenz praktisch unausweichlich.

Ein nicht ganz so eindeutiges Bild zeigen die Andauernverteilungen der Sommertage: an einzelnen Stationen werden die Andauern überschätzt, für die meisten Stationen jedoch unterschätzt, im Gesamtmittel um 0.15 Tage pro Jahr bei relativen Abweichungen von in der Regel bis zu 20%. Die Verteilung der Andauern ist allerdings nur für eine Station auf dem 1%-Niveau als unterschiedlich gesichert (siehe Abbildung 4.28 und Tabelle 4.3).



Abbildung 4.28: Differenzen in den mittleren Andauern der Sommertage ($T_{max} \ge 25^{\circ}$ C) zwischen simulierten und beobachteten Reihen 1976–2000.

Extreme der Temperatur

In Karte 4.29 ist das 90%-Quantil der Tagesmaxima der Lufttemperatur dargestellt. Insgesamt werden diese Quantile leicht überschätzt, an einzelnen Stationen um bis zu 1K. Die Abweichung ist damit gering. In ihrer räumlichen Verteilung lässt sich keine Systematik erkennen, was für räumlich zufällige Abweichungen spricht.

Die Abbildunden 4.30 und 4.31 zeigen die Differenzen von simulierten zu beobachteten Anzahlen von Sommer- bzw. Frosttagen. Die Sommertage werden um etwa 1.5 Tage pro Jahr überschätzt, für die Frosttage ist das Mittel nahezu ausgeglichen mit Abweichungen zwischen rund –8 und 11 Tagen (siehe Tabelle 4.4). Die Verteilungen der jahresweisen Anzahlen dieser Ereignistage unterscheiden sich auf dem 1%-Niveau bei keiner Station gesichert.

4.4.2 Auswertung für den Niederschlag

Mittlere Niederschläge (1. Moment). Die Karten 4.32, 4.33 und 4.34 zeigen die mittleren Niederschlagssummen über die Jahre, Sommer und Winter. Die rechten Karten zeigen jeweils die relativen Abweichungen in Prozent. Ähnlich wie bei der Temperatur stimmen die Jahressummen gut überein, die Abweichungen bewegen sich im Bereich $\pm 10\%$. Etwas größere Abweichungen finden sich für die



Abbildung 4.29: 90%-Quantil der Tagesmaxima der Lufttemperatur 1976–2000. Links für die Beobachtungen, rechts die absoluten Abweichungen der simulierten Quantile von den beobachteten.



Abbildung 4.30: Differenzen in den jahresweisen Anzahlen von Frosttagen ($T_{\min} \le 0^{\circ}$ C) zwischen simulierten und beobachteten Temperaturreihen.



Abbildung 4.31: Differenzen in den jahresweisen Anzahlen der Sommertage ($T_{\text{max}} \ge 25^{\circ}$ C) zwischen simulierten und beobachteten Temperaturreihen.

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	<i>p</i> -Wert	0.1	0.05	0.01
VS (KS)	1.35	-2.52 / 5.60	0.08 / 1.00	1.48	0.91	1	0	0
VF (KS)	-0.46	-7.92 / 11.32	0.02 / 1.00	-2.24	0.91	2	1	0

Tabelle 4.4: Tests für Extreme der Temperatur. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

VS (KS): Verteilung der jahresweisen Anzahlen der Sommertage ($T_{max} \ge 25^{\circ}$ C), KS-Test, [Tage].

VF (KS): Verteilung der jahresweisen Anzahlen von Frosttagen ($T_{\min} \leq 0^{\circ}$ C), KS-Test, [Tage].

Sommer- und Winterniederschlägen, bei denen einzelne Stationen Unterschätzungen von bis zu 20% aufweisen.

Den mittleren Niederschlag statistisch zu vergleichen stellt grundsätzlich vor die gleichen Probleme wie bei der Temperatur. Ich habe mich in diesem Fall allerdings, da die Verteilung der täglichen Niederschläge beim besten Willen nicht als normalverteilt angesehen werden kann, für den parameterfreien Mann-Whitney-Test entschieden (siehe Anhang B.5), der prüft ob zwei Stichproben aus Grundgesamtheiten mit gleichem Median stammen. Er hat seinerseits den Nachteil, dass die Persistenz der Reihe und damit die verringerte Anzahl von Freiheitsgraden nicht berücksichtigt werden kann. Da die Niederschlagsreihe jedoch nur eine relativ geringe Persistenz aufweist, nehme ich diese Voraussetzung als gegeben an. Auch wenn der Test Mediane vergleicht, sind sowohl in den Tabellen als auch in den Abbildungen Mittelwerte dargestellt, die ein gewöhnlicheres Maß für mittlere Verhältnisse geben als die Mediane.

Trotz der geringeren relativen Abweichung bei den Jahressummen verglichen mit Sommer- und Winterniederschlägen werden auch auf dem 1%-Niveau für 24 der 63 Stationen verschiedene Mittelwerte der Tagesniederschläge gesichert. Für die Sommer- und Winterniederschläge sind es nur 10 bzw. 12 Stationen. Auch wenn die Abweichungen gesichert sind, sind ihre Beträge klein: Sie überschreiten im Mittel in allen drei Fällen die 7% nicht (siehe hierzu Tabelle 4.5). Im Gebietsmittel weichen die Jahresniederschläge weniger als 2% ab (!). Diese Differenz ist, ebenso wie für die Gebietsmittel von Sommer- und Winterniederschlag, statistisch nicht zu sichern.

Insgesamt ergeben sich damit zwar zum Teil statistisch gesicherte Abweichungen, die jedoch erfreulich klein sind.

Streuung (2. Moment) der Niederschläge. Um die Wiedergabe der interannuellen Variabilität des Niederschlags in der Simulation zu prüfen, zeigt Karte 4.35 die Standardabweichungen der Jahresniederschläge. Die Abweichungen liegen für die meisten Stationen zwischen $\pm 20\%$ und scheinen räumlich zufällig verteilt zu sein.

Wie bei der Temperatur wurde auch hier für die statistische Untersuchung der Streuung anstatt direkter Maße wie der Standardabweichung das Spektrum als Bemessung für das 2. Moment der Reihen verwendet. Ebenfalls wurde sowohl dieses 2. Moment für die vollständige Reihe wie das lediglich ihrer Rauschkomponente verglichen. In beiden Fällen lassen sich für keine Station signifikante Unterschiede feststellen (siehe Tabelle 4.5). Die Standardabweichungen der täglichen Niederschläge werden im Mittel über alle Stationen sowohl für die komplette Reihe als auch für ihr Rauschen um weniger als 1% verfehlt.



Abbildung 4.32: Mittlere Jahresniederschläge 1976–2000. Links die Mittel aus den Beobachtungen, rechts die relativen Abweichungen zwischen Beobachtung und Simulation. An den Gebirgsstationen fallen allerdings Niederschläge, die mit der hier verwendeten Farbskala bis 1000mm nicht erfasst werden.



Abbildung 4.33: Mittlere Sommerniederschläge 1976–2000. Links die Mittel aus den Beobachtungen, rechts die relativen Abweichungen zwischen Beobachtung und Simulation.



Abbildung 4.34: Mittlere Winterniederschläge 1976–2000. Links die Mittel aus den Beobachtungen, rechts die relativen Abweichungen zwischen Beobachtung und Simulation.

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
MP (MW)	-1.48	-9.79 / 8.66	0.00 / 0.94	-1.93	0.17	37	31	24
SP (MW)	0.76	-18.23 / 20.82	0.00 / 0.96	0.42	0.18	22	18	10
WP (MW)	-6.12	-21.45 / 20.43	0.00 / 0.93	-5.87	0.72	29	19	12
SpP (KS)	-0.46	-15.09 / 20.79	1.00 / 1.00	-5.95	1.00	0	0	0
SpPR (KS)	-0.82	-15.73 / 20.25	1.00 / 1.00	-6.76	1.00	0	0	0
VP (KS)	0.03	0.01 / 0.09	0.00 / 0.68	0.02	0.01	51	48	42
VPf (KS)	0.05	0.02 / 0.12	0.00 / 0.54	0.02	0.16	54	52	45

Tabelle 4.5: Tests klassischer Statistiken des Niederschlags. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

MP (MW): Mittlerer Niederschlag, Mann-Whitney-Test, [%].

SP (MW): Mittlerer Sommerniederschlag, Mann-Whitney-Test, [%].

WP (MW): Mittlerer Winterniederschlag, Mann-Whitney-Test, [%].

SpP (KS): Standardabweichung und Spektrum der täglichen Niederschläge (2. Moment), KS-Test, [%].

SpPR (KS): Standardabweichung und Spektrum des Rauschens der täglichen Niederschläge (2. Moment), KS-Test, [%].

VP (KS): Verteilung der täglichen Niederschläge, KS-Test,[].

VPf (KS): Verteilung der täglichen Niederschläge (ohne Trockentage), KS-Test,[].



Abbildung 4.35: Standardabweichung der Jahresniederschläge 1976–2000. Links die Standardabweichung aus den Beobachtungen, rechts die relativen Abweichungen zwischen Beobachtung und Simulation.

Verteilung des Niederschlags. Für den Niederschlag enthält Tabelle 4.5 zum Schluss die Kolmogorov-Smirnov-Statistik, die den in gewisser Weise allgemeinsten Vergleich zweier Stichproben ermöglicht: den der gesamten Verteilung, ohne sich auf spezielle Momente o. ä. zu beschränken. Ich habe diesen Vergleich in zwei Varianten durchgeführt: einmal unter Einschluss der niederschlagsfreien Tage – hier lässt sich auch auf dem 1%-Niveau für 42 der der 63 Stationen eine Verschiedenheit von beobachteter und simulierter Niederschlagsverteilung feststellen. Zum anderen unter Ausschluss der niederschlagsfreien Tage, was das Bild allerdings auch nicht besser macht: Auf dem 1%-Niveau enthalten sogar 45 Stationen gesichert verschiedene Niederschlagsverteilungen in ihren Simulationsreihen.

Persistenz des Niederschlags

Für die Untersuchung der Persistenz in den Niederschlagsreihen werden die Dekorrelationszeiten (Abbildung 4.36) für die einzelnen Stationen sowie die mittleren Längen von Trockenperioden (Abbildung 4.37) betrachtet. Auch beim Niederschlag wird die Dekorrelationszeit an vielen Stationen unterschätzt mit Abweichungen bis zu einem halben Tag. Die mittlere Unterschätzung liegt etwa bei 1/10 Tag bzw. 2.5%.



Abbildung 4.36: Differenzen der Dekorrelationszeiten in den simulierten bzw. beobachteten Niederschlagsreihen 1976–2000.

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
VTrP (KS)	-0.05	-0.56 / 0.22	0.00 / 1.00	-0.08	1.00	10	6	3

Tabelle 4.6: Persistenz des Niederschlags. Beschreibung der Tabelle am Anfang vonAbschnitt 4.4.

VTrP (KS): Verteilung der Andauern trockener Tage, KS-Test, [Tage].

Die Abweichungen bei den mittleren Trockenperioden sind hingegen recht ausgegeglichen verteilt, im Mittel unterscheiden sie sich um 0.05 Tage. Auf dem 1%-Niveau sind lediglich drei Stationen, für die sich die Verteilungen der Periodenlängen von Simulation und Beobachtung statistisch unterscheiden lassen (siehe Tabelle 4.6).

Extreme des Niederschlags

Spätestens seit dem Elbe-Hochwasser im Sommer 2002 sind extreme Niederschlagsereignisse in dieser Region regelmäßiger Untersuchungsgegenstand. In Karte 4.38 sind die in der Referenzperiode beobachteten 90%-Quantile der täglichen Niederschläge dargestellt. Die Abweichungen zwischen Beobachtung und Simulation liegen für die meisten Stationen unter 10%. Das ist ein sehr wichtiges Ergebnis, da die Abschätzung extremer Ereignisse für Folgeuntersuchungen



Abbildung 4.37: Differenzen in den mittleren Längen der Trockenperioden in den simulierten und beobachteten Reihen.

möglichst zuverlässig sein sollte.

Ein Vergleich für Tage ohne Niederschlag findet sich in Abbildung 4.39. Auch hier werden die Anzahlen sehr befriedigend reproduziert, mit einer mittleren Abweichung von etwa 5 Tagen (entsprechend rund 3%). Dem entspricht, dass nur für 6 Stationen die Verteilungen der jährlichen Anzahlen der Trockentage sich bei einer Irrtumswahrscheinlichkeit von 1% statistisch unterscheiden lassen (siehe Tabelle 4.7).

Ein vergleichbar gutes Bild zeichnen die Starkregentage (Abbildung 4.40). Sie werden im Mittel nur um 0.2 Tage pro Jahr unterschätzt mit maximalen Abweichungen von vier Tagen bzw. 20% an einzelnen Stationen. Selbst auf dem 10%-Niveau lassen sich die Anzahlen nur an 6 Stationen unterscheiden, auf dem 1%-Niveau keine einzige mehr (siehe Tabelle 4.7).

4.4.3 Zusammenfassung

Diese Auswertung zeigen, dass die ausgewählte Simulation sowohl für Temperatur als auch für den Niederschlag eine sehr gute Approximation an die tatsächlich beobachteten Reihen darstellt. Der größte Schwachpunkt ist vielleicht die Überschätzung der mittleren Wintertemperatur um etwa 0.4K (im Mittel). Auch beim Niederschlag lassen sich bei den Gesamt-, Sommer- und Winterniederschlägen statistisch signifikante Abweichungen feststellen, die jedoch die Grö-



Abbildung 4.38: 90%-Quantil der Niederschläge 1976–2000. Links die Quantile aus den Beobachtungen, rechts die relativen Abweichungen zwischen Beobachtung und Simulation.



Abbildung 4.39: Differenz zwischen simulierter und beobachteter Anzahl der Trockentage pro Jahr.



Anzahl der Starkregentage Abweichung

Abbildung 4.40: Differenz zwischen simulierter und beobachteter Anzahl der Starkregentage pro Jahr.

	Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
VTr (KS)	4.49	-32.08 / 25.08	0.00 / 0.99	-5.16	0.08	14	8	6
VSr (KS)	-0.22	-3.96 / 3.08	0.02 / 0.99	-1.44	0.04	5	4	0

Tabelle 4.7: Extreme des Niederschlags. Beschreibung der Tabelle am Anfang vonAbschnitt 4.4.

VTr (KS): Verteilung der jahresweisen Anzahlen trockener Tage, KS-Test, [Tage].

VSr (KS): Verteilung der jahresweisen Anzahlen von Starkregentagen, KS-Test, [Tage].
ßenordnung von 10% nicht überschreiten. Die Verteilungen der täglichen Niederschläge weisen für einen beträchtlichen Anteil der Stationen signifikante Verschiedenheiten auf, für die Temperatur gilt das nur für wenige Stationen. Die Persistenzen werden eher unter- denn überschätzt, was sich zum Teil dem Ansatz des Umordnens schuldet. Sehr befriedigend wiedergegeben werden hingegen die 90%-Quantile der Verteilungen von Niederschlag und Temperatur, die für Extremereignisse von besonderer Relevanz sind. Dazu passt die gute Wiedergabe der Ereignistage wie Sommer-, Frost- und trockenen Tagen.

Die meisten untersuchten Statistiken zeigen keine räumlich systematischen Abweichungen, was den Eindruck aus Abschnitt 4.3 bestätigt: Die räumliche Struktur der untersuchten Klimatologie wird überzeugend reproduziert.

Diese statistisch rigorosere Untersuchung einer einzelnen Simulation bestätigt das befriedigende Bild, das die Auswertung des Ensembles der 1000 Simulationen hinterlassen hatte.

4.5 Untersuchung der geforderten Eigenschaften

4.5.1 Regressionsparameter

Die Eigenschaft *ii*) in Teil 2.3.3 beschreibt die Regressionsvorgabe (bestehend aus absolutem Glied und Steigung) an die Jahresmittel der Temperatur, die an den repräsentativen Stationen erfüllt werden muss. Die erlaubte Abweichung zwischen der Regressionsgerade durch die simulierten Jahresmittel der Temperatur und der Vorgabe wurde auf 0.15K gesetzt.

Für die Mittelwerte als einem dem absoluten Glied äquivalenten Maß siehe Abbildung 4.4, insbesondere die Darstellungen für die repräsentativen Stationen mit den Nummern 8, 23, 42 und 52: Die simulierten Mittel weichen nur innerhalb der gesetzten Toleranz von den tatsächlich beobachteten Mitteln ab.

Für die Steigung liefert Abbildung 4.41 die Untersuchung des linearen Trends in den Jahresmitteln der Temperatur über den Simulationszeitraum nach. Auffällig sind recht große Schwankungen von bis zu 0.5K an einigen Stationen (hierbei ist allerdings zu bedenken, dass die Trends aus lediglich 25 Punkten geschätzt werden). An den repräsentativen Stationen hingegen bewegen sich die Abweichungen innerhalb der Toleranz – die Abweichungen sind für diese Stationen noch einmal in 4.42 dargestellt.

Die Regressionsvorgabe kann damit als erfüllt angesehen werden.

4.5.2 Jahresgang

Realistische Jahresgänge sind eine der wichtigsten Eigenschaften, die den Simulationsreihen abverlangt werden, siehe Abschnitt 2.2.1. Dort wurden drei Bedingungen für realistische Jahresgänge formuliert:



Abbildung 4.41: Trends über die Simulationsperiode in den Jahresmitteln der Temperatur.



Trend Jahresmittel Tmit Abweichung

Abbildung 4.42: Abweichungen der Trends in den Jahresmitteln der Temperatur über die Simulationsperiode für die repräsentativen Stationen.



Abbildung 4.43: Jahresgang für die Temperatur an den vier repräsentativen Stationen entsprechend drei Simulationen und der (verschobenen) Trainingsreihe (in schwarz).

- 1. Die aus beobachteter und simulierter Reihe extrahierten Jahresgangssignale sollen in Phase sein.
- 2. Die Periodogramme der beiden Reihen sollen als spektrale Realisierungen des selben Prozesses aufgefasst werden können.
- 3. Die Werte des Rauschens in den Reihen, das nach Eliminierung von Jahresgang und Trend übrig bleibt, sollen als Stichproben der gleichen Grundgesamtheit aufgefasst werden können.

Die Abbildungen 4.43 und 4.44 untersuchen die erste Bedingung: Dargestellt sind für die vier repräsentativen Stationen die Jahresgangsignale der Temperatur und des Niederschlags von drei zufällig ausgewählten Simulationen, die in verschiedenen Farben übereinandergelegt wurden. Zuoberst, in schwarz eingezeichnet ist der Jahresgang der zugehörigen Trainingsreihe, deren Beginn vom 1.1.1951 auf den 1.1.1976 verschoben wurde, so dass ihr Anfang mit dem Beginn der Simulationsreihe zusammenfällt. Für die Temperatur erhält man ein sehr klares Signal, dessen Phase für alle dargestellten Reihen übereinstimmt. Das Signal des Niederschlags ist weniger klar (wie üblich), auch hier erkennt man jedoch übereinstimmende Phasen.

Die zweite Bedingung wird in den Abbildungen 4.45 und 4.46 untersucht. Dargestellt sind exemplarisch die Periodogramme (siehe Abschnitt B.4) der Trai-



Abbildung 4.44: Wie Abbildung 4.43, jedoch für den Niederschlag.

ningsreihe sowie einer zufällig ausgewählten Simulationsreihe für Potsdam. Sowohl für Niederschlag (4.46) als auch für die Temperatur (4.45) ist in beiden Spektren deutlich die Spitze zu erkennen, die die Jahresperiode anzeigt. Statistisch lassen sich die Periodogramme mit der Kolmogorov-Smirnov-Statistik vergleichen (siehe Abschnitt B.4), die bemisst, ob die beiden Periodogramme als spektrale Realisierungen des selben Prozesses aufgefasst werden können. Die p-Werte für alle Stationen sind im unteren Teil der Abbildungen dargestellt – der Test kann keine Unterschiede feststellen. Dargestellt ist hier nur die Auswertung einer Simulation. Das Bild für die anderen Simulationen ist jedoch das gleiche.

In Abbildung 4.47 wird die dritte Bedingung ausgewertet. Dargestellt sind in Boxplots die *p*-Werte des KS-Tests (vergleiche Abschnitt B.3) für den Vergleich der Verteilungen des Rauschens. Unter der Nullhypothese, dass die simulierten und beobachteten Rauschen Stichproben der gleichen Grundgesamtheit darstellen, sind die *p*-Werte in [0, 1] gleichverteilt zu erwarten, was sich für die Temperatur in der Abbildung auch beobachten lässt. Für die Temperatur lassen sich also keine Unterschiede in den Verteilungen des Rauschens feststellen, beim Niederschlag treten hingegen signifikante Verschiedenheiten auf.



Abbildung 4.45: Periodogramm der Temperatur für Potsdam in der Trainingsreihe (links oben) und einer simulierten Reihe (rechts oben). Unten die *p*-Werte des KS-Tests für den Vergleich der beiden Periodogramme.



Abbildung 4.46: Wie Abbildung 4.45, jedoch für den Niederschlag.



Abbildung 4.47: *p*-Werte aus dem KS-Test zum Vergleich der Verteilungen der Rauschen aus den Trainings- und simulierten Reihen.

4.5.3 Persistenz

Realistische Persistenz ist ebenfalls eine wichtige Forderungen an die Simulationsreihen, spezifiziert in Abschnitt 2.2.1 unter Verwendung der Schätzwerte der Autokorrelationsfunktion. In den Tabellen 4.2 und 4.5 werden unter anderem Testergebnisse für den Vergleich der Spektren aus beobachtetem und simuliertem Rauschen für Temperatur und Niederschlag dargestellt. Da das Spektrum und die Autokorrelationsfunktion einander gleichwertige Beschreibungsmöglichkeiten der linearen Abhängigkeit innerhalb einer Zeitreihe bieten, vergleichen diese Tests implizit auch die Autokorrelationsfunktion, siehe hierzu Abschnitt B.4. Die Auswertung in dieser Tabelle bezieht sich auf eine einzelne Simulation, für die in keinem Fall irgendein Unterschied zwischen den Spektren festgestellt werden kann. Dieses Bild ist für die anderen Simulationen das gleiche. Die Forderung nach realistischer Persistenz kann also in diesem Experiment als erfüllt angesehen werden.

Bis auf die Verteilung des Rauschens der Niederschlagsreihen sind also alle für die Simulationsreihen geforderten Eigenschaften in diesem Kreuzvalidierungsexperiment erfüllt.

4.6 Vergleich mit REMO

Wie bereits im Einleitungskapitel 1 beschrieben, hat sich eine Vielzahl verschiedenster *downscaling*-Verfahren etabliert, von denen einige, insbesondere die dynamischen, eine viel höhere Komplexität als mein Ansatz aufweisen (meine Implementierung besteht aus etwa 2000 Zeilen Code, verglichen mit den mehreren 10000 Zeilen eines REMO oder CLM). Ein Vergleich mit anderen Modellen ist daher naheliegend.

Exemplarisch möchte ich einige der Auswertungen, die ich für die mit meinem Verfahren simulierten Daten vornehme, auch für Daten durchführen, die mit dem regionalen Modell REMO (Jacob, 2001) erzeugt wurden. Im Rahmen eines ähnlichen Experiments⁴ für ganz Deutschland wurden mit diesem Modell Gitterdaten für die Jahre 1950-2000 erzeugt (Jacob, 2005), die von Böhm am PIK entsprechend den Empfehlungen zur Nutzung der REMO-Daten⁵ auf die meteorologischen Stationen des Elbe-Einzugsgebiets interpoliert wurden. Sie können so direkt mit den Beobachtungen der Stationen einerseits, andererseits auch mit meinen simulierten Daten verglichen werden.

Gerechnet wurde hierbei mit einer horizontalen Auflösung von 10km, der Antrieb an den Rändern des Untersuchungsgebiets entstammte einem ECHAM5/MPI-OM-Lauf, der seinerseits für die Vergangenheit durch die gemessenen Einflussgrößen (z. B. CO₂) angetrieben wurde. Dadurch sind die Ergebnisse von REMO abhängig von den vom GCM gelieferten Antriebsdaten – eventuelle Diskrepanzen zwischen REMO-Simulation und Beobachtungen können also sowohl an Schwächen von REMO als auch an Schwächen des GCMs liegen. Andererseits, da das GCM seinerseits durch tatsächlich beobachtete Einflussgrößen angetrieben wird, sind die in diesem Experiment letztendlich von REMO erzeugten Simulationsreihen die realistischsten, die dieser Modellverbund produzieren kann, so dass ein Vergleich mit meinen Validierungs-Simulationen, die den Vorteil eines Antriebs haben, der tatsächlich im Untersuchungsgebiet beobachtet wurde, statthaft ist.

Besonderes Augenmerk verdient, wie üblich, der Niederschlag: Es ist allgemein bekannt, dass die Messung der Niederschlagsmenge mit dem Hellmann-Niederschlagsmesser den tatsächlichen Niederschlag – aufgrund von Verwehung, Verdunstung etc. – unterschätzt und damit für den Vergleich mit den REMO-Niederschlägen eine entsprechende Korrektur verlangt (Richter, 1995). Ich beschränke die Auswertungen in diesem Abschnitt allerdings auf Statistiken trockener Tage, für die diese Unterschätzung nicht relevant ist.

⁴ Das Experiment wird auf www.mpimet.mpg.de/fileadmin/staff/pfeifersusanne/ REMO_UBA/Klimaaenderungsworkshop.pdf vorgestellt.

⁵ Siehe http://www.mad.zmaw.de/fileadmin/extern/documents/REMO-UBA-Hinweise.pdf.



Abbildung 4.48: Mittel der Lufttemperatur 1976–2000. In den Boxen: meine Simulationen, rot das tatsächlich beobachtete Mittel, grün das Mittel entsprechend REMO.

4.6.1 Einige Vergleiche für die Temperatur

Der Boxplot 4.48 zeigt noch einmal das Temperaturmittel über die 25 simulierten Jahre. Die Werte von REMO, mit den grünen Punkten dargestellt, liegen systematisch zu hoch, wobei die Überschätzung mit wachsender Entfernung von der Mündung zunimmt und die Größenordnung von 1K erreicht.

Trotz der allgemeinen Überschätzung der mittleren Temperatur durch RE-MO werden die 90%-Quantile der täglichen Temperaturmaxima für viele Stationen erheblich unterschätzt (Abbildung 4.49), mit Abweichungen, die in Küstennähe besonders stark ausgeprägt sind.

Eine besonders deutliche Fehleinschätzung durch REMO zeigt Abbildung 4.50, in der die Trends in den Jahresmitteltemperaturen an den einzelnen Stationen dargestellt sind: REMO simuliert sie sämtlich als negativ mit einer Abnahme von bis zu rund 2K, während sie tatsächlich um etwa 1K zunehmen.

4.6.2 Einige Vergleiche für die trockenen Tage

In Abbildung 4.51 sind die mittleren Anzahlen der Trockentage pro Jahr dargestellt. Während sie in meinen Simulationen leicht überschätzt werden (für die eine ausgewählte Simulation um etwa 5 Tage im Mittel, siehe Abschnitt 4.4), zeigen die von REMO generierten Reihen eine deutliche Unterschätzung von rund



Abbildung 4.49: q90 der täglichen Temperaturmaxima 1976–2000. In den Boxen: meine Simulationen, rot das tatsächlich beobachtete Quantil, grün das Quantil entsprechend REMO.



Abbildung 4.50: Linearer Trend in den Jahresmitteltemperaturen 1976–2000. In den Boxen: meine Simulationen, rot der tatsächlich beobachtete Trend, grün der Trend entsprechend REMO.



Abbildung 4.51: Mittlere Anzahl der Trockentage pro Jahr 1976–2000. In den Boxen: meine Simulationen, rot der tatsächlich beobachtete mittlere Anzahl, grün die mittlere Anzahl entsprechend REMO.

30%, einem Klima mit zu häufigem Niederschlag entsprechend.

Dementsprechend werden auch die Trockenperioden durch REMO drastisch unterschätzt mit Abweichungen, die sich ebenfalls auf 30% belaufen können, siehe Abbildung 4.52.

4.6.3 Zusammenfassung

Zumindest für den betrachteten Zeitraum ist die Übereinstimmung zwischen Beobachtung und Simulation für die Simulationen aus dem hier vorgestellten statistischen Verfahren erheblich besser als bei den Ergebnissen von REMO. Während die betrachteten Temperaturstatistiken in den Simulationen aus meinem Verfahren befriedigend wiedergegeben werden, simuliert REMO insgesamt ein zu warmes Klima. Gleichzeitig zeigen die Trends in den Jahresmitteln der Temperatur nach REMO ein anderes Vorzeichen als die Beobachtungen.

Für den Niederschlag beschränke ich mich hier aufgrund des nicht unproblematischen Vergleichs von dynamisch simulierten und gemessenen Niederschlägen auf einfache Statistiken der Trockentage, die insgesamt in der REMO-Simulation bedeutend seltener sind als in den Beobachtungen. Auch für die Niederschlagsstatistiken ist die Übereinstimmung meiner statistischen Simulationen mit den Beobachtungen erheblich besser.



Abbildung 4.52: Mittlere Andauer der Trockentage 1976–2000. In den Boxen: meine Simulationen, rot die tatsächlich beobachteten mittleren Andauern, grün die mittleren Andauern entsprechend REMO.

Verallgemeinert man diesen Vergleich vorsichtig, so scheinen statistische Verfahren eher in der Lage zu sein, für Simulationsperioden, deren erwartete Klimatologie sich nicht zu gravierend von der des Beobachtungszeitraums unterscheidet, eine akkurate Approximation an die Klimatologie der Simulationsperiode zu liefern. Das macht sie zum Mittel der Wahl, wenn es um Entscheidungshilfe bei Fragen geht, die für oder durch die klimatische Entwicklung der nächsten Jahrzehnte relevant sind.

Dynamische Modelle, auch wenn der hier angestellte exemplarische Vergleich zu ihren Ungunsten ausfällt, sind hingegen theoretisch in der Lage, auch von der beobachteten völlig verschiedene Klimatologien zu simulieren (wenn brauchbare Randbedingungen verfügbar sind), für die die Statistik nicht mehr in Frage kommt. Außerdem ermöglichen sie eine Analyse der Abläufe im Witterungsgeschehen und erlauben damit, einem Verständnis der beteiligten Prozesse näher zu kommen.

Bis zur Entwicklung eines Modells, das alle Erwartungen befriedigt, scheint man also nicht umhin zu kommen, das Modellierwerkzeug anhand der gegebenen Fragestellungen zu wählen.

Kapitel 5

Die klimatische Zukunft des Elbe-Einzugsgebiets

Das Elbe-Einzugsgebiet umfasst eine Fläche von 148268 km², die in etwa die westliche Hälfte der Tschechischen Republik und große Teile Ost- und Norddeutschlands einschließt. Die Elbe selbst mündet nach 1091 Kilometern nördlich von Hamburg in die Nordsee.

Das Gebiet befindet sich in der Übergangszone zwischen dem atlantischen und dem kontinentalen Klima Europas. Abbildung 5.1 zeigt zunehmende Temperaturmittel 1951-2003 von der Mündung stromaufwärts, von den höheren Lagen in Harz, Erzgebirge und Thüringer Wald abgesehen. Die Niederschläge fallen in den kontinentaleren Bereichen der Tiefebene niedriger aus als in Meeresnähe, ergiebige Niederschläge finden sich lediglich in den Luv-Lagen der Mittelgebirge wie die Westseiten von Harz und Erzgebirge.

Insgesamt sind weite Teile des Einzugsgebiets niederschlagsarm, was zu einer besonderen Sensibilität dieser Region für Klimaänderungen führt – etwa die Hälfte des Elbe-Einzugsgebiets wird landwirtschaftlich genutzt, wofür ein ausreichendes Wasserangebot notwendig ist (siehe für eine detaillierte Untersuchung für den deutschen Teil des Elbe-Einzugsgebiets Wechsung et al., 2005).

Wie die klimatologische Strukturierung in den Karten zeigt, kann das Gebiet keinesfalls als klimatisch einheitlich angesprochen werden. Für Simulationen der klimatischen Zukunft ist daher eine räumliche Differenzierung erforderlich, wie sie mit der erweiterten Fassung des hier vorgestellten Verfahrens möglich ist.

5.1 Vorbereitung der Simulationen

Verwendet wurden die Stationsdaten, die bereits im Kreuzvalidierungsexperiment auftraten (siehe Kapitel 4): 63 Klimastationen im Elbe-Einzugsgebiet (sie-



Abbildung 5.1: Das beobachtete Klima im Elbe-Einzugsgebiet 1951–2003. Links die mittlere Temperatur, rechts die mittlere Jahressumme des Niederschlags, die an der Bergstationen allerdings die hier erfassten Mengen bis 1000mm deutlich übersteigen.

he Tabelle 4.1) mit täglichen Daten von 1951–2003. Auch die vier repräsentativen Stationen bleiben die selben: Grambeck, Potsdam, Gera und Churanov.

Der Simulationszeitraum umfasst die Jahre 2004–2055.

Zwei verschiedene Trendvorgaben für die Temperatur wurden als Antrieb verwendet, entsprechend einem ECHAM4-Lauf für das IS92A-Szenarium und einem ECHAM5-Lauf für das SRES-A2-Szenarium (Houghton et al., 2001; Böhm, 2006). Die Temperaturzunahmen, die die beiden Modelle simulieren, unterscheiden sich erheblich – ECHAM4 errechnet eine Erwärmung von weniger als 1K, bei ECHAM5 beträgt sie um die 2K.

Die ECHAM4-Trendvorgaben für die Jahresmittel der Temperatur betragen an den repräsentativen Stationen Grambeck: 0.68K, Potsdam: 0.82K, Gera: 0.87K und Churanov: 0.84K. Nach den Untersuchungen in Abschnitt 3.2 sind diese Trends für den Datensatz zulässig. Um die Temperaturtrends der einzelnen Gitterzellen des GCMs auf die Stationen zu übertragen, bin ich wie folgt vorgegangen: Für die Stationen Grambeck, Potsdam und Gera, die Cluster von Stationen repräsentieren, die räumlich einigermaßen zusammenhängend sind (siehe Karte 4.3) wurde die Gitterzelle bestimmt, in der die jeweilige Station liegt. Die Station bekam dann den über diese Zelle und die sie im 3x3-Quadrat umgebenden Zellen gemittelten Trend vorgegeben. Churanov vertritt die Gebirgsstationen der Region, die räumlich disparat verteilt sind, so dass es nicht sinnvoll erscheint, den Trend aus einer Gitterzelle bzw. dem umgebenden 3x3-Quadrat zu bestimmen. Daher habe ich für diese Station den Trend über lineare Regressionsbeziehungen zwischen ihrer beobachteten Temperaturreihe mit den Reihen der anderen drei Stationen verwendet, um eine den drei anderen Trendvorgaben entsprechende Vorgabe für Churanov zu erhalten.

Die absoluten Glieder wurden so gewählt, dass die vorgegebenen Regressionsgeraden auf dem Niveau anschließen, das 2003 an den vier Stationen erreicht worden war, siehe Abbildung 5.2.

Für die Trendvorgaben aus ECHAM5 wurde analog vorgegangen. An den repräsentativen Stationen ergaben sich 2.12K für Grambeck, 2.17K für Potsdam, für Gera 2.05K und für Churanov 1.91K. Diese Trendvorgaben sind streng genommen nach dem in Abschnitt 3.2 untersuchten Kriterium nicht zulässig, sie überfordern die Variabilität des Beobachtungsdatensatzes. Da im Rahmen des GLOWA-Elbe-Projektes Bedarf an Simulationen dieser Erwärmungen bestand, habe ich sie trotz der Verletzung des Begrenzungskriteriums durchgeführt. Wie Abschnitt 5.4 zeigt, sind die räumlichen Strukturen der Änderungen für beide Antriebe sehr ähnlich, das Verfahren erzeugt also auch jenseits der eigentlichen Grenze seiner Anwendbarkeit robuste Ergebnisse. Der Schwerpunkt der Auswertungen dieses Kapitels wird jedoch auf den Simulationen zum ECHAM4-Antrieb liegen.

Dass die Trendvorgabe, wie hier geschehen, für eine einzelne Station aus einem Flächenmittel übernommen werden darf, ist im Allgemeinen nicht gegeben. Besonders für die Temperatur, die keine physikalische Erhaltungsgröße darstellt, können Flächenmittel und Stationswert weit auseinanderliegen, ohne dabei physikalisch inkonsistent zu sein. Diese Trends sollten deshalb als eine sinnvolle und plausible Vorgabe für die Temperaturentwicklung angesehen werden, die jedoch auch in der Wirklichkeit des GCMs nicht unbedingt der Entwicklung an den repräsentativen Stationen entsprechen muss.

Die Parametereinstellungen des Verfahrens wurden aus dem Kreuzvalidierungsexperiment beibehalten (Blocklänge, Toleranz beim julianischen Datum etc.). Die Toleranz bei der Regressionsvorgabe entsprach einer maximalen Abweichung zwischen vorgegebener und simulierter Regressionsgeraden von 0.1K. Wiederum wurden jeweils 1000 Simulationen erzeugt.

5.2 Die Simulationen zur ECHAM4-Vorgabe

Die Analyse eines Ensembles von Simulationen vermittelt ein Maß für die Streubreite der klimatischen Entwicklungen unter den gegebenen Randbedingungen, kann jedoch unter Umständen sehr aufwändig sein. Im GLOWA-Elbe-Projekt beispielsweise dienen die mit dem Verfahren erzeugten Simulationsreihen als Input für hydrologische Modelle, deren Ergebnisse wiederum an andere Modelle weitergereicht werden. Mit dem ganzen Modellverbund 1000 Simulationen durchzuspielen ist schon aufgrund der damit verbundenen Rechenzeit illusorisch und auch nicht nötig – wie die vorangegangenen Auswertungen in Kapitel 4 zeigen, ist der Bereich, den die 1000 Simulationen aufspannen, für bestimmte Statistiken, vor allem solche bezüglich der Temperatur, eng. Es stellt sich also die Frage, wie aus den 1000 einzelne Simulationen ausgewählt werden können, die für bestimmte Fragestellungen das ganze Spektrum der 1000 gut widerspiegeln.

Die Fragestellung, die das gesamte GLOWA-Elbe-Projekt bewegt, sind die Auswirkungen klimatischer Veränderungen auf den Wasserhaushalt. Ich habe deshalb, wie bereits im Kreuzvalidierungsexperiment, die Simulationen auf einer Skala von trocken nach feucht angeordnet, wiederum anhand der klimatischen Wasserbilanzreihen (KWB, siehe Gleichung (4.1)) an den vier repräsentativen Stationen. Für jede der Stationen wurde die mittlere KWB sowie die Differenz der Mittelwerte aus zweiter und erster Hälfte des Simulationszeitraums bestimmt – zwei Parameter, die grob Niveau und zeitliche Entwicklung erfassen. Insgesamt wird eine Simulation also durch acht Parameter charakterisiert.

Um drei Simulationen auszuwählen, die bezüglich dieser Parameter die ganze Bandbreite der Simulationen erfassen, habe ich für jeden der acht Parameter das 5%- und das 95%-Quantil sowie den Median bestimmt und die Simulationen ausgewählt, die zu dem jeweiligen 8-Tupel im Sinne der euklidischen Distanz am nächsten liegen. Diese drei Simulationen, charakterisiert in Tabelle 5.1, werden im folgenden untersucht. Dabei sollte man nicht vergessen, dass der Mediansimulation für die Interpretation ein stärkeres Gewicht als den anderen beiden zukommt, die den Extrembereich der simulierten Bandbreite abdecken: Die Mediansimulation enthält in etwa, was im Mittel aus dem Ensemble der 1000 zu erwarten ist.

	Grambeck		Potsdam		Gera		Churanov		
Simulation	MW	Diff.	MW	Diff.	MW.	Diff.	MW.	Diff.	Beschr.
5%	41	-68	-131	-66	-86	-68	482	-70	trocken
50%	46	-11	-112	-31	-68	-39	520	-54	mittel
95%	81	54	-100	19	-48	30	555	23	feucht

Tabelle 5.1: Charakterisierung der drei ausgewählten Simulationen: Für jede der vier repräsentativen Stationen mittlere Jahressumme der KWB sowie die Differenzen zwischen der mittleren Jahressumme von 2030–2055 bzw. 2004–2029 [mm].

In manchen Abbildungen, die Ergebnisse aller drei Simulationen zusammenfassen, verwende ich rot für die trockene, blau für die feuchte und schwarz für die Mediansimulation.

5.3 Auswertung der ausgewählten Simulationen zur ECHAM4-Vorgabe

Der Schwerpunkt dieser Auswertung wird auf Änderungen im atmosphärischen Wasserangebot liegen, also z. B. auf den Entwicklungen mittlerer Niederschlagsmengen, von Dürreperioden und Starkregenereignissen. Dabei soll auch das Wechselspiel zwischen den verschiedenen meteorologischen Größen und seine Bedeutung für den Wasserhaushalt betrachtet werden, z. B. die mit steigender Temperatur einhergehende Zunahme der potentiellen Verdunstung.

Beginnen möchte ich mit einer kurzen Untersuchung der Temperatur, der sich Auswertungen von Niederschlag und klimatischer Wasserbilanz anschließen werden. Die Tests aus Anhang B werden in der gleichen Weise verwendet wie für die Auswertungen in Kapitel 4.

5.3.1 Die Zukunft der Temperatur

Als Einstimmung zeigt Abbildung 5.2 die über alle Stationen gemittelten Jahresmitteltemperaturen der Beobachtungsperiode und für den Simulationszeitraum entsprechend den drei Simulationen. Der Vorgabe entsprechend sind sich die Regressionsgeraden zu den drei Simulationen sehr einig.

Karte 5.3 zeigt die 1951–2003 beobachteten Mittelwerte der Temperatur und die Veränderungen zum Mittel 2004–2055. Entsprechend der Regressionsvorgabe liegen diese Änderungen um 1K, siehe Tabelle 5.2. Auffällig ist, dass die stärksten Änderungen im Mittellauf der Elbe simuliert werden, der Gegend, in der bereits in der Beobachtungsperiode die höchsten Mitteltemperaturen beobachtet wurden. All diese Änderungen sind statistisch signifikant, siehe Tabelle 5.2 (die Tabelle ist Abschnitt 4.4 entsprechend aufgebaut).

Ein ähnliches Bild zeigt das 90%-Quantil der Tagesmaxima der Temperatur (Karte 5.4). Auch hier treten sowohl die hohen beobachteten Werte als auch die großen Zunahmen bevorzugt im Mittellauf auf.

In Karte 5.5 sind die mittleren Andauern der Sommertage im Beobachtungszeitraum gezeigt: Die hohen Andauern finden sich ebenfalls im Mittellauf, passend zu allgemein hohen Temperaturen in diesem Gebiet. Rechts davon sind ihre simulierten Änderungen dargestellt. Sie zeigen ein ausgeprägtes räumliches Muster mit der stärksten Zunahme im Mittellauf, die dort bis zu 0.5 Tage erreicht. Zu statistisch gesicherten Unterschieden führt diese Zunahme für eine Irrtumswahrscheinlichkeit von 1% allerdings bei keiner einzigen Station (dies entspricht dem Ansatz des Verfahrens, sich für die Persistenz an der Persistenz der Beobachtungsreihen zu orientieren). Insgesamt nimmt auch die Anzahl der Sommertage zu, siehe Tabelle 5.2. Im Gegensatz zur Entwicklung der Andauern ist die Zunahme der Sommertage für praktisch alle Stationen statistisch signifikant.



Abbildung 5.2: Jahres- und Gebietsmittel der Temperatur in Beobachtungs- und Simulationsperiode entsprechend den drei ausgewählten Simulationen (blau=feucht, schwarz=Median, rot=trocken).



Abbildung 5.3: Veränderung des Mittelwerts der Temperatur. Links die beobachteten Mittel 1951–2003, rechts die Differenz vom simulierten Mittel 2004–2055 zum beobachteten Mittel. Median-Simulation.



Abbildung 5.4: Veränderung des q90 des Maximums der Temperatur. Links die beobachteten Quantile 1951–2003, rechts die Differenz vom simulierten Quantil 2004–2055 zum beobachteten Quantil. Median-Simulation.



Abbildung 5.5: Veränderung der mittleren Andauer der Sommertage. Links die beobachteten Mittel 1951–2003, rechts die Differenz vom simulierten Mittel 2004–2055 zum beobachteten. Median-Simulation.

		Abw. aller Stat.		<i>p</i> -Werte	Gebiet		♯ Stationen <i>p</i> <		
Sim.	. Statistik Mittel Min./Max.		Min./Max.	Abw.	p-Wert	0.1	0.05	0.01	
5%	MT (t)	0.99	0.83 / 1.08	0.00 / 0.00	0.99	0.00	63	63	63
50%	MT (t)	1.02	0.86 / 1.09	0.00 / 0.00	1.02	0.00	63	63	63
95%	MT (t)	1.00	0.83 / 1.10	0.00 / 0.00	1.00	0.00	63	63	63
5%	VSP (KS)	0.29	0.10 / 0.52	0.02 / 1.00	0.25	0.50	12	5	0
50%	VSP (KS)	0.34	-0.04 / 0.56	0.05 / 1.00	0.33	0.64	2	1	0
95%	VSP (KS)	0.27	0.04 / 0.44	0.05 / 1.00	0.27	0.29	7	1	0
5%	VS (KS)	12.82	0.51 / 16.95	0.00 / 0.38	13.76	0.00	62	61	61
50%	VS (KS)	10.59	0.60 / 14.18	0.00 / 0.04	10.47	0.00	63	63	61
95%	VS (KS)	10.73	0.43 / 15.62	0.00 / 0.38	10.80	0.00	61	61	60

Tabelle 5.2: Tests von Statistiken der Temperatur. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

MT (t): Mittelwert der Temperatur, t-Test, [K].

VSP (KS): Verteilung der Andauern von Sommertagen, KS-Test, [Tage].

VS (KS): Verteilung der jahresweisen Anzahlen der Sommertage ($T_{max} \ge 25^{\circ}$ C), KS-Test, [Tage].

In den Karten dargestellt sind lediglich die Veränderungen entsprechend der Mediansimulation. Wie Tabelle 5.2 jedoch zeigt, sind die Unterschiede zwischen den ausgewählten drei Simulationen gering, so dass die Betrachtung des Medians ausreichend ist. Die geringe Variationsbreite liegt einfach daran, dass die Vorgaben an die Temperaturentwicklung in allen Simulationen die gleiche ist.

Zusammenfassend lässt sich also sagen, dass alle drei Simulationen eine Erwärmung zeigen, die im Mittellauf der Elbe besonders ausgeprägt ist. Dies gilt sowohl für das Mittel als auch für den heißen Randbereich der Verteilung. Bei den Andauern der Sommertage lässt sich im Mittel eine leichte Zunahme feststellen, die statistisch allerdings nur in Ausnahmefällen gesichert ist.

5.3.2 Die Zukunft des Niederschlags

Für die Entwicklung des Niederschlags wird bei der Simulation keine äußere Randbedingung vorgegeben – die Streubreite zwischen den einzelnen Simulationen wird daher größer ausfallen als bei der Temperatur.

Das zeigt sich schon in Abbildung 5.6 mit den über alle Stationen gemittelten Jahresniederschlägen der Beobachtungsperiode und für den Simulationszeitraum entsprechend den drei Simulationen. Auch wenn den genauen Verläufen der Ausgleichsgeraden nicht zu viel Gewicht beigemessen werden darf – bei den enormen Schwankungsbreiten der Jahressummen sind ihre Parameter statistisch nicht zu sichern – unterscheiden sie doch feuchte, trockene und Mediansimulation.

Zusätzlich zu den Karten der Mediansimulation zeige ich daher Boxplots, die die Auswertungen für alle drei Simulationen zusammenfassen. Dargestellt sind



Gebietsmittel der Jahresniederschläge

Abbildung 5.6: Jahresniederschläge, im Gebietsmittel. Beobachtungs- und Simulationsperiode entsprechend den drei ausgewählten Simulationen.

dabei die Veränderungen, relativ und absolut, die sich entsprechend den drei Simulationen von Beobachtungs- zu Simulationsperiode ergeben.

Die ersten Darstellungen betreffen die mittleren Jahres-, Sommer- und Winterniederschläge. Die Jahresniederschläge zeigen für Median- und feuchte Simulation einen gemittelten Rückgang um knapp 4% (Abbildung 5.8), der für die Mediansimulation in Küstennähe kaum, weiter elbaufwärts etwas stärker ausgeprägt ist (Karte 5.7). Trotz der relativ kleinen Änderung sind die Unterschiede für trockene und Mediansimulation statistisch für viele Stationen und im Gebietsmittel signifikant, siehe Tabelle 5.3. Die feuchte Simulation zeigt hingegen weitgehend unveränderte Niederschläge an.

Ein im Grunde ähnliches Bild zeichnen die Sommerniederschläge: Alle drei Simulationen zeigen abnehmende Sommerniederschläge (siehe Abbildung 5.10), die für feucht und Mediansimulation um etwa 8% im Mittel abnehmen, für die trockene Simulation sogar um 11%. Die simulierten Unterschiede zwischen den Simulationen können an einzelnen Stationen erheblich variieren und in Einzelfällen sowohl positive als auch negative Entwicklungen umfassen. Generell lässt sich in dieser Abnahme kein räumliches Muster erkennen (vergleiche Karte 5.9).

Das ist beim Winterniederschlag anders. Die Karte für die Mediansimulation (Abbildung 5.11) zeigt deutlich, dass die Niederschläge in Küstennähe zu-



Abbildung 5.7: Veränderung der mittleren Jahresniederschläge. Links die beobachteten mittleren Jahressummen 1951–2003, rechts die Differenz von simulierten mittleren Summen 2004–2055 zu beobachteten. Median-Simulation. Auch hier: Die Niederschläge der Gebirgsstationen überschreiten die hier dargestellten 1000mm.



Abbildung 5.8: Veränderung der mittleren Jahresniederschläge als Boxplots.



Abbildung 5.9: Veränderung der mittleren Sommerniederschläge. Links die beobachteten Summen 1951–2003, rechts die Differenz von simulierten Summen 2004–2055 zu beobachteten. Median-Simulation.



Abbildung 5.10: Veränderung der mittleren Sommerniederschläge als Boxplots.



Abbildung 5.11: Veränderung der mittleren Winterniederschläge. Links die beobachteten Summen 1951–2003, rechts die Differenz von simulierten Summen 2004–2055 zu beobachteten. Median-Simulation.

nehmen, während die Niederschläge im tschechischen Teil des Gebiets abnehmen. Die über das Gebiet gemittelte Änderung ist entsprechend klein, wie Abbildung 5.12 zeigt, auch für die trockene Simulation. Hingegen nehmen die Winterniederschläge in der feuchten Simulation um im Mittel rund 10% zu. Alle Simulationen zeigen die räumliche Struktur, die in Karte 5.11 dargestellt ist.

Trotz der geringen relativen Unterschiede sind die Veränderungen für einen Großteil der Stationen signifikant, siehe Tabelle 5.3.

Ebenfalls geringe Veränderung zeigen sich im rechten Rand der Niederschlagsverteilung bei der Betrachtung des 90%-Quantils: Dieses ändert sich für alle drei Simulationen an allen Stationen nicht mehr als etwa 5% (Abbildung 5.14), es nimmt dabei im deutschen Teil östlich der Elbe ab, westlich davon zeigt sich eine Zunahme (Karte 5.13). Während in der Mediansimulation die Veränderungen im Mittel negativ sind, nehmen in den beiden anderen Simulationen die Werte des Quantils im Mittel zu.

Von besonderer Bedeutung für Land-, Forst- und Wasserwirtschaft sind Perioden und Tage ohne Niederschlag. In den Abbildungen zur mittleren Trockenperiode bzw. ihrer Änderung im Simulationszeitraum (Karte 5.15 und Boxplot 5.16) zeigt die Mediansimulation im Mittel eine leichte Zunahme der Trockenperioden um 0.03 Tage (siehe Tabelle 5.3). Räumlich differenziert betrachtet zeigt sich, dass es in Küstennähe sogar eine leichte Abnahme um bis zu 0.1 Tage ge-



Abbildung 5.12: Veränderung der mittleren Winterniederschläge als Boxplots.



Abbildung 5.13: Veränderung des q90 des Niederschlags. Links die beobachteten Quantile 1951–2003, rechts die Differenz der simulierten Quantile 2004–2055 zu beobachteten. Median-Simulation.



Abbildung 5.14: Veränderung des q90 der Niederschläge als Boxplots.

ben kann, während sich mit zunehmender Entfernung von der Küste die Trockenperioden um bis zu 0.2 Tage verlängern. Dieses Bild passt zur Entwicklung der Winterniederschläge, die in der Mediansimulation in Küstennähe zu- und im kontinentaleren Teil des Gebiets abnehmen. Statistisch lassen sich die Unterschiede kaum sichern.

Zu dieser räumlichen Struktur passt die Karte zur mittleren Anzahl von trockenen Tagen pro Jahr, Abbildung 5.17. Auch hier zeigt sich, dass die Trockentage in Küstennähe leicht abnehmen, weiter elbaufwärts nehmen sie, wie auch im Gesamtmittel, zu. Diese Tendenz zeigt sich in allen drei Simulationen, ebenfalls in allen Simulationen zeigt sich die räumliche Struktur der Veränderung (siehe Abbildung 5.18). Im Gegensatz zu den Andauern der Trockentage sind die Veränderungen bei der Anzahl der Trockentage an einigen Stationen signifikant.

Zusammenfassend lässt sich sagen, dass der Gesamtniederschlag nur leicht zurückgeht. Der Rückgang findet vor allem im Sommer statt, für den an allen Stationen ein Rückgang zu erwarten ist, der am Unterlauf der Elbe am schwächsten ausgeprägt ist. Dieser Gradient verstärkt sich im Winter noch, für den in Küstennähe sogar mit einer leichten Zunahme der Niederschläge zu rechnen ist, während sie für Mittel- und Oberlauf abnehmen.

Dem entspricht eine leichte Abnahme der Trockenperioden an der Küste, während sie im kontinentaleren Teil länger werden. Auch die Anzahl der Trockentage, die insgesamt zunimmt, wächst am Mittel- und Oberlauf stärker als am



Abbildung 5.15: Veränderung der mittleren Trockenperiode. Links die beobachteten Mittel 1951–2003, rechts die Differenz der simulierten Mittel 2004–2055 zu beobachteten. Median-Simulation.



Abbildung 5.16: Veränderung der mittleren Trockenperiode als Boxplots.



Abbildung 5.17: Veränderung der mittleren Anzahl von Trockentagen pro Jahr. Links die beobachteten Mittel 1951–2003, rechts die Differenz der simulierten Mittel 2004–2055 zu beobachteten. Median-Simulation.



Abbildung 5.18: Veränderung der mittleren Anzahl von Trockentagen pro Jahr als Boxplots.

		Abw. aller Stat.		<i>p</i> -Werte	Gebiet		♯ Stationen <i>p</i> <		
Sim.	Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
5%	MP (MW)	-3.51	-8.22 / 0.72	0.00 / 0.98	-3.38	0.00	57	53	41
50%	MP (MW)	-3.50	-6.68 / 0.40	0.00 / 0.62	-3.32	0.00	52	47	40
95%	MP (MW)	0.36	-4.00 / 4.71	0.00 / 1.00	0.66	0.79	10	8	1
5%	SP (MW)	-11.46	-22.12 / 0.47	0.00 / 0.00	-11.59	0.00	63	63	63
50%	SP (MW)	-8.61	-16.20 / 1.24	0.00 / 0.01	-8.52	0.00	63	63	63
95%	SP (MW)	-8.05	-16.48 / 0.84	0.00 / 0.06	-8.00	0.00	63	62	53
5%	WP (MW)	4.08	-5.64 / 14.04	0.00 / 1.00	5.10	0.08	30	20	10
50%	WP (MW)	2.58	-5.29 / 9.72	0.00 / 1.00	3.19	0.12	18	16	8
95%	WP (MW)	9.44	0.89 / 15.99	0.00 / 0.91	10.02	0.01	43	36	21
5%	VTrP (KS)	0.02	-0.09 / 0.16	0.02 / 1.00	-0.05	1.00	9	6	0
50%	VTrP (KS)	0.03	-0.11 / 0.18	0.03 / 1.00	0.04	0.84	5	1	0
95%	VTrP (KS)	-0.03	-0.13 / 0.06	0.33 / 1.00	-0.05	1.00	0	0	0
5%	VTr (KS)	5.95	-1.58 / 13.68	0.00 / 0.98	4.66	0.04	16	13	7
50%	VTr (KS)	5.21	0.48 / 10.58	0.00 / 1.00	2.31	0.23	14	11	3
95%	VTr (KS)	1.40	-2.58 / 8.18	0.06 / 1.00	2.54	0.33	2	0	0

Tabelle 5.3: Tests von Statistiken des Niederschlags. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

MP (MW): Mittlerer Niederschlag, Mann-Whitney-Test, [%].

SP (MW): Mittlerer Sommerniederschlag, Mann-Whitney-Test, [%].

WP (MW): Mittlerer Winterniederschlag, Mann-Whitney-Test, [%].

VTrP (KS): Verteilung der Andauern trockener Tage, KS-Test, [Tage].

VTr (KS): Verteilung der jahresweisen Anzahlen von trockenen Tagen, KS-Test, [Tage].

Unterlauf. Nur sehr geringe Änderungen scheinen im rechten Rand der Niederschlagsverteilung stattzufinden – die 90%-Quantile von simulierten und beobachteten Reihen unterscheiden sich nur minimal, wobei sich das räumliche Muster einer Abnahme im Nordosten und einer Zunahme im Südwesten andeutet.

Auch wenn sich viele der hier vorgestellten Unterschiede sichern lassen, sind sie doch in der Regel klein, simulierte und beobachtete Niederschlagsklimatologie unterscheiden sich nur geringfügig. Da der Niederschlag allerdings nicht alleine für die Wassermenge, die zwischen Atmosphäre und Boden ausgetauscht wird, verantwortlich ist, muss für eine vollständige Betrachtung das Wechselspiel des Niederschlags mit den sonstigen atmosphärischen Zustandsgrößen untersucht werden. Eine Größe, die dieses Wechselspiel zusammenfasst, ist die schon mehrfach erwähnte klimatische Wasserbilanz (KWB), die die Differenz zwischen Niederschlag und potentieller Verdunstung angibt und damit ein Maß für die Wassermenge ist, die der Boden aus der Atmosphäre erhält bzw. maximal an sie abzugeben hat. Diese Größe ist Gegenstand des nächsten Abschnitts.

5.3.3 Die Zukunft der klimatischen Wasserbilanz

Da die KWB für den Wasserhaushalt eine entscheidende Rolle spielt, werden die Veränderungen in allen drei Simulationen gegenüber der Beobachtungsperiode in eigenen Karten dargestellt. Beispielsweise zeigt Karte 5.19 die mittleren Jahressummen der KWB, wie sie im Beobachtungszeitraum beobachtet wurden bzw. die Änderungen der mittleren Jahressummen in den drei Simulationen. Bereits in der Beobachtungsperiode zeigt sich die KWB als sehr gering, mit weiten Teilen besonders im Mittellauf und im böhmischen Becken, die nahe an Null oder sogar darunter sind. In Küstennähe ist die Bilanz hingegen leicht positiv. Ausgeprägte "Überschüsse" finden sich lediglich an den Gebirgsstationen, an denen durch die geringeren Temperaturen wenig von den außerdem reichlichen Niederschlägen verdunsten kann.

In den drei weiteren Karten des Tableaus sind die Differenzen der simulierten mittleren Jahressummen zu den beobachteten dargestellt. Die Simulation zum 5%-Quantil (siehe Abschnitt 5.2) zeigt erwartungsgemäß die stärkste Abnahme in der Wasserbilanz, die wiederum besonders stark am Mittellauf und im böhmischen Becken ausgeprägt ist. Die Median- und besonders die 95%-Simulation zeigen diese räumliche Struktur noch stärker, wobei insgesamt die Verringerung der mittleren Summe nicht ganz so stark ausgeprägt ist. Dieses Gefälle in der Abnahme entspricht der Auswahl der Simulationen, die den von den 1000 Simulationen abgedeckten Bereich auf der feucht-trocken-Skala erfassen sollen. Auch wenn sich der Betrag der Abnahme zwischen den Simulationen unterscheidet, ist keiner Simulation zufolge mit einer Stabilisierung oder gar Zunahme der KWB zu rechnen. Selbst in der feuchten Simulation finden sich Abnahmen von bis zu 60mm. Relativ zu den grob 600mm Jahresniederschlag, die z. B. in Potsdam bisher beobachtet werden, ist eine solche Verringerung erheblich. Nach Tabelle 5.4 sind die Unterschiede für trockene und Mediansimulation für fast alle Stationen signifikant, für die feuchte auf dem 5%-Niveau immerhin an rund 2/3 der Stationen. Die Mittelwerte der über alle Stationen gemittelten Reihen erweisen sich für alle Simulationen als vom Mittelwert der Beobachtungsperiode signifikant verschieden.

Dazu passen die in Karte 5.20 dargestellten Entwicklungen des 10%-Quantils der Wasserbilanz. Auch hier lässt sich in allen Simulationen eine Verschiebung ins Negative feststellen, die v.a. für die 5%-Quantils-Simulation ausgeprägt erscheint. Von dieser Verschiebung sind die Gebirgsstationen besonders betroffen.

Betrachtet man die Entwicklung der mittleren Verweildauern der KWB unter Omm (Karte 5.21), so fallen für die Beobachtungsperiode die hohen Andauern im Nordosten des Mittellaufs sowie im böhmischen Becken auf – passend zu den allgemein niedrigen Bilanzen in diesen Gegenden. Gemäß aller drei Simu-





Abbildung 5.19: Veränderung der mittleren Jahressummen der KWB entsprechend den drei ausgewählten Simulationen gegenüber der Beobachtungsperiode, deren Summen oben links dargestellt sind.



Abbildung 5.20: Veränderung des 10%-Quantils der KWB entsprechend den drei ausgewählten Simulationen gegenüber der Beobachtungsperiode, deren Quantile oben links dargestellt sind.

lationen scheint sich diese Strukturierung in den nächsten Jahrzehnten zu verstärken, wobei die Zunahme dieser "Bilanz-Trockenperioden" der Auswahl der Simulationen entsprechend am stärksten in der trockenen ausgeprägt ist. Statistisch sind diese Unterschiede allerdings nicht zu sichern.

Die Entwicklung hin zu einem trockenen Klima zeichnet sich nicht nur beim Vergleich zwischen Simulations- und Beobachtungsreihen ab: Auch innerhalb der simulierten Reihen zeigen sich für die trockene und die Mediansimulation deutliche Tendenzen einer abnehmenden Wasserbilanz. Die Karten in Abbildung 5.22 zeigen die linearen Trends in den Jahressummen der KWB für die beobachteten und simulierten Reihen. In der Beobachtungsperiode sind Zu- und Abnahme im Gebietsmittel einigermaßen ausgeglichen, wobei die Küstengegend sowie die Berglagen eine leichte Zunahme verzeichnen. Am Mittellauf und im böhmischen Becken nimmt die Bilanz hingegen ab. Gemäß der 5%-Quantils-Simulation finden sich im ganzen Gebiet Abnahmen über die Simulationsperiode. Die Mediansimulation zeigt ein ähnliches räumliches Muster wie in der Beobachtungsperiode mit leichten Zunahmen an der Küste und an einigen Gebirgsstationen. Im Gesamtgebiet dominiert allerdings die Abnahme. Insgesamt eine positive Entwicklung findet sich lediglich in der feuchten Simulation beim Mittel über alle Stationstrends, die besonders in Küstennähe positiv ausfallen. Der Trend der über alle Stationen gemittelten Reihen hingegen ist auch hier negativ. Nach Tabelle 5.4 ist in der Mediansimulation keiner der simulierten Trends gesichert, für die trockene Simulation ergeben sich an einigen wenigen Stationen signifikante Trends, in der feuchten sind es noch weniger. Die geschätzten Trends wurden hierfür mit dem Test aus Abschnitt B.2 gegen die Nullhypothese "kein Trend" geprüft.

Wenn auch statistisch hier nicht zu sichern, muss die Möglichkeit einer sich auch in Zukunft fortsetzenden Abnahme der KWB nach diesen Ergebnissen in Betracht gezogen werden.

5.4 Vergleich der EH4- und EH5-Simulationen

In diesem Abschnitt werden anhand von Karten zu den Mediansimulationen beider Ensemble wichtige Statistiken, die mittlere Zustände, Persistenz und Extreme betreffen, dargestellt. Damit soll insbesondere die physikalische Plausibilität der Simulation zur ECHAM5-Vorgabe geprüft werden, der eine Regressionsvorgabe zugrunde liegt, die sich außerhalb des in Abschnitt 3.2 definierten Anwendungsbereichs des Verfahrens befindet. Abbildung 5.23 beispielsweise zeigt die langjährigen Mittel der Temperatur, die in beiden Simulationen im Mittellauf der Elbe am höchsten liegen, für die ECHAM5-Trends um etwa 1K höher entsprechend der stärkeren vorgegebenen Erwärmung.

Ebenfalls übereinstimmende Muster zeigen die Karten zum mittleren Jahres-



Abbildung 5.21: Veränderung der Verweildauer der KWB unter 0mm entsprechend den drei ausgewählten Simulationen gegenüber der Beobachtungsperiode, deren Verweildauern oben links dargestellt ist.



Trend q5 Simulation



Trend q50 Simulation Trend q95 Simulation mm mm -100 -50 -100 -50 0 50 100 150

Abbildung 5.22: Lineare Trends in den Jahressummen der KWB in der Beobachtungsperiode (oben links) und entsprechend den drei Simulationen 2004–2055.

		Abw. aller Stat.		<i>p</i> -Werte	Gebiet		\sharp Stationen $p <$		
Sim.	Statistik	Mittel	Min./Max.	Min./Max.	Abw.	p-Wert	0.1	0.05	0.01
5%	MW (t)	-77.49	-135.56 / -35.88	0.00 / 0.03	-77.49	0.00	63	63	58
50%	MW (t)	-70.60	-103.94 / -35.46	0.00 / 0.07	-70.60	0.00	63	62	61
95%	MW (t)	-41.76	-69.95 / 15.69	0.00 / 0.38	-41.76	0.00	48	41	33
5%	V0mmWP (KS)	-0.01	-0.14 / 0.11	0.03 / 1.00	-0.01	1.00	1	1	0
50%	V0mmWP (KS)	0.05	-0.08 / 0.19	0.18 / 1.00	-0.12	0.60	0	0	0
95%	V0mmWP (KS)	0.01	-0.13 / 0.11	0.11 / 1.00	-0.03	1.00	0	0	0
5%	TrW (t)	-83.59	-168.46 / 35.46	0.01 / 0.40	-40.51	0.29	23	8	1
50%	TrW (t)	-17.87	-101.52 / 57.52	0.14 / 0.40	-53.83	0.21	0	0	0
95%	TrW (t)	52.94	-47.11 / 253.58	0.05 / 0.40	-10.52	0.39	10	1	0

Tabelle 5.4: Tests von Statistiken der Wasserbilanz. Beschreibung der Tabelle am Anfang von Abschnitt 4.4.

MW (t): Mittlere Wasserbilanz, *t*-Test, [mm].

V0mmWP (KS): Verteilung der Andauern der Tage mit KWB< 0mm, KS-Test, [Tage]. TrW (t): Lin. Trend in den Jahressummen der KWB, *t*-Test, [mm].



Abbildung 5.23: Langjährige Mittel der Temperatur 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).


Abbildung 5.24: Langjährige Mittel der Jahresniederschläge 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).

niederschlag in Abbildung 5.24 mit besonders hohen Niederschlägen in Küstennähe und den Berglagen. Für die Simulation zur ECHAM5-Vorgabe fallen sie insgesamt geringfügig höher aus.

Dieses Bild wiederholt sich bei den mittleren Winterniederschlägen in Abbildung 5.25, wohingegen die Sommerniederschläge in Abbildung 5.26 ein umgekehrtes Bild zeichnen: Hier sind die simulierten Niederschläge entsprechend der Simulation mit dem ECHAM4-Antrieb geringfügig höher. Die jahreszeitliche Variabilität ist beim Niederschlag also für die größere vorgegebene Erwärmung etwas ausgeprägter. Die räumliche Verteilung der Niederschläge ist in allen gezeigten Karten gut übereinstimmend.

Ebenfalls gut übereinstimmend sind die räumlichen Verteilungen der mittleren Andauern von Sommertagen, gezeigt in Abbildung 5.27. Passend zu den hohen Temperaturen am Mittellauf der Elbe finden sich in dieser Region auch die längsten Andauern, die erwartungsgemäß für die Simulation mit dem ECHAM5-Antrieb etwas höher ausfallen.

Auch für die 90%-Quantile der täglichen Maxima der Temperatur zeigt die Simulation der ECHAM5-Vorgabe eine etwas wärmere Verteilung als bei ECHAM4, wiederum in beiden Simulationen übereinstimmend mit den höchsten Werten im Mittellauf (siehe Abbildung 5.28).

Die mittleren Anzahlen der Trockentage pro Jahr sind in Abbildung 5.29 dargestellt. Wie schon in den vorangegangenen Karten gleichen sich die räum-



Abbildung 5.25: Langjährige Mittel der Winterniederschläge 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).



Abbildung 5.26: Langjährige Mittel der Sommerniederschläge 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).



Abbildung 5.27: Mittlere Andauer von Sommertagen ($T_{max} \ge 25^{\circ}$ C) 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).



Abbildung 5.28: 90%-Quantil von T_{max} 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).



Abbildung 5.29: Mittlere Anzahl der Trockentage pro Jahr 2004–2055 entsprechend den Mediansimulationen zu ECHAM5-Trends (links) und ECHAM4-Trends (rechts).

lichen Strukturen in beiden Simulationen, die Simulation mit dem ECHAM5-Antrieb gibt dabei insgesamt mehr Trockentage für die Simulationsperiode an als die Simulation für ECHAM4.

Zusammenfassend lässt sich sagen, dass die räumliche Verteilung der dargestellten Statistiken für beide Simulationen vergleichbar ist. Unterschiede ergeben sich entsprechend der verschiedenen vorgegebenen Temperaturentwicklungen bei den Mittelwerten der Temperatur, mit der Differenzen im rechten Randbereich der Temperaturverteilung und bei den Verweildauern von Sommertagen korrespondieren. Der Niederschlag zeigt in der "wärmeren" Simulation für die ECHAM5-Vorgabe ein wenig ausgeprägtere Unterschiede zwischen Sommer- und Winterniederschlag als für ECHAM4. Insgesamt sind die Niederschläge gemäß dem ECHAM5-Antrieb etwas reichlicher, was zugleich mit mehr trockenen Tagen einhergeht. Der Niederschlag in der wärmeren Simulation scheint damit tendenziell seltener, dafür heftiger aufzutreten, die Unterschiede sind jedoch sehr gering.

Trotz der unterschiedlichen Antriebe und der Tatsache, dass die Simulation mit der ECHAM5-Vorgabe den durch das Begrenzungskriterium aus Abschnitt 3.2 erlaubten Bereich verlässt, stellen die beiden Simulationen robuste und zueinander konsistente Entwicklungen dar und es ergeben sich zumindest aus den hier dargestellten Statistiken keine Hinweise darauf, dass die physikalischen Zusammenhänge der Beobachtungsperiode unter dem deutlich verschiedenen Temperaturregime der ECHAM5-Vorgabe nicht mehr verwendbar wären. Dies ist insofern wichtig, als diese Simulation die Grundlage für hydrologische Modellierungen im GLOWA-Elbe-Projekt liefert.

5.5 Zusammenfassung

Die vorangegangenen Auswertungen der Simulationen mit dem ECHAM4-Antrieb zeigen, dass generell mit einem trockeneren Klima zu rechnen ist. Dies kommt weniger durch stark abnehmende Niederschläge als vielmehr durch das Zusammenwirken von potentieller Verdunstung, die ihrerseits stark von der Temperatur abhängt, und Niederschlag zustande. Die Temperatur nimmt gemäß den Vorgaben um grob 1K zu, mit einer besonders ausgeprägten Erwärmung im Mittellauf der Elbe. Die Niederschläge zeigen insgesamt sehr kleine Änderungen, was Mittelwerte, Andauern oder Extreme anbelangt. Im Gesamtmittel nehmen sie um wenige Prozent ab, die sich vor allem einer Abnahme der Sommerniederschläge schulden, die einigermaßen gleichmäßig auf das Gebiet verteilt ist. Die Winterniederschläge nehmen im Mittel um wenige Millimeter zu, wobei sich eine deutliche räumliche Strukturierung feststellen lässt: nahe der Küste verzeichnet sich eine Zunahme, weiter Elb-aufwärts nehmen auch die Winterniederschläge ab. Diese Schlussfolgerungen lassen sich auch für die Simulation mit dem Antrieb von ECHAM5 ziehen.

Betrachtet man die klimatische Wasserbilanz entsprechend den Simulationen zur ECHAM4-Vorgabe, so erweist sich die Simulationsperiode gemäß aller drei Simulationen als trockener als der Beobachtungszeitraum, mit Differenzen in den mittleren Jahressummen, die selbst für die feuchte Simulation noch Beträge von 60mm an einzelnen Stationen erreichen. Bedenkt man den bereits in der Beobachtungsperiode geringen atmosphärischen Wassereintrag (Potsdam z. B. erhält rund 600mm Niederschlag im Jahr), so ist diese Entwicklung bedenklich, besonders, wenn man sich die Darstellung der Trends der KWB in den drei Simulationen vergegenwärtigt: Die Trends der über alle Stationen gemittelten Reihen der KWB sind für alle drei Simulationen negativ. Für die höheren Temperaturen der Simulation gemäß des ECHAM5-Antriebs verschärft sich diese Situation noch (hier nicht gezeigt). Wie aus den Temperatur- und Niederschlagskarten plausibel wird, ist auch die Abnahme der KWB im Mittellauf besonders ausgeprägt.

Den Simulationen zufolge scheinen sich die bereits in der Beobachtungsperiode abzeichnenden Tendenzen weiter fortzusetzen: Allgemein fallen die Veränderungen in Küstennähe moderater aus als weiter im Landesinnern. Warme Temperaturen und Trockenheit werden in Zukunft vor allem am Mittellauf noch häufiger anzutreffen sein als bisher.

Kapitel 6

Zusammenfassung und Ausblick

Die vorliegende Arbeit entstand im Rahmen eines Projekts, das sich mit Untersuchungen über die Auswirkungen des Klimawandels der nächsten Jahrzehnte auf den Wasserhaushalt des Elbe-Einzugsgebiets befasst – einer Region, die im Allgemeinen nicht besonders niederschlagsreich zu nennen ist, mitunter jedoch mit katastrophalen Überschwemmungen zu kämpfen hat. In dieser Region Planungsrichtlinien für die Wasserbewirtschaftung der Zukunft zu entwickeln, setzt zuverlässige und hoch aufgelöste Abschätzungen der klimatischen Entwicklung – insbesondere in Bezug auf den Niederschlag – voraus.

Gesucht war ein Verfahren, das solche Abschätzungen liefert, ohne auf nur selten verfügbare Großrechner angewiesen zu sein. Die Anwendung sollte einfach bleiben und insbesondere ohne zu komplexe Vorgaben auskommen, die häufig nicht verfügbar sind und sich darüber hinaus nicht unbedingt als zuverlässig erweisen. In Frage kam daher nur ein statistischer Ansatz, dem eine der simpelsten denkbaren Vorgaben gemacht wird: Die lineare Entwicklung während der Simulationsperiode für eine ausgewählte Bezugsgröße, die für das Klima der Untersuchungsregion besonders charakteristisch ist.

Simulationsreihen, die mit einer solchen Vorgabe erzeugt werden, sind alles andere als vollständig bestimmt – viele klimatologisch sinnlose Simulationen sind denkbar, die mit dieser Vorgabe verträglich sind. Besonders die anderen meteorologische Variablen außer der Bezugsgröße sind durch diese Vorgabe nur denkbar geringfügig eingeschränkt. Neben dem vorgegebenen Trend und Mittelwert müssen die simulierten Reihen daher noch weitere Eigenschaften aufweisen: Dazu gehören realistische Jahresgänge und Erhaltungsneigungen, außerdem zwei Typen von Konsistenz – zum einen Konsistenz zwischen den einzelnen Beobachtungsvariablen für jeden Zeitschritt, zum anderen Konsistenz der räumlichen Felder der Beobachtungsvariablen. Mangels externer beschränkender Randbedingungen für Eigenschaften dieser Art in der Simulationsperiode werden die Simulationsreihen so konstruiert, dass diese Eigenschaften im Sinne der Beobachtungsreihen realistisch wiedergegeben werden, d. h. dass die Eigenschaften der simulierten und beobachteten Reihen statistisch miteinander verträglich sind.

Um die realistischen Konsistenzen sicherzustellen, benutzt das entwickelte Verfahren einen kombinatorischen Ansatz: Die Simulationsreihen werden aus Abschnitten der Beobachtungsreihen zusammengesetzt, wobei die Abschnitte die Beobachtungen aller untersuchten Stationen enthalten. Auf diese Weise bestehen die Simulationsreihen aus Variablenkombinationen und Feldern, die in der Beobachtungsperiode bereits aufgetreten sind und damit physikalisch Sinn machen. Diesem Ansatz liegt die Annahme zugrunde, dass sich bestimmte in der Beobachtungsperiode aufgetretene Witterungsabfolgen so oder zumindest ähnlich auch in der Simulationsperiode wiederholen können.

Mit diesem Ansatz lässt sich das Ziel des Verfahrens interpretieren als die Suche nach einer Funktion, die jedem Abschnitt der Simulationsperiode Beobachtungen eines Abschnitts der Beobachtungsperiode zuordnet. Diese Funktion wird aus heuristischen Überlegungen so konstruiert, dass die Trend- und Mittelwertvorgabe und die übrigen angesprochenen Eigenschaften für die entstehende Simulationsreihe sichergestellt sind. Stochastische Elemente in der Konstruktion ermöglichen die Erzeugung großer Ensembles von Simulationen, die eine Abschätzung der Bandbreite möglicher Entwicklungen erlauben.

Dieses Verfahren wurde in einem Validierungsexperiment geprüft, in dem zunächst der Einfluss interner Parametereinstellungen untersucht wurde. Mit den dabei gefundenen Einstellungen wurde eine Kreuzvalidierung durchgeführt, für die der Beobachtungsdatensatz des Elbe-Einzugsgebiets in zwei voneinander unabhängige 25-Jahres-Perioden unterteilt wurde. Mit den Daten der ersten wurde die Klimatologie der zweiten simuliert und mit der tatsächlich beobachteten verglichen. Insgesamt fällt die Übereinstimmung zwischen Simulation und Beobachtung sehr befriedigend aus, auch für den Niederschlag, der traditionell schwierig zu modellieren ist. Ausgewertet wurden klassische Statistiken wie langjährige Mittel und Varianzen, Ouantile sowie Persistenzmaße und Ereignistage. Erfreulich gut werden die für extreme Ereignisse relevanten Statistiken reproduziert, ebenso wie die räumlichen Verteilungen der untersuchten Statistiken. Soweit Abweichungen auftraten – z. B. bei der interannuellen Varianz der Temperatur oder beim Jahresgang des Niederschlags - ergaben sich diese aus der Stationarität des Ansatzes, der bei der bekanntermaßen instationären Klimaentwicklung begrenzte Gültigkeit hat. Ein kurzer exemplarischer Vergleich mit Simulationen des dynamischen Modells REMO für den gleichen Zeitraum zeigt, dass die Übereinstimmung zwischen Simulation und Beobachtung mit dem statistischen Ansatz erheblich besser als mit dem dynamischen Ansatz ist.

Simulationen des Klimas der nächsten Jahrzehnte für das Elbe-Gebiet zeigen an, dass mit einem trockeneren Klima zu rechnen ist. Dies liegt zum einen an den leicht zurückgehenden Niederschlägen, v.a. jedoch an der mit der steigenden Temperatur zunehmenden potentiellen Verdunstung. Interessanterweise scheinen sich dabei die Tendenzen, die sich bereits in der Beobachtungsperiode abzeichnen, fortzusetzen: Die Küstennähe verzeichnet einen Zunahme v.a. der Winterniederschläge, weiter flussaufwärts nehmen die Niederschläge ab. Die Sommerniederschläge werden insgesamt weniger. Für die bereits in der Beobachtungsperiode besonders heiße Region um den Mittellauf der Elbe nehmen die Temperaturen am meisten zu. Das Zusammenwirken von Niederschlag und (u. a.) Temperatur führt dazu, dass auch der "feuchte" Rand des Simulationsensembles eine im Mittel geringere Wasserbilanz angibt als in der Beobachtungsperiode. Diese Entwicklung scheint sich auch innerhalb des Simulationszeitraums fortzusetzen, die Trends der Wasserbilanz der über das ganze Gebiet gemittelten Reihen sind negativ. Für ein Gebiet wie das Elb-Einzugsgebiet muss eine solche Erwartung bedenklich stimmen – an dessen Stationen fallen geringe Jahresniederschläge von meist 600–700mm, die für Wassernutzungsbereiche wie Forst- und Landwirtschaft nicht gerade üppig sind.

Insgesamt scheinen sich der gewählte Ansatz und seine Umsetzung zu bewähren: Die vorgestellte Validierung erbrachte überzeugende Ergebnisse und die Zukunfts-Simulationen zeigen plausible Klimaentwicklungen an. Dabei darf nicht aus dem Blick verloren werden, dass dieses Verfahren statistisch arbeitet, d. h. dass die Zusammenhänge zwischen und in den einzelnen Beobachtungsgrößen aus der Beobachtungsperiode zugleich für den Simulationszeitraum als gültig angenommen werden. Diese Annahme ist freilich nur zu rechtfertigen, wenn diese beiden Klimatologien einen gewissen Ähnlichkeitsbereich nicht verlassen. Finden wesentliche Veränderungen statt, etwa eine grundsätzliche Umstellung der atmosphärischen Zirkulation, ist dieses Verfahren nur noch sehr eingeschränkt zu verwenden.

Eine Weiterentwicklung könnte daher z. B. darauf hinarbeiten, den dynamischen Aspekt der Zirkulation in die Simulation mit eingehen zu lassen. Ließe sich etwa aus einem GCM-Lauf eine zukünftige Statistik der Großwetterlagen ableiten, so wäre das ein möglicher Prediktor für die Auswahl der Abschnitte der Beobachtungsperiode, aus denen die Simulationsreihen zusammengesetzt werden. Eine weitere Möglichkeit wäre, die Länge der Abschnitte von typischen Verweildauern verschiedener Großwetterlagen leiten zu lassen. Diese stärkere Verknüpfung des Ansatzes mit einer angenommenen dynamischen Entwicklung könnte dazu beitragen, den Einsatzbereich des Verfahrens entscheidend zu erweitern, selbst wenn die inhärenten Beschränkungen natürlich bestehen bleiben: Letztlich steht den Simulationsreihen nur die klimatische Variabilität zur Verfügung, die in den Beobachtungsreihen bereits enthalten ist. Weiterentwicklungen könnten jedoch dazu beitragen, diese Variabilität noch besser auszunutzen.

Neben der Einbeziehung dynamischer Aspekte in das Verfahren könnte die

Weiterentwicklung dahin gehen, Kopplungen zwischen simuliertem Klima und z. B. erwarteten Landnutzungsänderungen zu berücksichtigen. Die einfachste Herangehensweise hierbei wäre vermutlich die Entwicklung eines Nachbearbeitungsverfahrens, das sich an die Erzeugung der Simulationsreihen anschließt: Gelänge es, ein statistisches Modell zu spezifizieren, das Veränderungen der Landnutzung in eine Klimaantwort übersetzt, könnten damit die ohne Berücksichtigung der Kopplung erzeugten Simulationsreihen nachkorrigiert werden.

Auch wenn solche angedachten Erweiterungen den Einsatzbereich des Verfahrens noch erheblich erweitern können, steht mit der bisher entwickelten Fassung bereits ein Werkzeug zur Verfügung, das es ermöglicht, mit relativ geringem Aufwand physikalisch plausible Simulationen zu erzeugen. Die Idee, das Geschehen der Vergangenheit behutsam auf die Zukunft zu übertragen, scheint Sinn zu machen.

Anhang A

Mathematische Begriffe und Notationen

A.1 Allgemeine mathematische Konzepte

Dieser Abschnitt rekapituliert einige mathematische Konzepte, die den Ausführungen in dieser Arbeit zugrunde liegen. Begriffe wie Mengen und deren Elemente, Tupel, und Funktionen gehören zu den mathematischen Grundlagen und verfügen über eine ausgereifte standardisierte Notation, siehe z. B. Fischer und Kaul (1990). Listen spielen eine wesentliche Rolle in funktionalen Programmiersprachen wie Haskell, in deren Kontext sich ebenfalls eine standardisierte Notation entwickelt hat, siehe z. B. Thompson (1999); Doets und van Eijck (2004).

Mengen. Eine Menge *M* ist eine Zusammenfassung verschiedener Objekte. Ein Objekt *m* aus einer solchen Zusammenfassung ist ein Element der Menge $M, m \in M$. Eine geläufige Notation ist, die Elemente einer Menge in geschweifte Klammern einzuschließen: $M = \{m_1, m_2, ...\}$. Die Reihenfolge der Aufzählung der Elemente einer Menge innerhalb der geschweiften Klammern ist unerheblich für die Charakterisierung der Menge, $\{m_1, m_2, m_3\} = \{m_2, m_1, m_3\}$.

Mengenbildungen (englisch: *set comprehension*) ermöglichen die Darstellung bestimmter Mengen mit Hilfe anderer Mengen, beispielsweise die Quadratzahlen $Q = \{x^2 \mid x \in \mathbb{Z}\}.$

Eine Menge \tilde{M} ist eine Teilmenge von M, $\tilde{M} \subset M$, wenn alle Elemente aus \tilde{M} auch Elemente von M sind.

Die Menge aller Teilmengen einer Menge M wird als Potenzmenge, P(M), der Menge M bezeichnet.

Tupel. Tupel sind Zusammenfassungen einer festen Anzahl von Objekten, die aus unterschiedlichen Mengen stammen können. Sie stellen aus den Objekten ein einheitliches Objekt her, für das die Ausgangsobjekte in runden Klammern zusammenfasst werden, z. B. $(a, b) \in A \times B$. Die Reihenfolge der Zusammenfassung ist hier für die Charakterisierung des Tupels entscheidend, $(a, b) \neq (b, a)$.

Listen. Eine Liste ist ebenfalls eine geordnete Zusammenfassung mehrerer Elemente, die jedoch im Gegensatz zum Tupel aus der selben Menge stammen müssen. Eine gegebene Liste endlicher Länge lässt sich daher immer als Tupel interpretieren, nicht jedoch umgekehrt. Die Länge einer Liste ist nicht festgelegt wie beim Tupel, sondern kann durch hinzufügen bzw. entfernen von Elementen verändert werden. Die Elemente einer Liste *L* werden in eckigen Klammern zusammengefasst, $L = [l_1, l_2, ...]$. Auch Listen können wie Mengen leer sein, [], eine endliche Anzahl von Elementen enthalten, $[l_1, l_2, ..., l_n]$, $n \in \mathbb{N}$, oder unendlich viele. Im Gegensatz zu Mengen ist die Anordnung der Elemente in der Liste charakterisierend für die Liste, $[a, b, c] \neq [a, c, b]$, ebenso wie das mehrfache Auftreten eines Elements, $[a, b] \neq [a, a, b]$.

Analog zur Mengenbildung lassen sich mit Hilfe der Listenbildung (englisch: *list comprehension*) Listen in Abhängigkeit von den Elementen anderer Listen angeben, z. B. die Liste aller geraden positiven Zahlen $G = [2n \mid n \leftarrow [1,2,...]]$. Das " \leftarrow " entspricht also dem " \in " bei Mengen. Um Listen aufeinanderfolgender ganzer Zahlen darzustellen, hat sich die Schreibweise [*a..b*], *a*, *b* $\in \mathbb{Z}$ bewährt, z. B. [3..6] = [3,4,5,6].

Analog zu den Teilmengen ist eine Liste \tilde{L} eine Teilliste von L, $\tilde{L} \subset L$, wenn alle Elemente aus \tilde{L} auch Elemente von L sind.

Funktionen. Eine Funktion bildet jedes Element einer Menge *A* auf genau ein Element einer Menge *B* ab, $f : A \to B$. *A* wird als Definitionsbereich bezeichnet. Beispielsweise ordnet $f : \mathbb{R} \to \mathbb{R}_+$, $f(x) = x^2$ einer reellen Zahl ihr Quadrat zu. Die Klammern enthalten in dieser Schreibweise stets das Argument der Funktion, also ein Element der Definitionsmenge. Besonders für die algorithmische Beschreibung des Verfahrens ist es vorteilhaft, diese Schreibweise etwas erweitert zu benutzen, indem die Klammern nicht nur einzelne Elemente, sondern auch Teilmengen der Definitionsmenge oder Listen, deren Elemente aus der Definitionsmenge stammen, umschließen können. Die Funktion wird in diesen Fällen auf jedes Element der übergebenen Menge bzw. Liste angewandt. Bekommt die Funktion *f* von eben beispielsweise die Liste [1,2,3] übergeben, so wäre das Bild dieser Liste unter *f* die Liste [1,4,9], *f*([1,2,3]) = [1,4,9]. Allgemein lässt sich die-

se Erweiterung mit der set bzw. list comprehension darstellen:

$$f(M) = \{f(m) \mid m \in M\},\$$

$$f(L) = [f(l) \mid l \leftarrow L].$$

Ich verwende in dieser Arbeit die Begriffe "Operator" und "Funktion" synonym.

A.2 Statistische Konzepte

Bei der Spezifizierung der Fragestellung werden einige statistische Konzepte verwendet, die ich im folgenden kurz einführen möchte. Eine allgemeine und vollständigere Darstellung findet sich in vielen gängigen Statistiklehrbüchern, z. B. von Storch und Zwiers (1999), Schönwiese (2000) oder Taubenheim (1969). Der Einfachheit halber werde ich mich hier auf den diskreten Fall beschränken.

Ergebnisraum und Ereignis. Die Menge der möglichen Ergebnisse eines beliebigen Zufallsexperiments bilden den sogenannten Ergebnisraum \mathcal{E} , etwa die 4 möglichen Kopf-Zahl-Kombinationen beim zweimaligen Werfen einer Münze, $\mathcal{E} = \{KK, KZ, ZK, ZZ\}$. Eine Untermenge des Ergebnisraums, $E \in P(\mathcal{E})$, wird als Ereignis bezeichnet.

Wahrscheinlichkeit eines Ereignisses. Die Wahrscheinlichkeitsfunktion p ist eine Funktion, die den Ereignissen eine reelle Zahl aus dem Intervall [0, 1] zuordnet,

$$p: P(\mathscr{E}) \to [0,1].$$

Für das Ereignis $E = \{KZ\}$ beim zweimaligen Werfen einer (fairen) Münze wäre die Wahrscheinlichkeit z. B. $p(E) = 1/2 \cdot 1/2 = 1/4$.

Zufallsvariablen und ihre Wahrscheinlichkeitsverteilung. Häufig ist es zweckmäßig, eine reelle Funktion zu betrachten, die von den Ereignissen abhängt:

$$X: P(\mathscr{E}) \to \mathbb{R}$$

und als *Zufallsvariable* bezeichnet wird. Bei den Münzwürfen könnte man beispielsweise Kopf die 1 und Zahl die 0 zuweisen und dann jeweils die Summe bilden (z. B. $X(\{KZ\}) = X(\{ZK\}) = 1)$. X ist variabel, weil sie von den Ereignissen aus $P(\mathscr{E})$ abhängt und zufällig, weil diese Ereignisse nicht vorhersehbar sind.

Der Wertebereich einer Zufallsvariablen wird auch als *Grundgesamtheit* bezeichnet. Sind die Wahrscheinlichkeiten der Ereignisse bekannt (z. B. p(KZ) = p(ZK) = 1/4 beim zweimaligen Münzwurf), so lassen sich damit auch die Wahrscheinlichkeit für die Werte der Zufallsvariable angeben,

$$f_X: \mathbb{R} \to [0,1],$$

z. B. für das zweimalige Münzwerfen $f_X(1) = p(ZK) + p(KZ) = 1/4 + 1/4 = 1/2$ (die Wahrscheinlichkeiten, die den zu einem bestimmten Wert der Zufallsvariablen gehörenden Ereignissen entsprechen, addieren sich auf). In der Notation wird die Zufallsvariable X von einem konkreten Wert x unterschieden, den sie bei einem Experiment annehmen kann. Die Funktion f_X wird als *Wahrscheinlichkeitsverteilung* der Zufallsvariable X bezeichnet, x als *Realisierung* der Zufallsvariable X.

Stichproben. Mehrfache Realisierungen einer Zufallsvariablen bilden eine *Stichprobe* aus der Grundgesamtheit. Eine Stichprobe vom Umfang *n* werde mit

$$\vec{x} = (x_1, x_2, \dots, x_n)$$

bezeichnet. Ihre Auswertung kann Rückschlüsse auf Grundgesamtheit und Verteilungsfunktion der Zufallsvariablen erlauben, aus der die x_i , $i \leftarrow [1..n]$ realisiert sind, indem z. B. bestimmte Parameter der Zufallsvariablen (etwa ihr Mittelwert) anhand der Stichproben *geschätzt* werden (für den Mittelwert durch Berechnung des arithmetischen Mittels). Der Begriff "Schätzen" macht deutlich, dass eine Stichprobe keine definitiven Aussagen über die wahren Parameter ihrer Zufallsvariablen erlaubt.

Testen einer Nullhypothese. Ein statistischer Test dient der Prüfung, ob eine Stichprobe einer bestimmten Annahme (in der statistischen Terminologie *Nullhypothese*) über die Zufallsvariable widerspricht, beispielsweise: Die Stichprobe wurde aus einer Zufallsvariable mit Mittelwert μ realisiert. Ein Test besteht aus drei Stufen:

- 1. Mittels der Nullhypothese wird ein statistisches Modell konstruiert, d. h. es wird die Wahrscheinlichkeit jeder möglichen Stichproben bestimmt, $f_{\vec{X}_0} : \mathbb{R}^n \to [0, 1]$, unter der Annahme, dass die Nullhypothese zutrifft.
- 2. Die sogenannte *Irrtumswahrscheinlichkeit* $\alpha \in [0, 1]$ wird festgelegt. α gibt dabei das Risiko an, bei der Durchführung des Tests die Nullhypothese abzulehnen, obwohl sie wahr ist. (Dazu äquivalent ist die Festlegung eines *Signifikanzniveaus* 1 α , das die Sicherheit angibt, mit der die Ablehnung der Nullhypothese korrekt ist.)

3. Durchführung des Tests: Im Prinzip kann nun $f_{\vec{X}_0}(\vec{x})$ mit α verglichen werden. Unterschreitet die Wahrscheinlichkeit der beobachteten Stichprobe unter Annahme der Nullhypothese die Irrtumswahrscheinlichkeit α , ist die Nullhypothese abzulehnen, anderenfalls lässt sich lediglich sagen, dass die Stichprobe keine Handhabe für die Ablehnung bietet.

Üblicherweise wird dieser Vergleich nicht direkt für die Stichprobe, sondern für eine sogenannte *Statistik S* durchgeführt – \vec{X}_0 lässt sich oft nicht vollständig angeben. *S* ist dabei eine neue, von der Stichprobe abhängige, Zufallsvariable,

$$S: \mathbb{R}^n \to \mathbb{R},$$

für die ebenfalls eine Wahrscheinlichkeitsverteilung bestimmt wird,

$$f_S : \mathbb{R} \to [0, 1].$$

Die Beobachtung einer Stichprobe \vec{x} wird also als Ereignis aufgefasst, dem die Zufallsvariable *S* eine reelle Zahl zuordnet.

Gilt $f_S(S(\vec{x})) < \alpha$, so ist die Nullhypothese abzulehnen. Aus dieser Bedingung lässt sich eine Wertemenge für *S* ableiten, zu der $S(\vec{x})$ gehören muss, soll die Nullhypothese abgelehnt werden.

Stochastische Prozesse. Eine wichtige Erweiterung der Konzepte von Zufallsvariable und Stichprobe stellen stochastische Prozesse und Zeitreihen dar, das sind geordnete Reihen von Zufallsvariablen und deren Realisierungen. Die Ordnung wird dabei durch einen Index erzeugt, der die Zeitpunkte der Realisierungen angibt. Ich verwende die übliche Notation X_i für die Zufallsvariable einer Beobachtung zu Zeitpunkt *i* aus einem solchen Prozess und x_i für eine Realisierung dieser Zufallsvariablen. Im allgemeinen kann die Zufallsvariable des Zeitpunkts *i*, X_i , von den Zufallsvariablen aller anderen Zeitpunkte, $X_{i'}$, $i' \neq i$, abhängen. Ein Beispiel hierfür sind tägliche Temperaturbeoachtungen an einer meteorologischen Station: Der Wert des nächsten Tages wird von denen der vorangegangenen abhängen (Sommertage sind in der Regel warm), exakt vorhersagen lässt er sich jedoch nicht aus ihnen – das verhindern die zufälligen Fluktuationen im Wettergeschehen.

Bezeichne *I* die Liste der Zeitpunkte zu den Beobachtungen in einer Zeitreihe, so kann eine Zeitreihe als eine Liste von Paaren (i, x(i)) aufgefasst werden, wobei $x : I \to \mathbb{R}$ einem Zeitpunkt $i \leftarrow I$ eine reellwertige Beobachtung $x(i) \in \mathbb{R}$ zuweist. Eine Zeitreihe hat damit allgemein die Form $[(i, x(i)) | i \leftarrow I], x : I \to \mathbb{R}$. Für ein einzelnes Paar aus dieser Liste ist die abkürzende Schreibweise x_i üblich, für die ganze Reihe verwende ich die Listenschreibweise $[x_i]$ (*I* ist in dieser Notation nicht näher spezifiziert).

Anhang B

Statistische Tests

Dieser Abschnitt rekapituliert einige Standard-Verfahren der statistischen Datenanalyse, die für diese Arbeit verwendet wurden. Eine ausführliche Darstellung geben gängige Lehrbücher wie z. B. von Storch und Zwiers (1999), Schönwiese (2000) oder Taubenheim (1969).

Viele der vorgestellten Tests setzen "i.i.d" Daten voraus (independent and identically **d**istributed): Jedes Elemente der untersuchten Stichprobe ist die Realisierung einer Zufallsvariable, die von den Zufallsvariablen aller anderen Elemente unabhängig ist (independent). Alle Zufallsvariablen haben ein und die selbe Verteilung (identically distributed).

Insbesondere die erste Eigenschaft dieser Voraussetzung ist bei meteorologischen Zeitreihen, die als Realisierung eines stochastischen Prozesses aufgefasst werden können, in der Regel nicht gegeben. Die Beobachtung eines bestimmten Tages wird vielmehr ganz bestimmt von den Beobachtungen der vorangegangenen Tage abhängen, z. B. ist es unwahrscheinlich, dass einem warmen Sommertag unmittelbar ein Frosteinbruch folgt. Bei der Anwendung der Tests muss diesem Umstand sorgfältig Rechnung getragen werden.

Viele Tests sind nur für bestimmte Familien von Verteilungen formuliert (wie die Normalverteiltheit beim *t*-Test, s.u.). Ob die beobachtete Stichprobe mit der jeweiligen Familie verträglich ist, muss im Einzelfall geprüft werden.

In der Statistik-Programmiersprache R (R Development Core Team, 2004) sind die meisten der hier verwendeten Werkzeuge bereits implementiert oder sehr leicht umzusetzen.

B.1 Der *t*-Test für den Vergleich zweier Mittelwerte

Häufig stellt sich die Frage, ob die Verschiedenheit der Mittelwerte zweier Stichproben noch als zufällig angesehen werden kann oder die Mittelwerte tatsächlich signifikant verschieden sind. Das für diese Fragestellung etablierte Testverfahren ist der *t*-Test (von Storch und Zwiers, 1999; Taubenheim, 1969), der allerdings Normalverteiltheit und i.i.d. Stichproben voraussetzt.

Mit diesem Test kann die Verschiedenheit eines aus einer Stichprobe \vec{x} vom Umfang *n* geschätzten Mittelwerts $m(\vec{x}) = 1/n \sum_{i=1}^{n} x_i$ zum Mittelwert μ_0 einer hypothetischen Normalverteilung oder die Verschiedenheit der Mittelwerte zweier Stichproben geprüft werden. Die Nullhypothese, die der Test prüft, besagt, dass die Stichprobe aus einer Grundgesamtheit mit Mittelwert μ_0 stammt bzw. dass die beiden Stichproben aus Grundgesamtheiten mit gleichem Mittelwert stammen. Der Test liefert eine Wahrscheinlichkeit dafür, dass diese Annahme zutrifft. Erweist sich die Wahrscheinlichkeit für die beobachtete(n) Stichprobe(n) bezogen auf eine gewählte Irrtumswahrscheinlichkeit als zu klein, muss die Nullhypothese abgelehnt werden. Im anderen Fall lässt sich lediglich sagen, dass gegen die Nullhypothese nichts einzuwenden ist.

Im ersten Fall, wenn außer μ_0 auch die Standardabweichung σ_0 der normalverteilten Grundgesamtheit bekannt ist, kann einfach die Testgröße

$$u = \frac{m(\vec{x}) - \mu_0}{\sigma_0} \sqrt{n}$$

verwendet werden. Sie folgt aufgrund des zentralen Grenzwertsatzes einer Standardnormalverteilung mit Mittelwert 0 und Standardabweichung 1. Somit genügt es, die Quantile einer interessierenden Irrtumswahrscheinlichkeit der Standardnormalverteilung zu bestimmen und zu prüfen, ob |u| deren Betrag überschreitet. In diesem Fall wäre die Verschiedenheit des Stichprobenmittels zum hypothetischen Mittel mit der gewählten Irrtumswahrscheinlichkeit gegeben.

Etwas komplizierter wird es, wenn die Standardabweichung der hypothetischen Grundgesamtheit nicht bekannt ist und aus der Stichprobe geschätzt werden muss. In diesem Fall wird die Testgröße

$$t = \frac{m(\vec{x}) - \mu_0}{s(\vec{x})} \sqrt{n}$$

mit der aus der Stichprobe geschätzten Standardabweichung

$$s(\vec{x}) = \sqrt{1/(n-1)\sum_{i=1}^{n} (x_i - m(\vec{x}))^2}$$

gebildet. Diese Testgröße folgt der sogenannten *t*-Verteilung. Damit kann der Test wie im obigen Fall durchgeführt werden, nur dass hier anstelle der Quantile der Normalverteilung die Quantile der *t*-Verteilung zu bestimmen sind.

Die *t*-Verteilung ist nur von der Anzahl der Freiheitsgrade bei der Bestimmung von *t* abhängig. Besteht die Stichprobe aus i.i.d. Daten, so ist diese Anzahl der Freiheitsgrade n - 1 (ein Freiheitsgrad geht durch die Bestimmung des Mittelwerts $m(\vec{x})$ "verloren"). Genau an diesem Punkt wird es möglich, den Test so zu modifizieren, dass er einer eventuellen Abhängigkeit der Elemente der Stichprobe untereinander Rechnung trägt – wie es für meteorologische Zeitreihen nötig ist. Die Idee ist hierbei, die Testgröße *t* wie im i.i.d. Fall zu bestimmen, allerdings beim Vergleich mit den Quantilen der *t*-Verteilung eine geringere Anzahl von Freiheitsgraden zugrunde zu legen. Dazu wird der sogenannte effektive Stichprobenumfang n^{eff} bestimmt, der angibt, wie viele wirklich voneinander unabhängige Elemente die Stichprobe enthält.

Bei einer Zeitreihe $[x_i]$ bietet sich für die Bemessung der reiheninternen Abhängigkeit die Autokorrelationsfunktion an, für die ein Schätzer bereits bei den Gleichungen zur Dekorrelationszeit (3.1) eingeführt wurde, siehe Gleichung (3.2). Der effektive Stichprobenumfang ergibt sich in diesem Fall zu (von Storch und Zwiers, 1999)

$$n^{\text{eff}} = \frac{n}{1 + 2\sum_{l=1}^{n-1} \frac{n-l}{n} \operatorname{acf}_{[x_i]}(l)}.$$

Für die Temperaturreihen ist der Unterschied zwischen n und n^{eff} erheblich: von den rund 9000 Werten der Reihen des Kreuzvalidierungsexperiments (siehe Kapitel 4) erweisen sich nur etwa 100 als unabhängig. Dem entspricht eine Verbreiterung der *t*-Verteilung gegenüber der i.i.d.-Annahme. Die statistische Trennung einer Mittelwertschätzung aus autokorrelierten Daten ist also schwieriger als für unkorrelierte Daten bei gleichem Stichprobenumfang.

Soll der *t*-Test dazu verwendet werden, die geschätzten Mittel zweier Stichproben \vec{x}_1 und \vec{x}_2 zu vergleichen, etwa die einer beobachteten und die einer simulierten Reihe, so müssen die Standardabweichungen beider Reihen in den Test eingehen. Die erweiterte Testgröße ist dann

$$t = \frac{|m(\vec{x}_1) - m(\vec{x}_2)|}{s_\Delta} \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

wobei

$$s_{\Delta}^{2} = \frac{(n_{1} - 1)s(\vec{x}_{1})^{2} + (n_{2} - 1)s(\vec{x}_{2})^{2}}{n_{1} + n_{2} - 2}$$

Sie folgt einer *t*-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden bzw. im Fall autokorrelierter Daten mit $n_1^{\text{eff}} + n_2^{\text{eff}} - 2$ Freiheitsgraden.

B.2 Tests für lineare Trends

Soll geklärt werden, ob in einer Zeitreihe ein signifikanter linearer Trend sichtbar ist, trifft man auf eine Schwierigkeit, die die Anwendung des üblichen *t*-Tests für Regressionskoeffizienten verhindert. Dieser prüft, ob der Regressionskoeffizient, der die lineare Abhängigkeit einer abhängigen von einer erklärenden Variablen beschreibt, signifikant von 0 verschieden ist. Er setzt allerdings die Normalverteiltheit beider Variablen voraus. Für Trends in Zeitreihen, bei denen die erklärende Variable der Zeitindex ist, ist diese Voraussetzung nicht erfüllt – der Zeitindex ist gleichverteilt.

Ein linearer Trend in einer Zeitreihe kann als eine lineare Korrelation der Werte der Zeitreihe mit den Zeitindizes aufgefasst werden. Ein verteilungsunabhängiges Maß der Korrelation ist Spearman's rangbasiertes ρ , das im Prinzip dem normalen Pearsonschen Korrelationskoeffizienten entspricht, nur dass anstelle der eigentlichen Werte in der Stichprobe deren Ränge verwendet werden. Um $\hat{\rho}$ (den Schätzwert von ρ) zu beurteilen, wird es mit Fisher's *z*-Transformation "normalisiert",

$$z = \frac{1}{2}\log\frac{1+\hat{\rho}}{1-\hat{\rho}},$$

d. h. dass *z* auch für kleinere Stichprobenumfänge in sehr guter Näherung einer Normalverteilung folgt. Deren Mittelwert ist dabei $\mu_z = 1/2\log((1+\rho)/(1-\rho))$, und die Streuung $\sigma_z^2 = 1/(n-3)$, die lediglich vom Stichprobenumfang abhängt.

 $z' = (z - \mu_z)/\sigma_z$ folgt damit einer Standardnormalverteilung. Die Nullhypothese, $\rho = 0$ und damit $\mu_z = 0$, muss also abgelehnt werden, wenn die Standardnormalverteilung an der Stelle $z' = z\sqrt{n-3}$ eine gewählte Irrtumswahrscheinlichkeit unterschreitet.

Nach von Storch und Zwiers (1999) ist dieser Test weniger anfällig für Autokorrelation in den Reihen als etwa der *t*-Test für den Mittelwert.

Möchte man die Trends in zwei verschiedenen Zeitreihen miteinander vergleichen, kann man nutzen, dass die Differenz zweier normalverteilter Zufallsvariablen wiederum normalverteilt ist. Die *z*-transformierten Spearmanschen Korrelationskoeffizienten der beiden Zeitreihen, z_1 und z_2 , werden daher zur Bildung der Größe

$$u = \frac{z_1 - z_2}{\sigma_u}$$

verwendet, wobei

$$\sigma_u^2 = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}$$
Annahme der Nullhypothese, dass die eind eine Standerdnermelwerteilung V

ist. u befolgt unter Annahme der Nullhypothese, dass die Trends in beiden Zeitreihen identisch sind, eine Standardnormalverteilung. Die Nullhypothese muss also abgelehnt werden, wenn die Standardnormalverteilung an der Stelle u eine gewählte Irrtumswahrscheinlichkeit α unterschreitet.

B.3 Der Kolmogorov-Smirnov-Test für den Vergleich zweier empirischer Verteilungen

Um die Nullhypothese, dass zwei Stichproben der gleichen Grundgesamtheit entstammen, zu untersuchen, wird üblicherweise der Kolmogorov-SmirnovTest (KS-Test) verwendet. Er hat den großen Vorteil, dass keinerlei Annahmen über die Art der zugrundeliegenden Verteiltheit der Grundgesamtheit getroffen werden müssen. Auch die z. B. beim χ^2 -Test, der für diese Fragestellung ebenfalls häufig verwendet wird, notwendige Einteilung der Stichprobenelemente in Klassen ist nicht erforderlich.

Die Testgröße ist der maximale Abstand, den die beiden empirischen kummulativen Wahrscheinlichkeitsfunktionen,

$$\hat{F}_{X_i}(x) = \frac{\#(x'_i < x)}{n_i}, \ i \leftarrow [1, 2]$$

zueinander haben, wobei $\#(x'_i < x)$ die Anzahl der Elemente aus Stichprobe *i* angibt, die kleiner als *x* sind und n_i den jeweiligen Stichprobenumfang bezeichnet. Das Dach auf dem *F* weist darauf hin, dass es sich um einen Schätzer der wahren kummulativen Wahrscheinlichkeitsfunktion handelt. Die Testgröße *D* ist dann

$$D = \max_{x} |\hat{F}_{X_1}(x) - \hat{F}_{X_2}(x)|.$$

Je verschiedener die empirischen kummulativen Verteilungen, desto größer wird *D* ausfallen.

Es lässt sich zeigen, dass unter der Nullhypothese die Wahrscheinlichkeit, einen bestimmten Wert D' der Prüfgröße zu beobachten,

$$p(D' > D) = 2e^{-2nD^2}$$

beträgt. *n* berechnet sich dabei aus den beiden Stichprobenumfängen zu $n = n_1 n_2 / (n_1 + n_2)$. Damit kann die Nullhypothese nach Festlegen einer Irrtumswahrscheinlichkeit α abgelehnt werden, wenn *D* die für α maximal erlaubte Differenz

$$D_{\rm sig} = \sqrt{\frac{\ln(\alpha/2)}{2n}}$$

überschreitet.

B.4 Der KS-Test für den Vergleich zweier Spektren

Interessiert die Autokorrelationsstruktur einer Zeitreihe, bieten sich Autokorrelationsfunktion und Spektrum aus einer Fourier-Zerlegung für deren Beschreibung an. In der Tat sind das zwei ineinander überführbare und damit gleichwertige Beschreibungsarten. Um Aussagen darüber treffen, ob die aus zwei Zeitreihen geschätzten Spektren Prozesse mit gleichem Spektrum beschreiben, benötigt man zunächst einen Schätzer für das Spektrum. Der einfachste (wenn auch im allgemeinen nicht der beste) ist das Periodogramm, das sich aus den Koeffizienten der Fourier-Zerlegung einer Zeitreihe $[x_i]$,

$$x_i = a_0 + \sum_{j=1}^q \left(a_j \cos(2\pi\omega_j t) + b_j \sin(2\pi\omega_j t) \right),$$

ergibt, wobei $q = \lceil n/2$ die abgerundete ganze Zahl der halben Zeitreihenlänge n/2 bezeichnet und $\omega_j = j/n$, j = 1...q die Fourier-Frequenzen darstellen. Die Amplituden des Periodogramms für die Frequenzen sind

$$\operatorname{per}_{[x_i]}(\omega_j) = \frac{n}{4}(a_j^2 + b_j^2), \ j = -\lceil \frac{n-1}{2} \dots \lceil \frac{n}{2}$$
(B.1)

und geben die Varianz in der Zeitreihe an, die auf die Frequenz ω_j entfällt.

Die auf [0, 1] normierten kummulativen Summen cs $_{j'}$ der Amplituden,

$$cs_{j'} = \sum_{i=\lceil (n-1)/2}^{j'} I_i / \sum_j I_j, \ j' = \lceil \frac{n-1}{2} \dots \lceil \frac{n}{2} \rceil$$

können durch

$$\hat{F}_{CS}(x) = \begin{cases} 1: & x > \lceil n/2 \\ cs_j, \ j = \operatorname{argmin}_{j'} j' - x \ge 0: & x \in (\lceil \frac{n-1}{2}, \lceil \frac{n}{2}) \\ 0: & x < \lceil (n-1)/2 \end{cases}$$

als empirische kummulative Wahrscheinlichkeitsfunktion interpretiert werden. Damit lässt sich der KS-Test für den Vergleich zweier Periodogramme anwenden.

B.5 Der Wilcoxon- oder Mann-Whitney-Test für den parameterfreien Vergleich der Mediane zweier Stichproben

Der Wilcoxon- oder Mann-Whitney-Test verwendet eine rangbasierte Testgröße. Hierzu werden die beiden Stichproben \vec{x} und \vec{y} zu einer gemeinsamen zusammengefasst,

$$\vec{z} = (x_1, x_2, \dots, x_{n_x}, y_1, y_2, \dots, y_{n_y}).$$

Die Testgröße S besteht aus der Summe der Ränge R_i , die die Elemente von \vec{x} in der gemeinsamen Stichprobe innehaben,

$$S = \sum_{i=1}^{n_x} R_i.$$

Kombinatorische Argumente liefern die Verteilung von *S* unter der Nullhypothese, dass die Stichproben zu Grundgesamtheiten mit gleichem Median gehören. Für eine gewählte Irrtumswahrscheinlichkeit α kann diese Nullhypothese abgelehnt werden, wenn

$$S < \kappa_{(1-\alpha)/2}$$
 oder $S > S_{\max} - \kappa_{(1-\alpha)/2}$,

wobei

$$\kappa_{(1-\alpha)} = \frac{n_1(n_1+n_2+1)}{2} - Z_{\alpha} \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}$$

mit Z_{α} dem α -Quantil der Standardnormalverteilung.

Anhang C

Clusteranalyse

Die Clusteranalyse dient der Klassifizierung von Beobachtungen eines Datensatzes *D*, bei der einander ähnliche Beobachtungen der selben Klasse zugeordnet werden. Die Beobachtungen werden üblicherweise durch ein *m*-Tupel numerischer Variablen charakterisiert, in meinem Kontext z. B. durch die Temperatur-, Luftdruck- und Niederschlagsmessung eines Tages. Die Clusteranalyse spezifiziert dann eine Funktion

class: $\mathbb{R}^m \to \{1, 2, \dots, n_c\},\$

die jedem Element des \mathbb{R}^m eine von n_c Klassen zuweist. Identifiziert die Clusteranalyse in einem Datensatz n_c Klassen ähnlicher Beobachtungstupel, so können die einzelnen Klassen durch den Schwerpunkt ihrer jeweiligen Beobachtungen repräsentiert werden,

$$cm_i = \frac{1}{n_i} \sum_{d \in D_i} d, \ D_i = \{d \mid d \in D, \ class(d) = i\}$$
 (C.1)

wobei $n_i = |D_i|$ die Anzahl der Beobachtungen in Klasse $i, i \leftarrow [1..n_c]$ bezeichnet. Dadurch wird eine Generalisierung des Datensatzes möglich: Seine wesentlichen Eigenschaften lassen sich durch einen reduzierten Satz von Beobachtungen darstellen, indem eine prototypische Beobachtung für jede Klasse bestimmt wird. Außerdem können zwei Beobachtungen d_1 und d_2 als ähnlich identifiziert werden, wenn class $(d_1) =$ class (d_2) , sonst als unähnlich. Dabei können auch Beobachtungen d', die nicht in D enthalten sind, einer Klasse zugeordnet werden:

class(
$$d'$$
) = argmin dist^e(d' , cm_i), $d' \notin D$.
 $i \leftarrow [1..n_c]$

Neben der Clusteranalyse dienen verschiedene Bauarten neuronaler Netze (siehe z. B. Bauer et al., 1999) und fuzzy-rule-Ansätze (z. B. in Bardossy et al., 1995) einer solchen Generalisierung. Auch für nichtnumerische Variablen – ordinale oder kategorische – existieren Clusterverfahren, die jedoch für diesen Kontext nicht relevant sind.

Dieser Abschnitt wird zunächst die grundlegenden Strategien der Clusteranalyse vorstellen und auf die damit verbundenen Schwierigkeiten und Probleme in der Anwendung eingehen – eine umfassende Darstellung findet sich z. B. in Steinhausen und Langer (1977). Anschließend wird das für diese Fragestellung ausgewählte Verfahren beschrieben.

C.1 Verfahren der Clusteranalyse

Für die Clusteranalyse gibt es zwei grundsätzlich verschiedene Ansätze: die hierarchischen und die nicht-hierarchischen Verfahren. Anhand des Beispieldatensatzes mit 20 Punkten in Abbildung C.1 sollen sie erläutert und veranschaulicht werden.

C.1.1 Hierarchische Verfahren

Ein wichtiges Merkmal der hierarchischen Verfahren ist, dass sie nicht mit den Beobachtungen selbst, sondern mit deren Ähnlichkeiten bzw. Distanzen untereinander arbeiten. Die Kenntnis der Nachbarschaftsbeziehungen zwischen den Beobachtungen reicht daher für ihre Anwendung aus.

Zu Beginn wird jeder Beobachtung eine eigene Klasse zugewiesen. Der erste Schritt fasst die beiden Beobachtungen, die einander am ähnlichsten zueinander sind, zu einem Cluster zusammen. Jeder weitere Schritt fasst nun die zwei ähnlichsten der Cluster zusammen, die sich aus dem vorigen Schritt ergeben haben, wobei je nach Fragestellung verschiedene Maße existieren, die die Ähnlichkeit bzw. Distanz zwischen zwei Clustern definieren, siehe Abschnitt C.2.

Abbildung C.1 veranschaulicht solch ein agglomeratives Verfahren mit einem Dendrogramm, an dessen Ende sich alle Beobachtungen in einem Cluster befinden.

Bei der Auswahl einer geeigneten Clusteranzahl kann ein Dendrogramm wie in Abbildung C.1 Anhaltspunkte geben – der große Sprung zwischen der Einteilung in ein bzw. zwei Cluster deutet auf zwei Cluster in dem Datensatz hin, was den visuellen Eindruck in Abbildung C.1 bestätigt.

C.1.2 Nicht-hierarchische Verfahren

Im Gegensatz zu den hierarchischen Verfahren arbeiten die nichthierarchischen direkt mit den Beobachtungen. Sie sortieren einen Satz von n Beobachtungen in eine in der Regel vorgegebene Anzahl von Clustern n_c so ein, dass ein bestimmtes Kriterium erfüllt wird. Üblicherweise minimiert



Abbildung C.1: Links: Zweidimensionaler Datensatz mit 20 Punkten zur Illustration der Clusteranalyse. Rechts: Dendrogramm einer hierarchischen Clusteranalyse für den Beispieldatensatz. Die senkrechte Achse bemisst dabei die Abstände zwischen den einzelnen Clustern.

man dabei die Summe aller clusterinternen quadratischen Abweichungen,

$$G = \sum_{i=1}^{n_c} \sum_{d \in D_i} (d - \mathrm{cm}_i)^2, \ D_i = \{d \mid d \in D, \ \mathrm{class}(d) = i\}, \ i \leftarrow [1..n_c].$$
(C.2)

Praktisch bedeutet dies, dass für eine gegebene Einteilung die Schwerpunkte cm_i , $i \leftarrow [1..n_c]$ berechnet werden und dann für jede Beobachtung geprüft wird, ob ein Wechsel in eine andere Klasse *G* verringert. In diesem Fall wechselt die Beobachtung die Klasse und die betroffenen Schwerpunkte werden neu berechnet. Findet kein Austausch mehr statt, gilt die erreichte Einteilung als optimal im Sinne von *G*. Dieses Verfahren ist unter den Namen *Minimaldistanzverfahren* oder *k-means*-Algorithmus (zum Beispiel Hartigan und Wong, 1979) bekannt. Für ein solches iteratives Umsortieren im Sinne einer Zielfunktionsminimierung gibt es viele Spielarten. Beispielsweise kann anstatt des Mittels der Median zur Bestimmung der Schwerpunkte verwendet werden, oder die Zielfunktion selbst kann je nach Fragestellung von der hier vorgestellten abweichen.

Die Starteinteilung kann dabei als ein Satz von Schwerpunkten vorgegeben werden, denen dann die jeweils nächsten Punkte z.B. über die euklidische Distanz zugeordnet werden. Alternativ können die Startschwerpunkte zufällig



Abbildung C.2: Die ersten Schritte des k-means-Verfahrens, ausgehend von 3 zufällig aus dem Datensatz gezogenen Startschwerpunkten. Die schwarz umrandeten Punkte geben jeweils die Schwerpunkte an.

aus den Datenpunkten gezogen werden. Erste Iterationsschritte sind in Abbildung C.2 illustriert.

Diese Verfahren garantieren nicht, ein globales Extremum der Zielfunktion zu erreichen – die Abhängigkeit von Starteinteilung und der Reihenfolge der Austausche können den Algorithmus in ein lediglich sekundäres Extremum konvergieren lassen. Siehe die Abschnitte C.2.3 und C.2.4 für diese Abhängigkeiten.

Für die Bestimmung einer geeigneten Clusteranzahl kann *G*, ausgewertet für verschiedene Clusteranzahlen, Anhaltspunkte geben. In Abbildung C.3 ist *G* in Abhängigkeit von der Clusteranzahl für den gezeigten Datensatz dargestellt. Ein deutlicher Knick für die Clusteranzahl 2 zeigt an, dass einerseits eine größere Clusteranzahl unwesentlich mehr Information bringt, andererseits durch eine kleinere Clusteranzahl eine erhebliche Menge an Information verloren geht. Dies entspricht dem visuellen Eindruck vom Plot des Datensatzes.

C.2 Bemerkungen zu den Verfahren

C.2.1 Wahl des Distanzmaßes in der hierarchischen Clusteranalyse

Wie erwähnt gehen nicht die Beobachtungen selbst, sondern deren Distanzen zueinander in die hierarchische Analyse ein. Für die Berechnung der Distanzma-



Abbildung C.3: Zielfunktion G für verschiedene Clusteranzahlen (rechts) für den Beispieldatensatzes (links). Der Knick liefert einen deutlichen Hinweis auf die optimale Clusteranzahl.

trix wird häufig die euklidischen Norm verwendet, üblich sind jedoch ebenso ihr Quadrat oder die Maximums-Norm.

Nicht trivial ist die Wahl eines Distanzmaßes, das den Abstand zwischen zwei Klassen bemisst, die mehr als eine Beobachtung enthalten. Wurden im ersten Schritt die beiden Beobachtungen j und k mit dem kleinsten Abstand zusammengefasst, so kann der Abstand dieses mit zwei Beobachtungen belegten Clusters zu einer Beobachtung i nach

$$d_{i,jk} = \min(d_{ij}, d_{ik}) \tag{C.3}$$

berechnet werden. Dies entspricht der kleinsten Distanz aller Paare mit Beobachtungen aus jeweils einem der beiden Cluster ("single linkage"). Ebenso kann die maximale Distanz verwendet werden ("complete linkage").

Bei einem späteren Schritt treten dann entsprechend die Cluster j und k an die Stelle der einzelnen Beobachtungen des ersten Schritts.

Ein anderes häufig benutztes Maß ist die Minimal-Varianz-Distanz:

$$d_{i,jk} = \frac{(n_i + n_j)d_{ij} + (n_i + n_k)d_{ik} - n_i d_{jk}}{n_i + n_j + n_k}.$$
 (C.4)

Sie begünstigt die Bildung hypersphärischer Cluster ähnlicher Größe, für die



Abbildung C.4: Abhängigkeit der Ergebnisse einer hierarchischen Clusterung vom verwendeten Ähnlichkeitsmaß.

der Schwerpunkt bereits sehr viel repräsentativer ist als beispielsweise beim "single-linkage"-Verfahren.

Die erzeugten Einteilungen hängen stark von der Wahl des verwendeten Distanzmaßes ab, wie Abbildung C.4 zeigt.

C.2.2 Repräsentation der Klassen bei hierarchischen Verfahren

In der Regel soll durch die Klassifizierung eine Generalisierung der Beobachtungen erreicht werden, wobei sich der Schwerpunkt eines Clusters als dessen Prototyp anbietet. Die Fähigkeit der Prototypen, ihre Klasse eindeutig zu repräsentieren, ist für mein Verfahren absolut notwendig. Die hierarchischen Verfahren werden diesem Anspruch nicht gerecht, da unter Umständen zwei Cluster zusammengefasst werden können, zu deren gemeinsamem Schwerpunkt einzelne Beobachtungen dieser Cluster weiter entfernt sind als zum Schwerpunkt eines dritten Clusters – die hierarchische Analyse arbeitet mit Ähnlichkeiten zwischen Beobachtungen und nicht mit den Beobachtungen selbst, Schwerpunkte werden bei ihr nie direkt verwendet.



Abbildung C.5: Unterschiedliche Klasseneinteilung bei einer nicht-hierarchischen Clusterung mit vorgegebenem N = 3. Die Unterschiede resultieren aus den verschiedenen Startschwerpunkten.

C.2.3 Abhängigkeit von der Starteinteilung bei nicht-hierarchischen Verfahren

Nicht-hierarchische Verfahren mit einer fest vorgegebenen Clusteranzahl n_c optimieren eine gegebene Startklassifizierung im Sinne von G. Das Ergebnis dieser Optimierung kann von der Starteinteilung abhängen, wie Abbildung C.5 zeigt. Im Sinne einer objektiven Klassifizierung ist diese Abhängigkeit problematisch.

Hierarchische Verfahren haben dieses Problem nicht.

C.2.4 Abhängigkeit von der Anordnung des Datensatzes

Auch bei gleicher Starteinteilung können sich unterschiedliche Klassifizierungen ergeben, wenn die Beobachtungen in unterschiedlichen Anordnungen der Analyse unterzogen wird. Dies resultiert aus der Pfadabhängigkeit des Austauschverfahrens, bei dem nach jedem Austausch die Schwerpunkte aktualisiert werden. Abbildung C.6 zeigt ein Beispiel für diesen Effekt, bei dem die Startschwerpunkte zufällig gezogen wurden, für beide Clusterungen aber die selben waren.

Dieses Problem tritt bei den hierarchischen Verfahren nicht auf, da durch die Verwendung der Distanzmatrix die Reihenfolge der Beobachtungen im Daten-



Abbildung C.6: Unterschiedliche Klasseneinteilung bei einer nicht-hierarchischen Clusterung mit gleichen Startschwerpunkten (zufällig aus den Punkten gezogen, schwarz umrandet). Die Unterschiede resultieren aus den verschiedenen Anordnungen des Datensatzes.

satz keine Rolle spielt.

Wie die letzten beiden Abschnitte zeigen, sollte die nicht-hierarchische Clusteranalyse nicht ohne weiteres benutzt werden, da die Ergebnisse von Starteinteilung und Anordnung des Datensatzes abhängen. Die hierarchischen Verfahren haben hingegen den Nachteil, dass die Repräsentation der erzeugten Klassen durch den Schwerpunkt uneindeutig sein kann. Raffiniertere Distanzmaße wie die Mahalanobis-Distanz (Steinhausen und Langer, 1977) bei der Zuordnung neuer Beobachtungen zu den Klassen können diesen Nachteil teilweise auffangen, jedoch nicht grundsätzlich aus der Welt räumen.

Das Verfahren, das hier zur Anwendung kommt, besteht daher aus einer Kombination hierarchischer und nicht-hierarchischer Methoden.

C.3 Eine Kombination der beiden Ansätze

Für die Klassifizierung wird hier eine hierarchisch-nicht-hierarchische Methode benutzt: Den Anfang macht das hierarchische Verfahren, das dazu verwendet wird, eine Starteinteilung für den nicht-hierarchischen Teil zu finden. Die Unabhängigkeit von Starteinteilung und und Reihenfolge der Beobachtungen bei den hierarchischen Verfahren wird so kombiniert mit der eindeutigen Repräsentierbarkeit der Klassen durch ihren Schwerpunkt.

Für beide Teile werden Methoden genutzt, die in der Numerical Algorithms Group–Library für die Programmiersprache C implementiert sind (NAG, 2005). Für die Distanzmatrix wird die euklidische Norm verwendet.

Bei der iterativen Zusammenfassung zu Clustern wird die oben beschriebene Minimal-Varianz-Distanz benutzt, die zu Clustern führt, die eine hypersphärenähnliche Gestalt haben – dies bedeutet insbesondere, dass die Repräsentanzfähigkeit eines solchen Clusters durch seinen Schwerpunkt bereits relativ gut ist. Diese Klassifizierung geht als Starteinteilung in die nichthierarchische Analyse ein. Wesentliche Änderungen finden jetzt nicht mehr statt, jedoch wird hierbei sichergestellt, dass jede Beobachtung sich in dem Cluster befindet, zu dessen Schwerpunkt sie am nächsten ist.

Anhang D

STL – Zerlegung von Zeitreihen in Saisonale- und Trendkomponente mit Loess

Häufig lässt sich eine geophysikalische Zeitreihe $[x_i]$ auffassen als Realisierung eines stochastischen Prozesses $[R_i]$ (das Rauschen), dem eine saisonale Komponente $[S_i]$ (in der Meteorologie z. B. der Jahresgang) und ein langfristiger Trend $[T_i]$ aufgeprägt sind. Der hinter der Zeitreihe sich verbergende Prozess lässt sich dann für Zeitschritt *i* darstellen als

$$X_i = T_i + S_i + R_i. \tag{D.1}$$

Es gibt verschiedene Arten, eine Zeitreihe gemäß Gleichung D.1 zu zerlegen. Der Trendanteil kann beispielsweise durch das Anpassen eines Polynoms an die Reihe oder durch Filterung mit gleitenden Mitteln approximiert werden. Der Jahresgang wird im Spektrum einer Fourier-Zerlegung der Reihe sichtbar, eine übliche Technik, ihn zu bestimmen, ist die Mittelung über alle Elemente gleichen julianischen Datums und, falls notwendig, einer Glättung der Reihe dieser Mittel.

Ein Ansatz, der alle drei Komponenten auf einen Streich liefert, ist das in Cleveland et al. (1990) beschriebene STL-Verfahren (*Seasonal Trend Decomposition Procedure based on Loess*), das in R (R Development Core Team, 2004) implementiert ist und eine für den Anwender sehr einfache Zerlegung ermöglicht, siehe Abbildung D.1. Er findet u. a. in hydrologischen Fragestellungen häufige Anwendung, siehe z. B. Elek und Màrkus (2004); Montanari et al. (1997). Das Verfahren ist unparametrisch, d. h. es müssen keine Annahmen, etwa über die Familie des Prozesses oder die genaue Form der saisonalen Komponente getroffen werden. Sechs Parameter müssen vom Anwender eingestellt werden, jedoch existieren für fünf davon Standardeinstellungen, die sich entweder eindeutig aus der



Abbildung D.1: Eine STL-Zerlegung der Potsdamer Temperaturreihe 1951–2003 in Trend, Jahresgang und Rauschen. Von oben nach unten: die originale Temperaturreihe, der Jahresgang *S*, die Trendkomponente *T* und das Rauschen *R*.

Problemstellung ableiten lassen oder unabhängig von der untersuchten Reihe gesetzt werden können. Unter diesen fünf ist beispielsweise die Länge der saisonalen Periode, bei täglichen Daten mit Jahresgang 365. Der sechste Parameter muss tatsächlich vom Anwender gesetzt werden und regelt den Anteil an Variabilität, die durch die saisonale bzw. die Rauschkomponente aufgenommen werden soll.

Das Verfahren besteht aus zwei Schleifen, einer inneren und einer äußeren, wobei die äußere die Funktion hat, die Bestimmung von t_i und s_i robust zu machen, wenn Ausreißer in der Reihe deren Anpassung destabilisieren. Bei jedem Durchlauf der äußeren Schleife werden Gewichte für die einzelnen Elemente der Zeitreihe so angepasst, dass der Einfluss der Ausreißer verringert wird. Für meinen Fall: Da die Qualität des Datensatzes für das Elbe-Einzugsgebiet sehr hoch ist, rechne ich für meine Untersuchungen nicht mit Ausreißern, zumindest mit keinen, die in die Anpassung der drei Komponenten t_i , s_i und r_i nicht mit eingehen sollten. Die äußere Schleife schalte ich daher aus.

Die innere Schleife zerlegt durch eine iterative Verknüpfung verschiedener Glättungsoperatoren die Reihe in die drei Komponenten $[t_i]$, $[s_i]$ und $[r_i]$. Zum Einsatz kommen dabei gleitende Mittel und loess-Operatoren (**lo**cally weighted
regr**ess**ion), die lokal an eine Zeitreihe Polynome (für STL bis zum 2. Grad) anpassen (Cleveland und Devlin, 1988). Dadurch entsteht eine geglättete Reihe, die auch (interpolierte) Information zwischen den einzelnen Messungen der Reihe enthält.

Wichtig für die innere Schleife sind die Teilreihen, die aus den Elementen bestehen, die zu einer bestimmten Position in der saisonalen Periode gehören, also in meinem Fall die Teilreihe bestehend aus allen 1. Januaren, die Teilreihe aus allen 2. Januaren usw.. Schaltjahre wurden dabei auf 365 Tage verkürzt, um in allen Jahren einheitliche Perioden zu haben.

Die innere Schleife besteht aus den folgenden Schritten:

- 1. Von der Zeitreihe wird der Trend abgezogen, $[x_i] [t_i]$. Im ersten Iterationsschritt ist der Trend dabei auf 0 gesetzt, d. h. die unveränderte Zeitreihe wird an den nächsten Schritt übergeben.
- 2. Glätten der oben erwähnten Teilreihen: Jede Teilreihe wird mit einem loess-Operator geglättet, dessen Fensterbreite vom Anwender vorgegeben werden muss. Die geglätteten Teilreihen werden wieder zusammengesetzt und ergeben zusammen ein Signal, dass aus der saisonalen Komponente besteht, eventuell überlagert von noch langskaligeren Schwankungen. Der Rauschanteil der Originalreihe ist in diesem Signal jedoch nicht mehr vorhanden.

Die Einstellung der Fensterbreite für die Glättung der Teilreihen legt fest, wie viel von der Variabilität in der Zeitreihe der saisonalen bzw. der Rauschkomponente zugeschrieben wird – je größer, desto glatter die entstehende saisonale Komponente und desto mehr kleinskaligere Variabilität verbleibt im Rauschanteil. Diese Einstellung ist die einzige, für die keine Standardempfehlung existiert. Da ich mich entschieden habe, den Jahresgang als ein periodisches Signal anzusehen, das sich innerhalb der maximal 53 Jahre nicht ändert, habe ich dieses Fenster gewissermaßen auf unendlich gesetzt, d. h. dass alle Teilreihen einfach gemittelt werden. Die saisonale Komponente ist damit für alle Jahre identisch und weist keine langskaligeren Fluktuationen mehr auf.

- 3. Tiefpass-Filter des bisherigen Signals: Durch zwei gleitende Mittel und einen loess-Filter werden die langfristigen Schwankungen jenseits der saisonalen Periode eliminiert. Was jetzt noch übrig bleibt, ist tatsächlich die saisonale Komponente $[s_i]$. Mit meiner Art der Glättung im vorangegangenen Schritt ist dieser Tiefpass natürlich überflüssig.
- 4. Bestimmung der Trendkomponente durch Entfernen der saisonalen Komponente aus der Zeitreihe, $[x_i] - [s_i]$, und anschließendes Glätten mit einem loess-Filter, für dessen Fensterbreite die Autoren eine Standardemp-

STL

fehlung in Abhängigkeit von der Periodenlänge und dem Glättungsfenster in Schritt 2 geben.

Diese Trendkomponente geht dann in den ersten Schritt der nächsten Iteration ein. In der Regel genügen zwei Iterationen, bis sich die extrahierten Komponenten nicht mehr ändern.

Zusammenfassend noch einmal die sechs einzustellenden Parameter und meine Einstellungen:

- 1. Anzahl der äußeren Schleifen: 0, da ich nicht mit verfälschenden Ausreißern rechne.
- 2. Anzahl der inneren Schleifen: 2, Empfehlung der Autoren, sichert bereits die Konvergenz.
- 3. Länge der Periode: 365, ergibt sich aus der Natur.
- 4. Fensterlänge des loess-Operators für die Glättung der Teilreihen: ∞, da ich einen einheitlichen Jahresgang für die interessierenden Zeiträume erwarte.
- 5. Fensterlänge für den loess-Operator bei der langskaligen Glättung der saisonalen Komponente: erübrigt sich, siehe 4.
- 6. Fensterlänge für den loess-Operator, der die Trendkomponente glättet: 1.5 * 365, Empfehlung der Autoren, gestützt auf eine Frequenzanalyse der verwendeten Filter.

Insgesamt bietet das Verfahren eine sehr kompakte Art der Zerlegung einer Zeitreihe, die außer der Periodenlänge keine weiteren Annahmen benötigt (abgesehen von der Entscheidung über die Verteilung der Variabilität in saisonaler und Rauschkomponente). Die Implementierung in R macht es außerdem sehr komfortabel zu bedienen.

Anhang E

Der Datensatz des Elbe-Einzugsgebiets

Für das Elbe-Einzugsgebiet stehen tägliche Daten von 1951–2003 von 148 Klimaund 683 Niederschlagsstationen zur Verfügung. Das Messtupel eines Tages bestand hierbei aus: Tagesmaximum der Lufttemperatur, Tagesmittel der Lufttemperatur, Tagesminimum der Lufttemperatur, der Niederschlagssumme, der mittleren relativen Feuchte, dem mittleren Luft- und Wasserdampfdruck, der Sonnenscheindauer, dem mittleren Bedeckungsgrad, der Globalstrahlung und der mittleren Windgeschwindigkeit. Für die Niederschlagsstationen wurden die fehlenden Messgrößen von den Klimastationen aus interpoliert.

E.1 Probleme mit meteorologischen Daten

Grundsätzlich steht man bei der Interpretation meteorologischer Beobachtungsdaten vor dem Problem zu entscheiden, ob ein beobachteter Effekt in den Daten tatsächlich aus dem untersuchten System stammt oder ob es sich um einen Artefakt handelt. So beobachtet man beispielsweise bei Stationen, die sich in oder nahe größeren Städten befinden, einen verhältnismäßig starken Anstieg der Tagesminima der Lufttemperatur während des 20. Jahrhunderts, der sich zum Teil auf die allgemeine Erwärmung zurückführen lässt, außerdem jedoch auf das Wachstum der Städte in dieser Zeit, die durch ihre eigene Wärmeabgabe besonders die Temperaturminima anheben. Für lange meteorologische Messreihen können diese verschiedenen Quellen für Trends in den Daten oft nur schwer voneinander getrennt werden.

Außer allmählichen Einflussentwicklungen wie dem Städtewachstum können sich die Messbedingungen selbst abrupt ändern. Zu solchen Änderungen gehören beispielsweise eine Stationsverlegung, der Austausch von Messgeräten oder eine Änderung der Bebauung in der unmittelbaren Umgebung der Station. Oft finden sich anlässlich solcher Veränderungen Sprünge in den Reihen. Wenn die Geschichte einer Station verfügbar ist und eine Inhomogenität in den Reihen mit einer Veränderung der Messbedingungen zusammenfällt, können die Reihen behutsam korrigiert werden. Schwieriger ist es, wenn die Stationsgeschichten nicht verfügbar sind oder das Datenmaterial so umfangreich ist, dass eine Einzelfallprüfung nicht durchführbar ist.

Immer wieder fehlen in den Reihen längere Zeitabschnitte, beispielsweise wegen zeitweise aufgegebenen Stationen, Kriegen oder ausgebrannten Archiven. Das Füllen solcher Lücken kann sehr schwierig und unter Umständen unmöglich sein, besonders wenn Lücken an vielen Stationen gleichzeitig existieren.

Auch bei der Übertragung alter Daten auf digitale Datenträger können Fehler auftreten: einfache Schreibfehler beim digitalisieren der alten Messblätter oder technische Unzulänglichkeiten, beispielsweise beim Einlesen alter Lochkarten.

Nicht zuletzt sind natürlich auch die Beobachter nicht unfehlbar, so dass mit einfachen Ablese- und Schreibfelern ebenfalls zu rechnen ist.

Techniken, die solche Inhomogenitäten erkennen und teilweise korrigieren können, werden deshalb kontinuierlich weiterentwickelt. Siehe Peterson et al. (1998) für eine Übersicht und Feng et al. (2004) für eine Anwendung auf Stationsdaten aus China. Der hier verwendete Datensatz wurde von Österle am PIK, wie in den folgenden Abschnitten beschrieben, aufbereitet (Österle et al., 2006).

E.2 Fehlererkennung und Vervollständigung von meteorologischen Daten

Die Vervollständigung einer Messreihe besteht aus zwei Phasen: erstens der Detektion und Eliminierung von fehlerhaften Messungen, zweitens der Ergänzung der Lücken.

E.2.1 Fehlerhafte Messungen

Der erste Schritt der Fehlererkennung besteht darin, zu prüfen, ob die Daten ihren möglichen Wertebereich überschreiten. Beispielsweise können Niederschlagssummen nicht negativ werden und der Bedeckungsgrad bewegt sich zwischen 0/8 und 8/8. Für Größen wie den Luftdruck oder die Temperatur müssen hingegen die Wertebereiche subjektiv festgelegt werden, in denen der Jahresgang berücksichtigt wird. Diese hängen dann natürlicherweise von der Lage der betrachteten Station ab. Für die Sonnenscheindauer lässt sich mit empirischen Formeln eine Obergrenze in Abhängigkeit von Datum im Jahr, geografischer Breite und Berglage angeben, deren Überschreitung eine Fehlmessung anzeigen würde. Der zweite Schritt besteht aus einer Konsistenzprüfung: Wenn beispielsweise für das Tagesmaximum der Lufttemperatur ein geringerer Wert registriert wurde als für das Minimum oder das Mittel, handelt es sich offensichtlich um eine Fehlmessung. Analog wurde mit Bedeckungsgrad und Sonnenscheindauer verfahren.

Drittens begegnet man bei Daten, die ursprünglich auf Lochkarten aufgezeichnet und danach auf magnetische Datenträger übertragen wurden, häufig über längere Zeiträume unveränderten Werten, z. B. 1000 hPa für den Luftdruck oder 0 für Bedeckungsgrad, Windgeschwindigkeit und Sonnenscheindauer. Diese können mit einer in einem gleitenden Fenster der Länge 15 Tage berechneten Streuung identifiziert werden.

Zuletzt wird die räumliche Konsistenz der Daten untersucht. Das setzt ein hinreichend dichtes Stationsnetz voraus. Nachdem die Daten jahresgangbereinigt wurden, wird jede Messung an einer Station auf ihre Verträglichkeit mit den zeitgleichen Messungen an den unmittelbar benachbarten Stationen geprüft. Dazu werden zum einen mit den Korrelationskoeffizienten zwischen den Reihen der zu prüfenden und den Nachbarstationen gewichtete Mittel von Messungen eines Tages berechnet, zum anderen lineare Regressionsgleichungen für die Zusammenhänge zwischen den Daten der zu prüfenden Station und ihren Nachbarstationen aufgestellt. Eine Messung der zu prüfenden Station wird als fehlerhaft markiert, wenn die Differenzen zwischen der Messung und dem gewichteten Mittel, den aus der Regression erhaltenen Werten oder auch direkt mit den Messungen der Nachbarstationen gewisse Schwellenwerte überschreiten. Diese Schwellenwerte werden z. T. konstant, z. T. in Abhängigkeit von den Streuungen in den Daten oder den Residuen der Regression definiert.

Anschließend werden die Lücken in den Reihen, seien sie durch fehlende oder wie oben beschrieben als fehlerhaft erkannte Daten bedingt, geschlossen soweit möglich.

E.2.2 Datenergänzung

Für die Ergänzung der Daten einer Station wird diejenige Nachbarstation benutzt, deren Daten am höchsten mit denen der betroffenen Station korrelieren. Der entsprechende Wert wird dabei um die langjährige mittlere Monatsmitteldifferenz korrigiert und eingesetzt. Für Messgrößen, deren Wertebereich beschränkt ist, wurde statt der Differenz der Quotient der langjährigen Monatsmittel benutzt, mit dem der Messwert der bestkorrelierten Nachbarstation dann multipliziert wurde. Dabei wurde durch zusätzliche Koeffizienten verhindert, dass der Wertebereich verlassen wurde (etwa beim Bedeckungsgrad).

E.3 Homogenitätsanalyse und Homogenisierung

Der Vervollständigung der Reihen schließt sich die Homogenitätsanalyse und Homogenisierung an. Dabei geht es um die Detektion und Eliminierung von nichtnatürlichen Einflüssen auf die Messdaten, um in den Daten möglichst nur das reine natürliche Signal eine eventuellen Klimawandels zu erhalten. Zu nichtnatürlichen Einflüssen zählen beispielsweise Stationsverlegungen, bauliche Veränderungen im Umfeld der Station, Änderungen der Messtermine etc..

Die von Österle durchgeführte Homogenitätsanalyse verwendete zwei Ansätze: zum einen wurde, wenn verfügbar, anhand der Stationsgeschichten untersucht, ob eine dokumentierte Veränderung der Messbedingungen zu Sprüngen in der Reihe führte. Dazu wurden die beiden jahresgangbereinigten Teilreihen von vor und nach der Veränderung mittels des *t*-Tests auf miteinander verträgliche Mittelwerte geprüft. Zeigte dieser Test eine Inhomogenität an, so wurden die Daten auf das Niveau nach der Veränderung korrigiert. Hierbei wurde die Entwicklung in den Daten einer benachbarten Station, die diese Inhomogenität nicht aufweisen, mit einbezogen.

Zum anderen wurden jeweils für eine zu prüfende und ihre Nachbarstationen gleitende *t*-Tests für zwei Teilreihen, die jeweils aus 48 Monatsmitteln bestanden, durchgeführt. Die sich so ergebende Reihe von *t*-Werten zeigt den Zeitpunkt von auftretenden Inhomogenitäten durch Extrema an. Treten diese sowohl bei der zu prüfenden als auch gleichzeitig bei der Mehrheit der Nachbarstationen auf, so wird die Inhomogenität als klimatologisch anerkannt, andernfalls wie oben beschrieben korrigiert.

Insgesamt steht nach diesen Analysen und Korrekturen ein homogener, vollständiger Datensatz mit einer selten hohen Stationsdichte zur Verfügung, den man sich sowohl für Klimaanalysen als auch als Grundlage für Simulationen optimaler kaum wünschen kann.

Literaturverzeichnis

- Bardossy, A., Duckstein, L., und Bogardi, I. (1995). Fuzzy Rule Based Classification of Atmospheric Circulation Patterns. *International Journal of Climatology*, 15:1087–1097.
- Bauer, H.-U., Herrmann, M., , und Villmann, T. (1999). Neural maps and topographic vector quantization. *Neural Networks*, 12:659–676.
- Beersma, J. und Buishand, T. (2003). Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. *Climate Research*, 25:121–133.
- Benestad, R. (2004). Tentative probabilistic temperature scenarios for northern Europe. *Tellus*, 56A:89–101.
- Blender, R. und Fraedrich, K. (2003). Long time memory in global warming simulations. *Geophysical Research Letters*, 30(14):1769–1772.
- Böhm, U. (2006). Annual mean temperatures for the Elbe catchment according to ECHAM4-IS92-A simulations and according to ECHAM5-SRES-A2 simulations. Personal Communication at Potsdam Institute for Climate Impact Research, Germany.
- Böhm, U., Kücken, M., Ahrens, W., Block, A., Hauffe, D., Keuler, K., Rockel, B., und Will, A. (2006). CLM – the climate version of LM: Brief description and long-term applications. *COSMO Newsletter*, (6).
- Cleveland, R., Cleveland, W., McRae, J., und Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1):3–33.
- Cleveland, W. und Devlin, S. (1988). Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83:596–610.
- Déqué, M. (2004). Uncertainties in PRUDENCE simulations: Global high resolution models. Technical report, DMI.

- Diez, E., Primo, C., Garcia-Moya, J., Gutierrez, J., und Orfila, B. (2005). Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus*, 57(A):409–423.
- Doets, K. und van Eijck, J. (2004). *The Haskell Road to Logic, Math and Programming*. King's College Publications, Cambridge, UK.
- DVWK (1996). Ermittlung der Verdunstung von Land- und Wasserflächen. DVWK Merkblätter zur Wasserwirtschaft 238, DVWK.
- Elek, P. und Màrkus, L. (2004). A long range dependent model with nonlinear innovations for simulating daily river flows. *Natural Hazards and Earth System Sciences*, 4:277–283.
- Enke, W. und Spekat, A. (1997). Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, 8:195–207.
- Feng, S., Hu, Q., und Qian, W. (2004). Quality control of daily meteorological data in China, 1951–2000: A new dataset. *International Journal of Climatology*, 24:853–870.
- Fischer, H. und Kaul, H. (1990). *Mathematik für Physiker 1*. BG Teubner, Stuttgart.
- Fox-Rabinovitz, M., Côté, J., Dugas, B., Déqué, M., und McGregor, J. (2006). Variable resolution general circulation models: Streched-grid model intercomparison project (SGMIP). *Journal of Geophysical Research*, 111:1–21.
- Fraedrich, K. und Blender, R. (2003). Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models. *Physical Review Letters*, 90(10):108501(1–4).
- Fraedrich, K., Raible, C., und Sielmann, F. (2003). Analog Ensemble Forcasts of Tropical Cyclone Tracks in the Australian Region. *Weather and Forecasting*, 18:3–11.
- Frei, C., Christensen, J., Déqué, M., Jacob, D., Jones, R., und Vidale, P. (2003). Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *Journal of Geophysical Research*, 108(D3).
- Gerstengarbe, F.-W. (2002). Angewandte Statistik PIK-Weiterbildungsseminar 2000/2001. PIK Report 75, Potsdam Institute for Climate Impact Research.
- Gerstengarbe, F.-W. und Werner, P. (2005). Katalog der Großwetterlagen Europas (1881–2004) nach Paul Hess und Helmut Brezowsky, 6. verbesserte und

ergänzte Auflage. PIK Report 100, Potsdam Institute for Climate Impact Research.

- Giorgi, F., Bi, X., und Pal, J. (2004). Mean, interannual variability and trends in a regional climate change experiment over Europe. I. Present-day climate (1961–1990). *Climate Dynamics*, 22:733–756.
- Giorgi, F., Marinucci, M., und Bates, G. (1993). Development of a second generation regional climate model (RegCM2). Part I. Boundary-layer and radiative transfer processes. *Monthly Weather Review*, 121(2):794–813.
- Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J., Johns, T., Mitchell, J., und Wood, R. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16:147–168.
- Grell, G., Dudhia, J., und Stauffer, D. (1995). A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). Technical Report 398, NCAR.
- Hartigan, J. A. und Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistic*, 28:100–108.
- Hess, P. und Brezowsky, H. (1977). Katalog der Großwetterlagen Europas 1881– 1976, 3. verbesserte und ergänzte Auflage. Bericht des DWD 15 (113), Deutscher Wetterdienst.
- Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der Linden, P., Dai, X., Maskell, K., und Johnson, C. (2001). *Climate Change 2001, The Scientific Bases. Contribution of Working Group I to the Third Assessment Report of the IPCC.* Cambridge University Press, Cambridge.
- Jacob, D. (2001). A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorol. Atmos. Phys.*, 77:61–73.
- Jacob, D. (2005). REMO CLIMATE OF THE 20th CENTURY RUN, UBA PRO-JECT, 0.088 DEGREE RESOLUTION, RUN NO. 006210, 1H DATA. CERA-DB "REMO_UBA_C20_1_R006210_1H". http://cera-www.dkrz.de/WDCC/ui/ Compact.jsp?acronym={REMO_UBA_C20_1_R006210_1H}.
- Kidson, J. und Thompson, C. (1998). Comparison of statistical and model-based downscaling techniques for estimating local climate variations. *Journal of Climate*, 11:735–753.

- Kotlarski, S., Block, A., Böhm, U., Jacob, D., Keuler, K., Knoche, R., Rechid, D., und Walter, A. (2005). Regional Climate Model Simulations as Input for Hydrological Applications: Evaluation of Uncertainties. *Advances in Geosciences*, 5:119–125.
- Lorenz, P. und Jacob, D. (2005). Influence of regional scale information on the global circulation: A two-way nesting climate simulation. *Geophysical Research Letters*, 32(L18706).
- Mitchell, J. und Ewins, P. (2003). The Scientific and Technical Review 2002/3. Technical report, Met Office, Exceter, UK.
- Montanari, A., Rosso, R., und Taqqu, M. (1997). Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation and simulation. *Water Resources Research*, 33(5):1035–1044.
- Murphy, J. (1999). An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate. *Journal of Climate*, 12:2256–2284.
- Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M., und Stainforth, D. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430:768–772.
- NAG (2005). *NAG C Library*. The Numerical Algorithms Group Limited, Vienna, Austria. ISBN 978-1-85206-206-4.
- Österle, H. (2005). Annual mean temperatures of Germany according to ECHAM4 present climate simulation. Personal Communication at Potsdam Institute for Climate Impact Research, Germany.
- Österle, H., Gerstengarbe, F.-W., und Werner, P. (2006). Ein neuer meteorologischer Datensatz für Deutschland, 1951–2003. http://www.meteo.physik. uni-muenchen.de/dkt/abstracts/klimaparameter/oesterle_p.pdf.
- Peterson, T., Easterling, D., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Foerland, E., Hanssen-Bauer, I., Alexandersson, H., Jones, P., und Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18(13):1493–1517.
- R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

- Richter, D. (1995). Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Meßfehlers des Hellmann-Niederschlagsmessers. *DWD*, 194.
- Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Manzini, L. K. E., Rhodin, A., Schlese, U., Schulzweida, U., und Tompkins, A. (2003). The atmospheric general circulation model ECHAM 5. PART I: Model description. *MPI-Report*, 349.
- Ruddiman, W. (2001). Earth's climate: past and future. Freeman, New York, USA.
- Ruiz-Barradas, A. und Nigam, S. (2006). IPCC's Twentieth-Century Climate Simulations: Varied Representations of North American Hydroclimate Varaibility. *Journal of Climate*, 19(16):4041–4058.
- Schönwiese, C. (2003). Klimatologie. UTB, Stuttgart, 2nd edition.
- Schönwiese, C. D. (2000). *Praktische Statistik für Meteorologen und Geowissenschaftler*. Gebr. Borntraeger, Berlin, 3rd edition.
- Semenov, M. und Barrow, E. M. (1997). Use of a stochastic Weather Generator in the Development of Climate Change Scenarios. *Climatic Change*, 35:397–414.
- Steinhausen, D. und Langer, K. (1977). *Clusteranalyse Einführung in Methoden und Verfahren der automatischen Klassifikation*. Walter de Gruyter, Berlin.
- Stock, M., Gerstengarbe, F., Kartschall, T., und Werner, P. (2005). Reliability of climate change impact assessments for viticulture. *Acta Hort. (ISHS)*, 689:29– 40.
- Taubenheim, J. (1969). *Statistische Auswertung geophysikalischer und meteorologischer Daten*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig.
- Thompson, S. (1999). *The Craft of Functional Programming*. Addison-Wesley, Boston, MA, 2nd. edition edition.
- Timbal, B. und McAvaney, B. (2001). An analogue-based method to downscale surface air temperature: Application for Australia. *Climate Dynamics*, 17:947–963.
- van den Dool, H. (1989). A New Look at Weather Forecasting through Analogues. Monthly Weather Review, 117:2230–2247.
- von Storch, H., Hewitson, B., und Mearns, L. (2000). Review of Empirical Downscaling Techniques. In Iversen, T. und Hoiskar, B., editors, *Regional climate development under global warming*, number 4 in General Technical Report, pages 29–46, Jevnaker, Torbjornrud, Norway. RegClim.

- von Storch, H. und Zwiers, F. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, UK.
- Wechsung, F., Becker, A., und Gräfe, P. (2005). GLOWA-Elbe I: Integrierte Analyse der Auswirkungen des globalen Wandels auf Wasser, Umwelt und Gesellschaft im Elbegebiet. PIK Report 95, Potsdam Institute for Climate Impact Research.
- Werner, P. und Gerstengarbe, F.-W. (1997). Proposal for the development of climate scenarios. *Climate Research*, 8(3):171–182.
- Wessel, P. und Smith, W. (2006). *The Generic Mapping Tools GMT*. University of Hawai'i at Manoa, NOAA.
- Wilby, R. und Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, 21:530–548.
- Wilby, R., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., und Wilks, D. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, 34:2995–3008.
- Wild, M. (2005). ECHAM5 T106L31 control run of CSCS/ETH, monthly mean values. CERA-DB "ECHAM5_T106L31_CTRL_CSCS_ETH_MM". http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ECHAM5_ T106L31_CTRL_CSCS_ETH_MM.
- Wilks, D. (1999). Interannual variability and extrem-value characterisitcs of several stochastic daily precipitation models. *Water Resources Research*, 34:2995–3008.
- Zorita, E. und von Storch, H. (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate*, 12:2474–2489.

Danksagungen

Ganz herzlichen Dank schulde ich meinen beiden Betreuern am PIK, den Herren Prof. F.-W. Gerstengarbe und Prof. P.C. Werner. In den knapp drei Jahren in ihrer Arbeitsgruppe erhielt ich alle Unterstützung, die ich mir wünschen konnte, Tür und Ohr standen mir jederzeit und immer konstruktiv offen. Besonders dankbar bin ich Herrn Gerstengarbe für ein wunderschönes Konzert in Potsdam.

Herrn Prof. K. Fraedrich möchte ich herzlichst dafür danken, dass er auf die Distanz Hamburg-Potsdam die Doktorvaterschaft für diese Arbeit übernommen hat – und für die viele für mich immer sehr inspirierende Zeit, die wir bei mehreren Treffen in Hamburg verbracht haben.

Die bisherige Zeit am PIK wäre nur halb so schön gewesen ohne meine Mitdoktoranden Malaak Kallache und Henning Rust, von denen ich viel über Statistik gelernt habe und deren Gesellschaft ich ungemein genieße. Außergewöhnlichen Dank an Dr. Nicola Botta, der, obwohl mit meinem Arbeitsfeld nicht verbunden, viel Zeit mit mir und meiner Arbeit verbrachte und nicht müde wurde, mir meine Probleme und deren Kommunikation zu erklären.

Herzlichen Dank den Korrekturlesern: am PIK – Henning Rust, Dr. Nicola Botta, Dr. Uwe Böhm und Martin Kücken; ehemals aus Hagen – Hagen Hagemann; an der Humboldt-Universität zu Berlin – Holger Gerhard (außerdem: Kompetenz für Satz, Layout und Duden); im Süden – mein Vater.

Unabhängig von dieser Arbeit danke ich meinen Eltern, Geschwistern und Freunden für all die schöne gemeinsame Zeit und ihr immer so unterstützendes Dasein, das ich als etwas sehr besonderes empfinde.