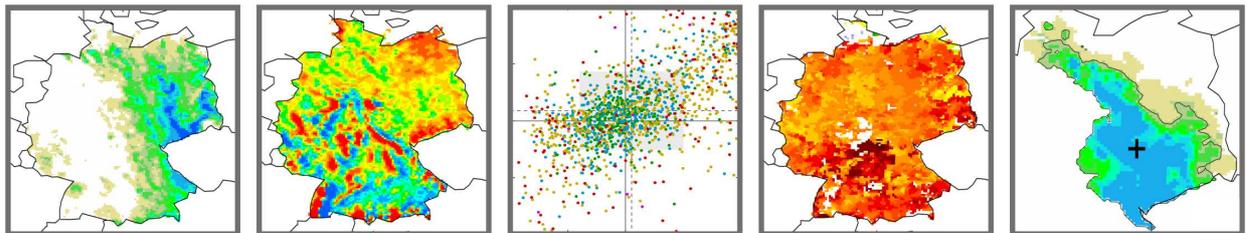


Verifikation der Niederschlagsvorhersage für Deutschland von 2001-2004

Marcus Paulat



Dissertation zur Erlangung des Grades

„Doktor der Naturwissenschaften“

Institut für Physik der Atmosphäre

Johannes Gutenberg-Universität in Mainz, Deutschland

Fachbereich Physik, Mathematik und Informatik

Mainz, den 30.November 2007

D77 - Mainzer Dissertation

Inhaltsverzeichnis

Abstract	III
Zusammenfassung	V
1 Einleitung	1
1.1 Stand der Wissenschaft	4
1.2 Ziele dieser Arbeit	10
2 Datengrundlage	13
2.1 Das COSMO-Modell	13
2.2 Das Globalmodell des ECMWF	17
2.3 Tägliche Stationsmessungen	17
2.4 Stündliche Stationsmessungen	21
2.5 Das Radarkomposit	22
2.6 Die Niederschlagsklimatologie	25
2.7 Flusseinzugsgebiete in Deutschland	28
3 Disaggregation	29
3.1 Gitterung der Stationsmessungen	29
3.2 Räumliche Mittelung des Radarkomposits	32
3.3 Technik der Disaggregation	33
3.4 Fehlerquellen der Disaggregation	35
3.5 Zusammenfassung der Disaggregation	42
4 Mittlere Niederschlagssummen	45
4.1 Jährliche Beobachtungen	45
4.2 Saisonale Beobachtungen	47
4.3 Vergleich zwischen Modellen und Beobachtungen	49
4.4 Modellvergleich zwischen aLMo und LM	54
4.5 Einfluss des prognostischen Niederschlags	57

5	Intensitäten und Tagesgänge	61
5.1	Häufigkeiten	61
5.2	Intensitäten	66
5.3	Der Tagesgang	73
6	Gitterpunkt-basierte Verifikation	79
6.1	Definition von Standardfehlermaßen	79
6.2	Auswertung der Tagessummen	82
6.3	Auswertung der Stundensummen	90
6.4	Zusammenfassung und Fazit	94
7	Entwicklung eines neuen Qualitätsmaßes	97
7.1	Definition von SAL	97
7.2	Idealisierte Beispiele	100
7.3	Anwendung auf reale Niederschlagsfelder	107
7.4	Fazit	110
8	SAL: statistische Auswertung	113
8.1	SAL für Tagessummen: Verifikation des aLMo	113
8.2	Vergleich zwischen aLMo und Globalmodell des ECMWF	122
8.3	Vergleich zwischen aLMo und LM	123
8.4	SAL für Stundensummen: Verifikation des aLMo	125
8.5	Kurzes Fazit	130
9	Zusammenfassung und Ausblick	131
9.1	Zusammenfassung	131
9.2	Ausblick	134
A	Häufigkeiten von Tagessummen	137
B	SAL: Abhängigkeiten vom Schwellwert	139
B.1	aLMo-Tagessummen	139
B.2	ECMWF-Tagessummen	140
	Zeichentabelle	141
	Literaturverzeichnis	145

Abstract

The verification of precipitation forecasts from numerical weather prediction models is essential for the future improvement of quantitative precipitation forecasts (QPF). The goal of this study is the development of new methods for the verification of precipitation forecasts and their application to the regional model of MeteoSwiss (COSMO-aLMo) and the global model of the European Centre for Medium-Range Weather Forecasts (ECMWF). To this end, a new observational dataset with 1-hour temporal resolution has been produced for Germany, and the novel quality measure „SAL“ for QPF validation has been developed and applied.

The observational dataset with high temporal and spatial resolution is based upon a disaggregation technique developed earlier within the Mesoscale Alpine Program. The idea is to combine the high temporal resolution from radar data (hourly) and the fairly high precision with respect to the amount of precipitation from rain gauge measurements (within the range of measurement errors). This dataset offers several new possibilities for QPF verification. For the first time, an area-wide investigation of the diurnal cycle of precipitation could be performed. It is shown that during winter, when there is no diurnal cycle in precipitation, COSMO-aLMo performs very well. However, during summer the diurnal cycle of precipitation of COSMO-aLMo and the disaggregated dataset differ substantially. In the model, maximum precipitation typically occurs too early at 11-14 UTC compared to 15-20 UTC in the observations and is distinctly overestimated by a factor of 1.5.

A novel quality measure SAL has been developed to address some of the shortcomings of gridpoint-based methods for the verification of the new generation of high-resolution numerical weather prediction models. SAL is based on the identification of precipitation objects (threshold dependent) within a catchment and calculates differences between model and observations with respect to the structure (S), amplitude (A) and location (L) of precipitation in the catchment. SAL has been tested with synthetic and real examples. On the one hand SAL can identify the model deficiencies that were found with standard error measures (e.g. associated with the diurnal cycle and the simulation of too many weak precipitation events). On the other hand SAL offers additional insight into a model's QPF performance and highlights model qualities and deficiencies that are particularly relevant for the prediction of intense small-scale events.

Verification with SAL is done for daily and hourly QPFs from the regional model COSMO-aLMo and the coarser-scale global ECMWF-model. For COSMO-aLMo, SAL shows that the QPF quality of intense events is better than for weak events,

which is a positive model characteristic with respect to social implications of QPF. The comparison of the two models shows that the global ECMWF-model often simulates too widespread precipitation. This is not surprising given the coarser horizontal resolution of the model. However, unlike standard error measures, SAL is able to identify this shortcoming, also in a statistical sense.

The methods developed in this study are very useful for the verification of QPFs with high temporal and spatial resolution. The use of the disaggregated observational dataset and of the quality measure SAL leads to new insights into the quality of QPFs and offers relevant future applications, in particular for the verification of the new generation of numerical weather prediction models, which explicitly simulate the life-cycle of deep convection cells.

Zusammenfassung

Die Verifikation numerischer Modelle ist für die Verbesserung der Quantitativen Niederschlagsvorhersage (QNV) unverzichtbar. Ziel der vorliegenden Arbeit ist die Entwicklung von neuen Methoden zur Verifikation der Niederschlagsvorhersagen aus dem regionalen Modell der MeteoSchweiz (COSMO-aLMo) und des Globalmodells des Europäischen Zentrums für Mittelfristvorhersage (engl.: ECMWF). Zu diesem Zweck wurde ein neuartiger Beobachtungsdatensatz für Deutschland mit stündlicher Auflösung erzeugt und angewandt. Für die Bewertung der Modellvorhersagen wurde das neue Qualitätsmaß „SAL“ entwickelt.

Der neuartige, zeitlich und räumlich hoch-aufgelöste Beobachtungsdatensatz für Deutschland wird mit der während MAP (engl.: Mesoscale Alpine Program) entwickelten Disaggregierungsmethode erstellt. Die Idee dabei ist, die zeitlich hohe Auflösung der Radardaten (stündlich) mit der Genauigkeit der Niederschlagsmenge aus Stationsmessungen (im Rahmen der Messfehler) zu kombinieren. Dieser disaggregierte Datensatz bietet neue Möglichkeiten für die quantitative Verifikation der Niederschlagsvorhersage. Erstmals wurde eine flächendeckende Analyse des Tagesgangs des Niederschlags durchgeführt. Dabei zeigte sich, dass im Winter kein Tagesgang existiert und dies vom COSMO-aLMo gut wiedergegeben wird. Im Sommer dagegen findet sich sowohl im disaggregierten Datensatz als auch im COSMO-aLMo ein deutlicher Tagesgang, wobei der maximale Niederschlag im COSMO-aLMo zu früh zwischen 11-14 UTC im Vergleich zu 15-20 UTC in den Beobachtungen einsetzt und deutlich um das 1.5-fache überschätzt wird.

Ein neues Qualitätsmaß wurde entwickelt, da herkömmliche, gitterpunkt-basierte Fehlermaße nicht mehr der Modellentwicklung Rechnung tragen. SAL besteht aus drei unabhängigen Komponenten und basiert auf der Identifikation von Niederschlagsobjekten (schwellwertabhängig) innerhalb eines Gebietes (z.B. eines Flusseinzugsgebietes). Berechnet werden Unterschiede der Niederschlagsfelder zwischen Modell und Beobachtungen hinsichtlich Struktur (S), Amplitude (A) und Ort (L) im Gebiet. SAL wurde anhand idealisierter und realer Beispiele ausführlich getestet. SAL erkennt und bestätigt bekannte Modelldefizite wie das Tagesgang-Problem oder die Simulation zu vieler relativ schwacher Niederschlagsereignisse. Es bietet zusätzlichen Einblick in die Charakteristiken der Fehler, z.B. ob es sich mehr um Fehler in der Amplitude, der Verschiebung eines Niederschlagsfeldes oder der Struktur (z.B. stratiform oder kleinskalig konvektiv) handelt.

Mit SAL wurden Tages- und Stundensummen des COSMO-aLMo und des ECMWF-Modells verifiziert. SAL zeigt im statistischen Sinne speziell für stärkere (und damit

für die Gesellschaft relevante Niederschlagsereignisse) eine im Vergleich zu schwachen Niederschlägen gute Qualität der Vorhersagen des COSMO-aLMo. Im Vergleich der beiden Modelle konnte gezeigt werden, dass im Globalmodell flächigere Niederschläge und damit größere Objekte vorhergesagt werden. Das COSMO-aLMo zeigt deutlich realistischere Niederschlagsstrukturen. Diese Tatsache ist aufgrund der Auflösung der Modelle nicht überraschend, konnte allerdings nicht mit herkömmlichen Fehlermaßen gezeigt werden.

Die im Rahmen dieser Arbeit entwickelten Methoden sind sehr nützlich für die Verifikation der QNV zeitlich und räumlich hoch-aufgelöster Modelle. Die Verwendung des disaggregierten Datensatzes aus Beobachtungen sowie SAL als Qualitätsmaß liefern neue Einblicke in die QNV und lassen angemessenere Aussagen über die Qualität von Niederschlagsvorhersagen zu. Zukünftige Anwendungsmöglichkeiten für SAL gibt es hinsichtlich der Verifikation der neuen Generation von numerischen Wettervorhersagemodellen, die den Lebenszyklus hochreichender konvektiver Zellen explizit simulieren.

Kapitel 1

Einleitung

Wasser ist eine der Grundvoraussetzungen für das Leben auf unserer Erde. Der Wasserkreislauf wird maßgeblich von der Atmosphäre bestimmt, indem diese Wasser in Form von Wasserdampf von der Erdoberfläche aufnimmt, speichert und transportiert. Durch weitere Prozesse, wie z.B. Hebung feuchter Luftmassen, kondensiert Wasserdampf zu flüssigem Wasser oder gefriert zu Eis und kann durch Niederschlag wieder zur Oberfläche gelangen. Damit regelt die Atmosphäre große Teile der Verfügbarkeit von Wasser.

Alleine aus diesem Grund ist die Vorhersage der Vorgänge in der Atmosphäre und speziell der Niederschlagsbildung von großer gesellschaftlicher Bedeutung (z.B. für Land- und Wasserwirtschaft, Verkehr, Energiewirtschaft). Dies zeigt sich noch klarer an den Personen- und wirtschaftlichen Schäden, die durch Wetterereignisse entstehen können. Die Elbeflut 2002 war dafür ein eindrucksvolles Beispiel. Ebenso kann aber auch das Ausbleiben von Niederschlägen dramatische Folgen haben, was nur ein Jahr später im sehr tockenen und heißen Sommer 2003 europaweit zu spüren war.

Die Weiterentwicklung von numerischen Wettervorhersagemodellen und Beobachtungssystemen hat in den letzten Jahrzehnten zu einer deutlichen Qualitätssteigerung der Vorhersagen einzelner meteorologischer Parameter, aber auch des Vorhersagehorizontes geführt. Dies wird anhand Abb. 1.1 für Vorhersagen des Globalmodells des ECMWF (engl.: European Centre for Medium-Range Weather Forecast) deutlich. Gezeigt ist die zeitliche Entwicklung der Anomalie-Korrelation der 500 hPa-Geopotentialfläche für operationelle 3-, 5-, und 7-Tage-Vorhersagen (aufgeteilt in Nord- und Südhemisphäre von 1980-2003). Es ist eine generelle Verbesserung dieser Vorhersagen zu erkennen. Die Anomalie-Korrelation der 3-Tage-Vorhersagen konnte für die Nordhemisphäre von 85 auf 95 % verbessert werden (5-Tage-Vorhersagen von 65 auf 85 %, 7-Tage-Vorhersage von 45 auf 70 %). Die Vorhersage für die Südhemisphäre war lange Zeit schlechter, konnte aber kontinuierlich an die Vorhersagegüte der Nordhemisphäre herangeführt werden. Abbildung 1.1 zeigt auch, dass heute eine 5-Tage-Vorhersage genauso gut ist, wie eine 3-Tage-Vorhersage vor etwa 20 Jahren (7-Tage-Vorhersagen sind heute etwas besser, als 5-Tage-Vorhersagen vor etwa 20 Jahren). Abbildung 1.2 zeigt die Zunahme der Modellprognosegüte des Deutschen Wetterdienstes (DWD) für verschiedene Parameter innerhalb der letzten 16 Jah-

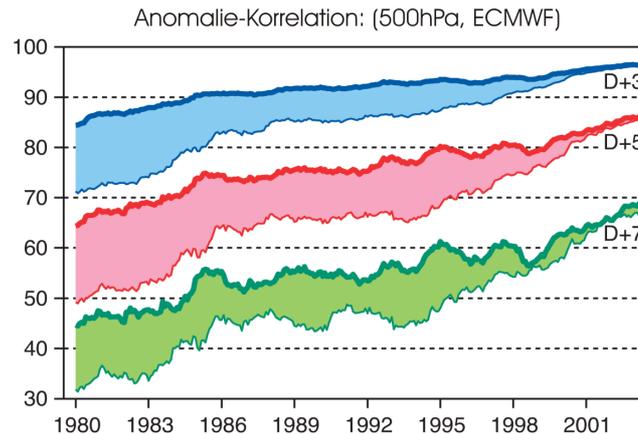


Abbildung 1.1: Zeitliche Entwicklung von Anomalie-Korrelationen der 500 hPa-Geopotentialfläche aus operationellen ECMWF-Vorhersagen für 3 Tage (Blau), 5 Tage (Rot) und 7 Tage (Grün). Die obere Kurve zeigt jeweils den Verlauf für die Nordhemisphäre, die untere für die Südhemisphäre, die Differenz daraus ist entsprechend eingefärbt. (aus ECMWF, 2004)

re (Balzer, 2002). Es zeigt sich, dass z.B. die Qualität der Temperaturvorhersage deutlich zugenommen hat (um etwa 48%). Ähnliches gilt für andere Parameter wie Windstärke (49%) oder mit Abstrichen der Wolkenbedeckungsgrad (32%). Dagegen geht die Zunahme der Prognosegüte für Windspitzen (ca. 6%) und insbesondere den Niederschlag (ja/nein: ca. 1%) gegen null. Internationale Studien, wie z.B. die Beurteilung der Niederschlagsvorhersage innerhalb der WGNE (engl.: Working Group on Numerical Experimentation), kommen zu dem Ergebnis, dass die numerischen Wettermodelle noch weit davon entfernt sind, die perfekte Vorhersage für den Niederschlag zu produzieren (Ebert et al., 2003).

Es stellt sich die Frage, warum gerade die aus gesellschaftlicher und wirtschaftlicher Sicht so wichtige, quantitative Niederschlagsvorhersage (QNV) in den letzten Jahren nicht verbessert werden konnte. Die Defizite werden derzeit in vier übergeordneten Bereichen gesehen (Hense et al., 2003):

- (a) unvollständige Modellierung der Komponenten des Wasserkreislaufs,
- (b) nicht aufgelöste Strukturen und Fehler in Anfangsfeldern (Ausgangsdaten einer Modellvorhersage),
- (c) Mängel bei der Datenassimilation und
- (d) grundsätzliche Probleme beim Verständnis und der Interpretation von Vorhersagemodellen.

Aufgrund des sehr breit gefächerten Aufgabenfeldes, das zur Verbesserung der QNV bearbeitet werden muss, konnten Anstrengungen einzelner Wissenschaftler bisher nicht den gewünschten Erfolg bringen. Deshalb wurde 2004 von der Deutschen Forschungsgemeinschaft (DFG) ein Schwerpunktprogramm (SPP) ins Leben gerufen, das die Anstrengungen an Universitätsinstituten, weiteren Forschungseinrichtungen

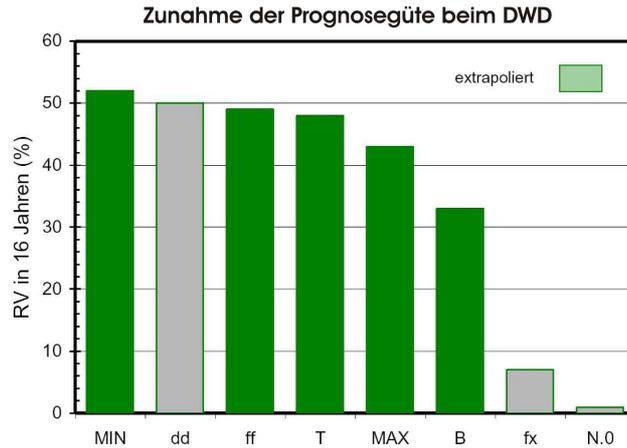


Abbildung 1.2: Zunahme der Prognosegüte (RV: Reduction of Variance) des DWD innerhalb der letzten 16 Jahre für verschiedene Modellvariablen: Tagesminimumtemperatur (MIN), Tagesmaximumtemperatur (MAX) Durchschnittstemperatur (T), Windrichtung (dd), Windstärke (ff), Wolkenbedeckungsgrad (B), Windspitzen $> 12\text{ms}^{-1}$ (fx) und Niederschlag ja/nein (N.0). (aus Balzer, 2002)

und des DWD bündeln und koordinieren soll. Dieses SPP mit dem Namen „Quantitative Niederschlagsvorhersage“ beinhaltet die Bearbeitung der folgenden drei übergeordneten Zielsetzungen (Hense et al., 2003):

- I) Identifikation der für die Defizite verantwortlichen physikalischen und chemischen Prozesse bei der QNV,
- II) Bestimmung und Ausschöpfung der Potenziale vorhandener und neuer Daten und Prozessbeschreibungen zur Verbesserung der QNV,
- III) Bestimmung der Vorhersagefähigkeit von Wettervorhersagemodellen durch statistisch-dynamische Analysen der QNV.

Innerhalb des SPP haben sich seit dem Start 2004 fünf Themengebiete herauskristallisiert, welche jeweils von mehreren Wissenschaftlern (bzw. wissenschaftlichen Instituten) mit unterschiedlichen Ansätzen bearbeitet werden (Homepage des SPP, 2007). Abbildung 1.3 zeigt schematisch die einzelnen Themengebiete: Konvektion/Topographie, Datenassimilation/Stochastik, Mikrophysik, Verifikation und die im SPP eingebetteten Messkampagnen COPS (engl.: Convective and Orographically-induced Precipitation Study; Wulfmeyer et al., 2007) und GOP (engl.: General Observation Period; Crewell et al., 2005; GOP-Homepage, 2007). Die vorliegende Arbeit ist dem Themenblock Verifikation unter dem Codenamen VERIPREG (engl.: Verification of Precipitation over Germany; Wernli et al., 2003) zugeordnet und trägt vor allem zum Erreichen der SPP-Ziele II und III bei.

Welches sind nun die Gründe, warum man sich gerade mit der Verifikationsthematik beschäftigen sollte, und wie kann dies zur Verbesserung der QNV beitragen? Die Wichtigkeit der Verifikation von Wettervorhersagen ist sehr schön in einem Artikel von Damrath (2002) erklärt. Dort wird geschrieben, dass „jeder Anwender einer

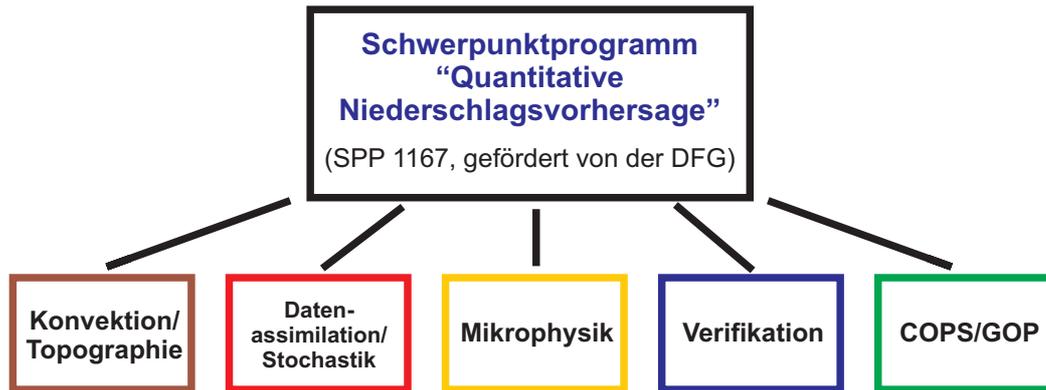


Abbildung 1.3: Schemazeichnung zu übergeordneten Themengebieten im Schwerpunktprogramm „Quantitative Niederschlagsvorhersage“ (SPP 1167, gefördert von der DFG; Homepage des SPP, 2007).

Vorhersage daran interessiert ist, die Verlässlichkeit der Vorhersage zu kennen, um darauf aufbauend eine für ihn sinnvolle Strategie der Nutzung der Vorhersage zu entwickeln“. Weiter wird gesagt, dass die Vorhersagenden und Modellierer darauf angewiesen sind, typische und systematische Modellfehler zu identifizieren, um diese entweder zu eliminieren oder durch statistische Verfahren (z.B. „Model Output Statistics“ (MOS)) im Nachhinein zu modifizieren. Dabei ist auch wichtig zu verstehen, dass gerade bei Wetterdiensten die Verantwortlichen Kenntnis darüber besitzen müssen, ob sich finanzielle Investitionen ausgezahlt haben (oder sich auszahlen werden). Die Verifikation der Wettervorhersage spielt dabei eine zentrale Rolle. Der Begriff Verifikation hat dabei seinen Ursprung im Lateinischen und leitet sich aus „verus“ (wahr) ab, womit Verifikation also der Vergleich von Wettervorhersage mit der Wahrheit bedeutet.

Im Folgenden wird zunächst der aktuelle Stand der Wissenschaft zum Thema Verifikation der QNV dargestellt, bevor dann die Ziele dieser Arbeit erläutert werden.

1.1 Stand der Wissenschaft

Gegenwärtige regionale numerische Wettervorhersagemodelle produzieren Niederschlagsvorhersagen mit hoher räumlicher (≤ 10 km) und zeitlicher Auflösung (≤ 1 h). Diese Vorhersagen zeigen oft kleinskalige Strukturen wie einzelne konvektive Zellen oder schmale Regenbänder, welche mit früheren Modellgenerationen nicht simuliert werden konnten. Mit der Implementierung operationeller, Wolken auflösender Modelle (< 3 km und ohne Konvektionsparametrisierung) geht die Tendenz dahin, immer kleinskaligere Strukturen im Niederschlagsfeld simulieren zu können.

An Vorhersagezentren wie dem DWD oder dem ECMWF werden zur Verifikation häufig kontinuierliche (z.B. RMSE, engl.: Root Mean Square Error; Simmons und Hollingsworth, 2002) oder kategorische (basierend auf einer Kontingenztabelle, z.B. POD, engl.: Probability of Detection; Wilks, 1995; Ebert, 2007) Fehlermaße verwendet, um die Vorhersagequalität aufzuzeigen. Diese Maße sind dafür geeignet, Vorhersagefelder zu verifizieren, die synoptisch-skalige Strukturen aufweisen

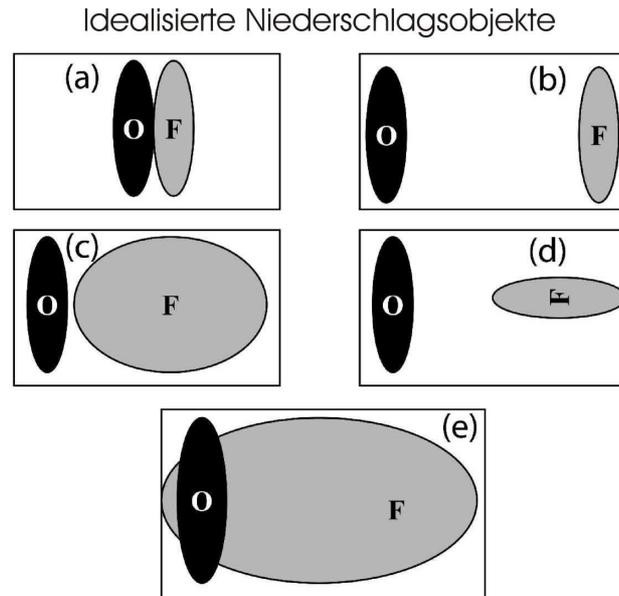


Abbildung 1.4: Idealisierte Niederschlagsfelder zur Veranschaulichung des Problems der Niederschlagsverifikation mit gitterpunkt-basierten Fehlermaßen (aus Davis et al., 2006a): F = Vorhersage (engl.: Forecast), O = Beobachtungen (engl.: Observations).

(z.B. die 500 hPa-Geopotentialfläche), allerdings weniger für Parameter, die komplexe Strukturen auf Skalen kleiner als 100 km zeigen (z.B. der Niederschlag). Solche gitterpunkt-basierten Fehlermaße werden als problematisch für diese kleinskaligen Strukturen erachtet, weshalb in den letzten Jahren einige neuartige Ansätze vorgeschlagen wurden (z.B. Ebert und McBride, 2000; Casati et al., 2004; Davis et al., 2006a,b).

Das klassische Beispiel, den eingeschränkten Nutzen von gitterpunkt-basierten Fehlermaßen aufzuzeigen, ist das sogenannte „Double-Penalty-Problem“: eine in Menge, Größe und Zeit richtige Niederschlagsvorhersage, jedoch (eventuell sogar nur leicht verschoben) am falschen Ort, wird mit obigen Fehlermaßen als relativ schlecht bewertet. In solchen Situationen würde beispielsweise die *POD* für ein lediglich verschobenes Niederschlagsfeld in der Vorhersage das gleiche (schlechte) Ergebnis wie für ein völlig verpasstes Ereignis liefern. Anhand Abb. 1.4 aus Davis et al. (2006a) wird dies sehr deutlich: die Bilder (a)-(e) zeigen jeweils ein Gebiet mit einem vorhergesagten (F) und beobachteten (O) Niederschlagsfeld. Die Bilder (a)-(d) sind so konstruiert, dass es keinerlei Überschneidung der beiden Felder gibt, so dass $POD = 0$ gilt (die Vorhersage also in allen vier Fällen maximal schlecht ist). Nur in Bild (e) gibt es eine Überschneidung, so dass die $POD > 0$ wird und gegenüber den anderen Fällen als beste Vorhersage bewertet wird. Diese Aussage entspricht allerdings nicht dem subjektiven Eindruck, dass wohl Bild (a) die beste (wenn auch nicht perfekte) Vorhersage der fünf Beispiele zeigt. Die Auswertung anderer gitterpunkt-basierter Fehlermaße (z.B. des *RMSE*) würde ebenfalls wenig aufschlussreiche Ergebnisse liefern. Anzumerken ist auch, dass diese Verifikationsergebnisse keinen Hinweis auf die Art des Fehlers geben, welche z.B. Fehler in der Struktur ((c) und (e)) oder Lage (alle fünf Bilder) des Niederschlagsfeldes sein könnten. Diese Begebenheiten bekommen umso größere Bedeutung, je höher aufgelöst numerische Wettervorhersa-

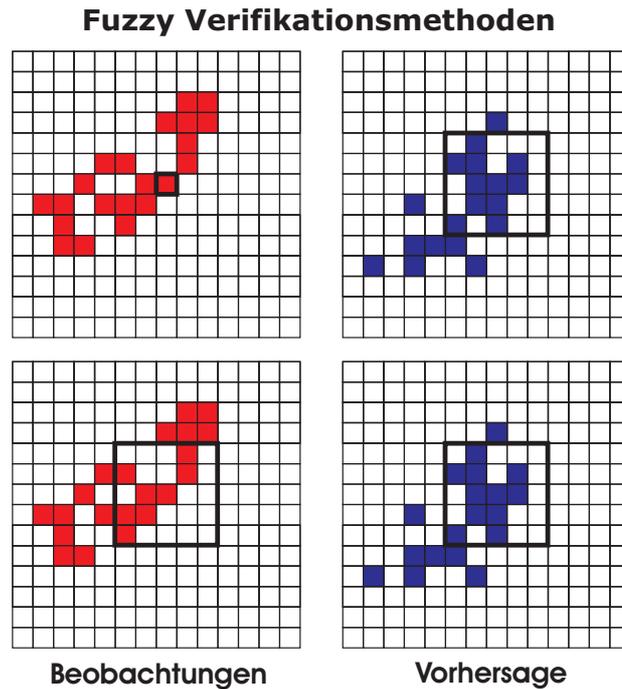


Abbildung 1.5: Idealisierte Bilder zur Beschreibung der Fuzzy Verifikationsmethoden (modifiziert aus Ebert, 2006). Die Quadrate stehen für Gitterboxen und zeigen Niederschlag an, wenn sie farbig sind (Rot: Beobachtungen, Blau: Vorhersage). Oben: Eine einzelne Gitterbox in den Beobachtungen im Vergleich zu einer Vorhersageumgebung, unten: Vergleich einer Umgebung in beiden Datensätzen (vgl. jeweils innerhalb der dick schwarz umrandeten Box).

gemodelle sind (in dieser Arbeit 7 km, seit Anfang 2007 bereits 2.8 km (z.B. operationell beim DWD)). Bei solch hohen Auflösungen liegen Niederschlagsvorhersagen vor, die in Struktur und Variabilität vergleichbar mit Radarinformationen sind. Es ist nicht endgültig erforscht, ob ein Niederschlagsfeld mit vielen kleinskaligen, konvektiven Zellen überhaupt deterministisch vorhergesagt werden kann. So könnte der bestmögliche Ansatz sein, die richtige Struktur des Niederschlags (z.B. einzelne konvektive Zellen) unabhängig der genauen Lage zu verifizieren.

Hier stellt sich die Frage nach der Grenze der Vorhersagbarkeit für den Niederschlag. Verschiedene theoretische Arbeiten besagen (z.B. Zhang et al., 2002, 2003, 2006; Walser et al., 2004; Walser und Schär, 2004; Hohenegger et al., 2006), dass die Grenzen auf Skalen unterhalb 100 km sehr schnell erreicht werden, da sich die Fehler in der Vorhersage kleinskaliger Strukturen durch nicht-lineare Skalenwechselwirkungen vervielfachen. Allerdings bedeutet das Vorhandensein konvektiver Zellen nicht automatisch, dass sie nicht vorhersagbar wären. Dies gilt ganz besonders für gebirgige Regionen (Walser und Schär, 2004). Außerdem gibt es für organisierte konvektive Systeme eine verbesserte Vorhersagbarkeit (Fritsch und Carbone, 2004). Es wird auch betont, dass die Vorhersagbarkeit stark von der Wettersituation und der Topographie abhängt.

Die oben erwähnten, neuen Ansätze zur Niederschlagsverifikation versuchen, das Double-Penalty-Problem zu vermeiden und die Grenzen der Vorhersagbarkeit mit zu berücksichtigen, indem sie nicht mehr den Niederschlag an einzelnen Gitterpunk-

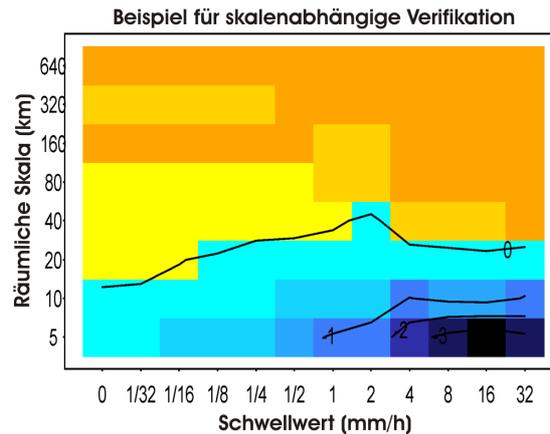


Abbildung 1.6: Ergebnisbild zur skalenabhängigen Verifikation aus Casati et al. (2004): Skill Score des mittleren quadratischen Fehlers (Farbskala: Blau bedeutet schlechter als der Zufall, Gelb keine Qualitätsaussage möglich, Orange gute bis sehr gute Qualität) in Abhängigkeit von der verwendeten Skala und der Intensität (Schwellwert).

ten vergleichen und Informationen über die Charakteristiken von Vorhersagefehlern geben. Man kann diese Verfahren in Fuzzy Verifikationsmethoden, skalenabhängige Techniken und objekt-orientierte Methoden unterteilen, wobei die Übergänge zwischen den Methoden oft fließend sind.

Unter Fuzzy Verifikationsmethoden versteht man, dass diese keine exakte Zuordnung von Vorhersage und Beobachtung benötigen, sondern eine gewisse Umgebung um einen betrachteten Gitterpunkt zulassen. Dies kann die Problematik der Vorhersagbarkeit auf kleinen Skalen, sowie die Unsicherheiten und die eventuell schlechte Repräsentativität der Stationsmessungen abschwächen. Die Umgebung wird gewählt, indem man mehrere benachbarte Gitter- und/oder Zeitpunkte um den eigentlich interessanten Punkt zur Verifikation hinzuzieht. Diese Umgebung kann dann mit kontinuierlichen und kategorischen Fehlermaßen oder aber über die Betrachtung von Wahrscheinlichkeiten verifiziert werden. Für Fuzzy Verifikationsmethoden gibt es zwei mögliche Ansätze zum Vergleich von Vorhersage mit Beobachtungen (Abb. 1.5):

1. Vergleich eines einzelnen Beobachtungswertes mit einer Umgebung aus der Vorhersage.
2. Vergleich identischer Umgebungen in beiden Datensätzen.

Ein Beispiel für Fuzzy Verifikation ist der „Fraction Skill Score“ (Roberts und Lean, 2005). Er vergleicht das Verhältnis von vorhergesagten und beobachteten Ereignissen innerhalb einer Umgebung. Dagegen vergleicht Atger (2001) nur einen Beobachtungswert mit einer Umgebung aus der Vorhersage und prüft, ob dieser Wert innerhalb der Umgebung vorkommt. Ein drittes Beispiel ist die Betrachtung der Umgebung in der Vorhersage (zeitlich und räumlich) als ein Ensemble (Theis et al., 2005). Jeder einzelne Gitterpunkt innerhalb der Umgebung repräsentiert eine deterministische Vorhersage. Betrachtet man nun einen Gitterpunkt, so zeigt das Ensemble aus der Umgebung die Wahrscheinlichkeitsdichtefunktion für dieses Ereignis an.

Beispiel für die CRA-Methode

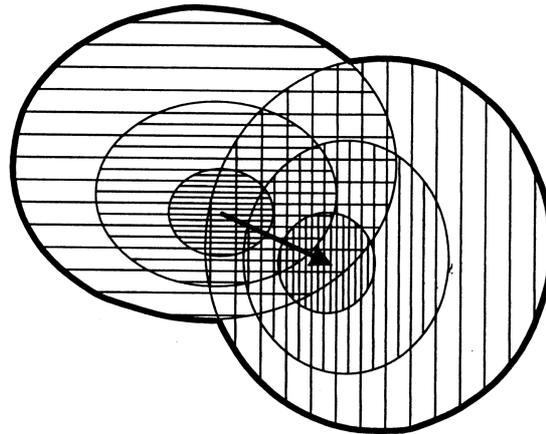


Abbildung 1.7: Idealisierte Niederschlagsobjekte für Vorhersage (längs gestreift) und Beobachtungen (quer gestreift) aus Ebert und McBride (2000). Die dicke schwarze Linie zeigt das CRA (engl.: Contiguous Rain Area). Der Pfeil zeigt den Verschiebungsfehler der Vorhersage gegenüber den Beobachtungen an.

Den Modellvorhersagen an einzelnen Gitterpunkten können demnach Wahrscheinlichkeitsaussagen für das Auftreten des Ereignisses hinzugefügt werden.

Wählt man verschieden große Umgebungen für denselben Fall, so kommt man automatisch in den Bereich der skalenabhängigen Verifikationsmethoden. Es ist damit möglich, Vergleiche zwischen verschiedenen Skalen in Abhängigkeit der Intensität (über die Wahl eines Schwellwertes) anzustellen. Abbildung 1.6 zeigt ein Beispiel einer solchen Abhängigkeit für den von Casati et al. (2004) verwendeten „Mean Squared Error Skill Score“. Casati et al. (2004) verwenden allerdings nicht eine Umgebungsmethode zur Verifikation, sondern kombinieren traditionelle Fehlermaße mit einer Aufspaltung in verschiedene Skalen durch eine Haar-Wavelet-Zerlegung binärer Bilder, welche aus den originalen Niederschlagsdaten gewonnen werden. Dies lässt eine Bewertung der Vorhersage in einem Gebiet zu, die besonders für räumlich nicht kontinuierlich verteilte Variablen wie den Niederschlag sinnvoll ist (z.B. bei konvektiver Wetterlage). Abbildung 1.6 zeigt für einen konvektiven Fall, dass die Vorhersagequalität für größere Skalen steigt. Zugleich sieht man, dass die Vorhersagequalität auf kleineren Skalen für höhere Schwellwerte (Intensitäten) abfällt. Die Aussage von Casati et al. (2004) für diesen Fall ist, dass kleine Skalen mit hohen Intensitäten (Starkniederschläge aus einzelnen Zellen) schlecht vorhergesagt wurden, dies aber hauptsächlich ein Verschiebungsfehler ist (konvektive Zellen am falschen Ort).

Eine der wichtigsten und am häufigsten zitierten Arbeiten aktueller Verifikationsmethoden ist die objekt-orientierte CRA-Verifikation von Ebert und McBride (2000). CRA (engl.: Contiguous Rain Area) beschreibt dabei ein Niederschlagsgebiet. Dieses Gebiet wird gewählt, indem ein durch die Wahl des Schwellwertes identifiziertes Objekt aus der Vorhersage so lange verschoben wird, bis es maximale Korrelation mit dem Beobachtungsobjekt erreicht (oder ein anderes Kriterium erfüllt). Diese Zuordnung von einzelnen Objekten in Modell und Beobachtungen ist eine Notwendigkeit dieser Methode, in der Praxis aber oftmals schwierig eindeutig durchführbar.

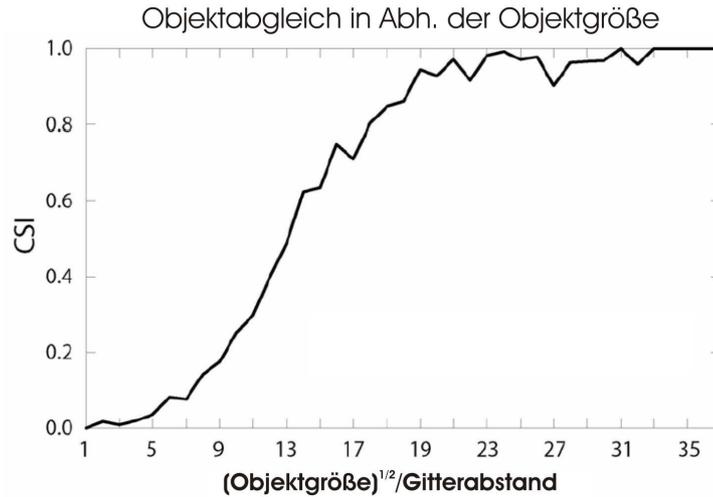


Abbildung 1.8: CSI für den Objektvergleich zwischen Beobachtungen und Vorhersage als Funktion der Objektgröße. (aus Davis, 2006a)

Abbildung 1.7 zeigt einen idealisierten Fall, indem Vorhersage und Beobachtungen bereits verschoben wurden und die schwarze dicke Linie schlussendlich das CRA definiert. Der Pfeil innerhalb des Bildes zeigt den Verschiebungsfehler an, der aus der Verschiebung des Vorhersageobjektes zu seinem ursprünglichen Ort errechnet wird. Dieser Verschiebungsfehler ist eine Komponente, die aus der Minimierung des *MSE* (engl.: Mean Squared Error) folgt. Die anderen *MSE*-Komponenten sind ein Volumen- und ein Strukturanteil:

$$MSE_{total} = MSE_{Verschiebung} + MSE_{Volumen} + MSE_{Struktur}$$

Die Struktur-Komponente des *MSE* ist dabei der übrig bleibende Anteil nach Aufspaltung in Verschiebungs- und Volumenkomponente. Weitere Vergleiche der beiden Objekte lassen Aussagen über verschiedene Objektmerkmale zu, z.B. Fläche, mittlere und maximale Intensitäten. Ein Nachteil der Methode ist, dass die Objekte mehr als 20 Gitterpunkte umfassen müssen, damit eine Aufspaltung des *MSE* möglich ist (Tartaglione et al., 2005).

Ein weiterer Ansatz zur objekt-orientierten Verifikation wurde von Davis et al. (2006a) vorgestellt. Die Objektidentifizierung ist dabei unterschiedlich zu Ebert und McBride (2000), allerdings ebenfalls basierend auf der Wahl eines Schwellwertes. Zudem wird das zu untersuchende Gebiet vorher festgelegt und nicht durch die Objektverschiebung definiert. Die Methode nach Davis et al. (2006) liefert weiterführende Untersuchungskriterien wie zum Beispiel die Orientierung der Objekte. Ein wichtiges Ergebnis ist aber, dass ein Abgleich von Objekten in Vorhersage und Beobachtungen sehr schwierig bis unmöglich wird, wenn die Objekte kleiner als 7 Gitterpunktlängen werden, was bei Davis et al. (2006a) etwa 150 km entspricht. Dies bedeutet, dass gerade konvektive Zellen, deren Vorhersage aufgrund auftretender Starkniederschläge von großer Wichtigkeit ist, nur schwierig mit dieser Methode untersucht werden können. Weitere Methoden beschäftigen sich daher damit, einzelne Zellen Stück für Stück zu Clustern zusammenzufügen, um zum einen nicht jede einzelne Zelle verifizieren zu müssen, zum anderen wiederum eine skalenabhängige Verifikation möglich

zu machen (Marzban und Sandgathe, 2006).

Vergleicht man nun Fuzzy mit objekt-orientierten Verifikationsmethoden, so stellt sich heraus, dass jede Methode Vor- und Nachteile hat, man also je nach Fragestellung entscheiden muss, welche Methode die (vermeintlich) bessere ist. Fuzzy Methoden können gut chaotisch geprägte Vorhersagefelder wie mehrere konvektive Niederschlagszellen behandeln und lassen eine Bewertung über verschiedene Skalen zu. Diese Methoden sind nur bedingt intuitiv (also nur bedingt nahe an der Wahrnehmung eines Betrachters).

Objekt-orientierte Methoden behandeln dagegen nur einzelne Objekte und lassen eine Bewertung auch nur auf dieser Skala zu. Allerdings sind sie sehr nahe an der subjektiven Einschätzung eines Betrachters und lassen verschiedene Parameter (je nach Fragestellung) zur Qualitätsprüfung zu. Der totale Fehler einer Vorhersage kann hierbei in verschiedene, der subjektiven Wahrnehmung entsprechende Fehler aufgespaltet werden (z.B. in Verschiebungs-, Volumen- und Strukturfehler bei Ebert und McBride (2000)).

1.2 Ziele dieser Arbeit

Hauptziel dieser Arbeit ist es, ein neues Qualitätsmaß für die Verifikation der QNV zu entwickeln. Dieses Qualitätsmaß wird mit „SAL“ bezeichnet und soll zusätzlich Einblick in die Charakteristiken und Fehler der QNV geben.

Ein weiteres wichtiges Ziel ist die Implementierung eines für Deutschland neuartigen, hoch-aufgelösten Niederschlagsdatensatzes. Dieser Datensatz wird durch eine während des MAP (engl.: Mesoscale Alpine Program; Binder und Schär, 1995) entwickelte Disaggregation erzeugt (Hagen et al., 2003). Dieser neue Beobachtungsdatensatz macht die Verifikation der QNV auf stündlicher Skala möglich. Die hier verwendeten Modelle (Kap. 2) können also auf täglicher und stündlicher Zeitskala mit konventionellen Fehlermaßen, als auch mit dem neu entwickelten SAL verifiziert werden.

Aus den im Abschnitt zuvor dargelegten Schwächen (bei allen Stärken) der vielfältigen neuen Ansätze zur Verifikation der QNV und den Fragen nach der Vorhersagbarkeit erscheint es nützlich, einen alternativen Ansatz für die Verifikation der Niederschlagsvorhersage zu entwickeln. Dieser soll einerseits die Struktur vorhandener Niederschlagsobjekte in Vorhersage und Beobachtungen betrachten, zum anderen aber unabhängig von einer direkten Zuordnung dieser Objekte sein. SAL soll verschiedene Charakteristiken einer Niederschlagsvorhersage (Struktur, Amplitude und Lage) in einem betrachteten Gebiet bewerten können. Dieses Qualitätsmaß beinhaltet folgende Eigenschaften, welche detailliert in Kap. 7 vorgestellt werden:

- (i) es macht Aussagen über die Qualität einzelner Niederschlagsvorhersagen in einem zuvor definierten Gebiet (z.B. ein Flusseinzugsgebiet), wobei die betrachteten, aufsummierten Niederschläge eine Dauer von 1 bis 24 Stunden haben.

- (ii) SAL ist nahe an einer subjektiven Einschätzung der Qualität von Niederschlagsvorhersagen.
- (iii) SAL setzt sich aus drei unabhängigen Komponenten zusammen (S , A und L).
- (iv) Es berücksichtigt und spezifiziert (im Gegenteil zu Ebert und McBride, 2000) die Struktur einzelner Niederschlagsereignisse (z.B. konvektive Zellen, frontale Niederschlagsbänder), welche als direktes Maß für die zu Grunde liegenden, physikalischen Eigenschaften gilt.
- (v) Es benötigt keine Zuordnung einzelner Niederschlagsobjekte zwischen Modellvorhersage und Beobachtungen.

Um die Verifikation mit SAL durchführen zu können, sind geeignete Datensätze notwendig (insbesondere für eine statistisch signifikante Zeitreihe). Betrachtet wird die QNV in dieser Arbeit für Deutschland, weshalb das COSMO-Modell (engl.: Consortium for Small-scale Modelling; Koppert et al., 2007) als hoch-aufgelöstes, lokales Modell und zusätzlich das Globalmodell des ECMWF verifiziert werden. Um auch die stündlich verfügbaren Modellvorhersagen verifizieren zu können, ist ein stündlicher Datensatz notwendig. Dieser muss zunächst erzeugt werden, da auf stündlicher Basis keine qualitativ ausreichenden Beobachtungen vorhanden sind. Die Idee der Disaggregation ist, Stationsmessungen („genaue“ Niederschlagsmenge) mit Radardaten (hohe zeitliche Variabilität) so zu kombinieren, dass ein zeitlich und räumlich hoch-aufgelöster Datensatz aus Beobachtungen erstellt wird. Details zur Technik und zu Fehlerquellen der Disaggregation sind in Kap. 3 ausführlich beschrieben. Mit den verfügbaren Datensätzen ist die eigentliche Verifikation nun detailliert möglich. Dabei wurden Vorhersagen und Beobachtungen zunächst jährlich und saisonal gemittelt, um einen ersten Eindruck über die Niederschlagsverteilung zu bekommen (Kap. 4). Dem schließen sich Auswertungen zu Häufigkeiten und Intensitäten an und führen schließlich zur Verifikation des Tagesgangs im Niederschlag (Kap. 5). Die Verifikation mit konventionellen Fehlermaßen folgt in Kap. 6 auf täglicher und stündlicher Zeitskala. Diese herkömmliche Verifikation wird hier durchgeführt, um Vergleiche mit anderen Ergebnissen herstellen und an konkreten Beispielen die Motivation zu neuen Ansätzen zeigen zu können. Nach der Einführung des neuen Qualitätsmaßes (Kap. 7) folgt abschließend die Verifikation mit SAL im statistischen Sinne (Kap. 8). Es wurden Untersuchungen auf stündlicher und täglicher Zeitskala durchgeführt. Dabei konnten Vergleiche zwischen den verwendeten Modellen gemacht, aber vor allem auch Unterschiede zwischen Jahreszeiten und in Abhängigkeit von Intensitäten (über die Wahl des Schwellwertes) und verschiedenen Einzugsgebieten betrachtet werden. Auf stündlicher Zeitskala konnte zudem eine mögliche Abhängigkeit von der Vorhersagezeit untersucht werden.

Kapitel 2

Datengrundlage

In diesem Kapitel werden die in dieser Arbeit verwendeten Datensätze beschrieben. Verifiziert werden die Niederschlagsvorhersagen aus dem COSMO-Modell und dem Globalmodell des ECMWF für Deutschland von 2001-2004. Dabei wurde wegen der besseren Datenverfügbarkeit auf das aLMo („alpines Modell“, Version des Schweizer Wetterdienstes „MeteoSchweiz“ (Abschn. 2.1)) zurückgegriffen und das LM („Lokal-Modell“, Version des DWD) nur zum Vergleich der beiden Modellversionen für 2004 herangezogen. Zur Untersuchung der Qualität der QNV mit verschiedenen Modellauflösungen wird das Globalmodell des ECMWF verwendet (Abschn. 2.2).

Verifiziert werden die verschiedenen Modellvorhersagen mit täglichen Stationsmessungen des DWD (Abschn. 2.3), sowie einem für Deutschland neuartigen, zeitlich und räumlich hoch-aufgelösten, disaggregierten Datensatz. Dieser disaggregierte Datensatz wird aus der Kombination der täglichen Stationsmessungen und dem Radarkomposit (Abschn. 2.5) für Deutschland erzeugt (Details Kap. 3). Als weitere Beobachtungsdatensätze stehen die Niederschlagsklimatologie für Deutschland (Abschn. 2.6) und Daten aus stündlichen Stationsmessungen des DWD zur Verfügung (Abschn. 2.4). Zusätzlich wurden Informationen zu den deutschen Flusseinzugsgebieten benötigt, um die Vorhersagen auch unter hydrologischen Gesichtspunkten verifizieren zu können.

2.1 Das COSMO-Modell

In diesem Abschnitt wird das COSMO-Modell genauer vorgestellt. COSMO steht für das Konsortium für kleinskalige Modellierung und ist ein 1998 gegründeter Zusammenschluß der Wetterdienste, die das Lokal-Modell verwenden (COSMO, 2007). Die teilnehmenden Wetterdienste kommen momentan aus Deutschland, Griechenland, Italien, Polen, der Schweiz und Rumänien (Stand: August 2007). Seit kurzem werden alle Lokal-Modelle entsprechend diesem Konsortium als COSMO-Modelle bezeichnet, z.B. trägt das LME (gebietserweiterte Version des LM (Steppeler et al., 2003)) vom DWD nun den Namen COSMO-EU, das aLMo der MeteoSchweiz den Namen COSMO-7. Das COSMO-Modell wird im Rahmen dieser internationalen Zusammenarbeit und der Hilfe anderer wissenschaftlicher Einrichtungen stetig weiter

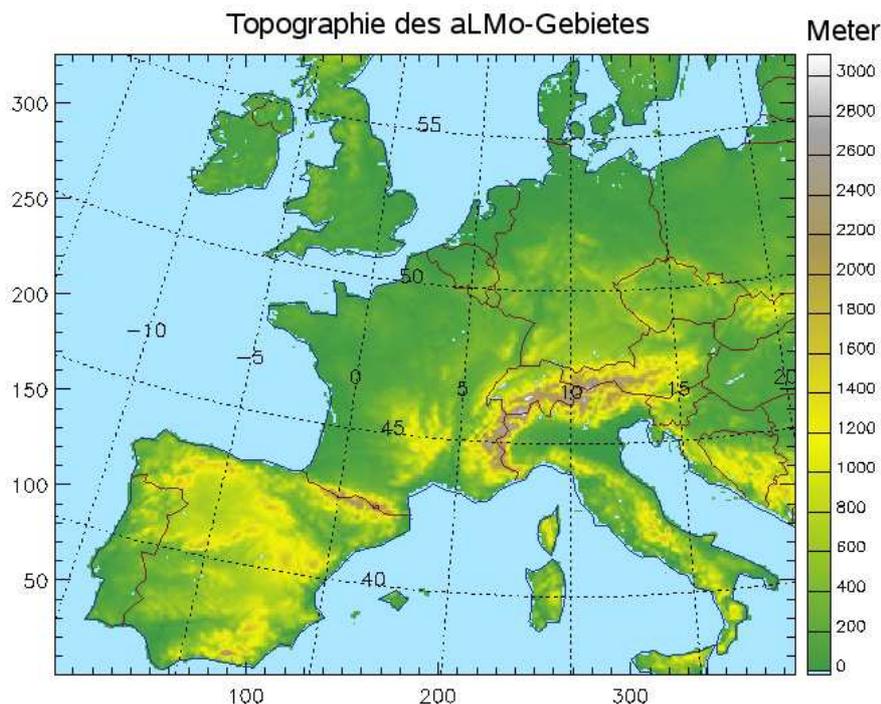


Abbildung 2.1: Modellgebiet des aLMo: Topographie (horizontale Auflösung: 7 km).

entwickelt. Die neue Namensgebung des Lokal-Modells wurde gegen Ende dieser Arbeit vorangetrieben, so dass hier aus technischen Gründen entschieden wurde, bei den Bezeichnungen LM und aLMo zu bleiben, zumal diese Modellversionen während der verwendeten Zeitreihe (2001-2004) auch diese Bezeichnungen trugen.

Zunächst war die Idee, das LM des DWD zu verifizieren. Allerdings ist es dem DWD aus Speicherplatzgründen nicht möglich, die Modellvorhersagen über Jahre zu speichern. Die LM-Zeitreihe für den Niederschlag war bei Beginn dieser Arbeit nur noch ab Oktober 2003 verfügbar. Um eine längere Zeitreihe für eine im Sinne der Statistik sinnvolle Verifikation zur Verfügung zu haben, wurde Kontakt mit der MeteoSchweiz aufgenommen. Dort waren glücklicherweise weitere Daten (ab 2001) verfügbar und so wurde entschieden, als lange Lokal-Modell-Zeitreihe aLMo-Vorhersagen zu verwenden (zumal beide Modellversionen quasi die gleiche Modellphysik haben, Unterschiede werden weiter unten beschrieben).

Das Modellgebiet des aLMo umfasst ganz West- und Mitteleuropa (Abb. 2.1) und ist damit gegenüber dem LM etwas nach Südwesten verschoben. Die horizontale Auflösung beträgt 7 km. Man erkennt deutlich die für Deutschland bekannte Topographie von der Norddeutschen Tiefebene über die Mittelgebirgsketten bis hin zu den Alpen im Süden. Allerdings wird auch ersichtlich, dass die realen Höhen der Gebirgsketten unterschätzt werden und z.B. das Mittelrheintal nicht komplett aufgelöst werden kann. In der Vertikalen hat das aLMo 45 Schichten (das LM 40 Schichten), wobei diese topographie-folgend definiert sind.

Das aLMo ist ein nicht-hydrostatisches kompressibles Regionalmodell und basiert auf den Hydro-Thermodynamischen Gleichungen, welche numerisch auf einem rotierten stereographischen Gitter gelöst werden. Prognostische Modellvariablen sind

Datensatz	Zeitraum	# fehlende Tage	in %
aLMo (00 UTC-Läufe)	2001 - 2004	5	0.3
LM (00 UTC-Läufe)	ab Okt. 2003 - 2004	0	0
ECMWF	Ende März 2001 - 2004	0	0
Radarkomposit	2001 - 2004	97	6.6

Tabelle 2.1: Vorhandene Zeitperiode und Datenlücken der Modelldatensätze und des Radarkomposits.

der dreidimensionale Wind, die Druckabweichung, die Temperatur, die turbulent-kinetische Energie, die spezifische Feuchte und der spezifische Wolkenwassergehalt. Hinzu kommen seit September 2003 der spezifische Wolkeneisgehalt und seit November 2004 die spezifischen Wassergehalte für Regen und Schnee (im LM seit April 2004). Vor der Einführung des prognostischen Niederschlags (Regen und Schnee) musste der Niederschlag diagnostisch aus dem Wolkenwasser- und Wolkeneisgehalt abgeleitet werden. Dabei hängt es von bestimmten Schwellwerten ab, wann Hydrometeore als Wolkenpartikel, wann als Niederschlagspartikel angesehen werden. Wenn niederschlagsbildende Hydrometeore vorhanden sind, so werden diese instantan (je nach Phase als Regen oder Schnee) aus der Wolke entfernt und dem darunter liegenden Gitterpunkt am Boden als Niederschlag zugeteilt. Der prognostische Niederschlag unterscheidet sich dabei vor allem vom diagnostischen Niederschlag, indem er mit dem Wind advehiert wird. (Schulz und Schättler, 2005)

Einige kleinskalige Prozesse, die nicht mit 7 km-Auflösung erfasst werden können, müssen parametrisiert werden:

- Strahlung,
- skaliger Niederschlag (Mikrophysik),
- flache, mittlere und hochreichende Feuchtekonvektion (Massenflussverfahren nach Tiedtke, 1989),
- partielle Bewölkung und
- Bodenprozesse.

Zweimal täglich, um 00 und 12 UTC, werden auf einem Supercomputer am nationalen Hochleistungsrechenzentrum der Schweiz in Manno in ca. 70 Minuten 48 h-Vorhersagen berechnet (ähnliches gilt für das LM vom DWD). Zu jeder vollen Stunde liefert das aLMo Vorhersagefelder. Tagessummen wurden erzeugt, indem die stündlichen Felder von 6-30 h (00 UTC-Läufe, wobei die ersten sechs Stunden als Modell-„Spin-up“ gelten) und 18-42 h (12 UTC-Läufe) aufsummiert wurden, wobei in dieser Arbeit meist nur die 00 UTC-Läufe betrachtet werden (Erklärung dazu in Kap. 4).

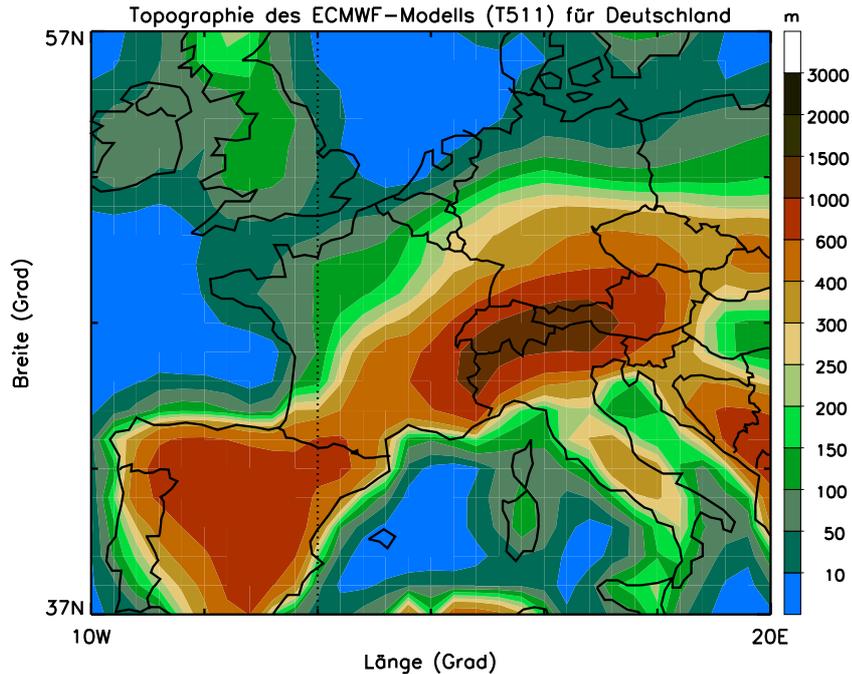


Abbildung 2.2: Ausschnitt der ECMWF-Topographie (Version T511) über Mittel- und Südwest-Europa.

Da das aLMo ein regionales Wettervorhersagemodell ist, benötigt es nicht nur Anfangsbedingungen aus Assimilationszyklen, sondern auch Randbedingungen aus einem Globalmodell. Dazu werden im COSMO aktuell die Vorhersagen des Globalmodells des ECMWF verwendet. Allerdings wurden diese Randbedingungen für Vorhersagen des aLMo erst ab dem 16.09.2003 operationell eingesetzt. Davor wurden die Randdaten vom GME (Globalmodell des DWD) verwendet (Hinweis: das operationell eingesetzte LM beim DWD verwendet bis heute die Randdaten des GME). Zwei weitere wichtige Änderungen im Bezug auf den Niederschlag wurden innerhalb der Zeitreihe vorgenommen: seit 31.10.2001 werden die Vorhersagen mit einer operationellen Nudging-Assimilation und seit 16.11.2004 auch mit dem prognostischen Niederschlag (Advektion fallender Niederschlagspartikel) durchgeführt (im LM wurde der prognostische Niederschlag am 26.04.2004 operationell eingeführt). Die Unterschiede im prognostischen Niederschlag, aber auch in der Wahl der Randbedingungen und das etwas verschobene Modellgebiet, machen gerade für 2004 einen Vergleich zwischen den beiden Lokal-Modellen interessant. Die stetige Weiterentwicklung des aLMo führt aber auch dazu, dass man kaum eine längere, homogene Zeitreihe zur Verfügung hat.

Tabelle 2.1 zeigt die Datenlücken der Lokal-Modell-Vorhersagen für 2001-2004. Es kann über eine fast vollständige Zeitreihe verfügt werden (aLMo 0.3%, LM keine Lücken).

2.2 Das Globalmodell des ECMWF

Das Globalmodell des ECMWF ist (im Gegensatz zum Lokal-Modell (Abschn. 2.1)) ein spektrales numerisches Modell und wird in Reading (England) betrieben. Es ist seit 1979 im operationellen Gebrauch und wird seitdem stetig weiterentwickelt (Informationen zum Modell und allgemein zum ECMWF finden sich auf der Homepage des ECMWF: www.ecmwf.int).

Die für diese Arbeit verwendete Modellversion ist T511. Dies entspricht einer horizontalen Auflösung von ungefähr 0.5 Grad oder 60 km. Die im Vergleich zur aLMo-Topographie (Abb. 2.1) gröbere Auflösung ist deutlich im Topographie-Ausschnitt für Mittel- und Südwest-Europa zu erkennen (Abb. 2.2). Die Alpen sind z.B. nur noch als ein großer Berg mit etwa 1500 m Höhe dargestellt, ansonsten sind in Deutschland keinerlei Mittelgebirgsregionen erkennbar. Dies hat entsprechenden Einfluß auf die Strukturen in der Niederschlagsvorhersage. Es können nur flächige Niederschläge erzeugt werden. Kleinskalige Phänomene, wie etwa konvektive Zellen, sind mit dem Globalmodell nicht vorhersagbar.

Um einen quantitativen Vergleich des Globalmodells mit dem Lokal-Modell machen zu können, wurden die ECMWF-Vorhersagen auf das 7 km-Gitter interpoliert. Tägliche Niederschlagssummen wurden wie für das Lokal-Modell von 6-30 h aufsummiert (00 UTC-Läufe). Aus Tab. 2.1 geht hervor, dass die Niederschlagsvorhersagen des ECMWF keine Datenlücken aufweisen, jedoch erst seit Ende März 2001 verfügbar sind. Dies liegt daran, dass die hier benötigten Vorhersagen aus 00 UTC-Läufen erst Ende März 2001 hinzukamen, zuvor nur 12 UTC-Läufe gerechnet wurden. Für diese Arbeit wurde auf die Möglichkeit verzichtet, die 18-42 h-Vorhersagen als Tagessummen zu verwenden, da dies einen Vergleich der Modelle schwieriger gemacht hätte, zum anderen der Hauptteil der Verifikation dem aLMo und nicht dem ECMWF-Modell gilt.

2.3 Tägliche Stationsmessungen

Der DWD hat in Deutschland in den Jahren 2001-2004 fast 4000 Stationen betrieben, an denen einmal täglich Niederschlag gemessen wurde. Dabei nimmt jedoch die Anzahl an Stationen bis heute ab und wird weiter sinken, was zum einen schlicht Kostengründe hat, zum anderen soll vermehrt auch ein zeitlich höher aufgelöstes Messnetz aufgebaut bzw. das alte umgewandelt werden. In Abb. 2.3 sind die Standorte der Niederschlagsmessstationen für alle vier Jahre in Deutschland dargestellt. Tabelle 2.2 zeigt zusätzlich die Anzahl an messenden Stationen in jedem der vier Jahre und die Anzahl an Stationen davon mit Messlücken. Diese sind zum Teil erheblich, allerdings oft auch nur über wenige Tage. Man kann gut die stetig abnehmende Anzahl an täglich messenden Stationen erkennen. Zu sehen ist aber auch, dass der DWD ein sehr dichtes und homogen über Deutschland verteiltes Messnetz betreibt. Allerdings ist es sehr wichtig zu erwähnen, dass diese Verteilung nicht repräsentativ mit der Höhe ist. Meist wird der Niederschlag in Tallagen, weniger in Höhenlagen gemessen. Dies führt bei der Gitterung dieser Daten (Abschn. 3.1) dazu, dass der

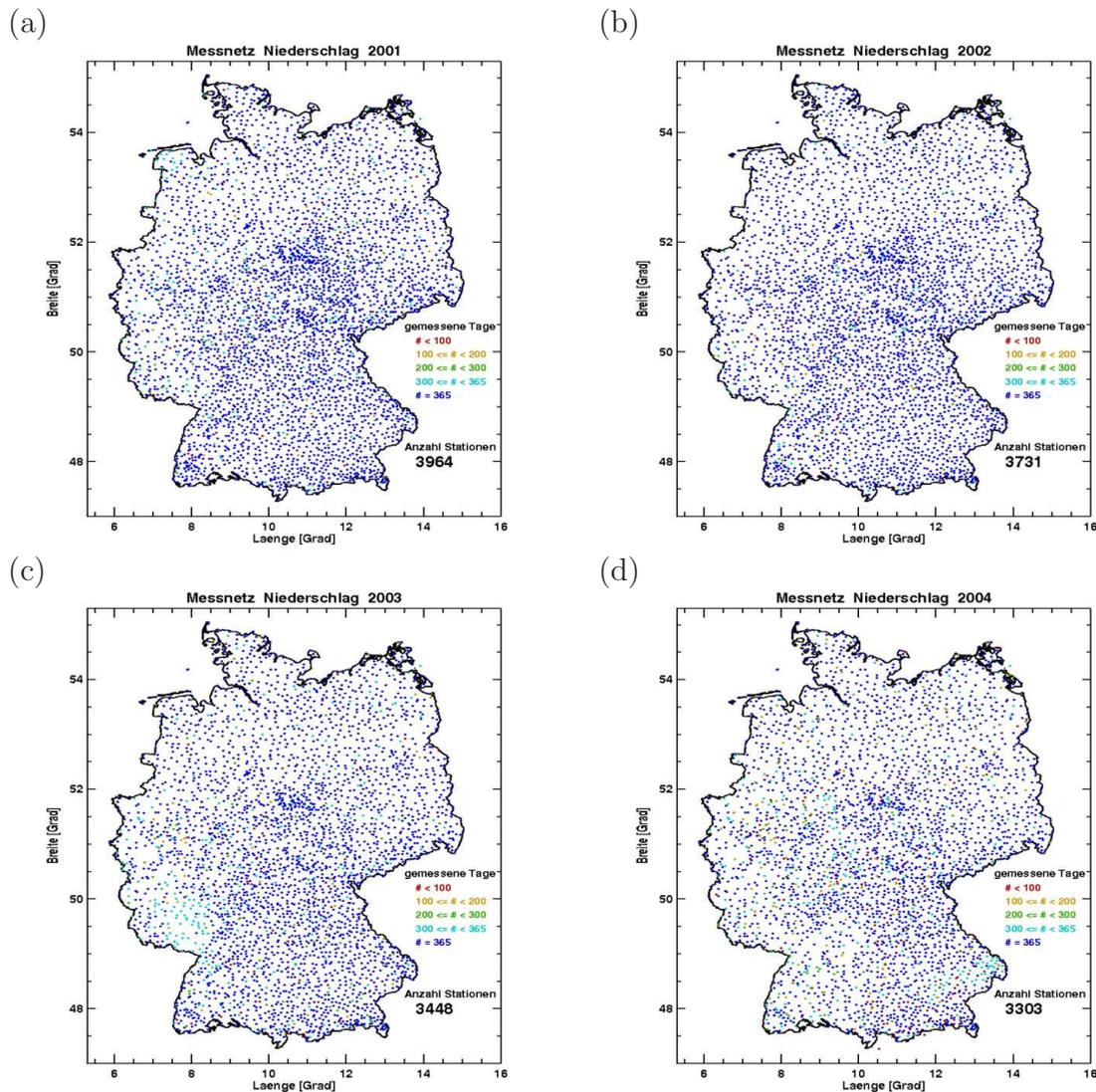


Abbildung 2.3: Stationsnetz für tägliche Niederschlagsmessungen in Deutschland: (a) 2001, (b) 2002, (c) 2003, (d) 2004. Farbkodierung zeigt Anzahl an Messlücken in Tagen an (Legende).

Niederschlag unterschätzt wird, da er eine klimatologische Korrelation mit der Höhe hat (je höher, desto mehr Niederschlag). Dieser Unterrepräsentativität wird mit Hilfe der Klimadaten (Abschn. 2.6) bei der Gitterung Rechnung getragen.

An den Stationen wird immer um 7:30 Uhr lokaler Zeit (also je nach Datum MEZ (UTC + 1 h) oder MESZ (UTC + 2 h)) die Niederschlagssumme der vergangenen 24 Stunden, meist mit dem Hellmann-Regenmesser (Abb. 2.4: Niederschlagsmesser aus Edelstahl mit einer Auffangfläche von 200 cm², entwickelt vom Meteorologen J. G. Hellmann im Jahre 1886), durch einen Beobachter gemessen. An Stationen mit automatischen Messungen sind die Messgeräte meist Niederschlagswaagen oder -wippen.

Messungen im Allgemeinen, aber vor allem beim Niederschlag, sind immer mit Messfehlern verbunden. Die Messfehler des Niederschlags sind vielfältig und zum Teil beträchtlich. Dies zeigen Untersuchungen von Sevruk (1985b) für die Schweiz und von Richter (1995) speziell anhand des Hellmann-Regenmessers. Dominierende

Datensatz	Zeitraum	# Stationen	davon mit Datenlücken
tägliche Messungen	2001	3964	355
	2002	3731	189
	2003	3448	428
	2004	3303	677
stündliche Messungen	2001	225	3
	2002	268	41
	2003	312	22
	2004	675	256

Tabelle 2.2: Anzahl an Stationsmessungen des Niederschlags und Datenlücken. Die großen Datenlücken bei den stündlich messenden Stationen kommen daher, dass im zweiten Halbjahr jeweils einige neue Stationen hinzukamen.

Verluste bei der Niederschlagsmessung sind demnach durch Benetzung und Windablenkung (wegen der dynamischen Beschleunigung über der Öffnung des Messgerätes) zu erwarten.

Systematische Fehler durch Benetzungs- und Verdunstungsfehler kommen infolge der Befeuchtung des Auffangtrichters und des Verweilens des Niederschlags in der Sammelkanne bis zur Messung vor. Dies ist unvermeidbar, wobei die Größe des Benetzungsfehlers von den Hafteigenschaften der Trichterfläche abhängt. Der maximale Verlust kommt natürlicherweise bei vollständiger Benetzung vor und hängt weiter davon ab, ob zwischen zwei Niederschlagsereignissen eine vollständige Abtrocknung der Trichterwand stattgefunden hat. Der Fehler ist zudem witterungsabhängig (Verdunstung). Ist es relativ kühl, feucht und bedeckt, so ist der mittlere Fehler kleiner als bei warmen, trockenen und wechselnd bewölkten Bedingungen. Hinzu kommt, dass die Benetzung wiederum abhängig von der gefallenem Niederschlagsmenge ist (50 % bei 0.5 mm, 100 % ab 3 mm). Die mittleren Fehler aus Benetzung und Verdunstung belaufen sich (mit Berücksichtigung von Abtrocknungszeiten zwischen zwei Ereignissen) insgesamt auf 0.22 mm pro Tag mit einer Streuung von 0.08 mm.

Der Windfehler ist allgemeingültig nur schwer korrigierbar, da hierfür routinemäßige Messungen der Windgeschwindigkeiten in Aufstellhöhe des Regennessers vonnöten wären. Fehler entstehen durch Hindernisse im Strömungsfeld (ein Überströmen bedeutet immer eine leichte Aufwärtskomponente durch den Wind) und sind abhängig von der Windgeschwindigkeit, sowie der Art der Niederschlagspartikel (Größenverteilung). Nach Untersuchungen von Richter (1995) sind 81.6 % des Gesamtniederschlags Regen, 5.2 % Schnee und 13.2 % eine Mischung aus beidem. Flüssige Niederschläge liefern dabei geringere Windfehler als feste Niederschläge, allerdings hängt dies auch von der Tropfengröße ab (kleinere Tropfen werden leichter verweht). Daraus resultieren recht unterschiedliche Windfehler für Winter und Sommer. Sie liegen für flüssige Niederschläge im Mittel bei 12.3 ± 3.1 % (Winter) und 5.6 ± 1.7 % (Sommer für große Tropfen). Es zeigt sich also ein klarer Jahresgang im Windfehler. Mischniederschläge (vor allem zwischen November und März) haben Fehler um 25-

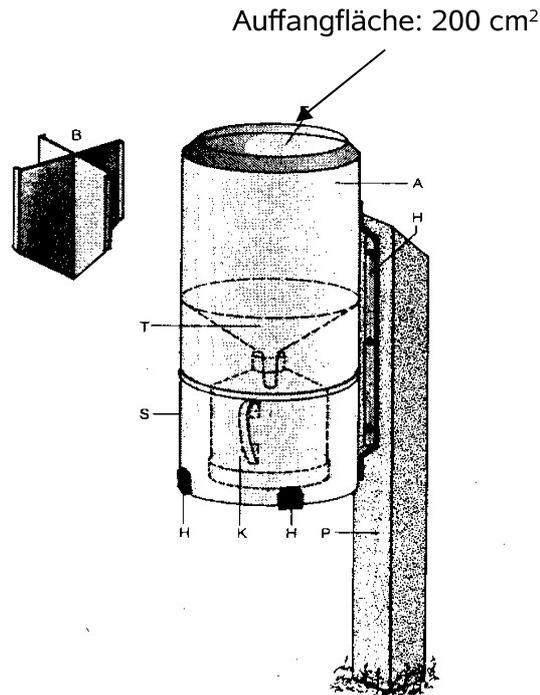


Abbildung 2.4: Hellmann-Regenmesser. *F: Auffangöffnung, A: abnehmbares Auffanggefäß mit eingebautem Trichter (T), K: Sammelkanne, S: Schutzgehäuse, H: Halterung, P: Montagepfahl, B: Schneekreuz.*

30 %, Schnee sogar Fehlerraten von 45-70 % am real gefallenen Anteil. 10 % werden zudem vom Hellmann-Regenmesser durch Verwehungen von Schnee gar nicht erfasst, dabei ist auch zu berücksichtigen, dass sich Schnee auf dem Rand des Regenmessers ansammeln und von Zeit zu Zeit als größere Menge in den Auffangbehälter fallen kann.

Langjährige Mittelwerte des gesamten Messfehlers eines Hellmann-Regenmessers werden für eine mäßig geschützt gelegene Station und nach Niederschlagsart gewichtet mit 11.6 % Windfehler + 5.1 % Benetzungs- und Verdunstungsfehler = 16.7 % angegeben (Richter, 1995).

Eine individuelle Korrektur jedes Hellmann-Regenmessers ist nicht möglich. Mit der Ableitung spezifischer, von der Art und Höhe des Niederschlags abhängiger Korrekturfunktionen für Tagessummen ist es jedoch möglich, Stationsdaten mit mittleren Fehlern zu korrigieren. Zudem wurde eine regionale Fehlerabhängigkeit in Deutschland gefunden, allerdings konnte keine Topographieabhängigkeit nachgewiesen werden.

Die vorliegenden Stationsdaten der Jahre 2001-2004 sind nach diesen Verfahren vom DWD qualitätskontrolliert worden. Zusätzlich wurde der Gitterung (Abschn. 3.1) eine Kontrolle der Daten auf Ausreißer und andere Auffälligkeiten vorangeschaltet. Eine Verifikation, wie in dieser Arbeit gemacht, sollte immer im Hinterkopf behalten, dass die als „bestmögliche Wahrheit“ angenommenen Messdaten ebenfalls mit größeren Fehlern behaftet sind, welche zwar durch Qualitätskontrollen vermindert, aber durch mittlere Fehlerannahmen für alle Stationen nicht vollständig eliminiert werden können.

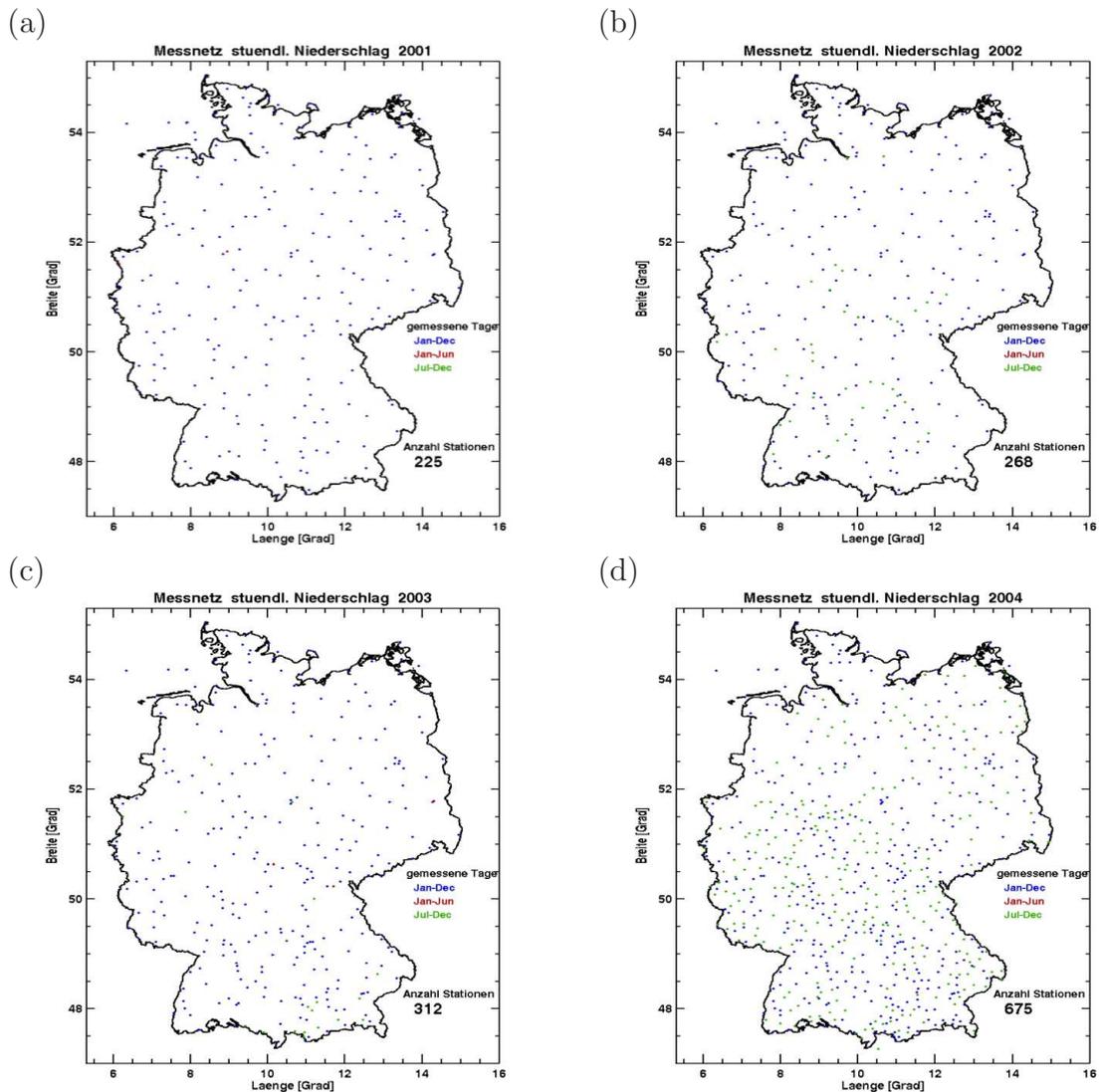
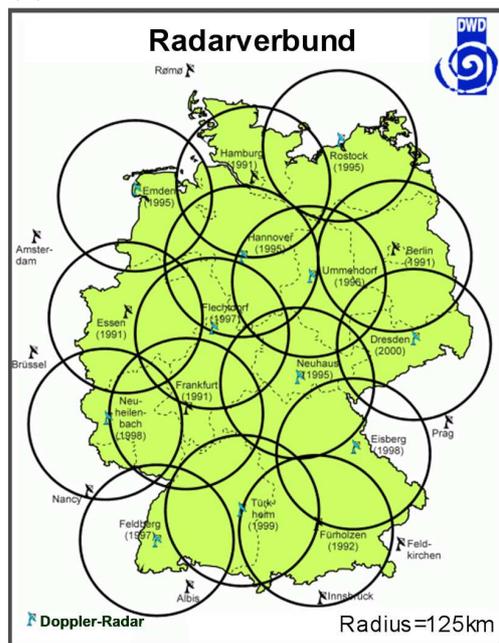


Abbildung 2.5: Stündlich messende Niederschlagsstationen in Deutschland: (a) 2001, (b) 2002, (c) 2003, (d) 2004. Farbkodierung zeigt die Messdauer an (Legende).

2.4 Stündliche Stationsmessungen

An einem Teil der Niederschlagsmessstationen (Abschn. 2.3) des DWD wird mit stündlicher Auflösung Niederschlag gemessen. Dieses zeitlich höher aufgelöste Messnetz wurde in den letzten Jahren deutlich ausgebaut (Tab. 2.2), so dass sich die Anzahl der Stationen von 2001 (ca. 200 Stationen) bis 2004 (über 600 Stationen) schon verdreifacht hat mit steigender Tendenz. Diese Niederschlagsstationen werden in der vorliegenden Arbeit vor allem zur Prüfung des disaggregierten Datensatzes (Kap. 3) und zum Vergleich des Tagesgangs (Abschn. 5.3) verwendet. Die stündlich messenden Stationen sind zwar sehr gleichmäßig über Deutschland verteilt (Abb. 2.5; mit dem Problem der Repräsentativität in Höhenlagen), jedoch reicht die Anzahl nicht aus, um eine qualitativ hochwertige Gitterung mit 7 km Auflösung durchzuführen. Deshalb bleibt die Disaggregation ein entscheidender Teil dieser Arbeit, um einen stündlichen Datensatz aus Beobachtungen zur Verifikation zu erstellen.

(a)



(b)

Radarstandort	operationell seit
München	1987
Frankfurt	1988
Hamburg	1990
Berlin-Tempelhof	1991
Essen	1991
Hannover	1994
Emden	1994
Neuhaus	1994
Rostock	1995
Ummendorf	1996
Feldberg (Schwarzwald)	1997
Eisberg	1997
Flechtdorf	1997
Neuheilenbach	1998
Türkheim	1998
Dresden	2000

Abbildung 2.6: Radarverbund für Deutschland: (a) Karte, (b) Start des operationellen Betriebs. Die schwarzen Kreise in (a) zeigen den Radius (125 km) um einen Standort für eine qualitativ brauchbare Messung an. (Quelle: DWD)

2.5 Das Radarkomposit

Der DWD betreibt seit einigen Jahrzehnten Einrichtungen zur Fernerkundung, zu denen auch der Radarverbund zählt (Radar engl.: Radio detecting and ranging). In Abb. 2.6 (a) sind die Radarstandorte für Deutschland eingezeichnet. Der schwarze Kreis um den jeweiligen Standort hat einen Radius von 125 km und zeigt die Entfernung für eine qualitativ brauchbare Radarmessung an. In Deutschland gibt es momentan 16 Radarstandorte (Abb. 2.6 (b)), an denen kontinuierlich mit Doppler-Radaren gemessen wird.

Aus den Radarmessungen (Volumenscan) jedes einzelnen Radars wird alle 15 Minuten ein Deutschlandkomposit mit einer Auflösung von $4 \cdot 4$ km erstellt (PC-Produkt), wobei den Regionen mit Radarüberlappungen immer der höhere Niederschlagswert zugeordnet wird. Wie in Abb. 2.6 (b) gut zu erkennen ist, reichen die 16 Radare für eine nahezu komplette Abdeckung über Deutschland aus. Es gibt aber einige kleine Regionen im Südosten, Nordosten und ganz im Norden Deutschlands, die qualitativ nicht gut abgedeckt sind.

Während Niederschlagsmessungen am Boden stets nur Punktmessungen sind, erlauben Radarinformationen eine relativ flächendeckende Wetterüberwachung. Zur räumlich hohen Abdeckung kommt ein weiterer entscheidender Vorteil der Radarmessung: die hohe zeitliche Auflösung.

Das Radar misst allerdings nicht direkt Niederschlagsmengen, sondern Rückstreusignale (Reflektivitäten) von in der Luft befindlichen Partikeln (abhängig vom Aggregatzustand und der Größe). Daraus folgt ein entscheidender Nachteil der Radarmessung, da diese Reflektivitäten mit Hilfe von z - R -Beziehungen (z : Reflektivitätsfaktor

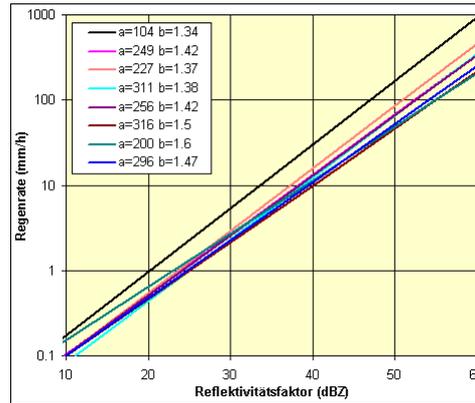


Abbildung 2.7: Verschiedene z - R -Beziehungen in Abhängigkeit von der gewählten Niederschlagspartikelverteilung bzw. Wetterlage (Hagen, 2004; Werte aus Aniol et al., 1980).

in mm^6m^{-3} (oft auch Reflektivität Z in dBZ), R : Niederschlagsmenge in mm) unter Annahme der Rayleigh-Streuung in Niederschläge umgerechnet, bzw. abgeschätzt werden müssen (das Radar unterschätzt z.B. den Niederschlag bei Schnee- oder Eisparkeln aufgrund des Brechungsindex). Dies ist mit teilweise deutlichen Unsicherheiten verbunden (Kap. 3). Die Schwierigkeit in der Umrechnung liegt vor allem darin, dass der Zusammenhang zwischen Regenrate und Reflektivitätsfaktor nicht linear ist. Beide Parameter hängen von der Größenverteilung der Regentropfen D ab ($z \sim \sum D^6$, $R \sim \sum D^{3.7}$). Im Allgemeinen wählt man daher eine empirische Beziehung der Form

$$z = aR^b. \quad (2.1)$$

Die beiden Konstanten a und b sind abhängig von der Niederschlagspartikelverteilung. Je nach Klimazone und Wetterlage (stratifomer oder konvektiver Niederschlag) finden sich über 60 verschiedene Werte für a und b (Battan, 1973). Aus Distrometer-Messungen (Aniol et al., 1980) wurden einige wetterlagenabhängige z - R -Beziehungen errechnet und in Abb. 2.7 dargestellt: (Schwarz) Warmluftzufuhr, (Rosa) Kaltluftzufuhr, (Türkis) Gewitter, und weitere (Zahlen für a und b in der Graphik). Die z - R -Beziehung von Marshall und Palmer (1948) ist wohl die bekannteste und wird noch heute als gängigste Abschätzung für stratiforme Niederschläge verwendet. Dabei wurde aus Messungen $a = 200$ und $b = 1.6$ gefunden, die theoretisch berechneten Konstanten liegen bei $a = 296$ und $b = 1.47$. Die dieser Arbeit zugrunde liegende z - R -Beziehung ist die vom DWD operationell verwendete mit den Konstanten $a = 256$ und $b = 1.42$ (Abb. 2.7).

Die vielen verschiedenen Ansätze zeigen, dass eine für alle Niederschlagsarten zufriedenstellende Umrechnung nicht mit einer z - R -Beziehung möglich ist. Das hat auch zur Folge, dass jeder Benutzer von Radardaten sich die für sich geeignetste z - R -Beziehung aussucht. Zusätzlich zur Unsicherheit aus der Wahl der z - R -Beziehung gibt es eine weitere Vereinfachung: die in Niederschläge umzurechnenden Reflektivitäten werden in einige wenige Reflektivitätsklassen unterteilt. Die in dieser Arbeit verwendeten Reflektivitätsklassen und deren Umrechnung in Niederschlagsmengen werden in Abschn. 3.2 beschrieben.

In Abb. 2.8 sind drei Beispiele für bereits in Niederschlagssummen umgerechnete

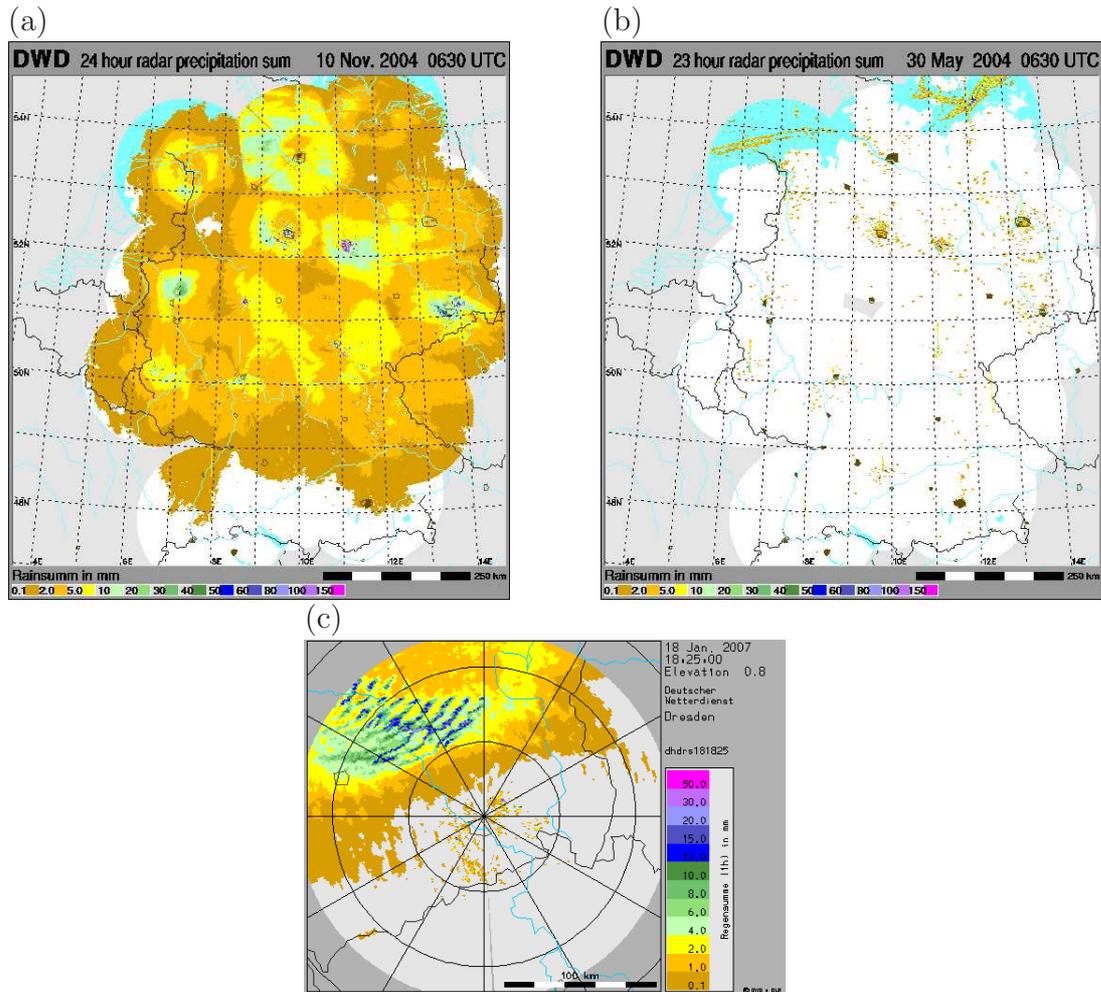


Abbildung 2.8: Beispiele zu Fehlerquellen bei der Radarmessung (Bilder von Martin Hagen).

Radarbilder dargestellt. In den zwei Kompositbildern (a) und (b), sowie dem Radarbild (c) vom Standort Dresden kann man gut einige Fehlerquellen der Radarmessung aufzeigen. Bild (a) (24 h-Summe) zeigt hohe ringförmige Niederschlagsmengen um einige der Radarstandorte an. Dies zeigt das Problem der Radarmessung in einer gewissen Höhe, in der Niederschlagspartikel schmelzen und damit zu hohe Reflektivitäten liefern (Schmelzschicht). Bild (b) zeigt einen eigentlich niederschlagsfreien Tag über Deutschland, jedoch werden vereinzelt punktförmige Niederschläge angezeigt. Diese resultieren aus Bodenechos, z.B. Rückstreuung von Rauchfahnen in Industriegebieten oder auch von Schiffen („Niederschlagssignale“ in Nord- und Ostsee). Erkennbar ist in diesem Bild auch, dass es auch Datenausfälle geben kann (kleine graue Fläche in der Mitte Deutschlands). In diesem Fall wird der niederschlagsfreie Tag vermutlich dazu genutzt, das Radar in Flechtdorf zu warten (vgl. Abb. 2.6(b)). Ein drittes Problem der Radarmessung ist, dass das Radar trotz zeitlich hoch-aufgelöster Messung schnell ziehende Niederschlagsstrukturen räumlich nicht vollständig erfasst (Bild (c)). Es entstehen dann wellenartige Strukturen verschieden hoher Niederschläge. Theoretisch ist es sogar möglich, dass eine konvektive Zelle schnell über einen Niederschlagsmesser am Boden zieht, dieser den

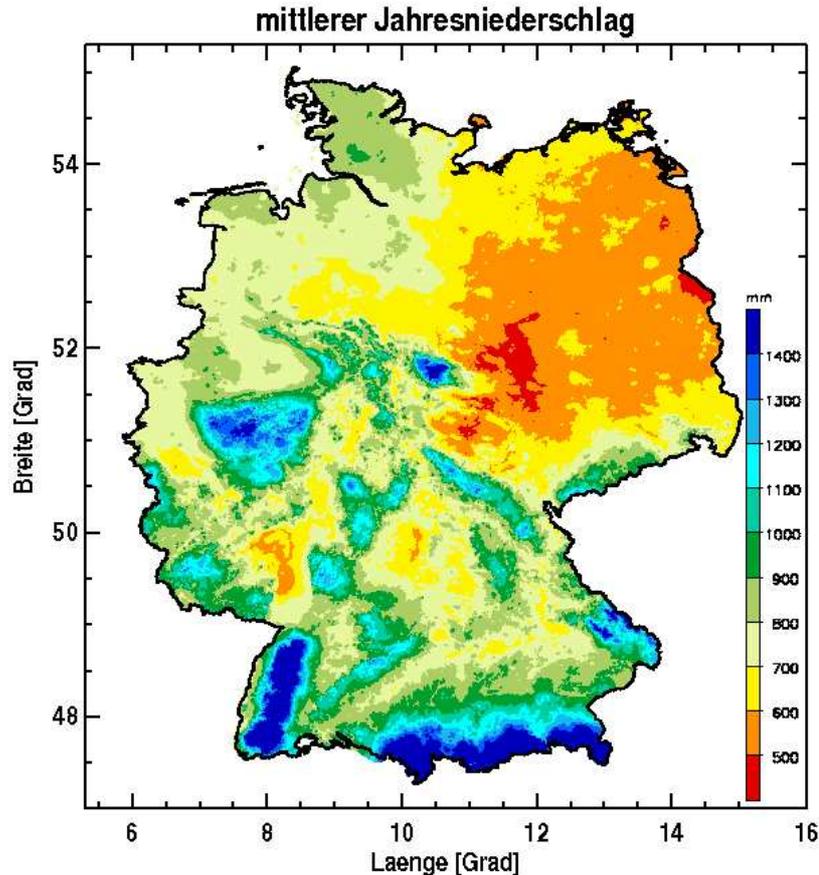


Abbildung 2.9: Mittlerer Jahresniederschlag für Deutschland (Klimatologie von 1961-1990).

Niederschlag erfasst, jedoch das Radar beim Volumenscan erst an diesen Punkt kommt, wenn die Zelle bereits weitergezogen ist. Ein weiteres, hier nicht bildlich dargestelltes Problem ist die Abschattung des Radarsignals vor allem an Bergen, aber auch (hohen) Gebäuden. Dahinter kann das Radar in gewissen Höhen keine Informationen sammeln.

In Tab. 2.1 ist die Verfügbarkeit des Radarkomposits für Deutschland für 2001-2004 aufgeführt. Es fehlen 6.6% an möglichen Tagen.

2.6 Die Niederschlagsklimatologie

Die hier verwendete Niederschlagsklimatologie beinhaltet Monatsmittel auf einem 1·1 km-Gitter. Die Datengrundlage liefert das Stationsmessnetz des DWD mit ca. 3000 Stationen, was einem mittleren Stationsabstand von 10 km und damit einem sehr dichten Messnetz entspricht. Diese Niederschlagsklimatologie umfasst 30 Jahre und reicht von 1961-1990 (Müller-Westermeier, 1995).

Die Stationsdaten müssen zur Gitterung auf das hoch-aufgelöste 1·1 km-Gitter sinnvoll interpoliert werden. Dabei ist es wichtig, die Repräsentativität der Stationsverteilung zu prüfen, da die Messungen in besonders gebirgigen Regionen meist in den Tallagen, weniger auf den Höhenlagen durchgeführt werden. Dies ist für die Gitter-

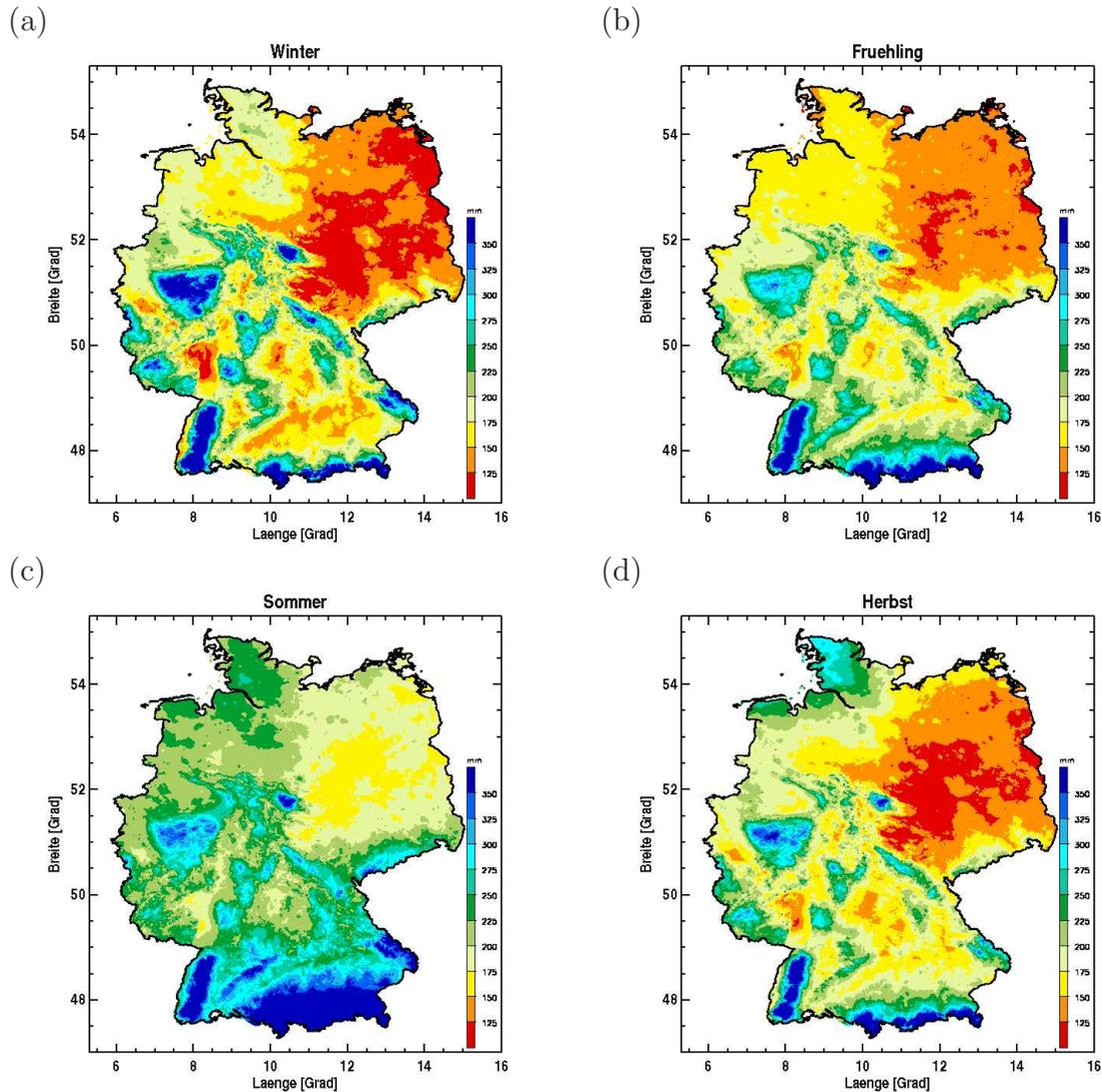


Abbildung 2.10: Saisonaler gemittelter Niederschlag (Klimatologie von 1961-1990): (a) Winter, (b) Fruehling, (c) Sommer, (d) Herbst.

rung aller Variablen von Bedeutung, die eine Korrelation mit der Höhe zeigen, also auch dem Niederschlag (allgemein steigt die Niederschlagsmenge mit der Höhe an). Wichtig ist auch, dass eine signifikante räumliche Korrelation zwischen den Stationen besteht, was bei dem dichten Messnetz in Deutschland erfüllt ist.

Nach Müller-Westermeier (1995) bringt eine räumliche Interpolation der statistischen Kenngrößen klimatischer Parameter noch keine befriedigende räumliche Darstellung. Dies liegt daran, dass das Messnetz nicht repräsentativ genug verteilt ist, um die Einflüsse der Topographie hinreichend genau darzustellen. Insbesondere ist der Einfluss der Messungen einzelner Bergstationen auf das Umfeld zu groß, während Gebirge, bei denen keine Stationen in höheren Lagen vorhanden sind, oft kaum einen klimatischen Einfluss erkennen lassen.

Dieses Repräsentativitätsproblem tritt vor allem bei der Temperatur zu Tage. Bei der einfachen linearen räumlichen Interpolation des Niederschlags ist auf Grund des dichten Messnetzes aber bereits eine Kartenstruktur erreichbar, die Ähnlichkeiten mit der Topographie in Deutschland hat. Dennoch wurde auch für die Übertra-

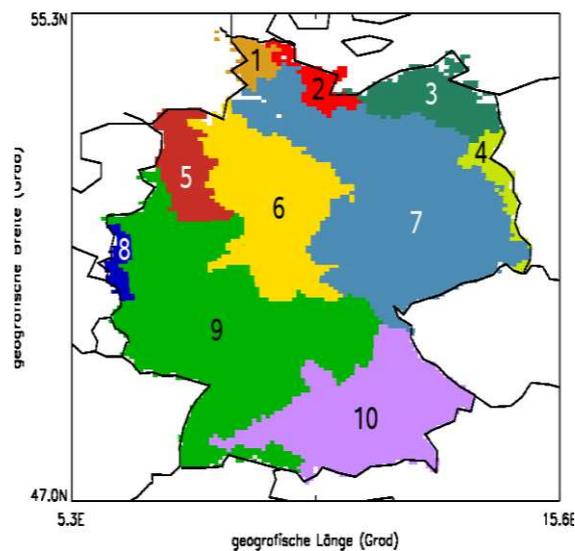


Abbildung 2.11: Karte der großen Flusseinzugsgebiete in Deutschland: 1) Eider, 2) Schlei/Trave, 3) Warnow/Peene, 4) Oder, 5) Ems, 6) Weser, 7) Elbe, 8) Maas, 9) Rhein, 10) Donau.

gung des Niederschlags in die Fläche die Korrelation mit der topographischen Höhe berücksichtigt. Dieses Verfahren wird auch „Kriging-Verfahren“ genannt (nach D.G. Krige: geostatistisches Interpolationsverfahren, das auf Grundlage der topographischen Verteilung die räumliche Verteilung der betrachteten Daten abschätzt).

Wenn man sich mit Niederschlag (insbesondere auch der Verifikation der Niederschlagsvorhersage) befasst, ist ein erster Blick auf die Klimatologie sehr sinnvoll. Man bekommt einen ersten Eindruck, wie die Verteilung des Niederschlags für Deutschland im langjährigen oder aufgeteilt in jahreszeitliche Mittel aussieht. Abb. 2.9 zeigt den mittleren Jahresniederschlag der vorliegenden Klimatologie. Auffällig ist, dass insbesondere große Teile Ostdeutschlands relativ trocken sind mit mittleren Niederschlägen unter 400 mm. Die Maxima finden sich in den Deutschen Mittelgebirgen (z.B. südlich des Ruhrgebietes, im Harz oder im Schwarzwald) und dem Alpenvorland sowie den Alpen selbst. Die maximalen mittleren Jahresniederschläge betragen über 3700 mm. Jahreszeitlich aufgeteilt (Abb. 2.10) zeigen sich einige Unterschiede. Das Bild des Jahresmittel mit Maxima in den Mittelgebirgen und Ostdeutschland als relativ trockener Region findet sich am besten im Winter wieder. Es zeigt sich, dass der Winter gegenüber den anderen Jahreszeiten im Osten besonders trocken ist (ähnlich im Herbst, jedoch nicht so ausgedehnt), die Maxima in den Gebirgen besonders feucht sind (vor allem in der Mitte Deutschlands). Im Sommer ist das Bild nicht mehr so klar strukturiert. Die allgemeine Wetterlage ist (klimatologisch gesehen) durch konvektive Niederschlagsereignisse dominiert (im Gegensatz zu frontalen Niederschlägen im Winter und Herbst). Das Alpenvorland und der Schwarzwald stehen mit hohen Niederschlägen gegenüber dem restlichen Deutschland stark heraus. Die Maxima in den übrigen Mittelgebirgsregionen sind nicht mehr so deutlich, allerdings fällt dafür in den im Winter trockenen Regionen deutlich mehr Niederschlag (im Osten z.B. um 200 mm gegenüber teilweise weit unter 100 mm im Winter).

Frühling und Herbst liegen mit den klimatologischen Niederschlagsmengen im allgemeinen zwischen den Werten in Winter und Sommer.

2.7 Flusseinzugsgebiete in Deutschland

Um die Verifikation auch auf der Ebene einzelner Flusseinzugsgebiete machen zu können, wurden über die Bund/Länder-Arbeitsgemeinschaft Wasser (LAWA) Daten dieser Einzugsgebiete angefragt und bereitgestellt. Die größten deutschen Flusseinzugsgebiete sind in Abb. 2.11 dargestellt. In dieser Arbeit wird auf die Gebiete der Ems, Weser, Donau, des Rheins und besonders der Elbe Bezug genommen (Kap. 6, 7 und 8). Es gilt anzumerken, dass zwar auch über die Grenzen Deutschlands Flusseinzugsgebiets-Daten vorhanden gewesen wären, jedoch dort keine Beobachtungsdaten aus Stationsmessungen vorlagen.

Durch Kenntnis der Begrenzung eines jeden Flusseinzugsgebietes war es möglich, Masken zu erstellen, die jeden Gitterpunkt der Modell- und Beobachtungsdaten eindeutig einem Gebiet zuordnen (oder ausschließen). Im Sinne der Niederschlagsverifikation macht dies nun Aussagen über die Qualität der QNV für einzelne Flusseinzugsgebiete möglich. Ein Hydrologe ist etwa daran interessiert, ob die Vorhersage in einem bestimmten Gebiet in einem gewissen Rahmen zutreffend ist oder tendenziell die Niederschlagsmenge oder -häufigkeit unter-, bzw. überschätzt wird. Dies kann ernsthafte Auswirkungen auf Hochwasserwarnungen haben.

Kapitel 3

Disaggregation

In diesem Kapitel wird die Disaggregation (Hagen et al., 2003; Litschi, 2005) vorgestellt, die als Resultat einen räumlich und zeitlich hoch-aufgelösten Datensatz aus Beobachtungen (Stationsmessungen und Radar) für Deutschland liefert. Dieser disaggregierte Datensatz ist neuartig für Deutschland und Voraussetzung für die Verifikation auf der stündlichen Zeitskala in dieser Arbeit. Die Technik der Disaggregation wurde während des MAP (Binder und Schär, 1995) entwickelt.

Voraussetzung für die Disaggregation ist, dass Niederschlagswerte aus Stations- und Radarmessungen auf einem einheitlichen Gitter vorliegen. Für diese Arbeit handelt es sich um das in Abschn. 2.1 beschriebene 7 km-Gitter des aLMO. In Abschn. 3.1 dieses Kapitels wird ausführlich die Gitterung der Stationsdaten nach der Methode von Frei und Schär (1998) beschrieben. In Abschn. 3.2 folgt dann die Mittelung der räumlich und zeitlich hoch-aufgelösten Radardaten auf das aLMO-Gitter. Abschnitt 3.3 zeigt den eigentlichen technischen Teil der Disaggregation. Abschnitt 3.4 zeigt mögliche Fehlerquellen der Disaggregation und den Umgang damit auf. Dabei werden einige statistische Aussagen vor allem zum Verhältnis von täglichen gegitterten Stations- und Radardaten getroffen. Abschließend folgt eine Zusammenfassung (Abschn. 3.5) über die Verwendungsmöglichkeit des disaggregierten Datensatzes für diese Arbeit.

3.1 Gitterung der Stationsmessungen

Um die verschiedenen numerischen Wettermodelle, hier vor allem das aLMO, verifizieren zu können, muss man sich überlegen, wie man die Stationsmessungen mit den Vorhersagen vergleicht. Dabei gibt es grundsätzlich zwei Möglichkeiten:

1. die Beobachtungen auf das Modellgitter interpolieren,
2. die Modelldaten zu den Stationen hin interpolieren.

Für diese Arbeit wird ersteres angewandt, d.h. die Stationsdaten werden mit einer geeigneten Interpolationsmethode auf das aLMO-Gitter gesetzt. Hintergrund dieser

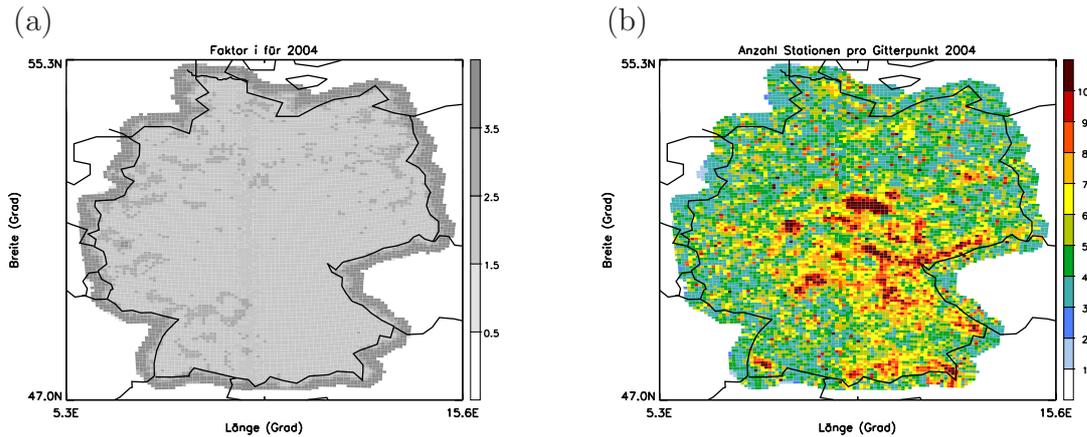


Abbildung 3.1: Statistik zur Gitterung der Stationsdaten auf das aLMO-Gitter: (a) Faktor i als Vielfaches des Gitterpunktabstandes, (b) Anzahl an gefundenen Stationen innerhalb des Suchradius aus (a) ($i \cdot 7 \text{ km}$).

Entscheidung ist vor allem, dass der Modellniederschlag immer ein Gebietsmittel (Im Gegensatz zu einer Punktmessung), nämlich das einer Gitterbox darstellt. Somit ist eine Gitterung der Beobachtungsdaten nicht nur für einen generellen Vergleich sinnvoll, sondern auch für einen fairen Vergleich (Osborn und Hulme, 1997).

Zur Gitterung der Stationsdaten wurde in dieser Arbeit eine modifizierte Version der Methode von Frei und Schär (1998) verwendet. Dieses Interpolationsverfahren zur Erstellung einer Niederschlagsklimatologie für die Alpen basiert auf dem SYMAP-Interpolationsalgorithmus von Shepard (1984) mit einem „Richtungs-Distanz-Gewichtungsverfahren“.

Als Ausgangsdatensatz wurden die Niederschlagsmessungen aller verfügbaren Stationsdaten von 2001-2004 (Abschn. 2.3) verwendet. Im folgenden Teil wird ausführlich die Verfahrensweise der Interpolationsmethode beschrieben.

Es wird für jedes Gebiet um einen aLMO-Gitterpunkt aus den dort vorhandenen Stationsdaten ein Niederschlagswert für diesen Gitterpunkt berechnet. Dabei werden die Stationen mit einem zu wählenden Radius um den Gitterpunkt herum bestimmt und dann auf geeignete Weise gewichtet. Die Gewichtung erfolgt aber nicht alleine aufgrund der Entfernung zum Gitterpunkt, sondern auch bei einer Häufung von Stationen in einer bestimmten Richtung zum Gitterpunkt. In diesem Fall werden die Messungen dieser gehäuft auftretenden Stationen weniger stark gewichtet. Der zweite wichtige Punkt ist, dass der Radius in Abhängigkeit zur regionalen Stationsdichte steht. Es wird eine Mindestanzahl von drei Stationen innerhalb des Suchradius vorausgesetzt. Die von Frei und Schär (1998) verwendete und hier übernommene Gewichtungsfunktion lautet

$$w_d(\tilde{r}_i) = \begin{cases} \frac{1 + \cos(\pi \times \tilde{r}_i)}{2} & \tilde{r}_i \leq l_i \\ 0 & \tilde{r}_i > l_i \end{cases}$$

mit

$$\tilde{r}_i = \sqrt{\left(\frac{x_s - x_g}{i\Delta x}\right)^2 + \left(\frac{y_s - y_g}{i\Delta y}\right)^2}$$

als skalierte Distanz zwischen den Stationskoordinaten (x_s, y_s) und den Koordinaten des aLMO-Gitterpunktes (x_g, y_g) . Der Faktor i der Skalierung wird dabei mit dem Gitterpunktabstand $(\Delta x, \Delta y)$ multipliziert und definiert die Größe des Suchradius um den Gitterpunkt $(i\Delta x, i\Delta y)$, wobei für $i = (1, 2, 3, 4)$ gilt. Ein Wert von z.B. $i = 2$ bedeutet, dass die radiale Gewichtungsfunktion (\tilde{r}) in einer Distanz (l_2) von $2 \cdot \Delta x$ vom Gitterpunkt auf 0 abfällt. Da man aber vom Gitterpunkt aus in beide Richtungen geht, werden Stationswerte über eine Gitterweite von $4 \cdot \Delta x$ in die Berechnung einbezogen. Allerdings nimmt das Gewicht außerhalb eines Radius von $\pm 1 \cdot \Delta x$ stark ab. Die effektive Auflösung des Gitters (bei $i = 2$) ist also zwischen $2 \cdot \Delta x$ und $4 \cdot \Delta x$ (14-28 km) und damit nicht identisch mit der Modellauflösung. Dies muss bei der Verifikation in dieser Arbeit immer berücksichtigt werden.

Abbildung 3.1 (a) zeigt den Faktor i an jedem Gitterpunkt für das Jahr 2004 (repräsentativ für 2001-2003). Überwiegend gilt für Deutschland $i = 2$, in einigen wenigen Regionen $i = 3$. Abb. 3.1 (b) zeigt die Anzahl an gefundenen Stationen pro Gitterpunkt innerhalb des verwendeten Radius für 2004. Die Methode fordert, dass mindestens drei Stationen in die Gitterung eingehen. Sollte dies selbst bei $i = 4$ nicht der Fall sein, wird der Gitterpunkt nicht besetzt. Dies kommt allerdings aufgrund des sehr dichten Messnetzes in Deutschland (Abschn. 2.3) nicht vor. Man sieht, dass die Erfüllung der Minimalforderung von drei Stationen innerhalb des Suchradius nur dann für alle Gitterpunkte möglich ist, wenn $i \leq 3$. Für Regionen mit dichterem Stationsnetz hätte theoretisch $i = 1$ ausgereicht. Allerdings hätte dies den Nachteil, dass die effektive Auflösung stark variieren würde. Das ist für die Anwendung in dieser Arbeit nicht wünschenswert, weshalb in den Einstellungen zur Gitterung ein minimaler Wert von $i = 2$ gesetzt wurde.

Ein Problem der Verteilung von Niederschlagsmessstationen ist, dass diese meist in Tallagen, weniger in höheren Regionen der Gebirge stationiert sind. Dies führt dazu, dass eine Gitterung zur Unterschätzung des Niederschlages neigt, da dieser im Mittel mit der Höhe steigt (Ekhart, 1948; Peck und Brown, 1962). Dieses Verhalten wird in der Methode nach Shepard (1968, 1984) nicht berücksichtigt. Um diese Problematik aber hier zu behandeln, wurde in das Gitterungsverfahren von Frei und Schär (1998) zusätzlich die Niederschlagsklimatologie (Abschn. 2.6, Müller-Westermeier, 1995) mit einbezogen. Dabei werden im Endeffekt die relativen Abweichungen zur Klimatologie gegittert, also Tagesniederschläge mit Monatsmitteln kalibriert (Widmann und Bretherton, 2000; New et al., 2000). Nach der Gitterung dieser Abweichungen wird die Klimatologie am entsprechenden Gitterpunkt wieder multipliziert. Bei der Gitterung insbesondere des Niederschlages als nicht kontinuierlich verteilte Messgröße muss man einige Dinge beachten. So werden durch die Gitterung mehrerer Stationsmessungen die Niederschlagsintensitäten abgeschwächt, zum anderen aber die Häufigkeit der Niederschlagsereignisse erhöht (Frei et al., 2003). Weiter ist zur Interpretation der später in dieser Arbeit aufgeführten Verifikationsergebnisse zu beachten, dass bereits die Stationsdaten (wie alle Messdaten, besonders aber Niederschlagsdaten) Fehlerquellen und Unsicherheiten aufweisen (Abschn. 2.3), aber dennoch die „bestmögliche Wahrheit“ darstellen.

Reflektivität dBZ	Niederschlag mm/h	Klassenmittel mm/h	Summe in mm/15 min
0 - 7	0 - 0.1	0	0
7 - 19	0.1 - 0.5	0.4	0.1
19 - 28	0.5 - 1.9	1.2	0.3
28 - 37	1.9 - 8.1	3.6	0.9
37 - 46	8.1 - 34	10	2.5
46 - 55	34 - 149	56	14
> 55	> 149	160	40

Tabelle 3.1: Niederschlagsraten aus Radarreflektivitäten (PC-Produkt des DWD): Reflektivitätsklasse (1.Spalte), Niederschlagsklasse (2.Spalte), angenommene mittlere Niederschlagsrate/h (3.Spalte), daraus folgende Niederschlagsrate in mm/15 min (4.Spalte).

3.2 Räumliche Mittelung des Radarkomposits

Das für diese Arbeit vom DWD zur Verfügung gestellte Radarkomposit für Deutschland ist das sogenannte PC-Produkt. In Tab. 3.1 sind die sechs Reflektivitätsklassen dieses Produkts aufgeführt, dazu noch die „Reflektivitätsklasse“ ohne Rückstreusignal (0 dBZ). Die Einteilung in sieben Reflektivitätsklassen wurde vom DWD wegen des Einsparens von Speicherkapazität vorgenommen.

Wie in Abschn. 2.5 beschrieben, benötigt man zur Umrechnung von Reflektivitäten zu Niederschlägen eine zR-Beziehung, welche hier die vom DWD operationell verwendete ist:

$$z = 256R^{1.42}. \quad (3.1)$$

In Tab. 3.1 sind die aus den Reflektivitätsklassen errechneten Niederschlagsklassen in mm/h aufgeführt. Nun sind jeweils Niederschlagsklassen immer einer Reflektivitätsklasse zugeordnet, das Radar kann natürlich aber auch alle Werte zwischen den Grenzen einer Reflektivitätsklasse messen. Es entsteht das Problem, dass nicht jede beliebige Reflektivität einem genauen Niederschlagswert zugeordnet ist. Um dieses Problem zu lösen, wird für jede Klasse eine typische mittlere Niederschlagsrate angenommen. Die hier verwendeten Werte sind in Tab. 3.1 dargestellt. Die Niederschläge aus dem Radarkomposit werden immer in mm/h angegeben. Zurückgerechnet auf viertelstündige Niederschlagssummen bedeutet das, dass genau die in der vierten Spalte von Tab. 3.1 resultierenden Werte vorkommen können (Stundensummen jeweils durch vier geteilt). Diese nun erhaltenen 15-minütigen Niederschlagsfelder sind wiederum die Ausgangswerte, um die Niederschlagssumme für eine Stunde zu errechnen. Dabei wurde die Niederschlagssumme von 6:30-6:30 UTC, 7:30-7:30 UTC, usw. gebildet. Das bedeutet, dass man für jedes Radarpixel aus dem Komposit vier Niederschlagsraten hat, aus denen die jeweilige Stundensumme errechnet wird.

Ein Beispiel: In der ersten Viertelstunde fällt kein Niederschlag, in der zweiten

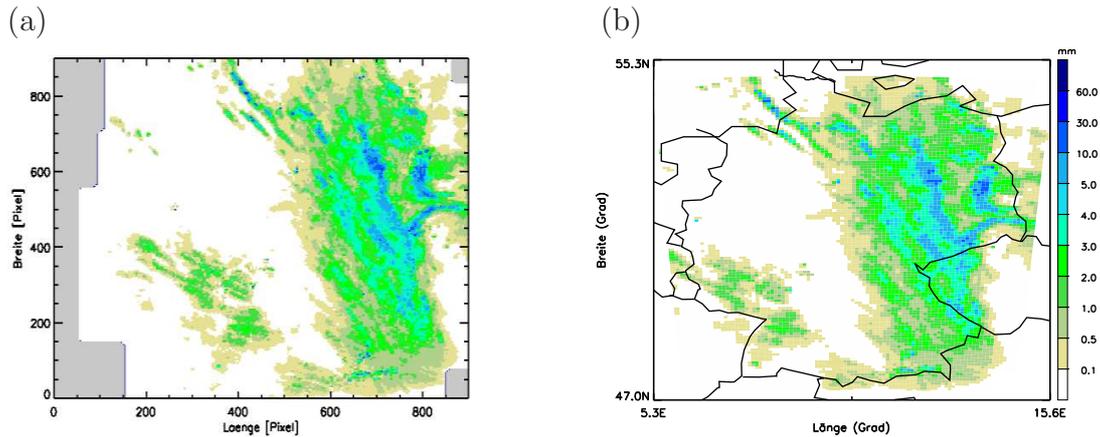


Abbildung 3.2: Beispiel zur Mittelung der Radardaten auf das aLMO-Gitter: (a) Radar-komposit in Originalauflösung auf einem kartesischen Gitter, (b) gegittertes Radarkomposit auf einem stereographischen Gitter.

0,3 mm, in der dritten 2,5 mm und in der letzten Viertelstunde noch einmal 2,5 mm. Die Niederschlagssumme für diese Stunde und dieses eine Radarpixel beträgt dann 5.3 mm.

Diese Stundensummen-Berechnung wurde im Vorfeld dieser Arbeit von Martin Hagen am DLR (Deutsches Zentrum für Luft- und Raumfahrt) vorgenommen. Der Datensatz aus dem Radarkomposit besteht schlussendlich aus stündlichen Niederschlagssummen auf einem 4·4 km Gitter (geographische Koordinaten) für ganz Deutschland.

Um die Radardaten nun auf das hoch-aufgelöste, identische Gitter wie die Stationsdaten am Boden (aLMO-Gitter, Abschn. 3.1) zu setzen, wurde einfach das arithmetische Mittel aller Radarpixel, die zu dem jeweils am nächsten liegenden aLMO-Gitterpunkt gehören, für jede Stunde berechnet. In Abb. 3.2 ist anhand eines Beispiels die Gitterung der Radarpixel dargestellt. Man erkennt gut, dass die Auflösung von 4 km im originalen Radardatensatz mit dem Setzen der Werte auf das aLMO-Gitter (7 km) etwas abnimmt. Zum anderen ist aber auch gut zu sehen, dass diese einfache Mittelung der Radardaten ein sehr ähnliches Bild wie in den Originaldaten ergibt.

3.3 Technik der Disaggregation

Die bereits am Anfang dieses Kapitels erwähnte Disaggregation wird hier detailliert erläutert. Genauer gesagt handelt es sich um eine zeitliche Disaggregation. Dabei ist es Voraussetzung, dass die benötigten Datensätze aus Stationsmessungen und des Radarkomposits für Deutschland auf dem gleichen Gitter vorliegen zu haben, wie es in den vorhergehenden Abschn. 3.1 und 3.2 durchgeführt wurde.

Wie in Abb. 3.3 schematisch dargestellt, hat die Disaggregation zum Ziel, einen zeitlich und räumlich hoch-aufgelösten Datensatz aus Beobachtungen zu erzeugen. Die Idee dabei ist, die jeweiligen Vorteile der Stations- und Radardaten zu kombinieren. Aus den (gemittelten) Radardaten bekommt man die hohe zeitliche Auflösung,

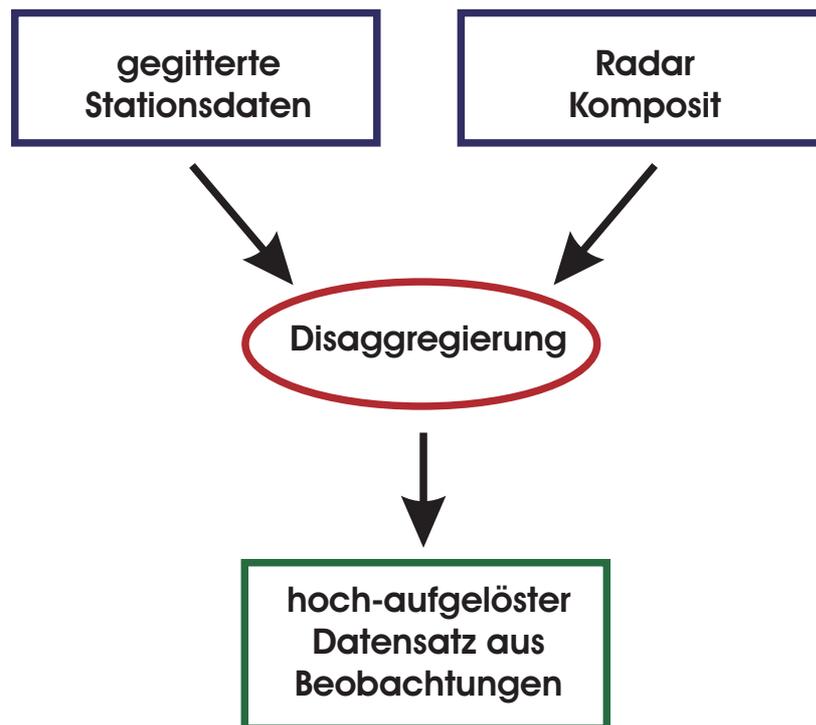


Abbildung 3.3: Schema zur Disaggregation.

aus den (gegitterten) Stationsdaten die genaue Niederschlagsmenge. Der mathematische Weg der Disaggregation ist einfach und für den Niederschlag $R(i, j)$ an jedem Gitterpunkt (i, j) wie folgt definiert:

$$R_{dis,h}(i, j) = R_{rad,h}(i, j) \cdot \frac{R_{obs,d}(i, j)}{R_{rad,d}(i, j)}, \quad (3.2)$$

wobei rad, h den Index für die stündliche Radarsumme bezeichnet, rad, d die Tagessumme des Radars, obs, d die Tagessumme der gegitterten Bodendaten und dis, h die resultierende, stündliche Niederschlagssumme. Die Disaggregation ist anhand eines Beispiels vom 12. August 2002 in Abb. 3.4 dargestellt. Die Bilder (a) und (b) zeigen die gemittelten Radarniederschläge auf dem 7 km-Gitter (stündliche und tägliche Summe), Bild (c) die Tagessumme aus den gegitterten Stationsdaten. Das aus der Disaggregation resultierende Feld in Bild (d) zeigt die stündliche Niederschlagssumme für den gleichen Zeitraum wie Bild (a). Vergleicht man nun (d) mit (a), so leuchtet zunächst ein, dass in der Mitte und in Teilen Westdeutschlands kein Niederschlag in (d) sein kann, wenn auch kein Niederschlag in (a) vorhanden ist (Gl. 3.2). Insgesamt sollte im Vergleich von (d) mit (a) zumindest die Niederschlagsstruktur sehr ähnlich sein. Dies ist auf den ersten Blick auch der Fall.

Schaut man an diesem Tag auf den niederschlagsreichen Osten Deutschlands, so fällt auf, dass (d) im Vergleich zu (a) in der Fläche etwas weniger, in den Maxima aber etwas mehr Niederschlag hat. Die höheren Maxima erklären sich damit, dass die Niederschläge in der Tagessumme am Boden (c) im Vergleich zum Radar (b) höher sind. Im Südwesten dagegen fällt auf, dass in (d) weniger Niederschlag aus der Disaggregation folgt, als in (a) zu erkennen ist. An diesem Tag ist für diese Region

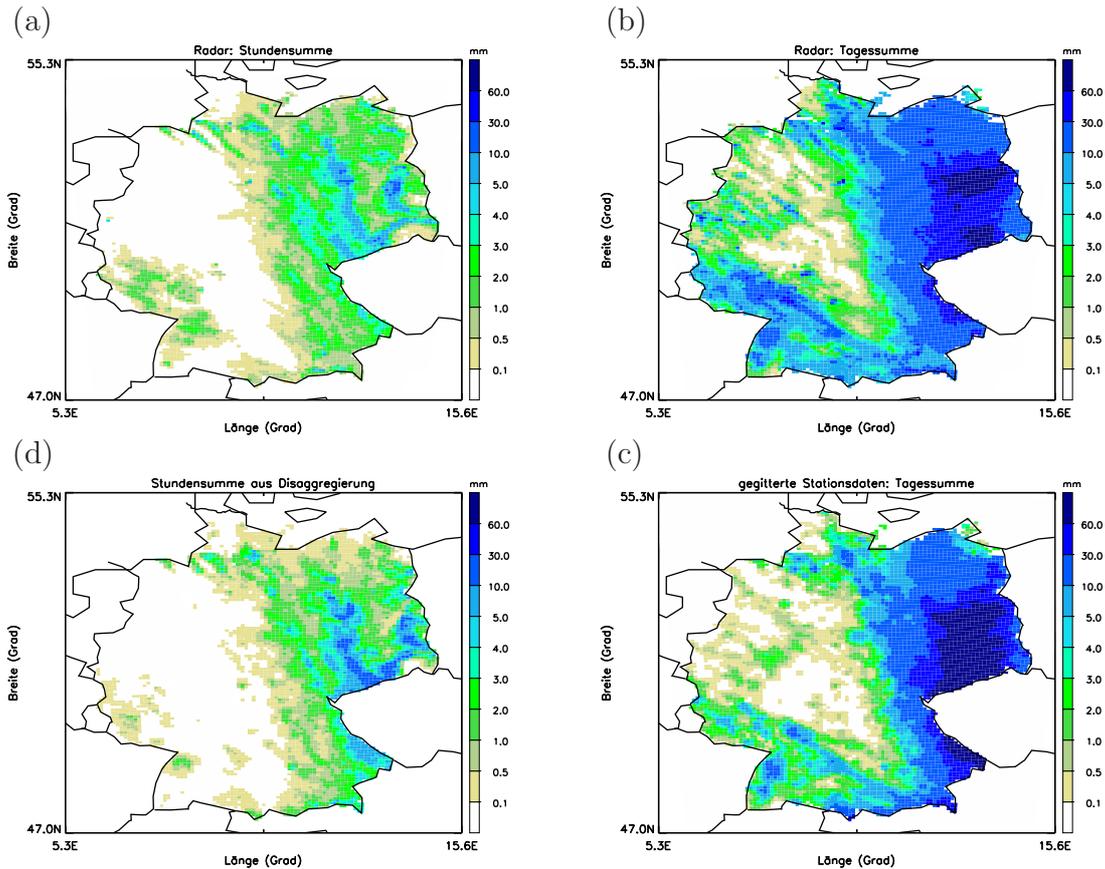


Abbildung 3.4: Beispiel zur Disaggregation (12. August 2002, 7km-Gitter): Gl. 3.2 ist hier gleichbedeutend mit $(d)=(a) \cdot (c)/(b)$ (an jedem Gitterpunkt), (a) stündliche Radarsumme, (b) Tagessumme Radar, (c) Tagessumme Bodenmessungen, (d) disaggregierte stündliche Niederschlagsmenge (gleicher Zeitpunkt wie (a)).

eine Überschätzung der Niederschlagsmenge im Radar sichtbar. Warum auch in der räumlichen Struktur die Bilder (d) und (a) nicht identisch sind, kann nur mit Blick auf die restlichen 23 einzelnen Bilder der stündlichen Radarsummen geklärt werden. Der relativ hohe Niederschlag im Südwesten im Radar verteilt sich offensichtlich nur zu geringen Teilen auf die ausgewählte Stunde. Vermutlich werden einige andere Stundensummen (zumindest eine) deutlich höhere Werte aufzeigen (im Radar), womit nach Gl. 3.2 auch entsprechend hohe Anteile am Tagesniederschlag dorthin disaggregiert werden. Näheres zu dieser Problematik und anderen Fehlerquellen folgt in Abschn. 3.4.

Hat man schlussendlich für jeden Tag und an jedem Gitterpunkt die Disaggregation durchgeführt, so erhält man (wie in Abb. 3.4 (d) für eine Stunde dargestellt) einen stündlichen Datensatz mit der hohen zeitlichen Auflösung aus den Radardaten und der genauen Niederschlagsmenge aus den Stationsdaten.

3.4 Fehlerquellen der Disaggregation

Wie im vorhergehenden Abschnitt schon angedeutet, hat die Technik der Disaggregation Fehlerquellen, die zu diskutieren und quantifizieren sind. Für die Dis-

		Niederschlagsmessung		
		Missing	0	>0
Radarmessung	Missing	0 %	4 %	4 %
	0	0 %	23 %	4 %
	>0	0 %	15 %	54 %

Abbildung 3.5: Grafik der verschiedenen möglichen Fälle (Litschi, 2005) und deren prozentualer Anteil bei der Disaggregation.

aggregation entscheidend und gut zu untersuchen ist die Beziehung zwischen den Tagessummen aus gegitterten Stations- und Radardaten, die in Gl. 3.2 als Verhältnis

$$C(i, j) = \frac{R_{obs,d}(i, j)}{R_{rad,d}(i, j)} \quad (3.3)$$

vorkommt. Mit diesem Verhältnis lassen sich einige einfache Untersuchungen zur Genauigkeit der Niederschlagsmengen des Radardatensatzes machen. Wie bereits in den Abschn. 2.3 und 2.5 beschrieben, haben beide Beobachtungsdatensätze ihre Fehlerquellen. Da aber an den Stationsdaten mehrfach Qualitätsprüfungen vorgenommen und bei der räumlichen Analyse direkt vor der Gitterung lokale Zufallseffekte zum Teil herausgemittelt wurden, werden die gegitterten Stationsdaten in dieser Arbeit als „bestmögliche Wahrheit“ angesehen. Deutlich mehr Unsicherheiten liegen in der Radarmessung, weshalb hier systematische Abweichungen in der Niederschlagsmenge erwartet werden.

Untersucht man zunächst die Tagessummen $R_{obs,d}(i, j)$ und $R_{rad,d}(i, j)$ in Bezug zueinander (dabei wird im Folgenden immer von den Datensätzen mit 7 km-Auflösung gesprochen), so bekommt man neun mögliche Fälle, die in Abb. 3.5 (Litschi, 2005) in Form von farbigen Boxen dargestellt sind:

Rot: Fälle, in denen gegitterte Stationsdaten (an einen Gitterpunkt) nicht verfügbar sind. Dieser Fall tritt allerdings aufgrund der hohen Dichte der Niederschlagsstationen in Deutschland nicht auf.

Blau: Aufgrund von Ausfällen eines oder mehrerer Radare kann es vorkommen, dass keine Radardaten verfügbar sind (4% der Fälle). Zum anderen gibt es den Fall, dass das Radar keinen Niederschlag misst, jedoch am Boden Niederschlag beobachtet wird (meist durch Radarabschattung, 4%).

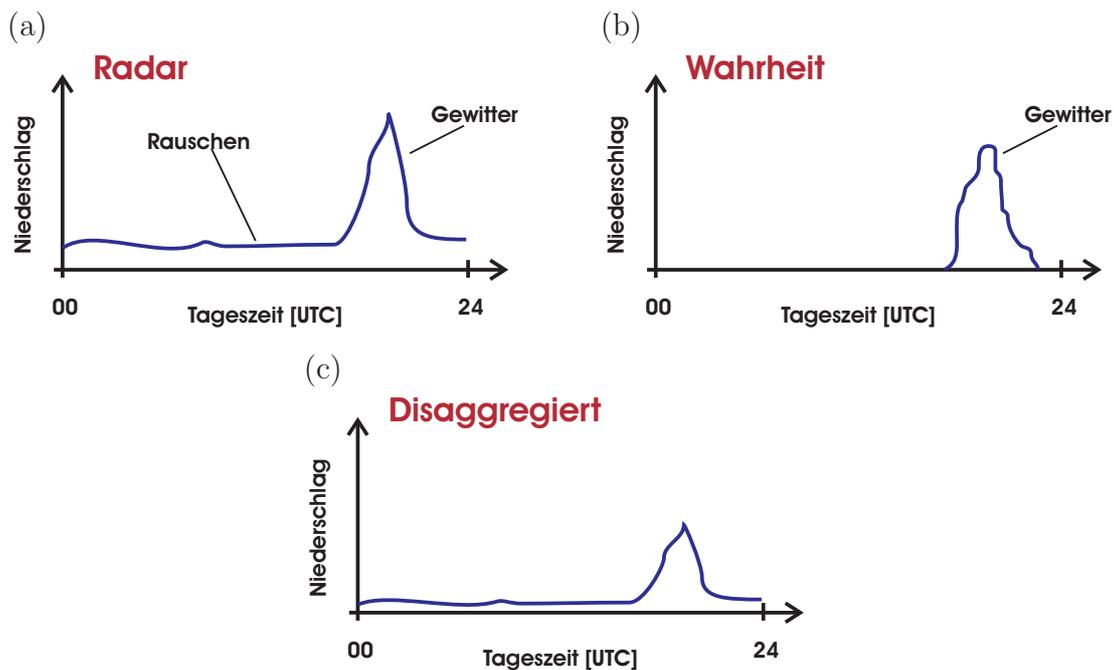


Abbildung 3.6: Skizze zur Verdeutlichung des Problems der Disaggregation (an einem Gitterpunkt), dass wegen Radarrauschens der tatsächlich gefallene Niederschlag über den Tag fehlerhaft verteilt wird: (a) Radarsignal, (b) Wahrheit (nicht bekannt, da am Boden nur die 24 h-Summe gemessen wird), (c) Ergebnis aus Disaggregation.

Dies sind alle Fälle, in denen nicht disaggregiert werden kann und somit Lücken im disaggregierten Datensatz entstehen.

Grün: Fälle, für die die Disaggregation durchgeführt werden kann. In dem Fall, dass gegitterte Stations- und Radardaten Niederschlag anzeigen, kann direkt gemäss Gl. 3.2 verfahren werden (54%). Wenn das Radar Niederschlag misst, aber am Boden kein Niederschlag gefallen ist (Verdunstung, Schmelzschicht, Bodenechos (Abschn. 2.5): 15%), werden alle Stundensummen an diesem Tag auf 0 mm gesetzt (für den entsprechenden Gitterpunkt). In 23% aller Fälle haben beide Datensätze keinen Niederschlag (alle Stundensummen werden auf 0 mm gesetzt).

Insgesamt ist die Disaggregation in Bezug zu allen Gitterpunkten der Zeitreihe 2001-2004 folglich zu 92% durchführbar.

Eine weitere Fehlerquelle ist gegeben, obwohl gegitterte Stations- und Radardaten beide Niederschlag anzeigen (Abb. 3.5, grüne Box rechts unten). Dieses Problem ist anhand einer Skizze in Abb. 3.6 für einen Gitterpunkt an einem Tag erläutert. Dabei zeigt Bild (a) den Verlauf des Radarsignals über einen Tag, der aus angenommenem Rauschen und einem klaren (Gewitter-)Signal besteht. Nun liegt aus den gegitterten Stationsdaten nur die Information über die Tagessumme, nicht aber über die Verteilung für den Tag vor. Bild (b) zeigt nun den wahren Verlauf des Niederschlags am gleichen Gitterpunkt wie beim Radar an. Die einzige Information darüber in

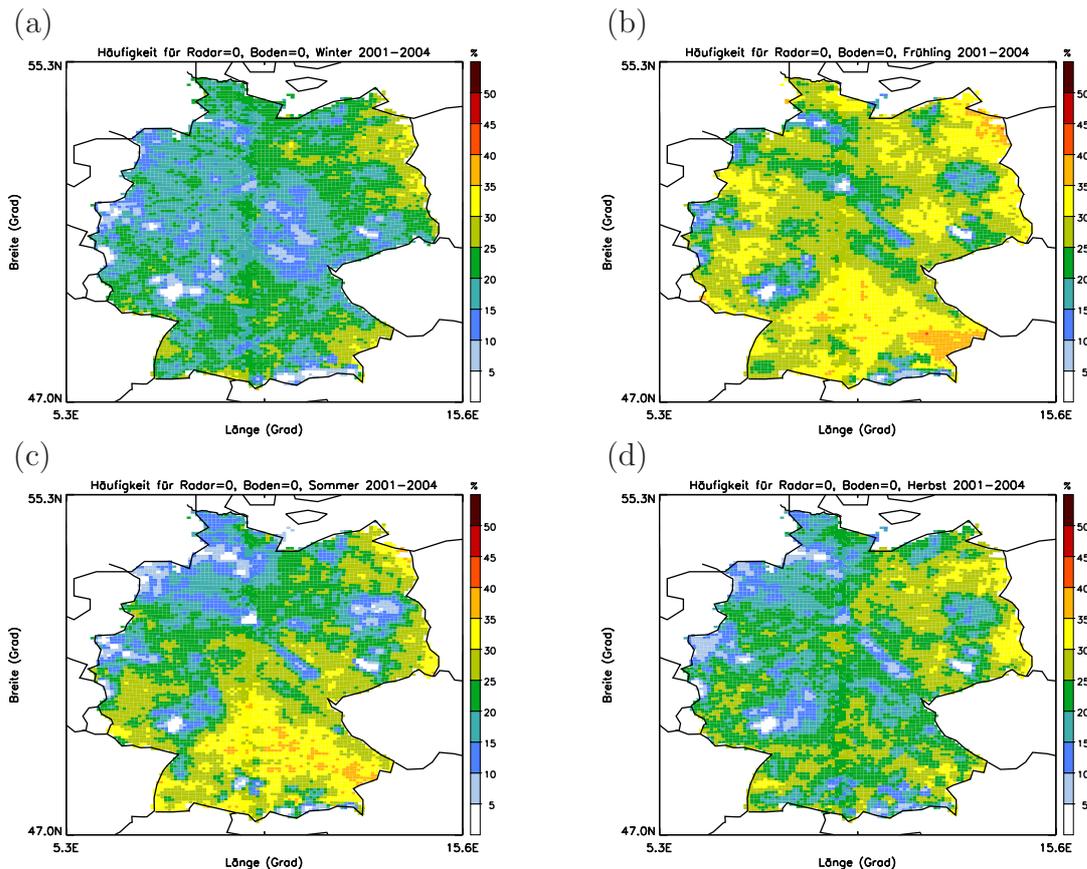


Abbildung 3.7: Statistik über die Jahre 2001-2004 für jeden Gitterpunkt für den Fall gegitterte Stationsdaten = Radar = 0 mm in %: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

den gegitterten Stationsdaten ist aber, dass eben eine gewisse Summe an Niederschlag in 24 Stunden gefallen ist. Die Disaggregation erzeugt aus den vorhandenen Informationen für diesen Tag und Gitterpunkt den Niederschlagsverlauf wie in Bild (c) gezeigt. Das Gewittersignal ist klar und am richtigen Zeitpunkt erkennbar, jedoch in etwas abgeschwächter Form, da aufgrund des Radarrauschens ein Teil der Tagessumme am Boden auf den restlichen Tag verteilt wurde. Dieses Problem ist nicht einfach zu umgehen und muss in dieser Arbeit immer im Hinterkopf behalten werden. Allerdings ist das hier als Fehlerquelle identifizierte Radarrauschen nicht flächendeckend und ständig vorhanden, es bedeutet in der Regel auch nur sehr kleine Niederschläge. Außerdem kommt es auch nur dann zu Problemen, wenn Rauschen und Niederschlag am Boden an einem Tag zusammentreffen. Oft ist es der Fall, dass dieses Radarrauschen aus Bodenechos oder ähnlichem herrührt, aber gar kein Niederschlag fällt (was für die Disaggregation weiter kein Problem ist). Eine größere Problematik besteht nur dann, wenn wie in Abb. 3.6 das wahre Niederschlagssignal tatsächlich nur einen kurzen Zeitraum abdeckt und am gleichen Tag deutliches Rauschen vorhanden ist.

In den nun folgenden Vergleichen zwischen Tagessummen aus gegitterten Stations- und Radardaten gehen alle verfügbaren Daten der gesamten Zeitreihe (2001-2004) ein. Abbildung 3.7 zeigt für jede Jahreszeit den prozentualen Anteil an Tagen, an denen es in beiden Datensätzen zum gleichen Zeitpunkt keinen Niederschlag gab

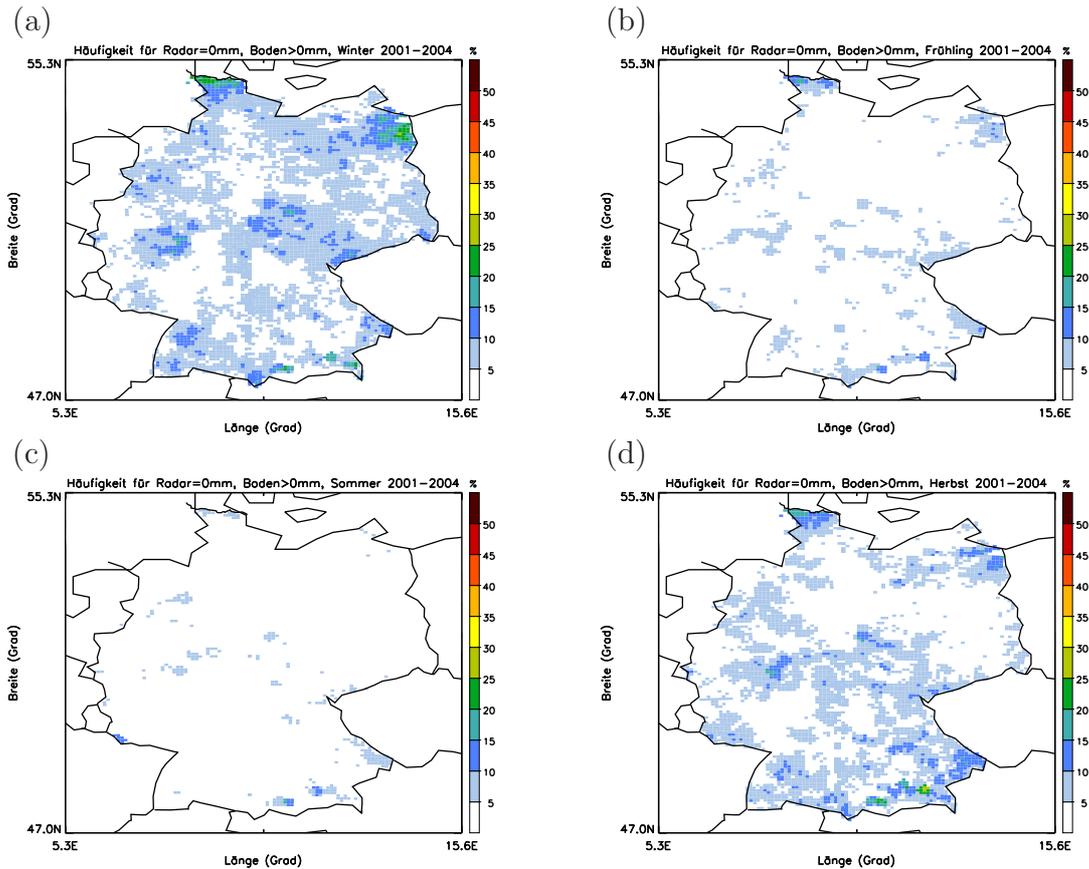


Abbildung 3.8: Wie Abb. 3.7 für den Fall gegitterte Stationsdaten > 0 mm, Radar = 0 mm, der in erster Linie durch Radarabschattung bedingt ist.

($R_{obs,d}(i, j) = R_{rad,d}(i, j) = 0$ mm/Tag). Man sieht, dass dieser Fall am häufigsten (bis zu 45 %) im Frühling auftritt und im Sommer ähnliche Häufungen im Süden Deutschlands zu finden sind. Somit finden sich im Herbst und insbesondere im Winter die wenigsten dieser Fälle. Auffällig sind in allen Jahreszeiten Regionen mit 0-5 % niederschlagsfreier Tage in der Nähe einiger Radarstandorte. Dies ist gut mit Bodenechos in der Nähe dieser Radare erklärbar, die quasi ständig zu leichten Reflektivitäten (und damit Niederschlägen) führen.

Abbildung 3.8 zeigt die Fälle mit $R_{obs,d}(i, j) > 0$ mm und $R_{rad,d}(i, j) = 0$ mm für alle Jahreszeiten. Diese Fälle treten in der Mehrheit im Winter und Herbst auf, sind also im Frühling und Sommer fast kein Thema. Gründe dafür sind

- Radarabschattung, d.h. bei tiefen Wolken und nicht hochreichend konvektiven Niederschlagssystemen in der vom Radar abgewandten Seite von Hindernissen (Gebirge),
- Nebel, den das Radar nicht sieht, der aber als Niederschlag am Boden gemessen werden kann.

Der Effekt von Abschattung oder Nebel für die Disaggregation ist quasi umgekehrt dem aus Abb. 3.6. In dem Fall, dass einige Stunden im Radar kein Niederschlag gemessen wird, obwohl eventuell Niederschlag am Boden vorkommt, wird eine

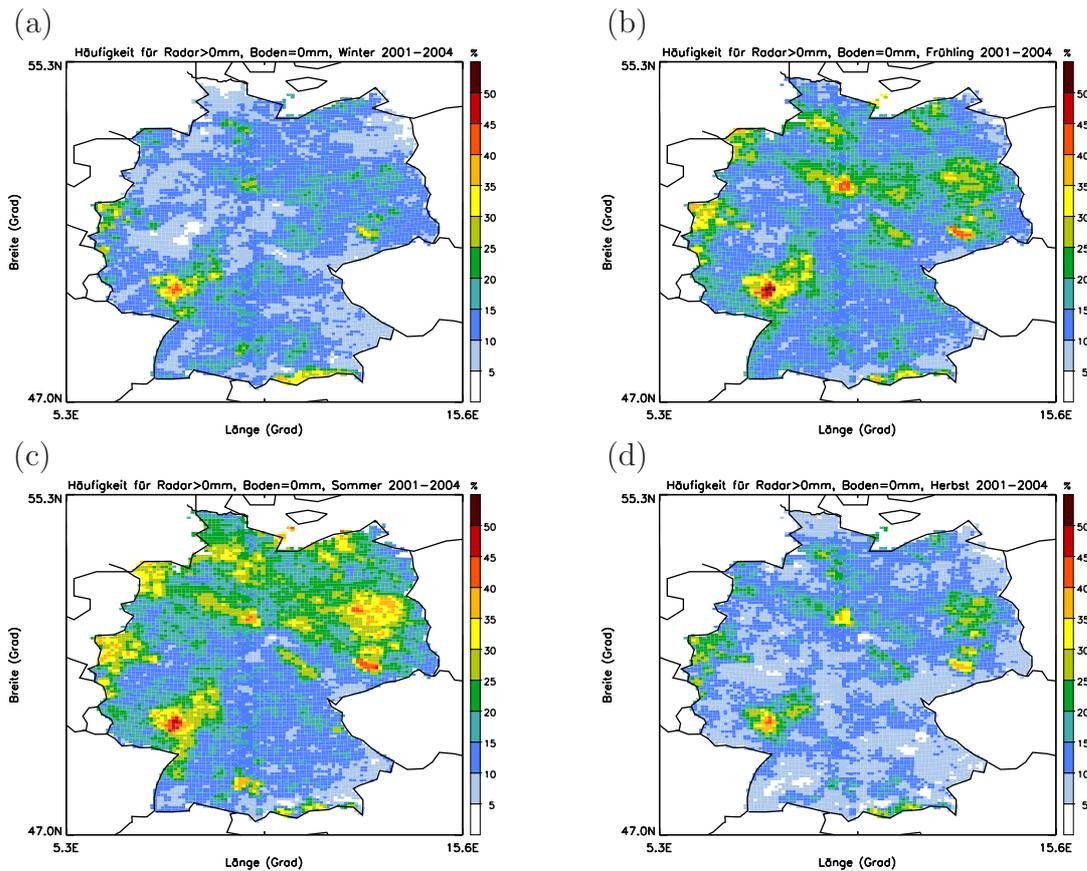


Abbildung 3.9: Wie Abb. 3.7 für den Fall gegitterte Stationsdaten = 0 mm, Radar > 0 mm (Erklärungen im Text).

Verstärkung einzelner Niederschlagssignale die Folge sein. Wie zuvor in Abb. 3.5 gezeigt, sind diese Fälle in Deutschland aber relativ selten. Leicht ausgeprägte Maxima (bis 15 %) finden sich wie zu erwarten in den Mittelgebirgsregionen und in Richtung der Alpen. Vor allem im Winter werden aber auch zwei größere Maxima im Norden und im Nordosten deutlich. Dies sind genau die beiden Regionen, in denen die Radarabdeckung qualitativ nicht ausreichend ist (Abschn. 2.5).

Der zu Abb. 3.8 umgekehrte Fall ist in Abb. 3.9 dargestellt ($R_{obs,d}(i, j) = 0$ mm und $R_{rad,d}(i, j) > 0$ mm). Diese Fälle kommen flächendeckend zwischen 10 und 25 % vor, im Frühling und vor allem Sommer auch bis 30 %. Maxima bis über 50 % finden sich direkt in der Umgebung der Radarstandorte, sowie ein weiteres ausgeprägtes Maximum im Südwesten (gleiche Regionen wie für Minima in Abb. 3.7). Für diesen Fall gibt es verschiedene Erklärungen:

- Bodenechos in den Radardaten,
- Probleme der Radarmessung in der Schmelzschicht,
- Problem mit dem unteren Schwellwert beim Radar: das Radar erkennt eventuell Wassertropfen und identifiziert diese als Niederschlag, welche aber gar nicht den Boden erreichen (bei 7 dBZ Radarreflektivität ist ziemlich sicher, dass die daraus errechnete Niederschlagsmenge nicht den Boden erreicht),

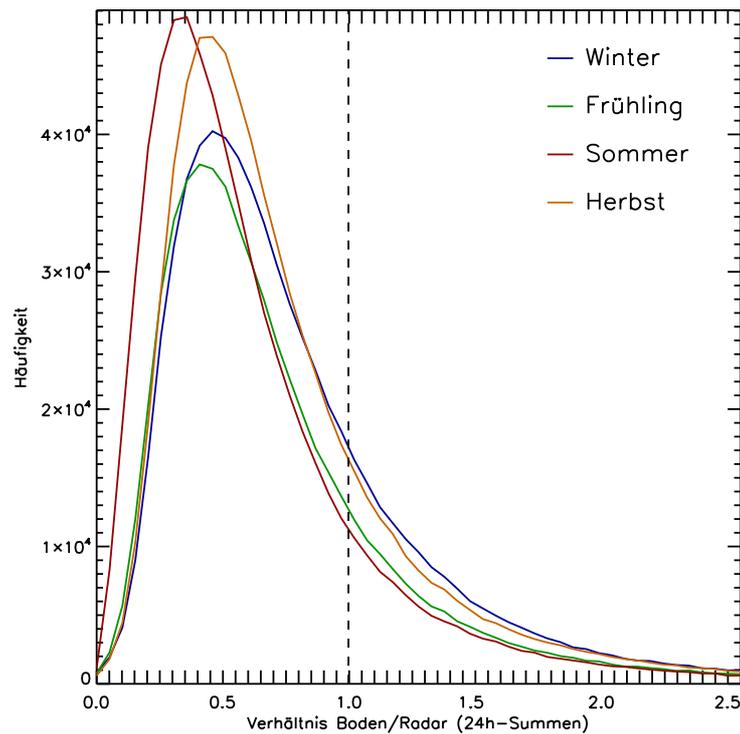


Abbildung 3.10: Häufigkeitsverteilung für das Verhältnis gegitterte Stationsdaten/Radar für 24 h-Summen (an jedem Gitterpunkt) aufgeteilt in Jahreszeiten der Zeitperiode 2001-2004. Es gehen alle Werte mit gegitterten Stationsdaten ≥ 2 mm und Radar > 0 mm ein.

- ähnlich der vorherigen Erklärung die Verdunstung von fallenden Niederschlagspartikeln.

Diese vielfältigen Probleme aus der Radarmessung sind für die Disaggregation glücklicherweise nicht von Bedeutung. Da die Tagessumme am Boden keinen Niederschlag angibt, werden automatisch auch alle resultierenden Stundensummen gleich null gesetzt ($R_{dis,h}(i,j) = 0$). Zu erwähnen ist, dass mit diesen Fehlerquellen die Maxima an den Radarstandorten erklärbar sind, nicht aber das stark ausgeprägte Maximum im Südwesten. Es ist schwer festzustellen, warum gerade dort das Radar Niederschlag sieht, wenn doch am Boden kein Niederschlag gemessen wurde. Über die genauen Gründe (Ballungsraum Rhein-Main-Gebiet, das Frankfurter Radar hat schlechte Qualität) kann an dieser Stelle nur spekuliert werden.

Um nun ein Gefühl dafür zu bekommen, ob das Radar generell Niederschläge überschätzt, unterschätzt oder es keinen eindeutigen Trend gibt, wurde das Verhältnis $C(i,j)$ (Gl. 3.3) untersucht. Dabei wurden nur Fälle mit $R_{obs,d}(i,j) \geq 2$ mm/Tag und $R_{rad,d}(i,j) > 0$ mm/Tag berücksichtigt, damit ein möglicher Zusammenhang eindeutig den größeren Niederschlagsmengen zugeordnet werden kann (bei geringeren Niederschlägen ist die absolute Differenz zwischen gegitterten Stations- und Radardaten nicht sehr groß).

Abbildung 3.10 zeigt die Häufigkeiten für alle Jahreszeiten in Abhängigkeit von $C(i,j)$ für die ganze Zeitperiode 2001-2004. Der Wert $C(i,j) = 1$ steht für identische Tagessummen in beiden Datensätzen. Für $C(i,j) > 1$ gilt, dass das Radar den

Niederschlag unterschätzt, bei $C(i, j) < 1$ entsprechend überschätzt. Das Maximum für alle Jahreszeiten liegt deutlich im Bereich der Überschätzung bei 0.45 in Winter, Frühling und Herbst, für die Sommermonate sogar bei 0.35. Dies bedeutet eine Überschätzung des Niederschlags um mehr als die doppelte Menge. Unterschätzungen kommen zwar vor, sind aber wesentlich seltener, womit ein eindeutiger Zusammenhang der beiden Tagessummen klar wird. Dieser Zusammenhang stellt allerdings ebenfalls kein großes Problem für die Disaggregation dar, da beim Disaggregieren die „genaue Menge“ aus den gegitterten Stationsdaten gewonnen wird.

Das Problem dieser Überschätzung liegt in der ausführlich beschriebenen Schwierigkeit, die Reflektivitäten in Niederschlagsmengen umzurechnen (auch Reflektivitäten, die nicht von Niederschlagspartikeln kommen). Die in Abschn. 3.2 beschriebenen mittleren Niederschlagswerte für jede Reflektivitätsklasse können einen weiteren Anteil an der deutlichen Überschätzung ausmachen. Die mittleren Klassenwerte sind Erfahrungswerte und müssen von Klasse zu Klasse einen mehr oder weniger großen Niederschlagsbereich abdecken (z.B. gilt zwischen 1.9-8.1 mm immer der Klassenmittelwert 3.6 mm, eventuell ist dieser aber etwas zu hoch gewählt).

3.5 Zusammenfassung der Disaggregation

Ziel der Disaggregation ist es, einen Datensatz aus Beobachtungen in zeitlich und räumlich hoher Auflösung zu erstellen. Dabei spielte für diese Arbeit die entscheidende Rolle, dass Niederschlagsvorhersagen des aLMO verifiziert werden sollten. Das aLMO produziert stündliche Vorhersagen mit einer räumlich Auflösung von 7 km. Genau diesen Vorgaben in der Auflösung sollte daher auch der disaggregierte Datensatz entsprechen. Dazu wurden Stationsdaten und Radardaten für Deutschland der Jahre 2001-2004 kombiniert, indem diese zunächst mit geeigneten Methoden (Abschn. 3.1 und 3.2) auf das 7 km-Gitter gesetzt wurden. Die eigentliche zeitliche Disaggregation kombinierte nun die beiden gegitterten Beobachtungsdatensätze so, dass die „genaue Niederschlagsmenge“ aus den gegitterten Stationsdaten, die hohe zeitliche Auflösung aus den Radardaten entnommen wurde.

Beide Datensätze beinhalten systematische Fehlerquellen basierend auf den Messmethoden. Dabei wurde gezeigt, dass die deutlich größeren und vielfältigeren Fehlerquellen bei der Radarmessung liegen, weshalb die Stationsdaten (nach beschriebenen Qualitätskontrollen) als „bestmögliche Wahrheit“ angesehen werden können. Aufgezeigt wurden einige Schwächen der Disaggregation bzw. Fälle, in denen aufgrund der Datenlage oder -qualität nicht disaggregiert werden konnte (8%). Auch wurde darauf hingewiesen, dass es Fälle gibt, in denen zwar disaggregiert wird, jedoch das Ergebnis mit Vorsicht betrachtet werden muss (Abb. 3.6).

Da die meisten und bedeutendsten Fehlerquellen in der Überschätzung der Niederschlagsmenge im Radar liegen (Abb. 3.10), spielt gerade diese Fehlerquelle für den disaggregierten Datensatz keine Rolle, da diese Information nicht in diese Methode eingeht.

Als letzte Qualitätsprüfung des resultierenden disaggregierten Datensatzes soll ein Vergleich zwischen Intensitäten zwischen Radar, disaggregierten Datensatz und Mes-

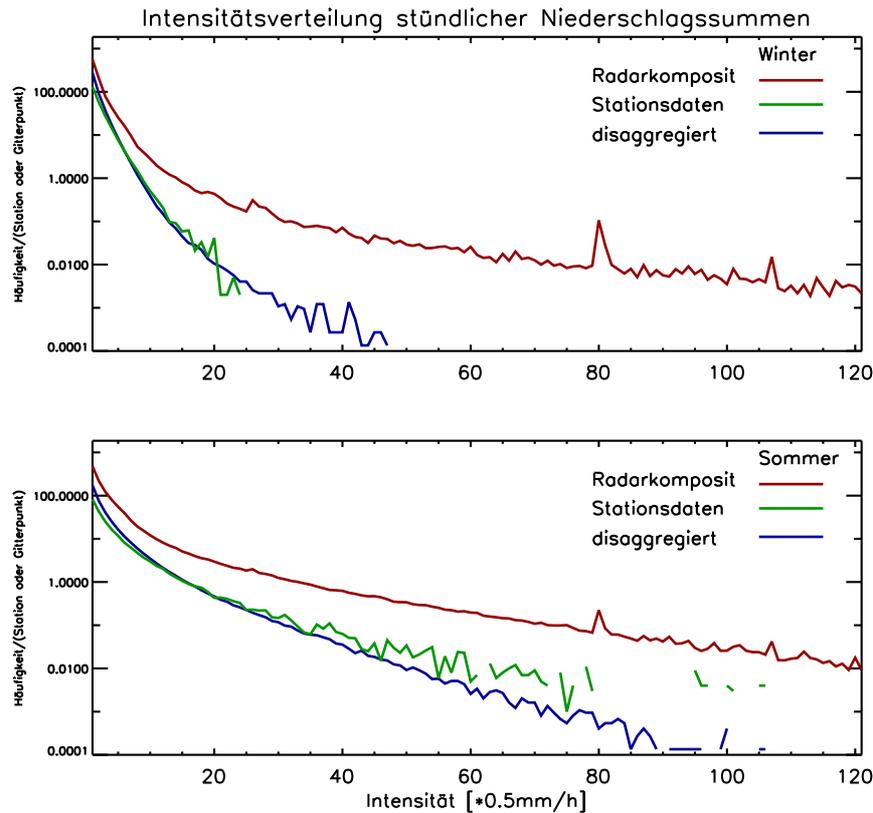


Abbildung 3.11: Häufigkeitsverteilung für stündliche Niederschlagsintensitäten für Radar (rot), disaggregierten Datensatz (blau) und stündliche Messungen (grün): Winter (oben) und Sommer (unten) 2001-2004. Die Häufigkeiten wurden pro 0.5 mm-Intensitätsintervall aufsummiert (und mit Stations- bzw. Gitterpunktzahl normiert). Intervall 1: $0.1 \text{ mm/h} \leq \text{Niederschlagswert} < 0.5 \text{ mm/h}$, Intervall 2: $0.5 \text{ mm/h} \leq \text{Niederschlagswert} < 1 \text{ mm/h}$, usw.

sungen (Abschn. 2.4) auf stündlicher Basis betrachtet werden. Abb. 3.11 zeigt diese Intensitätsverteilungen für Winter (oben) und Sommer (unten). Dazu wurden Intensitätsintervalle vorgegeben, für welche dann Häufigkeiten aus der Gesamtanzahl an Niederschlagsereignissen ($R(i, j) > 0,1 \text{ mm/h}$) gezählt wurden. Die Intervalle sind in 0.5 mm-Schritte eingeteilt, wobei das erste Intervall I_1 für $0.1 \text{ mm/h} \leq I_1 < 0.5 \text{ mm/h}$, das zweite Intervall I_2 für $0.5 \text{ mm/h} \leq I_2 < 1 \text{ mm/h}$, das dritte Intervall I_3 für $1 \text{ mm/h} \leq I_3 < 1.5 \text{ mm/h}$ definiert ist, usw. (abgeschnitten bei Intensitäten $> 60 \text{ mm/h}$). Die drei Datensätze wurden zur Vergleichbarkeit jeweils mit der Anzahl an vorhandenen Datenpunkten (Gitterpunkten oder Stationen) normiert. Man sieht, dass

1. die Radarintensitäten in Winter und Sommer immer oberhalb der anderen Datensätze liegen, die Aussage aus Abschn. 3.4 über die Überschätzung der Niederschläge im Radar also in allen Intensitäten bestätigt wird,
2. die Intensitätskurven für den disaggregierten Datensatz und die stündlichen Messungen sehr gut übereinstimmen,

3. im Winter relativ schwache Niederschläge < 3 mm/h im disaggregierten Datensatz überschätzt werden,
4. im Sommer relativ schwache Niederschläge < 5 mm/h im disaggregierten Datensatz überschätzt werden,
5. im Sommer relativ starke Niederschläge > 12 mm/h im disaggregierten Datensatz unterschätzt werden.

Die Über- und Unterschätzungen können sehr gut damit erklärt werden, dass diese typische Merkmale zwischen gegitterten, also gebietsgemittelten Datensätzen (disaggregierter Datensatz) und Punktmessungen (stündliche Messungen) sind. Außerdem spielt die schlechte Repräsentativität der stündlichen Messungen in Höhenlagen (wo im Mittel auch häufiger Niederschlag fällt) eine Rolle für die genannten Unterschiede bei schwächeren Intensitäten.

Die Schlussfolgerung aus der Erstellung und Untersuchung des für Deutschland neuartigen, disaggregierten Datensatzes ist, dass nach eingehenden Tests und Abwägen der Fehlerquellen ein qualitativ hochwertiger und für die Verifikation hoch-aufgelöster Niederschlagsvorhersagen sehr nützlicher Datensatz vorliegt.

Kapitel 4

Mittlere Niederschlagssummen

Wie in den vorhergehenden Kap. 2 und 3 detailliert beschrieben, liegen nun die zur Niederschlagsanalyse benötigten Datensätze (basierend auf Beobachtungen und Modellen) in geeigneter Form, d.h. gegittert und zeitlich hoch aufgelöst vor.

Einen guten Einstieg und Überblick bekommt man, indem man zuerst die mittleren Jahresniederschläge und deren saisonale Aufteilung betrachtet (Abschn. 4.1 und 4.2). In Abschn. 4.1 werden zunächst die in dieser Arbeit verwendeten gegitterten Stationsdaten (Beobachtungen) der Jahre 2001-2004 mit der vorhandenen 30-jährigen Klimatologie (Abschn. 2.6) verglichen, um die vier Jahre in die langjährige Niederschlagsverteilung in Deutschland einordnen zu können. Es kann aber nicht erwartet werden, dass ein Mittel über vier Jahre die gleichen Werte liefert wie eine 30-jährige Zeitreihe. Ebenso wird der Vergleich für die saisonale Verteilung in Abschn. 4.2 gemacht. In Abschn. 4.3 werden aLMo-Vorhersagen (00 und 12 UTC-Läufe) für den gleichen 4-Jahres-Zeitraum mit den Beobachtungen verglichen, indem erst allgemein auf die Niederschlagsverteilung, dann etwas detaillierter auf die Differenzfelder zwischen Modell und Beobachtungen eingegangen wird.

In Abschn. 4.4 geht es um die Untersuchung möglicher Unterschiede im mittleren Jahresniederschlag zwischen den Vorhersagen des aLMo und des Globalmodells des ECMWF sowie zwischen aLMo und LM, wobei letzterer Vergleich wegen der Datenverfügbarkeit nur für 2004 möglich ist. Zum Abschluss dieses Kapitels wird der Vergleich zwischen aLMo und LM in Abschn. 4.5 im Hinblick auf die Einführung des prognostischen Niederschlags fortgeführt.

4.1 Jährliche Beobachtungen

Um die in dieser Arbeit verwendete und auszuwertende Zeitreihe von 2001-2004 für den Niederschlag in Deutschland gut einordnen zu können, wurde als erstes der mittlere Jahresniederschlag aus den Beobachtungen mit der 30-jährigen Klimatologie des DWD für die Jahre 1961-1990 verglichen.

Abbildung 4.1 zeigt den Jahresniederschlag der einzelnen Jahre und den mittleren Jahresniederschlag für Deutschland aus den Beobachtungen für 2001-2004. Ver-

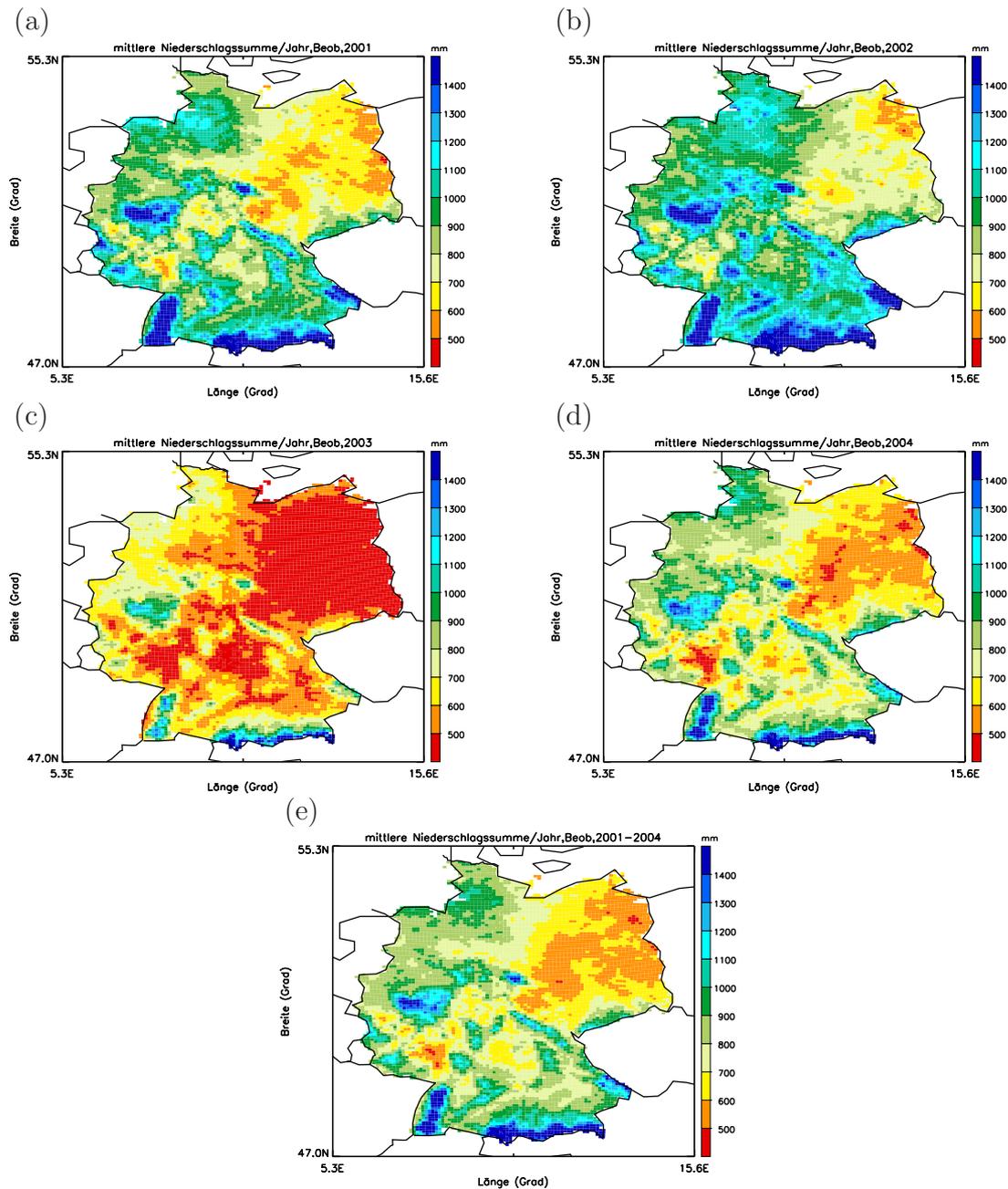


Abbildung 4.1: Jahresniederschläge (in mm) aus Beobachtungen für Deutschland: (a) 2001, (b) 2002, (c) 2003, (d) 2004, (e) Mittel aus 2001-2004.

gleicht man die einzelnen Jahre miteinander, so ist 2003 deutlich das trockenste Jahr gewesen. Das feuchteste Jahr war 2002, gefolgt von 2001 und 2004. Am nächsten am 4-jährigen Mittel liegt 2004.

Vergleicht man nun diese Jahresmittel aus Abb. 4.1 mit der Klimatologie aus Abschn. 2.6, so erkennt man eine sehr gute Übereinstimmung der Niederschlagsverteilung beider Jahresmittel (trotz der starken jährlichen Schwankungen). Die trockensten Regionen finden sich in beiden Datensätzen in ganz Ost- und Teilen Südwestdeutschlands. Die Niederschlagsmaxima finden sich in den Deutschen Mittelgebirgsregionen und den Alpen bzw. dem Alpenvorland. Leichte Unterschiede findet man im Norden, wo das 4-jährige Mittel etwas feuchter ist, und im Südwesten, der ge-

Datensatz	Σ Jahr [mm/Jahr]	Winter [%]	Frühling [%]	Sommer [%]	Herbst [%]
Klimatologie (1961-1990)	703	22.8	23.6	30.4	23.2
Beobachtungen (2001-2004)	828	23.7	20.9	28.5	26.9
aLMo (00 UTC) (2001-2004)	859	22.9	21.6	32.1	23.4
aLMo (12 UTC) (2001-2004)	837	23.1	21.6	32.8	22.5
ECMWF (2002-2004)	811	24.9	20.6	29.6	24.9

Tabelle 4.1: Mittlerer Jahresniederschlag in mm über alle Gitterpunkte, sowie die prozentualen Anteile für die Jahreszeiten.

genüber der Klimatologie etwas trockener ist.

Zusätzlich sind für den Vergleich mit der Klimatologie und später in diesem Kapitel mit den aLMo-Vorhersagen in Tab. 4.1 die über Deutschland gemittelten Jahresniederschläge aufgeführt. Insgesamt findet man mit 828 mm einen für die 4-jährige Zeitreihe um 115 mm höher liegenden Wert als in der Klimatologie (703 mm). Wegen des sehr trockenen Jahres 2003 innerhalb der Zeitreihe ist es überraschend, dass die Klimatologie im Mittel trockener ist (die Jahre 2001 und 2002 scheinen diese Trockenheit jedoch mehr als nur zu kompensieren). Es könnte aber auch sein, dass der höher liegende Wert in der 4-jährigen Zeitreihe generell mit den langjährigen Niederschlagstrends zusammenhängt. Über Nordeuropa wird aktuell eine Zunahme der Winter- und zum Teil auch der Herbstniederschläge beobachtet (IPCC, 2007).

Tabelle 4.2 gibt zusätzlich einen Einblick in die Minimal- und Maximalwerte der Datensätze. Die klimatologisch ermittelten Extremwerte werden von den Beobachtungen 2001-2004 nicht ganz erreicht, vor allem das Maximum ist um ca. 800 mm/Jahr niedriger. Dies scheint ein Widerspruch zu den Aussagen über die Werte in Tab. 4.1 zu sein. Allerdings kann dies dadurch begründet sein, dass beide Datensätze auf einem verschieden aufgelösten Gitter liegen, vor allem die Intensitäten der Stationsdaten für das 7 km-Gitter deutlich mehr verschmiert werden.

4.2 Saisonale Beobachtungen

Einen etwas differenzierteren Einblick gibt die Aufteilung des Jahresmittels in jahreszeitlich getrennte Werte. Abbildung 4.2 zeigt diese saisonal gemittelten Niederschlagsverteilungen für alle vier meteorologischen Jahreszeiten. Um einen schnellen Vergleich der einzelnen Jahreszeiten mit dem Jahresniederschlag machen zu können,

Datensatz	Niederschlags- minimum [mm/Jahr]	Niederschlags- maximum [mm/Jahr]
Beobachtungen (2001-2004)	460	2424
Klimatologie (1961-1990)	383	3270
aLMo (00 UTC) (2001-2004)	429	3760
aLMo (12 UTC) (2001-2004)	441	3653
ECMWF (2002-2004)	577	1605

Tabelle 4.2: Minimum und Maximum des mittleren Jahresniederschlags für Deutschland.

wurden die Konturwerte der Niederschlagsskala geviertelt. Somit kann direkt veranschaulicht werden, welche Jahreszeit überproportional zur Jahressumme beiträgt (würde jede Jahreszeit gerade genau ein Viertel des Jahresniederschlags beitragen, sähen alle Bilder in Abb. 4.2 exakt wie das Jahresmittel in Abb. 4.1 aus).

Erkennbar ist sofort, dass im Sommer (Abb. 4.2 (c)) fast in ganz Deutschland mehr als ein Viertel des Jahresniederschlags beobachtet wird (bis auf wenige Maxima in Mittelgebirgen). Auch im Herbst (Abb. 4.2 (d)) wird vor allem in den Mittelgebirgen und Regionen an der Nordseeküste ein höherer Anteil an Niederschlägen beobachtet. Dagegen ist der Frühling (Abb. 4.2 (b)) in ganz Deutschland am trockensten. Im Winter (Abb. 4.2 (a)) ist in Ostdeutschland und den niedriger gelegenen Regionen in der Mitte und im Süden Deutschlands der Niederschlagsanteil deutlich unter einem Viertel des Jahresniederschlags, während sich diese Tatsache auf den Höhenlagen der Mittelgebirge gerade umgekehrt.

Auch auf jahreszeitlicher Basis ist ein kurzer Vergleich zur Niederschlagsverteilung der Klimatologie sinnvoll. Winter, Sommer und Herbst passen von der Niederschlagsverteilung her generell gut zu den entsprechenden Jahreszeiten der Klimatologie (Abb. 2.10). Im Frühling gibt es den Unterschied, dass im saisonalen Mittel zwischen 2001-2004 im Norden weniger Niederschlag beobachtet wurde. Nimmt man die jahreszeitlichen Werte aus Tab. 4.1 dazu, wird deutlich, dass für 2001-2004 vor allem im Sommer mit 28.5%, aber auch im Herbst mit 26.9% jeweils mehr als ein Viertel der Jahressumme an Niederschlag beobachtet wird. Dies gilt in der Klimatologie nur für den Sommer mit sogar 30.4%. Im Gegensatz zu 2001-2004 folgt aus der 30-jährigen Klimatologie ein relativ gleichmäßiger Niederschlagsanteil für Winter, Frühling und Herbst am Jahresniederschlag.

Im Vergleich zum langjährigen Niederschlagsklima in Deutschland ist die untersuchte Zeitreihe eher feuchter, jedoch schließen die vier Jahre 2001-2004 sowohl ein außergewöhnlich feuchtes, als auch ein sehr trockenes Jahr mit ein. Deshalb kann da-

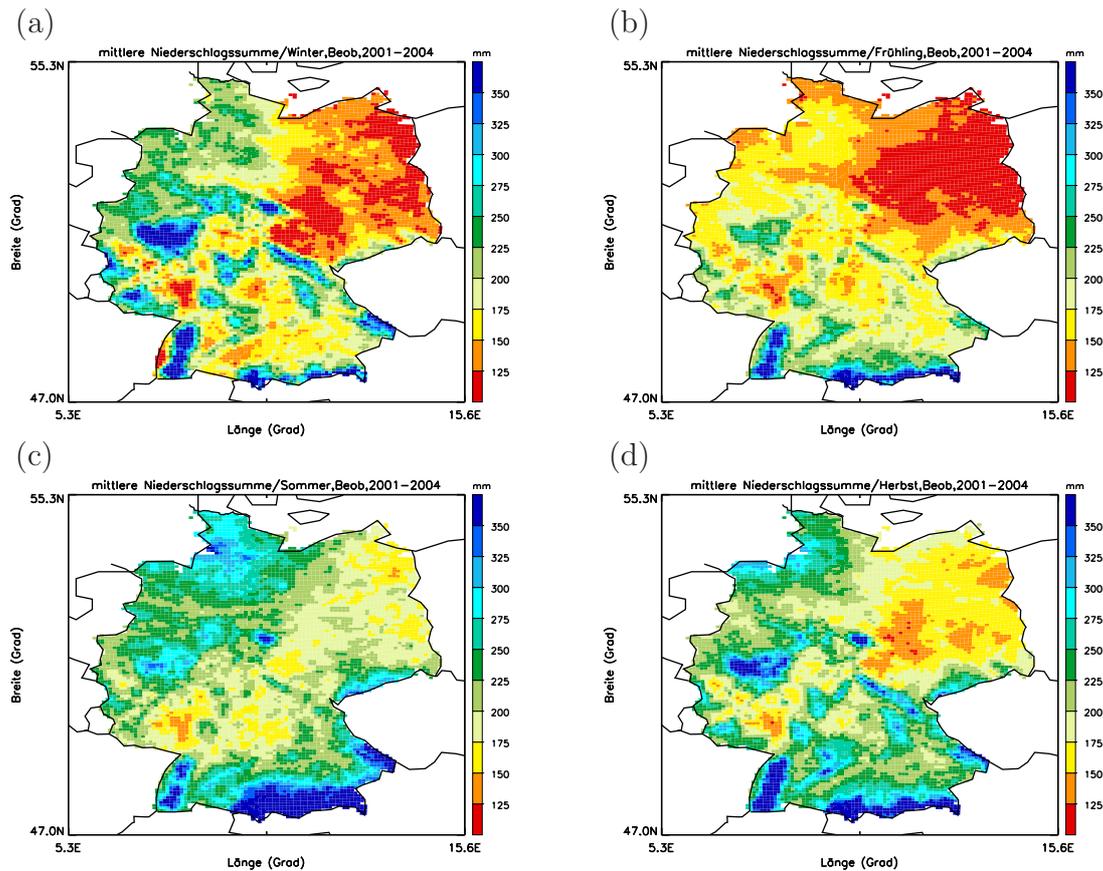


Abbildung 4.2: Saisonale gemittelte Niederschlagssummen aus Beobachtungen für 2001-2004: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

von ausgegangen werden, dass die Verifikationsergebnisse in dieser Arbeit Auskunft über das Modellverhalten zum gesamten Spektrum der klimatischen Bedingungen in Deutschland geben.

4.3 Vergleich zwischen Modellen und Beobachtungen

In diesem Abschnitt wird ein erster grundlegender Schritt zur Verifikation vor allem des aLMo (Abschn. 2.1), aber auch des ECMWF-Globalmodells (Abschn. 2.2) getan, indem mittlere Jahresniederschläge und ihre Aufteilung in Jahreszeiten mit Beobachtungen verglichen werden. Dabei wurde auch untersucht, ob es für diesen ersten Verifikationsschritt einen Unterschied der aLMo-Vorhersagen mit Startpunkt 00 UTC (Tagessumme aus Vorhersagezeit: 6-30 h) und 12 UTC (Tagessumme aus Vorhersagezeit: 18-42 h) gibt.

Das aLMo

Abbildung 4.3 (a) zeigt den mittleren Jahresniederschlag für 2001-2004 aus den 00 UTC-Läufen, Abb. 4.3 (b) für 12 UTC-Läufe. Vergleicht man zuerst einmal diese

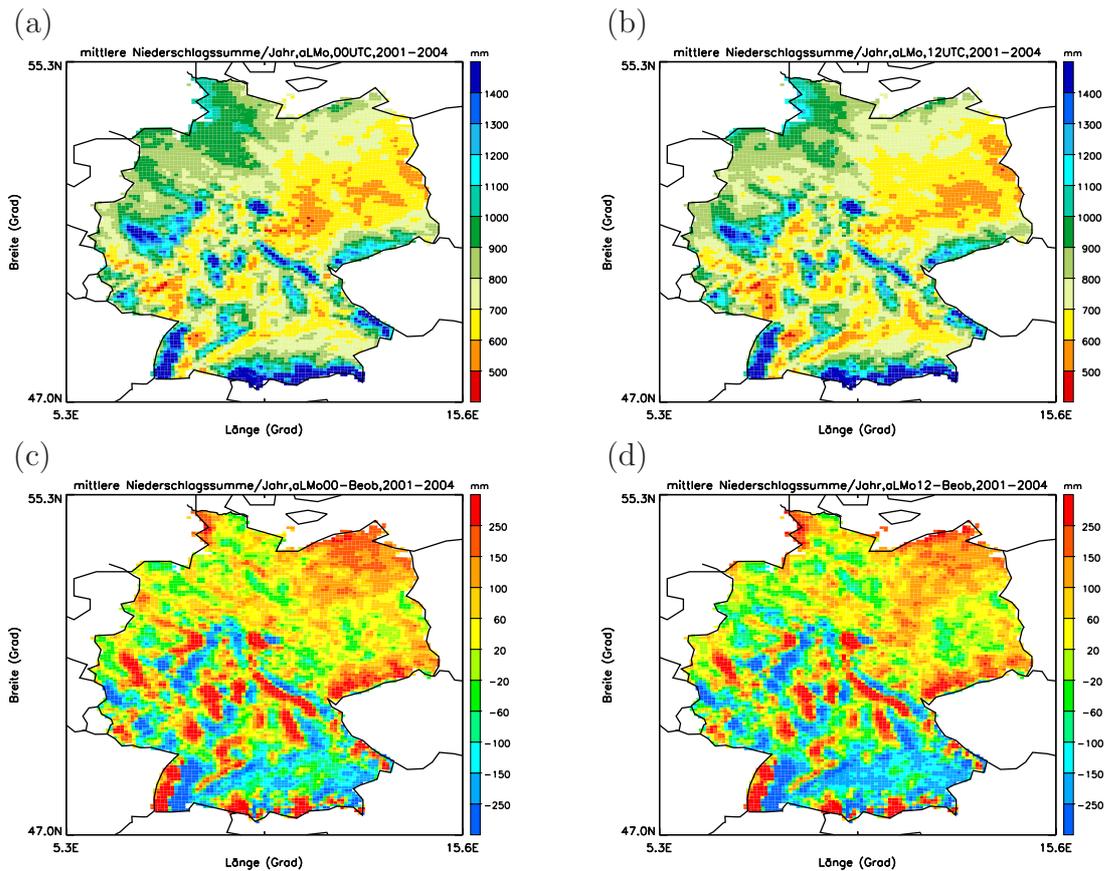


Abbildung 4.3: Mittlere Jahresniederschläge für 2001-2004: (a) aLMO-Vorhersagen (00 UTC-Läufe, 6-30 h), (b) aLMO-Vorhersagen (12 UTC-Läufe, 18-42 h), (c) Differenz der 00 UTC-Läufe und den Beobachtungen (Abb. 4.1 (a)), (d) Differenz der 12 UTC-Läufe und den Beobachtungen.

beiden mittleren Jahresniederschläge, so erkennt man kaum Unterschiede. Es soll im Folgenden also ausreichen, den mittleren Jahresniederschlag aus aLMO-Vorhersagen der 00 UTC-Läufe zu behandeln.

Vergleicht man nun Abb. 4.3 (a) mit den Beobachtungen aus Abb. 4.1 (e), so scheint das aLMO die Niederschlagsverteilung in Deutschland grundsätzlich gut wiederzugeben mit den relativ trockenen Regionen in Ost- und Südwestdeutschland sowie den Niederschlagsmaxima in den Mittelgebirgen und dem Voralpenland. Tabelle 4.1 gibt weiter Auskunft darüber, dass der über Deutschland gemittelte Jahresniederschlag recht gut vorhergesagt wird. Im aLMO gibt es nur eine leichte Überschätzung. Aus Tab. 4.2 bekommt man jedoch schon eine erste Information darüber, dass das aLMO im Maximum deutlich höhere Jahresniederschläge vorhersagt, als im selben Zeitraum beobachtet wurden (Maximum um ca. 1300 mm/Jahr überschätzt). Auch die Minima sind etwas extremer, weshalb man sagen kann, dass die Modellfehler generell größere räumliche Variationen zeigen. Diese Unterschiede sind auch darauf zurückzuführen, dass es Unterschiede in der effektiven Auflösung zwischen Modell und Beobachtungen gibt. Je nach Region und damit Stationsdichte liegt diese effektive Auflösung in der gegitterten Analyse bei 14-28 km (vgl. Abschn. 3.1), also geringer als die Auflösung im Modell. Deshalb erwartet man in den Beobachtungen

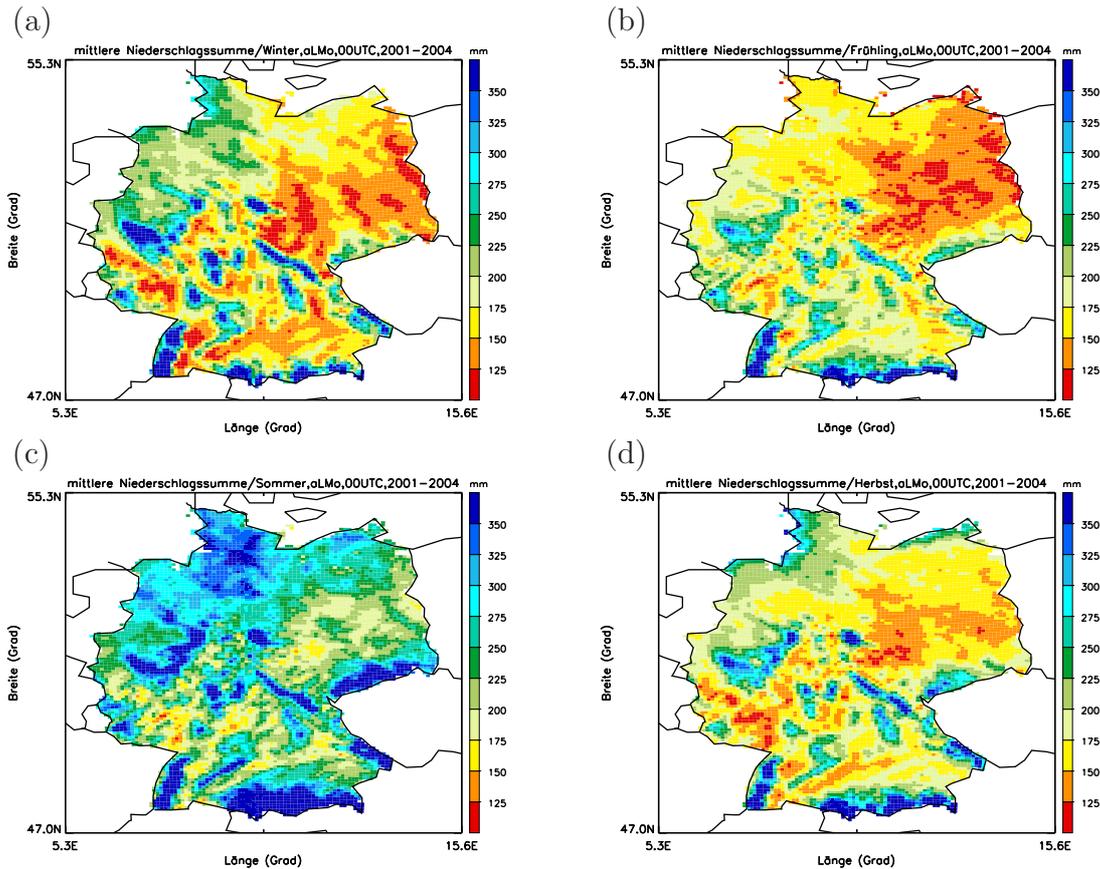


Abbildung 4.4: Saisonale gemittelte Niederschlagssummen von aLMO-Vorhersagen der 00 UTC-Läufe für 2001-2004: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

etwas glattere Niederschlagsfelder.

Um etwas mehr ins Detail zu gehen, wurde die Differenz der mittleren Jahresniederschläge von aLMO und Beobachtungen an jedem Gitterpunkt gebildet (Abb. 4.3 (c) für 00 UTC-Läufe und (d) für 12 UTC-Läufe). Es ist deutlich zu erkennen, dass über ganz Deutschland verteilt Unterschätzungen (blaue Farbgebung) und Überschätzungen (rote Farbgebung) des Niederschlags in der aLMO-Vorhersage vorkommen. Besonders auffallend dabei sind die bandartigen Strukturen von Überschätzungen im Luv und Unterschätzungen im Lee der Mittelgebirgsregionen (bei typischer westlicher Anströmrichtung). Weiterhin auffällig ist die Niederschlagsüberschätzung in weiten Teilen Ostdeutschlands (60 bis über 250 mm/Jahr) und die Niederschlagsunterschätzung in großen Teilen des Donaeinzugsgebietes (20 bis ca. 250 mm/Jahr). Zur besseren Darstellung wurde die Skala in den Abb. 4.3 (c) und (d) oberhalb von 250 mm/Jahr und unterhalb von -250 mm/Jahr nicht fortgeführt. Maximale Über- bzw. Unterschätzungen zwischen einem Gitterpunkt im aLMO und dem gleichen Gitterpunkt in den Beobachtungen betragen +1940 mm/Jahr bzw. -952 mm/Jahr. Diese oder ähnlich hohe Werte findet man vor allem an Gitterpunkten in den Mittelgebirgsregionen, wo die erwähnte „Luv/Lee-Problematik“ der aLMO-Vorhersagen maximal groß ist.

Wie schon in Abschn. 4.1 für die Beobachtungen wurde auch für die Vorhersagen des aLMO eine Aufteilung der mittleren Jahresniederschläge in saisonal gemittelte Nie-

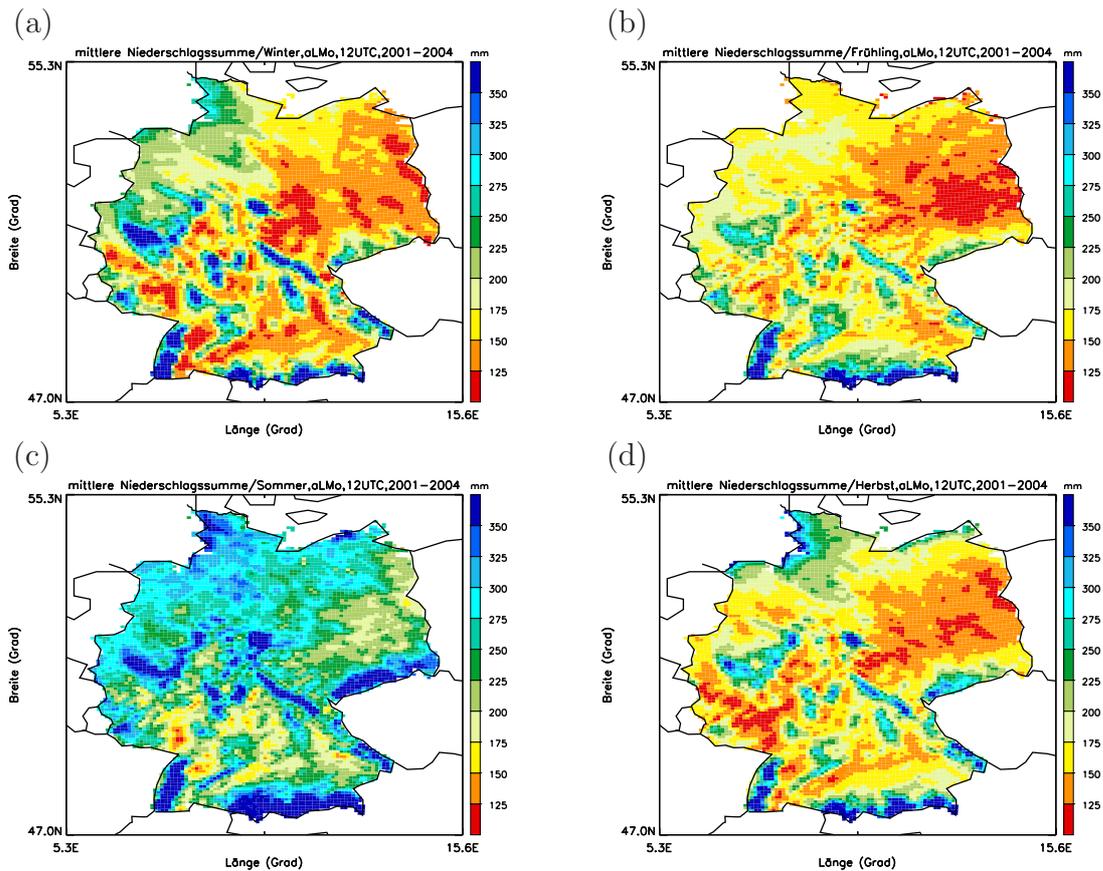


Abbildung 4.5: Wie Abb. 4.4 für 12 UTC-Läufe.

erschlagssummen vorgenommen. Abbildung 4.4 zeigt die Niederschlagsverteilungen der Jahreszeiten für aLMO-Vorhersagen der 00 UTC-Läufe mit gleicher Farbskala wie für die Jahresniederschläge, jedoch mit durch vier geteilten Konturwerten (vgl. Abschn. 4.1). Gleiches findet sich in Abb. 4.5 für 12 UTC-Läufe.

Ein Vergleich zwischen den jeweiligen Niederschlagsverteilungen aus beiden Modellläufen zeigt kaum Unterschiede (ähnlich wie schon in den mittleren Jahresniederschlägen). Unterschiede finden sich bei genauerem Hinsehen z.B. für den Sommer in Norddeutschland (jeweils Bild (c)), wo die um 00 UTC gestarteten Vorhersagen mehr Niederschlag liefern als für 12 UTC-Läufe, oder im Herbst im Osten Deutschlands, wo wiederum die Vorhersagen aus 00 UTC-Läufen etwas mehr Niederschlag liefern.

Im Fortlauf dieses Abschnitts werden aber aufgrund der relativ geringen Unterschiede zwischen den saisonal gemittelten Niederschlagsfeldern der beiden Modellläufe nur noch die 00 UTC-Läufe näher betrachtet. Dies wird auch durch die prozentualen Anteile der jahreszeitlichen Niederschläge am Jahresniederschlag aus Tab. 4.1 unterstützt. Die Vorhersagen aus beiden Läufen liefern vergleichsweise vernachlässigbare Unterschiede. Im Vergleich zu den beobachteten Werten ist der Niederschlagsanteil im Winter und Frühling in sehr ähnlicher Größenordnung, im Sommer wird der Niederschlag für Deutschland um ca. 13% überschätzt, im Herbst dagegen mit ca. 13% unterschätzt.

Vergleicht man die jahreszeitlich aufgeteilten Niederschlagsverteilungen mit dem

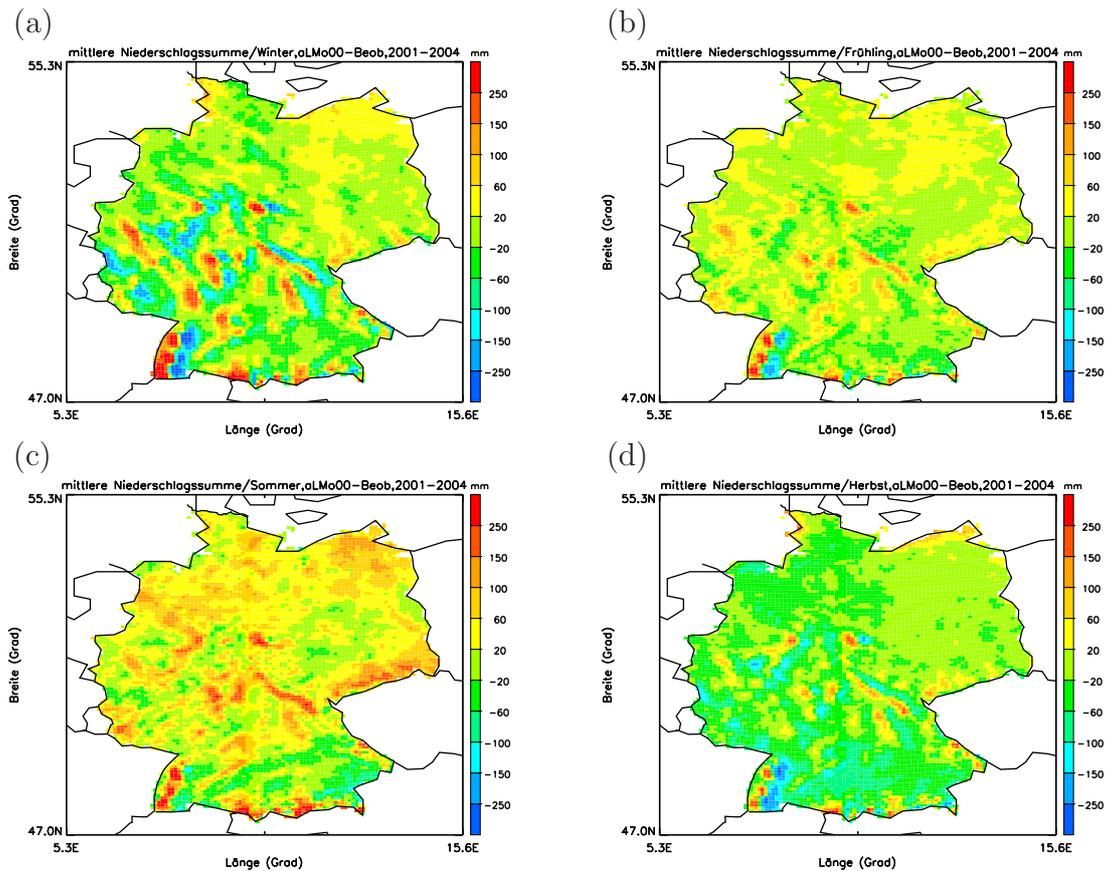


Abbildung 4.6: Saisonale gemittelte Niederschlagsdifferenzen zwischen aLMo (00 UTC-Läufe) und Beobachtungen an jedem Gitterpunkt für 2001-2004: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

Jahresmittel aus Abb. 4.3 (a), so stellt man wie zuvor schon für die Beobachtungen fest, dass im Sommer (Abb. 4.4 (c)) anteilig der meiste Niederschlag fällt (vgl. auch Tab. 4.1). Der Winter ist, ebenfalls ähnlich zu den Ergebnissen aus Abschn. 4.1, in Ostdeutschland und den tiefergelegenen Regionen relativ trocken, dagegen in den Mittelgebirgsregionen relativ feucht. Ein sehr ähnliches Bild zeigt sich im Herbst (Abb. 4.4 (d)), der bis auf ein paar relativ trockene Regionen in Teilen der Mitte Deutschlands ungefähr dem Bild des mittleren Jahresniederschlags entspricht, was einem Beitrag von einem Viertel der Jahressumme ausmacht (vgl. auch Tab. 4.1). Der Frühling (Abb. 4.4 (b)) ist dagegen fast überall in Deutschland die trockenste Jahreszeit im Modell. Dies entspricht dem Ergebnis aus den Beobachtungen (Abschn. 4.1).

Ein etwas detaillierteres Bild für den Vergleich zu den Beobachtungen bekommt man, wenn man erneut die Differenz aus aLMo und Beobachtungen an jedem Gitterpunkt (diesmal für jede Jahreszeit) berechnet. Abbildung 4.6 zeigt die vier resultierenden Differenzfelder, wobei hier die Farbgebung und Levelinteilung im Unterschied zu den Niederschlagsverteilungen aus Abb. 4.4 die gleiche wie für Differenzen des Jahresmittels ist.

Die für die Jahresniederschläge gefundene Problematik mit Überschätzungen im Luv und Unterschätzungen im Lee der Mittelgebirge (ausgehend vom Modell) findet sich

hier in abgeschwächter Form wieder (besonders im Winter (a), aber auch im Herbst (d) und etwas schwächer im Frühling (b)). Dabei tritt diese Problematik in allen Jahreszeiten besonders stark in der Schwarzwaldregion auf.

Sowohl für aLMo als auch für die Beobachtungen wurde der Sommer als feuchtester Monat identifiziert. Die Differenz im Sommer (Abb. 4.6 (c)) zeigt, dass das aLMo den Niederschlag bis auf die Donauregion zu großen Teilen überschätzt, in einigen Regionen sogar deutlich mit über 60 mm/Sommer bis zu einem Maximalwert von 600 mm/Sommer.

Zusammenfassend kann festgestellt werden, dass aus Untersuchungen der mittleren Jahresniederschläge und deren Aufteilung in Jahreszeiten für die Modellvorhersagen des aLMo zwei Hauptprobleme herausgestellt werden können (vgl. COSMO, 2006):

- „Luv/Lee-Problematik“ in Mittelgebirgsregionen,
- Überschätzung des Niederschlags im Sommer.

Das Globalmodell des ECMWF

Im Folgenden wird das Globalmodell des ECMWF mit den Beobachtungen verglichen. Dabei wurden zur Jahres- und saisonalen Mittelung des Niederschlags wegen fehlender Daten in den ersten drei Monaten von 2001 nur die Jahre 2002-2004 verwendet. Aufgrund der deutlich niedrigeren horizontalen Auflösung des Globalmodells (ca. 60 km, im Gegensatz zu aLMo, LM und Beobachtungen mit 7 km) kann natürlich auch die Niederschlagsverteilung nur entsprechend flächig (ohne kleinskalige Strukturen) vorhergesagt werden.

Abbildung 4.7 zeigt allerdings dennoch eine ganz gute Übereinstimmung mit den Beobachtungen, sowohl für das Jahresmittel, als auch für die saisonale Verteilung (vgl. Tab. 4.1). Die relativ trockenen Gebiete liegen wie in den Beobachtungen vorwiegend in Ost- und Südwestdeutschland. Die Niederschlagsmaxima finden sich in den ausgedehnten Mittelgebirgsregionen im Westen und des Schwarzwaldes, sowie im Alpenvorland und den Alpen selbst. Kleinere Mittelgebirgsregionen stechen wegen der niedrigen Auflösung nicht so stark hervor (vgl. auch Maxima und Minima aus Tab 4.2: Extrema kommen nicht so stark zur Geltung) wie in den Beobachtungen, sind aber dennoch zu erkennen (z.B. Harz). Ebenfalls qualitativ gut vorhergesagt ist der mittlere Jahresniederschlag im Norden Deutschlands mit einer Zunahme hin zur Nordseeküste. Wie in den Beobachtungen können der Frühling als trockenste und der Sommer als feuchteste Jahreszeit identifiziert werden (Tab. 4.1).

4.4 Modellvergleich zwischen aLMo und LM

In diesem Abschnitt werden die beiden Lokal-Modelle aLMo und LM anhand des Jahresniederschlags für 2004 verglichen (vgl. Abschn. 2.1). Der Vergleich zwischen den stark verwandten Lokal-Modellen erlaubt es zu prüfen, ob die Unterschiede in

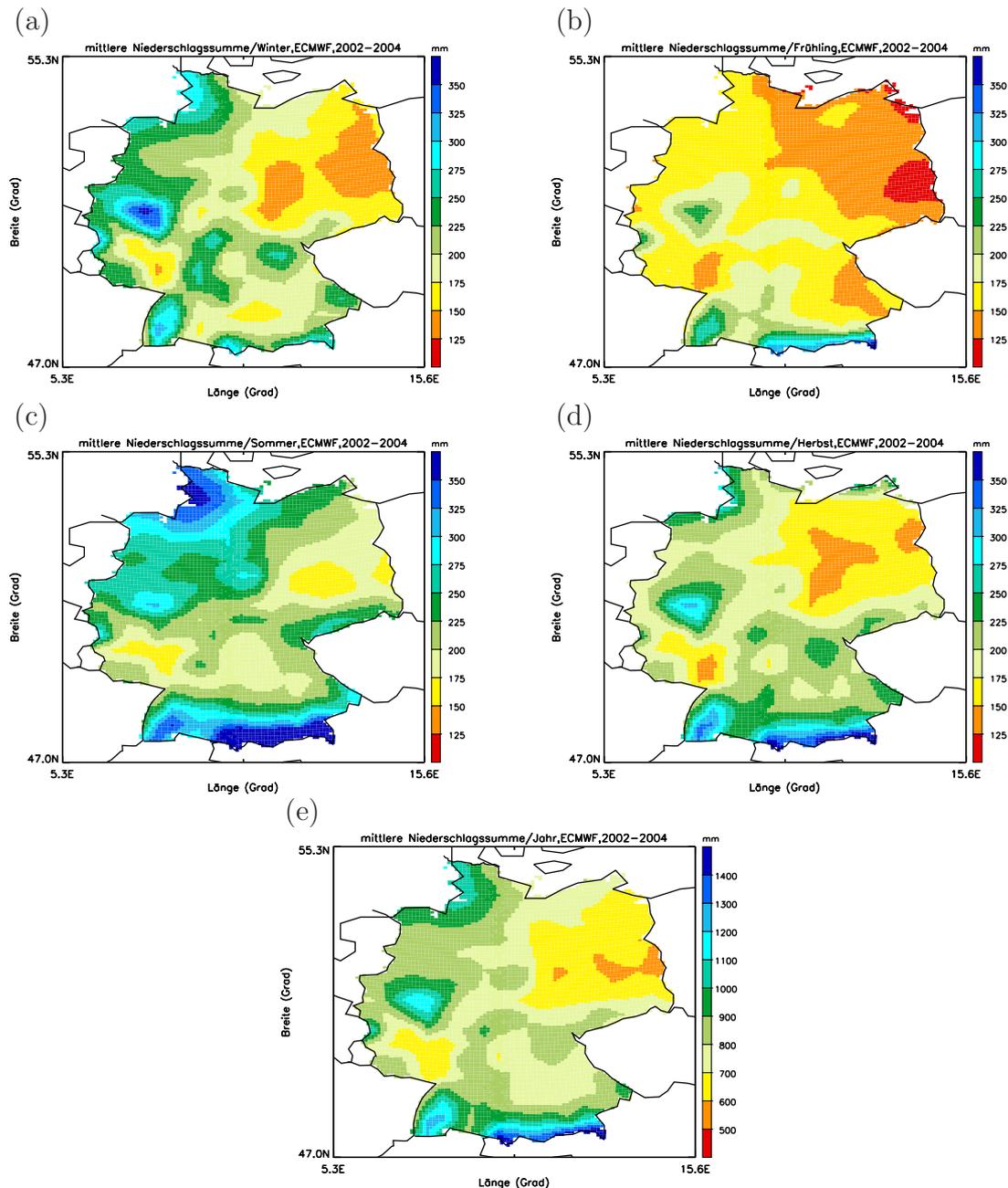


Abbildung 4.7: mittlerer Niederschlag für ECMWF-Vorhersagen in Deutschland für 2002-2004: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst und (e) das ganze Jahr.

den Anfangs- und Randbedingungen (Abschn. 2.1) und in der Einführung des prognostischen Niederschlags entscheidende Auswirkungen auf die Niederschlagsvorhersage haben.

Abbildung 4.8 zeigt den Jahresniederschlag für 2004 beider Modellversionen und deren Differenzen zu den Beobachtungen für Vorhersagen der 00 UTC-Läufe, Abb. 4.9 entsprechend für 12 UTC-Läufe.

Vorab ist zu sagen, dass der Jahresniederschlag der aLMO-Vorhersagen 2004 in der regionalen Verteilung repräsentativ für die vier Jahre ist (Abb. 4.3). Etwas auffällig sind nur die trockenen Regionen, die im Gegensatz zum 4-jährigen Mittel etwas ausgedehnter und trockener sind.

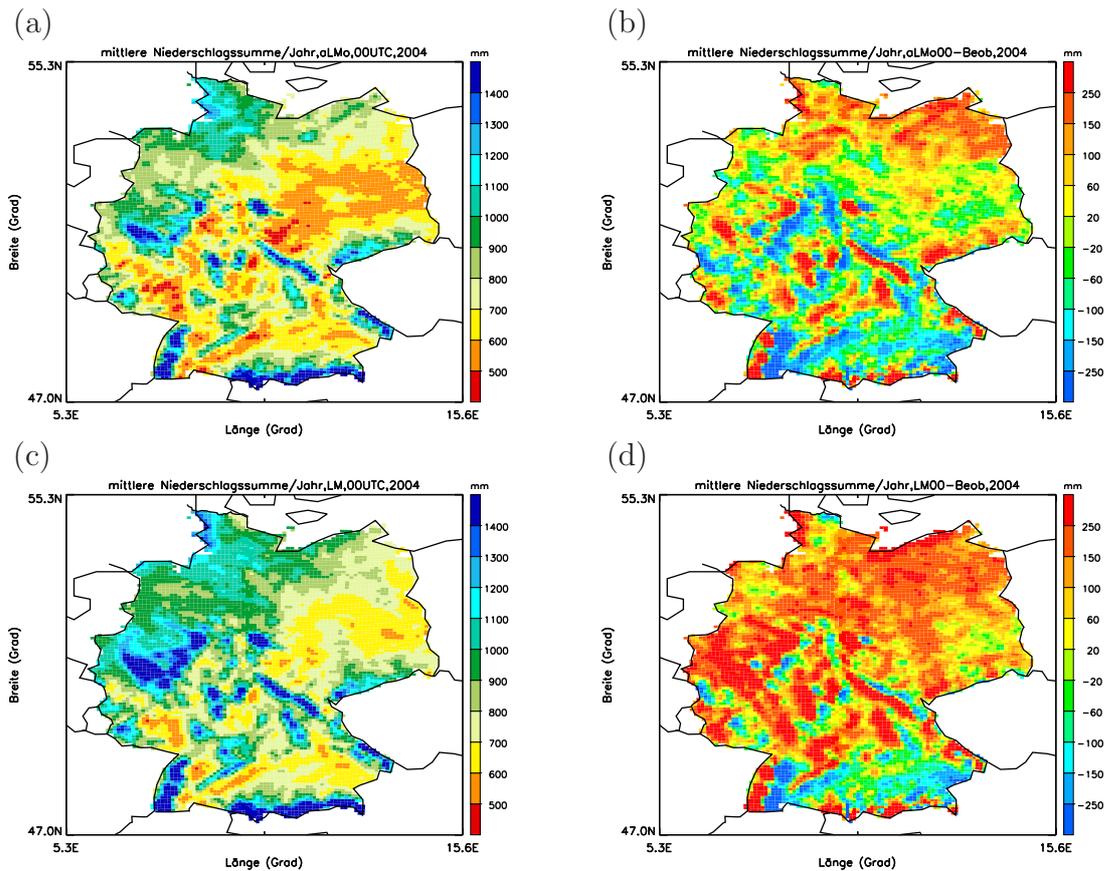


Abbildung 4.8: Modellvergleich zwischen *aLMo* und *LM* für den mittleren Jahresniederschlag in Deutschland für 2004: (a) *aLMo* (00 UTC-Läufe), (b) Differenz zwischen *aLMo* (00 UTC-Läufe) und Beobachtungen, (c) *LM* (00 UTC-Läufe), (d) Differenz zwischen *LM* (00 UTC-Läufe) und Beobachtungen.

Das *LM* zeigt deutliche Unterschiede. Sieht man sich die Differenzfelder in Abb. 4.8 an, so weist das *LM* gegenüber dem *aLMo* in Norddeutschland und in der Mitte deutlich größere Flächen mit hohen Niederschlagsüberschätzungen aus. Im Nordosten und in Süden sagen beide Modellversionen sehr ähnliche Niederschlagsmengen vorher.

Im Gegensatz zu Abschn. 4.3 können für 2004 auch deutliche Unterschiede zwischen den beiden Startzeitpunkten der Vorhersagen gefunden werden. Beim *aLMo* liegt der Hauptunterschied der 12 UTC-Läufe im Vergleich zu den 00 UTC-Läufen einerseits im Nordosten mit etwas weniger Niederschlagsüberschätzung, andererseits nimmt die Unterschätzung in der Donauregion in der Fläche deutlich zu. Die *LM* Vorhersagen der 12 UTC-Läufe unterscheiden sich im Osten kaum von den 00 UTC-Läufen, jedoch ist die oben erwähnte allgemeine Niederschlagsüberschätzung für den Norden und die Mitte in den 12 UTC-Läufen nicht mehr so ausgeprägt. Die *LM* Vorhersagen aus 12 UTC-Läufen scheinen im Mittel über Deutschland immer noch schlechter zu sein als beide Läufe des *aLMo* (mit Ausnahme von Süddeutschland, dort sind beide Modelle sehr ähnlich), jedoch deutlich besser als die *LM* Vorhersagen aus 00 UTC-Läufen.

Überraschend ist bei allen gefundenen Differenzen auch, dass gerade in Süddeutsch-

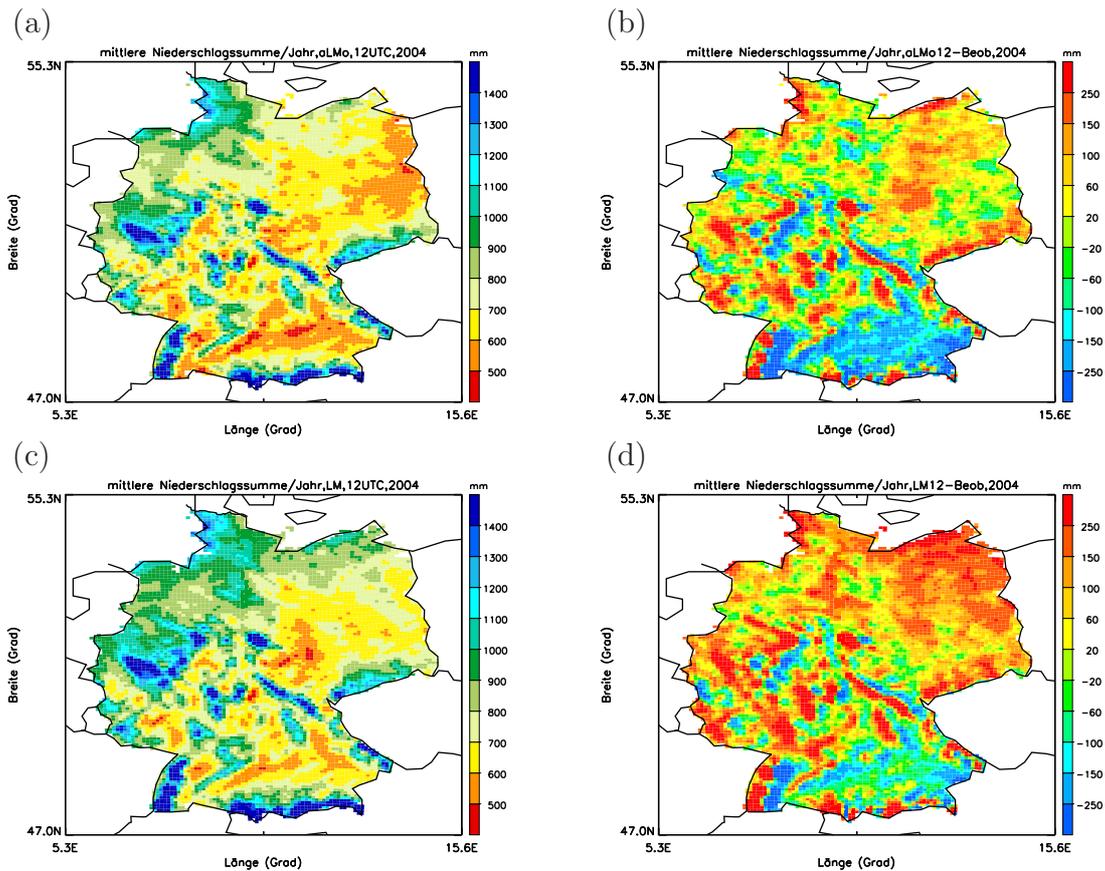


Abbildung 4.9: Wie Abb.4.8 für 12 UTC-Läufe.

land (mit Schwarzwald, Alpenvorland und Donauregion) beide Modelle ähnlich sind, bei den 12 UTC-Läufen speziell für die Donauregion sogar das aLMo etwas schlechter abschneidet, denn das aLMo ist eigentlich eine speziell für gebirgiges Terrain angepasste Version des LM.

4.5 Einfluss des prognostischen Niederschlags

Eine wichtige Neuerung in den Modellen aLMo und LM war die Einführung des prognostischen Niederschlags (Abschn. 2.1). Der prognostische Niederschlag ist im LM seit 25.04.2004 und im aLMo seit 16.11.2004 in den operationellen Betrieb eingebunden. Um einen möglichen Aufschluss über eine Verbesserung der Vorhersage mit prognostischem Niederschlag zu bekommen, wurden beide Modelle (aLMo und LM, 00 UTC-Läufe) für 2004 in drei Zeitperioden aufgeteilt und die Niederschlagssummen davon verglichen.

Abbildung 4.10 zeigt diesen Vergleich mit Differenzfeldern zunächst für die Zeit vom 01.01.2004 bis 25.04.2004, in der beide Modellversionen noch ohne prognostischen Niederschlag Vorhersagen liefern (Abb. 4.10 (a) und (b)). Die zweite Zeitspanne reicht vom 26.04.2004 bis 15.11.2004 (Abb. 4.10 (c) und (d)), in der im aLMo noch ohne, jedoch im LM bereits mit prognostischem Niederschlag gerechnet wurde. Anschließend folgt die dritte Zeitspanne vom 16.11.2004 bis 31.12.2004 (Abb. 4.10 (e))

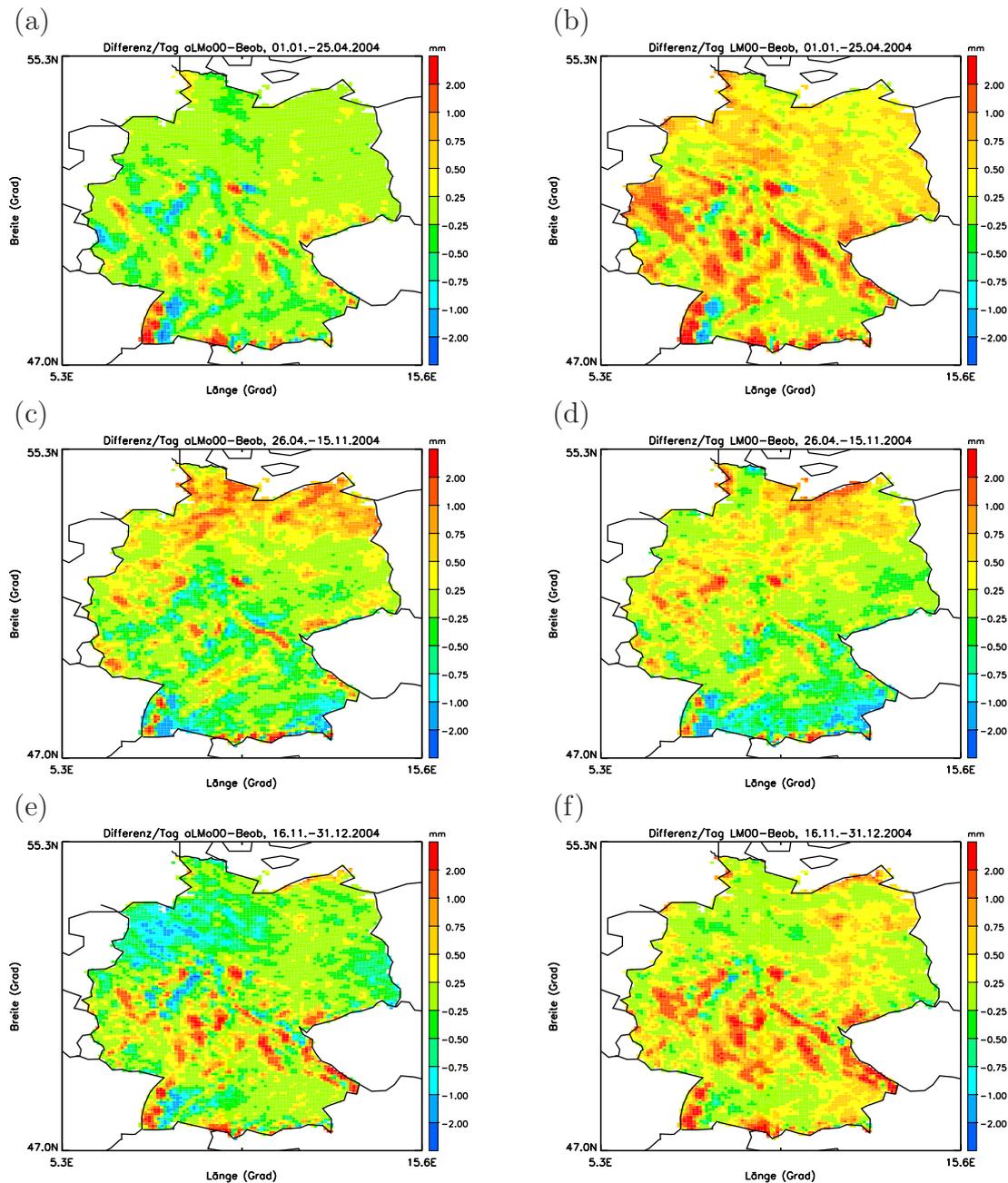


Abbildung 4.10: Niederschlagsdifferenzen für aLMo und Beobachtungen (linke Spalte) und LM und Beobachtungen (rechte Spalte) der 00 UTC-Läufe für 2004 bezüglich des prognostischen Niederschlags. Oben: 01.01.-25.04.2004 (beide Modelle ohne prognostischen Niederschlag). Mitte: 26.04.-15.11.2004 (LM mit, aLMo ohne prognostischen Niederschlag). Unten: 16.11.-31.12.2004 (beide Modelle mit prognostischem Niederschlag).

und (f)), in der beide Modellversionen mit prognostischem Niederschlag rechnen. Da die Zeitspannen verschieden lang sind, wurde in Abb. 4.10 immer mit der entsprechenden Anzahl an Tagen normiert. Aufgrund dessen, dass die einzelnen Zeitperioden recht kurz sind (und teils nur bestimmte, von verschiedenen Wetterbedingungen dominierte Jahreszeiten umfassen), soll hier nur ein kurzer, rein qualitativer Vergleich gemacht werden.

In der ersten Zeitspanne (116 Tage) ohne prognostischen Niederschlag sieht der Ver-

gleich ähnlich dem in Abschn. 4.4 für die Jahressumme aus, dass nämlich das LM in der Fläche deutlich mehr Überschätzung zeigt, wobei wiederum im süddeutschen Raum (insbesondere in der Schwarzwaldregion) etwas weniger Unterschiede zwischen den beiden Modellen sichtbar sind.

In der zweiten Zeitspanne (204 Tage), in der das LM nun mit, jedoch das aLMo weiter ohne prognostischen Niederschlag läuft, werden beide Niederschlagsfelder deutlich ähnlicher. Im Vergleich zu Abb. 4.10 (a) und (b) hat das LM insgesamt qualitativ zu den aLMo-Vorhersagen aufgeschlossen.

Die Statistik wird für die dritte Zeitspanne (46 Tage), in der beide Modellversionen mit prognostischem Niederschlag laufen, deutlich schlechter. Die Überschätzungen im LM sind jetzt jedoch wieder (zumindest in der Gesamtsicht auf Deutschland) höher als im aLMo. Im aLMo tritt nun (zusätzlich) das Problem einer Unterschätzung im Nordwesten auf. Zudem nehmen die Überschätzungen im Luv der Gebirge wieder leicht zu.

Vor allem im Sommer scheint der prognostische Niederschlag (im LM) eine deutliche Verbesserung zu bringen, wenn auch nicht alle Probleme damit gelöst sind. Diese Aussage ist vorsichtig zu bewerten, da nur Daten eines Jahres aufgeteilt in verschiedene Jahreszeiten verglichen werden konnten.

Kapitel 5

Intensitäten und Tagesgänge

Der für Deutschland neuartige, disaggregierte Datensatz (Abschn. 3) bietet einige neue Möglichkeiten zur Analyse des Niederschlags. Um von dem im vorhergehenden Kapitel beschriebenen Überblick mit Hilfe der mittleren Jahresniederschläge (und deren Aufteilung in Jahreszeiten) nun etwas mehr ins Detail zu gehen, werden die Niederschlagshäufigkeiten und -intensitäten betrachtet. In Abschn. 5.1 werden zunächst Häufigkeitskarten für Deutschland für tägliche und stündliche Niederschlagsereignisse ausgewertet. Aus dieser Gesamtheit an Niederschlagsereignissen folgt im anschließenden Abschn. 5.2 die Auswertung von Intensitätsverteilungen für Winter und Sommer. Aus diesen Intensitätsverteilungen können weiter Perzentile (50, 80, 90, 95 und 99 %) errechnet, interpretiert und mit aLMO Vorhersagen verglichen werden.

Abschnitt 5.3 handelt schließlich von der Verifikation des Tagesgangs im Niederschlag. Die Verifikation des Tagesgangs im aLMO wird mit dem disaggregierten Datensatz nun sehr gut möglich. Außerdem werden die gefundenen Ergebnisse mit Arbeiten von Dai et al. (1999) und Kaufmann et al. (2003) verglichen.

5.1 Häufigkeiten

Die Ergebnisse aus Kap. 4 liefern einen Überblick über die Niederschlagsmengen, die in Deutschland im Mittel gemessen und vorhergesagt werden. In diesem Abschnitt soll die Frage beantwortet werden, wie häufig es zu Niederschlägen kommt.

Tägliche Niederschlagsereignisse

Im ersten Schritt wurde jahreszeitlich aufgeteilt die relative Häufigkeit an Niederschlagsereignissen (Niederschlag ≥ 0.1 mm) an jedem Gitterpunkt bestimmt. Der Schwellwert (SW) ist mit 0.1 mm als der Messgrenze für Niederschlagsmengen gesetzt. Es gibt in der Literatur einige Hinweise darauf, dass Messungen dieser geringen Mengen (für Tageswerte) sehr sensitiv vom Beobachter abhängen (z.B. Frei et al., 2003). Deshalb wurde vor der Auswahl der im Folgenden gezeigten Bilder auch ein

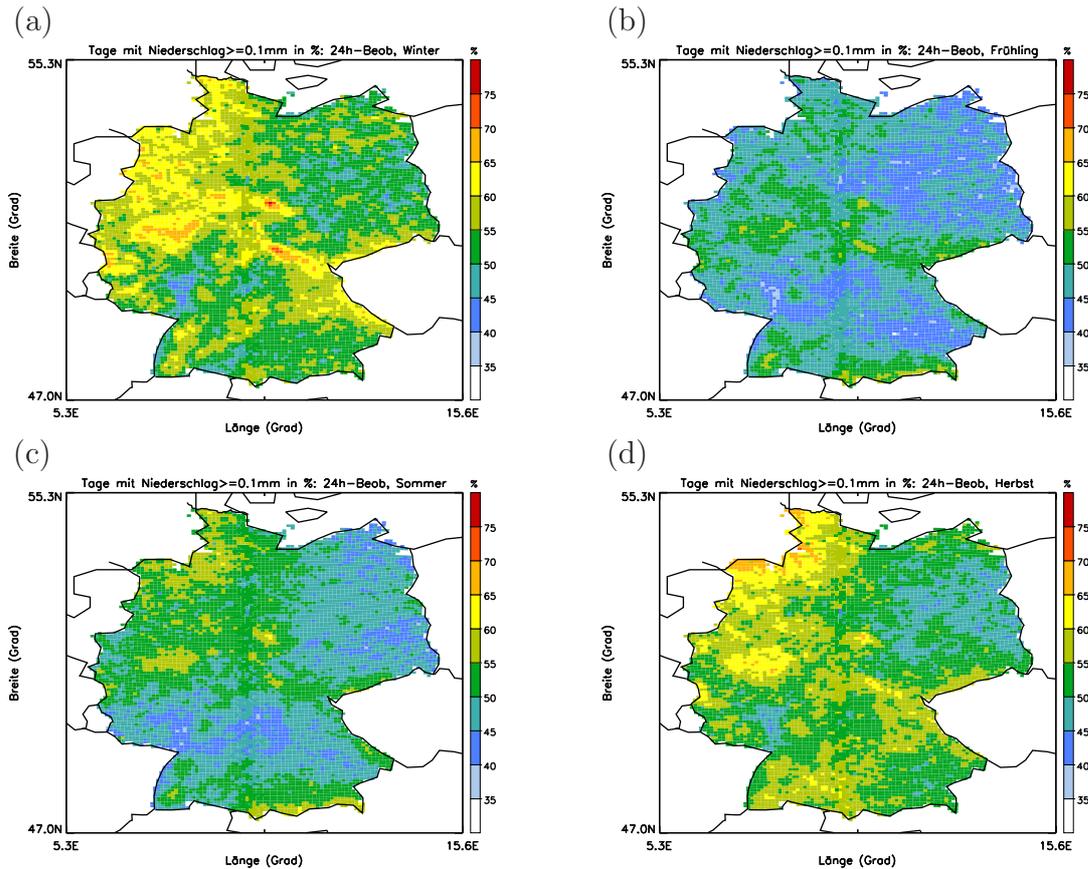


Abbildung 5.1: Beobachtete Häufigkeiten von Niederschlagstagen mit Niederschlagssummen ≥ 0.1 mm in Prozent (Gesamtheit sind alle Tage der jeweiligen Jahreszeit) an jedem Gitterpunkt für die Jahre 2001-2004: (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

Vergleich der Häufigkeitskarten zwischen den Schwellwerten 0.1 mm (Abb. 5.1) und 1 mm (Anh. A) gemacht. Dies ist insbesondere für die Häufigkeitskarten für Tagessummen von Bedeutung, da Tagessummen an vielen Stationen immer noch von Beobachtern und nicht automatisch gemessen werden. Allerdings zeigten die Vergleiche qualitativ die gleichen Verteilungen für Niederschlagshäufigkeiten (quantitativ natürlich einige Unterschiede), so dass, gerade auch im Hinblick auf den Vergleich zwischen Tages- und Stundensummen, der Schwellwert von 0.1 mm beibehalten wurde.

In Abb. 5.1 sind diese Häufigkeiten für tägliche Niederschlagssummen aus Beobachtungen dargestellt. Im Winter (Abb. 5.1 (a)) fällt deutschlandweit am häufigsten Niederschlag mit Maximalwerten zwischen 70 und 80 % (Harz und Thüringer Wald). Regionen mit ebenfalls relativ hohen Werten zwischen 60 und 70 % findet man in West- und Nordwestdeutschland und in den Hochlagen verschiedener Mittelgebirge. Relativ selten sind Niederschläge im Winter in Teilen Ostdeutschlands und am Oberrhein (40 bis 50 %). Im Mittel über ganz Deutschland gibt es im Winter an 50 bis 60 % der Tage Niederschlag.

Ein ähnliches Bild der Häufigkeitsverteilung zeigt sich im Herbst (Abb. 5.1 (d)). Unterschiede zum Winter finden sich in Ostdeutschland, wo im Herbst die Regionen mit relativ wenigen Niederschlagsereignissen noch ausgedehnter sind. Außerdem sind die

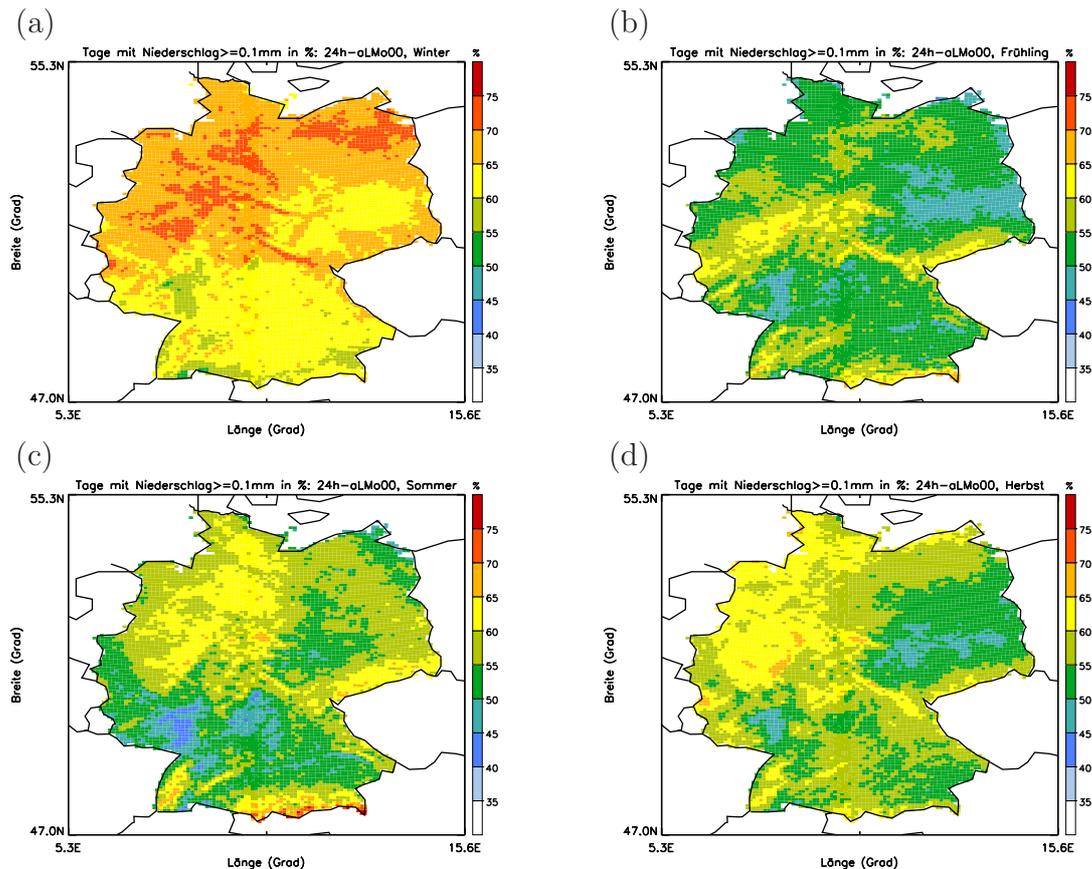


Abbildung 5.2: Wie Abb. 5.1 für aLMO Vorhersagen (00 UTC-Läufe).

Häufigkeitsmaxima kleiner als 70 %. An der Nordseeküste sind die Häufigkeiten im Herbst etwas höher als im Winter. Im Mittel findet man allerdings ebenfalls Häufigkeiten von Niederschlagstagen zwischen 50 und 60 %.

Im Frühling und Sommer (Abb. 5.1 (b) und (c)) sind dagegen weniger Niederschlagstage zu verzeichnen. Im Mittel fallen im Frühling an etwa 50 %, im Sommer an etwa 55 % der Tage Niederschlag.

Vergleicht man diese Häufigkeiten mit den Niederschlagssummen aus Abb. 4.2, so fällt auf, dass im Sommer zwar relativ selten Niederschlag fällt, jedoch dieser dann mit relativ hoher Intensität (da im Sommer die größten Niederschlagsmengen beobachtet werden). Auch im Herbst (Abb. 4.2 (d)) fällt, mit Ausnahme Ostdeutschlands, relativ viel Niederschlag, jedoch sind Niederschlagsereignisse im Mittel über Deutschland häufiger, haben also eine geringere Intensität als im Sommer. Diese im Vergleich zum Sommer geringeren Intensitäten finden sich mit der gleichen Argumentationskette aus Niederschlagshäufigkeit und Niederschlagsmenge auch für Winter und Frühling (Abb. 4.2 (a) und (b)).

Abbildung 5.2 zeigt die gleichen Häufigkeitsverteilungen wie Abb. 5.1 für aLMO-Vorhersagen aus 00 UTC-Läufen. In allen vier Jahreszeiten zeigen die Modellvorhersagen eine meist gute Übereinstimmung mit dem räumlichen Muster der Beobachtungen, allerdings auch eine Überschätzung der Niederschlagshäufigkeiten. In Winter, Frühling und Sommer (Abb. 5.2 (a), (b) und (c)) betragen diese Überschätzungen in weiten Teilen 10 %, in Norddeutschland im Winter sogar bis zu 20 %. Im

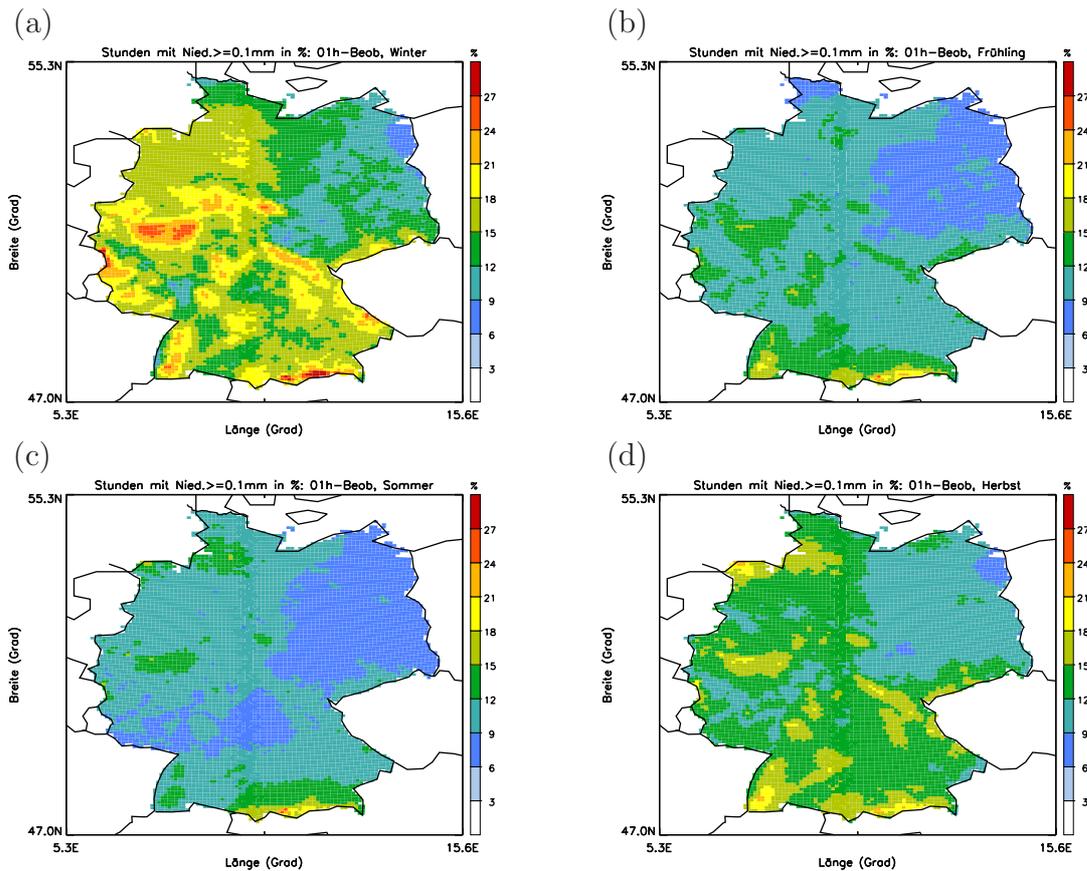


Abbildung 5.3: Wie Abb. 5.1 für stündliche Niederschlagssummen ≥ 0.1 mm.

Herbst (Abb. 5.2 (d)) zeigen die Vorhersagen ein ähnliches Bild wie die Beobachtungen (Abb. 5.1 (d)). Vor allem in West- und Ostdeutschland ist aber auch hier eine Überschätzung der Niederschlagshäufigkeit um mehrere Prozent erkennbar. Vergleicht man diese Ergebnisse aus den oben diskutierten Häufigkeitskarten mit den Werten in Tab. 4.1, so kann man die Schlussfolgerung ziehen, dass das aLMO im Mittel für Deutschland zu viele schwache Niederschläge simuliert. Dieses Ergebnis findet sich auch in Arbeiten über Niederschlagsklimatologien (Osborn und Hulme, 1997; Frei et al., 2003) und aktuell innerhalb neuer Ergebnisse aus dem COSMO (COMSO, 2006) wieder.

Stündliche Niederschlagsereignisse

Wie schon im Abschnitt zuvor werden hier Häufigkeiten für Niederschläge ≥ 0.1 mm ausgewertet, allerdings diesmal für stündliche Niederschlagssummen. Dazu wurde der disaggregierte Datensatz (Kap. 3) verwendet.

Abbildung 5.3 zeigt Häufigkeitskarten (prozentual zur Gesamtheit aller betrachteten Niederschlagsstunden) für Deutschland aus diesen disaggregierten Niederschlagssummen für alle Jahreszeiten. Es wird sofort ein deutlicher Unterschied zu den Häufigkeiten für Tagessummen (Abb. 5.1) erkennbar. Auf stündlicher Zeitskala liegen die Häufigkeiten von Niederschlagsereignissen deutlich niedriger ($< 30\%$). Dies ist aber nicht verwunderlich, da die Gesamtheit an möglichen Ereignissen um

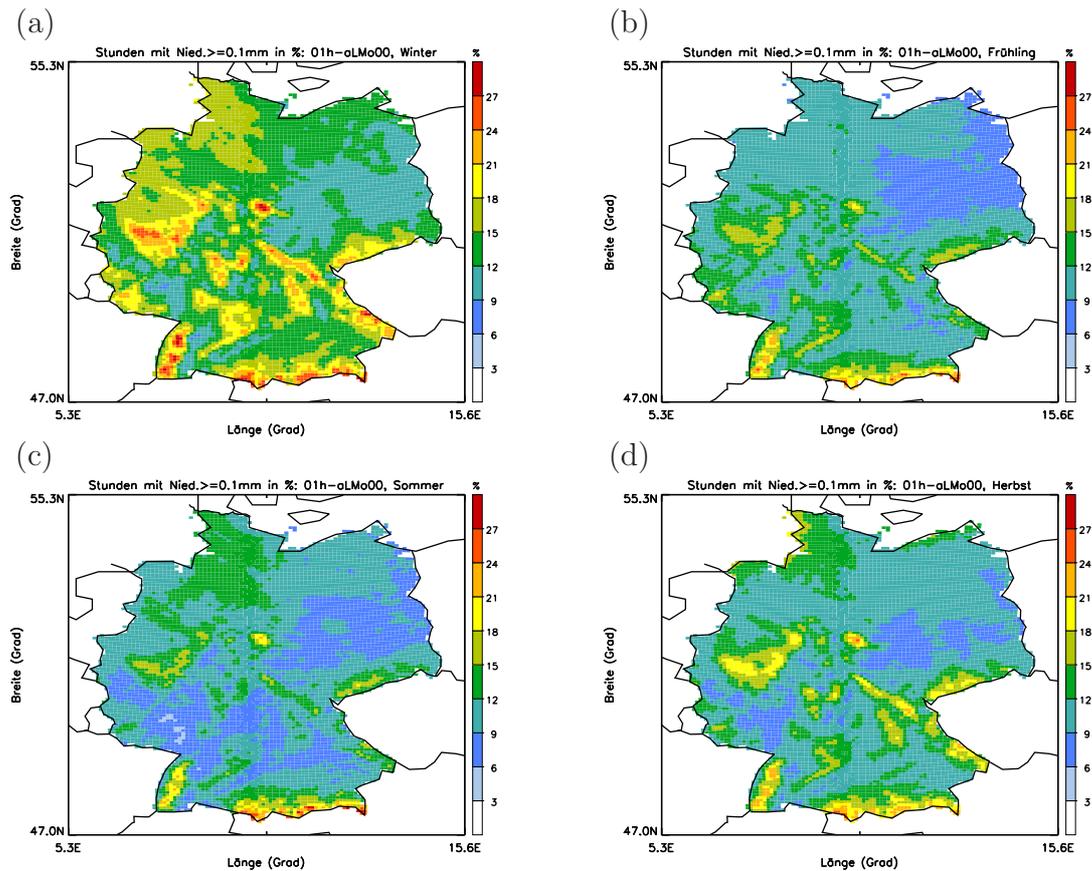


Abbildung 5.4: Wie Abb. 5.2 für stündliche Niederschlagssummen ≥ 0.1 mm.

das 24-fache steigt, dagegen aber nicht an jedem Niederschlagstag in allen 24 Stunden Niederschlag fällt.

Vergleichbar zu den Ergebnissen für Tagessummen ist in den Beobachtungen besonders im Winter, aber auch im Herbst (Abb. 5.3 (a) und (d)) die Häufigkeit an Ereignissen am größten (im Mittel über Deutschland). Ebenfalls ähnlich zu den Häufigkeiten aus Tagessummen ist der Frühling (Abb. 5.3 (b)) bei stündlichen Häufigkeiten einzuordnen. Im Sommer (Abb. 5.3 (c)) finden sich wiederum (zusammen mit dem Frühling) die niedrigsten Häufigkeiten (im Mittel über Deutschland). Allerdings fallen im Sommer stündlich noch seltener Niederschläge als im Frühling, was bei den Tagessummen umgekehrt der Fall war. Dies ist ein weiteres Indiz dafür, dass im Sommer die konvektiven Niederschläge relativ selten und kurzzeitig sind, dafür aber oft mit relativ hoher Intensität.

Im Gegensatz zu Häufigkeiten aus Tagessummen wurden keine so deutlichen Unterschiede zwischen stündlichen aLMo-Vorhersagen und Beobachtungen gefunden. Dies ist im Vergleich zu Abb. 5.4 für Häufigkeiten im aLMo gut herauszulesen. Im Winter sind im aLMo vor allem stärkere Kontraste zwischen Mittelgebirgen und tiefer liegenden Regionen im Vergleich zu den Beobachtungen erkennbar (besonders stark in der Schwarzwaldregion). Im Frühling, Sommer und Herbst gibt es in einigen Mittelgebirgsregionen und dem Alpenvorland Häufigkeitsmaxima von bis zu 30 % im aLMo. Außerdem sind in einigen kleineren Regionen die Häufigkeiten im Sommer im aLMo etwas höher (z.B. Teile Ostdeutschlands). Im Herbst zeigt sich außer in

den Maxima noch der Unterschied, dass die Häufigkeiten aus dem disaggregierten Datensatz (vor allem im Osten und Südwesten) bei 9 bis 15 % gleichmäßiger verteilt sind als im aLMo.

Es stellt sich nun die Frage, wie es sein kann, dass bei Tagessummen die Häufigkeiten vom aLMo teils deutlich überschätzt werden, diese Überschätzung aber nicht so deutlich auf stündlicher Basis zur Geltung kommt?

Drei Gründe können hierbei eine Rolle spielen (sind aber in ihrer Bedeutung schwierig zu quantifizieren):

1. das aLMo simuliert auf stündlicher Basis häufiger Niederschläge ≤ 0.1 mm. Diese gehen nicht in die Häufigkeitskarten aus Abb. 5.4 ein. Die Tagessumme aus mehreren solcher stündlichen Niederschläge kann aber 0.1 mm überschreiten und geht somit in die 24-stündigen Häufigkeiten ein.
2. Die Fehler in der Häufigkeit von Niederschlagsereignissen (Anzahl Tage mit Niederschlag) und der Dauer dieser Ereignissen kompensieren sich zum Teil gegenseitig.
3. Das Radar sieht wegen diverser Effekte Niederschläge, welche aber nicht am Boden beobachtet werden (über mehrere Stunden möglich). Wenn in einer Stunde am Tag z.B. ein hoher Niederschlag fällt, so wird dieser sowohl vom Radar als auch am Boden (tägliche Messung) gemessen, aber durch die Disaggregation auf (unter Umständen) alle Stunden am Tag aufgeteilt (Abb. 3.6).

5.2 Intensitäten

Nachdem im vorhergehenden Abschnitt die Häufigkeiten von Niederschlagsereignissen gezeigt wurden, soll hier beschrieben werden, welche Intensitäten diese Ereignisse haben.

Tägliche Niederschlagsereignisse

In diesem Abschnitt werden Intensitätsverteilungen für Tagessummen im Niederschlag betrachtet. Dazu wurden Intensitätsintervalle vorgegeben, für welche dann Häufigkeiten aus der Gesamtanzahl an Niederschlagsereignissen (Abschn. 5.1) gezählt wurden. Die Intervalle haben eine Breite von 0.5 mm, wobei das erste Intervall I_1 von 0.1 bis 0.5 mm/24h, das zweite Intervall I_2 von 0.5 bis 1 mm/24h, das dritte Intervall I_3 von 1 bis 1.5 mm/24h definiert ist (usw.).

Abbildung 5.5 zeigt die so erstellten Intensitätsverteilungen für Winter und Sommer im Vergleich zwischen gegitterten Beobachtungen und aLMo-Vorhersagen (00 UTC-Läufe). Aus diesen Intensitätskurven können mehrere Aussagen abgeleitet werden:

- relativ schwache Niederschläge (< 3 mm/24h) sind im Winter häufiger als im Sommer (Beobachtungen und aLMo),

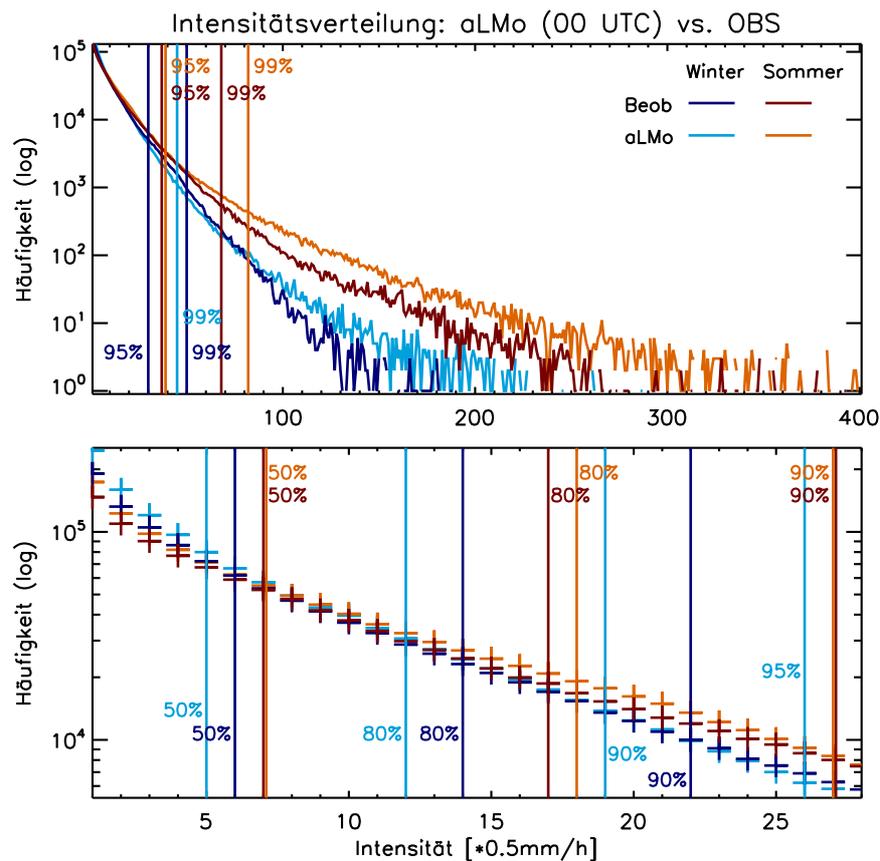


Abbildung 5.5: Häufigkeitsverteilung für tägliche Niederschlagsintensitäten für Winter (Blau) und Sommer (rötlich) für Deutschland (2001-2004): aLMO (00 UTC-Läufe) in heller, Beobachtungen in dunkler Farbgebung. Intervalleinteilung wie im Text beschrieben, wobei das untere Bild einen Ausschnitt der ersten 29 Intervalle zeigt. Eingezeichnet sind Perzentile für 50, 80, 90, 95 und 99 %.

- relativ starke Niederschläge (> 8 mm/24h) sind im Sommer häufiger als im Winter (Beobachtungen und aLMO),
- das aLMO simuliert mehr relativ schwache Niederschläge als beobachtet werden (im Winter bis zu einer Höhe von 3.5 mm/24h, im Sommer 4 mm/24h),
- das aLMO simuliert mehr relativ starke Niederschläge als beobachtet werden (im Winter ab 40 mm/24h, im Sommer ab 25 mm/24h).

Weiter sind in Abb. 5.5 Perzentile für 50, 80, 90, 95 und 99 % eingezeichnet. In Tab. 5.1 sind die jeweiligen Perzentilwerte für aLMO und Beobachtungen für Winter und Sommer aufgeführt (z.B. liegt das 80 %-Perzentil im Sommer bei den Beobachtungen zwischen 8 und 8.5 mm/24h. Das bedeutet, dass 80 % der täglichen Niederschläge kleiner als 8.5 mm/24h sind). Aus Tab. 5.1 ist abzulesen, dass (wie erwartet) alle Perzentile im Sommer größere Niederschlagswerte annehmen als im Winter. Interessant ist, dass die Perzentilwerte im Sommer im aLMO und den Beobachtungen fast identisch sind und erst bei 99 % das aLMO einen signifikant höheren Wert hat. Diese höheren Werte sollten aber nicht zwingend mit einer Überschätzung der hohen Intensitäten im aLMO erklärt werden. Grund dafür ist auch die eingeschränkte

Perz.	Winter [mm/h]		Frühling [mm/h]		Sommer [mm/h]		Herbst [mm/h]	
	Beob	aLMo	Beob	aLMo	Beob	aLMo	Beob	aLMo
50 %	2.5-3	2-2.5	2.5-3	2-2.5	3-3.5	3-3.5	2.5-3	2-2.5
80 %	6.5-7	5.5-6	6.5-7	6-6.5	8-8.5	8.5-9	7.5-8	6.5-7
90 %	10.5-11	9-9.5	10-10.5	9.5-10	13-13.5	13-13.5	11.5-12	10-10.5
95 %	14.5-15	12.5-13	14-14.5	13.5-14	18-18.5	19-19.5	16-16.5	14-14.5
99 %	24.5-25	22-22.5	25-25.5	24.5-25	33-33.5	39.5-40	27.5-28	25.5-26

Tabelle 5.1: Häufigkeitsperzentile für Niederschlagsintensitäten für alle Jahreszeiten aus täglichen Beobachtungen und aLMo Vorhersagen (Deutschland, 2001-2004).

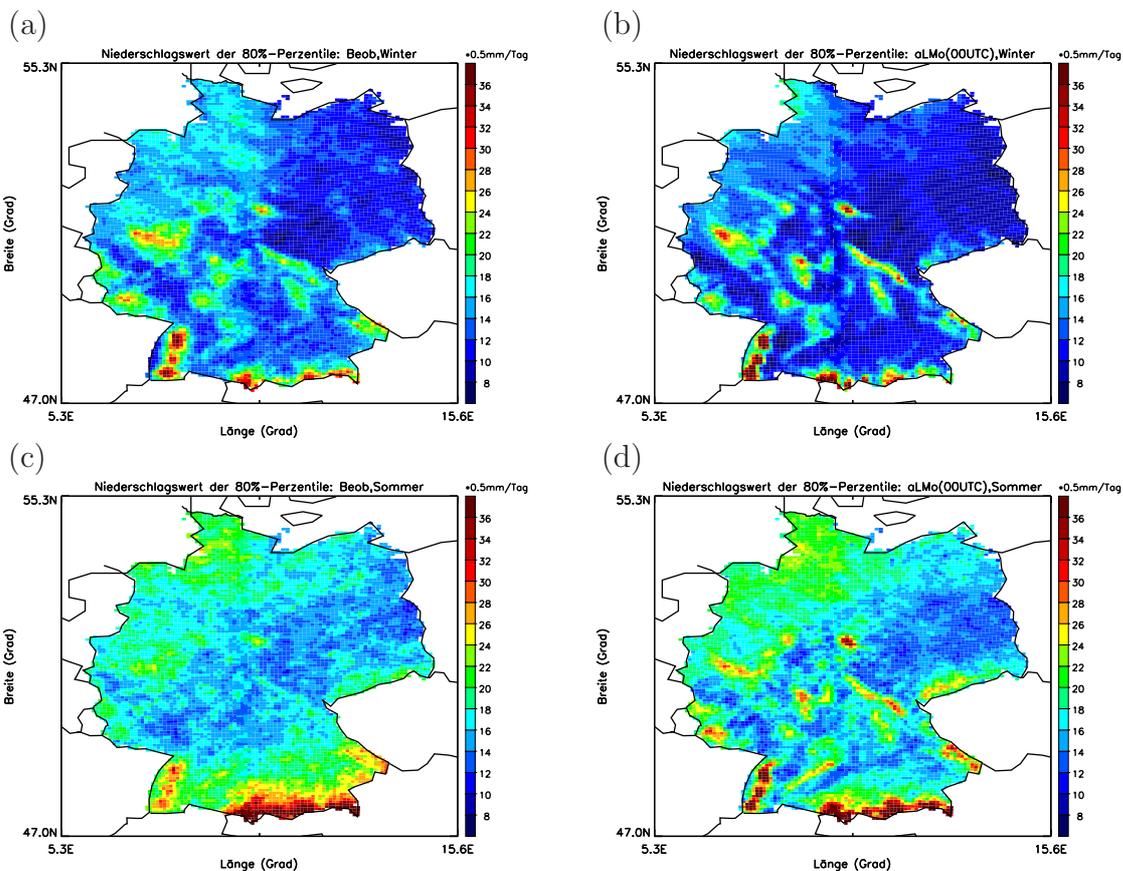


Abbildung 5.6: 80 %-Perzentil der Häufigkeit von täglichen Niederschlagsintensitäten an jedem Gitterpunkt (2001-2004). Skala zeigt Intervall für bestimmte Niederschlagsmengen an (Intervalle wie in Abb. 5.5): (a) Beobachtungen Winter, (b) aLMo (00 UTC-Läufe) Winter, (c) Beobachtungen Sommer, (d) aLMo (00 UTC-Läufe) Sommer.

Stationsdichte und damit effektive, niedrigere Auflösung als im aLMo. Dies hat zur Folge, dass die Intensitätsverteilung schneller abfällt als im Modell.

Im Winter liegen die Perzentilwerte des aLMo (wenn auch relativ wenig) unterhalb den Werten aus Beobachtungen. Die Begründung dafür findet sich in Abb. 5.5. Es gibt einen Bereich zwischen 12 und 40 mm/24h, in dem die Häufigkeiten aus Beobachtungen über denen des aLMo liegen.

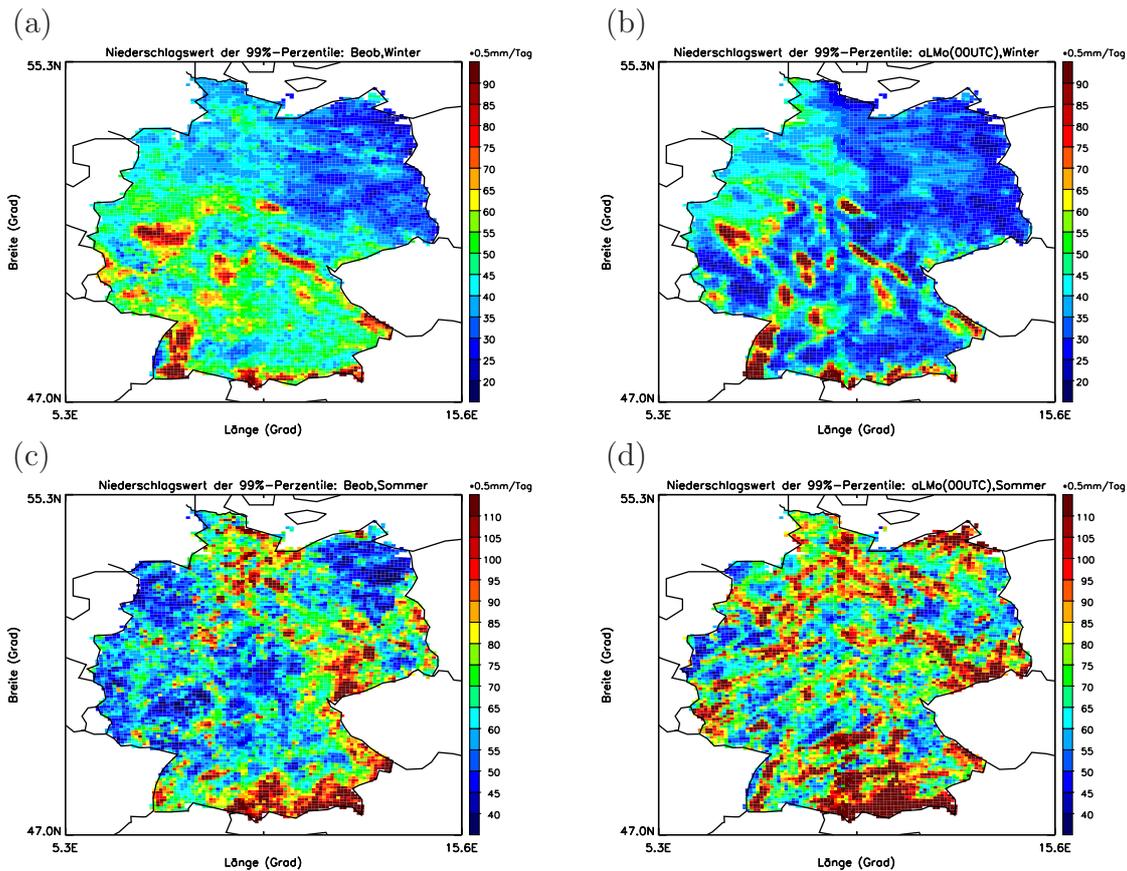


Abbildung 5.7: Wie Abb. 5.6 für das 99%-Perzentil.

Der Vollständigkeit halber sind in Tab. 5.1 auch Perzentilwerte für Frühling und Herbst aufgeführt. Den Erwartungen entsprechend sind diese für alle Perzentile sehr ähnlich und liegen zwischen den Werten für Sommer und Winter.

Abbildung 5.6 und 5.7 zeigen im Detail die räumliche Verteilung der Perzentile für 80 und 99 % für aLMO und Beobachtungen im Winter und Sommer. Zunächst ist wichtig zu erwähnen, dass die Farbskala in diesen Abbildungen gleich ist, aber unterschiedliche Konturwerte zugrunde liegen. Anhand Abb. 5.6 ist zu erkennen, dass den mittleren gefundenen Perzentilwerten aus Abb. 5.5 eine räumlich stark unterschiedliche Verteilung in Deutschland zugrunde liegt (für das 99 %-Perzentil beginnt die Verteilung zum Teil räumlich zufällig zu werden). Höhere Intensitäten für die Perzentile an jedem Gitterpunkt finden sich in Regionen, in denen es auch typischerweise mehr Niederschlag gibt (Mittelgebirgsregionen und Alpenvorland, Kap. 4). Im Winter findet man im aLMO deutliche Unterschätzungen im Nordwesten, dagegen eine Überschätzung der Maxima im Schwarzwald (und anderer Mittelgebirge). Im Sommer sind die Werte in den Mittelgebirgen viel zu stark ausgeprägt. Es ist unwahrscheinlich, dass hier die weiter oben erwähnte, eingeschränkte effektive Auflösung im Stationsnetz alleine verantwortlich ist, da hier Regionen betroffen sind, die über besonders dichte Messnetze verfügen (z.B. Schwarzwald und Harz, vgl. Abschn. 2.3). Erstaunlich ist auch, dass das 80 %-Perzentil im Sommer in Norddeutschland eher überschätzt, an der Donau eher unterschätzt wird (wobei für 99 % dann auch eine Überschätzung vorliegt).

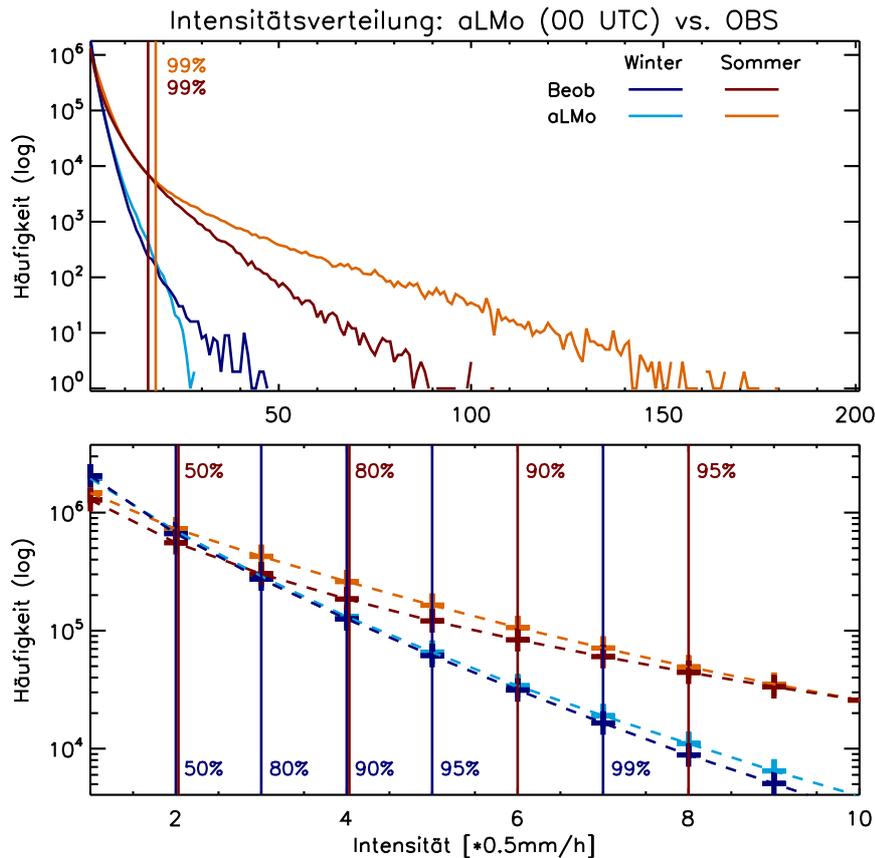


Abbildung 5.8: Wie Abb. 5.5 für stündliche Niederschlagsintensitäten, wobei das untere Bild hier einen Ausschnitt der ersten 10 Intervalle zeigt.

Von der Struktur her finden sich die Ergebnisse aus den Karten für das 80 %-Perzentil in den Karten für das 90 und 95 %-Perzentil wieder (nicht gezeigt), wie erwähnt nicht mehr so deutlich im Sommer für das 99 %-Perzentil (Abb. 5.7). Die Ähnlichkeiten in der Struktur gehen einher mit verschiedenen Extremwerten: Minimum des 80 %-Perzentils: 3.5 mm/24h, Maximum des 80 %-Perzentils: 37 mm/24h, Minimum des 90 %-Perzentils: 4.5 mm/24h, Maximum des 90 %-Perzentils: 44.5 mm/24h, Minimum des 95 %-Perzentils: 6 mm/24h, Maximum des 95 %-Perzentils: 68 mm/24h, Minimum des 99 %-Perzentils: 9.5 mm/24h, Maximum des 99 %-Perzentils: 112.5 mm/24h.

Stündliche Niederschlagsereignisse

Mit dem in Kap. 3 vorgestellten disaggregierten Datensatz ist es möglich, wie im Abschnitt zuvor für Tagessummen nun auf Stundensummen basierende Intensitätsverteilungen auszuwerten. Dazu wurden Abbildungen genau wie im Abschnitt zuvor erstellt.

Abbildung 5.8 zeigt die Häufigkeit stündlicher Intensitäten für Niederschlagsintervalle der Breite 0.5 mm/h (näheres im Abschnitt zuvor für Tagessummen). Es können einige sehr ähnliche Schlussfolgerungen wie für Tagessummen gezogen werden, mit dem Unterschied, dass die Intensitäten einerseits nicht so stark werden können, zum

Perz.	Winter [mm/h]		Frühling [mm/h]		Sommer [mm/h]		Herbst [mm/h]	
	Beob	aLMo	Beob	aLMo	Beob	aLMo	Beob	aLMo
50 %	0.5-1	0.5-1	0.5-1	0.5-1	0.5-1	0.5-1	0.5-1	0.5-1
80 %	1-1.5	1-1.5	1-1.5	1-1.5	1.5-2	1.5-2	1-1.5	1-1.5
90 %	1.5-2	1.5-2	1.5-2	1.5-2	2.5-3	2.5-3	1.5-2	2-2.5
95 %	2-2.5	2-2.5	2-2.5	2.5-3	3.5-4	3.5-4	2.5-3	2.5-3
99 %	3-3.5	3-3.5	3.5-4	4-4.5	7.5-8	8.5-9	4-4.5	4-4.5

Tabelle 5.2: Häufigkeitsperzentile für Niederschlagsintensitäten für alle Jahreszeiten aus stündlichen Beobachtungen und aLMo Vorhersagen (Deutschland, 2001-2004).

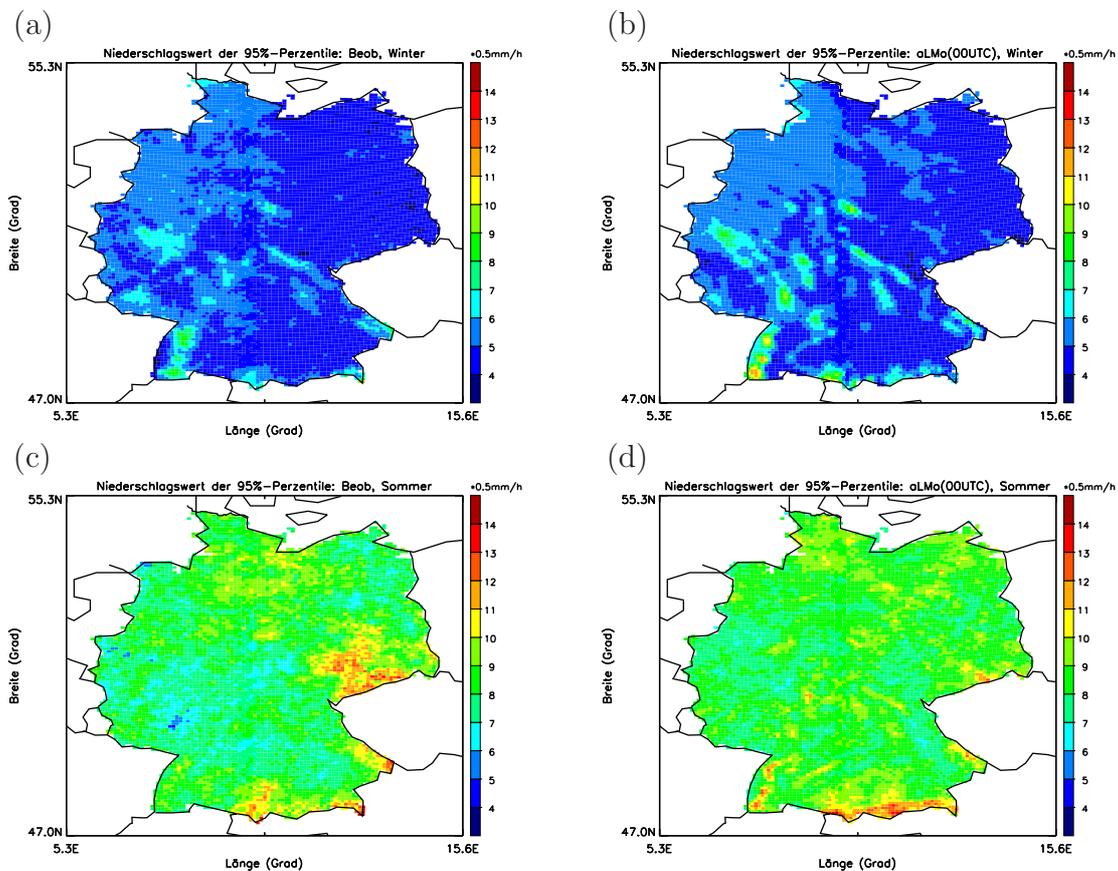


Abbildung 5.9: 95 %-Perzentil der Häufigkeit von stündlichen Niederschlagsintensitäten an jedem Gitterpunkt (2001-2004). Skala zeigt Intervall für bestimmte Niederschlagsmengen an (Intervalle wie in Abb. 5.5): (a) gegitterte Beobachtungen Winter, (b) aLMo (00 UTC-Läufe) Winter, (c) gegitterte Beobachtungen Sommer, (d) aLMo (00 UTC-Läufe) Sommer.

anderen besonders schwache Ereignisse häufiger sind (was natürlich aus der höheren möglichen Anzahl an Ereignissen folgt):

- schwache Niederschläge (< 1 mm/h) sind im Winter häufiger als im Sommer (disaggregierter Datensatz und aLMo),

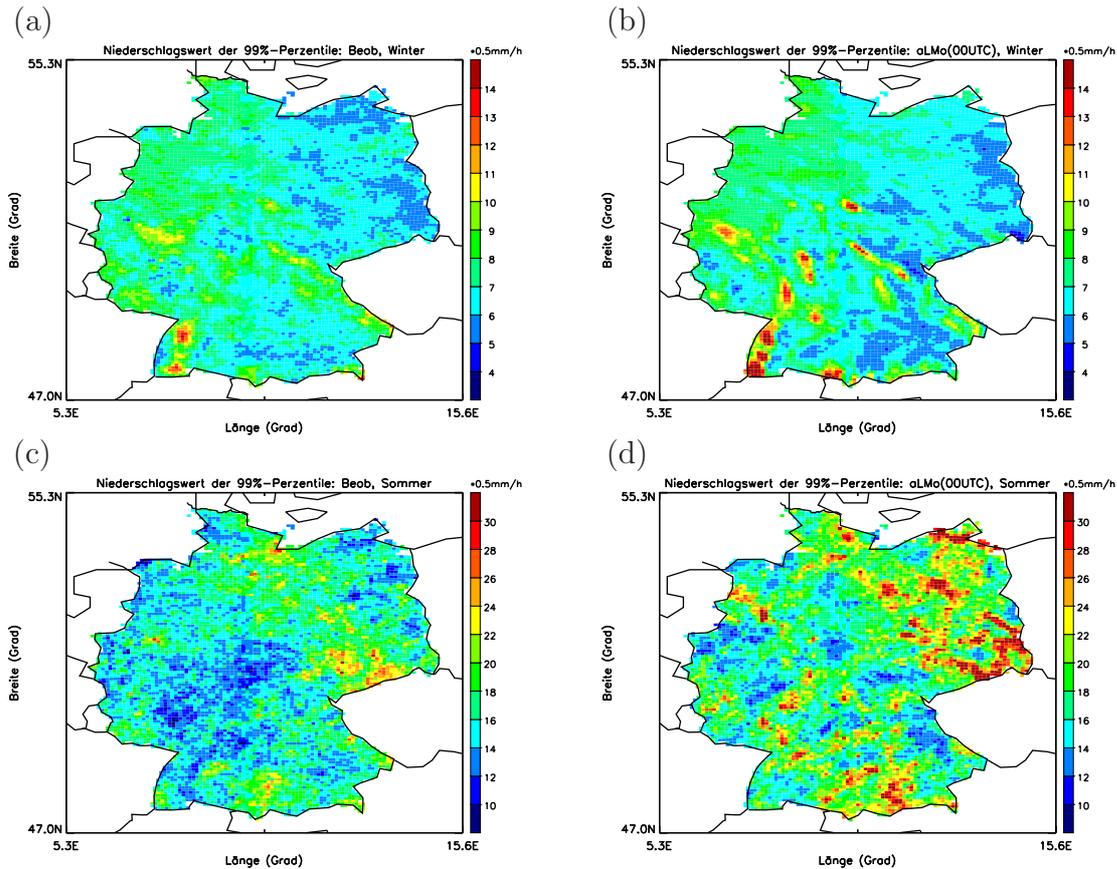


Abbildung 5.10: Wie Abb. 5.9 für das 99%-Perzentil.

- Niederschläge (> 1.5 mm/h) sind im Sommer häufiger als im Winter (Beobachtungen und aLMo),
- das aLMo simuliert im Sommer mehr relativ schwache (< 4 mm/h) Niederschläge als im disaggregierten Datensatz vorhanden, nicht jedoch im Winter,
- das aLMo simuliert im Sommer mehr relativ starke (> 10 mm/h) Niederschläge als im disaggregierten Datensatz vorhanden,
- im Winter sind im Gegensatz zu Tagessummen mehr relativ starke Niederschläge (> 11 mm/h) im disaggregierten Datensatz als vom aLMo vorhergesagt,
- Niederschläge zwischen 3 und 11 mm/h sind im aLMo im Winter häufiger.

Ebenfalls identisch zu den Auswertungen der Tagessummen sind in Tab. 5.2 Niederschlagsintervalle für verschiedene Perzentile aufgeführt (50, 80, 90, 95 und 99%). Aufgrund der wie für Tagessummen gewählten Intervallgröße von 0.5 mm (hier pro Stunde) ist es einleuchtend, dass auf stündlicher Zeitskala die Perzentilintervalle für aLMo und die disaggregierten Daten fast immer identisch sind (in Abb. 5.8 sind daher auch nur jeweils die Perzentile der disaggregierten Daten dargestellt, mit Ausnahme des 99%-Perzentils). Unterschiede gibt es aber auch auf stündlicher Basis zwischen den Jahreszeiten. So sind schwache Niederschläge im Winter deutlich

häufiger, während im Sommer stärkere Niederschläge dominieren. Erst für das 99 %-Perzentil zeigen sich auch Unterschiede zwischen aLMo (9 mm) und Beobachtungen (8 mm) im Sommer.

Analog zu den Auswertungen für Tagessummen zeigen die beiden Abb. 5.9 und 5.10 Perzentilkarten. Die Karten für die Verteilungen des 80 und des 90 %-Perzentils werden nicht gezeigt, da sich dort kaum Unterschiede in der Verteilung über Deutschland finden. Ähnlich wie bei den Tagessummen findet man auch hier im Winter die höchsten Werte in den Gebirgsregionen (Maximum des 95 %-Perzentils: 5.5 mm/h, Maximum des 99 %-Perzentils: 10.5 mm/h). Das aLMo und der disaggregierte Datensatz verhalten sich hierbei relativ ähnlich, wobei in den Verteilungen in weiten Teilen eine leichte Überschätzung im aLMo zu sehen ist.

Im Sommer sind wie zu erwarten höhere Werte zu finden (Maximum des 95 %-Perzentils: 8 mm/h, Maximum des 99 %-Perzentils: 33 mm/h), insbesondere auch in den flachen Regionen Deutschlands. Der Zusammenhang mit den Mittelgebirgen sticht allerdings nicht mehr so deutlich heraus wie für den Winter, mit Ausnahme Süddeutschlands. Deutlich wird im Vergleich der 95 %-Perzentile von Stunden- zu Tagessummen, dass auf stündlicher Zeitskala der Unterschied zwischen Winter und Sommer größer ist.

Auffällig sind in den disaggregierten Daten in Abb. 5.9 (c) und Abb. 5.10 (c) auch Maxima in der Region des Erzgebirges. Eine mögliche Erklärung könnten die über einige Tage im August 2002 sehr hohen Intensitäten sein, die zur Elbeflut führten. Allerdings müsste diese Struktur dann wohl auch in den Tagessummen auftauchen, da es sich um ein um mehrere Tage anhaltendes Ereignis handelte. Dies ist nicht der Fall. Eine Erklärung muss daher in dieser Arbeit offen bleiben (der Grund dafür könnte aber wegen fehlerhafter Radarwerte aus der Disaggregation herrühren).

5.3 Der Tagesgang

Ein weiteres Anwendungsgebiet des disaggregierten Datensatzes ist die Untersuchung des Tagesgangs im Niederschlag, wofür mindestens stündlich aufgelöste Daten benötigt werden.

Abbildung 5.11 und 5.12 zeigen mittlere Tagesgänge für Winter und Sommer aus drei Datensätzen: stündliche Stationsmessungen (Grün), disaggregierte Beobachtungen (Blau) und aLMo-Vorhersagen (00 UTC-Läufe, Rot). Beide Abbildungen sind in drei Bilder aufgeteilt: Häufigkeit an Niederschlagsereignissen (≥ 0.1 mm/h) pro Station bzw. Gitterpunkt (oben), mittlere Intensität dieser Ereignisse (Mitte) und Niederschlagsmenge pro Station bzw. Gitterpunkt (unten), wobei letzteres gerade das Produkt von Häufigkeit und Intensität ist.

Abbildung 5.11 zeigt die Tagesgänge im Winter. In allen Datensätzen ist kein signifikanter Tagesgang zu erkennen, was im Winter auch nicht erwartet wird. Bei der mittleren Intensität liegen die Werte in aLMo und disaggregiertem Datensatz fast identisch um 0.5 mm/h, in den Stationsmessungen etwas darüber bei 0.6 mm/h. Beim Vergleich der Werte aus Disaggregation und Stationsmessungen fällt auf, dass Niederschläge in den Stationsmessungen seltener vorkommen, aber mit etwas höhe-

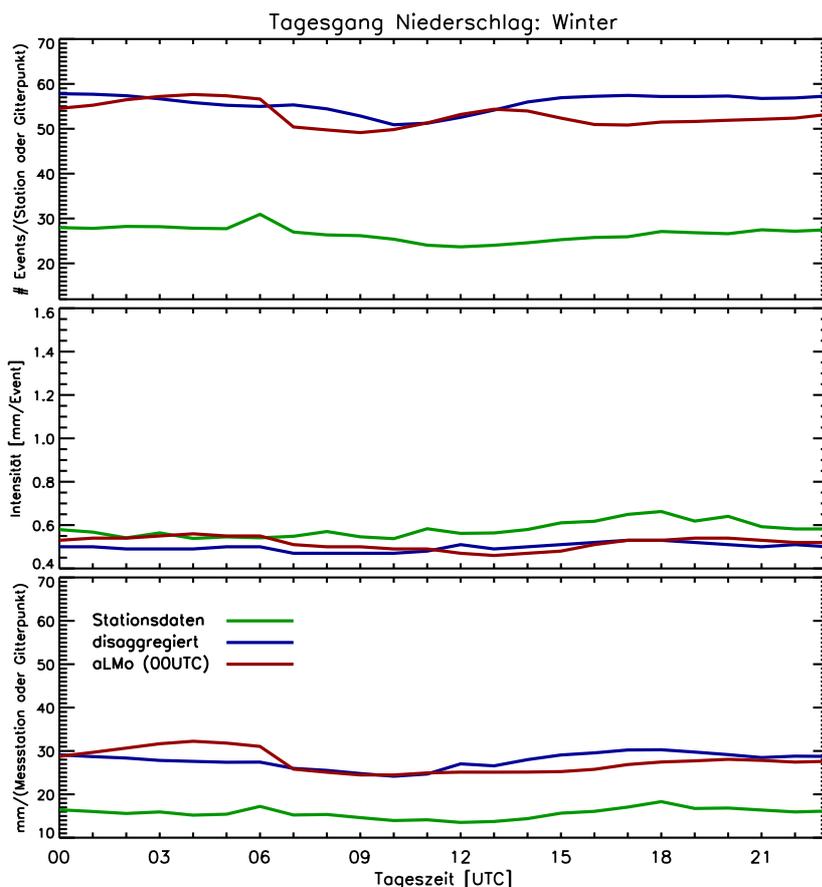


Abbildung 5.11: Tagesgang des mittleren Niederschlags im Winter aus stündlichen Messungen (Grün), disaggregiertem Datensatz (Blau) und aLMO-Vorhersagen (00 UTC-Läufe, Rot) in Deutschland. Oben: Häufigkeit von Niederschlagsereignissen (≥ 0.1 mm/h) pro Gitterpunkt bzw. Messstation, Mitte: Intensität pro Niederschlagsereignis, unten: Niederschlagsmenge (mm/h) pro Gitterpunkt bzw. Messstation.)

rer Intensität. Dies ist ein Resultat der unterschiedlichen räumlichen Repräsentativität von Gitterpunktwerten (Gebietsmittel) und Stationsmessungen (Punktmessungen). In Gebietsmittelwerten sind Niederschlagshäufigkeiten tendentiell höher als bei Punktmessungen, Intensitäten dagegen geglättet. Für die Niederschlagsmenge liegen die Stationsmessungen unterhalb der disaggregierten Beobachtungen. Der Grund dafür ist die Verteilung des stündlich messenden Stationsmessnetzes, dass für ein Deutschlandmittel nicht repräsentativ ist.

In allen drei Teilbildern in Abb. 5.11 ist zu sehen, dass das aLMO im Winter den „Tagesgang“ sehr gut simuliert. In Studien von Dai et al. (1999) für verschiedene Regionen der USA und Kaufmann et al. (2003) für die Schweiz finden sich vergleichbare Ergebnisse.

Abbildung 5.12 zeigt die gleichen Bilder wie in Abb. 5.11 für den Sommer. Aus dem Vergleich der beiden Beobachtungs-Datensätze folgt wiederum, dass wie für den Winter beschrieben hauptsächlich Gitterungseffekte, sowie die Repräsentativität des stündlichen Messnetzes für die Unterschiede verantwortlich sind. Der qualitative Verlauf ist aber in beiden Datensätzen sehr ähnlich. Alle drei Datensätze zeigen einen klaren Tagesgang. Wie erwartet finden sich die maximalen Niederschläge in

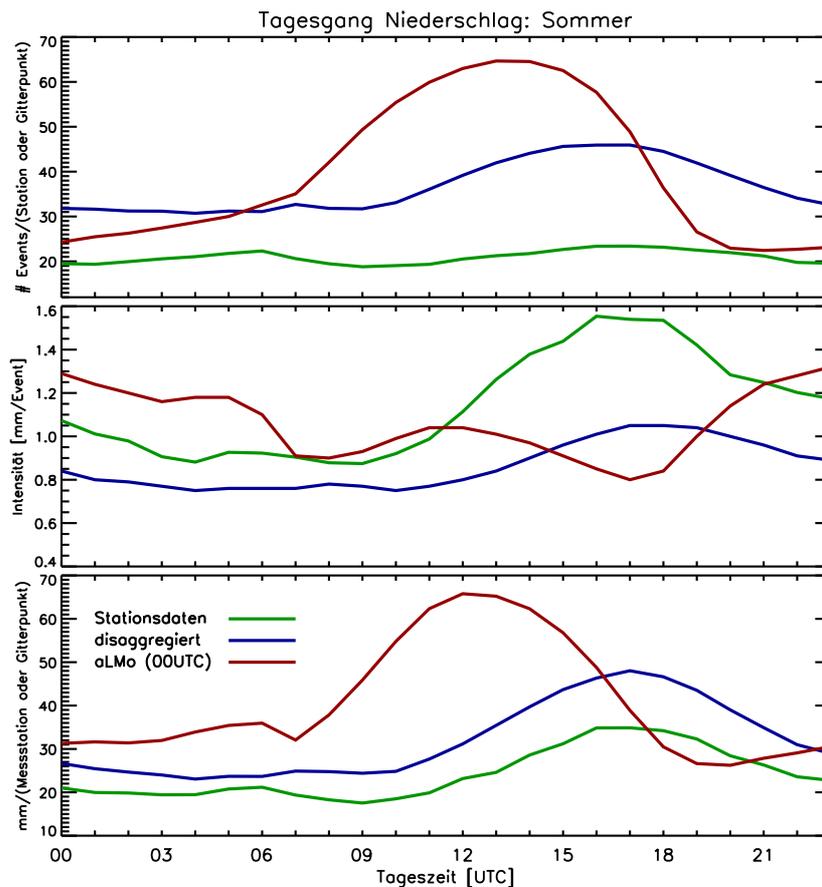


Abbildung 5.12: Wie Abb.5.11 für Sommer.

den Beobachtungen am späten Nachmittag bis Abend. Zu erwähnen ist, dass in den Stationsmessungen der Hauptanteil des Tagesgangs aus der Intensitätsverteilung herrührt, im disaggregierten Datensatz jedoch zusätzlich aus den Häufigkeiten. Dies ist ebenfalls ein Effekt der mangelnden Repräsentativität des Messnetzes auf stündlicher Basis. Wenn es in Mittelgebirgen und Alpen eine geringere Stationsdichte gibt, würde man genau dies erwarten.

Die Häufigkeiten liegen im Mittel deutlich unterhalb der Werte im Winter (für alle Datensätze), was mit den Ergebnissen aus Abschn. 5.1 konsistent ist. Dagegen liegen die mittleren Intensitäten aller drei Datensätze im Sommer (Stationsmessungen: ~ 1.2 mm/h, disaggregiert: ~ 0.9 mm/h, aLMO: ~ 1.1 mm/h) deutlich über denen im Winter, was ebenfalls mit den Ergebnissen aus Abschn. 5.2 konsistent ist.

Beim aLMO ist ein klarer Tagesgang in den Häufigkeiten mit einem deutlichen Maximum am späten Mittag zu erkennen. Im Intensitätsverlauf des aLMO ist kein typischer Tagesgang auszumachen, die intensiveren Niederschläge sind im Gegensatz zu den Beobachtungen sogar eher in der Nacht zu finden. Daraus resultiert (Abb. 5.12 unten), dass das aLMO zwar einen Tagesgang simuliert, aber

- zu früh am Tag,
- mit deutlich überschätztem Maximum (um einen Faktor 1.5),
- mit einem zu schnellen Ansteigen bis zum Maximum.

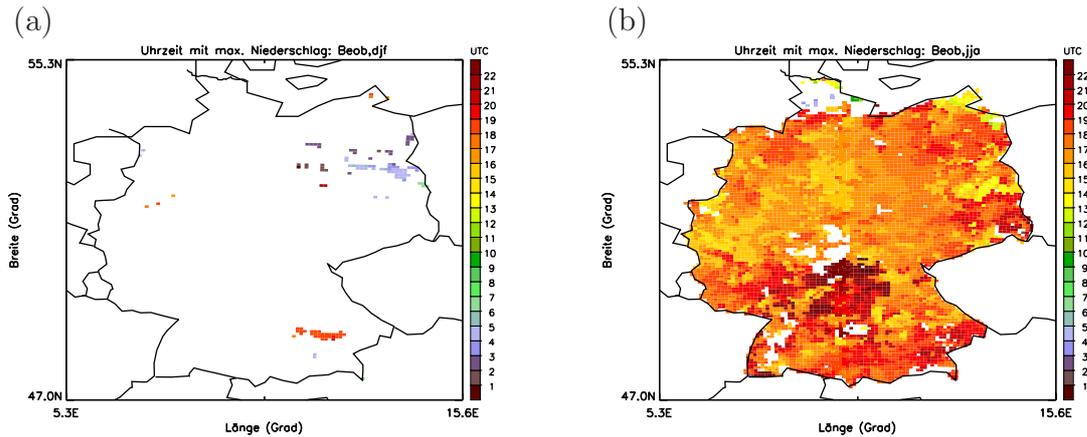


Abbildung 5.13: Uhrzeit (UTC) des Maximums im Tagesgang des Niederschlags für jeden Gitterpunkt aus Beobachtungen (Deutschland, 2001-2004): (a) Winter, (b) Sommer. Das Maximum ist als solches definiert, wenn es mindestens 30 % über dem Mittelwert des Tagesgangs am jeweiligen Gitterpunkt liegt, sonst wird keine Uhrzeit dargestellt.

Wie schon für den Winter finden sich diese Ergebnisse für den Sommer ebenfalls in sehr ähnlicher Weise in den Arbeiten von Dai et al. (1999) und Kaufmann et al. (2003).

Um sich einen Eindruck über die regionale Verteilung im Tagesgang des Niederschlags verschaffen zu können, wurden Deutschlandkarten angefertigt, die den Zeitpunkt der maximalen Niederschlagsmenge an jedem Gitterpunkt zeigen (im Mittel über die Jahre 2001-2004).

Abbildung 5.13 zeigt dies für den disaggregierten Datensatz für Winter (a) und Sommer (b). Um an einem Gitterpunkt einen Tagesgang zu identifizieren (wie im Mittel in Abb. 5.12 unten), soll folgendes Kriterium erfüllt sein:

$$\frac{R_{tgmax}(i, j) - \overline{R_{tg}(i, j)}}{\overline{R_{tg}(i, j)}} \geq 0.3, \quad (5.1)$$

wobei $\overline{R_{tg}(i, j)}$ den über den Tag gemittelten Niederschlagswert an einem Gitterpunkt (i, j) bezeichnet und $R_{tgmax}(i, j)$ den maximalen Niederschlagswert des mittleren Tagesgangs an einem Gitterpunkt. Alle Gitterpunkte, die dieses Kriterium nicht erfüllen, werden in Abb. 5.13 nicht dargestellt. Somit ergibt sich für den Winter eine fast datenfreie Karte für Deutschland, was bedeutet, dass es praktisch an allen Gitterpunkten keinen signifikanten Tagesgang gibt. Dagegen ist im Sommer in fast allen Regionen Deutschlands ein deutlicher Tagesgang erkennbar mit einem Niederschlagsmaximum am Nachmittag bis späten Abend. Ausnahmen sind hier Schleswig-Holstein und Teile in der Mitte und im äußersten Südwesten Deutschlands. In Schleswig-Holstein ist anzunehmen, dass dort auch im Sommer der dominierende Einfluss auf die Niederschläge aus Wechselwirkungen mit der Nord- und Ostsee herrühren (also eher ungebunden an typische Sommer-Konvektion). Etwas überraschend sind die Lücken in Mittel und vor allem im Südwesten Deutschlands (Schwarzwald). Offensichtlich ist dort der Niederschlag im Mittel relativ gleichverteilt, so dass die Bedingung aus Gl. 5.1 nicht erfüllt ist.

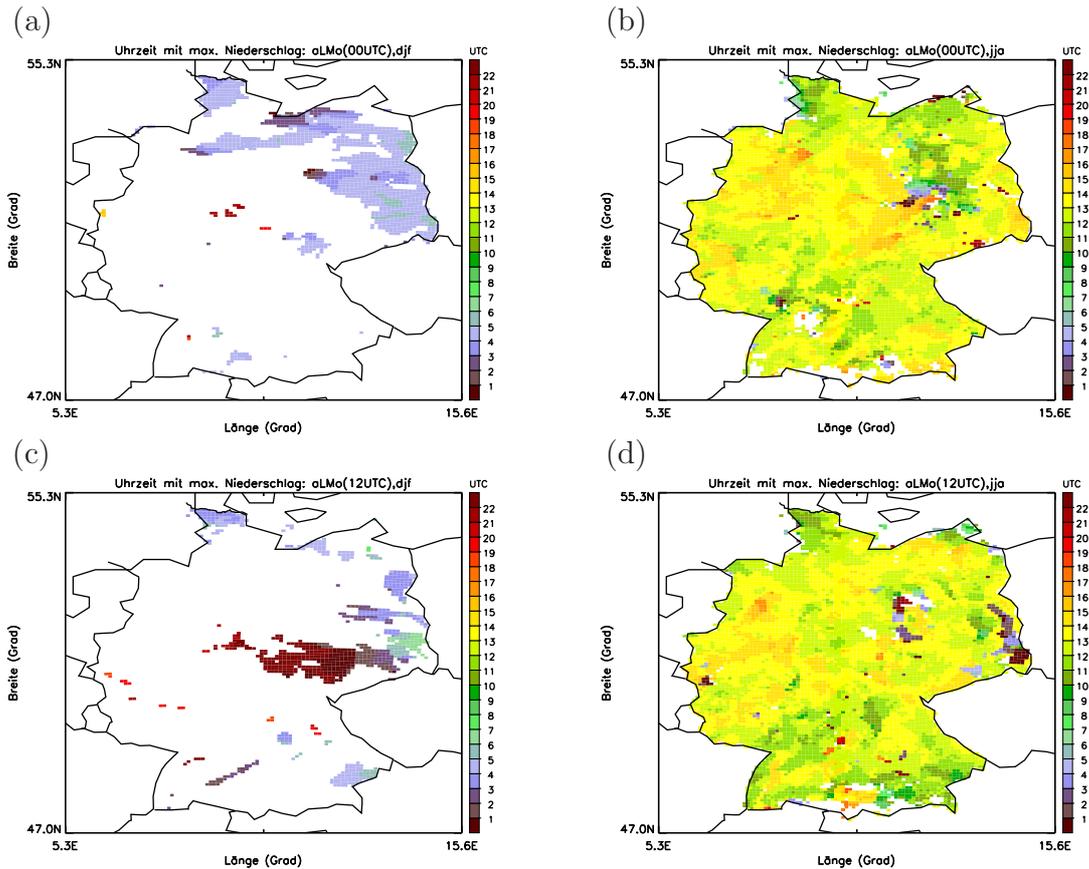


Abbildung 5.14: Wie Abb. 5.13 für aLMo Vorhersagen: (a) 00 UTC-Läufe Winter, (b) 00 UTC-Läufe Sommer, (c) 12 UTC-Läufe Winter, (d) 12 UTC-Läufe Sommer.

Ein Hinweis gilt der Tatsache, dass die zuvor gemachten Aussagen alle auf einer 4-jährigen Zeitreihe beruhen. Diese Zeitreihe ist nicht lang genug, um eventuelles „Rauschen“ an einzelnen Gitterpunkten zu filtern. Es sind also aus statistischen Gründen leichte sprunghafte Variationen in der Tageszeit mit maximalem Niederschlag möglich. Es ist wahrscheinlich, dass dies der Grund für die Strukturen im Winter (Abb. 5.14) ist. Einige wenige starke Ereignisse, zufällig zur gleichen Tageszeit, verursachen Spitzen im Tagesgang, die statistisch nicht signifikant sind.

Im Vergleich zum disaggregierten Datensatz werden in Abb. 5.14 die entsprechenden Zeiten für das Niederschlagsmaximum im Tagesgang an jedem Gitterpunkt für aLMo-Vorhersagen (00 und 12 UTC-Läufe) gezeigt. Im Winter ist sowohl für 00 als auch für 12 UTC-Läufe erkennbar, dass immer noch wenige, aber doch deutlich mehr Gitterpunkte als im disaggregierten Datensatz das Kriterium aus Gl. 5.1 erfüllen (Uhrzeiten in den Morgenstunden), vor allem in Regionen Ostdeutschlands. Aus diesen Regionen kommt der Hauptanteil des in Abb. 5.11 (unten) gefundenen schwachen Maximums am Morgen. Im Sommer findet sich in Deutschland ein klarer mittlerer Tagesgang für fast alle Gitterpunkte. Das Auftreten des Maximums ist aber

- in den Morgen- bis späten Mittagsstunden, also deutlich früher als in Abb. 5.13
- und etwas stärker uhrzeitlich gestreut.

Zusammenfassend zu sagen gibt es zwei Hauptprobleme in der Simulation des Tagesgangs im Niederschlag: er wird im aLMo einige Stunden zu früh vorhergesagt und dies mit einer deutlich zu grossen Amplitude. Diese deutlichen Fehler in der Vorhersage liegen vermutlich in der Parametrisierung des konvektiven Niederschlags. Im Mittel wird Konvektion, die später zu Niederschlag führt, offensichtlich zu früh und dann sich zu schnell entwickelnd simuliert. Bei der Vorhersage von Schauern und Gewittern besonders im Sommer sollte man dieses Modellfehlverhalten berücksichtigen.

Kapitel 6

Gitterpunkt-basierte Verifikation

Um einen Überblick über die Qualität eines Wettervorhersagemodells zu bekommen, gibt es verschiedene Ansätze aus der Statistik. Standardverfahren für die Verifikation des Niederschlags sind z.B. der auch für viele andere meteorologische Parameter verwendete RMSE („Root Mean Square Error“) oder diverse kategorische Fehlermaße (engl. „Error Scores“). Diese Fehlermaße basieren alle auf dem Vergleich von Gitterpunktwerten zwischen Beobachtungen und Modell (vgl. Kap. 1).

Dieses Kapitel soll keine umfassende Verifikation mit Standardmaßen beinhalten. Es soll gezeigt werden, wie Standardverifikation funktioniert (Grundlagen dazu in Abschn. 6.1) und wo ihre Stärken und Schwächen liegen. Dazu werden in Abschn. 6.2 Tagessummen und in Abschn. 6.3 Stundensummen des Niederschlags verifiziert. Es soll untersucht werden, ob es gemäß der Standardfehlermaße regionale und saisonale Unterschiede gibt, ob Trends in der Zeitreihe 2001-2004 vorhanden sind und/oder innerannuelle Variabilitäten vorkommen.

Am Ende dieses Kapitel steht ein kurzes Fazit zur Verifikation mit Standardfehlermaßen (Abschn. 6.4).

6.1 Definition von Standardfehlermaßen

In diesem Abschnitt werden die Definitionen der in dieser Arbeit zur Standardverifikation verwendeten Fehlermaße aufgeführt und beschrieben. Aus der Vielzahl an möglichen Fehlermaßen wurden einige wenige herausgegriffen: der „Bias Score“ (auch „Frequency Bias Index“ (*FBI*)), der „Equitable Threat Score“ (*ETS*) und der „Heidke Skill Score“ (*HSS*) zum Vergleich mit Ergebnissen von Damrath et al. (2000), sowie der *RMSE*.

Die Scores berechnen sich auf Basis der Kontingenztafel (Tab. 6.1) in Abhängigkeit eines zu wählenden Schwellwertes (Woodcock, 1976; Jolliffe und Stephenson, 2003; Ebert, 2007). Die Tabelle enthält für *A* die Anzahl der Fälle, in denen der Schwellwert sowohl in den Beobachtungen, als auch im Modell überschritten wird. *B* bezeichnet Fälle, für die der Schwellwert in den Beobachtungen nicht erreicht, im Modell jedoch überschritten wird. *C* bezeichnet den zu *B* umgekehrten Fall und

	Beobachtung > Schwellwert	
Vorhersage > Schwellwert	JA	NEIN
JA	A	B
NEIN	C	D

Tabelle 6.1: Kontingenztabelle: mögliche Fälle für die Beziehung zwischen Vorhersage und Beobachtung (abhängig von Schwellwert (SW)): A = „Treffer für Niederschlag $> SW$ “, B = „falsche Alarme“, C = „verpasste Ereignisse“, D = „Treffer für Niederschlag $< SW$ “.

D den Fall, für den der Schwellwert in keinem der beiden Niederschlagsfelder überschritten wird.

Die Definitionen der Fehler-Scores nach Ebert (2007) sind wie folgt:

$$FBI = \frac{A + B}{A + C}. \quad (6.1)$$

Beim FBI wird also das Verhältnis der Häufigkeiten zwischen den den Schwellwert überschreitenden Modellvorhersagen und Beobachtungen betrachtet.

Im Gegensatz dazu sind der ETS und HSS sogenannte „Skill-Scores“, also direkte Maße dafür, ob eine Vorhersage überhaupt die Qualität für eine sinnvolle Verwendung hat (meist im Bezug zum Zufall). Dabei bezieht sich der ETS auf die Aussage über die Qualität von den den Schwellwert überschreitenden Vorhersagen:

$$ETS = \frac{A - E}{A + B + C - E} \quad (6.2)$$

mit E als den den Zufall repräsentierenden Anteil:

$$E = \frac{(A + C)(A + B)}{A + B + C + D}.$$

Der HSS ist wie folgt definiert:

$$HSS = \frac{A + D - H}{A + B + C + D - H}. \quad (6.3)$$

Er bezieht im Gegensatz zum ETS auch die Fälle mit ein, für die die Vorhersagen und Beobachtungen den Schwellwert nicht überschreiten. Dabei ist H analog zu E in Gl. 6.3 der mit einbezogene Zufallsanteil:

$$H = \frac{(A + C)(A + B) + (D + C)(D + B)}{A + B + C + D}$$

Score	Wertebereich	perfekt	Aussage
FBI	$0 \dots \infty$	$= 1$	Verhältnis von vorhergesagten zu beobachteten „JA-Ereignissen“: $< 1 =$ Unterschätzung, $> 1 =$ Überschätzung
ETS	$-\frac{1}{3} \dots 1$	$= 1$	Wie gut war die Vorhersage für „JA-Ereignisse“ relativ zum Zufall: $0 =$ keine Eignung
HSS	$-\infty \dots 1$	$= 1$	Wie gut war die Vorhersage relativ zum Zufall: $0 =$ keine Eignung

Tabelle 6.2: Übersicht über die Wertebereiche und Aussagen der verwendeten Scores.

Eine Aussage der Skill-Scores *ETS* und *HSS* ist also immer schon unter Berücksichtigung von Zufallstreffern zu werten. Die Wertebereiche und Aussagen der drei zuvor definierten Standardfehlermaße sind in Tab. 6.2 aufgeführt.

Der *RMSE* dagegen ist ein vom Schwellwert unabhängiges Fehlermaß und ist definiert als mittlere Abweichung der absoluten Differenz zwischen Vorhersage und Beobachtung (Ebert, 2007):

$$RMSE(g) = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (F_t(g) - O_t(g))^2}, \quad (6.4)$$

wobei F für Vorhersage (engl. „Forecast“), O für Beobachtung (engl. „Observation“), t für den Zeitpunkt, N_t für die Anzahl der Zeitpunkte und g für den jeweiligen Gitterpunkt steht. In den Auswertungen werden zwei verschiedene Wege der *RMSE*-Berechnung für ein Einzugsgebiet verglichen.

Für die erste Variante ($RMSE_1$) wird an jedem Gitterpunkt des Einzugsgebietes der *RMSE* gemäß Gl. 6.4 berechnet und dann gebietsgemittelt:

$$RMSE_1(\text{Gebiet}) = \frac{1}{N_g} \sum_{g=1}^{N_g} RMSE(g) \quad (6.5)$$

mit N_g der Anzahl an Gitterpunkten im betrachteten Gebiet.

Für die zweite Variante wird der *RMSE* zunächst nicht an jedem Gitterpunkt, sondern für die gebietsgemittelten Niederschläge von F und O berechnet:

$$RMSE_2(\text{Gebiet}) = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} R_{diff,t}^2}, \quad (6.6)$$

wobei $R_{diff,t}$ die gebietsgemittelte Differenz

$$R_{diff,t}(Gebiet) = \frac{1}{N_g} \sum_{g=1}^{N_g} (F_t(g) - O_t(g))$$

bezeichnet.

Die Berechnung des $RMSE_2$ ist insofern fairer, als dass ein um wenige Gitterpunkte verschobenes, aber eben vorhergesagtes Niederschlagsereignis zwar durchaus zu einem gewissen $RMSE_2$ -Wert führen kann, dieser aber nicht so hoch ausfällt wie der $RMSE_1$ (es gilt immer $RMSE_1 \geq RMSE_2$).

6.2 Auswertung der Tagessummen

In diesem Abschnitt werden die Tagessummen aus Beobachtungen und Modellvorhersagen (00 UTC-Läufe des aLMO und ECMWF) mit den oben genannten Maßen verifiziert.

Regionale Verteilung des FBI und HSS

Abbildung 6.1 zeigt den FBI (Gl. 6.1) für das aLMO (00 UTC-Läufe) in seiner regionalen Verteilung für die Schwellwerte (SW) 0.1, 1, 4 und 8 mm für den Zeitraum 2001-2004. Es fällt auf, dass der FBI höhere Über- und Unterschätzungen in der Häufigkeit an Ereignissen aufzeigt, je höher der Schwellwert gewählt wird. Für $SW = 0.1$ mm liegt der FBI zwischen 0.9 und 1.3, im Nordosten an wenigen Gitterpunkten etwas darüber. Dies bedeutet im Mittel eine leichte Überschätzung der Häufigkeit schwacher Niederschlagsereignisse. Bereits bei $SW = 1$ mm wird die Verteilung des FBI strukturierter. Im Süden und der Mitte Deutschlands sind nun Bänder mit Überschätzungen im Luv und Unterschätzungen im Lee der Mittelgebirge zu erkennen. Diese Bänder werden noch ausgeprägter für $SW = 4$ mm und $SW = 8$ mm. Damit zeigt der FBI die bereits in Abschn. 4.3 erwähnte „Luv/Lee-Problematik“ in Mittelgebirgsregionen auf. Auffällig sind noch zwei weitere Punkte für höhere Schwellwerte:

- deutliche Überschätzungen ($FBI > 1.5$) im Nordosten und im Erzgebirge und
- deutliche Unterschätzungen ($FBI < 0.7$) in der Donauregion.

Abbildung 6.2 zeigt die saisonale Verteilung des FBI über Deutschland am Beispiel $SW = 1$ mm. Hier nicht gezeigte Untersuchungen zeigten, dass dieser Schwellwert repräsentativ für andere Schwellwerte ist ($SW = 1$ mm wurde wegen der guten Vergleichbarkeit zu Stundensummen gewählt, Diskussion dazu in Abschn. 6.3). Die in Abb. 6.1 gefundenen Bänder-Strukturen finden sich hier vor allem im Winter wieder (genaue Werte für Maxima und Minima sind abhängig vom Schwellwert, hier zwischen 0.5 und 1.7). Ebenso finden sich im Winter die beiden Regionen

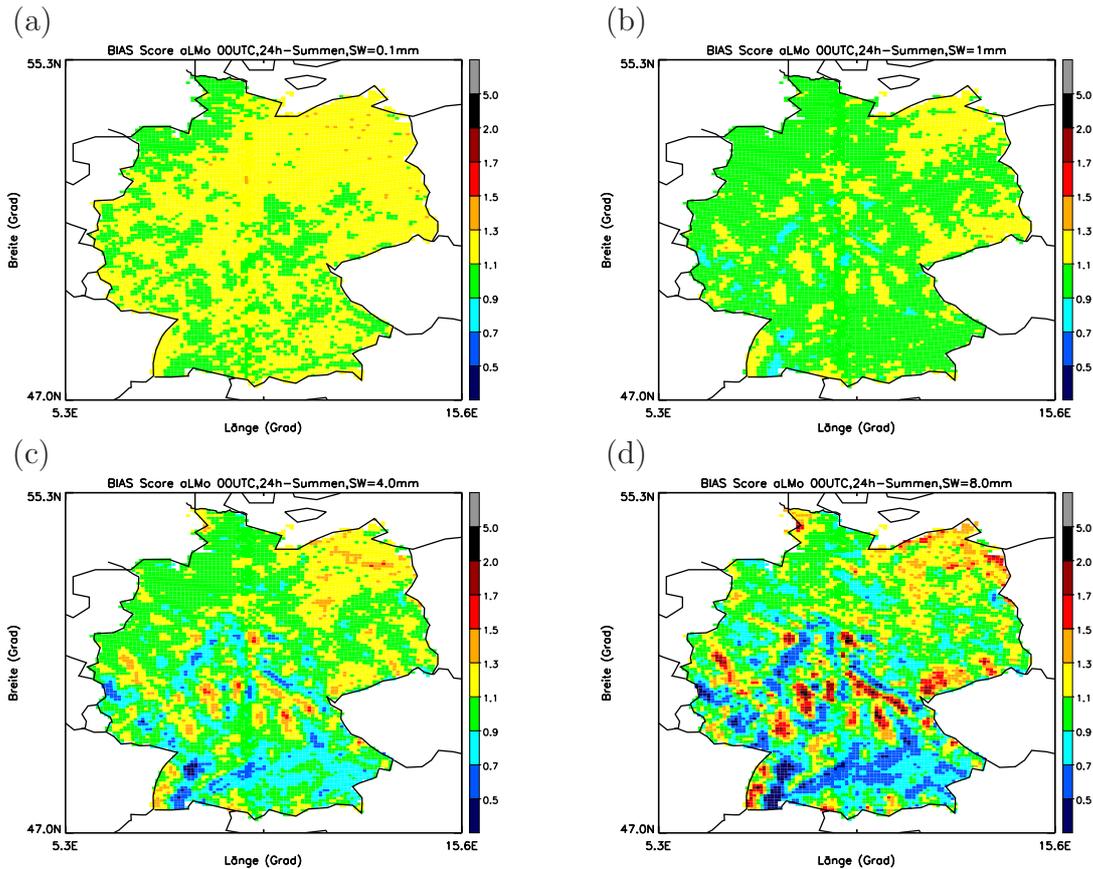


Abbildung 6.1: Bias Score (*FBI*) für verschiedene Schwellwerte (*SW*) für Tagessummen aus *aLMO*-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) von 2001-2004: (a) $SW = 0.1 \text{ mm}$, (b) $SW = 1 \text{ mm}$, (c) $SW = 4 \text{ mm}$, (d) $SW = 8 \text{ mm}$.

mit Überschätzung der Niederschlagshäufigkeiten im Nordosten und im Erzgebirge wieder. Im Frühling hat der *FBI* weniger Variabilität über Deutschland (zwischen 0.7 und 1.5, wobei die Überschätzungen überwiegen). Es gibt keine klaren Strukturen. Im Sommer dagegen fallen wieder die beiden erwähnten Regionen mit Überschätzung im Osten auf. Zudem ist das Luv/Lee-Problem im Schwarzwald gut zu erkennen. Der Herbst zeigt ein ganz anderes Bild. Fast in ganz Deutschland gibt es neutrale Werte ($FBI \approx 1$) oder sogar eine Tendenz zur Unterschätzung von Niederschlagsereignissen.

Weitere Informationen erhält man, wenn man den *HSS* für alle Jahreszeiten hinzuzieht (($SW = 1 \text{ mm}$, Abb. 6.3). Anders als beim *FBI* erkennt man hier deutliche Unterschiede zwischen Sommer und Winter, wobei der Winter insgesamt am besten abschneidet (bis auf wenige Regionen, z.B. Thüringen und Sachsen-Anhalt im Lee des Thüringer Waldes und des Harz). Die Vorhersagen im Sommer sind nach dem *HSS* am schlechtesten mit Werten fast überhalb unter 0.65, in manchen Regionen sogar deutlich unter 0.5. Im Mittel gleichen sich die Ergebnisse des *HSS* für Frühling und Herbst und liegen gerade zwischen den Maxima im Winter und Minima im Sommer. Dies ist auch ein deutlicher Unterschied zum *FBI* (Abb. 6.2), wo gerade im Herbst klare Unterschiede zu den anderen Jahreszeiten gefunden wurden. (Hinweis: der Unterschied in den Fehler-Scores zeigt, dass eine Aussage über die Qualität von

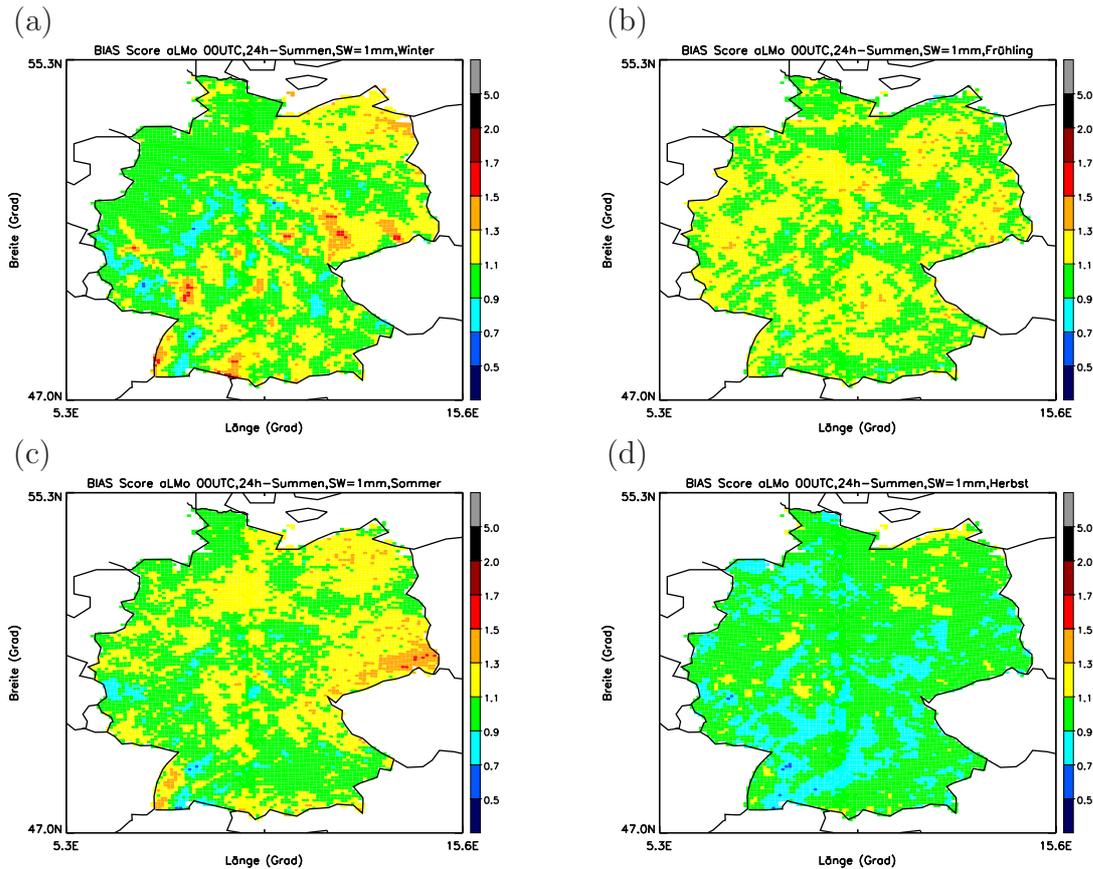


Abbildung 6.2: Bias Score (FBI) für Tagessummen aus aLMO-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) von 2001-2004 für $SW = 1$ mm (repräsentativ für andere SW): (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

Vorhersagen nicht alleine mit einem Score getroffen werden sollte. Es ist wichtig, je nach Fragestellung den oder die dafür richtigen Scores zu wählen.)

Zeitreihen für FBI, ETS und HSS

Abbildung 6.4 zeigt den *FBI*, den *ETS* und den *HSS* im Monats- und Flächenmittel über Deutschland für sechs verschiedene Schwellwerte: 0.1, 1, 2, 4, 8 und 16 mm (in Anlehnung an Abb. 4 bei Damrath et al. (2000)). Zunächst soll darauf hingewiesen werden, dass der Verlauf des *ETS* quasi identisch zu dem des *HSS* ist, wobei $HSS \geq ETS$ gilt. Dies wird deutlich, wenn man sich klar macht, dass beim *HSS* die Anzahl für D aus Tab. 6.1 mit einfließt, also richtige Vorhersagen, die somit eine positive Auswirkung auf das Ergebnis im *HSS* haben. Zum Vergleich mit Damrath et al. (2000) wurden in Abb. 6.4 sowohl der *ETS* als auch der *HSS* gezeigt, jedoch wird in der Auswertung nur auf den *HSS* eingegangen. Aus Abb. 6.4 für *FBI* und *HSS* kann man verschiedene Schlussfolgerungen ziehen:

1. je höher der Schwellwert, je schlechter die Ergebnisse und damit die Vorhersage (im Mittel),

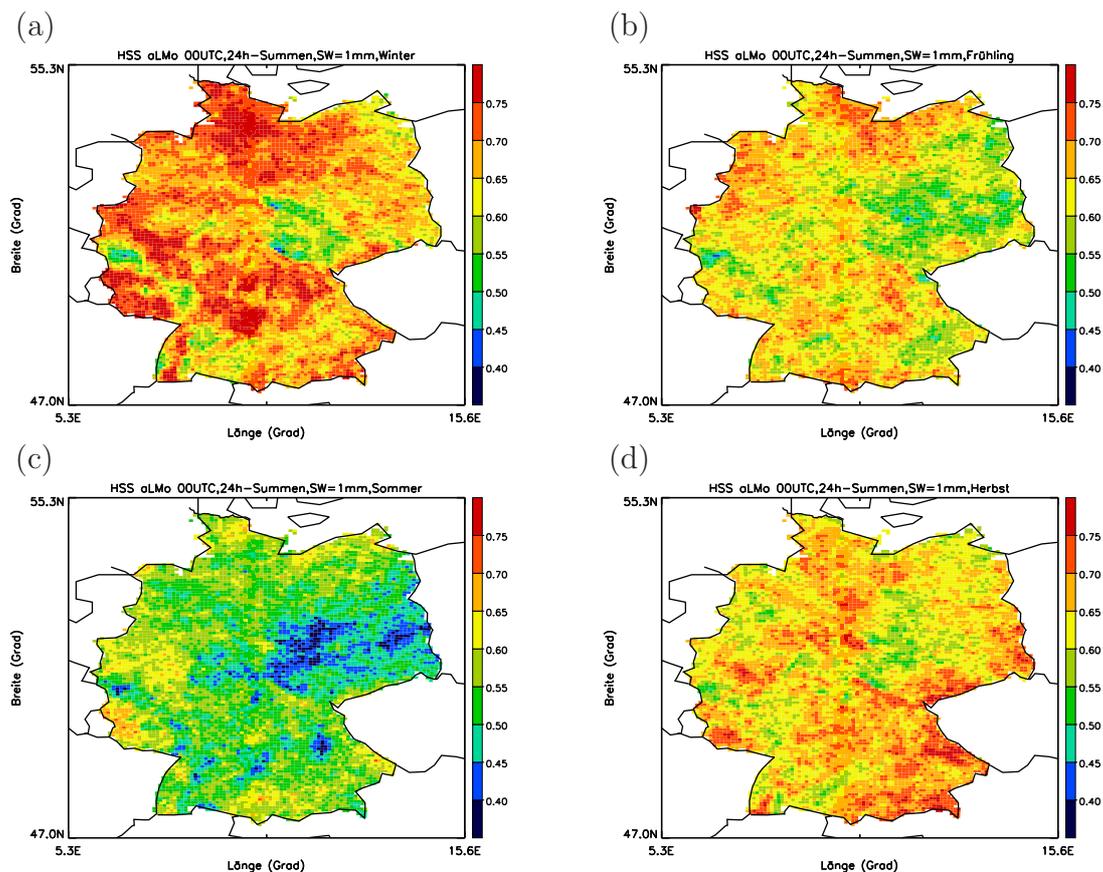


Abbildung 6.3: Wie Abb. 6.2 für HSS.

2. nahezu für die ganze Zeitreihe zeigt das aLMo eine leichte Überschätzung der Häufigkeit schwacher Niederschlagsereignisse ($1 \leq FBI < 1.4$, $SW = 0.1$ mm),
3. bei höheren Schwellwerten ist im Mittel über die Zeitreihe $FBI \simeq 1$, bei deutlich anwachsender Variabilität,
4. im Sommer zeigt der HSS für alle Schwellwerte (zumindest im Mittel) das schlechteste Ergebnis (Ausnahme 2001 für $SW \leq 4$ mm),
5. Maxima des HSS treten im Winter und Frühling auf (um 0.8 bei kleinen Schwellwerten),
6. auffällig (besonders hin zu hohen Schwellwerten) sind die tiefen Werte des HSS im Sommer 2003 (das aLMo hat in diesem sehr trockenen Sommer (Abb. 4.1) den Niederschlag nicht gut vorhergesagt, dies könnte auf ein Problem der Scores bei seltenen Ereignissen hinweisen).

Vorsichtig sollte man bei der Interpretation insbesondere bei höheren Schwellwerten sein, da dort die Anzahl an Ereignissen gering ist und somit keine quantitative Aussage gemacht werden kann. Eine Information über den in Abb. 6.3 angedeuteten jahreszeitlichen Verlauf des HSS bekommt man, wenn man aus den Monatswerten der vier Jahre einen mittleren Jahresgang erstellt. Abbildung 6.5 zeigt diesen Verlauf am Beispiel des HSS für $SW = 1$ mm. Die einzelnen Monatsmittel sind mit farbigen

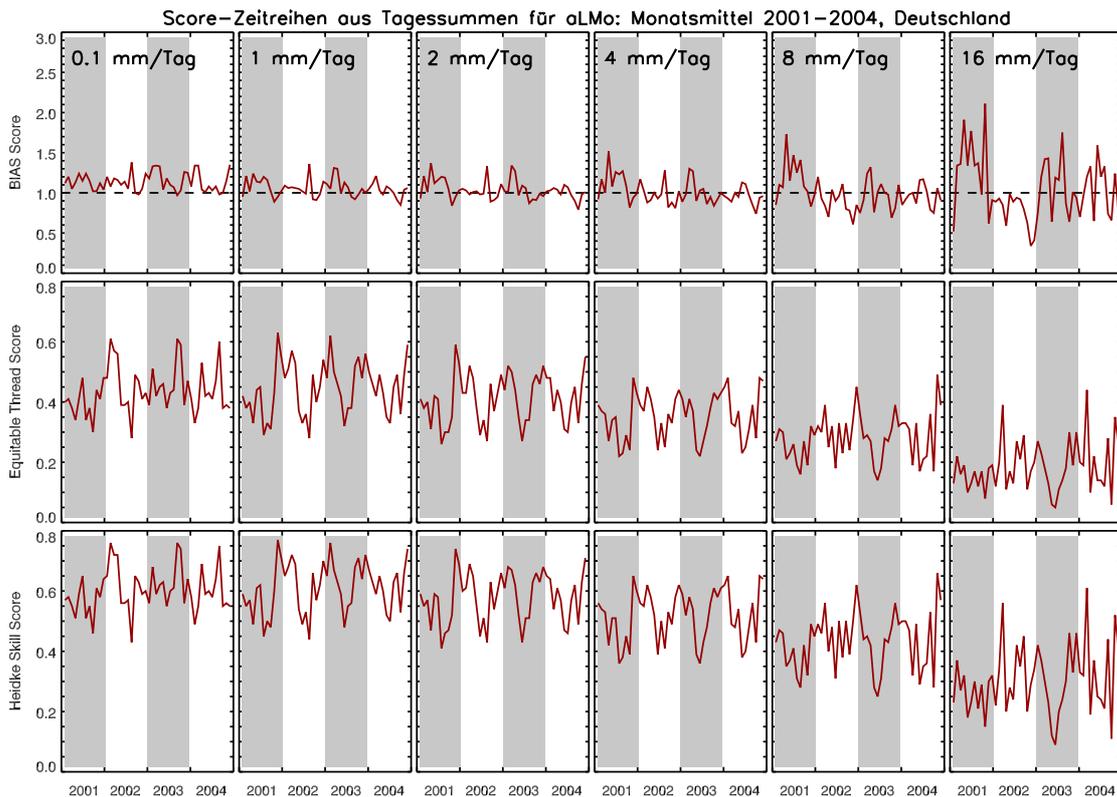


Abbildung 6.4: Zeitreihen der Monatsmittel des Bias Score (FBI, oben), Equitable Thread Score (ETS, Mitte) und Heidke Skill Score (HSS, unten) für Tagessummen aus aLMO-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) von 2001-2004. Schwellwerte von links nach rechts: 0.1, 1, 2, 4, 8 und 16 mm.

Kreuzen gekennzeichnet: (Rot) 2001, (Grün) 2002, (Blau) 2003 und (Gelb) 2004. Es ist ein klarer Jahresgang mit dem Minimum bei 0.47 im Juli und dem Maximum bei 0.68 im Dezember erkennbar. Die Schwankungen zwischen den Monatswerten der einzelnen Jahre sind relativ groß (bis zu 0.19 im April). Die Schwankung im Jahresgang im Herbst erklärt sich durch die kleine Statistik von gerade vier Werten pro Monat.

Vergleicht man nun Abb. 6.4 mit der analogen Abb. 4 aus Damrath et al. (2000), so muss zunächst darauf hingewiesen werden, dass dort das Europamodell (EM) und das Deutschmodell (DM) für die Zeit von Januar 1997 bis Juni 1999 verifiziert wurden. Die horizontale Auflösung des EM betrug 55 km, die des DM 14 km. Ein weiterer wichtiger Punkt ist, dass Damrath et al. (2000) im Gegensatz zur vorliegenden Arbeit die beiden Modelle direkt mit allen verfügbaren synoptischen Stationen, aber ohne dazwischen geschaltete Gitterung verifizierten. Die Modellvorhersagen wurden zum Vergleich an die Stationsorte interpoliert.

Damrath et al. (2000) finden ebenfalls leichte Überschätzungen der Häufigkeiten von Niederschlagsereignissen, für höhere Schwellwerte eine Überschätzung im Winter und eine Unterschätzung im Sommer. Ein Vergleich der beiden Abbildungen zeigt insgesamt, dass der Verlauf der drei Scores sehr ähnlich ist (für alle drei Modelle). Dies würde bedeuten, dass die Einführung des aLMO mit 7 km Auflösung keine weitere Verbesserung in der Niederschlagsvorhersage gebracht hat (das DM brachte

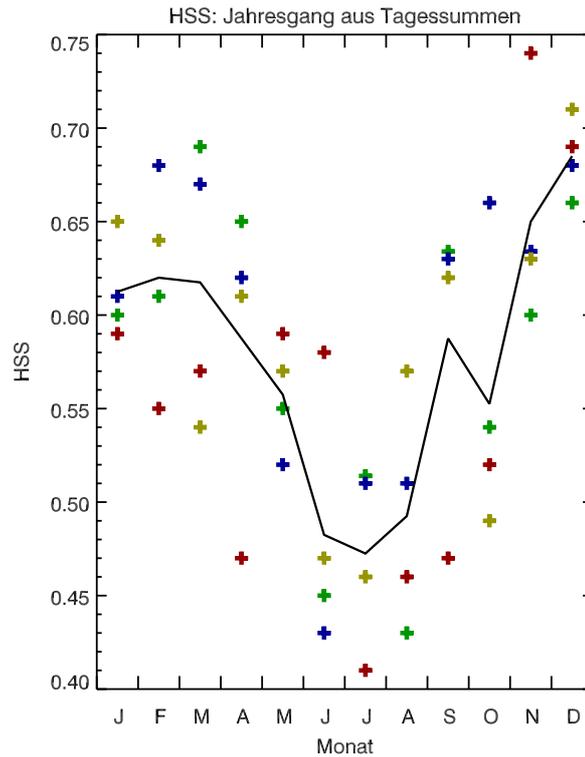


Abbildung 6.5: Mittlerer Jahresgang von 2001-2004 (schwarze Linie) des HSS ($SW = 1$ mm) für aLMO-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) für Deutschland. Die farbigen Kreuze zeigen die einzelnen Monatswerte der vier Jahre an: (Rot) 2001, (Grün) 2002, (Blau) 2003 und (Gelb) 2004.

gegenüber dem EM auch nur eine Verbesserung für $SW > 4$ mm). Mit gitterpunkt-basierten Fehlermaßen ist also keine Qualitätssteigerung zwischen Niederschlagsvorhersagen dieser verschieden hoch aufgelösten Modelle erkennbar.

Zeitreihen für verschiedene Flusseinzugsgebiete

Ein direkter Vergleich der Niederschlagsmengen (unabhängig von Schwellwerten) an jedem Gitterpunkt ist mit dem $RMSE$ möglich. Abbildung 6.6 zeigt Zeitreihen für Monatsmittel zweier verschiedener Varianten der $RMSE$ -Berechnung (oben: $RMSE_1$ (Rot) und $RMSE_2$ (Blau)) und des HSS (unten: $SW = 1$ mm) für verschiedene Flusseinzugsgebiete: Rhein (links), Elbe (Mitte) und Donau (rechts). Die Einzugsgebiete umfassen immer nur den deutschen Teil dieser Gebiete (Abschn. 2.7), da außerhalb keine Niederschlagsmessungen für diese Arbeit verfügbar waren. Folgende Aussagen kann man nach Abb. 6.6 (oben) für den $RMSE_1$ treffen:

- es gibt einen klaren Jahresgang im $RMSE_1$ mit Maxima im Sommer und Minima im Winter (lokaler konvektiver gegenüber flächigem Niederschlag),
- der besonders feuchte Sommer 2002 hat für jedes Gebiet den höchsten $RMSE_1$ der Zeitreihe (viel Niederschlag bedeutet auch mehr Möglichkeit für größere Abweichungen),

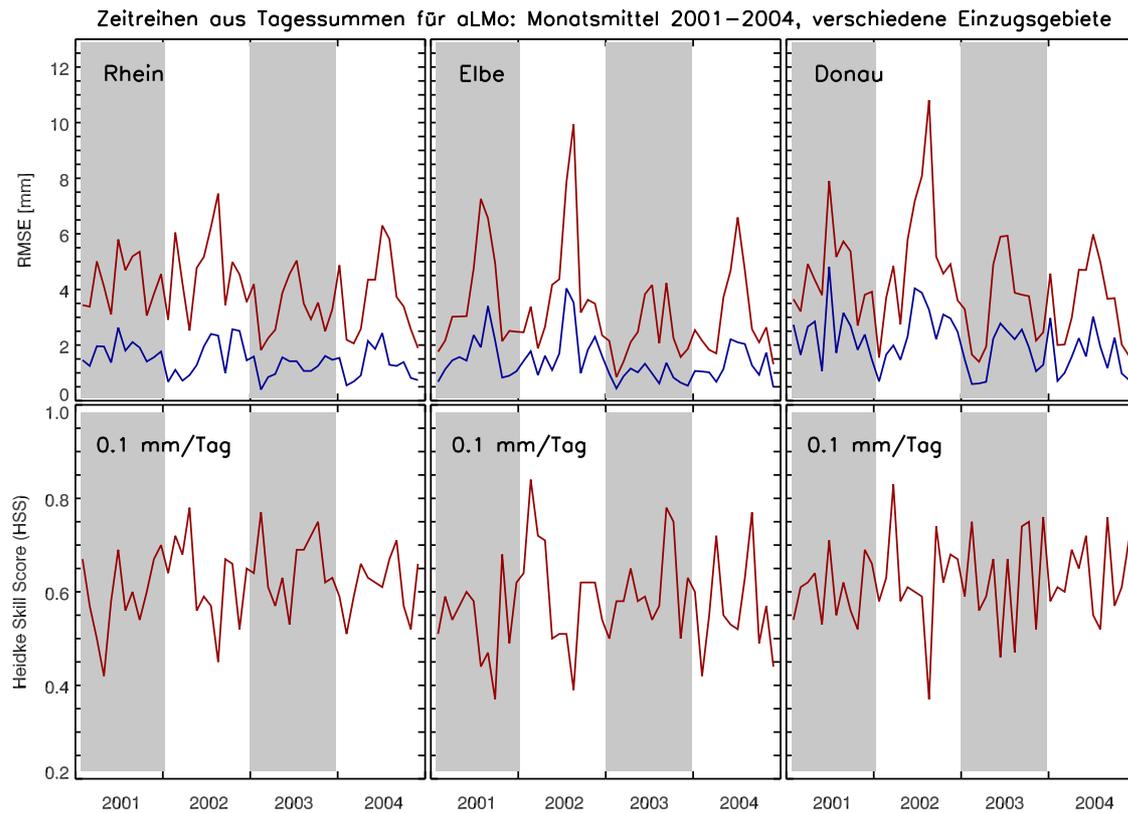


Abbildung 6.6: Zeitreihen der Monatsmittel für Tagessummen (2001–2004) des $RMSE_1$ (oben, Rot), $RMSE_2$ (oben, Blau) und HSS (unten, $SW = 0.1$ mm) aus aLMO-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) für drei Flusseinzugsgebiete: Rhein (links), Elbe (Mitte) und Donau (rechts).

- im Mittel liegt der $RMSE_1$ für alle Gebiete etwa gleich um 4 mm (bei unterschiedlich großen Schwankungen), regionale Unterschiede sind also relativ gering.

Vergleicht man den $RMSE_1$ (Gl. 6.5) mit dem ebenfalls dargestellten $RMSE_2$ (Gl. 6.6), so sind die Monatsmittel des $RMSE_2$ deutlich kleiner (im Mittel um 2 mm). Wie für den $RMSE_1$ gibt es auch für den $RMSE_2$ einen Jahresgang und im Mittel nur geringe Unterschiede zwischen einzelnen Flusseinzugsgebieten. Es zeigt sich, dass die Niederschlagsvorhersage des aLMO deutlich bessere Qualität hat, wenn man Niederschlagsmittel in einem Einzugsgebiet betrachtet. So kann verhindert werden, dass ein leicht verschoben vorhergesagtes Ereignis als Fehlvorhersage interpretiert wird. Eine nur wegen einer Verschiebung um wenige Gitterpunkte (und damit wenige Kilometer) fehlerhafte Vorhersage wird beim $RMSE_2$ nicht als schlecht gewertet. Dies kann insbesondere bei einer nicht kontinuierlich verteilten Variablen wie dem Niederschlag sinnvoll sein (und besonders bei konvektiven, also kleinskali- gen Ereignissen).

Wie für beide $RMSE$ -Versionen zeigt sich auch im HSS ein Jahresgang, allerdings mit deutlich mehr Variabilität zwischen den Monaten (vgl. mit Abb. 6.5). Der HSS zeigt im Mittel für alle drei Gebiete sehr ähnliche Werte um 0.6, welche auch schon in Abb. 6.4 im Deutschlandmittel gefunden wurden.

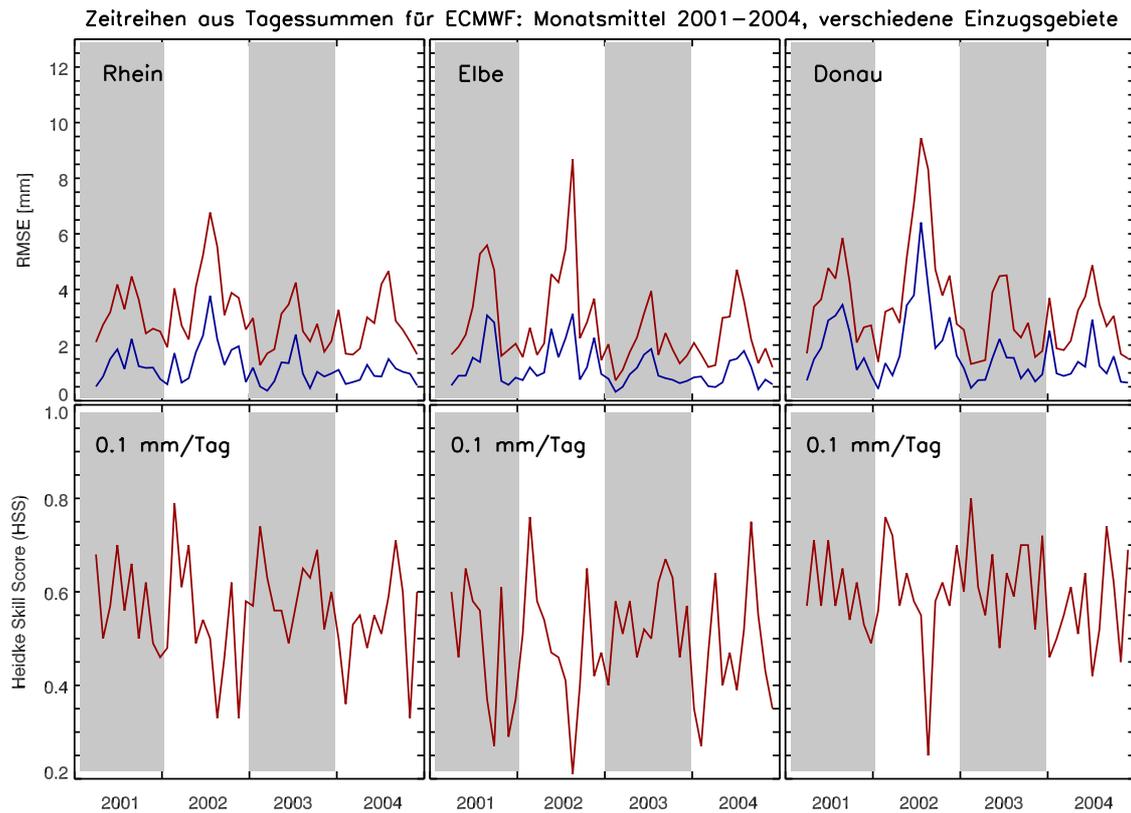


Abbildung 6.7: Wie Abb. 6.6 für ECMWF-Vorhersagen.

Zuletzt soll noch ein Vergleich mit der $RMSE$ - und HSS -Zeitreihen für das weniger hoch aufgelöste Globalmodell des ECMWF gemacht werden (Abb. 6.7). In den ersten drei Monaten sind hier keine Daten vorhanden. Für den $RMSE_1$ fällt auf, dass der Verlauf der Zeitreihen qualitativ sehr ähnlich dem des aLMO ist. Jedoch sind die Werte für das Globalmodell im Mittel über die Zeitreihe geringer (um 3 mm). Der $RMSE_2$ zeigt analog zu den aLMO-Ergebnissen niedrigere Werte als der $RMSE_1$ (Erklärung wie für aLMO). Der HSS für das Globalmodell zeigt dagegen etwas schlechtere Werte als im aLMO (der Verlauf ist aber sehr ähnlich). Vor allem für die Gebiete Rhein und Elbe liegt der Wert im Mittel um 0.1 unterhalb dem Wert im aLMO. Auch die Variabilität zwischen den Monaten ist etwas ausgeprägter. Es zeigt sich also für Tagessummen, dass gitterpunkt-basierte Verifikationsmaße kaum Verbesserungen in der Niederschlagsvorhersage bei Modellen mit höherer horizontaler Auflösung anzeigen. Für den $RMSE_1$ wird die Vorhersage des aLMO gegenüber des ECMWF-Modells sogar als schlechter bewertet (im Monatsmittel), dagegen im HSS als leicht verbessert. Sieht man allerdings von der Verifikation an einzelnen Gitterpunkten ab und vergleicht gebietsgemittelte Niederschläge ($RMSE_2$), so ist eine deutliche Verbesserung der $RMSE$ -Werte zu erkennen. Bei dieser Betrachtung liefert das aLMO im Vergleich zum ECMWF-Modell (Abb. 6.6 und Abb. 6.7) nun vergleichbar gute Niederschlagsvorhersagen.

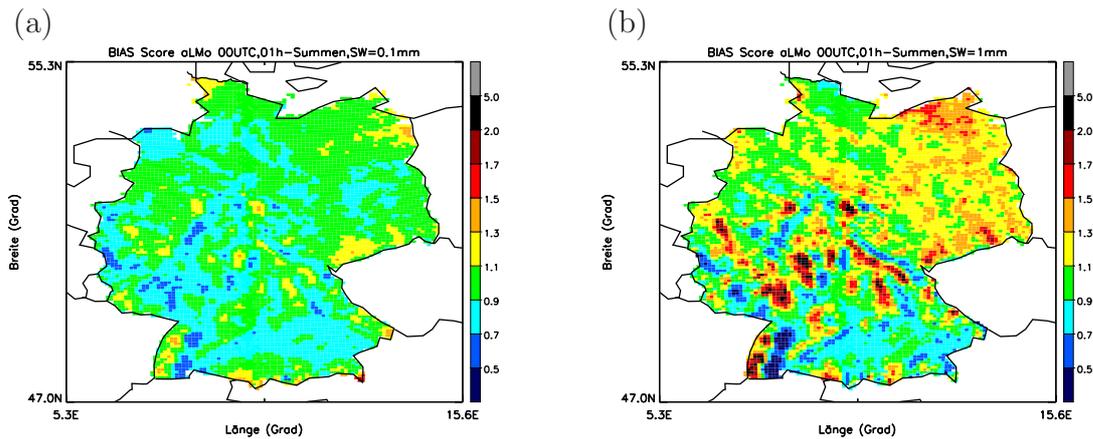


Abbildung 6.8: Bias Score (*FBI*) für zwei verschiedene Schwellwerte (*SW*) für Stunden­summen aus *aLMo*-Vorhersagen (00 UTC-Läufe, verglichen mit Beobach­tungen) von 2001-2004: (a) $SW = 0.1$ mm, (b) $SW = 1$ mm

6.3 Auswertung der Stunden­summen

Analog zu Abschn. 6.2 für Tagessummen werden in diesem Abschnitt Stunden­summen mit den ausgewählten Standardmaßen (Abschn. 6.1) verifiziert. Dazu wird der disaggregierte Datensatz (Kap. 3) herangezogen. Erwartet wird, dass die Fehler­maße für Stunden­summen schlechter als für Tagessummen abschneiden, da die zeitlich hohe Auflösung exakte Vorhersagen erschwert (bei Tagessummen kommt es nicht so sehr darauf an, ob ein Niederschlagsereignis wenige Stunden zu früh oder zu spät auftritt). Dabei gilt dies vor allem für den *FBI* und den *HSS* (der *ETS* wird nicht weiter betrachtet), weniger für den *RMSE*, da dieser aufgrund der kürzeren Zeitspanne für ein Niederschlagsereignis nicht mehr so hohe Werte annehmen kann (zumindest im Mittel).

Regionale Verteilung des *FBI* und *HSS*

Abbildung 6.8 zeigt den *FBI* für Stunden­summen und zwei verschiedene Schwell­werte (0.1 und 1 mm). Für Schwell­werte über 2 mm treten sehr hohe Minima und Maxima auf. Die Bilder sind dann nicht mehr sehr aussagekräftig, bzw. nicht mehr statistisch signifikant. Dies ist auch verständlich, da auf stündlicher Zeitskala Ereignisse mit relativ hohen Schwell­werten selten sind und diese öfter verpasst werden, bzw. zum falschen Zeitpunkt auftreten.

Für $SW = 0.1$ mm in den Stunden­summen (Abb. 6.8 (a)) findet man (im Gegen­satz zu Tagessummen) im Mittel eine Unterschätzung der Häufigkeiten an Nieder­schlagsereignissen (bis $FBI = 0.5$). Es wird auch die mehrfach erwähnte Luv/Lee-Problematik sichtbar, die für diesen kleinen Schwell­wert in den Tagessummen noch nicht sichtbar war. Diese Problematik in den Mittelgebirgsregionen verstärkt sich hin zu höheren Schwell­werten (Abb. 6.8 (b)). In Teilen Ostdeutschlands werden für $SW = 1$ mm die hohen Überschätzungen ($FBI > 1.5$) wie bei den Tagessummen wiedergefunden. Auffällig ist auch die schon in den Tagessummen identifizierte, mit

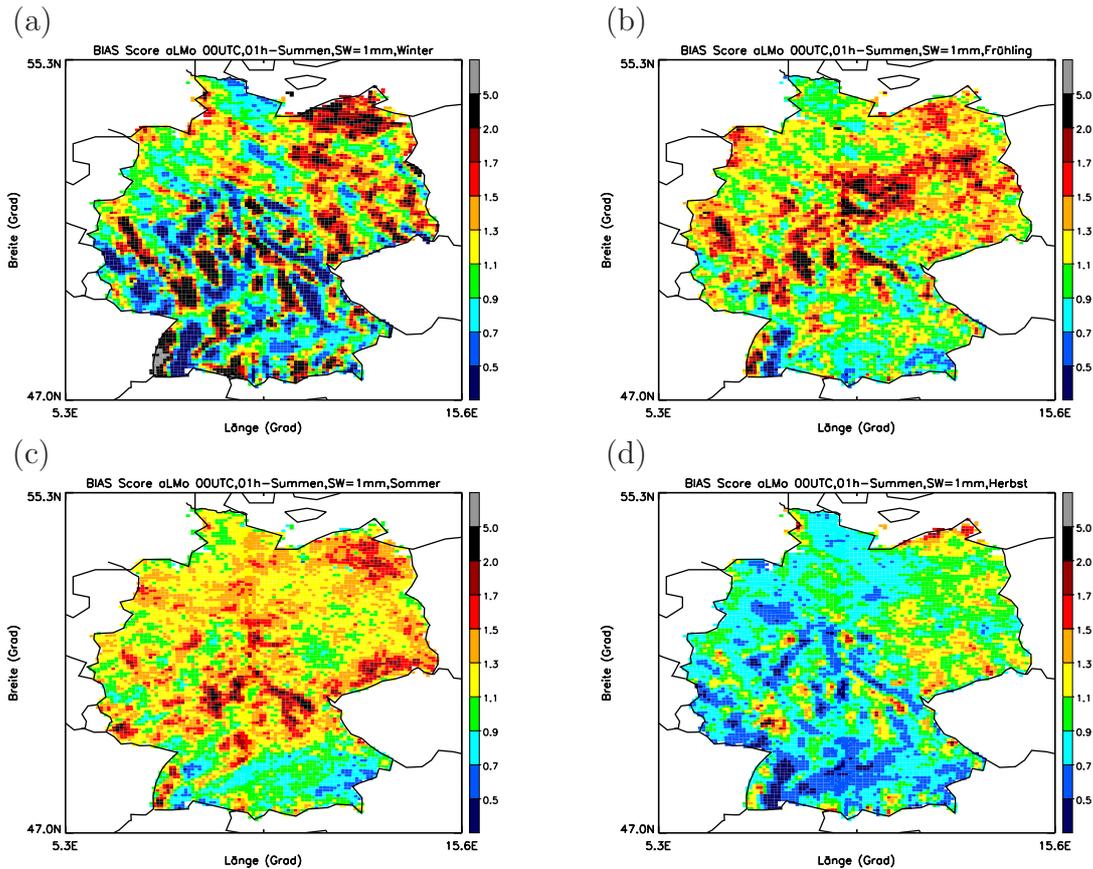


Abbildung 6.9: Bias Score (*FBI*) für Stundensummen aus *aLMO*-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) von 2001-2004 für $SW = 1$ mm (repräsentativ für andere SW): (a) Winter, (b) Frühling, (c) Sommer, (d) Herbst.

einer Unterschätzung von Niederschlagsereignissen > 1 mm charakterisierte Donau-region (für Stundensummen $FBI > 0.5$).

Vergleicht man weiter die in Abb. 6.9 gezeigten *FBI*-Werte für Stundensummen für einzelne Jahreszeiten mit den Tagessummen aus Abb. 6.2 ($SW = 1$ mm), so findet man in allen Jahreszeiten ähnliche Strukturen in der regionalen Verteilung (mit extremeren Werten bei Stundensummen). Erneut wird die Luv/Lee-Problematik sehr deutlich erkennbar (vor allem im Winter), zum anderen die deutlichen Überschätzungen nördlich des Mains im Frühling, noch etwas mehr nach Süden ausgedehnt im Sommer. Dagegen ist auch in den Stundensummen (mit Ausnahme der erwähnten Regionen in Ostdeutschland) eine meist klare Unterschätzung von Niederschlagsereignissen im Herbst sichtbar (Minima $FBI < 0.5$). Gerade im Frühling und Sommer stimmen die Ergebnisse (im Mittel deutliche Überschätzung) gut mit den Ergebnissen aus Abschn. 5.3 für den Tagesgang überein. Dort wurden deutlich zu hohe Häufigkeiten an Niederschlagsereignissen im *aLMO* gefunden, was auf mögliche Fehler der Parametrisierung von Konvektion schließen lässt.

Auch für den *HSS* finden sich schlechtere Ergebnisse für die Stundensummen bei gleichem Schwellert (1 mm, Abb. 6.10). Waren die Werte in den Tagessummen im Sommer am niedrigsten um 0.4, so liegen in dieser Größenordnung die besten *HSS*-Werte für Stundensummen (Herbst im Alpenvorland um 0.45). Vor allem im Win-

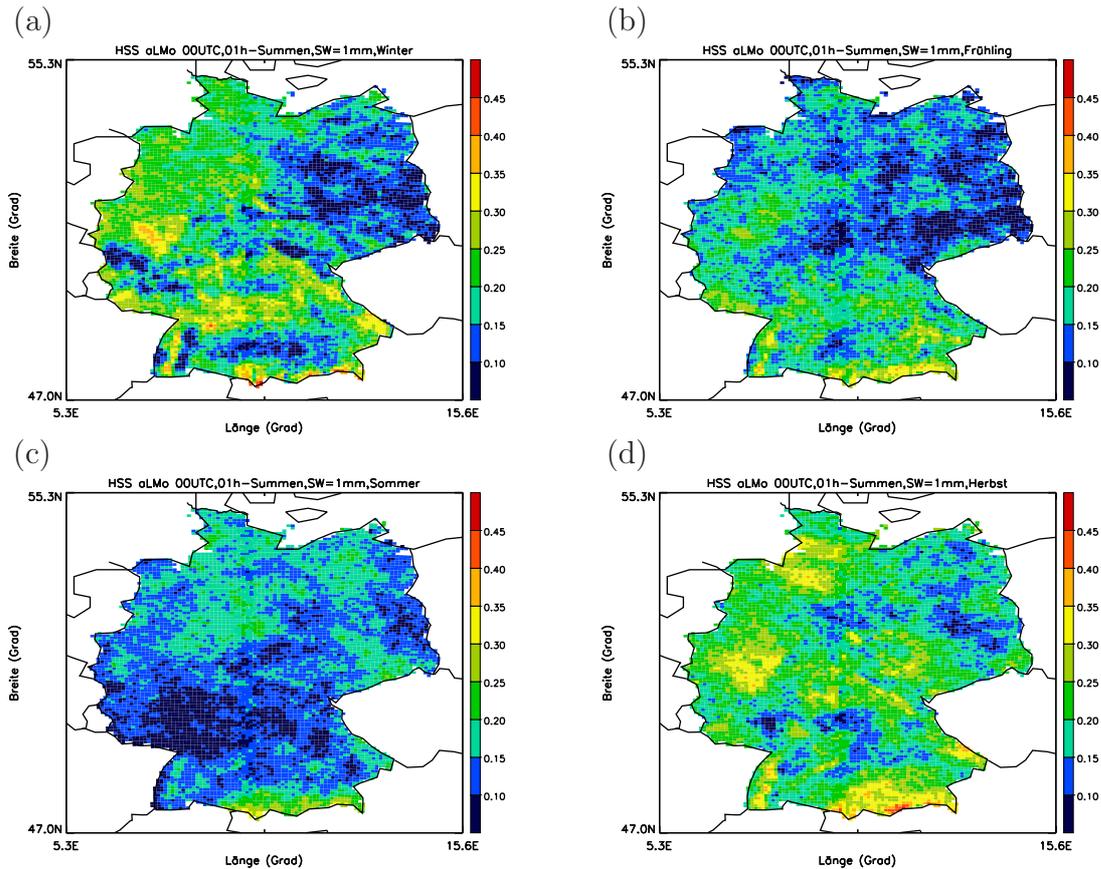


Abbildung 6.10: Wie Abb. 6.9 für HSS.

ter, Frühling und Herbst finden sich für Stundensummen ähnliche Strukturen in der regionalen Verteilung des *HSS* wie für Tagessummen. Auffällig ist jedoch, dass nicht mehr der Winter im Mittel das beste Ergebnis zeigt, sondern der Herbst. Am schlechtesten ist wiederum der Sommer, wobei in weiten Teilen Südwestdeutschlands Werte < 0.1 vorkommen (im Unterschied zu Tagessummen, wo das Minimum weiter im Osten lag).

Die Unterschiede (vor allem im Sommer und Herbst) zwischen Stunden- und Tagessummen basieren folglich auf der zeitlichen Verschiebung von wenigen Stunden zwischen vorhergesagten und beobachteten Ereignissen.

Zeitreihen für FBI und HSS

Abbildung 6.11 zeigt aus Stundensummen monatsgemittelte Zeitreihen für *FBI* und *HSS* (vgl. Abschn. 6.2 für Tagessummen, *ETS* wird hier nicht weiter diskutiert). Die Schwellwerte wurden im Vergleich zu den Tagessummen etwas kleiner gewählt: 0.1, 1 und 2 mm. Die gezeigte Skala wurde aber jeweils gleich gehalten.

Wie für Tagessummen findet sich auch hier ein Jahresgang (am deutlichsten im *HSS* sichtbar), der für alle drei Scores im Sommer die schlechtesten, im Winter die besten Werte anzeigt. Im Mittel erkennt man im Gegensatz zu Tagessummen für $SW = 0.1$ mm eine leichte Unterschätzung. Ausschlaggebend sind hier wohl vor allem die schwachen Niederschläge im Winter und Herbst (vgl. mit Häufigkeitskarten

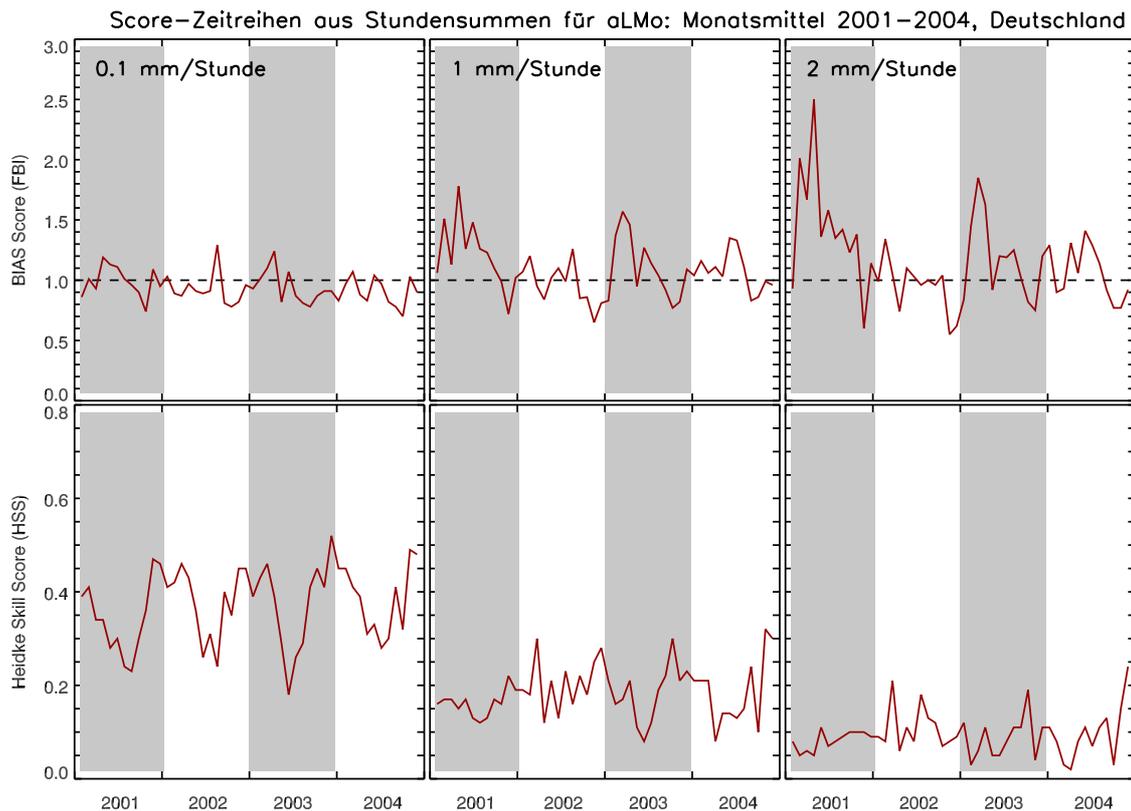


Abbildung 6.11: Zeitreihen der Monatsmittel des Bias Score (*FBI*, oben) und Heidke Skill Score (*HSS*, unten) für Stundensummen aus aLMO-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) von 2001-2004. Schwellwerte von links nach rechts: 0.1, 1 und 2 mm.

in Abb. 5.1 und Intensitätskurven in Abschn. 5.2). Hin zu größeren Niederschlägen (höher gewählten Schwellwerten) wird dann die Häufigkeit der Niederschlagsereignisse im Mittel leicht überschätzt (vgl. Intensitätskurven in Abschn. 5.2). Mit der Erhöhung des Schwellwertes steigt auch die Variabilität des *FBI* und vor allem die Maxima nehmen immer höhere Werte an.

Beim *HSS* liegen die Werte deutlich unterhalb denen für Tagessummen und gehen für $SW > 2$ mm gegen null. Damit wird deutlich, dass die Vorhersagequalität im Niederschlag auf kurzen Zeitskalen (stündlich im Vergleich zu täglich) stark abnimmt.

Zeitreihen für verschiedene Flusseinzugsgebiete

Zuletzt sollen im Rahmen der gitterpunkt-basierten Verifikation noch der *RMSE* und der *HSS* für Stundensummen mit den Werten für Tagessummen für verschiedene Flusseinzugsgebiete verglichen werden. Dazu wurden analog zu Abb. 6.6 Zeitreihen für Monatsmittel für zwei *RMSE*-Berechnungen und den *HSS* ($SW = 0.1$ mm) für den Rhein, die Elbe und die Donau erstellt (Abb. 6.12). Dabei ist zu beachten, dass für den *RMSE* für Stundensummen die Skala deutlich reduziert wurde (von 0 bis 1.6 mm, Tagessummen war bis 13 mm), da die absoluten Differenzen zwischen Vorhersage und Beobachtungen deutlich geringer sind. Auch für den *HSS* wurde die Skala verändert, allerdings nur um 0.2 nach unten verschoben.

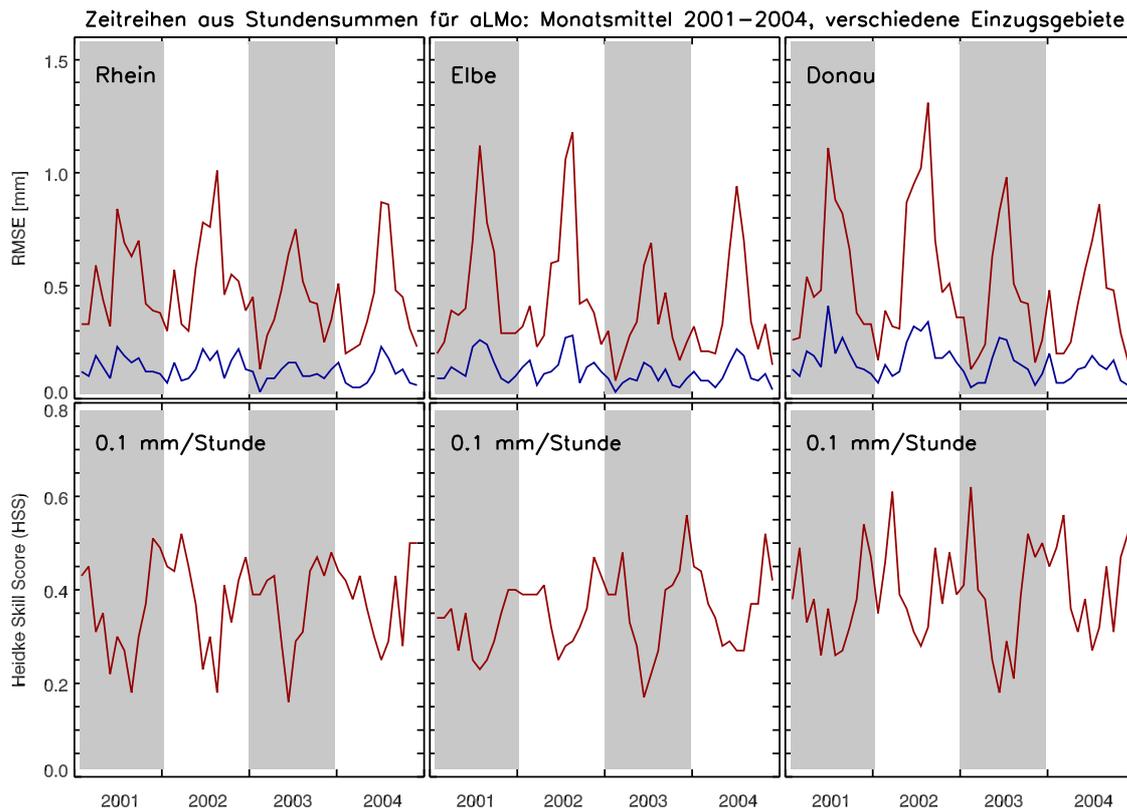


Abbildung 6.12: Zeitreihen der Monatsmittel für Stundensummen (2001-2004) des $RMSE_1$ (oben, Rot), $RMSE_2$ (oben, Blau) und HSS ($SW = 0.1$ mm, unten) aus aLMo-Vorhersagen (00 UTC-Läufe, verglichen mit Beobachtungen) für drei Flusseinzugsgebiete: Rhein (links), Elbe (Mitte) und Donau (rechts).

Es können die gleichen Aussagen über Verlauf und Zusammenhänge wie für Tagessummen getroffen werden, da sich beide Abbildungen für alle Gebiete sehr gleichen (die Absolutwerte liegen allerdings in allen drei Kurven unterhalb der Werte für Tagessummen).

6.4 Zusammenfassung und Fazit

In diesem Kapitel wurde gezeigt, wie herkömmliche Verifikationsmethoden für Niederschlagsvorhersagen angewendet werden und welche Ergebnisse ausgewählte Fehlermaße liefern. Weiter wurde auf Schwächen vor allem auch in der Methodik der gitterpunkt-basierten Verifikation eingegangen. Dabei wurde festgestellt, dass

- Fehler-Scores keine oder nur sehr kleine Verbesserungen in der Niederschlagsvorhersage beim hoch-aufgelösten aLMo im Vergleich zum Globalmodell des ECMWF zeigen (Tagessummen),
- der $RMSE$ für die Verifikation auf Gitterpunktbasis eher eine Verschlechterung der Vorhersagequalität mit dem aLMo im Vergleich zum Globalmodell liefert (Tagessummen),

- die Vorhersagequalität auf kürzeren Zeitskalen (Stunde gegenüber Tag) deutlich abnimmt.

Wie in der Einleitung (Kap. 1) beschrieben stellt sich die Frage, ob es überhaupt noch sinnvoll ist, Standardmaße für hoch-aufgelöste Modelle zu verwenden.

Einen ersten Hinweis darauf, dass speziell der Niederschlag nicht gitterpunktweise verifiziert werden sollte, um die Qualität von Vorhersagen fair zu behandeln, liefert der $RMSE_2$ (Gl. 6.6). Jedoch zeigt auch diese $RMSE$ -Berechnung keinen deutlichen Unterschied zwischen verschiedenen hoch aufgelösten Modellen (hier: aLMo und Globalmodell des ECMWF). Weiter können mit diesen Standardfehlermaßen nicht die genauen Charakteristiken der Modelldefizite aufgezeigt werden, ob z.B. der Fehler in der Amplitude oder nur einer Verschiebung des Niederschlagsfeldes liegt.

Es soll weiter darauf hingewiesen werden, dass Standardverifikation für den Niederschlag verwendet werden kann, wenn Vorhersagen gleicher räumlicher und zeitlicher Auflösung qualitativ verglichen werden sollen. Wenn zur statistisch signifikanten Auswertung solcher Standardmaße genügend Ereignisse vorhanden sind und sehr gute Werte errechnet werden, so darf die Vorhersage auch als sehr gut gewertet werden. Sind die Werte jedoch nicht besonders gut, heißt dies im Umkehrschluss nicht zwingend, dass die Vorhersage als schlecht interpretiert werden muss.

Kapitel 7

SAL: Entwicklung eines neuen Qualitätsmaßes für die Niederschlagsvorhersage

Wie in der Einleitung (Kap. 1) ausführlich beschrieben und anhand der Verifikation mit Standardmaßen (Kap. 6) bestätigt wurde, reichen herkömmliche Methoden nicht aus, um eine faire Aussage über die Qualität von Niederschlagsvorhersagen von hoch-aufgelösten Modellen zu machen. Deshalb war das Hauptziel dieser Arbeit, ein neues Qualitätsmaß für die Niederschlagsvorhersage zu entwickeln. Dieses neuartige Qualitätsmaß wird hier vorgestellt und heißt „SAL“. SAL soll eine Vorhersage in einem vorgegebenen Gebiet auf Strukturfehler (S), Amplitudenfehler (A) und Verschiebungsfehler (L engl.: Location) untersuchen.

Im folgenden theoretischen Teil wird die technische Entwicklung von SAL detailliert beschrieben. Dabei wird in Abschn. 7.1 die mathematische Definition von SAL eingeführt, in Abschn. 7.2 anhand idealisierter Niederschlagsfälle geprüft. In Abschn. 7.3 werden reale Beispiele der Niederschlagsvorhersage mit SAL verifiziert. Abschließend folgt ein kurzes Fazit zur Entwicklung und möglichen Verwendung von SAL (Abschn. 7.4).

7.1 Definition von SAL

SAL bezieht sich immer auf die Verifikation der Niederschlagsvorhersage in einem vorher zu wählenden Gebiet \mathcal{D} (z.B. ein Flusseinzugsgebiet) mit N_g Gitterpunkten in den Beobachtungen und im Modell. R bezeichnet ein Niederschlagsfeld. Zur Unterscheidung der Niederschlagsfelder in Beobachtungen und Modell werden R_{obs} und R_{mod} verwendet.

Die A-Komponente

A beschreibt die skalierte Differenz der gebietsgemittelten Niederschläge aus Modell und Beobachtungen:

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0.5[D(R_{mod}) + D(R_{obs})]}, \quad (7.1)$$

mit dem Gebietsmittel $D(R)$, wobei $D = \frac{1}{N_g} \sum_{(i,j) \in \mathcal{D}} R_{ij}$ und R_{ij} der Niederschlagswert an jedem Gitterpunkt ist.

A ist ein einfaches Maß für die Qualität einer Vorhersage im Hinblick auf die Niederschlagsmenge in einem Gebiet, ohne dabei kleinskalige Strukturen zu beachten. A ist im Bereich $[-2, \dots, +2]$ definiert, wobei $A = 0$ eine perfekte Vorhersage des Gebietsmittels bedeutet. $A = +1$ bedeutet dagegen, dass das Modell den Niederschlag im Gebietsmittel mit einem Faktor 3 überschätzt. $A = -1$ weist analog dazu auf eine Unterschätzung um den Faktor 3 hin. Überschätzungen um den Faktor 1.5 bzw. 2 führen zu $A = 0.4$ bzw. $A = 0.67$.

Die L-Komponente

Die L -Komponente besteht aus zwei Teilen: $L = L_1 + L_2$. Der erste Teil L_1 gibt die skalierte Entfernung zwischen den Schwerpunkten der Niederschlagsfelder aus Beobachtungen und Modell in Kilometer an:

$$L_1 = \frac{|\mathbf{x}(R_{mod}) - \mathbf{x}(R_{obs})|}{d}, \quad (7.2)$$

mit d als der größtmöglichen Entfernung zwischen zwei Punkten der Gebietsgrenze. Der Vektor $\mathbf{x}(R)$ gibt die Schwerpunktkoordinaten des Niederschlagsfeldes in \mathcal{D} an. Nach Gl. 7.2 liegen die Werte für L_1 im Bereich $[0, \dots, 1]$. L_1 gibt damit einen Einblick über die Qualität der Vorhersage in der Niederschlagsverteilung innerhalb \mathcal{D} . Für $L_1 = 0$ sind die Schwerpunktkoordinaten des vorhergesagten und beobachteten Niederschlagsfeldes identisch. Jedoch können viele verschiedene Niederschlagsverteilungen in einem Gebiet den gleichen Schwerpunkt haben, womit $L_1 = 0$ nicht notwendigerweise eine perfekte Vorhersage bedeutet. Ein Beispiel dafür sind Niederschläge an zwei gegenüber liegenden Seiten des Gebiets als beobachtetes Feld und eine Modellvorhersage, die genau dazwischen Niederschlag zeigt. Beide Niederschlagsfelder können den gleichen Schwerpunkt haben, obwohl die Niederschlagsverteilung sehr unterschiedlich ist (Diskussion dazu weiter unten mit idealisierten Beispielen).

Der zweite Teil L_2 wurde nun genau zur Verhinderung dieser nur vermeintlich perfekten Vorhersage für die L -Komponente eingeführt. Dabei benötigt man für L_2 und

später für die Struktur-Komponente S die Identifikation einzelner Niederschlagsobjekte. Dazu wurde ein einfacher, subjektiver Ansatz verwendet: der Niederschlagswert $R^* = f \cdot R^{max}$ definiert hier den Schwellwert, ab dem Objekte identifiziert werden. Damit ist der Schwellwert zur Identifikation von Objekten dynamisch (abhängig von R^{max}) gehalten, wobei R^{max} das Niederschlagsmaximum im Gebiet bezeichnet. Für den Faktor f wurde $1/15$ gewählt. Dieser Wert folgte aus der (subjektiven) Betrachtung verschiedener realer Niederschlagsfelder. Der Faktor $f = 1/15$ identifizierte dabei die Objekte in einem Niederschlagsfeld, welche auch subjektiv gewählt worden wären (weiter unten wird die Sensitivität von f auf SAL diskutiert).

Die Zuordnung der Gitterpunkte, die zu einem Objekt gehören, erfolgt durch einen zuvor zur Identifikation zusammenhängender Strukturen im Feld der Potentiellen Vorticity entwickelten Algorithmus (Wernli und Sprenger, 2007). Hat man einen Gitterpunkt mit $R_{ij} \geq R^*$ gefunden, so werden alle benachbarten Gitterpunkte dem Objekt hinzugefügt, bis das Schwellwertkriterium nicht mehr erfüllt wird. Die so gefundenen Objekte werden mit $\mathcal{R}_n, n = 1, \dots, M$ bezeichnet, wobei M die Anzahl an Objekten im Gebiet \mathcal{D} angibt. Die Schwerpunktkoordinaten der einzelnen Objekte werden dem entsprechend mit dem Vektor \mathbf{x}_n bezeichnet.

Die Identifizierung der Objekte erfolgt für Beobachtungen und Modell unabhängig. Sobald dies geschehen ist, wird die Niederschlagssumme für jedes Objekt berechnet: $R_n = \sum_{(i,j) \in \mathcal{R}_n} R_{ij}$. Die mittlere gewichtete Entfernung zwischen Objekten (repräsentiert durch deren jeweiligen Schwerpunkt \mathbf{x}_n) und dem Niederschlagsschwerpunkt im Gebiet ist definiert mit

$$r = \frac{\sum_{n=1}^M R_n \cdot |\mathbf{x} - \mathbf{x}_n|}{\sum_{n=1}^M R_n}. \quad (7.3)$$

Das Maximum für r kann dabei höchstens den Wert $d/2$ annehmen. Im Fall eines einzelnen Objektes im gesamten Gebiet wird $r = 0$ (Gl. 7.3). L_2 kann also nur größer null werden, wenn in mindestens einem Datensatz (Beobachtungen oder Modell) mehr als ein Objekt gefunden wird.

L_2 kann nun aus der Differenz zwischen r aus Beobachtungen und Vorhersage berechnet werden:

$$L_2 = 2 \cdot \frac{|r(R_{mod}) - r(R_{obs})|}{d}. \quad (7.4)$$

Der Faktor 2 ist dabei als Skalierung eingeführt, um den Wertebereich von L_2 dem von L_1 gleichzusetzen ($[0, \dots, 1]$).

Folglich kann die L -Komponente Werte zwischen 0 und 2 annehmen, wobei 0 für eine perfekte Vorhersage nur dann auftritt, wenn sowohl der Schwerpunkt der Niederschlagsvorhersage im Gebiet, als auch die mittleren Entfernungen dieses Schwerpunkts zu den Schwerpunkten der einzelnen Objekte mit den Werten aus den Beobachtungen übereinstimmen.

Die S-Komponente

Zuletzt folgt die Definition der Struktur-Komponente S . Die Grundidee ist, das Volumen skaliertes Niederschlagsobjekte miteinander zu vergleichen, um Informationen über die Größe und Form der Niederschlagsobjekte zu bekommen. Die S -Komponente ist dabei aber unabhängig von der A -Komponente, was später an verschiedenen Beispielen gezeigt wird. Zunächst muss also für die S -Komponente das skalierte Volumen V_n für jedes Objekt n berechnet werden:

$$V_n = \sum_{(i,j) \in \mathcal{R}_n} R_{ij} / R_n^{max}, \quad (7.5)$$

mit R_n^{max} als Niederschlagsmaximum innerhalb des Objekts (es gilt: $R_n^{max} \leq R^{max}$). Die Skalierung mit R_n^{max} ist der entscheidende notwendige Schritt, um S unabhängig von A zu machen.

Ist V_n für alle Objekte sowohl für Beobachtungen als auch für die Vorhersage berechnet, so kann daraus jeweils für jeden Datensatz das mittlere gewichtete Volumen aus allen skalierten Objekten berechnet werden:

$$V = \frac{\sum_{n=1}^M R_n V_n}{\sum_{n=1}^M R_n}. \quad (7.6)$$

Wie schon in Gl. 7.3 sind die Gewichtungen dabei proportional zur Niederschlagssumme R_n im jeweiligen Objekt n . S ist nun als die skalierte Differenz zwischen V aus Modell und Beobachtungen definiert (analog zu A , womit auch der Wertebereich von -2 bis $+2$ reicht und 0 eine perfekte Vorhersage im Sinne der S -Komponente bedeutet):

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5[V(R_{mod}) + V(R_{obs})]}. \quad (7.7)$$

Werte für S werden sehr groß, wenn im Modell beispielsweise großflächiger Niederschlag vorhergesagt wird, in der Realität aber kleine konvektive Ereignisse beobachtet werden. Negative Werte für S können im gerade umgekehrten Fall auftreten. Weitere Beispiele für negative und positive S -Werte sind in Abschn. 7.2 mit idealisierten Niederschlagsobjekten diskutiert.

Mit der Eigenschaft, Aussagen über die Qualität der vorhergesagten Niederschlagsstruktur machen zu können, ist S als Hauptkomponente von SAL anzusehen.

7.2 Idealisierte Beispiele

Um die Charakteristiken von SAL für Niederschlagsfelder besser verstehen zu können, ist eine Anwendung auf idealisierte Niederschlagsobjekte hilfreich. Zunächst sollen ein paar qualitative Aussagen über solche Beispiele getroffen werden, bevor dann mehrere verschiedene Objekte quantitativ mit SAL verglichen werden.

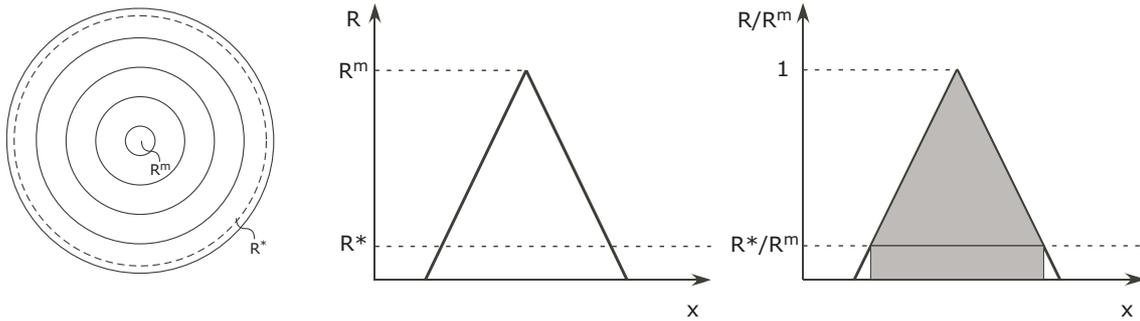


Abbildung 7.1: *Idealisiertes Niederschlagsobjekt mit Kegelform. Links: Niederschlagskonturen des Objekts mit dem Maximum R^m (im Text R^{max}) und dem Schwellwert R^* , der den für L und S verwendeten Teil des Objektes definiert. Mitte: Querschnitt durch die Mitte des Objekts. Rechts: gleicher Querschnitt, aber nach dem Skalieren mit R^m . Das zur Berechnung der S -Komponente verwendete Volumen ist grau hinterlegt.*

Qualitative Betrachtung

Der Einfachheit halber soll als erstes angenommen werden, dass in den Beobachtungen nur ein Niederschlagsobjekt im betrachteten Gebiet existiert. Dieses Objekt soll eine kreisrunde Grundfläche mit dem Niederschlagsmaximum genau in der Mitte haben. Steigt der Niederschlag vom Rand bis zum Maximum linear an, so ist dieses Objekt gleichbedeutend mit einem Kegel (Abb. 7.1). Das linke Bild zeigt die Konturen des Niederschlagsfeldes (1 Objekt) mit einem Maximum R^{max} und einem Schwellwert R^* , welcher den für die Berechnung der Komponenten L und S verwendeten Teil des Objektes definiert. Das mittlere Bild zeigt einen Querschnitt durch die Mitte des Objekts, das rechte Bild zeigt diesen Querschnitt nach der Skalierung mit R^{max} . Für die Berechnung von V (Gl. 7.6) werden nur die Gitterpunkte verwendet, für die $R_{ij} \geq R^*$ gilt (grau hinterlegt).

Dieses als Beobachtung verwendete, einzelne Objekt soll nun mit SAL für verschiedene idealisierte Beispiele als Modellvorhersagen (Abb. 7.2) untersucht werden. Diese Vorhersage-Objekte sollen ebenfalls alle kegelförmig sein, sich aber in ihrer Amplitude, Grundfläche oder Anzahl vom (beobachteten) Objekt in Abb. 7.1 unterscheiden. Für einzelne Objekte (Abb. 7.2) ist die Berechnung und Interpretation von L einfach (nur L_1 spielt eine Rolle: liegen die einzelnen Objekte im Modell und den Beobachtungen mit ihrem Maximum am gleichen Ort, so ist $L = 0$, andernfalls ist $L > 0$), weshalb hier vor allem auf die beiden Komponenten S und A eingegangen werden soll.

Beispiel 1:

Das Objekt in der Vorhersage (Abb. 7.2 (a)) hat die gleiche Grundfläche, aber eine kleinere Amplitude als das Objekt in den Beobachtungen (Abb. 7.1, $R_{mod}^{max} < R_{obs}^{max}$). A hängt nur vom Verhältnis der beiden Maxima ab, da der Schwellwert R^* für A keine Rolle spielt. Hier bedeutet dies, dass der gebietsgemittelte Niederschlag im Modell unterschätzt wird ($A < 0$).

Die im Modell und den Beobachtungen voneinander unabhängige Skalierung des

jeweiligen Objektes mit seinem Maximum führt zu zwei identischen Querschnitten und damit Volumen (jeweils das rechte Bild), was $S = 0$ bedeutet.

Gemäß SAL hat das vorhergesagte Niederschlagsfeld also die richtige Struktur ($S = 0$), während die Niederschlagsmenge unterschätzt wurde ($A < 0$).

Beispiel 2:

Es soll für die Vorhersage aus Abb. 7.2 (b) gelten, dass der Fehler in der Amplitude gerade null ist (die Unterschätzung der Amplitude wird durch die größere Ausdehnung des Objektes kompensiert: $A = 0$). Da jedoch die Grundfläche des Objekts im Modell deutlich größer ist, führt dies zu einem größeren skalierten Volumen (Abb. 7.2 (b, rechts)) als in den Beobachtungen (Abb. 7.1 (rechts)). Daraus folgt ein $S > 0$. In diesem Fall hat die Vorhersage also keinen Amplitudenfehler, aber einen positiven Strukturfehler.

Aus den beiden ersten Beispielen wird klar, dass A und S (wie in Abschn. 7.1 bereits erwähnt) unabhängig voneinander sind. Die Skalierung von V_n in Gl. 7.5 ist die Voraussetzung dafür, dass S trotz Amplitudenfehlers gleich null sein kann (Bsp. 1) oder ein Strukturfehler trotz richtiger Niederschlagsmenge vorliegt (Bsp. 2). Bis hierher scheint es verwunderlich, warum die Definition von S nicht alleine über die Grundfläche gemacht wird, da dies für die gezeigten Beispiele das gleiche Ergebnis bringen würde. Der Nutzen der etwas aufwändigeren Berechnung von S über das Volumen wird im nächsten Beispiel deutlich.

Beispiel 3:

Abbildung 7.2 (c) zeigt ein Niederschlagsobjekt (Vorhersage) mit derselben Grundfläche wie das in Abb. 7.1 (Beobachtungen), allerdings nicht mehr in Form eines zum Maximum linear zulaufenden Kegels, sondern nicht-linear und zum Maximum spitz zulaufend. Diese Form soll im weiteren Text mit „peakig“ beschrieben werden. Man kann annehmen, dass das peakige Modell-Objekt die gleiche Amplitude hat wie in den Beobachtungen, so dass $A = 0$ wird. Um S betrachten zu können, muss das Objekt wiederum mit R^{max} skaliert werden, was offensichtlich zu einem kleineren Volumen führt als in Abb. 7.1. Somit ist trotz der gleichen Grundfläche ein Strukturfehler mit $S < 0$ vorhanden. Vergleichbar dazu hätte ein zum Maximum hin abgeflachtes Objekt (gegenteilige Form von „peakig“) ein $S > 0$.

Dieses Beispiel zeigt, dass die S -Komponente nicht nur zwischen Objekten verschiedener Grundflächen, sondern auch verschiedener Niederschlagsstrukturen (peakig, flach) unterscheiden kann, auch wenn kein Amplitudenfehler vorhanden ist. Dies ist eine sehr nützliche Eigenschaft, da S sehr gut zwischen frontalen (typischerweise meist abgeflachte Objekte) und konvektiven (eher peakigen) Niederschlägen unterscheiden kann.

In den weiteren Beispielen sollen nun Fälle mit mehreren Objekten in einem Niederschlagsfeld betrachtet werden. Dabei wird zur Vereinfachung angenommen, dass die

Beobachtungen weiter ein einzelnes Objekt liefern (Abb. 7.1) und die Anzahl nur in der Vorhersage erhöht wird.

Beispiel 4:

Sind in der Vorhersage z.B. zwei identische Objekte identifiziert und gleich dem Objekt aus den Beobachtungen, so wird die Niederschlagsmenge gerade um den Faktor 2 vom Modell überschätzt: $A = 2/3$ (nach Gl. 7.1). S wird null, da alle Objekte die gleichen skalierten Volumen aufzeigen und im Gebiet immer das mittlere (gewichtete) Volumen errechnet wird. Die L -Komponente ist davon abhängig, wo sich die jeweiligen Objekte befinden. Dies wird im nächsten Beispiel näher diskutiert.

Beispiel 5:

Angenommen, in der Vorhersage seien wie in Bsp. 4 zwei Objekte mit (linearer) Kegelform, jedoch jeweils der halben Amplitude im Vergleich zu den Beobachtungen (Abb. 7.1) vorhanden, dann wären sowohl A , als auch S gleich null. Wenn die beiden Objekte in der Vorhersage nun jeweils den gleichen Abstand zum beobachteten Objekt hätten, allerdings genau in entgegengesetzter Richtung zum Schwerpunkt im Beobachtungsfeld, so wäre L_1 ebenso gleich null. Dies ist genau der Fall, für den SAL eine perfekte Vorhersage suggerieren würde, was aber nicht dem subjektiven Eindruck gleich käme. Deshalb wurde für L ein L_2 (Gl. 7.4) hinzugefügt, welches genau für diese Art einer nicht perfekten Vorhersage $L > 0$ liefert.

Wenn schon die beiden Objekte der Vorhersage nicht exakt wie oben beschrieben liegen, so bekommt man auch einen positiven Beitrag von L_1 . Diese Betrachtungen der L -Komponente können ebenso in Bsp. 4 gemacht werden.

Beispiel 6:

Als letztes Beispiel soll eine Situation mit zwei verschiedenen Objekten im Vorhersagefeld angenommen werden. Objekt 1 soll eine große Grundfläche wie in Abb. 7.2 (b), Objekt 2 dagegen eine kleinere Grundfläche haben und peakig sein (Abb. 7.2 (c)). A wird einen positiven Wert annehmen, außer die beiden Objekte haben eine deutlich kleinere Amplitude gegenüber dem beobachteten Objekt (Abb. 7.1). S kann positiv oder negativ sein, je nachdem welches Objekt mehr Einfluss durch die Gewichtung seiner Niederschlagsmenge in der mittleren Volumenberechnung hat (Gl. 7.6 und Gl. 7.7). Wenn das peakige Objekt eine kleine Grundfläche und/oder Amplitude hätte, würde klar das große Objekt dominieren und ein $S > 0$ folgen. Wenn das große Objekt eine sehr viel kleinere Amplitude gegenüber dem peakigen Objekt hat, kann $S < 0$ werden.

An diesem Beispiel kann man sehr schön sehen, dass die Gewichtung der Objekte mit ihrem Niederschlag dazu führt, dass die dominierenden Objekte auch den meisten Einfluss auf den Wert für S haben.

Bevor SAL quantifiziert werden soll, wird nochmal näher auf die Wahl des Schwellwertes und dessen Einfluss auf S und L eingegangen. Bei Ereignissen mit mehreren

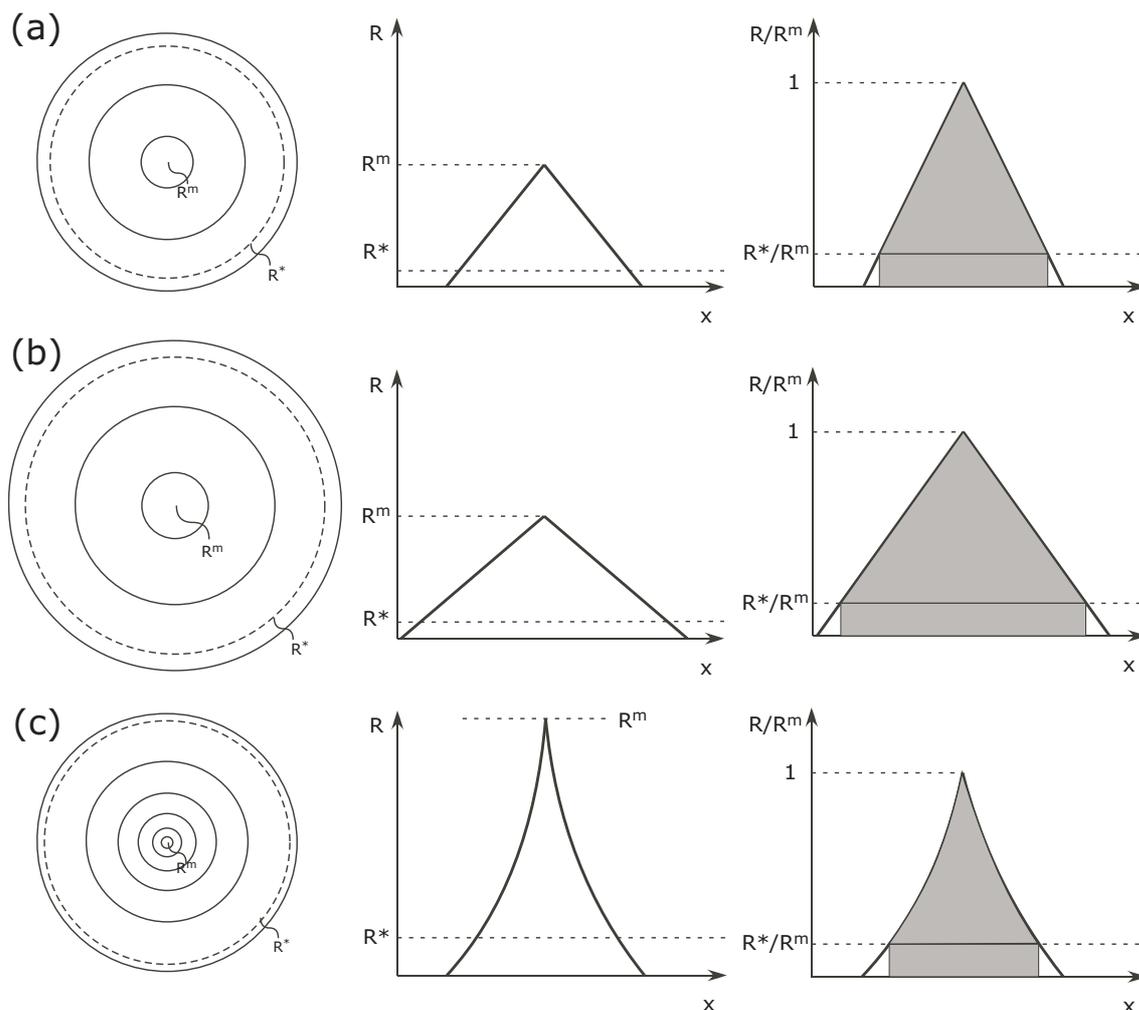


Abbildung 7.2: Wie Abb. 7.1 für drei verschiedene Objekte: (a) ein Kegel mit kleinerer Amplitude als in Abb. 7.1, (b) ein Kegel mit kleinerer Amplitude und größerer Grundfläche, (c) ein nichtlinear, spitz zulaufender Kegel („peakig“).

lokalen Niederschlagsmaxima im Gebiet kann die Wahl des Schwellwertes zu deutlich verschiedenen Objektzahlen (und damit Objektgrößen) führen. Dies kann die Werte für S und L_2 deutlich beeinflussen.

Dieses Problem wird hier als „Kamel-Effekt“ bezeichnet und ist schematisch in Abb. 7.3 illustriert. Je nachdem, wie klein das Minimum R^{min} zwischen den beiden Maxima ist, kann es sich um ein ($R^{min} > R^*$) oder zwei Objekte ($R^{min} < R^*$) handeln. Dies hat starke Auswirkungen auf S und L_2 . Angenommen, die beiden Maxima in Abb. 7.3 sind gleich in ihrer Amplitude und der darunter liegenden Grundfläche, so ist V doppelt so groß für den Fall eines einzelnen Objektes als im Fall mit zwei Objekten. Dies geschieht dadurch, dass R^{min} gerade etwas kleiner oder etwas größer als R^* ist. Diese Eigenschaft eines signifikanten Einflusses der Wahl des Schwellwertes auf SAL ist nicht wünschenswert. Im statistischen Sinne sollte dieser Kamel-Effekt allerdings kaum Bedeutung für SAL haben, da diese Situation relativ selten vorkommt (siehe auch Diskussion zur Sensitivität von SAL auf den Schwellwert in Abschn. 7.3).

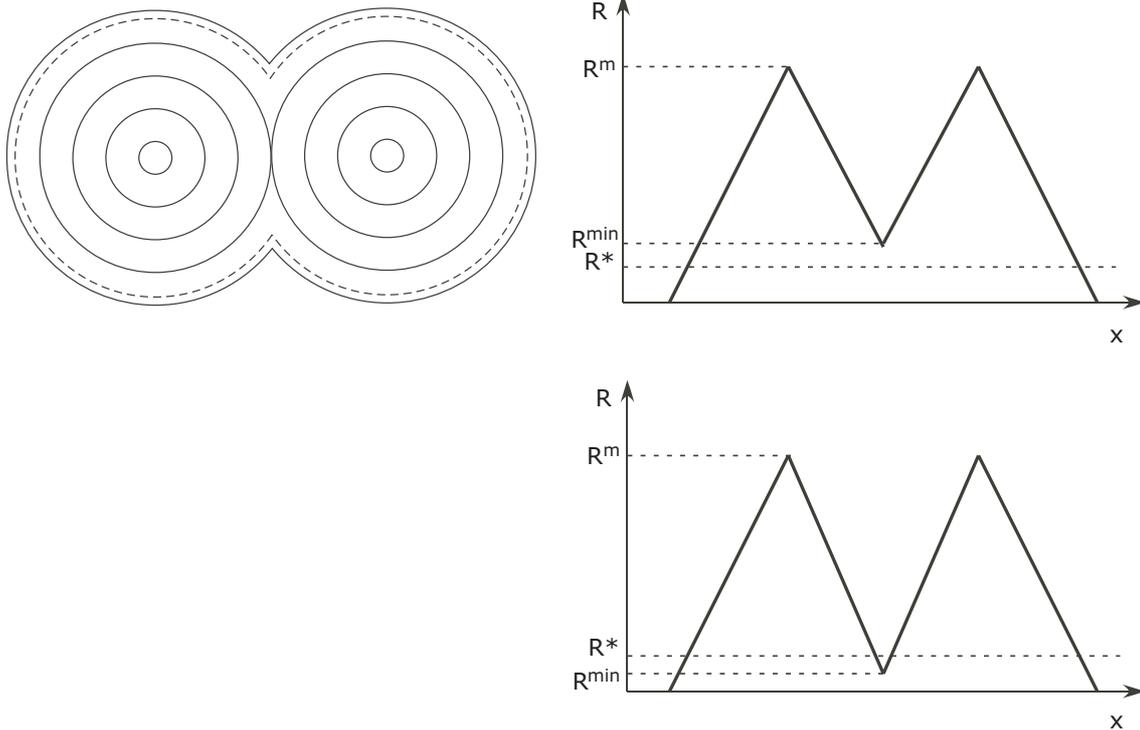


Abbildung 7.3: Schema wie Abb. 7.1, jedoch ohne Skalierung. Links: resultierendes Bild zweier etwas zusammengeschobener Kegel mit gleicher Grundfläche und Amplitude wie in Abb. 7.1. Rechts: Querschnitt durch die beiden Maxima. Je nach Wert für R^{\min} werden ein ($R^{\min} > R^*$) oder zwei ($R^{\min} < R^*$) Objekte identifiziert.

Quantitative Auswertung

In diesem Abschnitt soll SAL nun anhand konkreter Werte für verschiedene idealisierte Beispiele ausgewertet werden.

Abbildung 7.4 zeigt eine Auswahl von acht (bezeichnet mit A-H) verschiedenen idealisierten Niederschlagsfeldern mit einer Auflösung von $99 \cdot 99$ Gitterpunkten. SAL wurde nun auf alle möglichen Kombinationspaare dieser acht Bilder angewendet, die Ergebnisse dafür sind in Tab. 7.1 aufgeführt (Werte sind mit 100 multipliziert).

Die Einträge in Reihe A und Spalte C (Paar AC) zeigen z.B. S , A und L für A als Modell-Vorhersage und C als Beobachtungen. Die Diagonale muss daher immer $S = A = L = 0$ ergeben, weil die beiden Niederschlagsfelder identisch sind. Links der Diagonalen wurden keine Werte aufgeführt, da die Tabellenwerte antisymmetrisch für S und A , sowie symmetrisch für L zur Diagonalen sind.

Zunächst werden Objekte in den acht Bildern aus Abb. 7.4 beschrieben. A-D entsprechen einem linear zulaufenden Kegel (Prinzip wie in Abb. 7.1), wobei B eine deutlich größere Grundfläche als A hat, C gegenüber A verschoben ist und D gegenüber A eine gerade halbierte Amplitude aufweist. E und F haben jeweils 2 lokale Maxima gleicher Amplitude wie D, wobei diese in F weiter voneinander entfernt sind als in E. Das Objekt in G ist abgeflacht (breites Maximum in der Mitte und schnell abnehmend zum Rand hin), während H ein sogenanntes peakiges Objekt (Abb. 7.2 (c)) zeigt. Mit dem weiter oben eingeführten Schwellwert $R^* = R^{\max}/15$ finden sich für F zwei Objekte, sonst immer eines (auch in E).

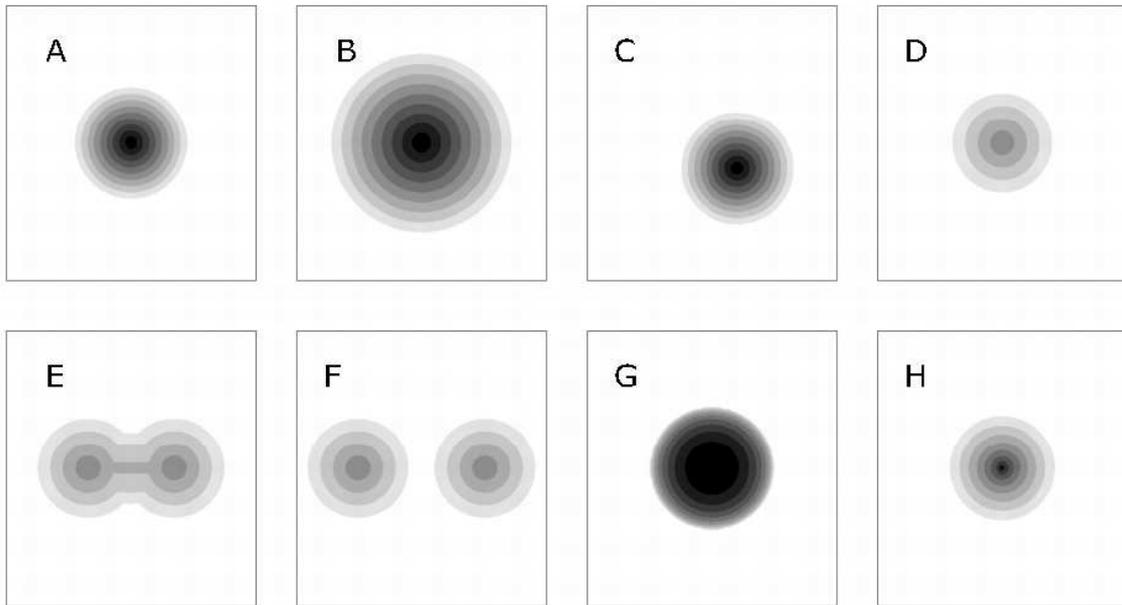


Abbildung 7.4: Beispiele für idealisierte Niederschlagsfelder in einem Gebiet (Auflösung: $99 \cdot 99$ Gitterpunkte). Graustufen zeigen Niederschlagshöhen an (je dunkler, je mehr Niederschlag).

Die quantitativen Ergebnisse in Tab. 7.1 stimmen mit den qualitativen Überlegungen weiter oben überein. Zunächst werden die Einträge diskutiert, für die nur eine Komponente von SAL ungleich null ist:

- die Paare AC und AF haben jeweils ein $L > 0$, aber aus verschiedenen Gründen: AC wegen der Teilkomponente L_1 (C ist verglichen mit A verschoben), AF wegen L_2 ,
- direkt daraus resultiert der Wert für L beim Paar CF, nämlich gerade die Summe aus $L(AC)$ und $L(AF)$.
- AD hat nur einen positiven Amplitudenfehler.
- AE ist die einzige Situation mit $A, L = 0$ und einem (negativen) Beitrag in S (hauptsächlich wegen der größeren Grundfläche in E).

Weiter werden Fälle betrachtet, in denen nur eine SAL-Komponente gleich null ist. Es gibt einige Kombinationspaare, für die S und A das gleiche Vorzeichen haben: AB, AG, AH, BD, BE, BG, BH, DE, DG, EG (S ist fast 0, da sich hier die S -Fehler von Grundfläche und Niederschlagsstruktur des Objektes kompensieren), EH und GH. Nur DH hat für S und A verschiedene Vorzeichen (die große Amplitude in H führt zu $A < 0$, jedoch ist wegen der peakigen Struktur in H $S > 0$ (obwohl die Grundfläche gleich ist)).

In den meisten Situationen mit S und A ungleich null kommt es also bei Unterschätzung der Amplitude zu einem negativen S , bei Überschätzung zu positivem S . Wichtig ist, dass diese Tatsache kein Ergebnis der mathematischen Definition,

	A	B	C	D	E	F	G	H
A	0/0/0	-88/-88/0	0/0/14	0/67/0	-67/0/0	0/0/18	-68/-67/0	40/38/0
B		0/0/0	88/88/14	88/135/0	24/88/0	88/88/18	23/25/0	117/116/0
C			0/0/0	0/67/14	-67/0/14	0/0/32	-68/-67/14	40/38/14
D				0/0/0	-67/-67/0	0/-67/18	-68/-120/0	40/-31/0
E					0/0/0	67/0/18	-1/-67/0	100/38/0
F						0/0/0	-68/-67/18	40/38/18
G							0/0/0	101/98/0
H								0/0/0

Tabelle 7.1: SAL (Format S/A/L, Werte sind mit 100 multipliziert) für alle möglichen Kombinationspaare aus Abb. 7.4. Werte links der Diagonalen wurden weggelassen, da S und A antisymmetrisch, L symmetrisch zur Diagonalen ist.

Fall	S	A	L	L_1	L_2
1	0.119	0.312	0.033	0.033	0.000
2	1.597	0.877	0.202	0.026	0.176
3	-0.430	-0.057	0.196	0.191	0.005
4	1.598	1.325	0.303	0.174	0.129

Tabelle 7.2: SAL für die vier in Abb. 7.5 und Abb. 7.6 gezeigten realen Fälle. Zusätzlich sind die Werte der beiden Teilkomponenten von L (L_1 und L_2) aufgelistet. Für den Schwellwert zur Objektidentifikation wurde $R^{\max}/15$ gewählt.

sondern der Wahl der Beispiele ist. Vergleicht man z.B. Bild A mit B, so kann man sich überlegen, was passiert, wenn man die Amplitude bei B kontinuierlich herabsetzen würde. Die A -Komponente würde sich natürlich ändern, nicht aber S .

Die restlichen Einträge in Tab. 7.1 lassen sich analog dazu auswerten. Auf einen Fall soll aber noch hingewiesen werden: EF. Die A -Komponente ist hier gleich null, weil das Niederschlagsfeld aus jeweils zwei identischen Objekten besteht, die in E nur zusammengeschoben wurden (was eine Summierung an sich überlappenden Gitterpunktwerten bedeutet). L ist größer null, weil der Schwerpunkt in E (ein Objekt) und F (zwei Objekte) zwar identisch ist, jedoch der Objektabstand zu diesem Schwerpunkt in F eine Rolle spielt (L_2 , Gl. 7.4). S ist deutlich größer null, weil in E vor allem die Grundfläche größer ist als in den Objekten von F (welche ja zur S -Berechnung gemittelt werden).

7.3 Anwendung auf reale Niederschlagsfelder

Nachdem SAL für idealisierte Beispiele ausführlich und erfolgreich getestet wurde, soll jetzt die Anwendung auf ausgewählte reale Niederschlagsereignisse erfolgen. Dabei wurden vier Fälle aus dem Sommer 2001 für das Einzugsgebiet der Elbe (deutscher Teil, Abschn. 2.7) ausgewählt und Vorhersagen des aLMo aus 00 UTC-Läufen mit den Beobachtungen (Abschn. 3.1) verglichen. Diese vier Fälle unterscheiden sich im Hinblick auf die meteorologischen Bedingungen und damit in der Niederschlags-

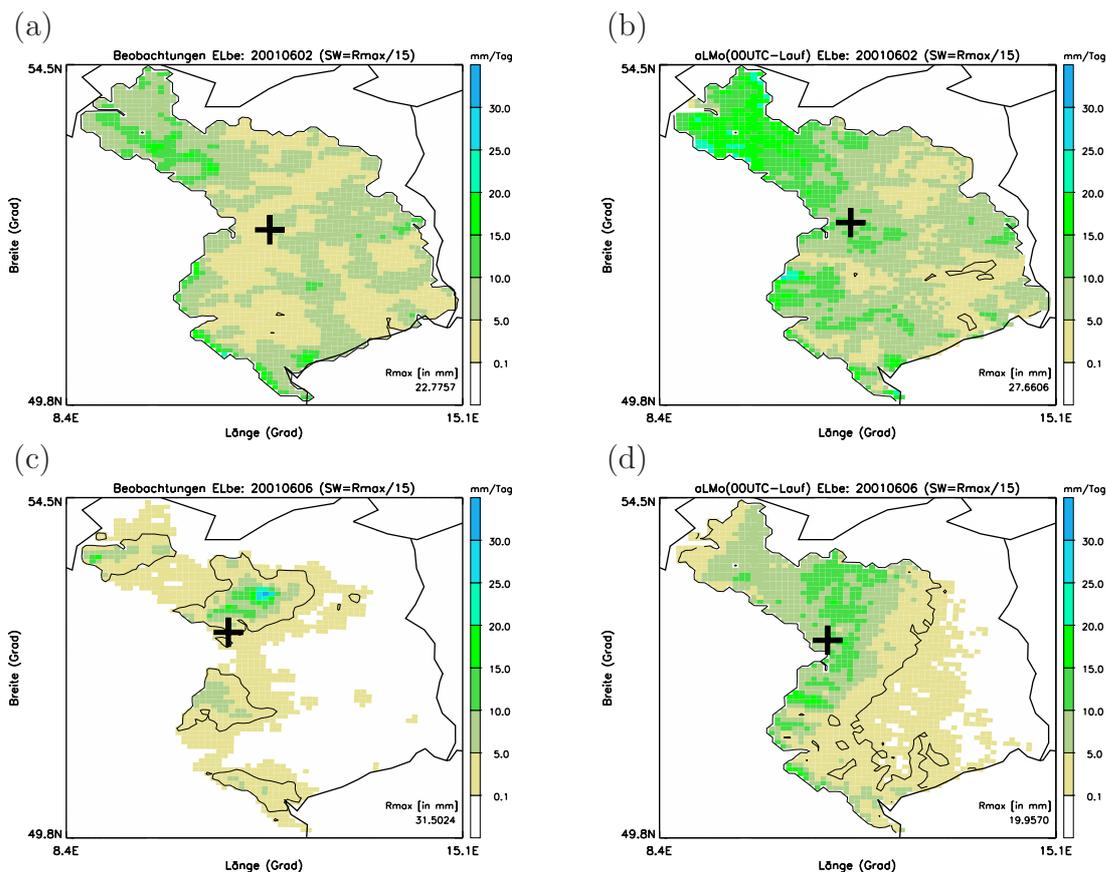


Abbildung 7.5: Ausgewählte Tagessummen aus Beobachtungen (links) und aLMO-Vorhersagen (00 UTC-Läufe, rechts) zum Test von SAL an realen Beispielen: 02.Juni 2001 (oben) und 06.Juni 2001 (unten). Die schwarze Konturlinie zeigt den Schwellwert zur Objektidentifizierung an ($R^* = R^{max}/15$), das schwarze Kreuz markiert den Niederschlagsschwerpunkt im Gebiet (L_1 , Gl. 7.2).

verteilung.

Die allgemeine Wetterlage für die gewählten Fälle soll kurz beschrieben werden (Berliner Wetterkarten, 2001):

- **1. Fall (02.Juni 2001):** Ein intensives Tiefdruckgebiet lag über Dänemark. Warm- und Kaltfronten zogen über Deutschland hinweg und führten zu weit verbreiteten Niederschlägen im Elbe-Gebiet (Abb. 7.5 (a)). Die maximale Temperatur lag bei 17°C.
- **2. Fall (06.Juni 2001):** Die Höchsttemperatur war weiterhin relativ niedrig (< 18°C). Ein heran nahendes Tiefdruckgebiet über der Nordsee löste vereinzelte Schauer aus (Abb. 7.5 (c)).
- **3. Fall (07.Juli 2001):** Die Warmfront eines Tiefdruckgebietes zog über Deutschland hinweg und führte zu sehr intensiven Niederschlägen (Abb. 7.6 (a)). Die Höchsttemperatur erreichte 32°C.

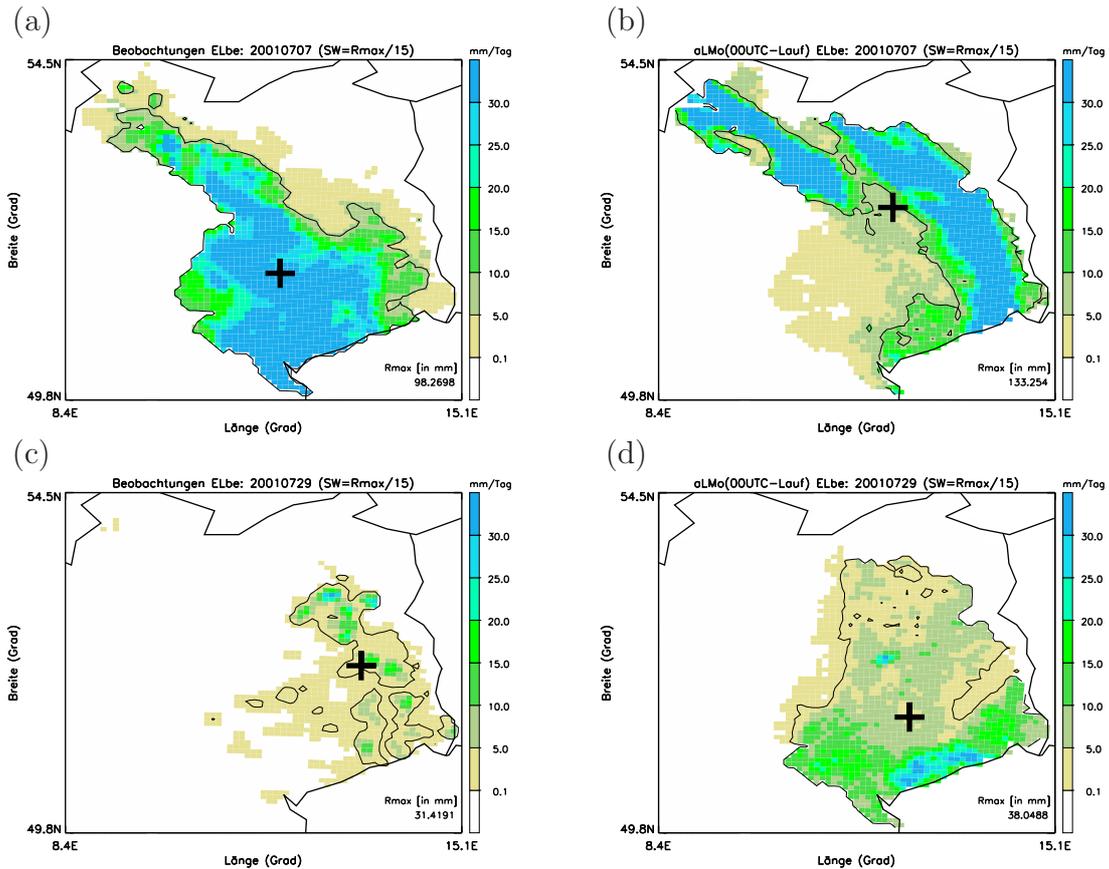


Abbildung 7.6: Wie Abb. 7.5 für 07. Juli 2001 (oben) und 29. Juli 2001 (unten).

- **4. Fall (29. Juli 2001):** Über West- und Mitteleuropa herrschte eine Hochdruckwetterlage mit Temperaturen bis 32°C . Es kam zu vereinzelt konvektiven Niederschlägen im Elbe-Einzugsgebiet (Abb. 7.6 (c)).

In Tab. 7.2 sind die Ergebnisse für SAL angewandt auf diese vier realen Fälle dargestellt.

Im 1. Fall ist der Niederschlag relativ homogen im gesamten Gebiet verteilt und nahezu alle Gitterpunkte R_{ij} haben Niederschlagswerte oberhalb des Schwellwertes $R^* = R^{max}/15$ (schwarze Konturlinie). Sowohl in den Beobachtungen, als auch im Modell, wird ein großes, ungefähr dem Einzugsgebiet entsprechendes Objekt identifiziert. Alle SAL-Komponenten sind relativ klein, womit die Qualität der Vorhersage als sehr gut bewertet wird. Der größte Fehler liegt in der Amplitude ($A = 0.312$). Dies liegt hauptsächlich an der Überschätzung des Niederschlags im nordwestlichen Teil des Gebietes. Die beiden Schwerpunkte des Gebietsniederschlags (L_1) liegen nah beieinander, weshalb L fast null ist.

Im 2. Fall ist der Niederschlag im Gebiet deutlich variabler verteilt, besonders in den Beobachtungen. Dort werden vier dominierende und ein paar kleinere Objekte identifiziert, im Modell nur ein großes Objekt. Dies zeigt sich im großen Wert für S . Das Modell kann die lokalen und eher peakigen Objekte in den Beobachtungen nicht wiedergeben. Der Niederschlag wird mit $A = 0.877$ zusätzlich deutlich überschätzt (vgl. Abschn. 7.2: Vorzeichen von S und A meist identisch). Die L -Komponente zeigt einen mittleren Wert an, obwohl der Schwerpunkt (L_1) gut getroffen wurde. Es gibt

aber dennoch einen (sinnvollen) Eintrag in L aus L_2 (mittlerer gewichteter Abstand der Objektschwerpunkte zum Gebietsschwerpunkt: Gl. 7.4).

Im 3. Fall ist in beiden Datensätzen ein dominierendes intensives Niederschlagsband zu erkennen, das aber in den Beobachtungen ausgedehnter ist. Diese Grundfläche des beobachteten Objektes ist genau der Grund, warum hier $S < 0$ ist. A ist nahe bei null, der Niederschlag wird im Gebietsmittel nur leicht unterschätzt. Wie aus Tab. 7.2 gut herauszulesen ist, spielt nur ein großes Objekt eine Rolle: L ist ungefähr gleich L_1 , wobei L eine Verschiebung des Niederschlagsbandes um etwa 20 % der maximalen Distanz zweier Randkoordinaten im Elbe-Gebiet anzeigt (ca. 100 km).

Im 4. Fall fällt in den Beobachtungen konvektiver Niederschlag aus einer Vielzahl an Objekten. Diese Wetterbedingungen kann das aLMO (an diesem Tag) nicht simulieren und sagt flächigen Niederschlag vorher. Dies führt insgesamt zu sehr hohen Werten von S und A . Auch im Sinne von L ist die Vorhersage schlecht, wobei der Beitrag aus beiden Anteilen (L_1 und L_2) in der etwa gleichen Größenordnung liegt.

Sensitivität des Schwellwertes

Abschließend soll noch eine kurze Studie zur Sensitivität des Schwellwertes auf die Komponenten S und L gemacht werden. Dazu wurde der Faktor f , welcher den Schwellwert abhängig von R^{max} bestimmt ($R^* = fR^{max}$), um den bis dato gewählten Wert ($f = 1/15$) für die vier realen Fälle variiert: $f = (1/13, \dots, 1/17)$. Folgende Ergebnisse wurden errechnet:

- **1. Fall:** S variiert um weniger als 1 %, L bis zu 10 %.
- **2. Fall:** S variiert um 1.5 %, L um 3 %.
- **3. Fall:** S springt für $f = 1/13$ von -0.430 auf -0.830, L von 0.196 auf 0.366 (Kamel-Effekt (vgl. Abb. 7.3)).
- **4. Fall:** S variiert um weniger als 1 %, L um 2 %.

Insgesamt ist SAL also sehr robust auf zumindest kleinere Änderungen im dynamisch gewählten Schwellwert. Vor allem der 3. Fall zeigt aber eine klare Sensitivität auf die Variation des Schwellwerts. Dies ist zum einen durch den Kamel-Effekt gut verstanden, zum anderen sind solche Beispiele statistisch gesehen von untergeordneter Bedeutung (der 3. Fall wurde auch speziell ausgewählt, um diesen Kamel-Effekt an einem realen Beispiel zeigen zu können).

7.4 Fazit

In diesem Kapitel wurde die Entwicklung von SAL detailliert beschrieben (sowohl konzeptionell, als auch mathematisch). Dabei war das Ziel, mit SAL eine Methode zu entwickeln, die speziell die Qualität von Niederschlagsvorhersagen in zuvor ausgewählten Gebieten untersucht.

Die Definition von SAL ist so gewählt, dass alle drei Komponenten unabhängig voneinander sind. SAL soll dabei nahe an einer subjektiven Verifikation orientiert sein und die wichtigsten Fragen zur Qualität einer Niederschlagsvorhersage beantworten:

S: Ist die Niederschlagsstruktur gut wiedergegeben? (z.B. konvektive oder frontale Ereignisse)

A: Stimmt die Niederschlagsmenge im Gebiet? (z.B. wichtig für Hydrologen)

L: Ist das Niederschlagsereignis am richtigen Ort vorhergesagt? (z.B. Gefährdung bestimmter Gebiete durch Starkniederschläge, Planung von Veranstaltungen)

Kommt man auf Abb. 1.4 aus der Einleitung (Kap. 1) zurück und betrachtet die fünf idealisierten Beispiele nun mit SAL, so zeigt sich, dass der subjektive Eindruck mit (a) als bester Vorhersage von SAL bestätigt wird (im Gegensatz zu gitterpunkt-basierten Fehlermaßen). In Struktur und Amplitude sind die Vorhersagen in (a), (b) und (d) gleich gut, jedoch klassifiziert die *L*-Komponente letztlich Bild (a) als insgesamt beste der drei Vorhersagen. In (c) und (e) kann man sich schnell klar machen, dass dort nicht nur *L* relativ schlecht ist, sondern auch *S* und *A* Fehler liefern.

Durch die vielen durchgeführten und teilweise hier gezeigten Fallstudien sowohl für idealisierte (Abschn. 7.2), als auch reale Niederschlagsfelder (Abschn. 7.3), wird SAL als sehr nützliches Qualitätsmaß zur Niederschlagsverifikation eingestuft.

Kapitel 8

SAL: statistische Auswertung

Nachdem SAL als neues Qualitätsmaß für die Niederschlagsvorhersage vorgestellt und nach detaillierten Untersuchungen als nützlich eingestuft wurde (Kap. 7), folgt in diesem Kapitel die statistische Auswertung für den Zeitraum von 2001-2004. SAL soll dabei die Niederschlagsvorhersagen der drei in dieser Arbeit verwendeten Modelle verifizieren (aLMo, LM, ECMWF-Globalmodell).

In Abschn. 8.1 werden Tagessummen betrachtet. Zunächst werden Vorhersagen des aLMo aus 00 UTC-Läufen mit Beobachtungen verglichen und dabei auch mögliche saisonale Unterschiede, sowie die Abhängigkeit vom Schwellwert untersucht. Weiter wird der mögliche Einfluss der betrachteten Gebietsgröße auf SAL-Werte anhand verschieden großer Flusseinzugsgebiete untersucht. In den beiden darauf folgenden Abschnitten werden Niederschlagsvorhersagen des aLMo mit denen des ECMWF-Globalmodells (Abschn. 8.2) und des LM (Abschn. 8.3) verglichen. Dabei zeigen sich vor allem beim Vergleich mit dem weniger hoch aufgelösten Globalmodell signifikante Unterschiede in der S -Komponente.

In Abschn. 8.4 wird SAL auf stündlicher Zeitskala angewendet. Dabei werden aLMo-Vorhersagen der 00 UTC-Läufe mit dem disaggregierten Datensatz (Kap. 3) verifiziert. Es werden Unterschiede zwischen Winter und Sommer, verschiedenen Schwellwertbereichen und Vorhersagezeiten am Beispiel des Elbe-Gebietes aufgezeigt. Zum Schluss folgt ein kurzes Fazit (Abschn. 8.5).

8.1 SAL für Tagessummen: Verifikation des aLMo

Ein erster Blick soll der Auswertung mit SAL für ganz Deutschland gelten. Abb. 8.1 zeigt SAL für Tagessummen aus aLMo-Vorhersagen (00 UTC-Läufe) verglichen mit gegitterten Stationsmessungen. Die gezeigte Abbildung wurde als Standard für die Darstellung von SAL gewählt und soll zunächst erläutert werden: jeder Punkt zeigt einen verifizierten Zeitbereich (in Abb. 8.1 sind dies Tagessummen für das Gebiet Deutschland). Dabei sind die Komponenten S und A gegeneinander aufgetragen und die Farbkodierung zeigt den jeweiligen L -Wert. Die Farbskala für L ist links oben im Bild eingefügt. In der Farbskala sind das arithmetische Mittel \bar{L} (schwarze Linie) und die Standardabweichung $\pm\sigma_L$ (graue Linien) eingezeichnet (Hinweis: die Farbskala

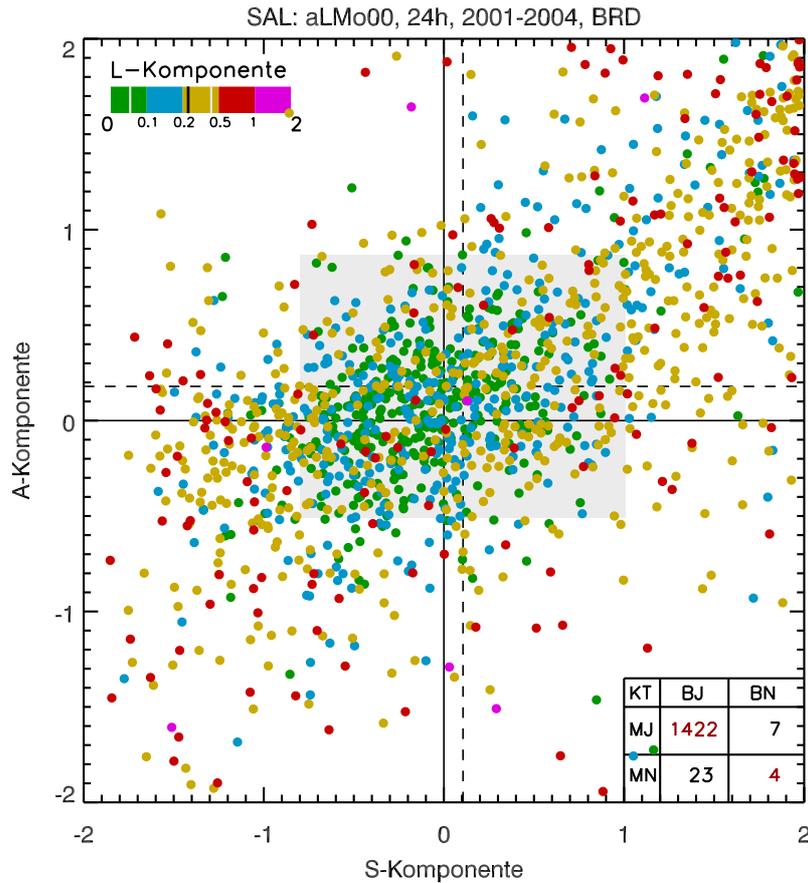


Abbildung 8.1: SAL für aLMO-Vorhersagen (00 UTC-Läufe) für ganz Deutschland. Jedes Symbol steht für SAL einer Tagessumme, wobei L farbkodiert ist (Grün = sehr gut bis Magenta = sehr schlecht, Legende links oben). Die Kontingenztabelle (KT, rechts unten, vgl. Tab. 6.1) zeigt die Anzahl an verifizierten Zeitpunkten (in Rot, wobei korrekt vorhergesagte Ereignisse für „kein Niederschlag“ gilt: $S = A = L = 0$). Die schwarzen Linien zeigen die perfekten Werte von S und A ($=0$, durchgezogen) und die arithmetischen Mittel \bar{S} und \bar{A} (gestrichelt). Die grau-schattierte Box zeigt die Standardabweichungen $\pm\sigma_S$ und $\pm\sigma_A$. In der Farblegende sind \bar{L} (schwarze Linie) und $\pm\sigma_L$ (graue Linien) eingezeichnet (Hinweis: die Farblegende ist nicht linear eingeteilt).

ist nicht linear eingeteilt). Die Mittelwerte von S und A sind jeweils durch schwarze gestrichelte Linien im Bild markiert (wobei alle Punkte ohne Gewichtung in die Berechnung eingehen, also schwache Niederschläge genauso gewertet werden wie starke Niederschläge). Die durchgezogenen schwarzen Linien zeigen perfekte Werte an, also $S, A = 0$. Durch eine graue Box im Hintergrund sind die Standardabweichungen von S und A eingezeichnet. Zuletzt ist rechts unten im Bild die Kontingenztabelle (KT) aufgeführt (vgl. Tab. 6.1), wobei der zugrunde liegende Schwellwert von 0.1 mm an mindestens einem Gitterpunkt im Gebiet überschritten werden muss, damit ein Niederschlagsereignis als solches identifiziert wird. Die mit Rot markierten Zahlen der KT zeigen die Anzahl der mit SAL verifizierten Werte an. Dabei werden aber nur Zeitpunkte mit Niederschlag in Vorhersage und Beobachtungen im Bild dargestellten Punkte (für Ereignisse mit null Niederschlag in beiden Datensätzen gilt direkt $S, A, L = 0$).

Abbildung 8.1 zeigt zunächst anhand der Summe der KT für Deutschland, dass von 1461 Tagen (4 Jahre) nur 5 Tage nicht verfügbar sind (wegen fehlender Vorhersagen (vgl. Tab. 2.1)). Weiter zeigt sich auch, dass nur 30 Fälle nicht mit SAL ausgewertet werden können, da das aLMO entweder den Niederschlag verpasst (23 Tage) oder an einem vollständig trockenen Tag Niederschlag produziert hat (7 Tage). Nur 4 Tage treten auf, in denen in beiden Datensätzen kein Niederschlag fällt.

In Abb. 8.1 erkennt man einen Zusammenhang zwischen S und A in der Weise, dass beide Komponenten häufig das gleiche Vorzeichen haben (vgl. Abschn. 7.2). Damit ergibt sich eine Charakteristik der Verteilung von Quadrant III nach Quadrant I mit einer deutlichen Häufung von SAL im Bereich um den Ursprung, also einer (mehr oder weniger) perfekten Vorhersage. Außer dieser Häufung in der Bildmitte ist auch eine etwas dichtere Anhäufung von Punkten im Quadrant I für $S \simeq 2$ sichtbar. Sehr gute Werte für die L -Komponente (grüne Punkte) finden sich bis auf wenige Ausnahmen im mittleren Bereich von Abb. 8.1. Gute Vorhersagen im Sinne von L bedeuten also zumeist auch gute Vorhersagen für S und A . Allerdings zeigt der Mittelwert von $\bar{L} > 0.2$ an, dass es ein Problem mit der richtigen Lage der Niederschlagsfelder gibt (für Deutschland). Die Mittelwerte von S und A liegen beide im positiven Bereich, was eine Überschätzung des Niederschlags (A -Komponente) im Zusammenhang mit im Mittel zu großen Niederschlagsobjekten im aLMO bedeutet. Die Streuung ist bei S etwas größer als bei A .

Was bedeuten nun die Einträge in einem der Quadranten im Sinne von SAL?

- **Quadrant I:** S und A größer null bedeutet, dass eine Überschätzung der Niederschlagsmenge im Zusammenhang mit zu großen Objekten im aLMO vorliegt. Auch Modell-Objekte mit im Vergleich zu den Beobachtungen abgeflachter Struktur (vgl. Kap. 7) fallen in diese Kategorie. Da der Mittelwert in Abb. 8.1 in diesen Quadranten fällt, ist dies das typischste Verhalten des aLMO. (mit 38.3% liegen die meisten Punkte in diesem Quadranten)
- **Quadrant II:** $S < 0$ und $A > 0$ sind Fälle mit Niederschlagsüberschätzung und gleichzeitig zu kleinen oder zu peakigen (vgl. Kap. 7) Objekten. Diese Objekte haben eine relativ geringere Ausdehnung, allerdings mit einer zu hohen Intensität gegenüber den beobachteten Objekten. (20.9% der Punkte liegen in diesem Quadranten)
- **Quadrant III:** S und A kleiner null bedeutet gerade den umgekehrten Fall von Quadrant I, dass nämlich das aLMO den Niederschlag unterschätzt im Zusammenhang mit zu kleinen Objekten, bzw. zu peakigen Objekten im Vergleich zu den Beobachtungen. (28.3% der Punkte liegen in diesem Quadranten)
- **Quadrant IV:** $S > 0$ und $A < 0$ sind Fälle mit Niederschlagsunterschätzung und gleichzeitig zu großen oder abgeflachten Objekten. Diese Objekte haben eine relativ große Ausdehnung, allerdings auch eine deutlich geringere Intensität als die beobachteten Objekte (Gegenteil zu Quadrant II). Dies kann z.B. auftreten, wenn schwacher stratiformer Niederschlag für einen Fall mit inten-

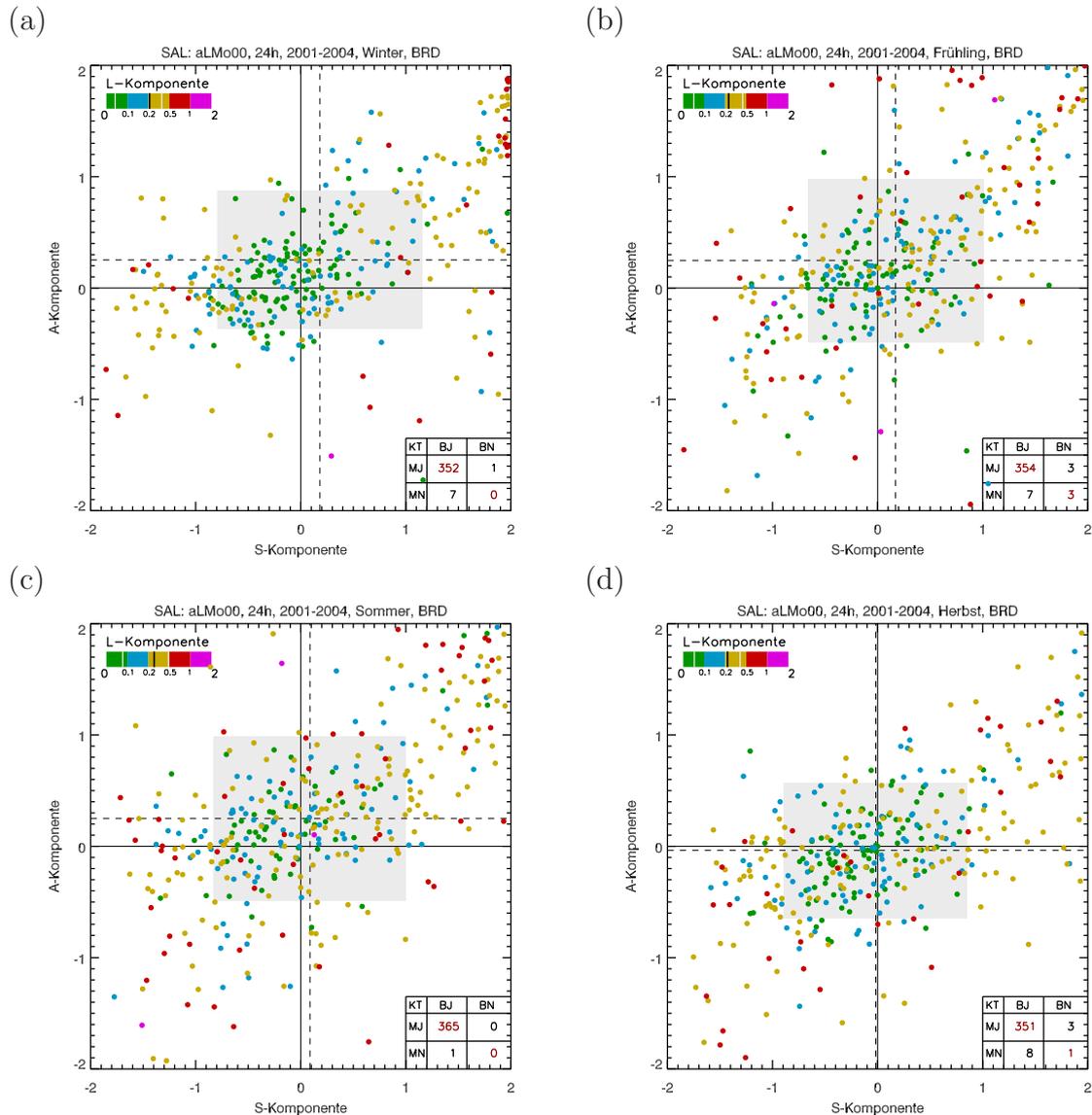


Abbildung 8.2: SAL saisonal aufgeteilt für Tagessummen aus aLMo-Vorhersagen (00 UTC-Läufe) für ganz Deutschland: (a) Winter, (b) Frühling, (c) Sommer und (d) Herbst.

siven kleinskaligen Schauern vorhergesagt wird. (mit 12.5% liegen in diesem Quadranten die wenigsten Punkte)

In den folgenden Abschnitten werden verschiedene Abhängigkeiten oder Unterschiede zwischen zwei Modellen mit dem SAL-Diagramm untersucht. Dabei werden meist Aussagen mit Hilfe der Mittelwerte und Standardabweichungen von S , A und L getroffen. Es wird auch aufgeführt, wieviele Vorhersagen gut im Sinne von SAL sind, wobei für „gute Vorhersagen,“ gilt: $S, A = [-0.5, \dots, 0.5]$, $L = [0, \dots, 0.2]$. In Abb. 8.1 sind dies 349 Vorhersagen (24.5%).

aLMo	\bar{S}		σ_S		\bar{A}		σ_A		\bar{L}		σ_L	
	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA
BRD	18	9	98	91	25	25	62	74	21	28	17	20
Ems	37	26	77	75	11	10	80	91	22	28	23	26
Weser	31	14	87	85	21	14	76	86	22	28	18	19
Donau	21	-8	88	92	19	17	78	100	22	28	19	19
Elbe	53	30	82	97	30	29	74	92	26	32	21	20
Rhein	26	15	90	90	24	29	66	83	22	29	16	19
ECMWF												
BRD	106	107	58	65	36	15	59	67	24	29	19	19
aLMo 2004												
BRD	17	1	99	90	22	19	66	55	22	25	17	19
LM 2004												
BRD	32	14	83	81	40	5	62	60	21	25	16	18

Tabelle 8.1: Mittel und Standardabweichung (mit 100 multipliziert) für S , A und L aus Tagessummen für Winter (DJF) und Sommer (JJA), verschiedene Einzugsgebiete (EZG) und Modelle: Deutschland (BRD) für aLMo, LM und ECMWF (jeweils 00 UTC-Läufe). Ems, Weser, Donau, Elbe und Rhein für aLMo (00 UTC-Läufe).

Saisonaler Vergleich

In Abb. 8.2 ist nun SAL für Tagessummen in Deutschland in Jahreszeiten aufgeteilt dargestellt. Folgende Aussagen können getroffen werden:

- Von Winter bis Herbst gibt es eine kontinuierliche Abnahme von \bar{S} von 0.18 bis -0.02.
- Im Winter, Frühling und Sommer gibt es eine identische Überschätzung (im Mittel) mit $A = 0.25$, im Herbst dagegen eine leichte Unterschätzung mit $A = -0.04$. Dies ist konsistent mit den saisonal gemittelten Niederschlägen aus Kap. 4.
- Im Winter ist L im Mittel am besten (0.21), im Sommer am schlechtesten (0.28). Auch die Streuung ist im Winter geringer (0.17) als im Sommer (0.2), was gut mit der Erwartung übereinstimmt, dass konvektive Zellen im Sommer sehr schwer am richtigen Ort vorhersagbar sind.
- Im Winter ist die Streuung für S am größten. Dies ist auch eine Folge der Häufung von Punkten bei $S \simeq 2$.
- In Frühling und Sommer, also den eher durch konvektiven Niederschlag geprägten Jahreszeiten, ist die Streuung von A am größten.

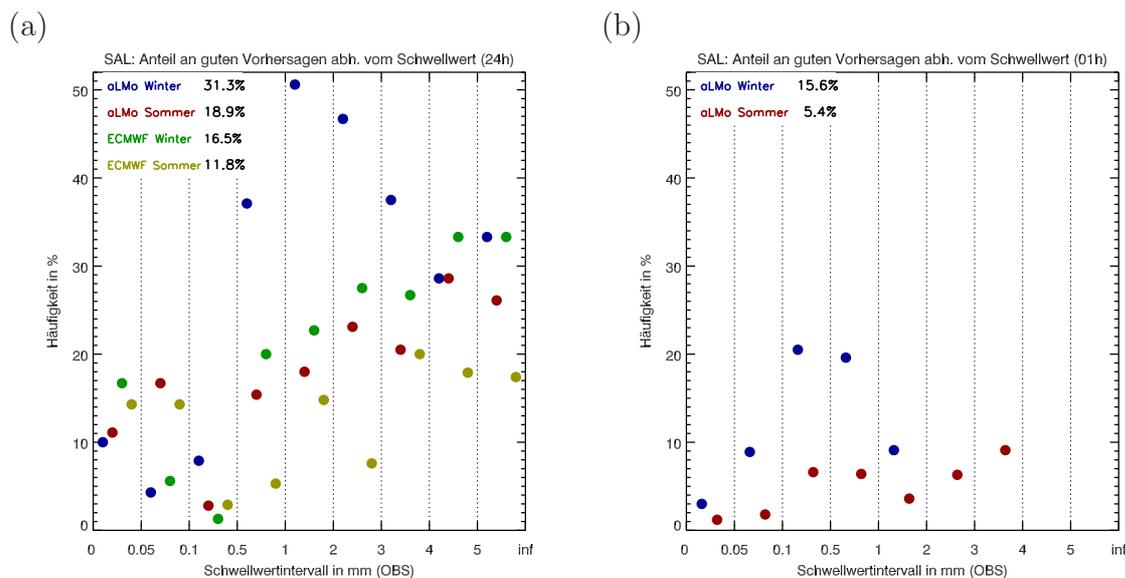


Abbildung 8.3: Prozentualer Anteil an „guten Vorhersagen“ (s. Text) innerhalb eines Schwellwertbereiches (vgl. Werte in Tab. 8.2) für Tagessummen (links) und Stundensummen (rechts). Der Schwellwertbereich bezieht sich auf den Schwellwert aus den Beobachtungen ($R_{obs}^* = R_{obs}^{max}/15$). Blau: aLMo-Vorhersagen im Winter, Rot: aLMo-Vorhersagen im Sommer, Grün: ECMWF-Vorhersagen im Winter, Gelb: ECMWF-Vorhersagen im Sommer (immer 00 UTC-Läufe). (Saisonaler Anteil an guten Vorhersagen im Bild oben links)

Die genauen Zahlen für Mittelwerte und Standardabweichungen finden sich für Winter und Sommer auch in Tab. 8.1. Die Diskussion der Tabelle folgt in den späteren Abschnitten an den entsprechenden Stellen. Zur erwähnten Häufung von Punkten mit hohen S -Werten (und damit einem größeren Einfluss auf den Mittelwert) kommt im Winter aber auch die größte Häufung an Punkten um den Ursprung, also nahezu perfekten Vorhersagen (zumal die meisten Punkte dort Grün und Blau gefärbt sind).

Abhängigkeit vom Schwellwert

In diesem Teilabschnitt soll geklärt werden, wie hoch der Anteil an guten Vorhersagen in den Jahreszeiten ist und ob es Unterschiede zwischen verschiedenen Schwellwertbereichen gibt. Dazu werden Sommer und Winter miteinander verglichen.

Abbildung 8.3 zeigt den prozentualen Anteil an guten Vorhersagen jeweils für verschiedene Schwellwertbereiche für Tagessummen (aLMo, ECMWF) und Stundensummen (aLMo), wobei die Werte für das Globalmodell und für Stundensummen an den entsprechenden Stellen interpretiert werden. Die Schwellwertbereiche gehen immer vom Schwellwert in den Beobachtungen ($R_{obs}^* = R_{obs}^{max}/15$) aus. Zusätzlich sind die Gesamtanzahlen an Vorhersagen in jedem Schwellwertbereich in Tab. 8.2 aufgeführt.

Zunächst kann man anhand der Gesamtanzahlen feststellen, dass für $R_{obs}^* < 1$ mm mehr Ereignisse im Winter als im Sommer auftreten, für $R_{obs}^* > 2$ mm deutlich mehr im Sommer (dies ist konsistent mit Ergebnissen für Häufigkeiten und Intensitäten

SW-Intervall in mm	aLMO 24h		ECMWF 24h		aLMO 01h	
	DJF	JJA	DJF	JJA	DJF	JJA
0-0.05	10	9	6	7	911	589
0.05-0.1	23	6	18	7	1254	490
0.1-0.5	89	36	76	35	3846	3000
0.5-1	70	39	55	38	194	1508
1-2	89	89	75	88	11	773
2-3	45	91	40	92	1	95
3-4	16	44	15	45	1	11
4-5	7	28	3	28	1	1
5-∞	3	23	3	23	1	1

Tabelle 8.2: Gesamtheit an Vorhersagen innerhalb eines Schwellwertbereiches (bezogen auf Beobachtungen, vgl. Abb. 8.3) für Winter (DJF) und Sommer (JJA) für aLMO- (Tages- und Stundensummen) und ECMWF-Vorhersagen (00 UTC-Läufe).

aus Kap. 5). Aus Abb. 8.3 kann man ablesen, dass der Gesamtanteil an guten Vorhersagen des aLMO im Winter bei 31.3 %, im Sommer nur bei 18.9 % liegt. Vergleicht man nun einzelne Schwellwertbereiche, so liegen die Anteile an guten Vorhersagen für $R_{obs}^* < 0.5$ mm unterhalb 12 %, mit Ausnahme des Bereichs $0.05 \leq R_{obs}^* < 0.1$ mm im Sommer, wobei dort die statistische Grundlage am schlechtesten ist (nur 6 Ereignisse). Für schwache Niederschläge gibt es also nur sehr wenige gute Vorhersagen. Dies kann damit erklärt werden, dass die Unterschiede zwischen Modell und Beobachtungen bei kleinen Niederschlagsmengen im Verhältnis schnell sehr groß werden können. Diese schwachen Ereignisse sind hauptsächlich die Ursache für die Häufung der Punkte bei $S \simeq 2$ (Abb. 8.2 (a)). Bestätigt wird diese Aussage mit zwei SAL-Diagrammen für verschiedene Intensitätsklassen in Anh. B.1. Diese Punkte sollten also nicht überbewertet werden, da sie meist nur bei schwachen Niederschlagsereignissen auftreten.

Für $R_{obs}^* \geq 0.5$ mm wird sichtbar, dass im Winter deutlich mehr gute Vorhersagen gemacht werden als im Sommer. Dieser Unterschied ist maximal für $1 \leq R_{obs}^* < 2$ mm (Winter: 51 %, was zugleich das Maximum ist; Sommer: 18 %). Das Maximum im Sommer für gute Vorhersagen liegt bei 29 % im Bereich $4 \leq R_{obs}^* < 5$ mm. Damit zeigt SAL (für aLMO-Vorhersagen), dass gerade im Sommer (aber auch im Winter) Niederschlagsereignisse mit höheren Intensitäten besser vorhergesagt werden können, als schwache Ereignisse.

SAL für verschiedene Flusseinzugsgebiete

Wie in der Einleitung (Kap. 1) erwähnt, liegt ein Anwendungsgebiet von SAL in der Hydrologie mit der Betrachtung von Flusseinzugsgebieten. Im Folgenden soll gezeigt werden, wie SAL statistisch für verschiedene (deutsche Teile der) Flusseinzugsgebiete aussieht und ob sich entscheidende Unterschiede zwischen verschiedenen

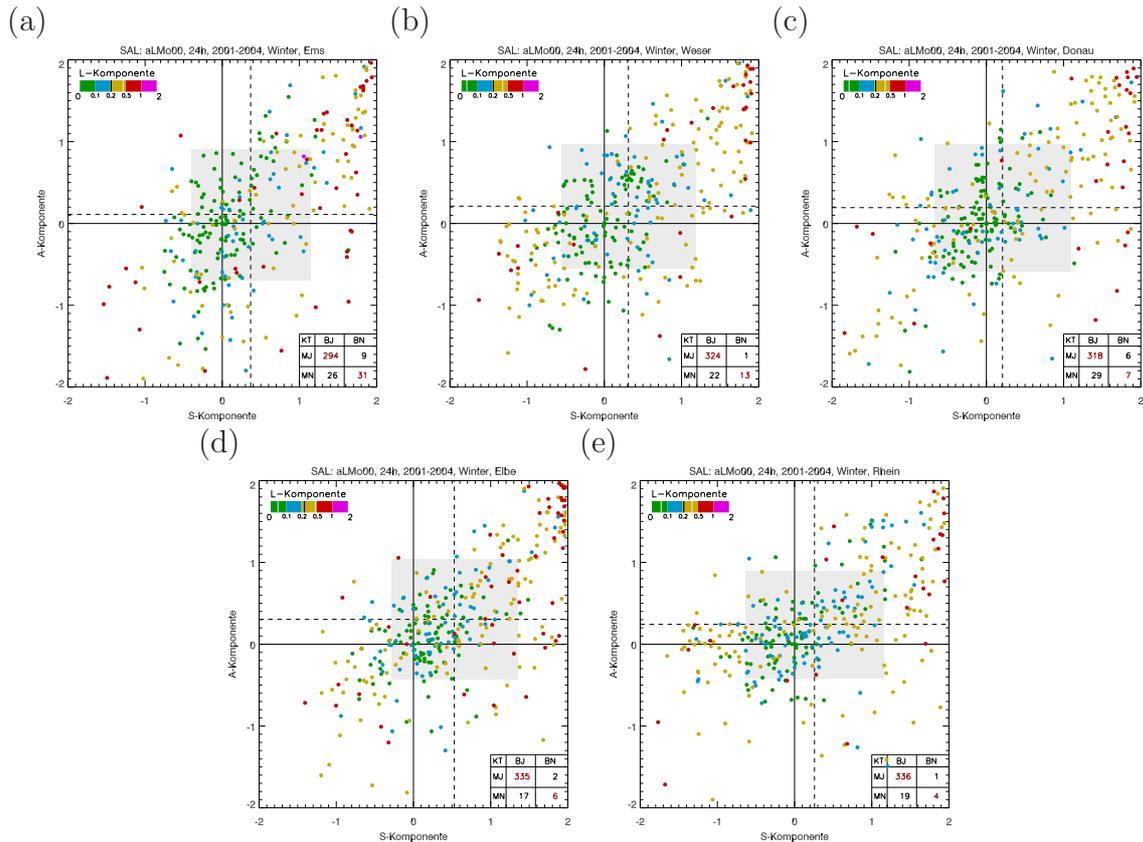


Abbildung 8.4: SAL für Tagessummen im Winter aus aLMO-Vorhersagen (00 UTC-Läufe) für verschiedene Einzugsgebiete (nach Fläche von klein nach groß sortiert): (a) Ems, (b) Weser, (c) Donau, (d) Elbe und (e) Rhein.

großen oder regional (topographisch) unterschiedlich gelegenen Gebieten zeigen. Abbildung 8.4 und 8.5 zeigen SAL für Winter und Sommer in den Einzugsgebieten Ems, Weser, Donau, Elbe und Rhein (von flächenklein nach -groß, vgl. Abschn. 2.7). Die Bilder zeigen alle die für SAL-Statistiken typische Charakteristik für aLMO-Vorhersagen. Zieht man nun die genauen Zahlen für die Mittelwerte und Standardabweichungen der drei Komponenten aus Tab. 8.1 hinzu, so folgt:

- \bar{S} und \bar{A} sind im Winter in allen Gebieten größer als im Sommer. Niederschlagsüberschätzungen gehen also meist einher mit zu großen Objekten im aLMO. Dies konnte mit herkömmlichen Fehlermaßen in Kap. 6 nicht gezeigt werden, dort ist im Mittel (Vergleich vor allem zur A-Komponente) eher der Niederschlag im Sommer als im Winter überschätzt,
- \bar{L} ist im Winter in allen Gebieten kleiner als im Sommer. Dies ist typisch für Verschiebungsfehler einzelner konvektiver Zellen im Gegensatz zu frontalen Ereignissen. (Hauptgrund für schlechtere Ergebnisse im Sommer aus gitterpunkt-basierten Fehlermaßen; vgl. auch Casati et al. (2004)),
- für die Streuung von S und L ist kein systematischer Unterschied zwischen Winter und Sommer erkennbar,

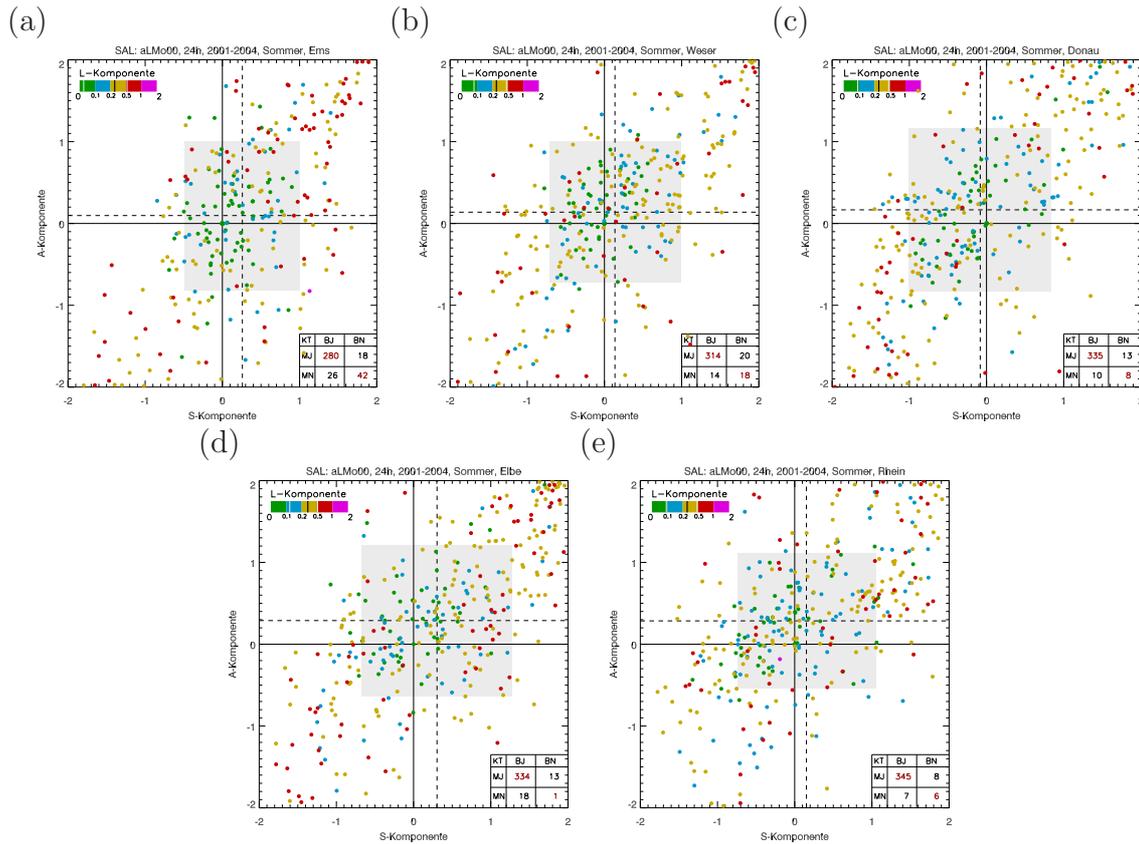


Abbildung 8.5: Wie Abb. 8.4 für den Sommer.

- die Streuung von A ist dagegen im Winter in allen Gebieten immer deutlich kleiner als im Sommer (höhere Intensitäten im Sommer machen automatisch auch größere Abweichungen möglich; vgl. mit höherer Variabilität bei höheren Schwellwerten aus Kap. 6),
- im Vergleich der Einzugsgebiete ist für \bar{S} kein klares Verhalten erkennbar, wobei für die Elbe der höchste Wert im Winter (0.53) und im Sommer (0.3) und für die Donau der einzige negative Wert im Sommer (-0.08) vorliegt (die S -Komponente zeigt also keine Abhängigkeit von der Gebietsgröße,
- ebenso zeigt auch \bar{L} keinen Zusammenhang mit der Größe der Gebiete bei Werten von 0.22 im Winter (Ausnahme wiederum Elbe mit 0.26) und um 0.28 im Sommer (Elbe mit 0.32),
- für \bar{A} zeigt sich im Sommer hingegen ein Zusammenhang in der Weise, dass der höchste Wert (0.29) in den beiden größten Gebieten (Elbe und Rhein) und der kleinste Wert (0.1) im kleinsten Gebiet (Ems) vorkommt (generelle Überschätzungen hatten sich auch im FBI (Kap. 6) gezeigt, dort allerdings für Häufigkeiten und deutlich nur für schwächere Niederschläge (vgl. Intensitätsabhängigkeit in Abb. 8.8 und 8.9 für Stundensummen)),
- in den Standardabweichungen aller drei Komponenten ist im Vergleich der Flusseinzugsgebiete kein Zusammenhang erkennbar,

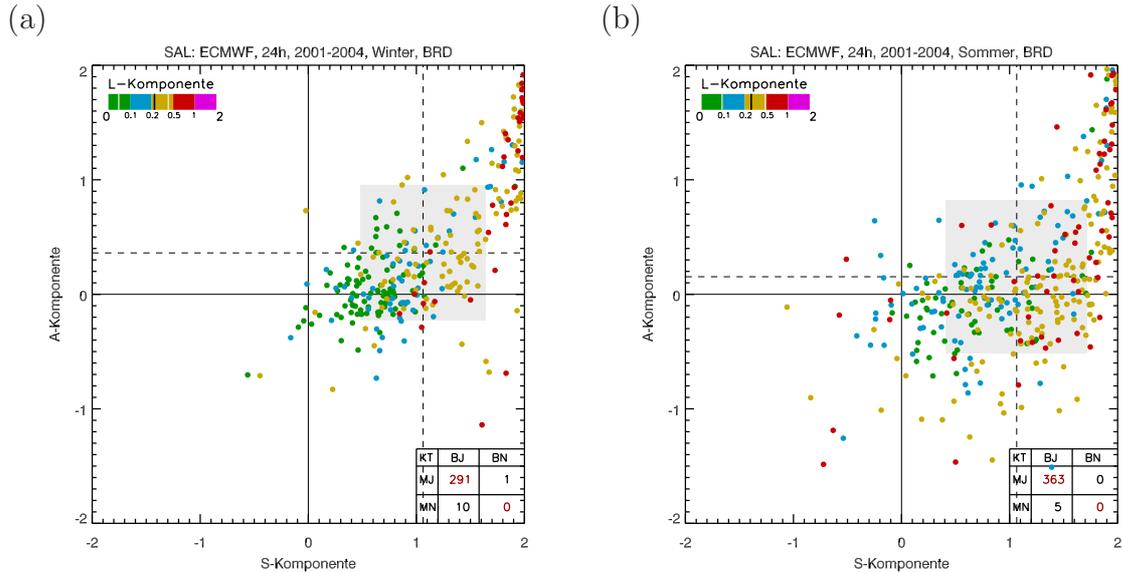


Abbildung 8.6: SAL für Tagessummen ECMWF-Vorhersagen für Deutschland: (a) Winter und (b) Sommer.

- die Anzahl an guten Vorhersagen ist in allen Gebieten im Winter deutlich besser als im Sommer (wie für ganz Deutschland in Abb. 8.3). Dabei liegen die Werte im Winter zwischen 19.4 (Donau) und 27 % (Elbe), im Sommer zwischen 11.6 % (Donau) und 19.6 % (Ems). Es ist auch hier keine Zusammenhang mit der Gebietsgröße erkennbar.

Aus diesen Ergebnissen folgt, dass die Größe des Einzugsgebietes keine entscheidende Rolle spielt, SAL also unabhängig von der Gebietswahl (zumindest in Deutschland) zu betrachten ist. Dieses Resultat findet sich auch in Kap. 6 für gitterpunkt-basierte Fehlermaße.

8.2 Vergleich zwischen aLMo und Globalmodell des ECMWF

In diesem Abschnitt sollen die aLMo- und ECMWF-Vorhersagen für Deutschland im Winter und Sommer mit Hilfe von SAL-Statistiken verglichen werden. Abbildung 8.6 zeigt SAL für die beiden Jahreszeiten aus Tagessummen von ECMWF-Vorhersagen. Verglichen mit den Bildern für aLMo-Vorhersagen (Abb. 8.2) ist ein deutlicher Unterschied erkennbar. Die gesamte Verteilung verschiebt sich nun für beide Jahreszeiten bis auf wenige Punkte in den positiven Bereich von S . Dies bedeutet, dass die Niederschlagsobjekte im Globalmodell deutlich zu groß sind oder zu wenig Struktur aufweisen (im Winter ist \bar{S} gegenüber dem aLMo um einen Faktor 6, im Sommer um einen Faktor 12 größer.). Die Erklärung hierfür ist einleuchtend und hängt mit der niedrigeren Auflösung im Globalmodell zusammen. Dieses Ergebnis für die S -Komponente ist nicht überraschend, zeigt allerdings die Nützlichkeit

von S , um Unterschiede zwischen Modellen mit verschiedenen Auflösungen zu identifizieren. Im Gegensatz dazu zeigen gitterpunkt-basierte Verifikationsmaße kaum Unterschiede zwischen den beiden Modellen (Kap. 6).

Unterschiede gibt es auch in der A -Komponente. \bar{A} des ECMWF-Modells ist wie beim aLMo in beiden Jahreszeiten positiv, liegt jedoch im Winter oberhalb, im Sommer unterhalb des aLMo-Mittels. Die Streuung für S und A ist deutlich geringer als im aLMo, was aber noch klarer zeigt, dass die im Schnitt schlechteren Vorhersagen des Globalmodells auch im Rahmen der Standardabweichung keine bessere Interpretation zulassen. Für die L -Komponente finden sich im Mittel nur geringe Unterschiede bei gleicher Streuung, allerdings fällt auf, dass es im Winter deutlich mehr gute L -Werte (grüne und blaue Punkte) gibt, als im Sommer. Die geringere Auflösung führt zu Vorhersagen flächigeren Niederschlages (und damit im Mittel eher mit für den Winter typischen Niederschlagsverteilungen). Wenn in weiten Teilen gleichmäßig starker (oder schwacher) Niederschlag fällt, so liegt der Schwerpunkt eher Richtung Mitte des Gebietes und die Verschiebung gegenüber den Beobachtungen kann nicht besonders deutlich ausfallen. Der Mittelwert für L liegt aber nicht etwa unterhalb dem des aLMo, sondern ist identisch (Sommer) bzw. leicht erhöht (Winter). Für diese (gewünschte) Tatsache ist vor allem die Teilkomponente L_2 (Gl. 7.4) verantwortlich.

Auch der Vergleich der Anzahl an guten Vorhersagen im Sinne von SAL zeigt, dass die Vorhersagequalität für den Niederschlag im aLMo besser als im ECMWF-Globalmodell ist (Tab. 8.2 und Abb. 8.3). Die Anzahlen an Vorhersagen innerhalb eines Schwellwertbereiches sind im Sommer fast identisch mit denen des aLMo, im Winter insgesamt weniger, da in den ersten drei Monaten des Jahres 2001 keine Daten des Globalmodells vorhanden sind (Abschn. 2.2). Im Winter gibt es nur noch halb so viele gute Vorhersagen im Globalmodell als im aLMo, mit 16.5% sind es sogar weniger als im Sommer für das aLMo. Im Sommer sind gute ECMWF-Vorhersagen ebenfalls seltener als im Winter mit nur noch 11.8%. Der Verlauf in Abhängigkeit vom Schwellwertbereich (Abb. 8.3) ist ebenfalls unterschiedlich zu dem des aLMo. Im Winter gibt es zwar ein Maximum ($2 \leq R_{obs}^* < 3$ mm) im ähnlichen Bereich wie beim aLMo, dieses fällt aber nicht sehr deutlich aus (die beiden oberen Klassen sollten nicht weiter interpretiert werden, da dort die Statistik mit je drei Vorhersagen schlecht ist). Im Sommer unterliegt der Verlauf im Globalmodell deutlich mehr Schwankungen zwischen den Schwellwertbereichen als im aLMo, einzig der Wert für den Bereich $3 \leq R_{obs}^* < 4$ mm zeigt fast identische Zahlen um 20%. (Vgl. auch SAL-Diagramme für das Globalmodell aufgeteilt in verschiedene Intensitätsklassen (Anh. B.2))

8.3 Vergleich zwischen aLMo und LM

In diesem Abschnitt soll ein kurzer Vergleich zwischen aLMo und LM für Deutschland im Jahr 2004 gemacht werden. Dafür ist in Abb. 8.7 SAL für Tagessummen der beiden Modelle für Winter und Sommer dargestellt.

Im Winter fällt auf, dass \bar{A} im LM eine größere Überschätzung anzeigt (ca. Faktor

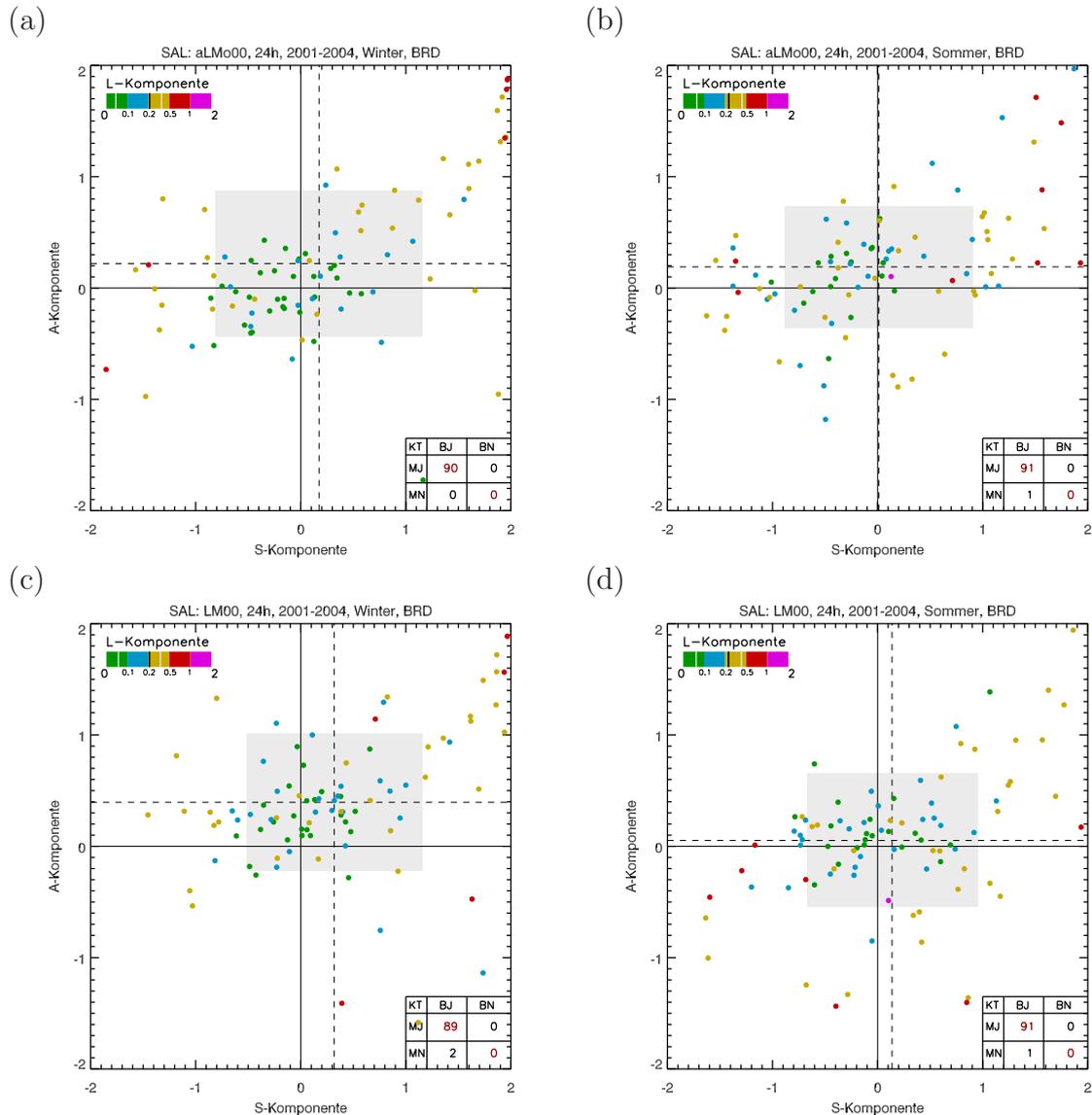


Abbildung 8.7: SAL für Tagessummen aus aLMo- (oben) und LM-Vorhersagen (unten) für Deutschland 2004 (jeweils 00 UTC-Läufe): Winter (links) und Sommer (rechts).

2, genaue Werte in Tab. 8.1). Dies konnte bereits in Abschn. 4.4 anhand der mittleren Niederschlagssummen gezeigt werden. Auch für \bar{S} zeigt sich ein ähnliches Bild, was die Tatsache bestätigt, dass Überschätzungen meist mit dem Problem zu großer Objekte im Modell zusammenfallen. Die Streuung für S und A ist im LM allerdings etwas geringer. Für die L -Komponente finden sich kaum Unterschiede.

Im Sommer zeigt sich für die mittlere Amplitude ein besserer Wert im LM. Dies geht einher mit einem gegenüber dem aLMo leicht erhöhten mittleren Strukturfehler. Genau gegenläufig dazu verhalten sich die entsprechenden Standardabweichungen. Dies könnte seine Ursache in der Einführung des prognostischen Niederschlags im April 2004 haben (Hinweis: im aLMo wurde diese Änderung erst im November 2004 durchgeführt, es zeigen sich auch kaum Unterschiede in \bar{A} für das aLMo zwischen Winter und Sommer; vgl. Abschn. 4.5). Für die L -Komponente finden sich auch hier kaum Unterschiede.

Die Anzahl an guten Vorhersagen beträgt im Jahr 2004 für das aLMo 36.7% im

Winter und 23.1% im Sommer. Diese Werte sind etwas höher als im Mittel über 2001-2004. Im LM ist zum einen der Wert im Winter etwas niedriger (34.8%), zum anderen im Sommer deutlich besser (30.8%). Dies kann wiederum durch die frühere Einführung des prognostischen Niederschlags in das LM begründet werden.

SAL ist also in der Lage, die in Kap. 4 gefundenen Unterschiede zwischen aLMo und LM anhand weniger Bilder zu identifizieren. Sicherlich ist dabei zu bedenken, dass die Interpretation der Unterschiede in diesem Abschnitt auf dem Wissen der zuvor gefundenen Ergebnisse basiert. Jedoch kann man daraus lernen, dass die Verifikation alleine mit SAL solche Schlüsse zulässt.

8.4 SAL für Stundensummen: Verifikation des aLMo

SAL soll nun auch auf stündlicher Zeitskala angewandt werden. Dazu wird der disaggregierte Datensatz herangezogen (Kap. 3).

Zunächst soll ein kurzer Vergleich der Anteile an guten Vorhersagen im Sinne von SAL (Abschn. 8.1) zwischen Tages- und Stundensummen für ganz Deutschland gemacht werden. Dazu wurden in Tab. 8.2 die Gesamtanzahlen an Vorhersagen für Schwellwertbereiche auf stündlicher Basis angefügt. Die Anzahlen sind verständlicherweise deutlich höher für die unteren Bereiche, ebenso ergeben sich im Sommer höhere Anzahlen für höhere Schwellwertbereiche, wobei die statistische Grundlage im Winter bei $R_{obs}^* > 2\text{ mm}$ und im Sommer bei $R_{obs}^* > 3\text{ mm}$ nicht mehr vorhanden ist. Die Anteile an guten Vorhersagen sind in Abb. 8.3 (rechts) aufgeführt und können somit direkt mit den Tagessummen (links) verglichen werden. Im Winter (15.6%) wie Sommer (5.4%) liegen die Anteile für Stundensummen insgesamt deutlich unterhalb denen für Tagessummen. Auch der Verlauf in Abhängigkeit der Schwellwertbereiche ist zwar sehr ähnlich, jedoch mit deutlich niedrigeren Werten im Winter bei maximal 21% und im Sommer bei 6-9% (mit Beachtung der statistischen Grundlage sollte eher der niedrigere Wert gewählt werden). Zudem ist der Verlauf hin zu kleineren Schwellwertbereichen verschoben. Diese Unterschiede kommen daher, dass es auf stündlicher Zeitskala sehr viel schwieriger ist, den exakten Zeitpunkt für ein Niederschlagsereignis vorherzusagen. Schon eine Verschiebung um eine Stunde gegenüber den Beobachtungen führt dazu, dass eine eigentlich gute Vorhersage als schlecht erachtet wird. Mit Blick auf die Grenzen der Vorhersagbarkeit insbesondere konvektiver Zellen im Sommerhalbjahr (Kap. 1) ist die Verifikation auf dieser Zeitskala ohne Zulassen zeitlich benachbarter Vorhersagen auch nicht besonders fair.

Saisonaler Vergleich für verschiedene Flusseinzugsgebiete

Tabelle 8.3 zeigt die Mittelwerte und Standardabweichungen aller drei Komponenten von SAL für Stundensummen und verschiedene Flusseinzugsgebiete (Winter und Sommer). Ein Vergleich mit den Werten für Tagessummen (Tab. 8.1) zeigt, dass zwar die Werte teilweise Unterschiede zeigen, sich jedoch wiederum kein systematischer

aLMO EZG	\bar{S}		σ_S		\bar{A}		σ_A		\bar{L}		σ_L	
	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA
Ems	13	14	63	67	3	5	74	84	20	21	27	28
Weser	10	12	76	82	2	5	77	96	20	22	21	23
Donau	11	-2	77	87	3	3	75	101	19	26	20	23
Elbe	30	19	81	86	8	10	80	101	23	27	22	24
Rhein	12	8	80	87	7	7	73	97	20	24	18	21

Tabelle 8.3: Mittel und Standardabweichung (mit 100 multipliziert) für S , A und L für Stundensummen aus aLMO-Vorhersagen (00 UTC-Läufe) für Winter (DJF) und Sommer (JJA) und verschiedene Flusseinzugsgebiete (EZG): Ems, Weser, Donau, Elbe und Rhein.

Zusammenhang mit der Gebietsgröße ergibt.

Aus diesem Grund soll sich die Auswertung für Stundensummen in den folgenden beiden Teilabschnitten auf das deutsche Gebiet der Elbe beschränken.

Verschiedene Schwellwertbereiche

Wie für Tagessummen wird zunächst ein saisonaler Vergleich für die Elbe zwischen Winter und Sommer gemacht. Dabei wurde SAL für die Stundensummen bereits in vier Schwellwertbereiche aufgeteilt: $0 < R_{obs}^* < 0.05$ mm, $0.05 \leq R_{obs}^* < 0.1$ mm, $0.1 \leq R_{obs}^* < 0.5$ mm und $0.5 \text{ mm} \leq R_{obs}^* < \infty$. In den Bildern für Winter (Abb. 8.8) und Sommer (Abb. 8.9) sieht man, dass die typische Charakteristik der SAL-Statistik für das aLMO (Häufung der Punkte entlang der positiv geneigten Diagonalen, Abschn. 8.1) auch für Stundensummen gilt.

Im Winter (Abb. 8.8) zeigt sich, dass die Verteilung mit zunehmendem Schwellwert im Mittel vor allem von positiven zu negativen A -Werten wandert, aber auch S von positiv Richtung null geht (für große Schwellwerte ist die statistische Grundlage im Winter nicht ausreichend). Dies hängt unter anderem mit einer starken Häufung (ähnlich wie bei den Tagessummen) bei $S > 1.5$ und $A > 1.5$ für kleine Schwellwerte zusammen. Dies bestätigt die für Tagessummen gefundene Aussage, dass diese massiven Überschätzungen von S und A meist für geringe Niederschläge resultieren. Im Sommer (Abb. 8.9) findet sich ein sehr ähnlicher Verlauf mit dem Unterschied, dass sich die Vorhersagen in der Anzahl auch statistisch signifikant in den höchsten Schwellwertbereich verteilen. Für die A -Komponente ist die Streuung im Sommer deutlich größer als im Winter. Auch findet man eine Häufung von Punkten bei $S < -1.5$ und $A < -1.5$ für Schwellwerte oberhalb 0.1 mm. Dies sind Fälle für relativ intensive Ereignisse in den Beobachtungen, welche vom Modell in dieser Menge und Struktur (konvektive Zellen zu klein und peakig) nicht vorhergesagt wurden. Für die L -Komponente lassen sich kaum Unterschiede finden. Im Mittel liegen die Werte ähnlich für Winter und Sommer um 0.25 und damit etwas oberhalb der Werte für Tagessummen. Dies zeigt, dass die Vorhersage der Niederschlagsverteilung über längere Zeiträume (Tagessummen) besser ist, als für relativ kurze Zeiträume (Stundensummen).

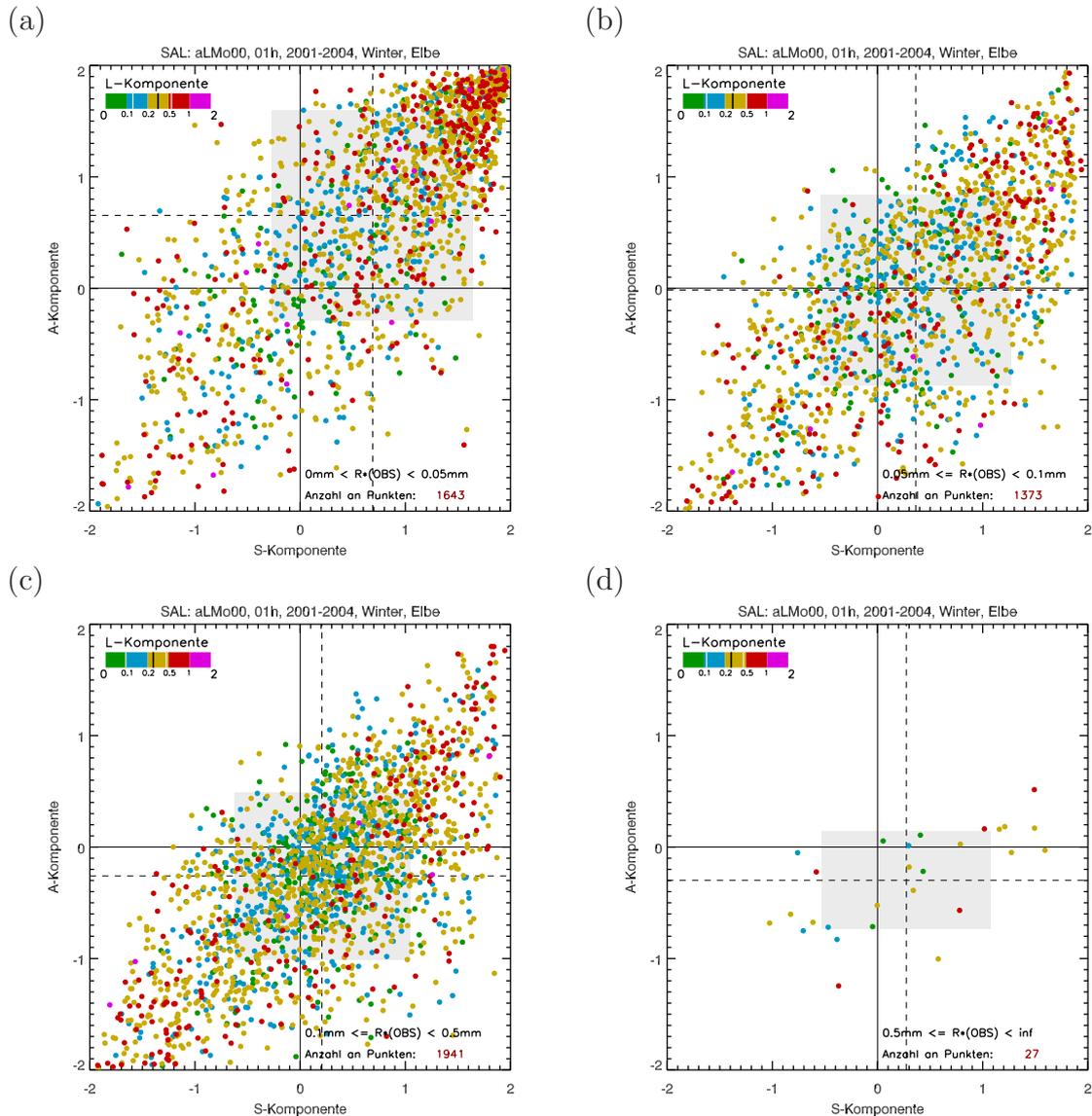


Abbildung 8.8: SAL für Stundensummen aus aLMO-Vorhersagen (00 UTC-Läufe) für die Elbe im Winter aufgeteilt in verschiedene Schwellwertbereiche (ausgehend von den Beobachtungen: $R_{obs}^* = R_{obs}^{max}/15$): (a) $0 < R_{obs}^* \leq 0.05 \text{ mm}$, (b) $0.05 < R_{obs}^* \leq 0.1 \text{ mm}$, (c) $0.1 < R_{obs}^* \leq 0.5 \text{ mm}$ und (d) $0.5 \text{ mm} < R_{obs}^* < \infty$.

Abhängigkeit von der vorhergesagten Tageszeit

Als letzter Punkt soll eine Untersuchung auf die Qualität von aLMO-Vorhersagen hinsichtlich der Vorhersagezeit gemacht werden. Dazu wurden wiederum die 00 UTC-Läufe und die Vorhersagezeit von 6-30 h (bzw. für den Winter 6-33 h) herangezogen. Da nur 00 UTC-Läufe betrachtet werden, entsprechen unterschiedliche Vorhersagezeiten automatisch auch unterschiedlichen Tageszeiten. Es wurde weiter der deutsche Teil der Elbe für Winter und Sommer betrachtet und entschieden, immer drei aufeinander folgende Stunden zu einer SAL-Statistik zusammenzufassen.

Abbildung 8.10 zeigt SAL für Winter und die Vorhersagezeit zwischen 6-9 h (a) und 30-33 h (b) (um mögliche Überlagerungen tageszeitlicher Unterschiede zu vermeiden, wurde für den Winter der Zeitraum exakt 24 h später herangezogen). Die

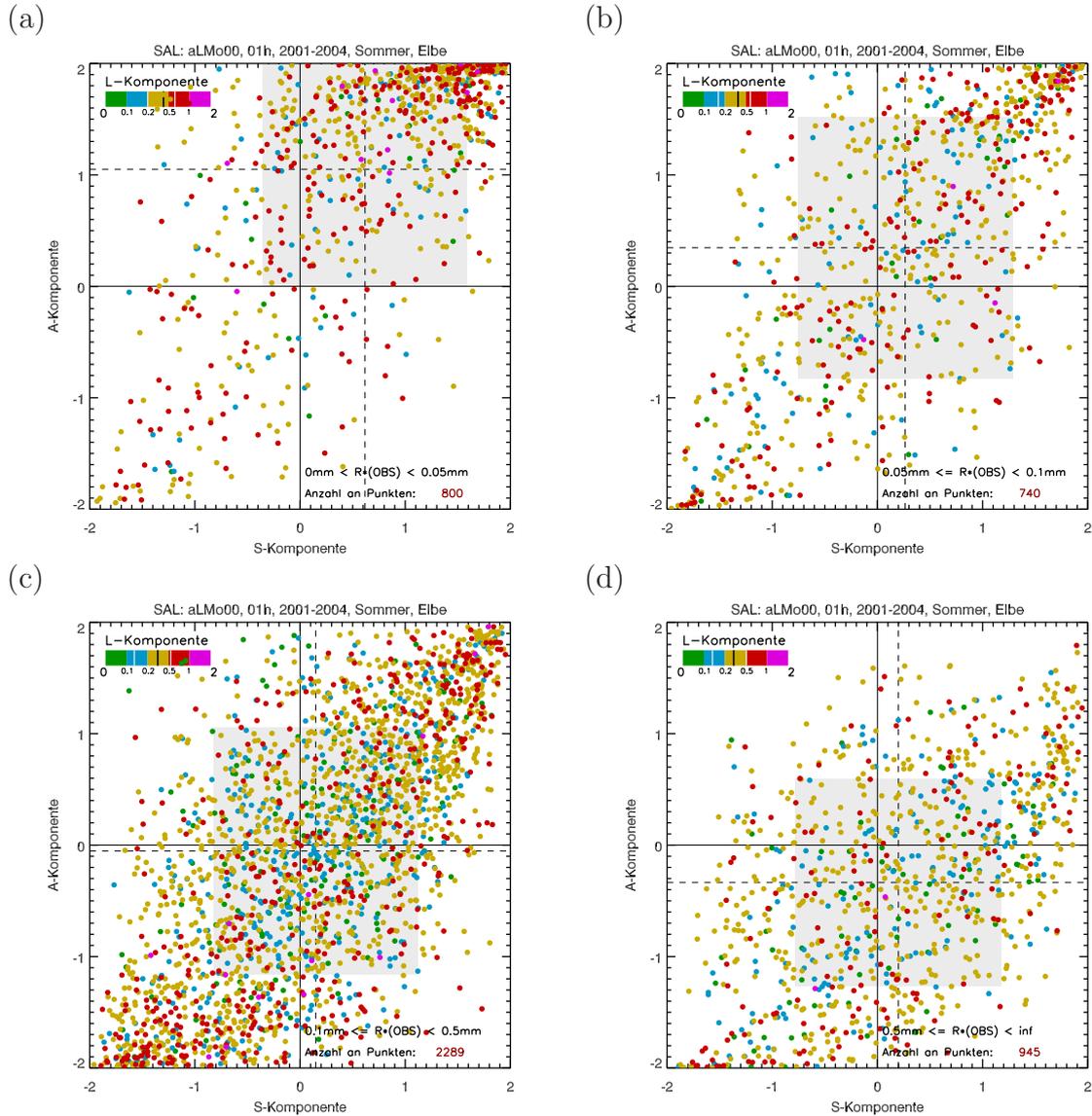


Abbildung 8.9: Wie Abb. 8.8 für Sommer.

Abhängigkeit von der Vorhersagezeit zeigt sich an der Streuung der beiden Bilder. Zu den späteren Zeitpunkten (30-33 h) ist vor allem die Standardabweichung in A angewachsen. Dies geht einher mit einer gestiegenen mittleren Überschätzung. Die L -Komponente zeigt kaum eine Veränderung. Die Unterschiede vor allem für A werden erst bei einem Zeitabstand von etwa 24 h Vorhersagezeit deutlich sichtbar. Das fällt auch auf, wenn man die Anzahl an guten Vorhersagen betrachtet: um 6-9 h liegt diese bei 13.8 % und sinkt allmählich auf unter 9.6 % (18-21 h). Um 30-33 h ist die Anzahl guter Vorhersagen schon fast auf nur noch die Hälfte gesunken (7.6 %).

Für den Sommer werden alle acht Bilder (Abb. 8.11) der 3 Stunden-Abschnitte gezeigt. Es ist sehr schön die Problematik bei der Simulation des Tagesgangs im aLMo sichtbar. Vergleicht man in diesen SAL-Diagrammen \bar{A} mit dem Tagesgang aus Abb. 5.12 (unten), so ist der Verlauf von Über- (mittags) und Unterschätzungen (abends) direkt nachvollziehbar. Für S zeigt sich fast der analoge Verlauf, was wiederum bestätigt, dass S und A meist das gleiche Vorzeichen haben (Abschn. 8.1). Die

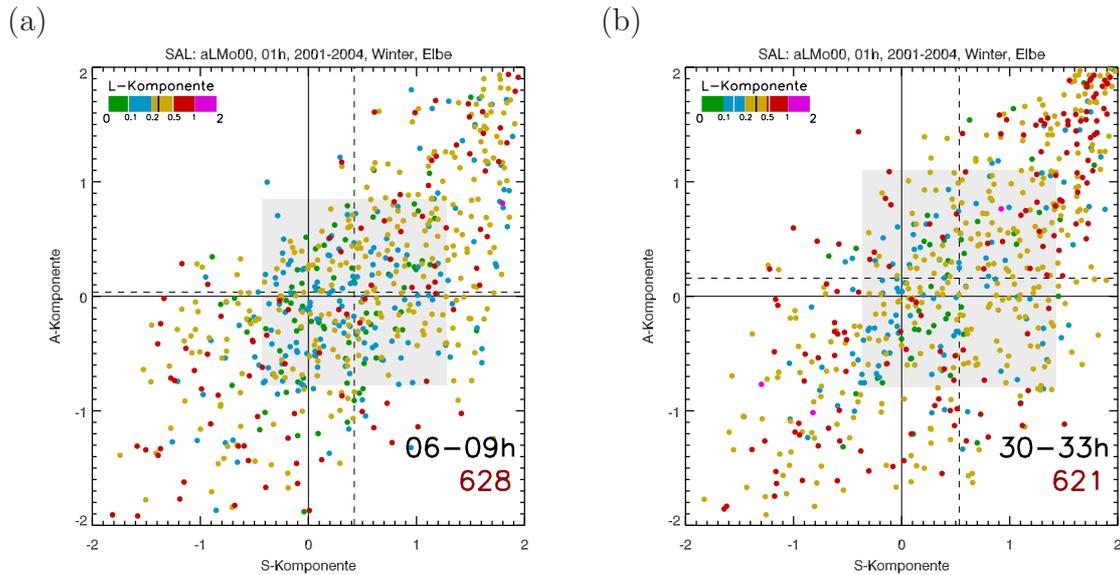


Abbildung 8.10: SAL in Abhängigkeit von der Vorhersagezeit für stündliche aLMO-Vorhersagen (00 UTC-Läufe) im Winter für die Elbe: (a) 6-9 h, (b) 30-33 h.

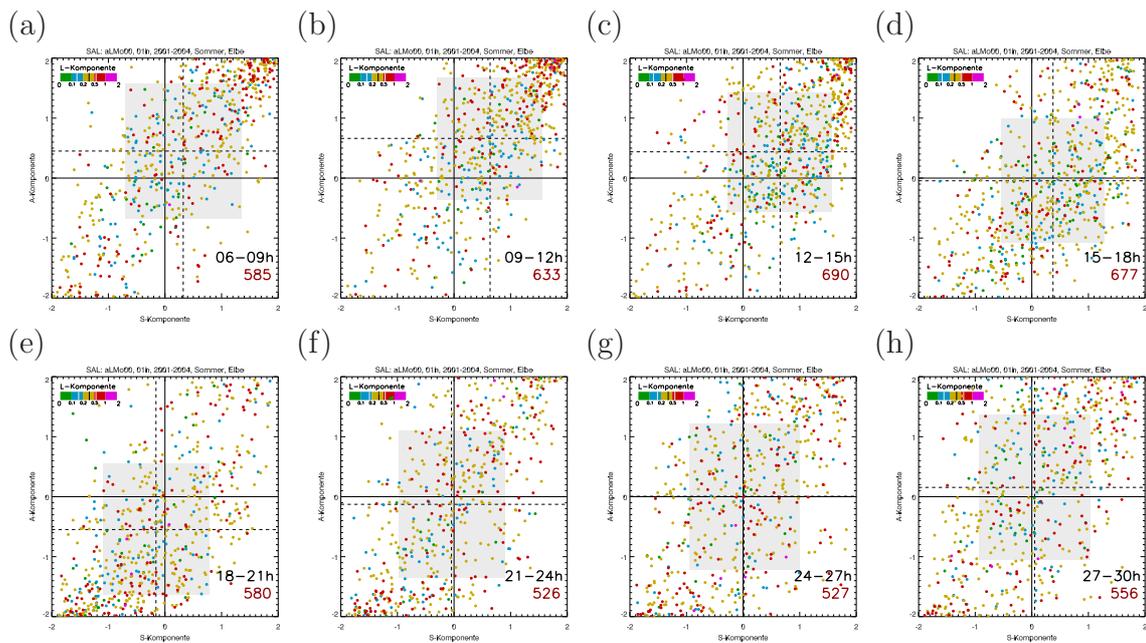


Abbildung 8.11: SAL in Abhängigkeit von der Vorhersagezeit für stündliche aLMO-Vorhersagen (00 UTC-Läufe) im Sommer für die Elbe: (a) 6-9 h, (b) 9-12 h, (c) 12-15 h, (d) 15-18 h, (e) 18-21 h, (f) 21-24 h, (g) 24-27 h, (h) 27-30 h.

Bilder wurden weniger zur Diskussion der L -Komponente ausgewählt, die sowohl im Mittel, als auch ihrer Standardabweichung kaum Unterschiede mit der Vorhersagezeit (Tageszeit) anzeigt.

Auffallend sind wieder Häufungen von Punkten anfangs bei $S > 1.5$ und $A > 1.5$ und später am Tag bei $S < -1.5$ und $A < -1.5$. Diese Zusammenhänge zeigten sich bereits in den Bildern zu Schwellwertabhängigkeiten (Abb. 8.8). Hier kann man nun erkennen, dass diese Probleme des aLMO mit verschiedenen vorhergesagten Intensitäten auf stündlicher Zeitskala mit der Problematik der Tagesgangsimulation

zusammenhängen.

Der Verlauf der Anzahl an guten Vorhersagen in Abhängigkeit von der Vorhersagezeit zeigt einen ähnlichen Zusammenhang wie im Winter (Abnahme mit der Vorhersagedauer). Die Werte sind jedoch wiederum deutlich niedriger bei 13.2% um 6-9 h und nur noch 2.5% um 27-30 h. Hier gibt es allerdings eine Überlagerung durch den Tagesgangfehler, da die Anzahl an guten Vorhersagen zwischenzeitlich auf 1.4% abfällt (18-21 h).

8.5 Kurzes Fazit

SAL wurde entwickelt (Kap. 7), um einen weiteren Blickwinkel auf die Qualität von Niederschlagsvorhersagen bieten zu können. In diesem Kapitel konnte anhand quantitativer, statistisch signifikanter Auswertungen mit SAL gezeigt werden, dass viele Modell-Eigenschaften und damit zum Teil auch -Defizite sichtbar werden (z.B. verschiedene Niederschlagsstrukturen begründet durch unterschiedliche Modellauflösungen, Tagesgang, usw.). SAL kann im Gegensatz zu gitterpunkt-basierten Fehlermassen viel detailliertere Aussagen liefern und die vorhandenen typischen Modellfehler besser charakterisieren.

SAL wird daher als sehr nützliches neues Qualitätsmaß auch für die statistische Verifikation von Niederschlagsvorhersagen von globalen und regionalen Modellen angesehen.

Kapitel 9

Zusammenfassung und Ausblick

In diesem Kapitel werden die wichtigsten Ergebnisse dieser Arbeit zusammengefasst. Weiter folgt ein Ausblick, der offen gebliebene Fragen anspricht und Ideen für anschließende Arbeiten liefert.

9.1 Zusammenfassung

In dieser Arbeit wurde die Niederschlagsvorhersage für Deutschland anhand der Zeitreihe 2001-2004 verifiziert. Dabei wurden Vorhersagen des (COSMO-)aLMO (für 2004 zusätzlich auch des LM) und des ECMWF-Globalmodells betrachtet. Die Verifikation wurde einerseits mit Tagessummen aus gegitterten Stationsdaten des DWD, andererseits mit einem für Deutschland neuartigen, zeitlich und räumlich hochaufgelösten Datensatz durchgeführt. Dieser disaggregierte Datensatz wurde aus der Kombination gegitterter Stations- und Radardaten erzeugt (Kap. 3) und detailliert auf Fehlerquellen untersucht. Der Datensatz ist sehr nützlich für die Verifikation der QNV auf der stündlichen Zeitskala und erlaubt flächendeckende Untersuchungen zum Tagesgang und zur Intensitätsverteilung des Niederschlags.

Zunächst wurde die mittlere Verteilung der Niederschlagsmengen in Deutschland in allen Datensätzen betrachtet. Dabei stellte sich ein erstes typisches Problem der aLMO-Vorhersagen heraus: in den Mittelgebirgsregionen traten bänderartige Strukturen von Überschätzungen im Luv und Unterschätzungen im Lee auf (Kap. 4). Des Weiteren konnte mit dem disaggregierten Datensatz ein zweites systematisches Problem bei der Simulation des Tagesgangs im Sommer erkannt werden. Das aLMO überschätzt den Tagesgang, der Niederschlag tritt zu früh ein und hört in den Abendstunden oft zu früh auf (Kap. 5). Hier konnte auch gezeigt werden, dass das aLMO generell (Winter und Sommer) zu viele relativ schwache Niederschläge vorhersagt. Diese drei typischen Modellschwächen wurden zeitgleich zu dieser Arbeit auch innerhalb des COSMO (2006) beschrieben und sind bis heute teilweise verbessert (der prognostische Niederschlag schwächt das Luv/Lee-Problem ab), jedoch nicht vollständig behoben worden.

Zur Motivation, einen neuen Ansatz für die Verifikation der QNV zu entwickeln, wurde ein Kapitel über die Verifikation mit herkömmlichen Standardmaßen eingefügt

(Kap. 6). Diese Ergebnisse lassen weiter auch einen Vergleich mit der neuen Methode zu. Dabei wurden einige gitterpunkt-basierte kontinuierliche und kategorische Fehlermaße ausgewählt und die Zeitreihe von 2001-2004 auf täglicher und stündlicher Zeitskala für verschiedene Einzugsgebiete ausgewertet. Mehrere Eigenschaften der betrachteten Niederschlagsvorhersagen konnten hiermit bestätigt oder neu erkannt werden:

- Im Mittel gibt es einen Jahresgang der Vorhersagequalität im aLMO und ECMWF-Globalmodell (Sommer schlechter als Winter).
- Bei der Betrachtung verschiedener Flusseinzugsgebiete wurden keine starken regionalen Unterschiede und keine Abhängigkeit von der Gebietsgröße gefunden.
- Das aLMO zeigt gegenüber dem ECMWF-Globalmodell keine Verbesserung in der QNV.

Besonders letzteres Ergebnis soll dabei mit Vorsicht betrachtet werden. In der Einleitung zu dieser Arbeit wurde bereits beschrieben, dass dieses Ergebnis durch das Double-Penalty-Problem bei der gitterpunkt-basierten Verifikation begründet ist: kleinskalige Phänomene (wie z.B. Niederschlag aus einzelnen konvektiven Zellen) werden nicht fair behandelt, da schon eine leichte Verschiebung des Vorhersagefeldes gegenüber den Beobachtungen maximale Fehler bringt.

Aus dieser Tatsache heraus motivieren sich viele neuartige Ansätze zur Niederschlagsverifikation. Ein weiterer Grund ist, dass eine einzige Zahl (aus Standardmaßen) nicht alle wichtigen Charakteristiken (z.B. Struktur, Verschiebung) des Fehlers einer Niederschlagsvorhersage beschreiben kann. Zusätzlich wurde in der Einleitung aufgezeigt, dass die Grenze der Vorhersagbarkeit für kleinskalige Prozesse schnell erreicht wird und es auch aus diesem Grund sinnvoller sein könnte, solche Prozesse nicht punktgenau, sondern lediglich innerhalb eines Gebietes zu verifizieren.

Aus diesen Gründen erschien es sinnvoll, sich mit der Problematik der Verifikationsmethodik der QNV zu beschäftigen und einen alternativen Ansatz zu entwickeln. Dieses neue Qualitätsmaß wurde mit SAL benannt und ausführlich in Kap. 7 eingeführt. SAL wurde auf künstlich erzeugte Niederschlagsfelder angewendet und ausführlich getestet. Die grundlegenden Eigenschaften von SAL wurden bereits in Kap. 1 aufgelistet:

- (i) SAL macht Aussagen über die Qualität einzelner Niederschlagsvorhersagen in einem zuvor definierten Gebiet.
- (ii) Es ist nahe an einer subjektiven Einschätzung der Qualität von Niederschlagsvorhersagen.
- (iii) SAL setzt sich aus drei unabhängigen Komponenten zusammen (S , A und L).

- (iv) Berücksichtigt und spezifiziert (im Gegenteil zu Ebert und McBride, 2000) wird auch die Struktur einzelner Niederschlagsereignisse (z.B. konvektive Zellen, frontale Niederschlagsbänder), welche als direktes Maß für die zu Grunde liegenden, physikalischen Eigenschaften gilt.
- (v) SAL benötigt keine Zuordnung einzelner Niederschlagsobjekte zwischen Modellvorhersage und Beobachtungen.

Mit diesen Eigenschaften ist SAL ein sehr nützliches Instrument für die Verifikation der Niederschlagsvorhersage. Die statistische Anwendung von SAL für die QNV von aLMo, LM und ECMWF-Globalmodell brachte folgende wichtige Ergebnisse (vgl. Kap. 8):

- Es wurde ein klarer Zusammenhang zwischen der S - und der A -Komponente in der Art gefunden, dass beide häufig das gleiche Vorzeichen haben, was z.B. für $S > 0$ und $A > 0$ bedeutet, dass Niederschlagsüberschätzungen mit zu großen und/oder abgeflachten Objekten zusammenhängen.
- Die Vorhersagen sind im Mittel im Winter besser als im Sommer. Dies gilt sowohl auf der täglichen als auch der stündlichen Zeitskala.
- Das aLMo liefert viele fehlerhafte Vorhersagen für schwächere Niederschlagsereignisse. Untersuchungen zu verschiedenen Intensitäten haben gezeigt, dass das aLMo für schwache Niederschläge weniger gute Vorhersagen liefert, als für relativ starke Niederschläge.
- Es besteht keine klare Abhängigkeit zur gewählten Einzugsgebietsgröße (in Deutschland).
- Das Globalmodell des ECMWF zeigt große Fehler in der S -Komponente. Die Objekte sind systematisch zu groß bzw. flächig. Dies ist aufgrund der niedrigeren Auflösung gegenüber dem aLMo nicht überraschend, konnte aber bisher mit herkömmlichen Fehlermaßen nicht gezeigt werden.
- Zwischen aLMo und LM finden sich Unterschiede für 2004, obwohl die Modellphysik grundsätzlich fast identisch ist. Dies liegt zum einen am unterschiedlichen Zeitpunkt der Einführung des prognostischen Niederschlags, zum anderen wird das aLMo mit Randdaten aus dem ECMWF-Globalmodell, das LM mit GME-Vorhersagen angetrieben.
- Auf der stündlichen Zeitskala finden sich deutlich weniger gute Vorhersagen, als auf der täglichen Zeitskala. Dies ist verständlich, da eine Zeitverschiebung eines eigentlich gut vorhergesagten Ereignisses als fehlerhafte Vorhersage gewertet wird.
- Die Qualität der QNV nimmt mit der Vorhersagezeit ab.

- Im Sommer ist in den SAL-Diagrammen sehr gut das Tagesgang-Problem zu erkennen (vor allem in A , aber auch in S).

SAL kann folglich alle bisher bekannten Schwächen des aLMo erkennen und liefert darüber hinaus Einblick in die Charakteristik der Fehler. Insbesondere ist eine klare Verbesserung der QNV im aLMo gegenüber dem Globalmodell des ECMWF sichtbar. Es erscheint sinnvoll, die Niederschlagsvorhersage gebietsweise zu betrachten (im Gegensatz zu gitterpunktweise), um eine innerhalb des Gebiets um wenige Gitterpunkte verschobene, subjektiv als gut eingeschätzte Vorhersage auch als gut zu bewerten. Wie schon in Kap. 7 beschrieben, kann SAL die wichtigsten Fragen zur Qualität einer Niederschlagsvorhersage beantworten:

S: Ist die Niederschlagsstruktur gut wiedergegeben? (z.B. kleinskalige, peakige konvektive oder flächenhafte, stratiforme Ereignisse)

A: Stimmt die Niederschlagsmenge im Gebiet? (z.B. wichtig für Hydrologen)

L: Ist das Niederschlagsereignis am richtigen Ort vorhergesagt? (z.B. Gefährdung bestimmter Gebiete durch Starkniederschläge, Planung von Veranstaltungen)

SAL liefert wichtige Hinweise auf typische, systematische Modellfehler. Dieses Kenntnis der Modelldefizite sollten durch zusätzliche Interpretation der Modelldaten in die Wettervorhersage einfließen.

9.2 Ausblick

Die Entwicklung von numerischen Wettervorhersagemodellen geht in den letzten Jahren hin zu Auflösungen, die in der Horizontalen kleiner als 3 km sind. Das COSMO-DE (Baldauf et al., 2006) ist solch ein hoch-auflösendes Modell und wird seit 2007 operationell beim DWD eingesetzt. Dabei wird hochreichende Feuchtekonvektion nicht mehr parametrisiert, sondern explizit berechnet.

Diese neue Modellversion konnte im Rahmen dieser Arbeit noch nicht verifiziert werden. Es wird aber interessant sein, gerade auch im Hinblick auf die sehr hohe Auflösung und das Ausschalten der Feuchtekonvektion, das COSMO-DE mit dem aLMo oder der neueren Version COSMO-EU zu vergleichen. Möglich erscheint, dass die Struktur einzelner Niederschlagsereignisse noch besser wiedergegeben wird und die S -Komponente dies dann zeigen sollte. Zudem sollte einer Weiterentwicklung der SAL-Komponenten offen gegenüber gestanden werden. Einzelkomponenten wie etwa L könnten mit anderen Ansätzen (Venugopal et al., 2005) verglichen und gegebenenfalls ersetzt werden.

Ein Schwachpunkt der SAL-Methode ist, dass im Gegensatz zu Davis et al. (2006a,b) keine Aussage über die Orientierung von Objekten gemacht wird. Dies führt dazu, dass im Sinne der S -Komponente die idealisierte Vorhersage aus Abb. 1.4 (d) genauso gut bewertet wird wie diejenigen in den Bildern (a) und (b). Dies entspricht aber nicht unbedingt der subjektiven Einschätzung, welcher SAL möglichst

nahe kommen will. Weitere Überlegung sollten die Wahl des Schwellwertes betreffen. In dieser Arbeit wird ein dynamischer Ansatz gewählt, der auf vielen subjektiven Einschätzungen begründet ist. Besser wäre jedoch eine Variante, die mathematisch oder meteorologisch gut begründet ist. Dazu könnte die Methode nach Davis et al. (2006a) gewählt werden, was allerdings wiederum die Wahl eines festen Schwellwertes voraussetzen würde. Dies ist hier eher nicht gewollt. Andere Überlegungen gehen in die Richtung, den Schwellwert abhängig vom mittleren Niederschlag und seiner Standardabweichung für ein Gebiet an einem Zeitpunkt zu machen.

Eine interessante Erweiterung für SAL wäre eine Fuzzy Betrachtung in der Zeit. Dabei könnten auf stündlicher Zeitskala Vorhersagen benachbarter Zeitpunkte zur Verifikation hinzugezogen werden. Beispielsweise könnte man für einen Zeitpunkt in den Beobachtungen Vorhersagen in einem Zeitfenster ± 2 h (also insgesamt 5 h) für den Vergleich zulassen. Die nach SAL beste Vorhersage sollte dann in die Statistik einfließen. Problematisch hierbei ist, dass nicht immer alle drei Komponenten für den gleichen Zeitpunkt den besten Wert angeben. Dies könnte gelöst werden, indem man den besten Zeitpunkt über den besten L -Wert bestimmt, bei dem also die Lage des vorhergesagten Niederschlagsfeldes die beste ist.

Ebenso könnte es sinnvoll sein, die Zeitreihe in wetterlagenabhängige Klassen zu unterteilen. Es könnte sich herausstellen, dass die Probleme der QNV hauptsächlich bei bestimmten Atmosphärenströmungen auftreten oder von einzelnen, schlecht vorhergesagten Parametern (etwa Luftfeuchtigkeit) abhängen. Dies würde weitere Interpretationshilfen für die Wettervorhersage liefern.

Zu guter Letzt ist es wünschenswert, die hier verwendete Zeitreihe weiter auszudehnen und dabei möglichst homogen zu halten, damit sich die Ergebnisse im statistischen Sinne weiter festigen, wobei es natürlich auch interessant ist, verschiedene Modellversionen zu gleichen Zeitpunkten zu betrachten, z.B. um die Qualität der Vorhersagen mit Einführung des prognostischen Niederschlags weiter zu quantifizieren.

An diese Idee schließt sich die Gegenüberstellung vieler verschiedener Modelle an. Ein Vergleich dieser Modelle mit SAL würde dann die bestmögliche Vorhersage für einen oder mehrere Zeitpunkte herausfiltern, wobei dies auch bedeuten kann, dass ein Modell beispielsweise das beste Ergebnis für die A -Komponente, jedoch ein anderes Modell das beste Ergebnis für z.B. die L -Komponente liefert (man also herausfindet, welches Modell welche Charakteristik in der QNV am besten simuliert). Einen für solche Untersuchungen sehr guten Datensatz findet man in den Vorhersagen aus vielen verschiedenen hoch-aufgelösten Modellen während MAP D-Phase (Rotach et al., 2005) im Sommer 2007.

Anhang A

Häufigkeiten von Tagessummen

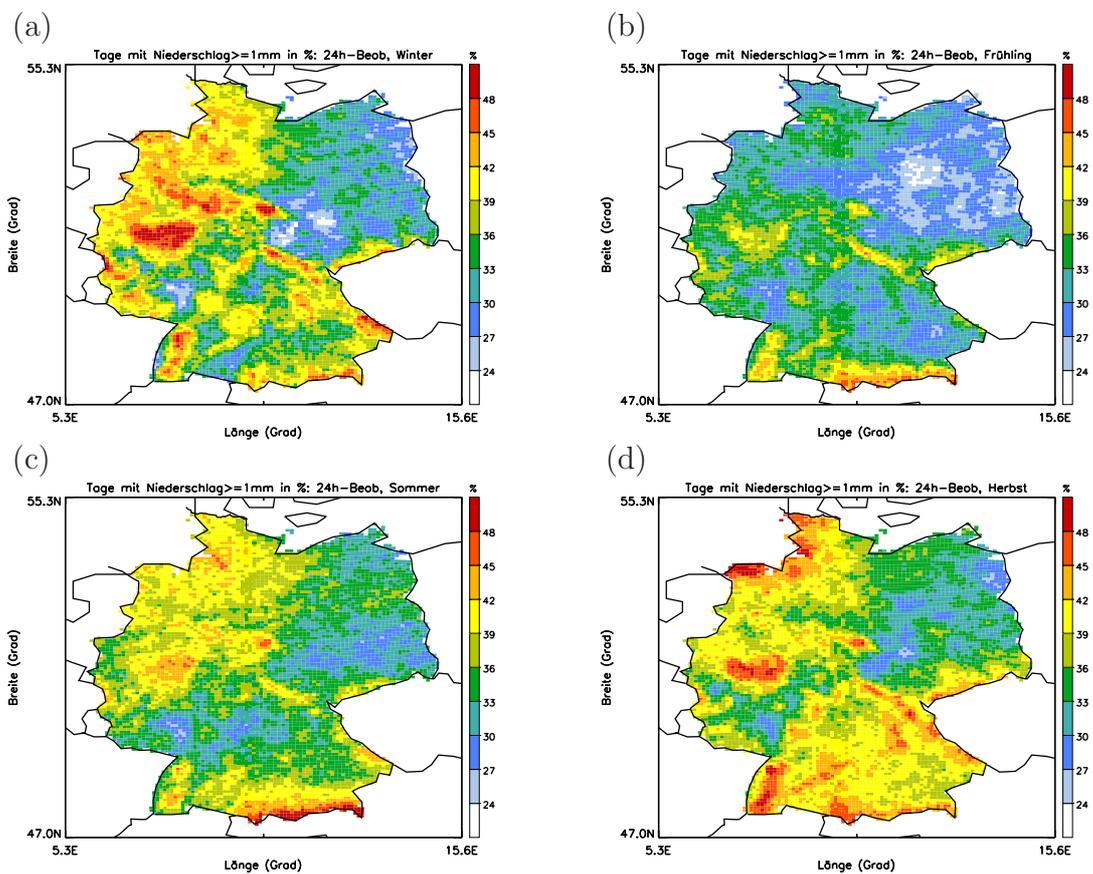


Abbildung A.1: Beobachtete Häufigkeiten von Niederschlagstagen mit Niederschlagssummen ≥ 1 mm in Prozent (Gesamtheit sind alle Tage der jeweiligen Jahreszeit) an jedem Gitterpunkt für die Jahre 2001-2004: (a) Winter, (b) Frühling, (c) Sommer und (d) Herbst.

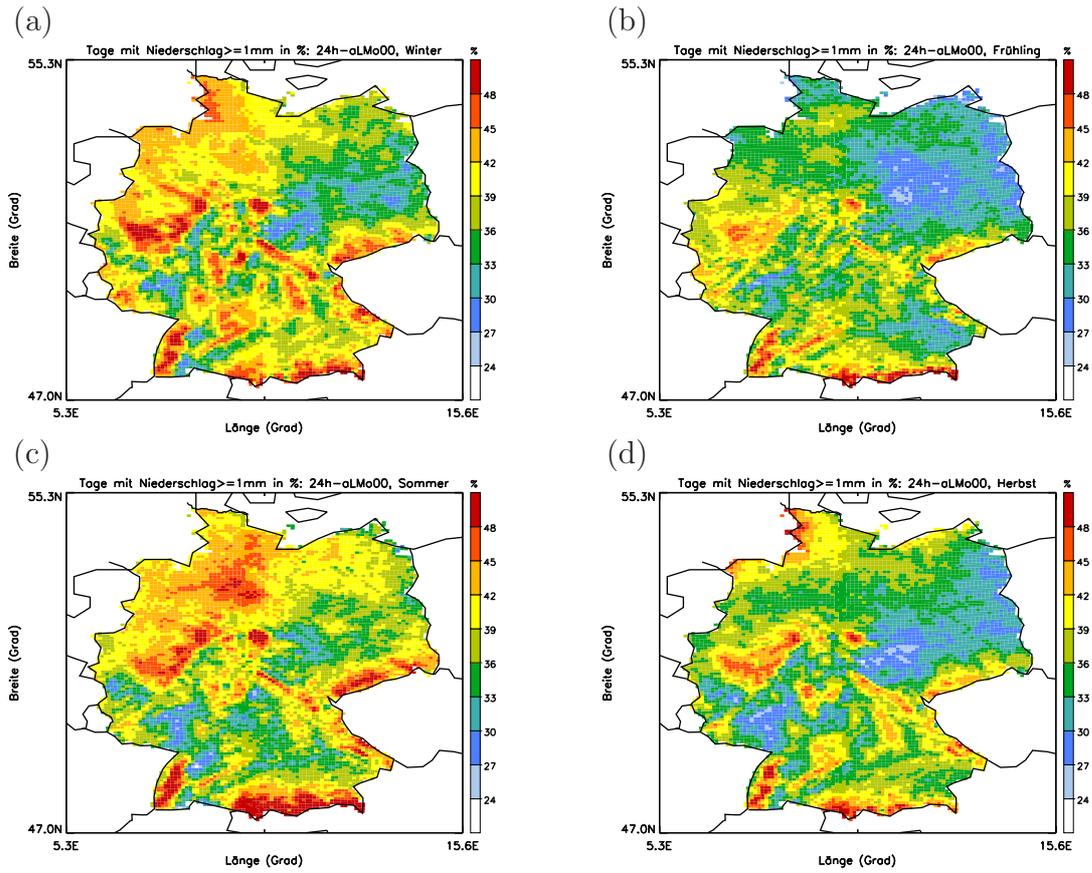


Abbildung A.2: Wie Abb. A.1 für aLMo Vorhersagen (00 UTC-Läufe).

Anhang B

SAL: Abhängigkeit vom Schwellwert

B.1 aLMO-Tagessummen: Winter

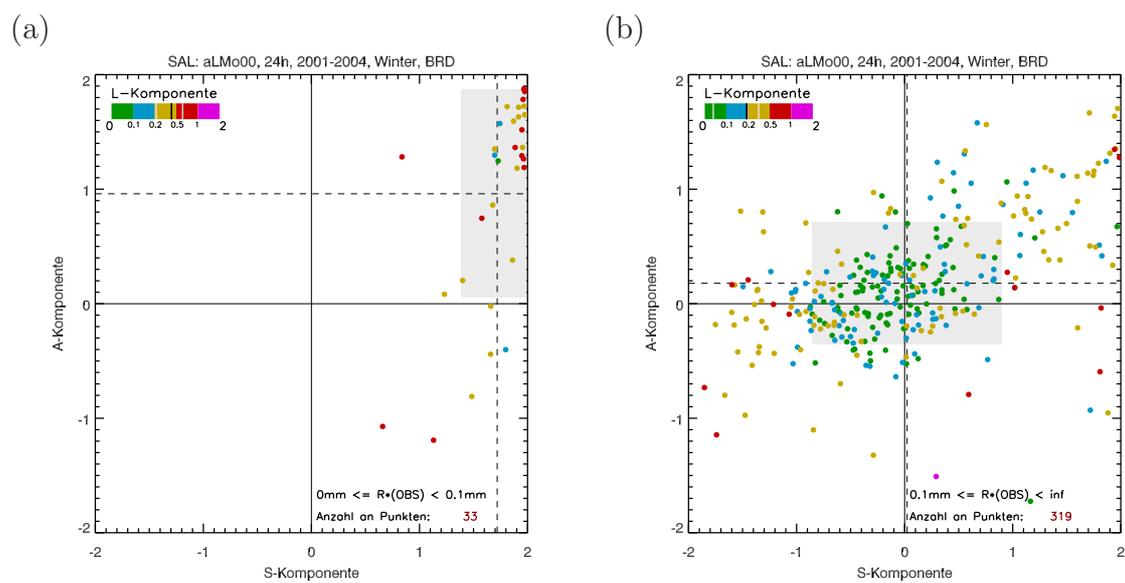


Abbildung B.1: SAL für Tagessummen aus aLMO-Vorhersagen (00 UTC-Läufe) für Deutschland im Winter aufgeteilt in zwei Schwellwertbereiche: (a) $0 < R_{obs}^* \leq 0.1 \text{ mm}$, (b) $0.1 \text{ mm} < R_{obs}^* < \infty$.

B.2 ECMWF-Tagessummen: Winter und Sommer

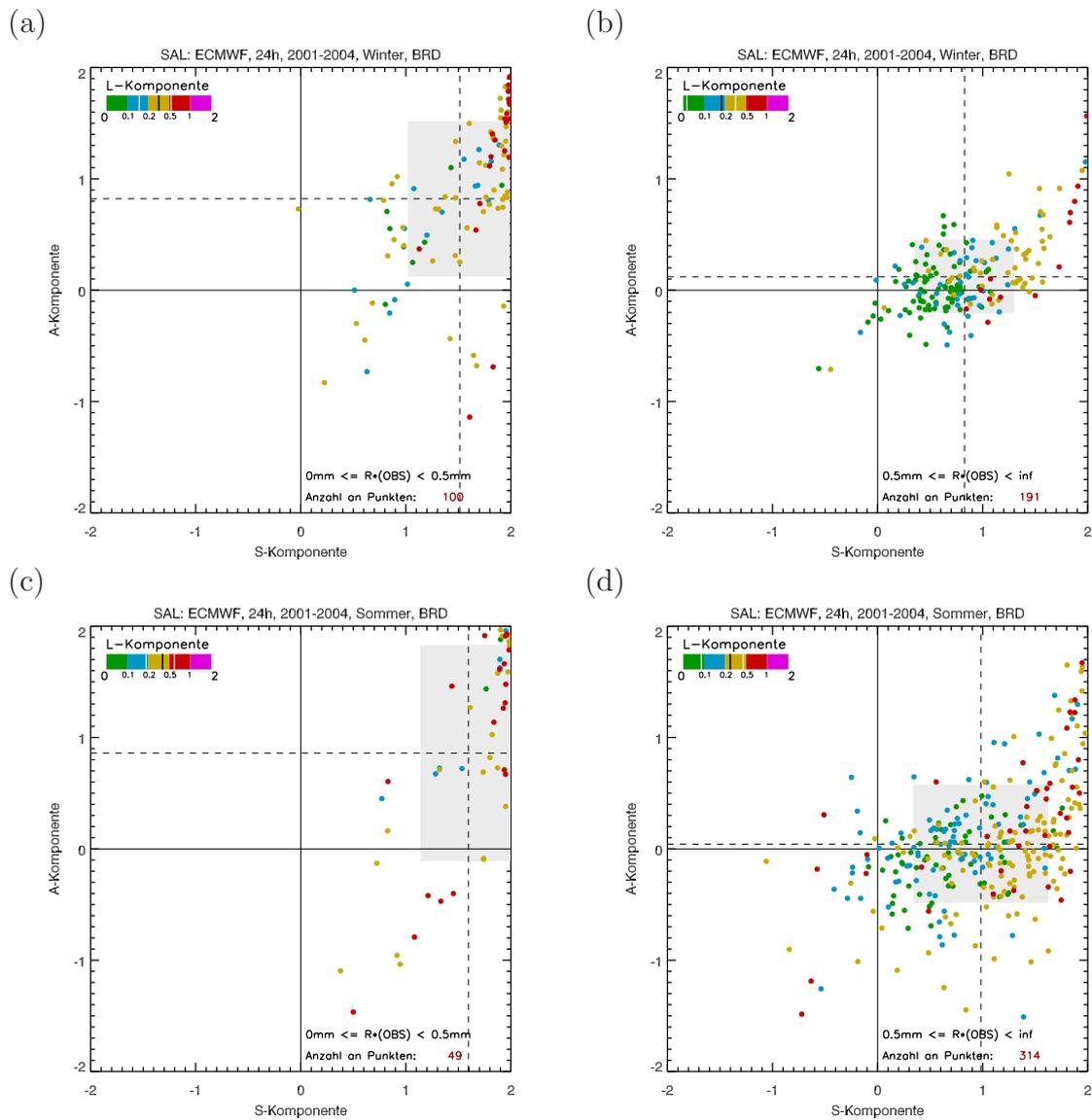


Abbildung B.2: SAL für Tagessummen aus ECMWF-Vorhersagen für Deutschland im Winter (oben) und Sommer (unten) aufgeteilt in zwei Schwellwertbereiche: (a) $0 < R_{obs}^* \leq 0.5\text{mm}$, (b) $0.5\text{mm} < R_{obs}^* < \infty$.

Zeichentabellen

Abkürzung	Bezeichnung
#	Anzahl
Abb.	Abbildung(en)
Abschn.	Abschnitt(e)
Anh.	Anhang/Anhänge
aLMo	alpines Lokal-Modell
COPS	engl.: Convective and Orographically-induced Precipitation Study
COSMO	engl.: Consortium for Small-scale Modelling
CRA	engl.: Contiguous Rain Area
DFG	Deutsche Forschungsgemeinschaft
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DWD	Deutscher Wetterdienst
ECMWF	engl.: European Centre for Medium Range Weather Forecast
Gl.	Gleichung(en)
GME	Globales Modell Europa
GOP	engl.: General Observation Period
Kap.	Kapitel
KT	Kontingenztafel
LAWA	Bund/Länder-Arbeitsgemeinschaft Wasser
LM	Lokal-Modell
LME	Lokal-Modell Europa
LMK	Lokal-Modell Kurzzeit
MAP	engl.: Mesoscale Alpine Program
MEZ	Mittel-Europäische Zeit
MESZ	Mittel-Europäische Sommerzeit
MOS	engl.: Model Output Statistics
QNV	Quantitative Niederschlagsvorhersage
Radar	engl.: Radio detecting and ranging
SPP	Schwerpunktprogramm

Abkürzung	Bezeichnung
SW	Schwellwert
SYMAP	engl.: Synagraphic Mapping System
Tab.	Tabelle(n)
UTC	engl.: Coordinated Universal Time
VERIPREG	engl.: Verification of Precipitation over Germany
WGNE	engl.: Working Group on Numerical Experimentation

Variable	Bezeichnung
σ	Standardabweichung
a	konstanter Faktor der z - R -Beziehung
A	Amplitudenfehler (SAL); KT: Anzahl Treffer für Niederschlag $>$ SW
b	konstanter Exponent der z - R -Beziehung
B	KT: Anzahl falscher Alarme
C	Verhältnis Radar/Bodenwert Tagessummen; KT: Anzahl verpasster Ereignisse
CSI	engl.: Critical Success Index
d	größtmögliche Entfernung zwischen zwei Gitterpunkten im Gebiet
D	Größenverteilung der Regentropfen; KT: Treffer für Niederschlag $<$ SW
$D(R)$	Gebietsmittel von R
\mathcal{D}	Menge aller Gitterpunkte im betrachteten Gebiet
E	in ETS : den Zufall repräsentierender Anteil
ETS	engl.: Equitable Threat Score
f	Faktor bei Schwellwertbestimmung (SAL)
F	engl.: Forecast (Vorhersage)
FBI	engl.: Frequency Bias Index
H	in HSS : den Zufall repräsentierender Anteil
HSS	engl.: Heidke Skill Score
I	Intervall
inf	engl.: infinity (unendlich)
L	Location-/Ortsfehler (SAL)
L_1	Teilkomponente von L (Verschiebungsfehler)
L_2	Teilkomponente von L (Hanteleffekt)
M	Anzahl an Niederschlagsobjekten im Gebiet
N_g	Anzahl Gitterpunkte

Variable	Bezeichnung
N_n	Anzahl Gitterpunkte in \mathcal{R}_n
N_t	Anzahl Zeitpunkte
O	engl.: Observation (Beobachtung)
POD	engl.: Probability Of Detection
r	mittler gewichtete Entfernung zwischen allen \mathbf{x}_n und \mathbf{x}
\tilde{r}	skalierte Distanz
R	Niederschlagsfeld
R^*	Schwellwert zur Definition von Objekten
R_{diff}	gebietsgemittelte Niederschlagsdifferenz
$R_{dis,h}$	disaggregierte Stundensumme
$R_{ij}, R(i, j)$	Niederschlag am Gitterpunkt (i, j)
R^{max}	Niederschlagsmaximum im Gebiet
R^{min}	Niederschlagsminimum im Gebiet
R_{mod}	Modell-Niederschlag
R_{obs}	beobachteter Niederschlag
$R_{obs,d}$	Tagessumme am Boden
$R_{rad,d}$	Tagessumme Radar
$R_{rad,h}$	Stundensumme Radar
R_d	mittlerer Tagesniederschlag
R_{dmax}	maximaler Niederschlag des Tagesgangs
$(R)MSE$	engl.: (Root) Mean Square Error
\mathcal{R}_n	Menge aller Gitterpunkte im Niederschlagsobjekt mit Index n
R_n^{max}	Niederschlagsmaximum in Objekt n
R_n	Niederschlagssumme in Objekt n
S	Strukturfehler (SAL)
V	gewichtetes Mittel aller skalierten Volumen der Objekte im Gebiet
V_n	skaliertes Volumen von Objekt n
w_d	Gewichtungsfunktion
x_g, y_g	Gitterpunktskoordinaten
x_s, y_s	Stationskoordinaten
\mathbf{x}	Schwerpunkt des Niederschlagsfeldes im Gebiet
\mathbf{x}_n	Schwerpunkt des Niederschlags in Objekt n
$\Delta x, \Delta y$	Gitterpunktsabstand
z	Reflektivitätsfaktor in mm^6m^{-3}
Z	Reflektivität in dBZ

Literaturverzeichnis

- Aniol, R., J. Riedl und M. Dieringer, 1980: Über kleinräumige und zeitliche Variationen der Niederschlagsintensität. *Meteorol. Rdsch.*, **33**, 50-56.
- Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlin. Proc. Geoph.*, **8**, 401-417.
- Baldauf, M., J. Förster, S. Klink, T. Reinhardt, C. Schraff, A. Seifert und K. Stephan, 2006: Kurze Beschreibung des Lokal-Modells Kürzestfrist LMK und seiner Datenbanken auf dem Datenserver des DWD. *DWD*, Geschäftsbereich Forschung und Entwicklung, Offenbach, Deutschland.
- Balzer, K., 2002: Wie gut sind unsere Wettervorhersagen. *DWD*, Geschäftsbereich Basisdienste, 21 pp.
- Battan, L. J., 1973: Radar observation of the atmosphere. *The University of Chicago Press*.
- Berliner Wetterkarten, 2001. *Verein BERLINER WETTERKARTE*, Berlin, Deutschland.
- Binder, P. und C. Schär, 1995: Mesoscale Alpine Programme: Design proposal. *Swiss Meteorol. Inst.*, Zürich, 77pp.
- Casati, B., G. Ross und D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.*, **11**, 141-154.
- COSMO-Newsletter No.6, 2006: <http://www.cosmo-model.org>
- COSMO-Homepage, August 2007: <http://cosmo-model.cscs.ch/public/default.htm>
- Crewell, S., K. Bumke, G. Dick, J. Fischer, M. Hagen, S. Keyn, C. Koziar, A. Mathes, G. Peters, J. Rost, C. Simmer, R. Sussmann und M. Wiegner, 2005: General observation period (GOP). *Newsletter SPP 02/2005*.
- Dai, A., F. Giorgi und K. E. Trenberth, 1999: Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *J. Geophys. Res.*, **104**(D6), 6377-6402.
- Damrath, U., G. Doms, D. Frühwald, E. Heise, B. Richter und J. Steppeler, 2000: Operational quantitative precipitation forecasting at the German Weather Service. *J. Hydrol.*, **239**, 260-285.
- Damrath, U., 2002: Verifikation von Wettervorhersagen. *Promet*, **28**(1/2), 8-16.

- Davis, C. A., B. Brown und R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772-1784.
- Davis, C. A., B. Brown und R. Bullock, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785-1795.
- Doms, G., J. Steppeler, und G. Adrian, 2002: Das Lokal-Modell LM. *Promet*, **27**(3/4), 123-128.
- ECMWF-Homepage, August 2007: <http://www.ecmwf.int>
- ECMWF-Newsletter No.101, 2004: <http://www.ecmwf.int>
- Ebert, E. E. und J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrol.*, **239**, 179-202.
- Ebert, E. E., U. Damrath, W. Wergen und M. E. Baldwin, 2003: The WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Am. Meteorol. Soc.*, **84**(4), 481-492.
- Ebert, E. E., 2006: Fuzzy forecast verification. 1st MAP D-Phase Scientific Meeting, Vortrag.
- Ebert, E. E., 2007: Forecast Verification - Issues, Methods and FAQ. *Homepage of the Joint Working Group on Verification*, http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html
- Ekhard, E., 1948: Die Niederschlagsverteilung in den Alpen nach dem Anomalienprinzip. *Geogr. Ann.*, **30**, 728-739.
- Frei, C. und C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, **18**, 873-900.
- Frei, C., J. Hesselbjerg Christensen, M. Déqué, D. Jacob, R. G. Jnes und P. L. Vidale, 2003: Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. *J. Geophys. Res.*, **108**(D3), 4124, doi:10.1029/2002JD002287.
- Fritsch, J. M. und R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season. *Bull. Am. Meteorol. Soc.*, **85**, 955-965.
- GOP-Homepage, August 2007: <http://gop.meteo.uni-koeln.de/gop/doku.php>
- Häckel, H., 1985: *Meteorologie*, Verlag Eugen Ulmer, Stuttgart.
- Hagen, M., C. Frei und C. Schär, 2003: Temporal disaggregation of rain-gauge analyses using radar. *Presentation at the Int. Conf. Alpine Meteorol. and MAP-meeting*, Brig, Schweiz.
- Hagen, M., 2004: Regenmessung mit Radar für VERIPREG. DLR Oberpfaffenhofen, Deutschland.
- Hense, A., G. Adrian, C. Kottmeier, C. Simmer und V. Wulfmeyer, 2003: Quantitative Niederschlagsvorhersage. *Proposal zum SPP 1167 der DFG*.

- Hense, A., G. Adrian, C. Kottmeier, C. Simmer, V. Wulfmeyer, A. Behrendt und A. Mathes, August 2007:
<http://www.meteo.uni-bonn.de/projekte/SPPMeteo/>
- Hohenegger, C., D. Lüthi und C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.*, **134**(8), 2095-2107.
- Intergovernmental Panel on Climate Change (IPCC), 2007: The IPCC 4th assessment report - Climate Change 2007.
- Kaufmann, P., F. Schubiger und P. Binder, 2003: Precipitation forecasting by a mesoscale numerical weather prediction (NWP) model: eight years of experience. *H E S S*, **7**(6), 812-832.
- Litschi, M., 2005: Kombination von Radar- und Pluviometerdaten zu stündlichen Niederschlagsfeldern für die Schweiz: Validierung und klimatologische Auswertung. *Inst. für Atm. und Klima*, ETH Zürich, Schweiz, Diplomarbeit.
- Marshall, J. und W. Palmer, 1948: The distribution of raindrops with size. *J. Meteorol.*, **5**, 165-166.
- Marzban, C. und S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, 824-838.
- Müller-Westermeier, G., 1995: Numerisches Verfahren zur Erstellung klimatologischer Karten. *Berichte des Deutschen Wetterdienstes*, 193.
- New, M., M. Hulme und P. Jones, 2000: Representing Twentieth-Century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217-2238.
- Osborn, T. J. und M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Climate*, **10**, 1885-1908.
- Peck, E. L. und M. J. Brown, 1962: An approach to the development of isohyetal maps for mountainous areas. *J. Geophys. Res.*, **67**, 681-694.
- Richter, D., 1995: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Meßfehlers des Hellmann-Niederschlagsmessers. *Berichte des Deutschen Wetterdienstes*, 194.
- Roberts, N. M., 2005: An investigation of the ability of a storm-scale configuration of the Met Office NWP model to predict flood-producing rainfall. *Forecasting Research Tech. Rept.*, **455**, 80 pp., Met Office, UK.
- Rotach, M., F. Bouttier, A. Buzzi, M. Dorninger, G. Frustaci, K. Mylne, R. Ranzi, E. Richard, C. Schär, M. Staudinger, H. Volkert und V. Wulfmeyer, 2005: MAP D-Phase: A forecast demonstration project in the framework of MAP. *Proposal to the WWRP-SSC*.
- Schulz, J.-P. und U. Schättler, 2005: Kurze Beschreibung des Lokal-Modells LME und seiner Datenbank auf dem Datenserver des DWD. *DWD, Geschäftsbereich Forschung und Entwicklung*, Offenbach, Deutschland.

- Sevruk, B., 1985b: Systematischer Niederschlagsmessfehler in der Schweiz, in Sevruk, B. (ed.), *Der Niederschlag in der Schweiz. Beiträge Geol. Schweiz - Hydrologie*, **31**, 65-75.
- Shepard, D. S., 1984: Computer mapping: the SYMAP interpolation algorithm, in Gaile, G. L. und C. J. Willmot (eds.), *Spatial Statistics and Models*, Dordrecht, pp. 133-145.
- Simmons, A. J. und A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteorol. Soc.*, **128**, 647-677.
- Jolliffe, I. T. und D. B. Stephenson, 2003: *Forecast Verification. A Practitioner's Guide in Atmospheric Science. Wiley and Sons Ltd.*, 240 pp.
- Steppeler, J., G. Doms, U. Schättler, H. W. Bitzer, A. Gassmann, U. Damrath und G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorol. Atmos. Phys.*, **82**, 75-96.
- Tartaglione, N., S. Mariani, C. Accadia, A. Speranza und M. Casaioli, 2005: Comparison of raingauge observations with modeled precipitation over Cyprus using contiguous rain area analysis. *Atmos. Chem. Phys.*, **5**, 2147-2154.
- Theis, S. E., A. Hense und U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorol. Appl.*, **12**, 257-268.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779-1800.
- Venugopal, V., S. Basu und E. Foufoula-Georgiou, 2005: A new metric for comparing precipitation patterns with an application to ensemble forecasts. *J. Geophys. Res.*, **110**(D8), 8111, doi:10.1029/2004JD005395.
- Walser, A., D. Lüthi und C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**(2), 560-577.
- Walser, A. und C. Schär, 2004: Convection-resolving Precipitation Forecasting and its Predictability in Alpine River Catchments. *J. Hydrol.*, **288**, 57-73.
- Wernli, H., M. Hagen und C. Frei, 2003: Novel approaches for the verification of precipitation forecasts from numerical models on short time-scales. *Antrag innerhalb des SPP „Quantitative Niederschlagsvorhersage“*, Mainz, Deutschland.
- Wernli, H. und M. Sprenger, 2007: Identification and ERA15 climatology of potential vorticity streamers and cut-offs near the extratropical tropopause. *J. Atmos. Sci.*, **64**, 1569-1586.
- Widmann, M. und C. S. Bretherton, 2000: Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the northwestern United States. *J. Climate*, **13**, 1936-1950.
- Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences: an introduction. Academic Press, San Diego.*

- Woodcock, F., 1976: The evaluation of Yes/No forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, **104**, 1209-1214.
- Wulfmeyer, V., A. Behrendt, C. Kottmeier, U. Corsmeier und L. Gantner, August 2007: <http://www.uni-hohenheim.de/spp-iop/>
- Zhang, F., C. Snyder und R. Rotunno, 2002: Mesoscale predictability of the surprise snowstorm of 24/25 January 2000. *Mon. Wea. Rev.*, **130**, 1617-1632.
- Zhang, F., C. Snyder und R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173-1185.
- Zhang, F., A. M. Odins und J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149-166.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende wissenschaftliche Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Marcus Paulat, Mainz, den 21.September 2007

Lebenslauf

Name: Marcus Paulat
Geburtstag: 31. Oktober 1977
Geburtsort: Delmenhorst

Ausbildung:

2004-2007 Promotion am
Institut für Physik der Atmosphäre
Johannes Gutenberg-Universität Mainz, Deutschland
Titel: „Verifikation der Niederschlagsvorhersage
für Deutschland von 2001-2004“

2003-2004 Diplomarbeit am
Institut für Physik der Atmosphäre
Johannes Gutenberg-Universität Mainz, Deutschland
Titel: „Berechnung und Analyse von CAPE für
Süddeutschland“

1998-2003 Studierender der Meteorologie am
Institut für Physik der Atmosphäre
Johannes Gutenberg-Universität Mainz, Deutschland

1988-1997 Abitur am
Max-Slevogt-Gymnasium, Landau/Pfalz, Deutschland
Leistungskurse: Mathematik, Physik, Erdkunde

Ersatzwehrdienst:

1997-1998 Dekanat der Evangelischen Kirche, Landau/Pfalz, Deutschland

Sprachen: Deutsch (Muttersprache), Englisch