

Merkmalsbezogene Verifikation hochaufgelöster Niederschlagsvorhersagen für Deutschland

Dissertation
zur Erlangung des Grades
„Doktor der
Naturwissenschaften“

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

MATTHIAS ZIMMER
geb. am 8. April 1979
in Dresden

Mainz, September 2010

Tag der Promotion: 9. Dezember 2010

D77 - Mainzer Dissertation

Inhaltsverzeichnis

Zusammenfassung	vii
Abstract	ix
1 Wissenschaftliche Basis und Fragestellungen	1
1.1 Quantitative Niederschlagsvorhersagen	4
1.1.1 Numerische Vorhersagemodelle	5
1.1.2 Beobachtungsdaten	9
1.2 Methoden zur Niederschlagsverifikation	11
1.2.1 Gitterpunktbezogene Bewertungsmaße	13
1.2.2 Räumliche Methoden	18
1.3 Ziele und Aufbau der Arbeit	23
2 Niederschlagsklimatologie für Deutschland	27
2.1 Beobachtete Niederschlagsverteilung	27
2.1.1 Jährliche Niederschlagsmengen	27
2.1.2 Saisonale Niederschlagssummen	29
2.1.3 Niederschlagshäufigkeiten	32
2.1.4 Wöchentlicher Zyklus	36
2.2 Vergleich zwischen Modellsimulation und Beobachtungen	42
2.2.1 Differenz der Niederschlagssummen	42
2.2.2 Unterschiede in den Niederschlagshäufigkeiten	45

2.2.3	Unterschiede in der Intensität der Niederschlagsereignisse	48
2.3	Der Tagesgang des Niederschlags	50
2.3.1	Beobachtungen	51
2.3.2	Modellsimulationen	53
3	Das Qualitätsmaß SAL	61
3.1	Definition und grundlegende Beschreibung	61
3.1.1	Definition von SAL	61
3.1.2	Das SAL Diagramm	65
3.2	Spatial Forecast Verification Methods Inter – Comparison Project	68
3.2.1	Einleitung	69
3.2.2	Geometrische Beispiele	69
3.2.3	Manipulierte Niederschlagsfelder	72
3.2.4	Reale Situationen von drei Modellen	74
3.2.5	Schlussfolgerungen	83
3.3	Quantifizierung zeitlicher Fehler	85
3.3.1	Motivation für die zeitliche Komponente	85
3.3.2	Methodische Vorgehensweise	87
3.3.3	Anwendungsbeispiel	88
4	MAP D-PHASE	97
4.1	Grundlegende Ausführungen	97
4.2	Daten und Methodik	99
4.3	Verifikationsergebnisse	102
4.4	Zusammenfassung der Ergebnisse	113
5	Langzeitevaluierung von Niederschlagsvorhersagen	119
5.1	Zeitliche Entwicklung der Vorhersagequalität	119
5.1.1	Traditionelle Bewertungsmaße	120
5.1.2	SAL Resultate	126

5.1.3	Abschließende Bemerkungen	130
5.2	Konvektive Zeitskala	131
5.2.1	Theoretische Betrachtungen	131
5.2.2	Daten und Methodik	134
5.2.3	Resultate	136
5.2.4	Schlussfolgerungen	140
6	Zusammenfassung und Ausblick	141
6.1	Zusammenfassung	141
6.2	Ausblick	144
A	Häufigkeiten von Tagessummen	147

Zusammenfassung

Die Verifikation bewertet die Güte von quantitativen Niederschlagsvorhersagen (QNV) gegenüber Beobachtungen und liefert Hinweise auf systematische Modellfehler. Mit Hilfe der merkmals-bezogenen Technik SAL werden simulierte Niederschlagsverteilungen hinsichtlich (S)truktur, (A)mplitude und (L)ocation analysiert. Seit einigen Jahren werden numerische Wettervorhersagemodelle benutzt, mit Gitterpunktabständen, die es erlauben, hochreichende Konvektion ohne Parametrisierung zu simulieren. Es stellt sich jetzt die Frage, ob diese Modelle bessere Vorhersagen liefern.

Der hochaufgelöste stündliche Beobachtungsdatensatz, der in dieser Arbeit verwendet wird, ist eine Kombination von Radar- und Stationsmessungen. Zum einen wird damit am Beispiel der deutschen COSMO-Modelle gezeigt, dass die Modelle der neuesten Generation eine bessere Simulation des mittleren Tagesgangs aufweisen, wenn auch mit zu geringen Maximum und etwas zu spätem Auftreten. Im Gegensatz dazu liefern die Modelle der alten Generation ein zu starkes Maximum, welches erheblich zu früh auftritt. Zum anderen wird mit dem neuartigen Modell eine bessere Simulation der räumlichen Verteilung des Niederschlags, durch eine deutliche Minimierung der Luv-/Lee Problematik, erreicht. Um diese subjektiven Bewertungen zu quantifizieren, wurden tägliche QNVs von vier Modellen für Deutschland in einem Achtjahreszeitraum durch SAL sowie klassischen Maßen untersucht. Die höher aufgelösten Modelle simulieren realistischere Niederschlagsverteilungen (besser in *S*), aber bei den anderen Komponenten tritt kaum ein Unterschied auf. Ein weiterer Aspekt ist, dass das Modell mit der größten Auflösung (ECMWF) durch den *RMSE* deutlich am besten bewertet wird. Darin zeigt sich das Problem des ‚Double Penalty‘. Die Zusammenfassung der drei Komponenten von SAL liefert das Resultat, dass vor allem im Sommer das am feinsten aufgelöste Modell (COSMO-DE) am besten abschneidet. Hauptsächlich kommt das durch eine realistischere Struktur zustande, so dass SAL hilfreiche Informationen liefert und die subjektive Bewertung bestätigt.

Im Jahr 2007 fanden die Projekte COPS und MAP D-PHASE statt und boten die Möglichkeit, 19 Modelle aus drei Modellkategorien hinsichtlich ihrer Vorhersageleistung in

Südwestdeutschland für Akkumulationszeiträume von 6 und 12 Stunden miteinander zu vergleichen. Als Ergebnisse besonders hervorzuheben sind, (i) dass je kleiner der Gitterpunktabstand der Modelle ist, desto realistischer sind die simulierten Niederschlagsverteilungen; (ii) bei der Niederschlagsmenge wird in den hoch aufgelösten Modellen weniger Niederschlag, d.h. meist zu wenig, simuliert und (iii) die Ortskomponente wird von allen Modellen am schlechtesten simuliert. Die Analyse der Vorhersageleistung dieser Modelltypen für konvektive Situationen zeigt deutliche Unterschiede. Bei Hochdrucklagen sind die Modelle ohne Konvektionsparametrisierung nicht in der Lage diese zu simulieren, wohingegen die Modelle mit Konvektionsparametrisierung die richtige Menge, aber zu flächige Strukturen realisieren. Für konvektive Ereignisse im Zusammenhang mit Fronten sind beide Modelltypen in der Lage die Niederschlagsverteilung zu simulieren, wobei die hoch aufgelösten Modelle realistischere Felder liefern. Diese wetterlagenbezogene Untersuchung wird noch systematischer unter Verwendung der konvektiven Zeitskala durchgeführt. Eine erstmalig für Deutschland erstellte Klimatologie zeigt einen einer Potenzfunktion folgenden Abfall der Häufigkeit dieser Zeitskala zu größeren Werten hin auf. Die SAL Ergebnisse sind für beide Bereiche dramatisch unterschiedlich für ein Modell mit parametrisierter hochreichender Konvektion. Für kleine Werte der konvektiven Zeitskala sind sie gut, dagegen werden bei großen Werten die Struktur sowie die Amplitude deutlich überschätzt.

Für zeitlich sehr hochaufgelöste Niederschlagsvorhersagen gewinnt der Einfluss der zeitlichen Fehler immer mehr an Bedeutung. Durch die Optimierung/Minimierung der L Komponente von SAL innerhalb eines Zeitfensters ($\pm 3h$) mit dem Beobachtungszeitpunkt im Zentrum ist es möglich diese zu bestimmen. Es wird gezeigt, dass bei optimalem Zeitversatz die Struktur und Amplitude der QNVs für das COSMO-DE besser werden und damit die grundsätzliche Fähigkeit des Modells die Niederschlagsverteilung realistischer zu simulieren, besser gezeigt werden kann.

Abstract

The verification assesses the goodness of quantitative precipitation forecasts (QPF) compared with observations and delivers hints to systematic model mistakes. With the help of the sign-related technology SAL simulated precipitation distributions are analysed concerning (S)tructure, (A)mplitude and (L)ocation. For some years numerical weather prediction models are used, with grid point distances which allow simulating deep convection without parameterization. Now it puts to itself the question whether these models deliver better predictions.

The highly resolved hourly observation data set, which is used in this work, combines radar and rain gauge measurements. On the one hand it is shown with the example of the German COSMO models that the models of the newest generation show a better simulation of the mean diurnal cycle, even if with too low maximum and something to late appearance. By contrast the models of the old generation deliver a too strong maximum which appears as too early considerably. On the other hand a better simulation of the spatial distribution of the precipitation is reached with the new model type, by a clear minimisation of the windward/lee problem. To quantify these subjective assessments, daily QPFs were examined by four models for Germany within an eight year time period by SAL as well as classical measures. The models with the higher resolution models simulate more realistic precipitation distributions (better in *S*), but for the other components no difference appears. Another aspect is that the model with the coarsest grid (ECMWF) is valued by the *RMSE* clearly as the best of all. Therein appears the ‘Double Penalty’ problem. The summary of the three components of SAL delivers the result that in particular for summer the most finely resolved model (COSMO-DE) takes the best values of all. Primarily it comes about by a more realistic structure, so that SAL delivers helpful information and confirms the subjective assessment.

In 2007 the projects COPS and MAP D-PHASE took place and offered the unique Opportunity to compare 19 models from three model categories concerning their forecast performance in South-western Germany for accumulation periods of 6 and 12 hours. When results particularly are to be emphasised that (i) the smaller the grid point distance of

the models is, the simulated precipitation distributions are the more realistically; (ii) with the amount of precipitation less precipitation, i.e. mostly not enough, is simulated in the models with higher resolution and (iii) the L component is simulated by all models most badly. The analysis of the prediction achievement of these model types for convective situations shows clear differences. With high-pressure situations the models are to be operated without a parameterization of deep convection not able to simulate this, while the models with a parameterization of deep convection realise the right amount, but too flat rain objects. For convective events in connection with fronts both model types are to be simulated the precipitation distribution but the models with the higher spatial resolution deliver more realistic fields. This weather condition-related investigation becomes even more systematic under the use of the convective timescale approach. A climatology first provided for Germany indicates the distribution closely follows a power law one of a power function to higher values. The results of SAL for a model with parameterization of deep convection are dramatically different for both regimes. For small values of the convective timescale it performs well, but for high timescale values it shows a clearly overestimation with high values for Structure as well as the Amplitude.

For temporal very highly resolved precipitation simulations the influence of the temporal errors wins more and more in meaning. By the optimisation/minimisation of the L component of SAL within a time window ($\pm 3h$) with the observation time is in the centre it is possible to do this. It is shown that with optimum time shift the structure and amplitude of the QPFs become better for the COSMO-DE and with it the basic ability of the model to simulate more realistically the precipitation distribution.

Kapitel 1

Wissenschaftliche Basis und Fragestellungen

Der Niederschlag stellt eine der wichtigsten meteorologischen Größen der Erdatmosphäre dar, denn er beeinflusst das menschliche Leben in verschiedenster Weise. Als einziger Bestandteil der Erdatmosphäre liegt Wasser in drei Aggregatzuständen (gasförmig, flüssig und fest) vor und wird durch verschiedene Umwandlungsprozesse innerhalb des hydrologischen Kreislaufs von einem in den anderen Zustand überführt. Zum einen stellt Wasser einen großen Einflussfaktor für viele Wirtschaftszweige dar, wie z.B. die Landwirtschaft oder die Stromgewinnung aus Wasserkraft (*Theis et al., 2005*). Auf der anderen Seite sind es Starkniederschlagsereignisse, die ein hohes Zerstörungspotential besitzen. Diese reichen von Überschwemmungen über Erdbeben bis hin zu Begleiterscheinungen wie Hagel, Blitzschlag und Starkwindereignissen. Ein gutes Verständnis der Vorgänge der niederschlagsbildenden Prozesse sowie deren Vorhersage ist deshalb nicht nur von wissenschaftlichem Interesse, sondern auch von gesellschaftlicher und ökonomischer Relevanz.

Die Wettervorhersage ist heutzutage eine Aufgabe, die in immer größerem Umfang auf Basis von numerischen Wettervorhersagemodellen bewerkstelligt wird, die die dynamischen und thermodynamischen Gleichungen lösen, um den zukünftigen Zustand der Atmosphäre zu bestimmen. Die Verbesserung und Optimierung der numerischen Wettervorhersagesysteme führte zu einer genaueren Beschreibung des Zustands der Atmosphäre für einen immer längeren Zeitraum im Voraus. In Abbildung 1.1 wird für Europa die zeitliche Entwicklung der Vorhersagegüte des deterministischen Globalmodells des ECMWF für den Zeitraum von 1980 - 2010 für Europa gezeigt. Dargestellt sind in blau das Monatsmittel und in rot das zwölfmonatige gleitende Mittel des Zeitpunkts bei dem der Wert der Anomaliekorrelation des 500 hPa Geopotentials unter 60 % fällt. Gut zu erkennen ist

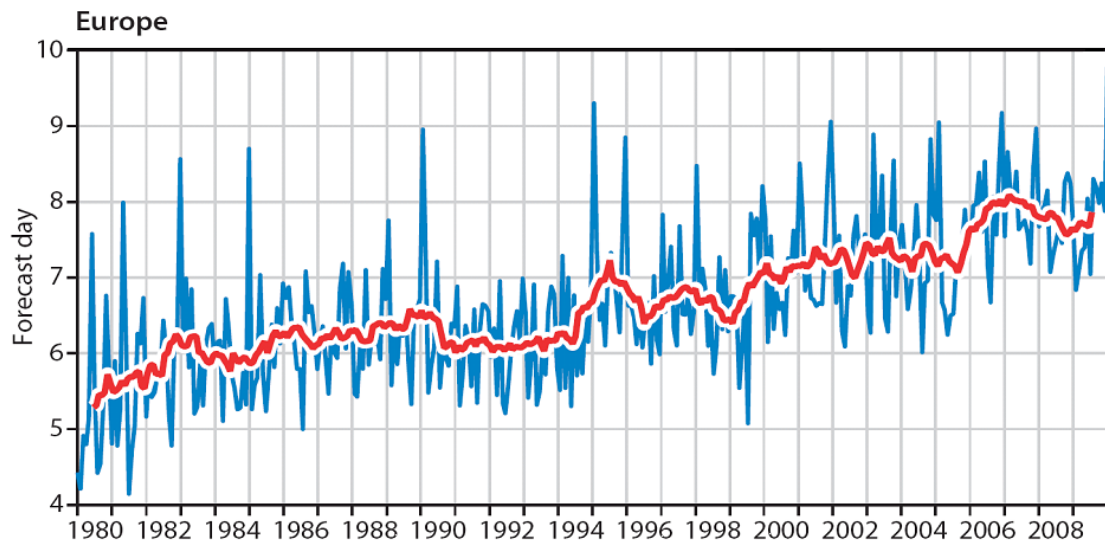


Abbildung 1.1. Zeitliche Entwicklung der Vorhersagegüte der deterministischen, operationellen Vorhersagen des ECMWF Modells seit 1980 (Richardson, 2010). Dargestellt ist das Monatsmittel (blau) und ein gleitendes Zwölfmonatsmittel (rot) des Vorhersagezeitpunkts in Tagen bei dem der Wert der Anomaliekorrelation des 500 hPa Geopotentials gemittelt über Europa unter 60 % fällt.

eine Verbesserung bzw. Verlängerung der Vorhersagehorizonts für das 500 hPa Geopotential. Lag in den 1980er Jahren der Vorhersagehorizont noch bei sechs Tagen so liegt diese Grenze aktuell bei annähernd acht Tagen. Trotzdem sollte gesagt werden, dass es immer wieder Einbrüche in der Qualität gab. Somit kann eine deutliche Verbesserung der Vorhersageleistung auf der mittelfristigen Zeitskala für Europa festgestellt werden. Allerdings sind trotz des sehr großen Aufwands, der zur Verbesserung der numerischen Vorhersagemodelle betrieben wird, die Erfolge beim Niederschlag eher gering. Besonders im Sommer, wenn die Niederschlagsbildung durch konvektive Prozesse dominiert wird, ist die Güte der Niederschlagsvorhersagen besonders schlecht (Ebert *et al.*, 2003; Fritsch and Carbone, 2004). Ein schönes Beispiel, welches die unterschiedliche Entwicklung in der Vorhersagegüte verschiedener meteorologischer Parameter zeigt, liefert die Abbildung 1.2 aus Balzer (2002). Hier ist dargestellt, wie sich die Qualität der Kurzfristvorhersage von Standard-Wetterelementen beim Deutschen Wetterdienst (DWD) zwischen 1984 und 2000 entwickelt hat. Dazu wird das Maß „Reduktion der Varianz“ (*RV*) verwandt, das die prozentuale Verbesserung der Vorhersage gegenüber einer Referenzvorhersage angibt. Im vorliegenden Fall ist dies die Vorhersageleistung zu Beginn der Untersuchungsperiode, um damit direkt die Veränderung aufzeigen zu können. Während die Qualität bei den Wind- (*dd*, *ff*), Bedeckungs- (*B*) und Temperaturvorhersagen (*MIN*,

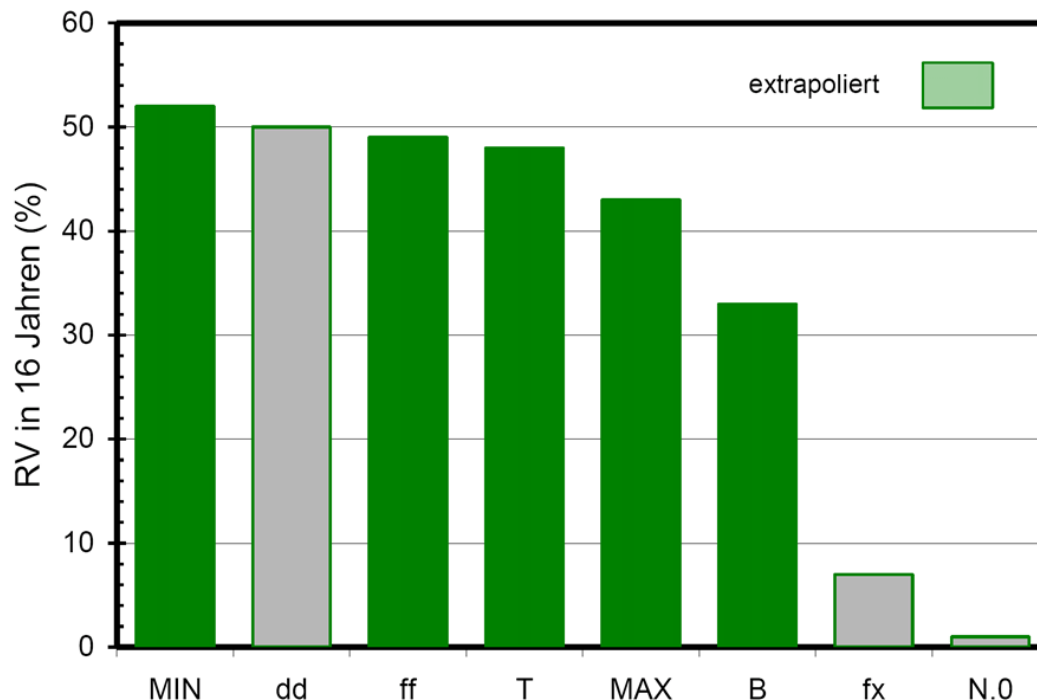


Abbildung 1.2. Zunahme der Prognosegüte von 1984 bis 2000 von Standard-Wetterelementen beim DWD, ausgedrückt durch die Reduktion der Fehlervarianz RV (Abb. 13 aus Balzer, 2002).

MAX und T) erheblich verbessert werden konnte, ist dies bei den Windböen ($> 12 \frac{m}{s}$, fx) und besonders beim Niederschlag (Ja/Nein; N.0) nicht zu erkennen.

Es verbleibt die Frage nach der Ursache, warum in den letzten Jahren bei vielen meteorologischen Parametern ein Verbesserung der Vorhersagequalität erzielt werden konnte, beim Niederschlag jedoch nicht. Aufgrund dieser ungeklärten Frage wurde durch die Deutsche Forschungsgemeinschaft ein Schwerpunktprogramm (SPP) zur Quantitativen Niederschlagsvorhersage (QNV) von 2004-2010 finanziert, um die atmosphärischen Prozesse die zur Niederschlagsbildung beitragen, besser zu verstehen und anschließend besser simulieren bzw. vorhersagen zu können. Innerhalb dieses SPP wurden drei wissenschaftliche Ziele verfolgt (Hense et al., 2003):

- Identifikation der für die Defizite verantwortlichen physikalischen und chemischen Prozesse bei der QNV.
- Bestimmung und Ausschöpfung der Potentiale vorhandener und neuer Daten sowie Prozessbeschreibung zur Verbesserung der QNV.

- Aufzeigen der Vorhersagefähigkeit von numerischen Atmosphärenmodellen mittels statistisch-dynamischer Untersuchungen der QNV.

Die vorliegende Arbeit ist im Projekt VERIPREG (*engl.: VERification of PREcipitation over Germany*, Wernli *et al.*, 2003) durchgeführt worden und ist im Rahmen des SPP dem Themenbereich Verifikation zugeordnet. Die Beiträge und Analyse dieser Arbeit liefern Hinweise zum Erreichen der wissenschaftlichen Ziele ii) und iii).

1.1 Quantitative Niederschlagsvorhersagen

Die Simulation von Niederschlag ist in der numerischen Wettervorhersage relativ schlecht umgesetzt, da die Entstehung und Entwicklung hoch komplex sowie nichtlinear ist. Diese Prozesse umfassen einen weiten räumlichen und zeitlichen Skalenbereich (*Richard et al.*, 2007), von der synoptischen Skala (ca. 1000 km und einer Dauer von bis zu mehreren Tagen) über konvektive Vorgänge (10 km und 30 Minuten) bis hin zur Bildung von Wolkentröpfchen (10 μm und einige Sekunden). Einige Prozesse laufen damit auf Skalen ab, die erheblich kleiner sind als die Größe der Gitterbox im Modell. Um diese sogenannten subgitterskaligen Prozesse in den Modellen behandeln zu können, werden diese parametrisiert. Allerdings sind diese Methoden eine vereinfachte Darstellung aufgrund von Grenzen der aktuellen Rechnerleistung oder das Fehlen des kompletten Prozessverständnisses. Ein weiteres großes Problem der QNV ist die Bereitstellung von sehr genauen Anfangsfeldern der Feuchteverteilung. Das Wasser in der Atmosphäre ist räumlich wie auch zeitlich sehr variabel und es ist schwierig, adäquate Beobachtungsdaten bereitzustellen, die ein relativ genaues Bild der Feuchteverteilung liefern können. Mit Hilfe von Methoden der Datenassimilation können diese Beobachtungsdaten in das Anfangsfeld der Modelle eingehen, sofern sie vorliegen.

Die vielen komplexen Abhängigkeiten der Prozesse, die die Niederschlagsbildung beeinflussen, können zu stark differenzierten und kleinskaligen Niederschlagsverteilungen führen. Aber auch Schauerereignisse im Winter stellen die NWV vor große Probleme, da diese eine Mehrzahl an Niederschlagstypen wie Schnee, gefrierender Regen oder Regen enthalten können, die ebenfalls parametrisiert sind. Konvektive Niederschlagsereignisse sind schwierig vorherzusagen, da sie sehr lokal und zeitlich begrenzt sein können sowie ein nichtlineares und chaotisches Verhalten (*Zawadzki*, 1973) aufweisen. Dadurch ist die Vorhersagbarkeit dieser Ereignisse stark limitiert (z.B. *Walser and Schär*, 2004; *Hohenegger et al.*, 2008). Besonders konvektive Situationen werden durch mesoskalige Prozesse dominiert, die auf Skalen ablaufen, die durch heutige NWV Modelle

nicht aufgelöst werden können. Um die Wechselwirkung dieser Vorgänge mit der skaligen, großräumigen, Strömung zu beschreiben, ist die Parametrisierung von hochreichender sommerlicher Feuchtkonvektion notwendig. Solche Parametrisierungen haben bei der Verwendung bei Maschenweiten von ca. 10 km erhebliche grundsätzliche Probleme (*Molinari and Dudek, 1992; Arakawa, 2004; Fritsch and Carbone, 2004*). Für die direkte Simulation von konvektiven Situationen ist es notwendig, höhere Auflösungen zu verwenden, wobei es strittig ist, welches Δx dazu benötigt wird. In der Studie von *Weisman et al. (1997)* konnte bei einer Gitterweite von 4 km gezeigt werden, dass die wichtigsten Eigenschaften und Charakteristiken von organisierter Konvektion explizit simuliert werden konnten. Dagegen wurde in der Publikation von *Bryan et al. (2003)* selbst bei einer Gitterweite von 125 m kein konsistentes Resultat für die Simulation einer 'Squall Line' erzielt. Bei *Skamarock (2004)* wurde anhand des Spektrums der kinetischen Energie (*Harris et al., 2001*) für verschiedene Versionen des WRF Modells (*Michalakes et al., 2001; $\Delta x = 22, 10$ und 4 km*) gezeigt, dass die effektive Auflösung dieses Modells bei ungefähr dem siebenfachen der Gitterweite ($7 \cdot \Delta x$) liegt. Damit kann erklärt werden, dass die explizite Simulation von hochreichender Konvektion mit einer Gitterweite von 4 km, trotz expliziter Behandlung nur ansatzweise für großräumige organisierte Konvektion möglich ist. Einer Verfeinerung der horizontalen Gitter muss ebenso eine bessere Beschreibung der Topographie einhergehen, da diese für die Simulation von Strömungen um Gebirge sowie des Niederschlags in der Nähe von strukturiertem Gelände entscheidend ist (z.B. *Benoit et al., 2002*). Zusätzlich müssen hochaufgelöste Beobachtungsdaten zur Verfügung stehen, damit der Anfangszustand bestmöglich bekannt ist.

1.1.1 Numerische Vorhersagemodelle

Nachdem die Herausforderungen der QNV beschrieben wurden, werden an dieser Stelle die drei NWV Modelle kurz vorgestellt, die weitestgehend für die Untersuchungen in dieser Arbeit verwandt wurden. Die NWV ist mathematisch betrachtet ein Anfangswertproblem und basiert auf den Grundlagen der Newton'schen Mechanik, so dass sich theoretisch die Zeitentwicklung der Atmosphäre bestimmen lässt. Damit ein Computer diese Aufgabe durchführen kann, muss die reale Atmosphäre auf ein Gitter abgebildet bzw. diskretisiert werden. Da die Gitterabstände nicht unendlich klein sein können aufgrund der aktuell verfügbaren Rechenleistung, kann diese Abbildung keine exakte Darstellung der Atmosphäre liefern. Die drei im Folgenden vorgestellten numerischen Atmosphärenmodelle umfassen alle aktuell für die operationelle Wettervorhersage genutzten deterministischen Modelltypen und stehen damit beispielhaft für ihren Modelltypus. Zuerst erfolgt die Beschreibung des deterministischen Globalmodells des Europäischen Zentrums für

Mittelfristvorhersagen (engl.: *European Centre for Medium Range Weather Forecasts*, ECMWF) und anschließend die beiden operationellen Regionalmodelle, COSMO-EU und COSMO-DE, des DWD.

Deterministisches Globalmodell des ECMWF

Das deterministische Globalmodell des ECMWF ist ein hydrostatisches, spektrales numerisches Vorhersagemodell. Dieses Modell wird seit 1979 operationell betrieben und dabei ständig weiterentwickelt und verbessert. Mit Hilfe dieses Modells kann aktuell die zeitliche Entwicklung von Wetterparametern wie Temperatur, Wind, Wasserdampf, Wolken und Niederschlag bis zu zehn Tagen im voraus bestimmt werden.

In dieser Arbeit werden Modellversionen mit zwei unterschiedlichen horizontalen und vertikalen Aufteilungen verwandt. Bis Ende Januar 2006 wurde die Modellversion mit der Bezeichnung T511 benutzt, was einer äquivalenten horizontalen Gitterweite von ca. 40 km entspricht. In der Vertikalen gibt es bis in eine Höhe von 0,1 hPa (ca. 65 km) 60 Modellschichten. Im Februar 2006 erfolgte eine Verbesserung der horizontalen Auflösung mit der Version T799, was einer horizontalen Maschenweite von ca. 25 km gleichzusetzen ist. Die Anzahl der vertikalen Schichten wurde auf 91 erhöht, wobei das oberste Niveau jetzt bei 0,01 hPa (ca. 80 km) liegt. In Abbildung 1.3 ist die Orographie des ECMWF Modells (Version T511) über Deutschland gezeigt. Die topographischen Charakteristiken sind sehr grob nachgestellt. Die Alpen sind als breites Band zu erkennen und besitzen nur eine Höhe von etwas mehr als 2000 m (reale Höhe des Mont Blanc ist 4807 m). Die

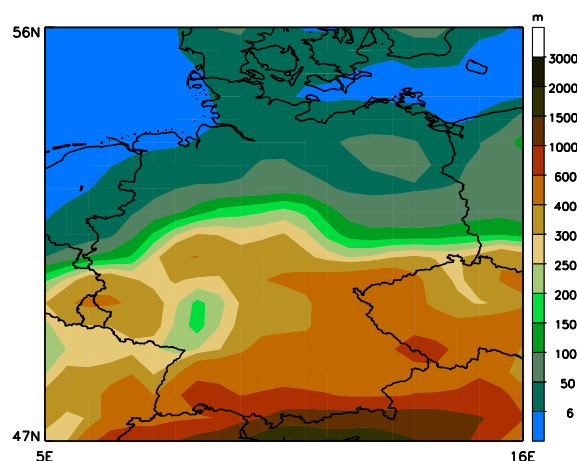


Abbildung 1.3. Topographische Höhe (m) des deterministischen Modells des ECMWF (Version T511) über Deutschland.

einzelnen deutschen Mittelgebirgsregionen sind nicht direkt dargestellt, so dass die Geländehöhe in Richtung der Nordsee stetig geringer wird. Einzige Ausnahme ist hier der nördliche Abschnitt des Oberrheingraben sowie das Mittelrheintal, die ansatzweise aufgelöst sind. Aufgrund der groben horizontalen Beschreibung der Atmosphäre durch dieses Modell müssen viele wichtige Prozesse, die sich auf Skalen abspielen die dieses Modell nicht beschreiben kann, parametrisiert werden. Dies hat zur Folge, dass mit diesem Modell keine direkte Simulationen von konvektiven Niederschlagsereignissen möglich sind.

COSMO-EU

Das Regionalmodell COSMO-EU wird seit Dezember 1999 operationell beim DWD betrieben und ist ein nichthydrostatisches Atmosphärenmodell, d.h. es besitzt eine prognostische Gleichung für die Vertikalbewegung. Diese Beschreibung wird notwendig, da die Maschenweite bei 7 km liegt und somit die Vertikal- und Horizontalbeschleunigungen von vergleichbarer Größe sind. Das Konsortium für kleinskalige Modellierung (*engl.: Consortium for Small-scale MOdelling*, COSMO) wurde 1998 gegründet, um ein nichthydrostatisches regionales numerisches Wettermodell zu entwickeln. Seither dient das COSMO-Modell den Mitgliedern sowohl in der operationellen Vorhersage als auch im Forschungsbetrieb. Das Modellgebiet umfasst annähernd ganz Europa sowie Teile Nordafrikas. Die Orographie des COSMO-EU ist in Abbildung 1.4 dargestellt. Im Gegensatz zum Modell des ECMWF (Abb. 1.3) sind die geografischen Strukturen, von der Norddeutschen Tiefebene über die Mittelgebirge bis hin zu den Alpen gut zu erkennen. Trotzdem entspricht die Orographie nicht ganz der Realität, so dass z.B. die Höhe der Alpen (ca. 3000 m) ebenfalls unterschätzt wird. Des Weiteren werden das Mittelrheintal und das Elbtal bei Dresden nicht komplett dargestellt. Die vertikale Darstellung der Atmosphäre wird durch 40 Schichten vorgenommen. Da das COSMO-EU ein Regionalmodell ist, benötigt es an den Rändern Informationen, die es beim DWD jede Stunde durch das Globalmodell GME (*Majewski et al., 2002*) erhält. Als prognostische Variablen werden Luftdruck, drei dreidimensionale Wind, Temperatur, Wasserdampf, Wolkenwasser, Wolkeneis, Regen und Schnee sowie die turbulente kinetische Energie bestimmt.

Trotz der gegenüber des ECMWF-Modells deutlich feineren Repräsentation der Atmosphäre müssen weiterhin noch eine Vielzahl an Prozessen, die auf kleineren Skalen stattfinden, parametrisiert werden. Die wichtigsten parametrisierten Prozesse beim COSMO-EU (*Schulz and Schättler, 2005*) sind:

- skaliger Niederschlag (Wolkenmikrophysik nach *Kessler, 1969*)
- Feuchtkonvektion (Massenflussschema nach *Tiedtke, 1989*)

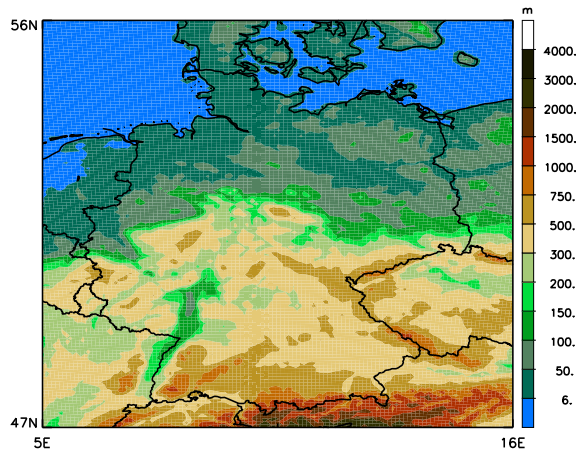


Abbildung 1.4. Topographische Höhe in m des COSMO-EU über Deutschland.

- Strahlung (Schema nach *Ritter and Geleyn, 1992*)
- partielle Bewölkung
- vertikale turbulente Flüsse (*Mellor and Yamada, 1974*)
- Prozesse in den obersten Schichten des Erdbodens

Die Repräsentation der Details der Orographie im COSMO-EU die einen Einfluss auf das Wettergeschehen haben, sollen es diesem Modell ermöglichen die Effekte der Land-Seewindzirkulation an den Küsten oder die Modifikation von Wolken und Niederschlägen an den Mittelgebirgen zu simulieren. Um diese und andere kleinräumige, wetterrelevante meteorologische Strukturen zu erfassen, besitzt das COSMO-EU eine Vorhersagezeit von bis zu 72 Stunden bzw. drei Tagen.

COSMO-DE

Seit April 2007 wird das COSMO-DE beim DWD operationell betrieben und besitzt eine horizontale Maschenweite von 2,8 km und 50 vertikale Schichten. Das COSMO-DE ist eine Anwendung des COSMO-Modellsystems, die insbesondere eine verbesserte Vorhersageleistung bei gefährlichen Wetterlagen erzielen soll, die im Zusammenhang mit hochreichender Feuchtkonvektion stehen (z.B. Super- und Multizellengewitter, Böenwalzen, mesoskalige konvektive Komplexe) oder aufgrund von Wechselwirkungen mit der feinskaligen Topographie (Bodennebel, Föhnstürme, Sturzfluten, etc.) auftreten (*Baldauf et al.,*

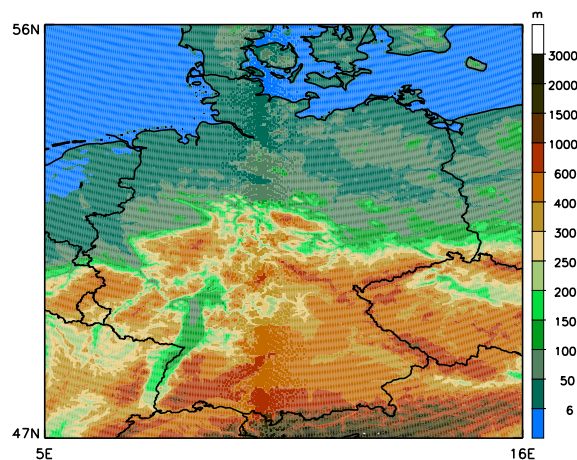


Abbildung 1.5. Topographische Höhe des COSMO-DE in m über Deutschland.

2010). In einem Modell das Vorhersagen auf einer solchen räumlichen wie zeitlichen Skala liefern soll, muss die hochreichende Feuchtkonvektion explizit simuliert werden. Darin besteht der wesentlichste Unterschied zum COSMO-EU, wobei die flache, nicht regnende, Konvektion weiterhin parametrisiert wird. Bei der Wolkenmikrophysikparametrisierung wird eine zusätzliche Klasse, der Graupel, mit berücksichtigt.

Das Modellgebiet umfasst neben Deutschland, auch die Schweiz und Österreich sowie Teile der übrigen Nachbarstaaten. Die Orographie im COSMO-DE über Deutschland ist in Abbildung 1.5 gezeigt. Bei einer Maschenweite von 2,8 km liefert die Topografie des COSMO-DE ein relativ realistisches Bild. Es werden viele Täler der Mittelgebirge und der Alpen (max. Höhe ist 4000 m) wesentlich genauer erfasst als beim COSMO-EU, so dass die Strömungsverhältnisse etwas realistischer simuliert werden sollten.

Das COSMO-DE erhält stündlich Randdaten vom COSMO-EU und es wird alle drei Stunden ein neuer Lauf gestartet, d.h. acht Läufe täglich. Die Länge der Vorhersageläufe beträgt 21 h. Da es für eine Kurzfristvorhersage notwendig ist, den Anfangszustand besonders gut zu kennen, enthalten die ersten drei Stunden einen Assimilationszyklus von Radardaten mittels eines Verfahrens, das 'Latent Heat Nudging' (LHN, Jones and Macpherson, 1997; Stephan et al., 2008) genannt wird. Die eigentliche Vorhersagelänge beträgt somit für jeden Lauf 18 Stunden.

1.1.2 Beobachtungsdaten

Die extrem heterogene räumliche sowie zeitliche Verteilung der meteorologischen Größe Niederschlag (z.B. Ebert et al., 2003) ist die Ursache dafür, dass selbst die Beobachtung

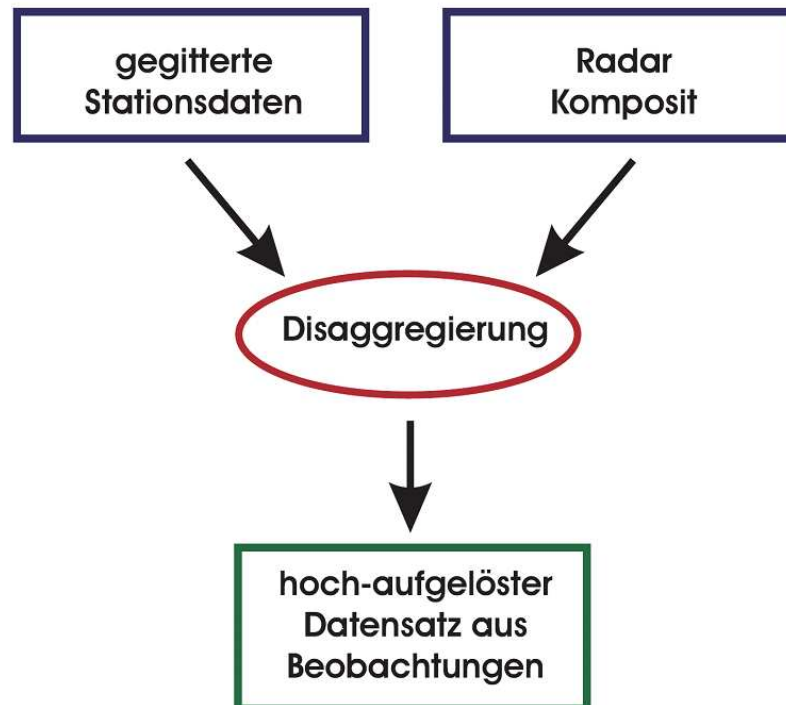


Abbildung 1.6. Schematische Darstellung der Disaggregation (Abb. 3.3 aus Paulat, 2007).

eine große Herausforderung darstellt. Besonders schwierig ist dies während des Sommers, wenn der Niederschlag konvektiv geprägt ist. Konvektive Niederschlagsereignisse sind häufig räumlich eng begrenzt (einige Kilometer im Durchmesser) mit einer Dauer von ca. einer Stunde und weisen dabei ein chaotisches Verhalten auf (Zawadzki, 1973). Ein Beispiel hierfür ist ein Niederschlagsereignis während COPS (*engl.: Convective and Orographic Precipitation Study*, Wulfmeyer et al., 2008; Wulfmeyer et al., 2010), als ein intensives und lokales konvektives Niederschlagsereignis während der IOP8b (*engl.: Intensive Observation Period*, IOP) im Schwarzwald auftrat. An keiner der bodengebundenen Niederschlagsmessstationen wurde Niederschlag registriert. Allerdings konnte mit Radarmessungen die Zelle deutlich erkannt und lokalisiert werden.

Der in dieser Arbeit zur Verifikation verwendete Beobachtungsdatensatz mit einer stündlichen Auflösung für Deutschland nutzt deshalb eine Kombination zweier Messstrategien wie in der Dissertation von Marcus Paulat beschrieben (Paulat, 2007; Paulat et al., 2008). In der vorliegenden Arbeit wird der Datensatz um vier Jahre auf jetzt acht Jahre erweitert (2001-2008). Im Folgenden werden die Definition und die Erstellung kurz beschrie-

Zeitraum	2001	2002	2003	2004	2005	2006	2007	2008
# Stationen	3964	3731	3448	3303	3081	2763	2215	2159

Tabelle 1.1. Anzahl von täglichen Stationsmessungen des Niederschlags durch den DWD im Zeitraum von 2001-2008.

ben.

Zur Erstellung des Datensatzes werden tägliche Stationsbeobachtungen mit stündlichen Radarkompositdaten kombiniert. In Abbildung 1.6 ist dieser Vorgang schematisch dargestellt. Die Idee, die dieser Methode zugrunde liegt, ist, die Vorteile beider Messmethoden zu nutzen und in einem Produkt zu vereinen. Die relativ genaue Messung der Menge durch die Niederschlagstöpfe sowie die ziemlich gute zeitliche Auflösung der Radarinformationen ermöglichen die Erstellung eines räumlich und zeitlich hochaufgelösten Beobachtungsdatensatz. Diese Vorgehensweise wird Disaggregation (*Binder and Schär, 1996*) genannt.

Zuerst wird aus den Stationsdaten unter Verwendung der Methode von *Frei and Schär (1998)* ein gegitterter Datensatz mit täglichen Summen erzeugt, der auf einem Gitter mit einer Maschenweite von 7 km vorliegt. In Tabelle 1.1 ist die Anzahl der Messstationen des DWD während des Achtjahreszeitraums (2001 ca. 4000 und 2008 ca. 2200) rückläufig, allerdings existiert weiterhin eine relativ hohe räumliche Abdeckung mit Messstationen. Aus dem Radarnetzwerkverbund des DWD mit seinen 16 Standorten wird aus den Reflektivitätsraten (Z , unterteilt in sieben Klassen; *Paulat et al., 2008*) unter Zuhilfenahme sogenannter empirisch ermittelter Z/R -Beziehungen die Regenrate bestimmt. Die stündlichen Radarregensummen werden durch Mittlung auf das selbe Gitter wie die Stationsdaten transformiert. Somit liegt für jeden Gitterpunkt die Tagessumme aus Stationsbeobachtungen sowie der Tagesgang aus den Radarbeobachtungen vor.

Die mathematische Formulierung der Disaggregation lautet:

$$R_{dis,h}(i, j) = R_{obs,d}(i, j) \cdot \frac{R_{rad,h}(i, j)}{R_{rad,d}(i, j)} \quad (1.1)$$

mit $R_{dis,h}$ der 1 h disaggregierten Summe, $R_{obs,d}$ der 24 h Summe der Stationsbeobachtungen, $R_{rad,h}$ der 1 h Radarsumme und $R_{rad,d}$ der 24 h Radarsumme an jeden Gitterpunkt (i, j) . Mit Hilfe des relativen Tagesgangs aus den Radarmessungen wird der Niederschlag der Stationsbeobachtungen anteilig über den Tag verteilt. Dabei können wie in *Paulat et al. (2008)* gezeigt, einige Fehler auftreten, wo die Disaggregation nicht richtig arbeitet. Zum Abschluss muss noch angemerkt werden, dass das oben beschriebene konvektive Ereignis IOP8b im disaggregierten Datensatz nicht enthalten ist, da dieses

Ereignis lokal sehr begrenzt war und an keiner Bodenstation festgestellt wurde. Für die mittlere Niederschlagsverteilung bzw. die statistische Analyse von Zeitreihen dürfte dies nicht sehr stark ins Gewicht fallen, allerdings muss dies bei Fallstudien, die lokale Konvektionsereignisse untersuchen, berücksichtigt werden.

1.2 Methoden zur Niederschlagsverifikation

Die Verifikation stellt einen wichtigen Punkt im Bereich der numerischen Wettervorhersage dar. Weshalb sie wichtig ist sowie als Motivation dient ein Zitat aus dem Editorial der Sonderausgabe für die Verifikation von numerischen Niederschlagsvorhersagen (*Burt et al., 2008*) in *Meteorological Applications*:

„It has been said that a forecast is like an experiment – given a set of conditions, one makes a hypothesis that a certain outcome will occur. The experiment is not complete until its outcome is determined, either proving or disproving the hypothesis. This is the act of forecast verification, i.e., determining whether the forecast was successful.“

Wird die numerische Wettervorhersage also als eine Art physikalisches Experiment betrachtet, so ist dieses erst dann abgeschlossen, wenn die Ergebnisse bzgl. ihrer Qualität bewertet sind. Diese Betrachtung ist völlig legitim, da sie auf der Lösung der Newton'schen Bewegungsgleichungen basiert und mathematisch gesehen ein Anfangswertproblem darstellt. Der Begriff der Verifikation leitet sich vom lateinischen Wort *verus* (wahr) ab, so dass in der Meteorologie die vorhergesagten Wetterelemente gegenüber der Wahrheit verglichen werden.

Die Kenntnis der Qualität des verwendeten Vorhersagesystems ist ein wichtiger Baustein bei der Nutzung der Vorhersagen. Darüber hinaus ist es interessant zu wissen, ob sich Veränderungen des Modellsystems positiv auf die Vorhersagequalität ausgewirkt haben. Die Verifikation kann Hinweise auf typische Fehler geben, so dass diese eliminiert werden können. Diejenigen die verantwortlich für die Weiterentwicklung der Wettermodelle sind, haben ein Interesse zu wissen, ob sich der dafür notwendige finanzielle Aufwand in der Qualität der Vorhersagen widerspiegelt. Dementsprechend ist für sie ein Vergleich mit Modellvorhersagen anderer Wetterdienste sehr interessant. Damit wird klar, dass verschiedene Nutzer der Vorhersagen auch an unterschiedlichen Aspekten der Vorhersagequalität interessiert sind, z.B. ist ein Modellentwickler daran interessiert welche typischen Fehler auftreten, wohingegen der Direktor eines Wetterdienstes wissen möchte, ob das Modell

besser geworden ist. Das Design des Verifikationssystems hängt ebenfalls stark vom Nutzer ab und sollte deshalb auf dessen Bedürfnisse/Interessen abgestimmt sein.

Auch die Frage was eine gute Vorhersage ausmacht, ist nicht einfach zu beantworten. Hierzu hat *Murphy* (1993) ein Essay vorgelegt und zwischen drei Arten unterschieden:

- **Stetigkeit** – Inwieweit stimmen die Aussagen eines Prognostikers mittels bekannter Informationen mit der eines Vorhersagesystems überein?
- **Qualität** – Inwiefern stimmen Vorhersage und Beobachtung (Wahrheit) überein?
- **Wert** – Wie groß ist der Effekt der Vorhersageinformation auf Entscheidungsträger, um einen deutlichen Nutzen aus der Vorhersage zu gewinnen?

In dieser Arbeit werden Vorhersagen im Sinne von „Qualität“ bewertet. Bei *Murphy* (1993) werden der „Qualität“ neun sogenannte Attribute zugeordnet. Welche dieser Eigenschaften für den Nutzer am wichtigsten ist, entscheidet welche Verifikationsstrategie zur Anwendung kommt.

1.2.1 Gitterpunktbezogene Bewertungsmaße

Zur Bewertung von deterministischen Wettervorhersagen wurden in den letzten annähernd 120 Jahren eine große Anzahl an Maßen entwickelt und vorgestellt. Die erste Veröffentlichung, die die Verifikation zum Thema hatte, war *Finley* (1884) in der die Vorhersage von Tornados für weite Gebiete der USA untersucht wurde. Mit dem Beginn der numerischen Wettervorhersage in den 1950er Jahren erfolgte eine stärkere Anwendung der Verifikation auf eine Vielzahl an Vorhersagen. Der Vergleich bei diesen Maßen erfolgt Gitterpunkt für Gitterpunkt und werden häufig traditionelle oder klassische Maße genannt. Grundsätzlich können für die Verifikation von deterministischen Vorhersagen zwei Typen von Variablen, kontinuierliche und kategorische, unterschieden werden, wobei für die Niederschlagsverifikation kategorische Maße häufiger zum Einsatz kommen. Anschließend werden zuerst einige ausgewählte kontinuierliche Maße und danach Beispiele für kategorische Maße vorgestellt. Einen guten Überblick über die verschiedenen klassischen Maße liefern die Textbücher von *Stanski et al.* (1989), *Jolliffe and Stephenson* (2003) und *Wilks* (2006).

Kontinuierliche Maße

Zunächst werden Gütemaße von kontinuierlichen Variablen vorgestellt, d.h. Variablen die einen spezifischen Wert oder Wertebereich annehmen können (z.B. einem Temperaturwert von 25,2°C oder eine Niederschlagssumme von 16 mm in 24 Stunden). Die Bewertung bei diesen Maßen basiert auf lokalen Differenzen von beobachteter und simulierter Niederschlagsmenge.

Mittlerer Fehler

Als erstes kontinuierliches Maß wird der mittlere Fehler (*engl.: Mean Error, ME*) vorgestellt. Er ist definiert als die mittlere Differenz zwischen Vorhersagen $R_{mod}(i, j)$ und Beobachtungen $R_{obs}(i, j)$ für alle Gitterpunkte innerhalb des betrachteten Untersuchungsgebiets:

$$ME = \frac{1}{N} \sum_{(i,j) \in D} [R_{mod}(i, j) - R_{obs}(i, j)], \quad (1.2)$$

mit N der Anzahl der Gitterpunkte im betrachteten Gebiet D . Für eine perfekte Vorhersage ist der ME gleich null. Positive bzw. negative Werte (in mm) sind mit Über- bzw. Unterschätzung der Niederschlagsmenge durch die Vorhersagen verbunden. Der ME ist ein absolutes Maß zur Beschreibung des Vorhersagefehlers und somit ist der Wert immer abhängig von der Hintergrundklimatologie, d.h. bei größeren Niederschlagsmengen ist zu erwarten, dass der Wert des ME auch höher ist. Damit lässt sich aufgrund des ME keine allgemeine Aussage über die Vorhersagequalität ableiten.

Wurzel des mittleren quadratischen Fehlers

Ein weiteres häufig gebrauchtes kontinuierliches Maß ist die Wurzel des mittleren quadratischen Fehlers ($RMSE$; *engl.: Root Mean Square Error*). Der $RMSE$ wird folgendermaßen berechnet:

$$RMSE = \left\{ \frac{1}{N} \sum_{(i,j) \in D} [R_{mod}(i, j) - R_{obs}(i, j)]^2 \right\}^{0,5}, \quad (1.3)$$

und ist immer positiv (in mm). Mit dem $RMSE$ wird keine Information über die Richtung des Fehlers angegeben, allerdings werden durch das Quadrieren große Differenzen zwischen Beobachtung und Vorhersagen stärker gewichtet als kleinere. Häufig wird auch der mittlere quadratische Fehler (*engl.: Mean Square Error, MSE = RMSE²*) verwendet.

Außerdem kann ein normierter $RMSE$ in der Form $nRMSE = RMSE/D(R_{obs})$ berechnet werden, mit $D(R_{obs})$ dem gebietsgemittelten Niederschlagswert der Beobachtungen. Mit Hilfe des $nRMSE$ können Niederschlagsvorhersagen von schwachen und intensiven Situationen miteinander verglichen werden.

Reduktion der Varianz

Das dritte hier vorgestellte kontinuierliche Maß ist die so genannte Reduktion der Varianz (RV). Die RV ist ein 'Skill Score' ($SKSC$), d.h. sie gibt an; wie groß die prozentuale Verbesserung des Vorhersagesystems gegenüber einer meist kostengünstigeren Referenzvorhersage (z.B. Klimatologie, Persistenz, Zufall, etc.) ist. Die grundsätzliche Form eines $SKSC$ lautet:

$$SKSC = \frac{SC - SC_{Ref}}{SC_{Perfekt} - SC_{Ref}} \cdot 100 \quad (1.4)$$

wobei SC ein beliebiges Maß ist und $SC_{Perfekt}$ der Idealwert des verwendeten Maßes darstellt. Im Fall, dass der ideale Wert des verwendeten Gütemaßes null ($SC_{Perfekt} = 0$) ist, lässt sich der $SKSC$ umformen in:

$$SKSC = \frac{SC_{Ref} - SC}{SC_{Ref}} \cdot 100 = \left(1 - \frac{SC}{SC_{Ref}}\right) \cdot 100$$

Die RV ist bei Verwendung des $RMSE$ als Grundbewertungsmaß wie folgt zu berechnen:

$$RV = \left(1 - \frac{RMSE}{RMSE_{Ref}}\right) \cdot 100 \quad (1.5)$$

Der Wertebereich ist zwischen $-\infty$ und 100. Ein Wert von 100 bedeutet ein perfekte Vorhersage, positive Werte geben einen Gewinn gegenüber der Referenzvorhersage an und für einen Wert von Null ist die Referenzvorhersage genauso gut wie die Vorhersage. Bei negativen Werten ist die Referenzvorhersage besser. Häufig wird die Reduktion der Varianz im Intervall von $-\infty$ und 1 angeben d.h. ohne eine Skalierung mit dem Faktor 100.

Kategorische Maße

Nachdem einige kontinuierliche Maße vorgestellt wurden erfolgt die Beschreibung von drei kategorischen Maßen. Mit Hilfe der Kategorisierung einer Variable kann das Auftreten bzw. Nichtauftreten eines bestimmten Ereignisses, wie etwa Niederschlag ja oder

Vorhersage	Beobachtung	
	Ja	Nein
Ja	a (Treffer)	b (Fehlalarm)
Nein	c (Verpassen)	d (korrekte Ablehnung)

Table 1.2. Die vier möglichen Ergebnisse für kategorisierte Vorhersagen eines binären Ereignisses: a (Treffer), b (Fehlalarm), c (Verpassen) und d (korrekte Ablehnung).

nein bewertet werden. Alle kontinuierlichen Variablen können kategorisiert werden, in dem bestimmte Schwellenwerte für das Über- bzw. Unterschreiten definiert werden. Als Beispiel kann hierfür das Überschreiten einer Niederschlagssumme von 1 mm in 24 Stunden dienen. In diesem Fall ist eine binäre Kategorisierung, d.h. eine Ja-Nein-Aussage, durchzuführen.

Eine nützliche Zusammenfassung des Überschreitens des Schwellenwerts von beobachteten und vorhergesagten Niederschlagsereignissen kann in Form einer Kontingenztabelle präsentiert werden. Sie stellt kein direktes Verifikationsmaß dar, ist allerdings die Basis für die Definition aller kategorischen Bewertungsmaße. Eine Kontingenztabelle für eine binäre Kategorisierung kann in folgender Art und Weise aufgestellt werden (Tabelle 1.2). Mit a wird die Anzahl der Ereignisse angegeben, die sowohl beobachtet als auch vorhergesagt wurden und mit d diejenigen die weder beobachtet noch vorhergesagt wurden. Die Summe von a und d gibt die Anzahl der korrekten Aussagen für den betrachteten Schwellenwert an. Dagegen werden mit b und c die falschen Vorhersagen gekennzeichnet. Die unterschiedlichen aus der Kontingenztabelle abgeleiteten Gütemaße bewerten und charakterisieren unterschiedliche Eigenschaften von Vorhersagen. An dieser Stelle werden die in dieser Arbeit verwendeten bzw. angesprochenen kategorischen Maße vorgestellt und beschrieben.

Häufigkeitsbias

Der Häufigkeitsbias (*engl.: Frequency Bias, FBI*) gibt das Verhältnis von vorhergesagten zu beobachteten Ereignissen an, ohne etwas über deren Genauigkeit, Güte oder deren räumliche Verteilung zu vermitteln. Der *FBI* ist definiert als das Verhältnis von positiver Vorhersage zu positiver Beobachtung:

$$FBI = \frac{a + b}{a + c}. \quad (1.6)$$

Der *FBI* nimmt Werte von 0 bis ∞ an. Wenn die vorhergesagten Niederschlagsereignisse gleich häufig wie die beobachteten Niederschlagsereignisse sind, so hat der *FBI* den per-

fechten Wert Eins. Bei Werten des FBI s kleiner als Eins, werden zu wenig und bei Werten größer als Eins zu viele Ereignisse vorhergesagt.

Heidke Skill Score

Der Heidke Skill Score (HSS ; *Doolittle*, 1888; *Heidke*, 1926) ist ein sehr häufig verwendetes Maß und ist wie der Name schon andeutet ein $SKSC$, d.h. er bewertet die Vorhersage bzgl. einer Alternativvorhersage. Im Fall des HSS wird der Zufall als Referenzvorhersage verwandt und die Trefferrate $TR = (a + d) / (a + b + c + d)$ dient als Grundbewertungsmaß. Damit lässt sich der HSS mit Hilfe der allgemeinen Definition eines $SKSC$ folgendermaßen schreiben:

$$HSS = \frac{TR - TR_{Ref}}{1 - TR_{Ref}} \quad (1.7)$$

Die Trefferrate der Zufallsvorhersage ergibt sich aus der Summe der Wahrscheinlichkeit der zufällig korrekt vorhergesagten Ereignisse und Nichtereignisse:

$$TR_{Ref} = \frac{(a + b)(a + c)}{n^2} + \frac{(b + d)(c + d)}{n^2}$$

mit $n = a + b + c + d$. Somit lässt sich der HSS ausschließlich durch Einträge der Kontingenztafel darstellen:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}. \quad (1.8)$$

Der Wertebereich des HSS befindet sich im Intervall von -1 bis $+1$ mit Eins als Wert für eine perfekte Vorhersage. Bei einem Wert des HSS von null ist die Zufallsreferenzvorhersage gleich gut und bei negativen Werten sogar besser als die Vorhersage.

Equitable Thread Score

Ein weiterer $SKSC$ ist der Equitable Thread Score (ETS), der auch als Gilbert Skill Score *Gilbert* (1884) bekannt ist und der als Grundbewertungsmaß den Thread Score ($TS = a / (a + b + c)$) verwendet. Wie im Fall des HSS dient der Zufall als Referenzvorhersage und besitzt in diesem Fall folgende Form:

$$TS_{Ref} = \frac{(a+b)(a+c)}{a+b+c+d}$$

Durch Einsetzen in die allgemeine Form des SKSC 1.2.1 ergibt sich für den ETS

$$ETS = \frac{a/(a+b+c) - TS_{Ref}}{1 - TS_{Ref}} = \frac{ad - bc}{n(b+c) + (ad - bc)} \quad (1.9)$$

wobei $n = a + b + c + d$ ist. Der maximale Wertebereich ist zwischen $-\frac{1}{3}$ und $+1$ definiert und hat mit 1 den perfekten Wert. Bei negativen Werten ist die Referenzvorhersage der Modellvorhersage vorzuziehen. Das Minimum des *ETS* ist abhängig von der beobachteten Niederschlagsklimatologie und je seltener es ist, desto näher liegt dieser Wert an null. Außerdem nähert sich der Wert des *ETS* für seltene Ereignisse dem *TS* an.

Die beiden in diesem Abschnitt vorgestellten kategorischen *SKSCs* (*HSS* sowie *ETS*), aber auch weitere können durch die Einträge der Kontingenztabelle berechnet werden. Darüber hinaus sind die *SKSC* proportional zum Term $ad - bc$, d.h. der Differenz aus dem Produkt der korrekten zu den falschen Aussagen und stehen somit mathematisch in Beziehung zueinander (z.B. Wilks, 2006). Der mathematische Zusammenhang zwischen beiden hier vorgestellten Maßen lautet: $ETS = HSS / (2 - HSS)$. Daraus ist abzulesen, dass der *ETS* immer geringere Werte als der *HSS* annehmen wird, außer im Fall einer perfekten Vorhersage $ETS = HSS = 1$. Allerdings wird der qualitative Verlauf der beiden Variablen gleich sein.

1.2.2 Räumliche Methoden

Die traditionellen gitterpunktbezogenen Methoden sind seit vielen Jahrzehnten in Gebrauch, allerdings sind ihre Ergebnisse problematisch zu interpretieren, wenn die Modelle der neuesten Generation mit sehr feinen horizontalen Gittern verifiziert werden. Eine Untersuchung der QNVs dieser Modelle scheint subjektiv realistischere Ergebnisse zu liefern. Allerdings sind die Werte der traditionellen Maße für diese Modelle schlechter als bei gröber aufgelösten Modellen (Rossa et al., 2008). Die Ursache dafür wird anhand von Abbildung 1.7 verdeutlicht. Ist ein Niederschlagsereignis in der Vorhersage gegenüber der Beobachtung versetzt (Abb. 1.7a-d), so wird durch die traditionellen Maße (z.B. *RMSE*) die Vorhersage zweimal bestraft. Zum einen an der Stelle wo kein Niederschlag simuliert aber beobachtet wird und zum zweiten an der Stelle wo welcher simuliert wird aber keine Beobachtung stattfindet. Dieses Phänomen wird 'Double Penalty' genannt und ist ein

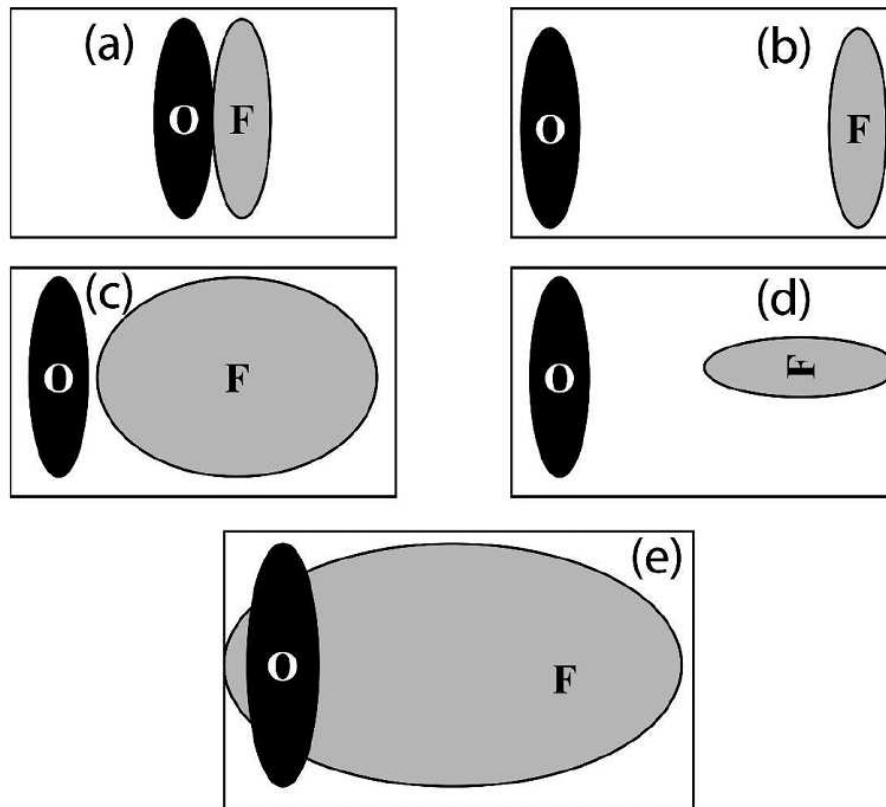


Abbildung 1.7. Ein schematisches Beispiel verschiedener Kombinationen von Vorhersage- und Beobachtungsfeldern zur Veranschaulichung der Problematik gitterpunktbezogener Verifikation. (Abb. aus Davis et al., 2006)

grundsätzliches Problem der traditionellen Methoden (Jolliffe and Stephenson, 2003). Die subjektive Analyse der fünf in Abbildung 1.7 dargestellten Paare führt zum Ergebnis, dass die Vorhersage im Teilbild (a) am besten ist, da nur eine geringe Verschiebung zwischen Beobachtung und Vorhersage besteht. Die Bewertung durch den *ETS* liefert für (a)-(d) einen Wert von Null, da keine Überlappung zwischen Vorhersage und Beobachtung existiert, aber für (e) einen Wert größer Null, d.h. es ist die gemäß *ETS* beste Vorhersage. Um diese 'Double Penalty'-Problematik zu beheben, wurden in den letzten Jahren verschiedene Verfahren entwickelt. Diese Methoden berücksichtigen die räumliche Verteilung der Vorhersagefelder und wollen außerdem eine Aussage über die physikalischen Ursachen der Fehler, sowie ergänzende Aussagen zu den Informationen der traditionellen Verifikationsmethoden liefern. Eine gute Übersicht über das gesamte Spektrum der Methoden kann Casati et al. (2008) entnommen werden. Aber auch Rossa et al. (2008), Ahijevych et al. (2009) sowie Gilleland et al. (2009) geben einen guten Überblick über die neu entwickel-

ten Verifikationsverfahren. In *Casati et al.* (2008) und *Gilleland et al.* (2009) werden die meisten dieser neuartigen Ansätze in vier Kategorien unterteilt: (i) Skalentrennung (oder -zerlegung), (ii) Umgebungs- oder Fuzzymethoden, (iii) Feldverformung und (iv) merkmalsbezogen (oder objekt-orientiert). Die ersten beiden Typen können verallgemeinert als Filter-Methoden bezeichnet werden, wobei die Methoden mit Skalentrennung mithilfe von Bandpassfiltern versuchen, die Vorhersageleistung auf verschiedenen physikalischen Skalen zu bewerten. Im Gegensatz dazu werden bei den Fuzzymethoden die Niederschlagsfelder geglättet werden. Vergleichbar dazu können die merkmalsbezogenen und die Methoden mit Feldverformung als räumliche Verschiebungsmethoden zusammengefasst werden, obwohl sie mehr Informationen über die Vorhersagegüte liefern als nur die reine Verschiebung. An dieser Stelle werden die Eigenschaften der vier Kategorien beschrieben und einige beispielhafte Verfahren angegeben.

(i) Ansätze mit Skalentrennung

Die grundsätzliche Idee dieser Methoden ist die Aufteilung von Beobachtung und Vorhersage in die Summe ihrer räumlichen Komponenten auf verschiedenen Skalen durch Verwendung von Bandpassfiltern (z.B. Fourier, Wavelets, etc.). Die Verifikation auf verschiedenen Skalen kann einen hilfreichen Einblick in die Charakteristik von NWV Modellen geben, um die mit verschiedenen physikalischen Prozessen verbundenen Phänomene auf unterschiedlichen Skalen zu bewerten. Der Begriff Skalen ist in diesem Zusammenhang mit physikalischen Prozessen verbunden wie großskalige Frontensysteme oder kleinräumige konvektive Schauer. Es kann z.B. angegeben werden, ab welcher räumlicher Skala der Übergang zwischen nützlichem und unnützlichem Bereich liegt. Beispiele der Anwendung von Wavelets liefern z.B. *Briggs and Levine* (1997), *Casati et al.* (2004) und *Casati and Wilson* (2007) um dann traditionelle Maße auf den verschiedenen Skalen zu bestimmen. Eine weitere Möglichkeit die Güte auf unterschiedlichen Skalen zu verifizieren, haben z.B. *Harris et al.* (2001) vorgestellt. Dabei wird die Fähigkeit der Vorhersage überprüft, beobachtete räumlich-zeitliche Strukturen der verschiedenen Skalen zu simulieren, durch die Verwendung von skaleninvarianten Parametern (z.B. die spektrale Energie).

(ii) Fuzzymethoden

Fuzzy Verifikationstechniken belohnen geringe Fehler durch eine lockere Handhabung der exakten Zuordnung zwischen Vorhersagen und Beobachtung. In *Ebert* (2008) wird eine Grundstruktur dieser Methoden unter Verwendung mehrerer Methoden beschrieben.

Einige dieser Verfahren bestimmen deterministische Standardmaße mit einer allgemeineren Definition für einen Treffer. Das Grundprinzip dieser Ansätze ist die Verwendung eines räumlichen Fensters der vorhergesagten und/oder beobachteten Felder, so dass gefilterte Felder entstehen. Meist bleiben die Spitzenwerte bei dieser Filterung erhalten, um die Bewertung von Extremereignissen zu ermöglichen. Durch iterative Vergrößerung der Fuzzyregion ist es möglich, Informationen über die Güte auf verschiedenen Skalen zu erhalten. In diesem Sinn ist der Begriff 'Skalen' anders zu verstehen als im Fall der Methoden der Skalentrennung, da hier die Skalen nicht unabhängig voneinander sind. Mit größer werdender Skala wird das gesamte Feld immer glatter, was in bessere Werte mündet.

(iii) Ansätze mit Feldverformung

Die erste Anwendung dieser Methoden in der meteorologischen Verifikation erfolgte durch *Hoffman et al.* (1995). Dieses Verfahren versucht, das vorhergesagte Feld dermaßen nach festgelegten Vorgehensweisen zu manipulieren, dass es optimal zum beobachteten passt und unterteilt den Fehler in Verschiebung, Amplitude und Residuum. Der Verschiebungs- und der Amplitudenfehler entsprechen Transformationen des Vorhersagefeldes durch Anwendung eines Geschwindigkeitsfeldes und eines skalaren Feldes, bis ein Optimierungskriterium (z.B. Maximierung der räumlichen Korrelation) erfüllt ist. Dieser Ansatz stellte einen Meilenstein für die Anwendung räumlicher Verifikationsmethoden dar, da hiermit der Fehler in physikalischen Einheiten (z.B. Verschiebung in Kilometer) angegeben werden konnte. Die Verifikationsergebnisse sind einfacher als die der traditionellen Maße zu interpretieren und beziehen sich auf spezifische physikalische Prozesse der Simulation (z.B. das Advektionsschema). Dieser Ansatz führte zu vielen weiteren daraus abgeleiteten Methoden. Bei der Methode des optischen Flusses (*engl.: Optical Flow*, *Mannstein et al.*, 2002), welche bei der automatischen Bilderkennung verwendet werden, werden die Änderungsvektoren zwischen Beobachtung und Vorhersage bestimmt, um den Verschiebungs- und den Amplitudenanteil am Fehler zu quantifizieren. Eine von vielen Anwendungen dieser Methode ist das Verfahren *DAS* (*Keil and Craig*, 2009).

(iv) Merkmalsbezogene Methoden

Diese Verfahren versuchen grundsätzlich besondere Strukturen oder Merkmale im Vorhersage- wie Beobachtungsfeld zu identifizieren, um anschließend einen Vergleich zwischen deren Eigenschaften (z.B. Position, Größe, Intensität, Orientierung, etc.) durchzuführen. Die Methoden unterschieden sich hauptsächlich durch den verwendeten Algorithmus zur Identifikation der Merkmale. Beispiele für diesen Typ sind *Ebert and McBride*

(2000) oder *Davis et al.* (2006). Einige Methoden passen nicht ganz zur Definition dieser Kategorie, *Nachamkin* (2004) untersucht die Verteilung von beobachteten zu simulierten Ereignissen und *Marzban and Sandgathe* (2006) verwenden Clusteranalysen.

Zwei der merkmalsbezogenen Verfahren werden im Detail vorgestellt, da auf sie an späterer Stelle in der Arbeit zurückgegriffen wird und sie eine ähnliche Vorgehensweise besitzen wie die in dieser Arbeit verwendete und weiterentwickelte Methode SAL (siehe Kapitel 3.1). Zuerst erfolgt die Beschreibung der CRA Methode (*Ebert and McBride*, 2000) und danach die Technik MODE (*Brown et al.*, 2004; *Davis et al.*, 2006; *Davis et al.*, 2009).

Die CRA Methode

Die merkmals-orientierte CRA Methode (*Ebert and McBride*, 2000) beschreibt den Verschiebungsfehler von „überlappenden“ Niederschlagsgebieten (*engl.: Contiguous Rain Area, CRA*) und zerlegt den gesamten Fehler in drei Komponenten, Verschiebung (*engl.: Displacement*), Menge (*engl.: Volume*) und Muster (*engl.: Pattern*). Sie ist damit angelehnt an die Methode von *Hoffman et al.* (1995). Zur Bestimmung der CRAs werden das beobachtete und vorhergesagte Niederschlagsfeld übereinandergelegt. Ein CRA ist begrenzt durch eine gemeinsame geschlossene Kontur, die durch einen festen Schwellenwert (z.B. 5 mm/24h) identifiziert werden kann. Daraus folgt, dass Beobachtung und Vorhersage direkt zugeordnet sind.

Die Bestimmung des Verschiebungsfehlers der CRAs der Vorhersage erfolgt durch so langes räumliches Verschieben bis der mittlere quadratische Fehler (*engl.: Mean Square Error, MSE*) minimal wird. Der Verschiebungsvektor zwischen Beobachtung und Vorhersage zeigt die Stärke und Richtung des Verschiebungsfehlers an. In Abbildung 1.8 ist das CRA (dicke schwarze Linie) und der Verschiebungsfehler (Pfeil) für ein idealisiertes Beispiel kreisförmiger Niederschlagsobjekte dargestellt.

Nach der Bestimmung des Verschiebungsfehler kann der gesamte *MSE* in drei Komponenten aufgespalten werden:

$$MSE = MSE_{disp} + MSE_{vol} + MSE_{pat}, \quad (1.10)$$

wobei der MSE_{tot} den Fehler der Originalvorhersage innerhalb der Verifikationsgebiets beschreibt. Die Differenz des *MSE* vor und nach der Verschiebung des Vorhersagefeldes gibt den Beitrag des Verschiebungsfehlers MSE_{disp} am gesamten Fehler an. Der Anteil MSE_{vol} berechnet sich aus der quadrierten Differenz der verschobenen Vorhersage und der Beobachtung innerhalb des CRA. Als Musterfehler wird der Anteil am gesamten Fehler bezeichnet, der nicht durch Verschiebung oder Menge beschrieben werden kann.

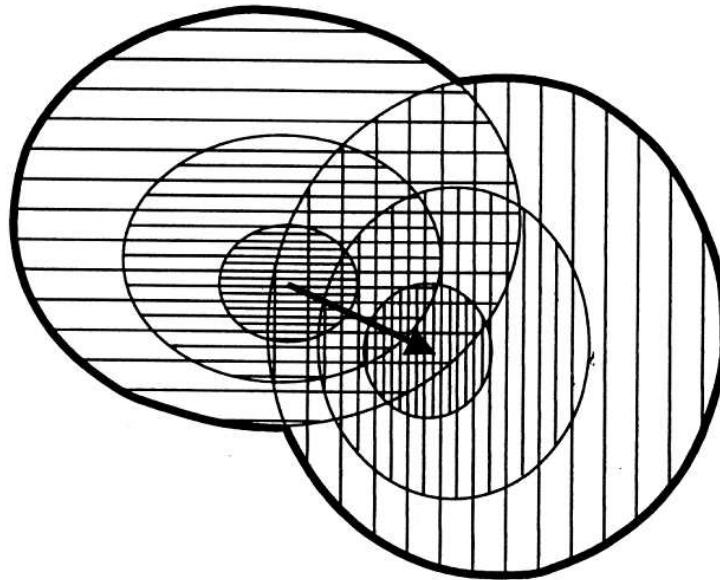


Abbildung 1.8. Prinzip der Methode der CRAs für ein idealisiertes Beispiel (Abb. aus Ebert and McBride, 2000). In der Beobachtung (quergestreift) und Vorhersage (längsgestreift) sind zusammenhängende kreisförmige Regenobjekte zu erkennen. Der Pfeil zeigt die Verschiebung der Vorhersage gegenüber der Beobachtung an und mit der dicken schwarzen Linie wird die Kontur bzw. die Grenzen der beiden CRAs dargestellt.

Die Aufspaltung des Fehlers in einzelne Komponenten ist eine nützliche Eigenschaft für Modellentwickler, um die numerischen Wettermodelle zu verbessern. Schwierigkeiten der Anwendbarkeit dieser Methode können auf Grund der direkten Zuordnung zwischen Objekten aus Beobachtungen und Vorhersagen auftreten, z.B. wenn das zu untersuchende Niederschlagsereignis sehr strukturiert ist (z.B. bei konvektiven Situationen).

Die MODE Technik

Als zweite objekt-orientierte Verifikationsmethode wird MODE (*engl.: Method for Object-based Diagnostic Evaluation*) an dieser Stelle ausführlicher beschrieben. Die Idee die dieser Technik zu Grunde liegt, ist die Bewertung die Ähnlichkeit von Beobachtung und Vorhersage unter Betrachtung verschiedener Aspekte. Die Anwendung dieser Methode enthält mehrere Schritte. Als erstes werden in den geglätteten Niederschlagsfeldern mit Hilfe eines Schwellenwertes Objekte bestimmt. Durch die Wahl der Glättung der Felder und des Schwellenwertes können verschiedene räumliche Skalen untersucht werden. Nach Bestimmung der Objekte werden diese für Beobachtung und Vorhersage mit Hilfe ei-

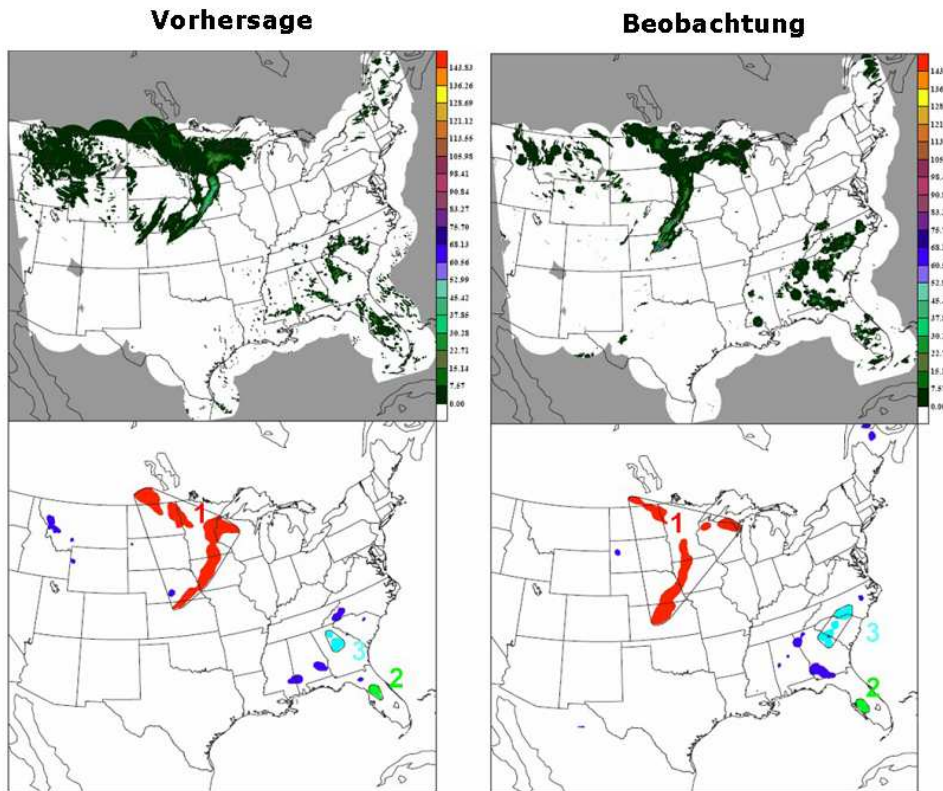


Abbildung 1.9. Reales Beispiel für die Identifizierung von Objekten mit Hilfe von MODE (Abb. aus Brown et al., 2009). Mit den selben Farben (rot, grün und cyan) und den dazugehörigen Zahlen werden die zugeordneten Objekte angegeben. Wohingegen die violetten Objekte nicht zugeordnet werden konnten.

nes 'Fuzzy'-Ansatzes vereinigt und anschließend erfolgt ein direktes Zuordnen zwischen beobachteten und vorhergesagten Objekten. Verschiedene Merkmale (z.B. Ort, Ausdehnung, Menge, Intensität, Form, etc.) der zugeordneten Objekte können ausgedrückt als kombiniertes Maß auf ihre Ähnlichkeit hin untersucht werden. Außerdem können die Merkmale von nicht zugeordneten oder zusammengefassten Objekten separat bestimmt werden.

In Abbildung 1.9 ist ein reales Beispiel der Anwendung von MODE für die Vereinigten Staaten von Amerika dargestellt. Im nördlichen mittleren Westen können in der Vorhersage wie auch der Beobachtung (obere Zeile in Abb. 1.9) größere zusammenhängende Niederschlagsereignisse ausgemacht werden. Außerdem sind noch im Südosten und Florida weitere Niederschlagsfelder zu erkennen. In der unteren Zeile sind dann die unterschiedlichen Objekte mit verschiedenen Farben gekennzeichnet. Diejenigen Objekte, die einander zugeordnet werden konnten besitzen die gleiche Farbe und wurden nummeriert. Außerdem gibt es Objekte (violett), die keinen entsprechenden Partner haben.

1.3 Ziele und Aufbau der Arbeit

Die Einführung der NWV Modelle mit Gitterweiten von $\Delta x = 2-4$ km sowie ohne Konvektionsparametrisierung in den operationellen Betrieb der Wetterdienste führt zu der Frage, ob diese bessere Vorhersagen liefern. Um diese Frage beantworten zu können, wird der hochaufgelöste Beobachtungsdatensatz (Kapitel 1.1.2) benötigt, welcher in Kapitel 2 analysiert und mit der Referenzperiode (1961-1990) sowie Modellvorhersagen verglichen wird. Als Werkzeug der Untersuchung der Modellvorhersagen wird die merkmalsbezogene Verifikationstechnik SAL verwendet. In Kapitel 3.1 werden die Definition sowie grundlegende Eigenschaften beschrieben und in Kapitel 3.2 werden verschiedene Eigenschaften von SAL, sowie die zusätzlichen Informationen die gegenüber den Ergebnissen traditioneller Maße gewonnen können, aufgezeigt. Die Erweiterung von SAL, um zeitliche Fehler von Niederschlagsvorhersagen mit kurzen Akkumulationszeiten (≤ 1 h) zu bewerten, wird in Abschnitt 3.3 vorgestellt und an einem Beispiel wird die grundsätzliche Anwendbarkeit dieser Erweiterung aufgezeigt.

In Kapitel 4 werden die QNVs von 19 Modellen im Rahmen der Demonstrationsperiode des MAP D-PHASE Projektes mit SAL analysiert und verglichen. Es werden 6- sowie 12-h Akkumulationen der 00 UTC Läufe genutzt. Zusätzlich erfolgt eine Bewertung der Güte für konvektive Niederschlagsereignisse in Hochdrucksituationen und Situationen mit synoptischen Antrieb. Das Kapitel 5 ist in zwei Abschnitte untergliedert. Im ersten Teil (Kapitel 5.1) werden die Zeitreihen der Qualität täglicher QNVs für vier Modelle über einen Zeitraum von acht Jahren untersucht. Dafür kommen neben SAL auch traditionelle Maße zum Einsatz um aufzuzeigen welche zusätzliche Information durch SAL zur Verfügung steht. Im zweiten Teil (Kapitel 5.2) wird der meteorologische Parameter der konvektiven Zeitskala definiert und beschrieben. Es wird die erste aus Beobachtungen abgeleitete Klimatologie dieser Größe für Deutschland vorgelegt. Zusätzlich dient die konvektive Zeitskala als Kriterium um verschiedene sommerliche Wettersituationen getrennt verifizieren zu können, um deren charakteristischen Fehler aufzuzeigen. Zusammenfassende Bemerkungen und zukünftige Anwendungsmöglichkeiten von SAL werden im abschließenden Kapitel 6 beschrieben.

Kapitel 2

Niederschlagsklimatologie für Deutschland

Nachdem in Kapitel 1.1.2 die Konstruktion des disaggregierten Niederschlagsdatensatzes vorgestellt wurde, soll im Folgenden die achtjährige Klimatologie für Deutschland betrachtet werden. Neben der Analyse erfolgt ein Vergleich mit Modellsimulationen. Doch zunächst wird die geografische Verteilung (Abschnitt 2.1) der mittleren, jährlichen und sowie saisonalen, Niederschlagssummen beschrieben. Zusätzlich werden Karten der Niederschlagshäufigkeit vorgestellt und es wird der Frage nachgegangen, ob ein wöchentlicher Zyklus in der Niederschlagsverteilung für Deutschland existiert. Anschließend erfolgt ein Vergleich der Eigenschaften stündlicher und täglicher Niederschlagssummen (Abschnitt 2.2) aus Beobachtungen und numerischen Modellsimulationen. Zum Abschluss wird in Abschnitt 2.3 der Tagesgang des Niederschlags behandelt.

2.1 Beobachtete Niederschlagsverteilung

Im Mittelpunkt dieses Abschnitts steht die grafische Darstellung der beobachteten mittleren Niederschlagsverteilung für Deutschland. Basis dafür bildet der Zeitraum von 2001 bis 2008 auf Grundlage von gegitterten täglichen Stationsdaten (Kapitel 1.1.2).

2.1.1 Jährliche Niederschlagsmengen

Zur Einordnung des Zeitraums 2001-2008, der in dieser Arbeit als Untersuchungsperiode zugrunde liegt, soll die mittlere jährliche Niederschlagssumme des in Kapitel 1.1.2 be-

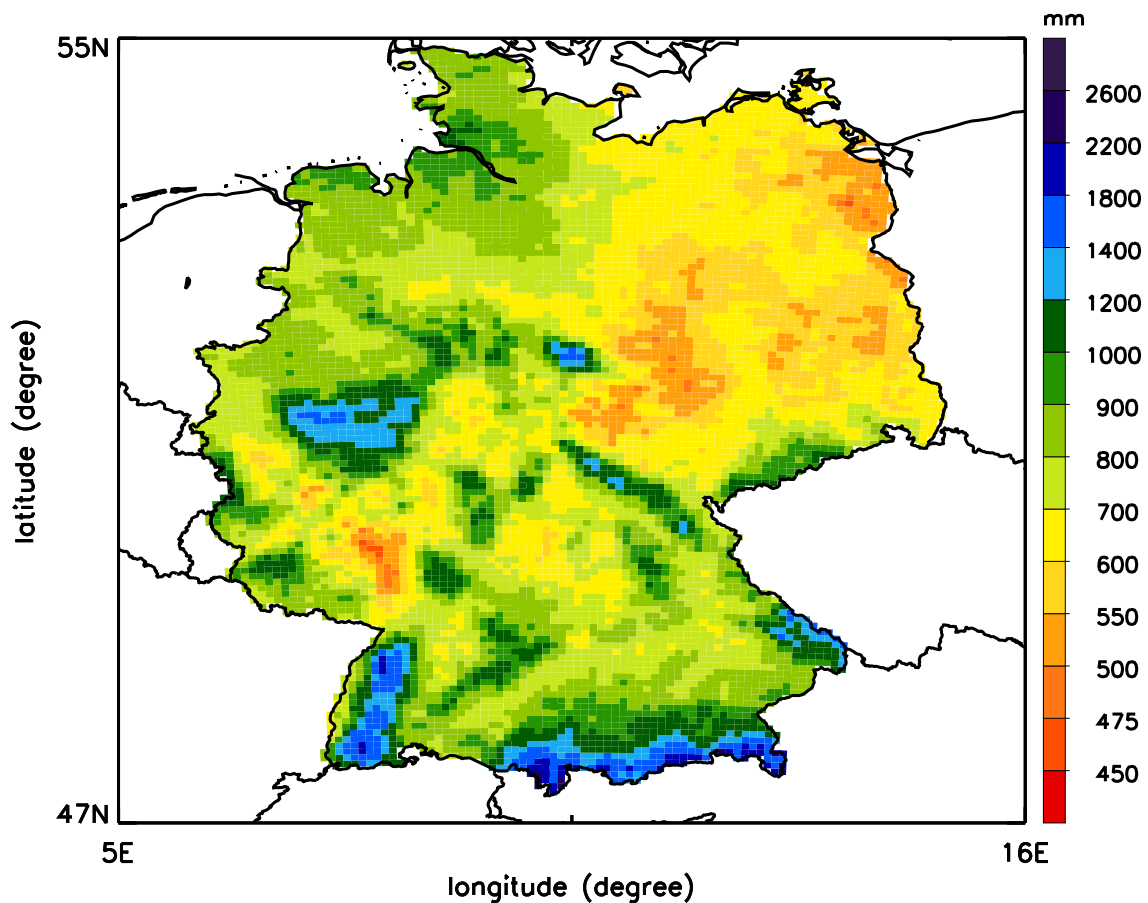


Abbildung 2.1. Karte des mittleren jährlichen Niederschlags [mm] aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008.

schriebenen Datensatzes mit den Beobachtungen der dreißigjährigen Klimatologie (1961-1990) des Deutschen Wetterdiensts (DWD) aus Müller-Westermeier *et al.* (1999) verglichen werden.

In Abbildung 2.1 ist die geografische Verteilung des mittleren jährlichen Niederschlags dargestellt. Diese wird aus gegitterten Stationsbeobachtungen berechnet. Die Konturenintervalle sind identisch mit denen die für den deutschen Klimaatlas gewählt wurden. Ein Vergleich dieses Achtjahreszeitraums mit der dreißigjährigen Klimatologie des deutschen Klimaatlas zeigt, dass dieser - klimatologisch betrachtet - relativ kurze Zeitraum, die charakteristischen Strukturen der geografischen Verteilung der jährlichen Niederschlagssumme für Deutschland beinhaltet. Die trockensten Regionen sind in beiden Zeiträumen die selben und umfassen weite Teile des Nordostens, sowie Rheinhessen im Südwesten Deutschlands mit weniger als 500 mm Jahresniederschlag. Die maximalen Werte werden

am Nordrand der Alpen, im Schwarzwald sowie in den deutschen Mittelgebirgen erreicht und übersteigen dort 1400 mm.

2.1.2 Saisonale Niederschlagssummen

Um einen genaueren Blick auf die geografische Verteilung des Niederschlags zu erhalten, soll die saisonale Verteilung betrachtet werden. Dies geschieht durch die Unterteilung der mittleren Jahressumme auf die vier meteorologischen Jahreszeiten. In Abbildung 2.2 sind die mittleren Niederschlagssummen für die einzelnen Jahreszeiten zusammengefasst.

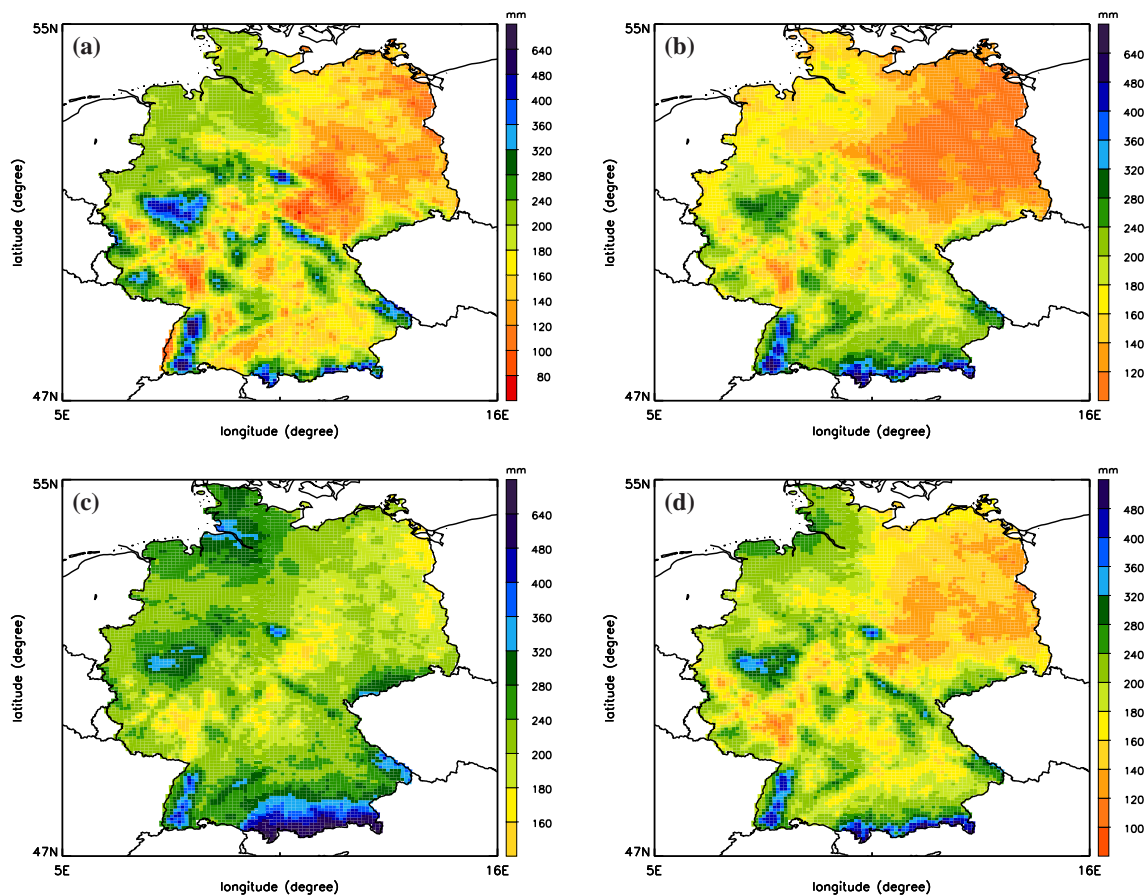


Abbildung 2.2. Karte der mittleren, saisonalen Niederschläge [mm] aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008: Winter (a), Frühling (b), Sommer (c) und Herbst (d).

Wie im Fall der mittleren Jahressumme erfolgte die Wahl der Konturenintervalle entsprechend der im deutschen Klimaatlas verwendeten Intervalle. Deutlich zu erkennen ist, dass im Sommer (Abb. 2.2c) für weite Gebiete Deutschlands die größte Niederschlagsmenge zu beobachten ist. Besonders auffällig ist dies für den Norden Deutschlands, aber auch am Nordrand der Alpen sind hohe Werte sommerlicher Niederschlagsmengen zu erkennen. Das sommerliche Niederschlagsmaximum an der Nordseeküste, das für den Achtjahreszeitraum konstatiert werden kann, tritt in der dreißigjährigen Periode des Klimaatlas erst im Herbst auf. Dies kann als Hinweis gewertet werden, dass die achtjährige Klimatologie für diese Region nicht ausreichend lang ist, um die jährliche Variation des Niederschlags in dieser Region zu beschreiben. Dies als eine Änderung des Klimas zu werten ist somit nicht möglich. Die Jahreszeit mit den geringsten Werten der Niederschlagssumme ist für einen Großteil Deutschlands der Frühling (Abb. 2.2b). Entlang des Rheins und im südöstlichen Harzvorland sowie im Lee des Thüringer Waldes werden im Winter (Abb. 2.2a) die geringsten Werte beobachtet.

Um einen genaueren Blick zu erhalten ob im Winter oder im Sommer in einer Region Deutschlands der meiste Niederschlag fällt, ist die Differenz der Menge zu betrachten. Regionen mit blauer Färbung zeigen in Abbildung 2.3 eine größere Niederschlagsmenge für den Sommer, währenddessen rötliche Farben eine größere winterliche Niederschlagsmenge darstellen. In weiten Teilen Deutschlands wird im Sommer die größte Niederschlagsmenge beobachtet. Nur die durch Staueffekte charakterisierten Mittelgebirge, z.B. Schwarzwald, Thüringer Wald oder der Harz, weisen ein Maximum im Winter auf. Neben den grafischen Vergleichen der beiden Datenreihen soll im Folgenden die Gegenüberstellung einiger charakteristischer Zahlen der Niederschlagsverteilung diskutiert werden. Tabelle 2.1.2 listet die Werte für die mittlere Jahressumme, Minimum und Maximum des Jahresniederschlags sowie die relative Verteilung für die einzelnen Jahreszeiten auf. Die mittlere Jahressumme des Achtjahreszeitraums ist mit 804 mm um 101 mm höher als die 30-jährige Klimatologie des DWD (703 mm). Gegenüber dem Wert der ersten vier Jahre (828 mm; *Paulat*, 2007) ist die Differenz zur DWD-Klimatologie geringer geworden. Generell bleibt festzuhalten, dass die mittlere Jahressumme im dreißigjährigen Datensatz beim achtjährigen Datensatz erheblich größer ist. Die Unterschiede treten bei der Betrachtung der Minimum- und Maximumwerte¹ noch deutlicher zu Tage. Der Minimumwert von 480 mm liegt um fast 100 mm über dem der DWD-Klimatologie, allerdings ist der Maximumwert annähernd 800 mm geringer, dies bedeutet eine geringere Variabilität im Datensatz. Möglicherweise kann diese Diskrepanz zwischen beiden Beobachtungsdatensätzen damit begründet werden, dass die DWD-Klimatologie auf einem 1x1 km Git-

¹Der Minimum- bzw. der Maximumwert ist derjenige Gitterpunkt, der im Jahresmittel der Zeitreihe den geringsten bzw. der höchsten Wert aufweist.

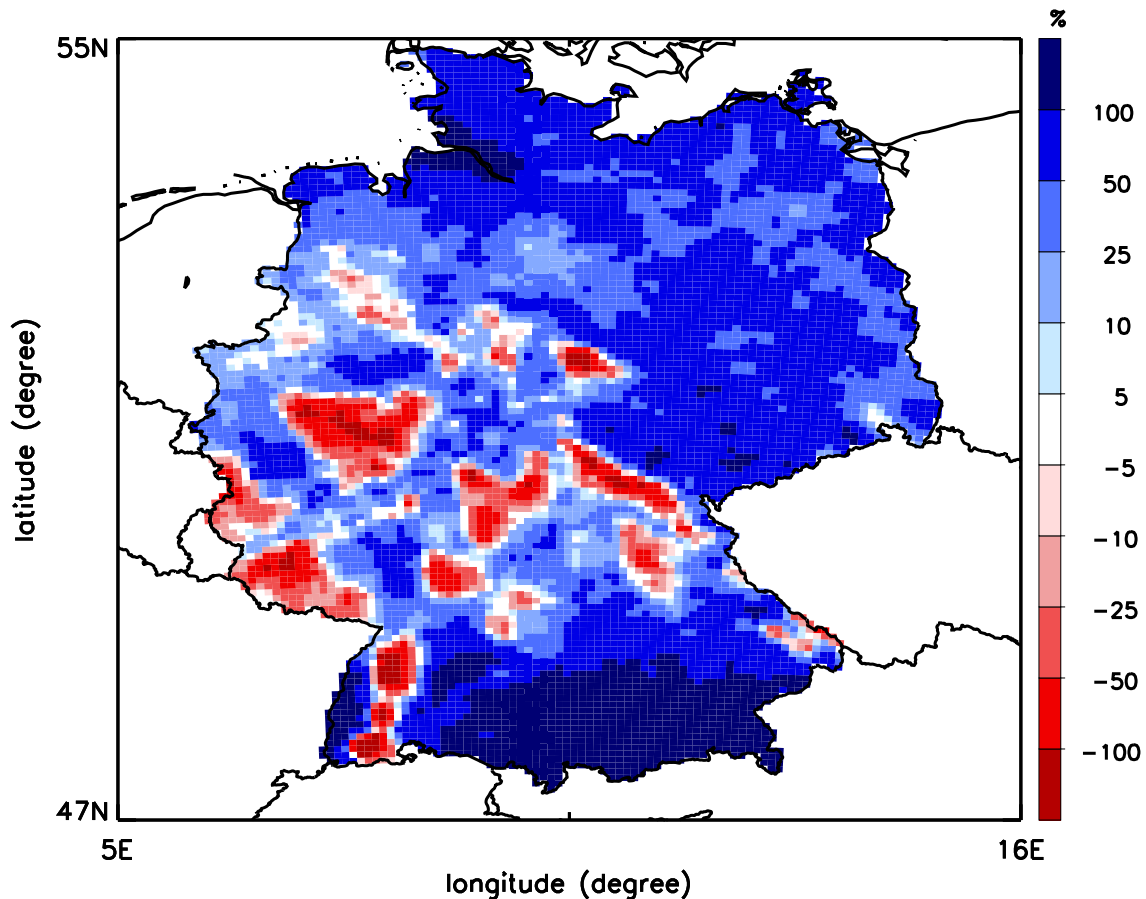


Abbildung 2.3. Karte der relativen Differenz des mittleren Sommer- und Winterniederschlags [%] aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008.

ter (Müller-Westermeier, 1995) vorliegt und der achtjährige Datensatz auf einem 7x7 km Gitter (Paulat et al., 2008). Die gröbere Auflösung sorgt dafür, dass besonders Intensitätsmaxima geglättet werden.

Beim Vergleich der relativen Anteile für die einzelnen Jahreszeiten zwischen beiden Datensätzen zeigt sich kaum ein Unterschied. Im Mittel gilt für Deutschland, dass im Sommer die größten Niederschlagssummen beobachtet werden und die restlichen Jahreszeiten jeweils gleich viel Niederschlag aufweisen.

Die auch von anderen Publikationen beschriebene Zunahme der Winter- und zum Teil der Herbstniederschläge für Deutschland (Schönwiese and Janoschitz, 2008) bzw. Nord-europa (IPCC, 2007) können mit Hilfe der Tabelle 2.1.2 grundsätzlich gezeigt werden,

	Klimatologie (DWD) (1961-1990)	Beobachtungen (2001-2008)
gemittelte Jahressumme [mm/Jahr]	703	804
Minimaler Gitterpunkt- wert des Jahresmittels [mm/Jahr]	383	480
Maximaler Gitterpunkt- wert des Jahresmittels [mm/Jahr]	3270	2428
Winter [%]	22,8	23,5
Frühling [%]	23,6	22,8
Sommer [%]	30,4	30
Herbst [%]	23,2	23,8

Table 2.1. Charakteristische Werte der Niederschlagsverteilung in Deutschland für die 30-jährige Klimatologie des DWD und des achtjährigen Datensatzes der gegitterten Stationsdaten.

da diese Jahreszeiten einen geringfügig höheren relativen Anteil am Jahresniederschlag aufweisen als in der 30-jährigen DWD Klimatologie.

2.1.3 Niederschlagshäufigkeiten

Um zusätzliche Informationen über die klimatologische Niederschlagsverteilung zu erhalten, soll im Folgenden die relative Häufigkeit von Niederschlagsereignissen untersucht werden. Außerdem wird die geografische Verteilung der Niederschlagsintensität anhand des 80 %- sowie des 95 %-Perzentils betrachtet.

Niederschlagsereignisse

Zunächst wird die relative Häufigkeit von täglichen Niederschlagssummen untersucht. Dazu werden die gegitterten Stationsmessungen benutzt. Als Niederschlagsereignis werden tägliche Niederschlagssummen gezählt, die einen Schwellenwert von 0,1 mm pro Tag am betrachteten Gitterpunkt überschreiten. In Abbildung 2.4 ist die Karte der jährlichen, relativen Häufigkeit von Niederschlagsereignissen dargestellt. Die Regionen Mittelgebirge, Deutsche Bucht und die Alpen (siehe Abschnitt 2.1) mit den größten Niederschlagshäufigkeiten (Häufigkeit größer als 55%) sind diejenigen mit den größten jährlichen Niederschlagssummen. Die Gebiete mit den geringsten Niederschlagssummen in

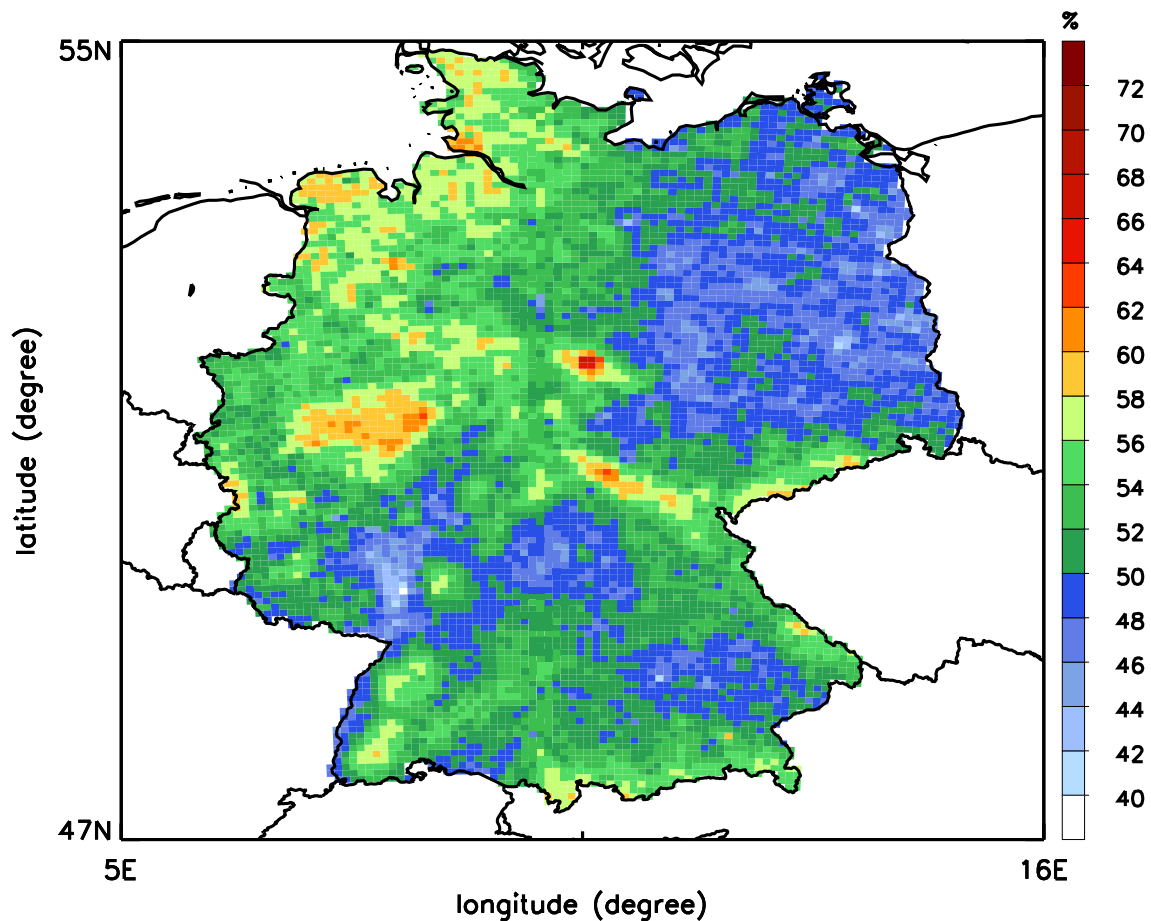


Abbildung 2.4. Geografische Verteilung der relativen Häufigkeit [%] von Tagen mit Niederschlagssummen $\geq 0,1$ mm pro Tag aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008.

Deutschland, der Nordosten und der nördliche Oberrhein, weisen eine relative Niederschlagshäufigkeit von weniger als 45% auf. Dies bedeutet, dass an weniger als jedem zweiten Tag Niederschlag in diesen Regionen beobachtet wird, wohingegen z.B. der Harz im Mittel an mindestens zwei von drei Tagen Niederschlag aufweist. Bei der Betrachtung der geografischen Verteilung von Ereignissen die mindestens 1 mm pro Tag (siehe Anhang Abbildung A.1) aufweisen, werden keine Unterschiede in der Struktur der regionalen Verteilung zu den Resultaten der Ereignisse von 0,1 mm pro Tag identifiziert. Geringe beobachtete Niederschlagsmengen (< 1 mm/d) können sehr stark vom Beobachter beeinflusst sein (z.B. *Frei et al.*, 2003). Allerdings scheint dies kaum bzw. gar keinen Einfluss auf die Verteilung der Niederschlagsereignisse zu haben. Die Regionen mit den häufigsten Ereignissen sind dieselben wie beim niedrigen Schwellenwert. Im Fall der Mit-

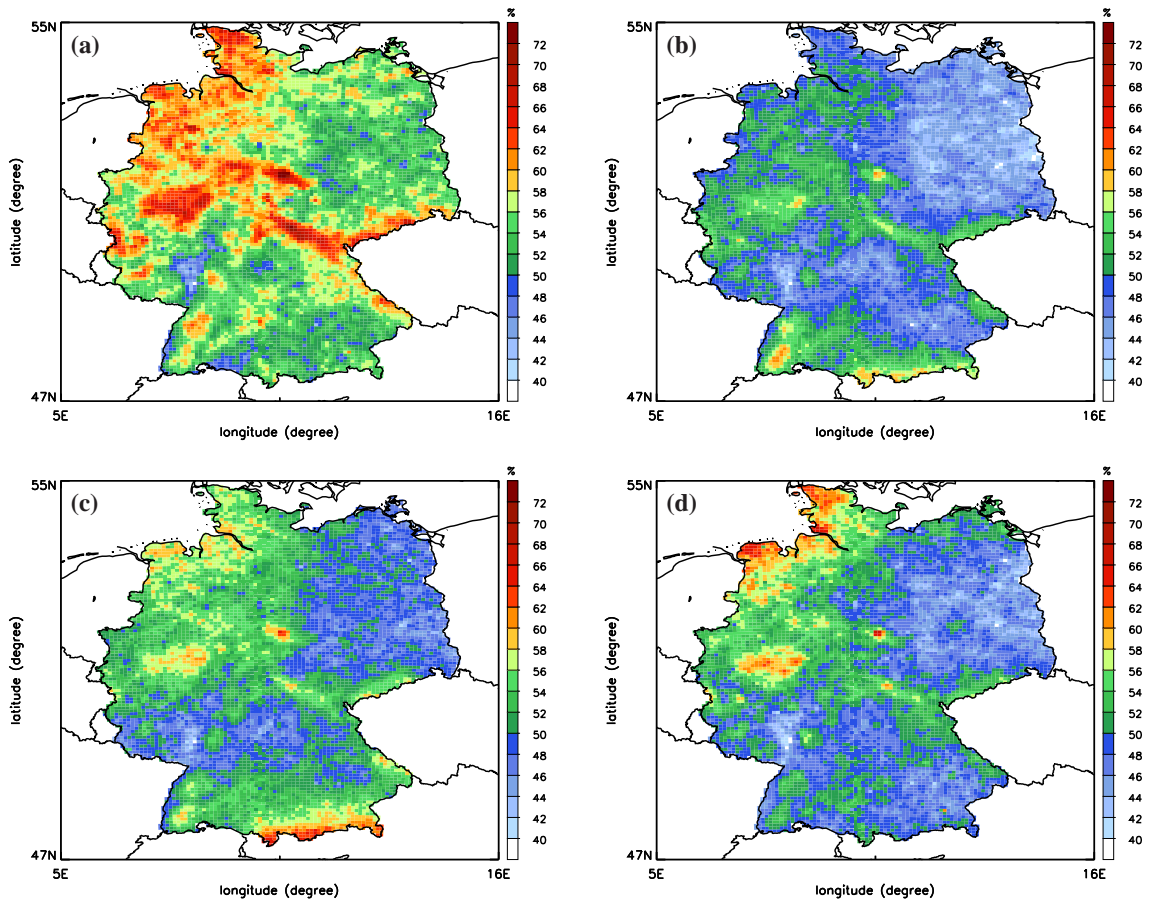


Abbildung 2.5. Wie Abbildung 2.4 nach Jahreszeiten unterteilt: Winter (a), Frühling (b), Sommer (c) und Herbst (d).

telgebirgsregionen liegen die Werte bei knapp über 40% und beispielsweise in der Region Rheinhessen bei unter 30%.

Da neben regionalen Unterschieden innerhalb der jährlichen Häufigkeitsverteilung auch Differenzen bezüglich der einzelnen Jahreszeiten zu erwarten sind, wird nachfolgend die Analyse mit saisonaler Unterteilung betrachtet. Hierfür wird in Abbildung 2.5 für jede Jahreszeit die geografische Verteilung der Häufigkeiten von täglichen Niederschlagssummen $\geq 0,1$ mm pro Tag dargestellt. Im Gegensatz zu den Niederschlagssummen ist der Winter (Abb. 2.5a) für viele Regionen in Deutschland die Jahreszeit mit der größten Häufigkeit von Niederschlagsereignissen. Eine Ausnahme bildet hier der Alpenraum mit einem Häufigkeitsmaximum im Sommer (Abb. 2.5c). Dieses Maximum wird durch sommerliche Gewitter aufgrund des orographischen Einflusses in der Region hervorgerufen (Finke and Hauff, 1996, Lugauer and Winkler, 2005). Ein weiterer interessanter

Aspekt kann für die Deutsche Bucht im Herbst (Abb. 2.5d) festgestellt werden. Die Erklärung dafür dürfte mit der im Herbst im Vergleich zum Festland noch relativ warmen Nordsee (Müller-Westermeier et al., 1999) zu suchen sein, so dass in dieser Region eine erhöhte Anzahl an Niederschlagsereignissen entstehen kann. Die Niederschlagshäufigkeit im Frühling (Abb. 2.5b) ist in Deutschland am geringsten, was einhergeht mit der Niederschlagsmenge.

Niederschlagsintensität

Entsprechend dem vorherigen Abschnitt wird an dieser Stelle die Intensität von Niederschlagsereignissen täglicher Summen aus gegitterten Stationsdaten untersucht. Dazu werden Deutschlandkarten des 80 %- sowie des 95 %-Perzentils berücksichtigt. Bei dieser Betrachtung werden ausschließlich die Sommer- und Winterwerte vorgestellt und diskutiert.

Die Abbildung 2.6 stellt die Karte der geografischen Verteilung des 80 %-Perzentil für Deutschland dar. Auf der linken Seite ist das Ergebnis für Winter und auf der rechten Seite für Sommer zu betrachten. Die Farbskala für beide Jahreszeiten ist identisch, so dass jahreszeitliche Unterschiede schnell und einfach erkannt werden können. Deutlich sichtbar sind die unterschiedlichen Regionen mit hohen Niederschlagswerten des 80 %-Perzentils. Im Winter (Abb. 2.6a) sind dies insbesondere die Mittelgebirge wie z.B. der Schwarzwald oder der Harz. Ebenfalls weisen diese Regionen auch die größten Niederschlagsmengen im Winter auf. Die Regionen (nördlicher Oberrhein sowie östlich und südöstlich des Harz-

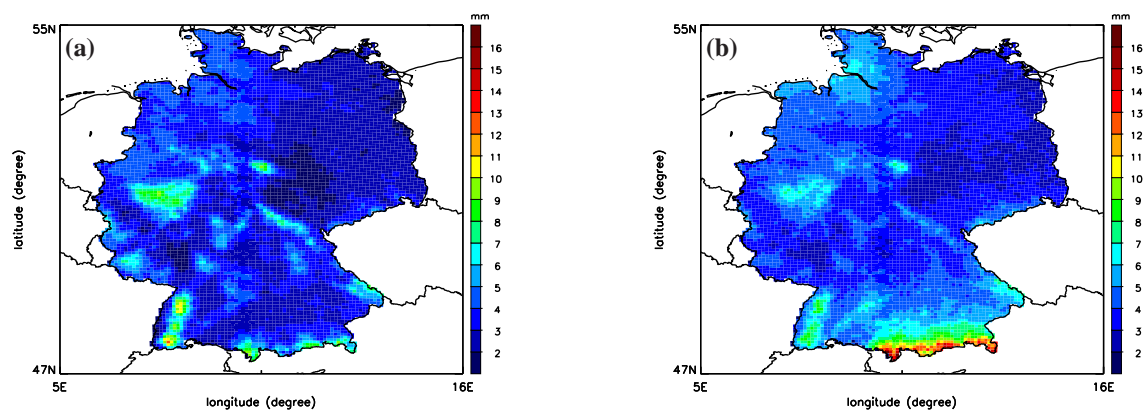


Abbildung 2.6. Geografische Verteilung des 80 %-Perzentils [mm] des Niederschlags für jeden Gitterpunkt der gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008 für Winter (a) und Sommer (b).

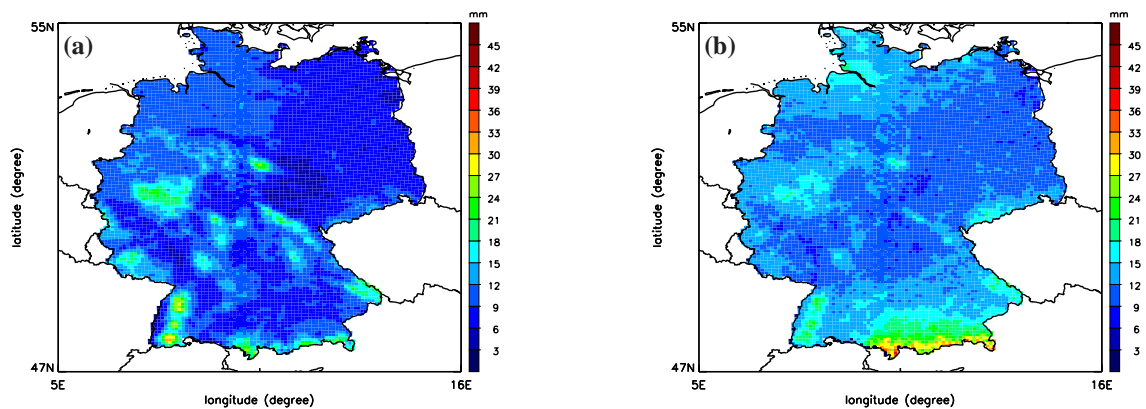


Abbildung 2.7. Wie Abbildung 2.6, jedoch für das 95 %-Perzentil.

zes) mit den geringsten Werten für das 80 %-Perzentil sind darüber hinaus die Regionen mit der geringsten Niederschlagssumme. Dagegen treten im Sommer (Abb. 2.6b) die höchsten Werte des 80 %-Perzentils an den Alpen auf. Die hier beobachteten Werte sind mit mehr als 12 mm pro Tag deutlich höher als die Maximalwerte im Winter, wo nur auf den Höhen des Schwarzwaldes ähnlich hohe Werte dokumentiert sind.

Nach der Darstellung und Diskussion der geografischen Verteilung der Werte des 80 %-Perzentil (Abb. 2.6) soll die Analyse für die Werte des 95 %-Perzentil (Abb. 2.7) angesprochen werden. Für den Winter (Abb. 2.7a) sind kaum Unterschiede in der geografischen Verteilung der Perzentilwerte zwischen beiden Karten zu erkennen, d.h. die Regionen mit den höchsten bzw. niedrigsten Werten sind identisch. Im Sommer (Abb. 2.7b) sieht das Bild etwas anders aus. Zwar bleibt der Alpenraum die Region mit den höchsten Werten, allerdings treten wie im Fall der Verteilung des 80 %-Perzentil jetzt kaum bzw. keine sekundären Maxima auf. Diese Regionen mit erhöhten Werten sind mit der Topografie verknüpft. Im Fall der Verteilung des 95 %-Perzentil sind diese Strukturen nicht mehr zu erkennen, sondern es existiert eine scheinbar zufällige Verteilung der Werte. Diese schwächer ausgeprägte räumliche Struktur konnte schon in der mittleren Niederschlagssumme des Sommers (Abb. 2.2c) festgestellt werden, was einen Hinweis dafür liefert, dass die stärksten Ereignisse (5 Prozent aller Ereignisse) den größten Anteil am Gesamtniederschlag leisten.

2.1.4 Wöchentlicher Zyklus

In diesem Abschnitt wird der Frage nachgegangen, ob es Wochentage gibt an denen besonders viel/wenig bzw. häufig/selten Niederschlag beobachtet wird. Der einzige bekann-

te Zyklus der eine siebentägige Periode besitzt ist die menschliche Aktivität in den westlichen Industrienationen (*Quaas et al.*, 2009). Anthropogene Effekte auf das Wetter und das Erdklima umfassen z.B. Hitzeinseln, Luftverschmutzung und der globale Klimawandel (*Albrecht*, 1989, *Rosenfeld*, 2000, *Rosenfeld et al.*, 2007). Eine mögliche Auswirkung ist der Einfluss von zusätzlichen Kondensations- und Eiskeimen in der Nähe von Ballungsräumen, die die Niederschlagsmenge und -häufigkeit verändern könnten. Anthropogene Aerosole können zur Zunahme der Wolkentropfenkonzentration führen, da mehr Partikel um die selbe Menge des Wolkenwassers konkurrieren. Dadurch kommt es zu einer Modifizierung der Wolkenalbedo (erster indirekter Aerosoleffekt; *Twomey*, 1974). Außerdem führt dies zu einer geringeren Regenbildungsrate sowie längeren Wolkenlebenszeiten mit einhergehendem Bedeckungsgrad (zweiter indirekter Aerosoleffekt; *Albrecht*, 1989). Modellstudien von *Lohmann and Feichter* (2005) zeigen, dass beide Effekte die gleiche Größenordnung haben, wobei beachtliche Unsicherheiten vorhanden sind. In wolkenmikrophysikalischen Modellstudien konnte *Grützun et al.* (2008) für anthropogen verschmutzte Situationen bestätigen, dass die Niederschlagsbildung verringert ist. Dagegen zeigen die Resultate von *Reutter* (2009), für eine durch Waldbrände stark verschmutzte Atmosphäre keine Verringerung der Niederschlagsmenge. Die Bildung von Niederschlag ist verlangsamt, da aber die Lebenszeit der Wolke länger wird, ist die Summe am Ende gleich.

Für verschiedene meteorologische Größen kann ein Wochengang beobachtet werden. Die Temperatur zeigt zum Beispiel einen solchen Zyklus für Deutschland (*Bäumer and Vogel*, 2007), Europa (*Laux and Kunstmann*, 2008) oder die Nordhemisphäre (*Gordon*, 1994). Darüber hinaus zeigen Luftschadstoffe, wie O₃, CO₂ und NO₂ sowie feste Aerosolpartikel einen Wochengang (z.B. *Beany and Gough*, 2002, *Jin et al.*, 2005, *Shutters and Balling*, 2006, *Bäumer et al.*, 2008).

Im Fall des Niederschlags wird in einigen Publikationen eine wöchentliche Periode festgestellt, z.B. in *Bäumer and Vogel* (2007) für 12 Wetterstationen (1991-2005) in Deutschland mit dem Dienstag als trockensten und Samstag als niederschlagsreichsten Tag. Die Untersuchung von *Cerveny and Balling* (1998) identifiziert einen schwachen wöchentlichen Zyklus für die nordöstliche Atlantikregion der USA (1979-1995 aus Satellitendaten abgeleitet) und in der Studie von *Bell et al.* (2008) für den Sommerniederschlag in den USA (1998-2005, aus Satellitendaten abgeleitet), wo allerdings der Tag des maximalen Niederschlags regional unterschiedlich ist. Dagegen gibt es auch eine Anzahl an Veröffentlichungen die einen Wochenzyklus nicht feststellen konnten. Dazu gehört z.B. der Kommentar von *Franssen Hendricks* (2008) auf *Bäumer and Vogel* (2007), in dem für die schweizer Städte Zürich und Lugano kein wöchentlicher Zyklus zu finden war, bezogen auf den Untersuchungszeitraum von 1864-2005 oder Teilzeiträumen. In der Arbeit

von *Laux and Kunstmann* (2008) konnte für 33 Wetterstationen verteilt über Deutschland kein statistisch signifikanter Wochengang für die Untersuchungsperiode 1991-2005 erkannt werden. In der Studie von *Schultz et al.* (2007) in der 219 Stationen in den USA im Zeitraum von 1951-1992 untersucht wurden, konnte ebenfalls kein wöchentlicher Rhythmus beobachtet werden.

Die Frage, ob der Niederschlag einem möglichen wöchentlichen Rhythmus unterliegt, kann mithilfe der gegitterten Stationsdaten für Deutschland, regional differenziert, untersucht werden. Hier sollten dann Regionen zu erkennen sein, die einen Zyklus haben oder keinen aufweisen. Dies könnte die unterschiedlichen Ergebnisse von *Bäumer and Vogel* (2007) und *Laux and Kunstmann* (2008) erklären, obwohl sie den selben Untersuchungszeitraum verwendeten, allerdings eine unterschiedliche Anzahl an Stationen (12 zu 33 Wetterstationen) mit einer etwas anderen Aufteilung über Deutschland.

Zur Beantwortung der Frage nach einem möglichen Wochengang und einer regionalen Unterscheidung werden die Niederschlagssumme und die -häufigkeit analysiert. Zur Untersuchung wird der komplette achtjährige Datensatz (2001-2008) auf die sieben Wochentage aufgeteilt. Anschließend werden die Niederschlagssumme und die Häufigkeit von Ereignissen mit mehr als 1 mm pro Tag für jeden Gitterpunkt im Datensatz bestimmt.

Niederschlagssumme

Die Abbildung 2.8 enthält Karten der geografischen Verteilung der mittleren, täglichen Niederschlagsmenge. Oben links ist der mittlere Tageswert dargestellt und die restlichen Karten zeigen die relativen Abweichungen für jeden einzelnen Wochentag an. Der Dienstag und Mittwoch zeichnen sich durch die größte negative Abweichung in der Nordhälfte Deutschlands aus. Für Süddeutschland ist kaum eine negative Abweichung zu erkennen, am ehesten noch am Sonntag. Die größte Niederschlagsmenge wird am Donnerstag beobachtet. Eine Besonderheit ist am Montag für Ostdeutschland zu vermerken. Hier tritt eine positive Abweichung auf, die mit der sogenannten 'Elbeflut' (*Rudolf and Rapp*, 2002) am 12. August 2002, ein Montag, in Verbindung steht. Es sind regionale Unterschiede auszumachen und ob es sich dabei um wirkliche oder rein zufällige Muster handelt, soll nachfolgend anhand einer statistischen Analyse untersucht werden.

Als erster Test wird untersucht, ob die Niederschlagsmenge eine Abhängigkeit vom Wochentag aufweist. Als Nullhypothese wird angenommen, dass kein signifikanter Unterschied der Niederschlagsmenge als Funktion des Wochentags existiert. Die beobachteten Niederschlagsmengen werden an jedem Gitterpunkt für jeden Tag der Woche aufsteigend sortiert (Tage ohne Niederschlag werden nicht betrachtet), um anschließend Mittelwert, Standardabweichung, Median und Interquartilsabstand zu berechnen. Die statis-

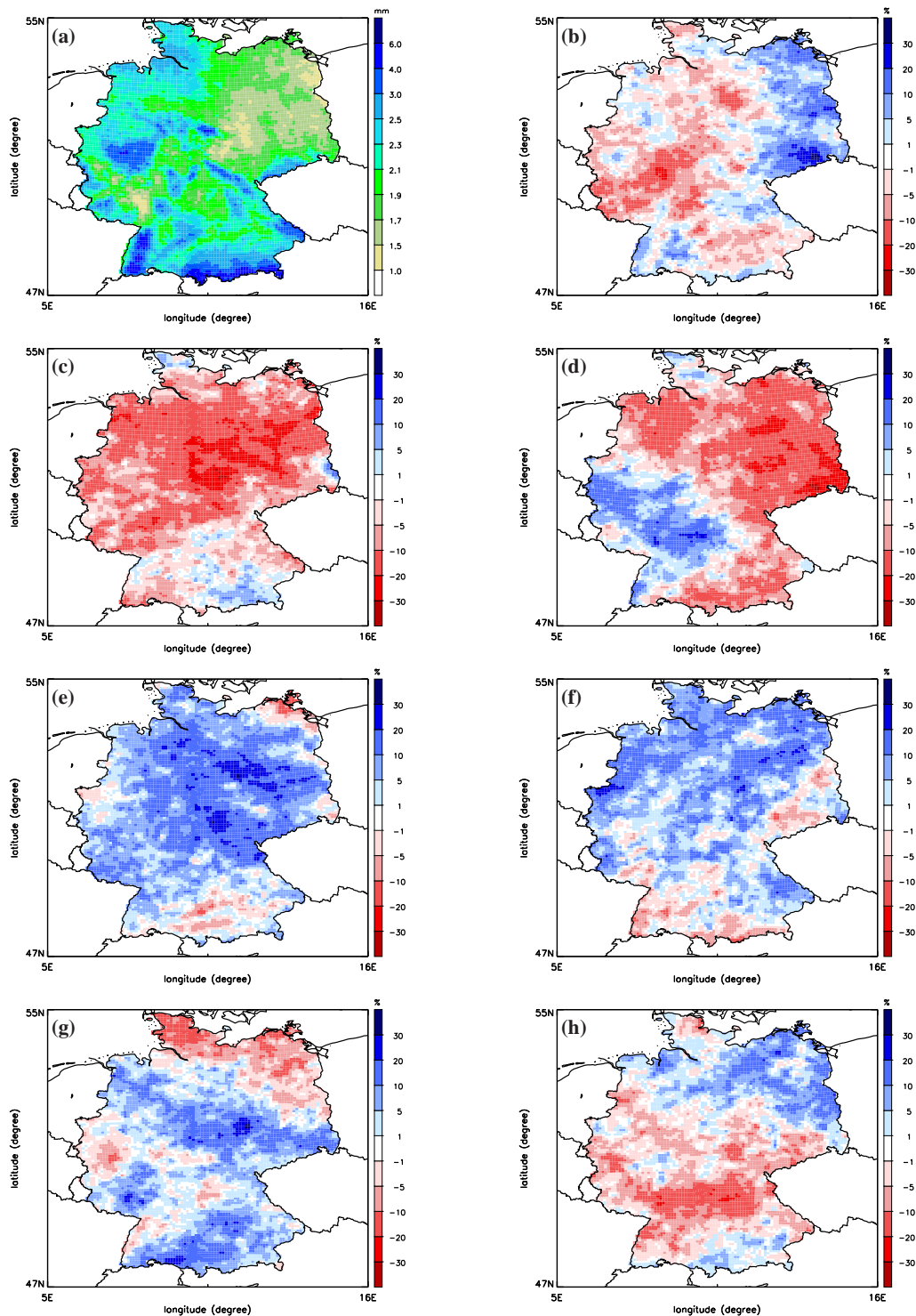


Abbildung 2.8. Geografische Verteilung der mittleren täglichen Niederschlagssumme aufgeteilt nach Wochentagen aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008. Die mittlere tägliche Niederschlagssumme in mm ist in (a) dargestellt und die relativen Abweichungen an den einzelnen Wochentagen, (b) Montag, (c) Dienstag, (d) Mittwoch, (e) Donnerstag, (f) Freitag, (g) Samstag und (h) Sonntag.

tische Untersuchung der Niederschlagsmenge hinsichtlich eines wöchentlichen Zyklus wird folgendermaßen durchgeführt, so dass an jedem Gitterpunkt des Beobachtungsdatensatzes die Differenz zwischen der mittleren Tagesmenge und der mittleren Summe des betreffenden Wochentags berechnet wird. Diese sieben Tageswerte werden mittels χ^2 -Test analysiert, um zu überprüfen ob eine signifikante Abweichung von der Annahme einer Gleichverteilung (Wilks, 2006) besteht. Die Analyse führt zu dem Ergebnis, dass an keinem Gitterpunkt ein signifikanter Unterschied für die Niederschlagsmenge auf dem 90 % Signifikanzniveau ermittelt wird. Damit kann die Nullhypothese nicht abgelehnt werden und geschlussfolgert werden, dass für die Niederschlagsmenge keine Präferenz für einen bestimmten Wochentag vorliegt.

Niederschlagsereignisse

Als zweite Eigenschaft wird für die Häufigkeit von Niederschlagsereignissen (≥ 1 mm/d) getestet, ob eine Abhängigkeit bzgl. der Wochentage vorliegt. Die Nullhypothese lautet in diesem Fall, dass keine signifikante Abweichung in der Häufigkeit existiert, d.h. es wird erwartet das an jedem Wochentag $\frac{1}{7}$ ($\approx 14,3\%$) aller Ereignisse auftreten sollten. Für jeden Gitterpunkt und jeden Wochentag wird die Differenz zwischen beobachteter und einem siebtel aller Ereignisse bestimmt. In Abbildung 2.9 ist das Ergebnis davon grafisch dargestellt. Die linke obere Karte zeigt die geografische Verteilung der mittleren Häufigkeit. Die anderen sieben Karten geben die relative Abweichung für die einzelnen Wochentage wieder. Grundsätzlich kann festgehalten werden, dass der Dienstag der Tag mit der geringsten Häufigkeit ist. Für Ostdeutschland ist auch der Mittwoch durch eine geringe Anzahl an Ereignissen gekennzeichnet. Der Sonntag ist für Süddeutschland ein Tag mit geringerer Häufigkeit. Dieses Muster ist vergleichbar mit dem der Niederschlagssumme. Mit mehr Ereignissen als den zu erwarteten Ereignissen sind der Freitag und der Samstag charakterisiert, wobei Freitag eher im Norden und Nordosten und am Samstag der Süden heraus stechen. Die Häufigkeit steht damit im Gegensatz zur Menge, bei der das Maximum am Donnerstag beobachtet wurde. Für jeden Gitterpunkt wurden die sieben Wochentagswerte unter der Verwendung des χ^2 -Tests getestet, um herauszufinden, ob sie signifikant von einer uniformen Verteilung abweichen (Wilks, 2006, Schultz et al., 2007). Bei dieser statistischen Analyse wurde an 499 Gitterpunkten eine Abweichung von der uniformen Wochentagsverteilung unter Verwendung des 90 % Signifikanzniveaus gefunden. Zu erwarten ist bei Verwendung des 90 % Signifikanzniveaus, dass die Nullhypothese an 741 Gitterpunkten ($\approx 10\%$ aller Gitterpunkte) per Zufall abgelehnt werden müsste. Die Daten sind also entsprechend der Nullhypothese konsistent, so dass damit belegt werden kann das die Niederschlagshäufigkeit keinem wöchentlichen Zyklus folgt.

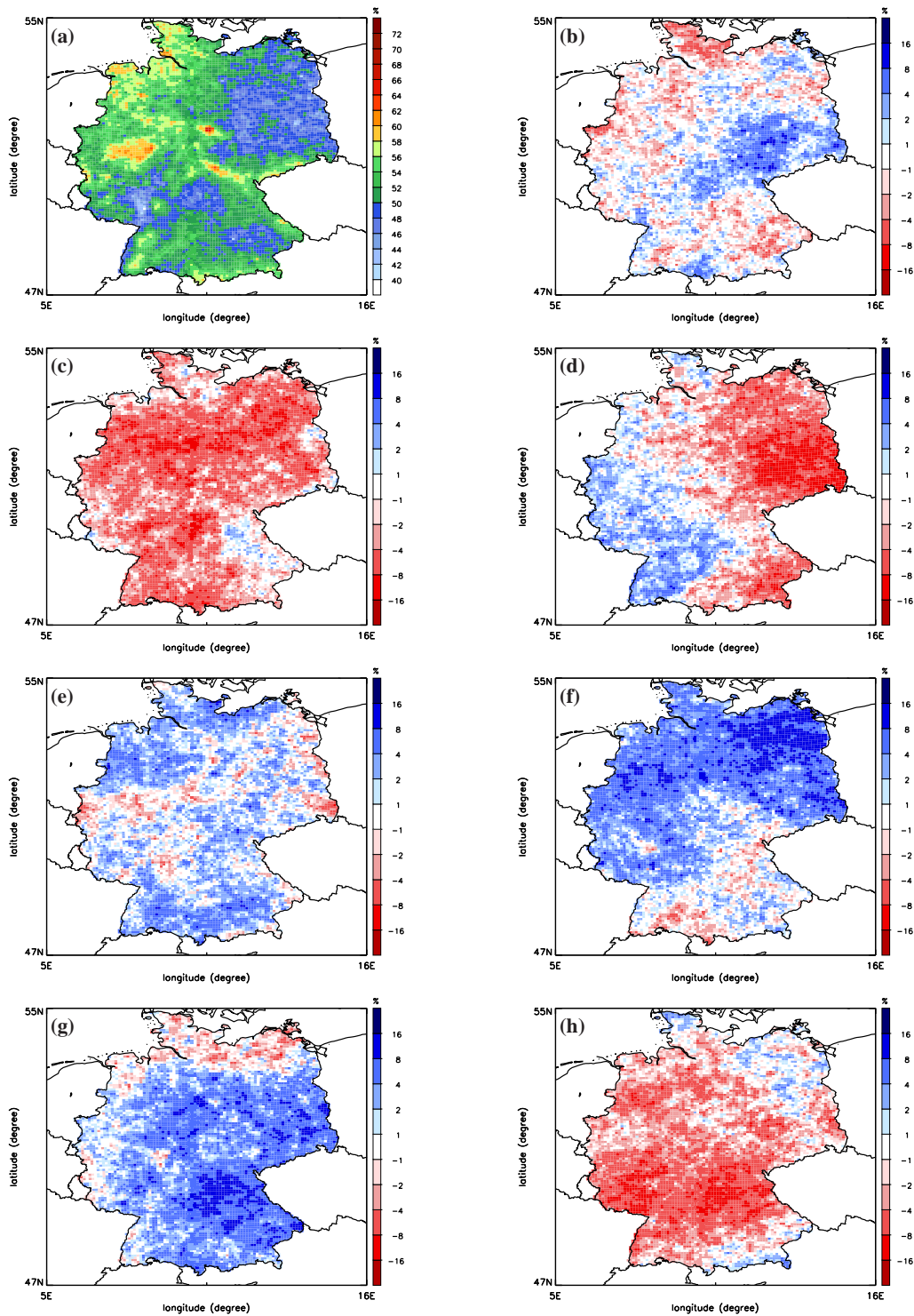


Abbildung 2.9. Wie Abbildung 2.8, jedoch für die Niederschlagshäufigkeit von Ereignissen von mehr als 1 mm pro Tag

Außerdem wurde die Analyse für den Wochengang der Niederschlagshäufigkeit noch für den Sommerzeitraum durchgeführt, da *Bell et al.* (2008) dafür eine Existenz in den USA aufzeigen konnten. Die statistische Untersuchung ist identisch wie die für das gesamte Jahr. Es wurden 644 Gitterpunkte bestimmt an denen ein statistisch signifikanter Wochengang berechnet wurde. Wie im Fall für das Gesamtjahr müssten bei Verwendung des 90 % Signifikanzniveaus 741 Gitterpunkte rein zufällig eine statistisch signifikant Abweichung zeigen, damit die Nullhypothese abzulehnen ist. Damit kann auch für den Sommer kein Wochengang in der Niederschlagshäufigkeit für Deutschland festgestellt werden.

2.2 Vergleich zwischen Modellsimulation und Beobachtungen

Nachdem im vorangegangenen Abschnitt die beobachtete Verteilung des Niederschlags für Deutschland untersucht wurde, soll an dieser Stelle ein Vergleich beobachteter und simulierter Niederschlagscharakteristiken vorgestellt werden. Für den Vergleich werden Simulationen des COSMO-EU sowie des COSMO-DE benutzt, so dass neben der Beobachtung auch der Vergleich zwischen den beiden Modellen möglich ist. Da das COSMO-DE erst seit 2007 operationell betrieben wird, erfolgt die Untersuchung ausschließlich für 2007 und 2008.

2.2.1 Differenz der Niederschlagssummen

Zunächst soll die Untersuchung der Differenz der jahreszeitlichen Niederschlagssummen erfolgen. Dafür werden die gerasterten Stationsbeobachtungen mit den Resultaten der Simulationen vom COSMO-EU und COSMO-DE verglichen. Im Fall des COSMO-EU werden Tagessummen, unter der Verwendung des 00 UTC Modelllaufs, berechnet, in dem die siebte bis dreißigste Vorhersagestunde akkumuliert wird. Bei der Berechnung der Tagessumme des COSMO-DE ist es notwendig, den 00 UTC und den 12 UTC Lauf zu verwenden, da die Vorhersagezeit nur 21 Stunden beträgt. Es wird jeweils Vorhersagestunde 7 bis 18 benutzt, so dass aus der Kombination beider Läufe die Tagessumme bestimmt werden kann.

Die Abbildung 2.10 enthält eine Darstellung der geografischen Verteilung der mittleren, relativen Differenzen der Niederschlagssummen für den Winter- (obere Zeile) sowie den Sommerzeitraum (untere Zeile) von COSMO-EU (linke Spalte) und COSMO-DE (rechte Spalte) der Jahre 2007 und 2008. Die Vorhersagen der beiden Modelle zeigen im Win-

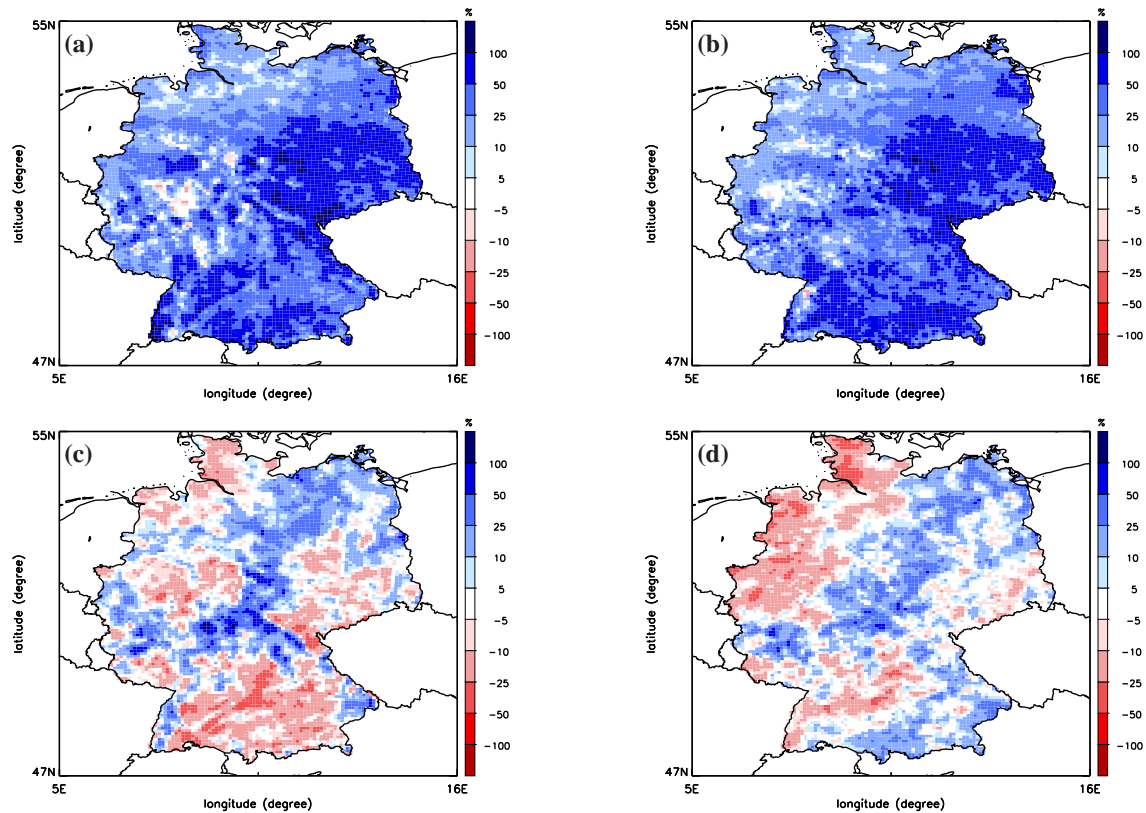


Abbildung 2.10. Geografische Verteilung der relativen Differenz der Niederschlagssumme für den Zeitraum 2007-2008. Die linke Spalte (a,c) zeigt die Ergebnisse für das COSMO-EU, die rechte Spalte (b,d) für das COSMO-DE sowie die obere Zeile für den Winter- (a,b), die untere für den Sommerzeitraum (c,d).

ter eine deutliche Überschätzung der Niederschlagssumme für weite Teile Deutschlands. Ausnahmen bilden jedoch die nordwestlichen Mittelgebirge, wobei das COSMO-EU auf den Höhenlagen eine stärkere Unterschätzung aufweist, als das COSMO-DE. Des Weiteren sind weitere Unterschiede zwischen beiden Modellen aufgrund orographischer Effekte im Bereich des Harzes, des Schwarzwaldes sowie des Thüringer Waldes zu erkennen. In diesen Regionen erzeugen die Simulationen des COSMO-DE ein realistischeres Bild der Niederschlagsverteilung, wenngleich für den Winter eine Überschätzung der Niederschlagssumme festzuhalten ist.

Das Bild der Modellsimulationen für den Sommerzeitraum ist etwas differenzierter. Es sind für beide Modelle Regionen vorhanden, die eine Über- bzw. eine Unterschätzung der Niederschlagssumme aufweisen. Diese Regionen sind nicht identisch in beiden Modellen. Im COSMO-EU sind es hauptsächlich die Gebiete an der deutschen Nordseeküste,

	Beobachtungen	COSMO-EU	COSMO-DE
Σ Jahr [mm/Jahr]	846	982	999
Minimum [mm/Jahr]	459	608	687
Maximum [mm/Jahr]	2404	2804	3296
Winter [%]	23,4	27,3	26,9
Frühling [%]	22,6	23,1	24,1
Sommer [%]	32	28,1	27,4
Herbst [%]	22,1	21,5	21,6

Table 2.2. Charakteristische Werte der Niederschlagsverteilung in Deutschland für den Zeitraum 2007-2008 für Tagessummen der gegitterten Stationsbeobachtungen, COSMO-EU und COSMO-DE.

Süddeutschlands sowie der Süden Nordostdeutschlands, die eine Unterschätzung aufweisen. Eine Überschätzung ist an den Westseiten vom Harz, dem Thüringer Wald und dem Schwarzwald zu erkennen. Beim COSMO-DE sind diese Regionen nicht zu identifizieren. Die Simulationen des COSMO-DE sind in Nordwest- und Norddeutschland durch eine Unterschätzung in der Niederschlagsmenge charakterisiert. Der südliche Abschnitt des Donaeinzugsgebiets ist im COSMO-DE im Vergleich zur Beobachtung zu nass, wohingegen das COSMO-EU hier zu geringe Mengen simuliert.

Ob die realistischere Simulation der Niederschlagsverteilung beim feiner und besser aufgelösten COSMO-DE im Bereich des Harzes, Thüringer Waldes und Schwarzwalds durch eine bessere und genauere Wiedergabe der Topografie ihre Ursache hat oder durch die explizite Simulation von hochreichender Konvektion, ist nicht eindeutig zu klären. Dieses sogenannte Luv/Lee genannte Problem (*Baldauf and Schulz, 2004; Elementi et al., 2005*) tritt im Winter wie im Sommer beim COSMO-EU auf, wobei es im Sommer erheblich stärker ausgeprägt ist. Für diesen Effekt wird hauptsächlich die Parametrisierung von hochreichender Konvektion verantwortlich gemacht. Ob das auch für den Winter gilt, wo eher selten mit konvektiven Situationen zu rechnen ist, könnte mit Hilfe der Unterteilung in konvektiven und gitterskaligen untersucht werden. Letztendlich bleibt festzuhalten, dass das COSMO-DE diese Charakteristik nicht zeigt, im Fall des Schwarzwaldes sogar eine Unterschätzung der Menge im Sommer aufweist, was Probleme bei der Auslösung von konvektiven Situationen andeutet (*Wilson et al., 1998; Seifert, 2008*).

Zusätzlich zur Betrachtung der räumlichen Verteilung mittels Abbildungen sollen im Folgenden anhand der Zahlen aus Tabelle 2.2 die Unterschiede zwischen den Modellsimulationen und den Beobachtungen herausgestellt werden.

Als erstes wird dabei die mittlere beobachtete Jahressumme 846 mm näher betrachtet. Es fällt auf, dass die Jahre 2007 und 2008 wohl mehr Niederschlag aufweisen als das achtjährige Mittel (804 mm; siehe Tabelle 2.1.2). Außerdem ist zu erkennen, dass die Modellsimulationen mit 982 mm (COSMO-EU) und 999 mm COSMO-DE die Beobachtungen deutlich überschreiten. Dasselbe Bild ist für den Minimal- und den Maximalwert zu berichten. Bei der relativen Verteilung des Niederschlags auf die einzelnen Jahreszeiten fällt auf, dass der Anteil im Winter in den Simulationen deutlich zu hoch ist und im Sommer zu gering. Dieses Ergebnis konnte bereits mit Hilfe der Abbildung 2.10 verdeutlicht werden. Dagegen liegen die Werte für die Übergangsjahreszeiten, Frühjahr und Herbst, auf vergleichbarem Niveau zu den Beobachtungen.

2.2.2 Unterschiede in den Niederschlagshäufigkeiten

Nachdem im vorangegangenen Abschnitt die Unterschiede in der Niederschlagsmenge der Modelle bzgl. der Beobachtungen sowie untereinander beschrieben wurden, erfolgt jetzt die Diskussion der Verteilung der Niederschlagshäufigkeit sowohl für tägliche als auch anschließend für stündliche Summen.

Tägliche Summen

Bei der Betrachtung der Häufigkeitsverteilung der Niederschlagsereignisse mit mehr als 1 mm/d wurden für die beiden Modelle, COSMO-EU und COSMO-DE, Tagessummen bestimmt. In der Abbildung 2.11 sind die Ergebnisse von Winter und Sommer für die gegitterten Stationsdaten, das COSMO-EU sowie das COSMO-DE dargestellt. Die Simulation der beiden Modelle zeigt für den Winter eine erhebliche Überschätzung der Niederschlagshäufigkeit. Im Fall des COSMO-DE ist diese Überschätzung etwas geringer, allerdings immer noch recht beachtlich. Diese Überschätzung in der Häufigkeit konnte ebenfalls in der Niederschlagsmenge beobachtet werden (Abschnitt 2.2.1). Auch im Sommer ist in beiden Modellen eine Überschätzung der Häufigkeit zu sehen. Obwohl bei der Niederschlagsmenge (Abb. 2.10c,d) ein relativ differenziertes Bild der Verteilung für Deutschland zu erkennen ist, tritt dieses hier nicht auf.

Stündliche Summen

Im Anschluss an die Beschreibung der Häufigkeit der täglichen Niederschlagssummen werden im Folgenden stündliche Daten verwendet. Als Beobachtungen werden die disaggregierten Daten (siehe Kapitel 1.1.2) benutzt. Als Vorhersagedaten werden die gleichen Stunden betrachtet, die zur Berechnung der Tagessummen von COSMO-EU und COSMO-DE verwendet wurden.

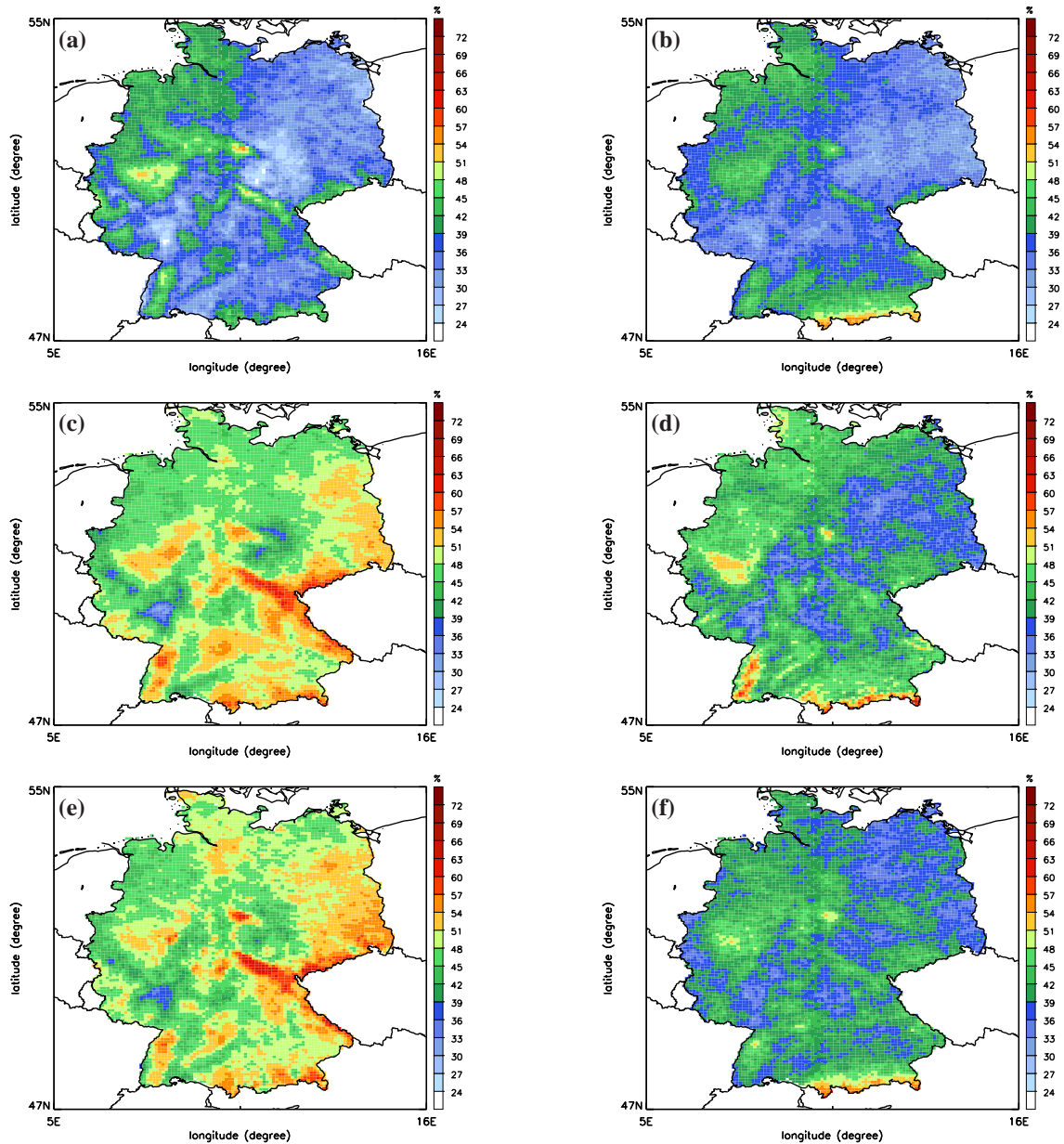


Abbildung 2.11. Geografische Verteilung der relativen Häufigkeit von Niederschlagsereignissen (≥ 1 mm) von täglichen Niederschlagssummen für den Zeitraum 2007-2008. Die linke Spalte (a,c,e) zeigt die Ergebnisse für den Winter-, die rechte Spalte (b,d,f) für den Sommerzeitraum. Die obere Zeile enthält die Ergebnisse der gegitterten Stationsdaten (a,b), die mittlere Zeile die des COSMO-EU (c,d) und die untere Zeile die des COSMO-DE (e,f).

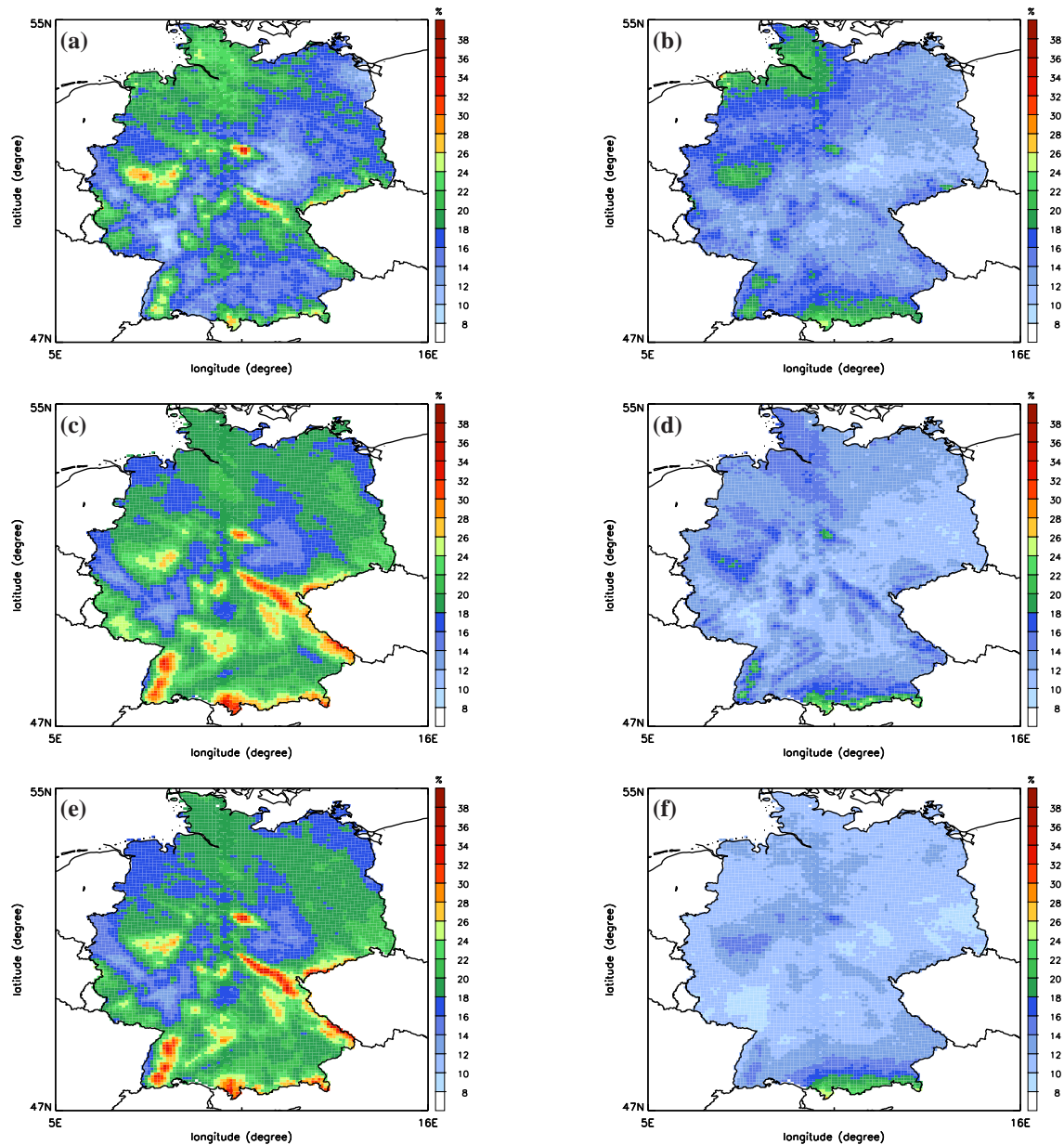


Abbildung 2.12. Wie Abbildung 2.11, jedoch für stündliche Summen und Ereignisse mit mehr als 0,1 mm pro Stunde.

Abbildung 2.12 zeigt die Ergebnisse für Winter (linke Spalte) und Sommer (rechte Spalte) mit den Beobachtungen in der oberen, das COSMO-EU in der mittleren und das COSMO-DE in der unteren Zeile. Die Simulationen kommen im Winter qualitativ gesehen relativ gut an die Beobachtungen heran, d.h. die geografische Position der Maxima und Minima

ist vergleichbar. Allerdings fällt bei der Betrachtung der Werte auf, dass sie in den Simulationen gegenüber den Beobachtungen leicht erhöht sind. Beim Vergleich der Ergebnisse der Modellsimulationen bleibt festzuhalten, dass die Struktur der Maxima in den Mittelgebirgen beim COSMO-DE den Beobachtungen ähnlicher ist als beim COSMO-EU. Dies dürfte auf die bessere und genauere Repräsentation der Topografie zurückzuführen sein. Das Bild für den Sommer sieht etwas anders aus. Besonders sticht das Fehlen des beobachteten Häufigkeitsmaximums im Bereich der deutschen Nordseeküste heraus, welches weder das COSMO-EU noch das COSMO-DE simulieren können. Dagegen sind beide Modelle in der Lage die erhöhte Niederschlagshäufigkeit an den Alpen zu erzeugen. Die Häufigkeitsmaxima im Schwarzwald, Harz und Thüringer Wald können im COSMO-EU qualitativ einigermaßen wiedergegeben werden, wohingegen beim COSMO-DE in diesen Gebieten eine erhebliche Unterschätzung auftritt.

2.2.3 Unterschiede in der Intensität der Niederschlagsereignisse

Nachfolgend zur Betrachtung der Niederschlagsmenge und der -häufigkeit soll nun die geografische Verteilung der 95 %-Perzentilwerte vorgestellt werden. Auch hier werden wiederum tägliche und stündliche Summen untersucht.

Tägliche Summen

Die hier betrachteten täglichen Summen wurden, ebenso wie in Abschnitt 2.2.1 beschrieben, berechnet. Abbildung 2.13 zeigt die räumliche Verteilung des 95 %-Perzentilwerts für gegitterte Stationsbeobachtungen sowie den Simulationen von COSMO-EU und COSMO-DE für Winter und Sommer.

Der Vergleich der zweijährigen Beobachtungszeitreihe mit dem kompletten Achtjahreszeitraum (siehe Abb. 2.7 in Abschnitt 2.1.3) zeigt kaum qualitative Unterschiede, so dass die grundlegenden Eigenschaften der Verteilung enthalten sind. Am ehesten sind diese im Sommer zu beobachten, was sich damit erklären lässt, dass die stärksten Ereignisse durch lokal auftretende Gewitter hervorgerufen werden. Aus diesem Grund sind die Maxima und Minima für den Zweijahreszeitraum ein wenig stärker ausgeprägt.

Die Modellsimulationen für den Winter (Abb. 2.7c,e) zeigen die Maxima an den beobachteten Orten, allerdings sind die 95 %-Perzentilwerte leicht erhöht. Im Sommer hat die geographische Verteilung beider Modelle (Abb. 2.7d,f) kaum eine Ähnlichkeit mit den Beobachtungsergebnissen, sondern die Werte sind größtenteils deutlich erhöht. Einzig die erhöhten 95 %-Perzentilwerte des COSMO-EU an den Alpen finden ein äquivalent in den Beobachtungen. Dagegen sind beim COSMO-DE (Abb. 2.7f) deutlich erhöhte Werte in

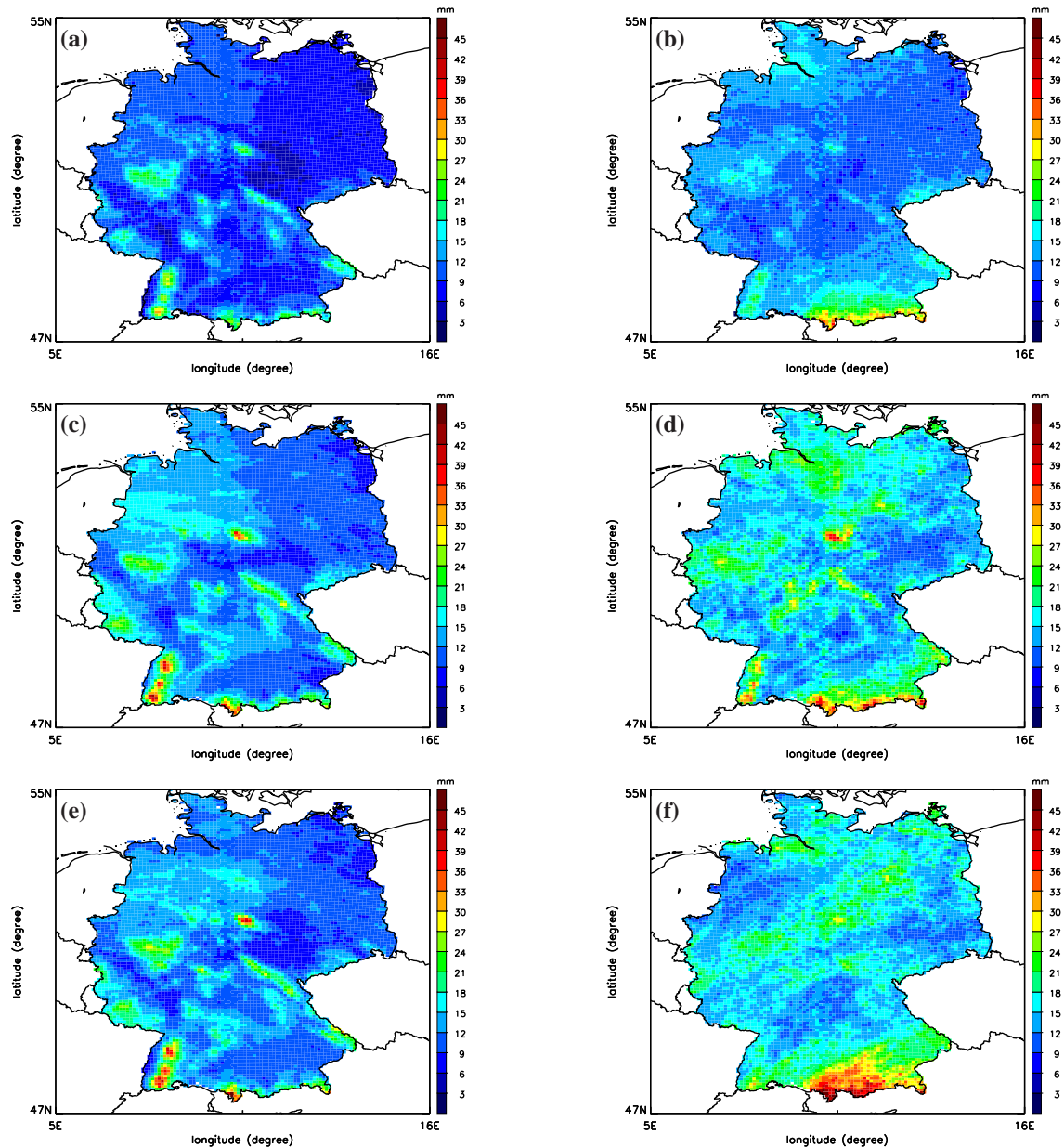


Abbildung 2.13. Wie Abbildung 2.11, aber für das 95 %-Perzentil.

den Alpen zu erkennen. Für den restlichen Teil Deutschlands sind die 95 %-Perzentilwerte in beiden Modellen deutlich höher als in den Beobachtungen.

Auch in den Ergebnissen für das COSMO-DE sind Merkwürdigkeiten zu konstatieren, besonders augenscheinlich tritt dies in Mecklenburg-Vorpommern auf, jedoch ohne die bandartigen Strukturen des COSMO-EU.

Stündliche Summen

Zum Abschluss dieses Teilkapitels erfolgt die Betrachtung der räumlichen Verteilung des 95 %-Perzentilwertes stündlicher Niederschlagssummen. Abbildung 2.14 zeigt Karten für die Ergebnisse, analog zu der Abbildung 2.11. Im Vergleich zu den Werten der täglichen Summen sind diese deutlich geringer, was auf die größere Anzahl von Ereignissen und die

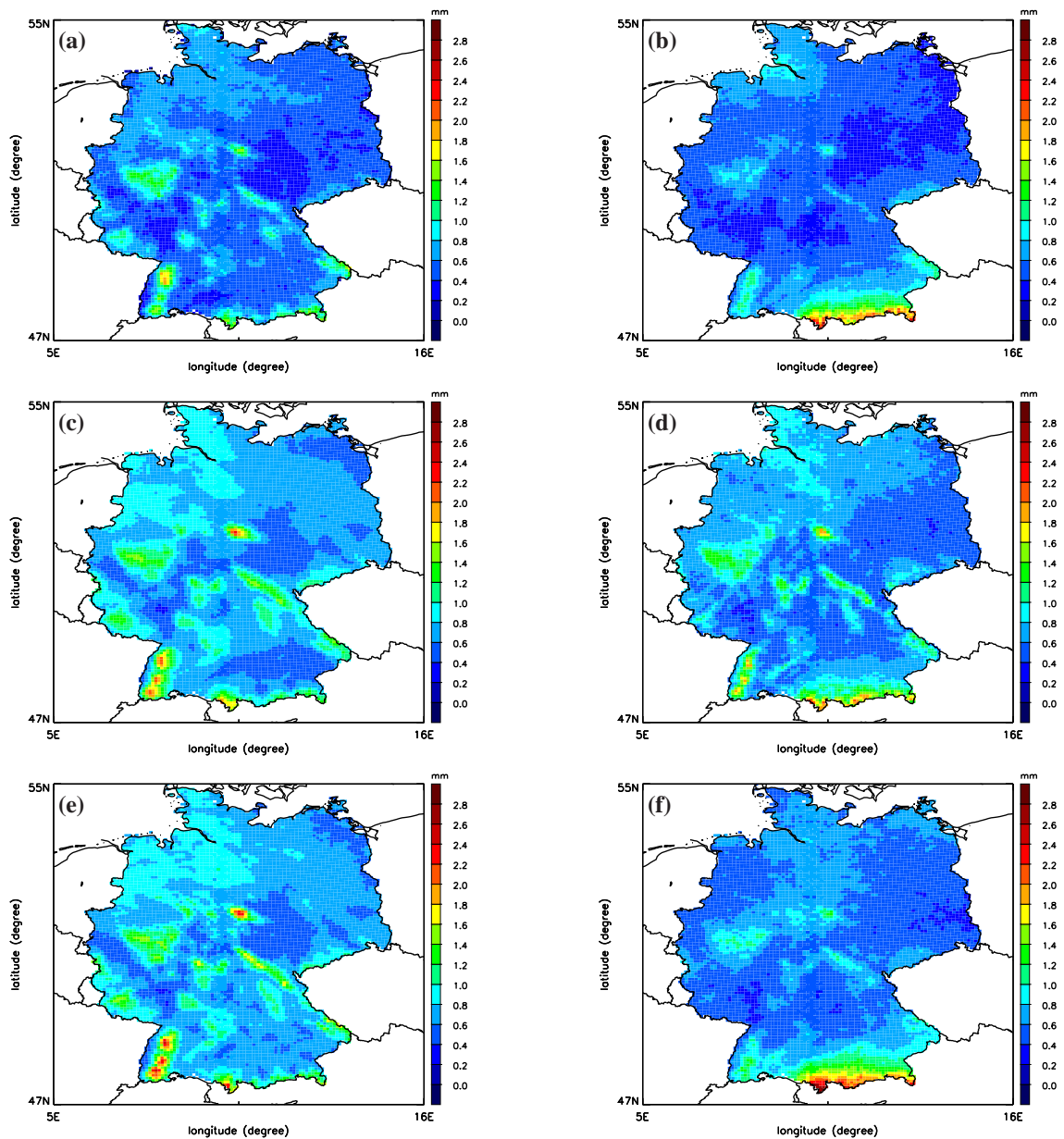


Abbildung 2.14. Wie Abbildung 2.11, jedoch für das 95 % Perzentil stündlicher Summen.

kürzere Akkumulationszeit zurückzuführen ist. Die stündlichen Ergebnisse für den Winter sind qualitativ mit den täglichen Summen vergleichbar, d.h. die Simulationen weisen leicht höhere Werte im Vergleich zu den Beobachtungen auf.

2.3 Der Tagesgang des Niederschlags

Der Tagesgang des Niederschlags, vor allem im Sommer, ist ein sehr markantes Merkmal. Die Simulation des Tagesgangs stellt für die numerische Wettervorhersage ein erhebliches Problem dar (z.B. *Dai et al.*, 1999). Zuerst soll an dieser Stelle beschrieben werden, welche Charakteristiken der beobachtete Tagesgang in Deutschland für verschiedene Jahreszeiten besitzt. Danach soll exemplarisch mit Hilfe von Simulationen zweier NWV Modelle, COSMO-EU und COSMO-DE, gezeigt werden, welche Schwierigkeiten es dabei gibt.

2.3.1 Beobachtungen

Ein weiterer interessanter Aspekt der mit Hilfe des disaggregierten Datensatzes betrachtet werden soll, ist der Tagesgang der Niederschlagssumme. Der Tagesgang ist insbesondere im Sommer ein charakteristisches Merkmal (z.B. *Janowiak et al.*, 2007) des Niederschlags. Für die Analyse wird zuerst an jedem Gitterpunkt der achtjährigen Klimatologie (2001-2008) nach Jahreszeiten unterteilt diejenige Stunde ermittelt, die im Mittel den meisten Niederschlag aufweist. Hauptsächlich werden der Sommer und der Winter betrachtet. Ob an einem Gitterpunkt tatsächlich ein Tagesgang beobachtet wird, soll unter Verwendung des χ^2 -Test, wie in Kapitel 2.1.4 im Fall des Wochengangs, untersucht werden. Als Nullhypothese wird angenommen, dass es keinen signifikanten Tagesgang gibt, d.h. das Auftreten der Niederschlagsmenge über den Tag entspricht einer uniformen Verteilung. Als Irrtumswahrscheinlichkeit wird $\alpha = 0,1$ verwandt. Kann die Nullhypothese mit dem benutzten α nicht abgelehnt werden, so kann von einer Existenz eines statistisch signifikanten Tagesgangs für diesem Gitterpunkt ausgegangen werden. Diese Vorgehensweise ist ein Unterschied zur Arbeit von *Paulat* (2007), in der die Existenz eines Tagesgangs angenommen wurde, falls der maximale stündliche Niederschlagswert das tägliche Mittel um mindestens 30 % übersteigt. Unter Zuhilfenahme des χ^2 -Tests ist somit eine Aussage über die Signifikanz des Tagesgangs möglich und erfolgt demnach nicht mithilfe eines subjektiven Kriteriums.

Die Ergebnisse dieser Analyse sind als Deutschlandkarte in Abbildung 2.15 für Sommer (a) und Winter (b) dargestellt. Diejenigen Gitterpunkte für die kein signifikanter Ta-

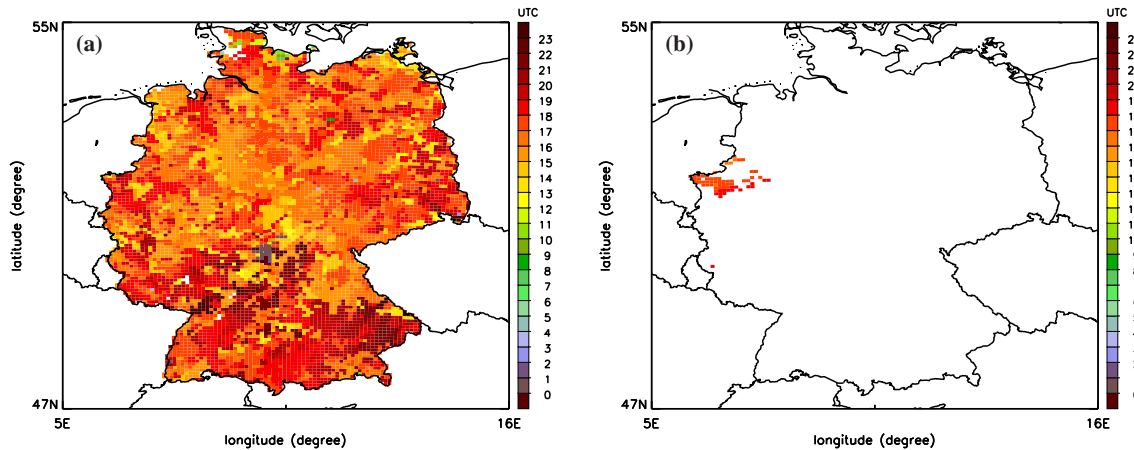


Abbildung 2.15. Geografische Verteilung der Uhrzeit [UTC] des Maximums im Tagesgang des Niederschlags für jeden Gitterpunkt des disaggregierten Beobachtungsdatensatzes (Deutschland, 2001-2008): (a) Sommer und (b) Winter. Ein Maximum wird nur dann dargestellt, wenn statistisch signifikant (Irrtumswahrscheinlichkeit $\alpha = 0,1$) ein Tagesgang am jeweiligen Gitterpunkt existiert, ansonsten wird keine Uhrzeit angegeben.

gesang bestimmt werden konnte, werden mit keiner Farbe dargestellt. In den Sommermonaten kann in weiten Teilen Deutschlands ein eindeutiger Tagesgang beobachtet werden. Das Maximum des Niederschlags wird zwischen dem späten Nachmittag und dem Abend ermittelt. An der Nordseeküste, vor allem in Schleswig-Holstein, wird kein Tagesgang im Niederschlag festgestellt. Die Erklärung hierfür dürfte der Land-See-Wind Effekt sein, der die Bodenerwärmung weniger stark werden lässt, so dass hier nicht die typische lokale Sommerkonvektion der hauptverantwortliche Prozess für die Niederschlagsentstehung ist. Für Süddeutschland, vor allem im Alpenvorland, ist die beobachtete Stunde des maximalen Niederschlags geringfügig später und teilweise sogar in den Morgenstunden, was durch die Orographie der Alpen und des Schwarzwaldes hervorgerufen wird (Finke and Hauff, 1996; Lugauer and Winkler, 2005). Vergleichbare Resultate für die mittleren Breiten konnten Dai et al. (1999), Janowiak et al. (2007) sowie Clark et al. (2007) für die USA und Kaufmann et al. (2003) für die Schweiz berichten.

Im Winter (Abb. 2.15b) dagegen wird nur im Westen Deutschlands ein statistisch signifikanter Tagesgang berechnet. Da er allerdings nur eine paar wenige Gitterpunkte betrifft, kann allgemein festgehalten werden, dass im Winter kein Tagesgang beobachtet wird. Nachdem die Stunde mit dem maximalen Niederschlag bestimmt und bezüglich der Signifikanz getestet ist, erfolgt die Betrachtung über die relative Amplitude des Tagesgangs.

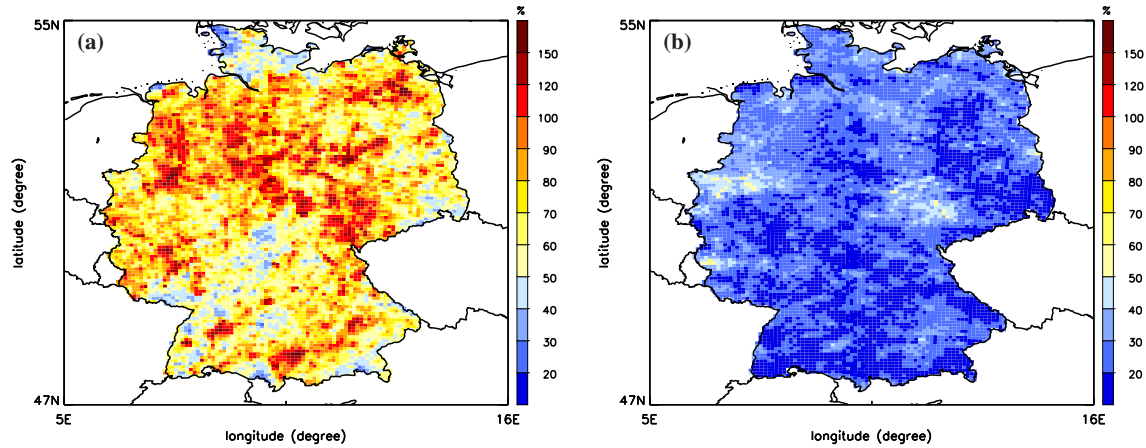


Abbildung 2.16. Geografische Verteilung der relativen Amplitude [%] im Tagesgang des Niederschlags für jeden Gitterpunkt des disaggregierten Beobachtungsdatensatzes (Deutschland, 2001-2008): (a) Sommer und (b) Winter.

Außerdem können mögliche regionale Unterschiede aufgezeigt werden. Zur Bestimmung der relativen Amplitude des Tagesgangs wird folgendes Kriterium benutzt:

$$R_{norm}(i, j) = \left(\frac{R_{tgmax}(i, j)}{R_{tg}(i, j)} - 1 \right) \cdot 100 \quad (2.1)$$

wobei R_{norm} den normierten maximalen Stundenwert in % am jeweils betrachteten Gitterpunkt (i, j) , $R_{tg}(i, j)$ den über den Tag gemittelten stündlichen Niederschlagswert in mm/h und mit $R_{tgmax}(i, j)$ den maximalen stündlichen Niederschlagswert angibt. Der normierte Maximalwert zeigt an, um wie viel Prozent das Niederschlagsmaximum den mittleren Niederschlagswert übersteigt. In Abbildung 2.16 sind Deutschlandkarten des normierten Tagesmaximums für Sommer (a) und Winter (b) dargestellt.

Der Erwartung entsprechend ist die Amplitude für den Sommer deutlich stärker als im Winter. Die Regionen mit den größten Amplituden im Tagesgang des Niederschlags befinden sich im Norden Deutschlands. Die Ausnahme hierbei stellt Schleswig-Holstein dar, wo für den Nordseeküstenbereich kein signifikanter Tagesgang bestimmt wurde.

2.3.2 Modellsimulationen

Ein wichtiger Aspekt in der Vorhersage von Niederschlag ist die Repräsentation des Tagesgangs. Nachdem für einen achtjährigen Zeitraum die Beobachtungsdaten hinsichtlich

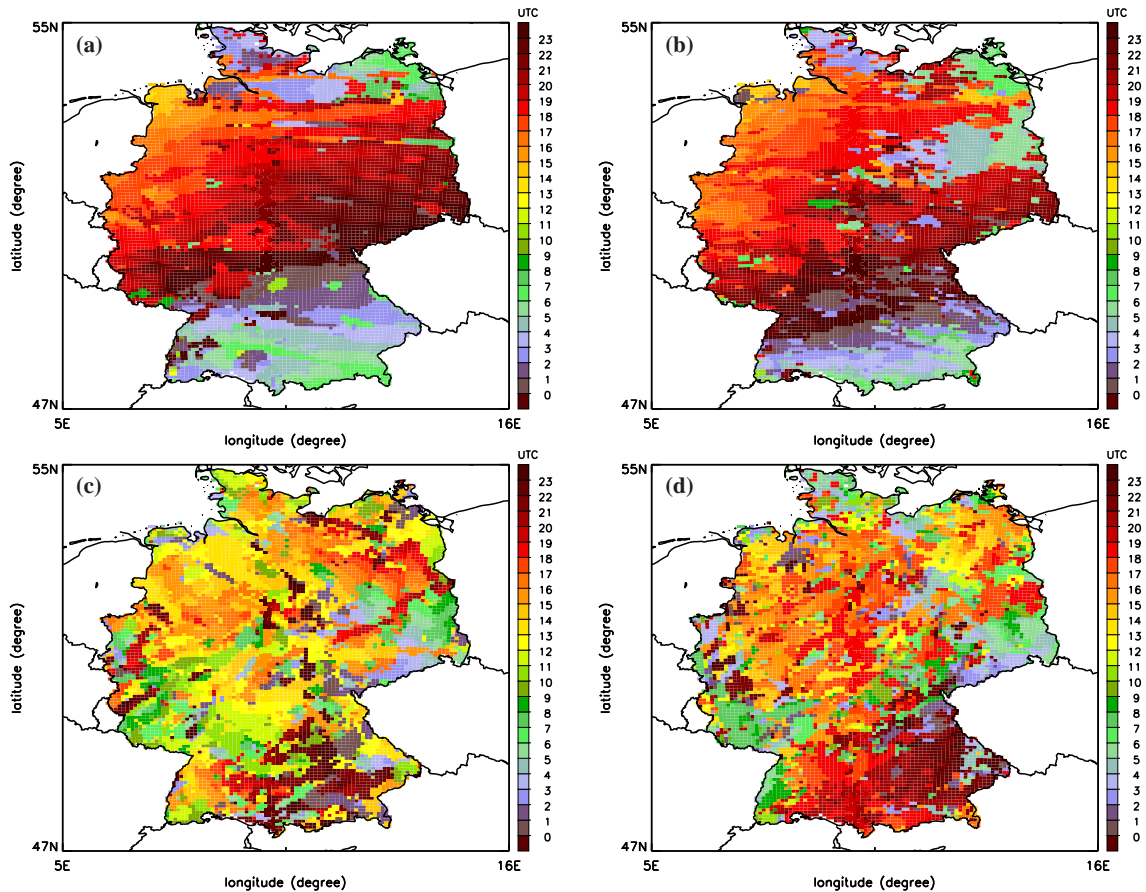


Abbildung 2.17. Geografische Verteilung der Uhrzeit [UTC] des Maximums im Tagesgang des Niederschlags für jeden Gitterpunkt von Modellsimulationen auf Basis der Jahre 2007 und 2008 für Deutschland: (a) Winter COMSO-EU, (b) Winter COSMO-DE, (c) Sommer COSMO-EU und (d) Sommer COSMO-DE. Ein Maximum wird dann als Maximum betrachtet wenn statistisch signifikant ($\alpha = 0,1$) ein Tagesgang am jeweiligen Gitterpunkt existiert, ansonsten wird keine Uhrzeit dargestellt.

der Stunde des maximalen Niederschlags sowie der Stärke untersucht worden ist, soll hier die Darstellung des Tagesgangs in NWV Modellen am Beispiel von COSMO-EU und COSMO-DE betrachtet werden. Hierfür wird nicht ein Achtjahreszeitraum analysiert, sondern es werden die Jahre 2007 und 2008 untersucht, da dass COSMO-DE erst seit 2007 operationell beim Deutschen Wetterdienst betrieben wird. Die Ergebnisse der Beobachtung aus dem vorherigen Abschnitt sind für die beiden Jahre 2007 und 2008 gültig (nicht gezeigt).

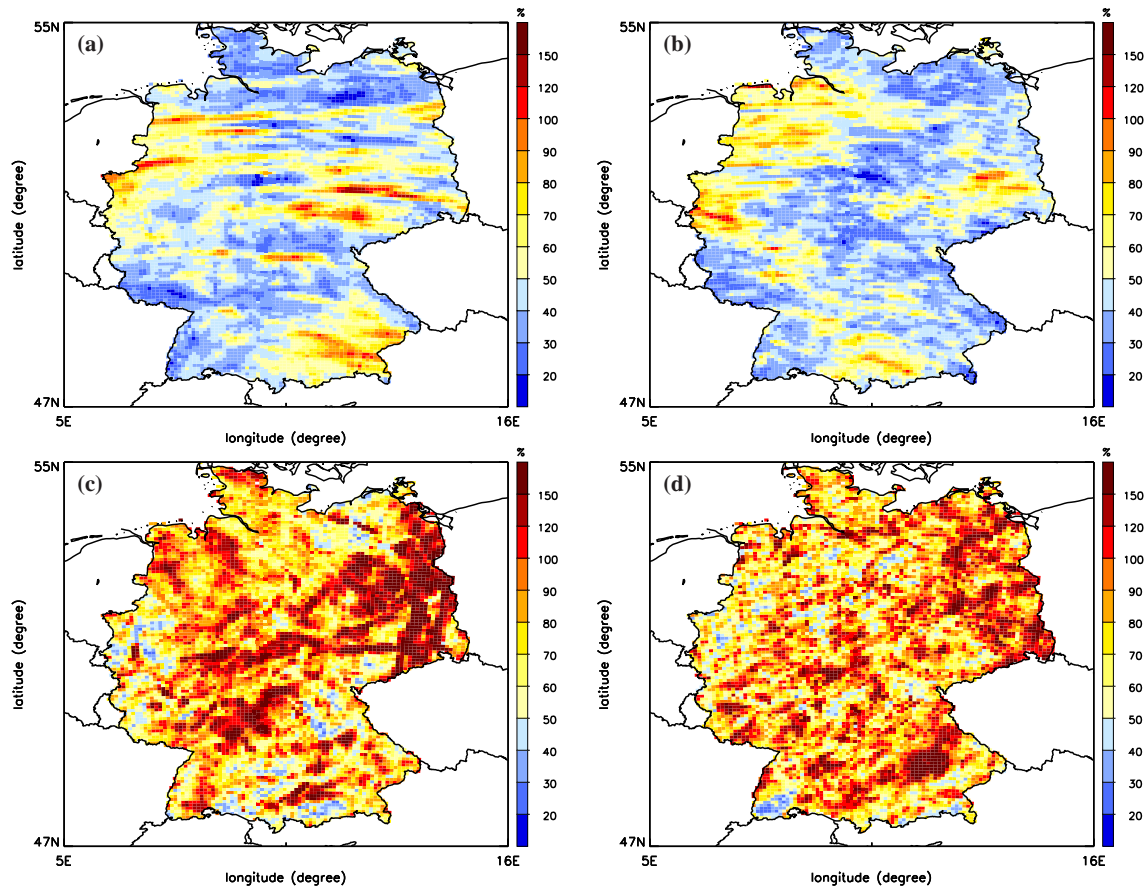


Abbildung 2.18. Wie Abbildung 2.17, jedoch für die relative Amplitude [%] des simulierten Tagesgangs beim Niederschlag.

Im Vergleich zu den Resultaten des disaggregierten Beobachtungsdatensatzes werden in Abbildung 2.17 die geografische Verteilung der Tageszeiten des Niederschlagsmaximums für COSMO-EU und COSMO-DE im Winter und Sommer dargestellt. Zur Bestimmung der Stunde des Niederschlagsmaximums wird für COSMO-EU die siebte bis 30. Vorhersagestunde der 00 UTC Läufe analysiert, wohingegen beim COSMO-DE eine Kombination der siebten bis 18. Vorhersagestunde aus den 00 und 12 UTC Läufen zur Anwendung kommt.

Die Ergebnisse der Modellsimulationen weisen im Winter einen signifikanten Tagesgang im Niederschlag auf, was ein Widerspruch zur Analyse der disaggregierten Niederschlagsdaten ist, wo kein Tagesgang existierte. Die Modellsimulationen untereinander zeigen kaum einen Unterschied und es wird für weite Teile Deutschlands die Stunde mit

der größten Niederschlagsmenge zwischen 16 und 22 UTC bestimmt. Vor allem im Süden Deutschlands, in Schleswig-Holstein und in Mecklenburg-Vorpommern tritt das Maximum während der frühen Morgenstunden auf.

Im Sommer wird in den Modellsimulationen ein signifikanter Tagesgang ermittelt. Allerdings gibt es zwischen beiden Modellen sowie bzgl. der Beobachtungen deutliche Unterschiede. Die Ergebnisse für das COSMO-EU zeigen für weite Gebiete Deutschlands das Auftreten der Stunde mit dem maximalen Niederschlag in den Morgen- und Mittagstunden (zwischen 08 und 15 UTC), d.h. ein deutlich zu frühes Auftreten, welches durch die Parametrisierung hochreichender Konvektion hervorgerufen wird (Clark *et al.*, 2007; Janowiak *et al.*, 2007). Neben der deutlichen tageszeitlichen Streuung ist die räumliche Verteilung erheblich inhomogener im Vergleich zu den disaggregierten Beobachtungen. Bei der Betrachtung der Ergebnisse für das COSMO-DE, das keine Parametrisierung hochreichender Konvektion beinhaltet, fällt es schwer einen bestimmten Tageszeitpunkt als Maximum zu erkennen. Zwei Zeitpunkte kommen besonders häufig vor. Zum einem morgens gegen 06 UTC (grünliche Farben) und zum anderen gegen 20 UTC (rot bis rotbraun). Damit zeigt sich, dass der Zeitpunkt des maximalen Niederschlags im Tagesgang beim COSMO-DE nur unzureichend simuliert wird. Einzig im Donaeinzugsgebiet kann die Uhrzeit des maximalen Niederschlags in den späten Abendstunden korrekt simuliert werden.

Nach dem Vergleich der Stunde des maximalen Niederschlags in Modellvorhersagen mit der Beobachtung, folgen an dieser Stelle die Vorstellung und der Vergleich sowie die Diskussion der relativen Amplitude des Tagesgangs. Die Berechnung der relativen Amplitude erfolgte mit Hilfe von Gleichung 2.1 ist damit analog zu der bei der Beobachtung. Die Darstellung der relative Amplitude ist in Abbildung 2.18 zu sehen. Bei beiden Modellen ist im Winter (Abb. 2.18a,b) die relative Amplitude erheblich größer als in der Beobachtung (Abb. 2.16b). Deshalb kann für beide Modelle ein Tagesgang im Winter diagnostiziert werden. Der Vergleich beider Modelle zeigt geringe Unterschiede in der Stärke der relativen Amplitude, wobei das COSMO-EU (Abb.2.18a) einen etwas ausgeprägteren Tagesgang simuliert. Auch in der geografischen Verteilung treten kleine Unterschiede auf und beim COSMO-EU können teilweise intensive Streifenmuster erkannt werden. Im Sommer simulieren die Modelle ebenfalls eine stärkere relative Amplitude als beobachtet wird. Besonders in der Mitte und in Osten Deutschlands treten diese starken Tagesgänge auf. Wie im Winter sind auch im Sommer die Überschätzungen der relativen Amplitude für das COSMO-EU stärker ausgeprägt. In Schleswig-Holstein konnte im Sommer nur ein schwacher Tagesgang beobachtet werden, dagegen wird in beiden Modellen ein starker Tagesgang festgestellt.

Gemittelter Tagesgang

Nachdem die geografische Verteilung des Tagesgangs im Niederschlag beschrieben worden ist, soll nun die Betrachtung und Analyse des für Deutschland gemittelten Tagesgangs erfolgen. In Abbildung 2.19 werden der Tageszyklus der Niederschlagsmenge und -häufigkeit ($\geq 0,1$ mm/h) für Winter und Sommer aufgezeigt. Die blauen Kurven geben die Ergebnisse des COSMO-EU an, die roten die des COSMO-DE und die grünen die der disaggregierten Beobachtungsdaten. Für die COSMO-EU Simulationen wurden die siebte bis 30. Vorhersagestunde der 00-UTC Läufe verwendet. Beim COSMO-DE erfolgte eine Kombination der siebten bis 18. Vorhersagestunde der 00- und 12-UTC Läufe. Die

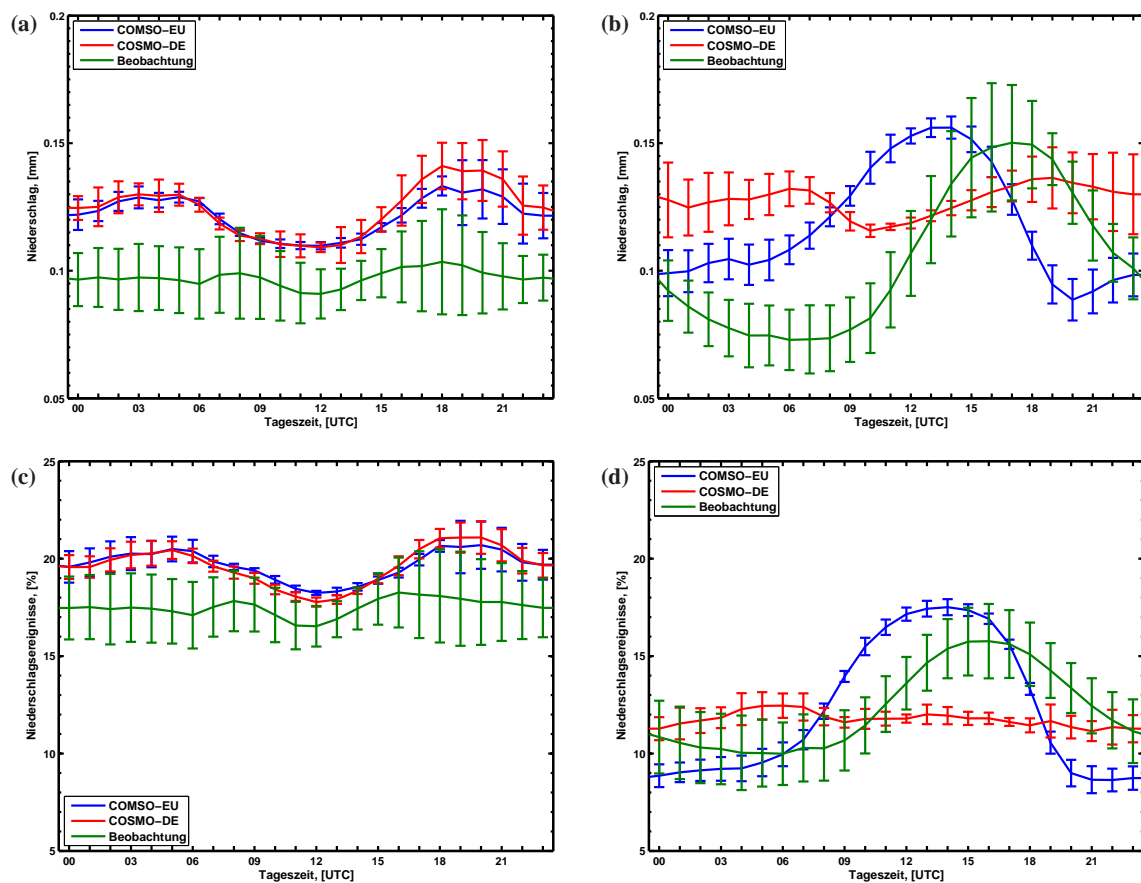


Abbildung 2.19. Mittlerer Tagesgang des Niederschlags in Deutschland (2007-2008) für das COSMO-EU (blau), das COSMO-DE (rot) und die disaggregierten Beobachtungsdaten (grün). In der oberen Reihe ist die Niederschlagsmenge für Winter (a) und Sommer (b) und in der unteren Reihe die Häufigkeit ($\geq 0,1$ mm/h) für Winter (c) und Sommer (d) dargestellt.

Fehlerbalken geben die Standardabweichung für die einzelnen Datensätze zur jeweiligen Stunde an.

Die obere Zeile in Abbildung 2.19 zeigt den Tagesgang der Niederschlagsmenge für Deutschland. Im Winter (Abb. 2.19a) kann kein signifikanter Tagesgang für die Beobachtungsdaten erkannt werden, was äquivalent zu den in Abbildung 2.17 gezeigten Ergebnissen ist. Dagegen zeigen beide Modellsimulationen im Winter einen Tagesgang mit bimodaler Ausprägung sowie eine deutlich zu hohe Niederschlagsmenge. Das stärkere Maximum tritt am Abend zwischen 18 und 20 UTC auf, wobei das COSMO-DE (rote Kurve) ein etwas deutlicher ausgeprägtes Maximum aufweist. Das zweite, weniger stark ausgeprägte Maximum, tritt in den Morgenstunden zwischen 02 und 05-UTC auf und zeigt keinen Unterschied zwischen beiden Modellsimulationen.

Im Sommer kann, analog zu dem in Abbildung 2.17 gezeigten Ergebnis, ein signifikanter Tagesgang in den Beobachtungsdaten erkannt werden. Das Maximum der Niederschlagsmenge tritt gegen 17 UTC auf (analog zu Abb. 2.15). Die Ergebnisse der Modellsimulationen zeigen eine unterschiedliche Realisierung des Niederschlagstagesgangs. Für die COSMO-EU Vorhersagen (blau) kann ebenfalls ein deutliches Maximum bestimmt werden, allerdings tritt dieses gegen 13-14 UTC auf und damit 3-4 Stunden früher als beobachtet. Neben dem zu frühen Auftreten ist auch die Amplitude zu groß. Diese Fehlercharakteristiken sind aus verschiedenen Publikationen wie z.B. von *Clark et al.* (2007) für das WRF-Modell, *Janowiak et al.* (2007) für das Eta-Modell und *Kaufmann et al.* (2003) für das COSMO-7 bekannt. Ein deutlich differenzierteres Bild ist bei den Ergebnissen des COSMO-DE (rot) zu erkennen. Hier sind zwei Maxima vorhanden, eines am frühen Morgen (06-07 UTC) und ein weiteres gegen 19 UTC. Das zweite Maximum findet in den Beobachtungen seinen Gegenpart mit dem Maximum zwei Stunden früher, jedoch zeigt das im COSMO-DE simulierte Maximum einen zu geringeren Wert. Dieses Maximum ist ein Hinweis darauf, dass das COSMO-DE den Tagesgang am Nachmittag und späten Abend besser simuliert als das COSMO-EU. Dieses Merkmal ist von *Clark et al.* (2007) für das WRF ohne Parametrisierung beschrieben worden. Der Unterschied zu den Ergebnissen der Verteilung der Stunde des maximalen Niederschlags (Abb. 2.17) kann mit dem ersten Maximum am Morgen erklärt werden.

Die Ergebnisse der Niederschlagshäufigkeit (untere Zeile in Abb. 2.19) spiegeln im Wesentlichen das Bild für die Summe wieder. Im Winter kann im quantitativen Verlauf kein Unterschied festgestellt werden, weder in der Beobachtung noch in den NWV Modellsimulationen. Die Situation im Sommer ist ein klein wenig anders, der qualitative Verlauf ist ungefähr gleich, allerdings können kleine Unterschiede erkannt werden. Die zeitliche Differenz zwischen den Vorhersagen des COSMO-DE verringert sich auf circa zwei Stunden, wobei das Maximum der beobachteten Niederschlagshäufigkeit etwas früher auftritt

als bei der Menge. Das COSMO-EU weist auch in der Niederschlagshäufigkeit eine Überschätzung auf. Dagegen kann im Tagesgang des COSMO-DE fast kein Maximum erkannt werden.

Zusammenfassung der Resultate

Der gegitterte tägliche Stationsdatensatz, der aktuell eine achtjährige Periode (2001-2008) umfasst, wurde in diesem Kapitel analysiert. Außerdem erfolgte für einen zweijährigen Teilzeitraum (2007-2008) die Darstellung einiger Parameter der Niederschlagsverteilung der stündlichen, disaggregierten Beobachtungsdaten. Außerdem erfolgte für die zweijährige Periode eine Analyse für stündliche und tägliche Summen der Modellsimulationen von COSMO-EU und COSMO-DE.

Die mittleren, jährlichen und saisonalen, Niederschlagssummen aus den Beobachtungen für Deutschland sind der dreißigjährigen Klimatologie des DWD sehr ähnlich. Unterschiede sind für die Minimum- und Maximumwerte zu konstatieren sowie ein deutlich höheres Jahresmittel. Die relative Verteilung auf die einzelnen Jahreszeiten weist kaum Unterschiede auf. Mit Hilfe der gegitterten Stationsdaten wurde außerdem der Frage nachgegangen, ob für die Niederschlagsverteilung in Deutschland ein Wochengang existiert. Es konnte weder für die Menge noch für die Anzahl der Ereignisse ein statistisch signifikanter Wochengang ermittelt werden.

Die zweijährige Niederschlagsklimatologie der Modellsimulationen zeigt deutlich einige Probleme, die bei numerischen Niederschlagsvorhersagen auftreten. Die Simulation des Niederschlags am richtigen Ort beim COSMO-EU (Lee-/Luv-Problem bei Gebirgen, z.B. Schwarzwald und Thüringer Wald), welche beim COSMO-DE erheblich verringert wurden. Im Winter kann die Position der Niederschlagsmaxima und -minima besser simuliert werden als im Sommer, wobei eine deutliche Überschätzung in der Menge auftritt. Im Sommer wird die Niederschlagsmenge besser simuliert, aber die Maxima und Minima befinden sich nicht an den beobachteten Orten.

Mit Hilfe des stündlichen disaggregierten Niederschlagsdatensatzes konnte die räumliche Verteilung der Stunde mit dem maximalen Niederschlag aufgezeigt werden. Im Sommer ist dieser Zeitpunkt meist in den Nachmittagsstunden und am Abend zu finden. Dagegen wird im Winter kein oder ein nur relativ schwach ausgeprägter Tagesgang beobachtet, der statistisch nicht signifikant ist. Die Simulation des Tagesgangs im COSMO-EU für den Sommer ist für weite Teile Deutschlands durch ein zu frühes Auftreten des maximalen Niederschlags und einer zu starken Ausprägung gekennzeichnet. Die COSMO-DE Simulationen zeigen ihr Maximum eher zu spät und haben eine zu schwache Ausprägung im

Tagesgang. Im Gegensatz zu den Beobachtungen zeigen die Modellsimulationen auch im Winter einen Tagesgang, der aber geringer ist als im Sommer. Allerdings kann mittels 90%-Signifikanzniveau der Tagesgang der Modelle statistisch als signifikant betrachtet werden.

Kapitel 3

Das Qualitätsmaß SAL

Zunächst erfolgt in Abschnitt 3.1 die Definition von SAL mit der Beschreibung der Modifikation der Methode gegenüber der in *Wernli et al. (2008)* beschriebenen Vorgehensweise, sowie weiteren Ausführungen zur Interpretation von SAL Resultaten. Daran anschließend werden die Ergebnisse von SAL auf einen Satz von Niederschlagsfeldern des *Spatial Forecast Verification Methods Inter-Comparison Project* in Abschnitt 3.2 vorgestellt. Der letzte Punkt in diesem Kapitel 3.3 ist die Beschreibung der Entwicklung eines Ansatzes zur Bewertung von zeitlichen Fehlern in hochaufgelösten QNVs mit SAL.

3.1 Definition und grundlegende Beschreibung

Die Verifikation von hochaufgelösten Niederschlagsvorhersagen stellt ein schwieriges Problem dar (Kapitel 1.1). Deshalb wurden in den letzten fünfzehn Jahren verschiedene neuartige Bewertungsmaße entwickelt (Kapitel 1.2.2), um der großen räumlichen Variabilität der Niederschlagsverteilung gerecht zu werden. Eine dieser neuartigen Methoden ist SAL, welches während der Promotion von Marcus Paulat (*Paulat, 2007; Wernli et al., 2008*) entwickelt wurde. SAL bewertet eine Niederschlagsvorhersage in einem vorher definierten Gebiet (z.B. einem Flusseinzugsgebiet) auf Struktur- (*S*), Amplituden- (*A*) und Verschiebungsfehler *L* (*engl.: Location*).

3.1.1 Definition von SAL

Die Definition von SAL ist in Abschnitt 2 bei *Wernli et al. (2008)* beschrieben und ausführlich diskutiert worden. SAL ist ein aus drei unabhängigen Komponenten bestehendes

merkmalbezogenes Qualitätsmaß für QNVs. Es untersucht deren Güte für die Eigenschaften Struktur S , Amplitude A und Location L in einem vorher festgelegten Untersuchungsgebiet, z.B. einem Flusseinzugsgebiet. Für die S - und L -Komponente ist es notwendig zusammenhängende Objekte sowohl in der Beobachtung als auch in den Vorhersagen zu identifizieren, wobei kein direktes Zuordnen von beobachteten und simulierten Objekten Voraussetzung ist. Dies ist ein Hauptunterschied zu den in Kapitel 1.2.2 vorgestellten Methoden CRA (Ebert and McBride, 2000) und MODE (Davis et al., 2006). Die Anwendung von SAL beinhaltet folgende drei Schritte:

1. Die Auswahl des Untersuchungsgebiets (d.h. ein Einzugsgebiet, ein Land, ein Modellgebiet oder anderes künstliches Gebiet) stellt einen wichtigen Schritt bei der Verifikation mit SAL dar. Außerdem muss die größtmögliche Distanz innerhalb des Untersuchungsgebietes bestimmt werden, diese Größe spielt bei der Berechnung von L eine bedeutende Rolle.
2. Für die Berechnung von S und L müssen sogenannte Niederschlagsobjekte definiert werden. In Wernli et al. (2008) wird ein Schwellenwert R^* bestimmt durch $R^* = \frac{1}{15}R^{max}$, wobei R^{max} der maximale Gitterpunktwert innerhalb des betrachteten Untersuchungsgebietes ist. Da diese Art der Bestimmung sehr sensitiv auf Ausreißer reagiert (z.B. einzelne Gitterpunkte mit einer sehr hohen Niederschlagsintensität), wird in dieser Arbeit das 95 %-Perzentil aller Gitterpunktwerte im Untersuchungsgebiet größer als 0,1 mm (R^{95}) anstelle von R^{max} als Schwellenwert verwendet. Somit ergibt sich:

$$R^* = \frac{1}{15}R^{95}. \quad (3.1)$$

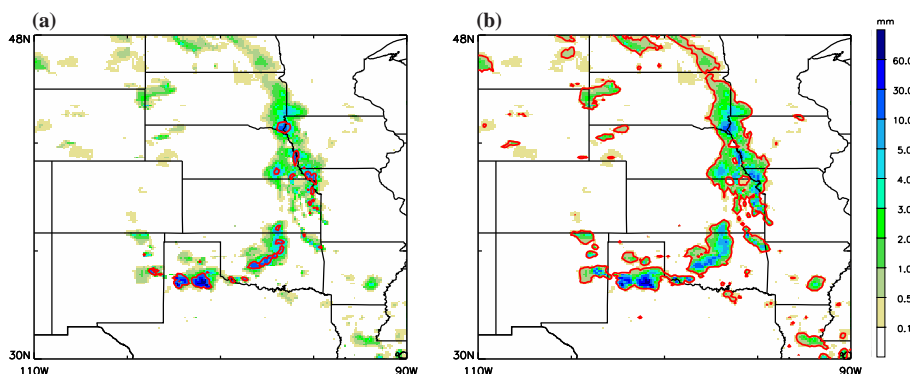


Abbildung 3.1. Beispiel (1. Juni 2005) für die unterschiedliche Schwellenwertbestimmung: Links mit R^{max} (a) und rechts mit R^{95} (b). Die roten Konturen markieren die identifizierten Objekte.

In Abbildung 3.1 ist ein Beispiel dargestellt, wie sensitiv die Bestimmung der Objekte sein kann. Das gewählte Beispiel ist der 1. Juni 2005 aus dem Beobachtungsdatensatz STAGE II (*Lin and Mitchell, 2005*) und zeigt die Niederschlagsituation im mittleren Westen der USA. Da einige wenige Gitterpunkte extrem hohe Werte aufweisen, werden im Fall der Bestimmung der Objektschwellenwerte unter Verwendung des Gitterpunkts mit dem maximalen Wert (Abb. 3.1a) nur die Gebiete mit der größten Intensität als Objekte identifiziert. Es entsteht der Eindruck, dass die so detektierten Objekte nicht die wahre Struktur der Niederschlagsverteilung wiedergeben. Dagegen liefert die Benutzung des 95 %-Perzentilwertes (Abb. 3.1b) ein Ergebnis, das dem einer subjektiven Analyse deutlich besser entspricht.

3. Als letzter Schritt steht dann die Kalkulation der drei Komponenten an, deren Definition im Anschluss folgt.

Amplitude

Die Amplitudenkomponente A entspricht der normierten Differenz der gebietsgemittelten Niederschlagssumme D aus der Vorhersage R_{mod} und Beobachtung R_{obs} :

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0,5[D(R_{mod}) + D(R_{obs})]}. \quad (3.2)$$

Mit Hilfe von A kann die Güte der QNVs im Hinblick auf die Niederschlagsmenge im Untersuchungsgebiet bestimmt werden, ohne dass die räumliche Verteilung beachtet wird. Der Wertebereich von A ist $[-2, \dots, +2]$, wobei mit $A = 0$ eine perfekte Vorhersage des Gebietsmittels verbunden ist. Bei einem Wert von $A = 1$ wird der Niederschlag in der Vorhersage um den Faktor 3 überschätzt. Weitere Werte können aus Tabelle 3.1 entnommen werden.

Die A Komponente ist vergleichbar mit dem mittleren Fehler (ME), allerdings besteht ein klarer Unterschied. Der ME ist ein absolutes Maß und A ist ein relativer Parameter. Beide besitzen immer das selbe Vorzeichen, aber eine Überschätzung der Niederschlagsmenge um 10 % (Faktor = 1,1) führt immer zu einem Wert von $A = 0,095$. Der ME dagegen ist abhängig von den zugrundeliegenden absoluten Niederschlagswerten.

Location

Die L Komponente bewertet den Verschiebungsfehler von QNVs und setzt sich aus zwei Teilen zusammen: $L = L_1 + L_2$. Der erste Term L_1 beschreibt die normierte Distanz der

Faktor	0,1	0,2	0,33	0,5	0,9	1,1	2	3	5	10
Amplitude A	-1,64	-1,33	-1	-0,67	-0,095	0,095	0,67	1	1,33	1,64

Tabelle 3.1. Werte von A für verschiedene Faktoren der Über-/Unterschätzung der Niederschlagsmenge. Ist der Faktor = 1,5, dann entspricht dies einer Überschätzung der simulierten Niederschlagsmenge um 50 % (z.B. $D(R_{mod}) = 3$ und $D(R_{obs}) = 2$ im Fall der Überschätzung um 50 %).

Schwerpunkte, sowohl des simulierten als auch des beobachteten Niederschlagsfeldes,

$$L_1 = \frac{|\mathbf{x}(R_{mod}) - \mathbf{x}(R_{obs})|}{d}, \quad (3.3)$$

mit d als dem Abstand zwischen den zwei am weitesten entfernten Gitterpunkten im Untersuchungsgebiet. Mit $\mathbf{x}(R)$ werden die Koordinaten der Schwerpunkte bezeichnet. Die Werte von L_1 befinden sich im Bereich von $[0, \dots, +1]$, wobei ein Wert von null bedeutet, dass der Schwerpunkt von Vorhersage und Beobachtung identisch ist. Der zweite Term L_2 misst den gemittelten Abstand des Schwerpunkts des gesamten Niederschlagsfeldes mit dem der einzelnen Niederschlagsobjekte. Nachdem Objekte in Beobachtung und Vorhersage bestimmt wurden, wird für jedes Objekt die Niederschlagsmenge ermittelt:

$$R_n = \sum_{(i,j) \in R_n} R_{ij}.$$

Die gewichtete mittlere Distanz zwischen dem Schwerpunkt der einzelnen Objekte (\mathbf{x}_n) und dem des gesamten Niederschlagsfeldes (\mathbf{x}) errechnet sich folgendermaßen:

$$r = \frac{\sum_{n=1}^M R_n |\mathbf{x} - \mathbf{x}_n|}{\sum_{n=1}^M R_n}. \quad (3.4)$$

Der Maximalwert von r beträgt $d/2$, d.h. die Hälfte des Abstands der am weitesten entfernten Gitterpunkte im betrachteten Gebiet. Die Summe im Nenner von Gleichung 3.4 ist nicht identisch mit der Summe aus Gleichung 3.2, da im zweiten Fall alle Gitterpunkte zur Berechnung verwendet wurden und nicht nur die der identifizierten Objektgitterpunkte. Jetzt kann L_2 als die Differenz von r der simulierten und beobachteten Niederschlagsfelder berechnet werden:

$$L_2 = 2 \cdot \left[\frac{|r(R_{mod}) - r(R_{obs})|}{d} \right]. \quad (3.5)$$

Der zweite Term liefert nur dann einen Beitrag, wenn in Beobachtung und/oder Vorhersage mehr als ein Objekt im Untersuchungsgebiet identifiziert werden kann. Der Faktor 2 wurde gewählt, damit der Wertebereich von L_2 mit $[0, \dots, 1]$ identisch zu L_1 ist. Die

komplette L Komponente kann Werte zwischen 0 und 2 annehmen, wobei 0 nur erhalten werden kann, wenn der Schwerpunkt der Vorhersage sowie der gemittelte Abstand zwischen den Objekten dem der Beobachtungen entsprechen. Es soll noch angemerkt werden, dass trotz der Verwendung von L_2 verschiedene Situationen die selben Werte für L liefern können. Außerdem ist die Definition von L nicht ausgelegt um Drehungen um den Schwerpunkt zu bewerten.

Struktur

Die grundsätzliche Idee für S ist der Vergleich der Volumen von normierten Niederschlagsobjekten. Wie in *Wernli et al. (2008)* anhand von Beispielen gezeigt, kann ein solcher Parameter Informationen über Größe und Form von Niederschlagsobjekten liefern. Für jedes Objekt wird ein skaliertes Volumen V_n als Summe der Gitterpunkte $R(i, j)$ des Objektes berechnet:

$$V_n = \sum_{(i,j) \in R_n} R(i, j) / R_n^{max}, \quad (3.6)$$

mit R_n^{max} dem maximalen Niederschlagswert innerhalb des Objekts. Das Volumen V_n wird separat für alle Objekte in Beobachtung und Vorhersage ermittelt. Anschließend wird das mittlere, gewichtete Volumen aus allen skalierten Objekten berechnet:

$$V = \frac{\sum_{n=1}^M R_n V_n}{\sum_{n=1}^M R_n}. \quad (3.7)$$

Daraus folgt die Definition von S als normierte Differenz von V , analog zur A Komponente:

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0,5[V(R_{mod}) + V(R_{obs})]}. \quad (3.8)$$

Der Wertebereich von S liegt wie bei der Amplitude zwischen -2 und +2. Positive Werte von S sind mit Vorhersagen verbunden, die gegenüber der Beobachtung zu große und/oder zu flächige Niederschlagsobjekte simulieren. Dagegen sind negative Werte von S charakteristisch für Situationen in denen die Vorhersagen zu kleine und/oder „spitze“ Objekte enthalten.

3.1.2 Das SAL Diagramm

Nach der Beschreibung von SAL als merkmalsbezogenes Qualitätsmaß liegt der Fokus in diesem Abschnitt auf der Darstellung der Verifikationsresultate. Die Auswertung für ein-

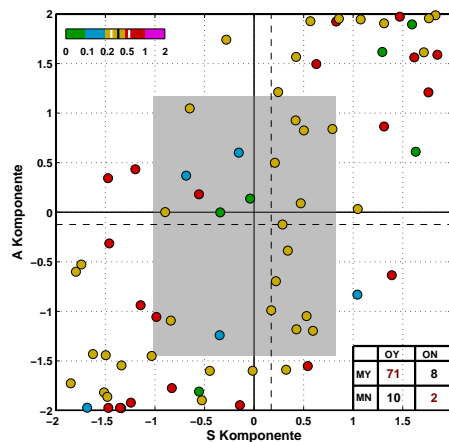


Abbildung 3.2. SAL Diagramm für aus Beobachtungen künstlich erzeugte QNVs. Jedes Symbol steht für das Resultat einer QNV, wobei die L Komponente farbkodiert ist (von Grün = sehr gut bis Magenta = extrem schlecht). Die Kontingenztabelle rechts unten liefert eine Aussage über den Verifikationszeitraum (Anzahl mit verifizierter Zeitpunkte, links oben, ohne Niederschlag, rechts unten, keine Beobachtung, rechts oben und keine Vorhersage links unten). Die schwarzen durchgezogenen Linien geben den perfekten Wert von $S = 0$ (vertikal) und $A = 0$ (horizontal) an und die gestrichelten Linien den jeweiligen Median. Die graue Box gibt durch das 25 %-Perzentil die untere und durch das 75 %-Perzentil die obere Grenze von S (horizontal) und A (vertikal) an. In der Farblegende für L (links oben, keine lineare Einteilung) sind der Median (schwarze Linie) und das 25 %-/75 %-Perzentil (weiße Linien) des Verschiebungsfehlers eingezeichnet.

ge wenige Vorhersagen, wie z.B. bei Fallstudienanalysen, lässt sich gut in Form von Tabellen bewerkstelligen. Bei der Visualisierung der Ergebnisse von SAL von statistischen bzw. klimatologischen Untersuchungen der Güte von Modellsimulationen ist es sinnvoll ein sogenanntes SAL Diagramm (Paulat, 2007; Wernli et al., 2008) zu verwenden. Die Informationen die ein solches Diagramm enthält und wie es zu interpretieren ist, wird im Folgenden beschrieben.

In Abbildung 3.2 ist ein SAL Diagramm der Ergebnisse einer Zufallsvorhersage¹ dargestellt. Diese kompakte Präsentation aller drei Komponenten liefert schnell und einfach einen Überblick über die Vorhersagegüte. In einem solchen Diagramm ist die S gegen die A Komponente aufgetragen. Jeder Punkt visualisiert das Verifikationsergebnis eines Akkumulationszeitraums. Die L Komponente ist farbkodiert (Legende links oben, keine lineare Unterteilung) und reicht von Grün = sehr gut bis Magenta = extrem schlecht). Au-

¹Die Generierung dieses Vorhersagedatensatzes erfolgte durch die zufällige Auswahl eines Zeitpunkts aus dem Beobachtungsdatensatz in der Art und Weise, dass jeder beobachtete Niederschlagszeitpunkt auch in zufälliger Auswahl als Vorhersagefeld dient (Wernli et al., 2008).

Berdem werden einige statistische Maße der drei Parameter dargestellt. Dies ist zum einen der Median von S (vertikal) und A (horizontal) als gestrichelte Linien sowie zum anderen die graue Box, welche durch das 25 %-Perzentil als unterer und das 75 %-Perzentil von S und A als obere Begrenzung festgelegt wird. Der Median von L ist durch eine schwarze und das 25 bzw. 75 %-Perzentil als weiße Linien in der Farblegende dargestellt. Zusätzlich liefert das SAL Diagramm durch die Kontingenztabelle (rechts unten) Informationen über die Anzahl der verifizierten Zeitpunkte. Die Anwendung von SAL, wie auch anderer merkmalsbezogener Qualitätsmaße, kann nur dann erfolgen, wenn sowohl in Beobachtung als auch in der Vorhersage Niederschlagsobjekte identifiziert werden können. Die Zahl links oben in der Kontingenztabelle gibt die Anzahl dieser Situationen im Untersuchungszeitraum an, die Zahl rechts unten die Anzahl derjenigen Zeitpunkte an denen korrekt kein Niederschlag ($S = A = L = 0$) vorhergesagt wurde. Diese Ereignisse werden in die Berechnung der statistischen Maße einbezogen und sind deshalb mit rot hervorgehoben. Dagegen können die anderen Situationen, in denen entweder die Beobachtung oder die Vorhersage kein Niederschlag enthält, nicht verifiziert werden. Im Allgemeinen sollte die Anzahl dieser Ereignisse erheblich geringer sein, damit die statistische Analyse und Interpretation der SAL Resultate sinnvoll ist. Es wird manchmal bei der Darstellung der SAL Diagramme auf die Angabe der Kontingenztabelle verzichtet.

Die genauere Betrachtung des SAL Diagramms (Abb. 3.2) zeigt, dass sich sehr wenige Punkte im Zentrum befinden. Wie es aufgrund der Konstruktion des Experiments zu erwarten ist, sind die Medianwerte von S und A ungefähr null. Der Median von L ist 0,4. Der Interquartilsabstand ist ein robustes Maß zur Abschätzung der Streuung der Ergebnisse. In diesem Fall sind die Werte für $S(=1,8)$ und $A(=2,6)$ sehr hoch. Bei der statistischen Analyse der SAL Resultate sollte nicht nur der Median, sondern ebenso der Interquartilsabstand der Komponenten betrachtet werden. Trotz der großen Streuung der Punkte in diesem SAL Diagramm treten die meisten Punkte im rechten oberen und im linken unteren Quadranten auf. Anschließend werden die einzelnen Quadranten genauer untersucht und werden dafür im Uhrzeigersinn, beginnend mit rechts oben, von I bis IV durchnummeriert.

- **Quadrant I:** S und A sind größer als null, d.h. die Überschätzung der Niederschlagsmenge ist mit zu großen und flächigen Niederschlagsobjekten verbunden. Bei der Zufallsvorhersage befinden sich 35 % in diesem Quadranten.
- **Quadrant II:** $S > 0$ und $A < 0$. Damit werden Situationen charakterisiert bei denen die Unterschätzung der Niederschlagsmenge bei gleichzeitig zu flächigen und/oder zu großen Niederschlagsobjekten auftritt (16 % der Punkte liegen in diesem Quadranten). Weist eine Modellvorhersage überdurchschnittlich viele dieser Ereignisse

auf, d.h. mehr als zufällig zu erwarten wären, dann kann dies als Hinweis gedeutet werden, dass großflächig zu schwache Niederschlagsereignisse, z.B. Niesel anstatt Regen, simuliert werden.

- **Quadrant III:** S und A sind kleiner als null. Damit ist die Unterschätzung der Niederschlagsmenge mit zu kleinen und/oder intensiven Niederschlagsereignissen verbunden (36 % aller Punkte befinden sich in diesem Quadranten). Dieser Quadrant ist der umgekehrte Fall von I.
- **Quadrant IV:** $S < 0$ und $A > 0$ und beschreiben damit Fälle in denen die Überschätzung der Niederschlagsmenge bei gleichzeitig zu kleinen oder „spitzen“ Niederschlagsobjekten vorhergesagt werden (13 % aller Punkte befinden sich in diesem Quadranten). Modelle die sehr viele Punkte in diesem Quadranten aufweisen, simulieren wahrscheinlich häufig sogenannte Gitterpunktstürme (*Bougeault and Geleyn, 1989*), die durch Defizite der Parametrisierung hoch reichender Konvektion verursacht werden.

In Quadrant I und III befinden sich fast dreiviertel alle Punkte. Dieses Charakteristikum der SAL Ergebnisse in dieser Zufallsvorhersage deutet darauf hin, dass das Auftreten von zu großen Niederschlagsobjekten bei zu viel Niederschlag ein sehr häufiges Fehlercharakteristikum darstellt. Ebenso gestaltet sich die Situation bei zu kleinen Objekten die häufig eine Unterschätzung der Niederschlagsmenge aufweisen.

3.2 Spatial Forecast Verification Methods Inter – Comparison Project*

Nachdem die grundlegenden Eigenschaften von SAL im vorangegangenen Abschnitt beschrieben wurden, folgt an dieser Stelle die Anwendung von SAL auf QNVs in Teilregionen der Vereinigten Staaten von Amerika (hauptsächlich im Mittleren Westen und im Süden der USA). Diese QNVs stammen aus dem Datensatzes des internationalen Vergleichsprojekts für räumliche Vorhersagestrategien (*engl.: Spatial Forecast Verification Methods Inter-Comparison Project, SVFM ICP; Ahijevych et al., 2009*). Der Datensatz enthält geometrische (3.2.2), manipulierte (3.2.3) und reale Niederschlagssituationen (3.2.4a) für die SAL sowie klassische Bewertungsmaße (Kapitel 1.2.1), wie der mitt-

*Die Ergebnisse dieses Abschnittes sind in *Weather and Forecasting* veröffentlicht wurden: Wernli, H., Hofmann, C. and M. Zimmer (2009). Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique, **24**, 1472-1484.

lere Fehler (*ME*), die Wurzel des mittleren quadratischen Fehlers (*RMSE*), dem Häufigkeitsbias (*FBI*) und dem Heidke Skill Score (*HSS*).

3.2.1 Einleitung

Zunächst erfolgt an dieser Stelle die Beschreibung des SVFM ICP Projekts und vor allem dessen Ziele. Die Fortschritte in der Wettervorhersage wie auch der Beobachtungssysteme haben in den letzten Jahren zu einer besseren Auflösung und Vorhersage geführt. Allerdings waren die Ergebnisse der klassischen Bewertungsmaße häufig schlechter, da diese Methoden Probleme haben kleinräumige Strukturen zu bewerten oder die Eigenschaften der Fehler aufzuzeigen (siehe Kapitel 1.1). Aus diesem Grund wurden in den letzten Jahren der wissenschaftlichen Forschung viele neuartige Verifikationsmethoden entwickelt. Die Grundidee des SVM ICPs ist der Vergleich der verschiedenen neuartigen Verifikationsmethoden, um den Nutzern Hinweise zu liefern, welche Methode für ihre Anwendung bzw. Problem am besten geeignet ist. Außerdem können die Charakteristiken der einzelnen Methoden aufgezeigt werden, sowie zu einem besseren Verständnis beitragen.

Bisher wurde SAL in *Wernli et al.* (2008) auf künstliche und eine große Anzahl operationeller QNVs des COSMO-7 Modells² für den deutschen Abschnitt des Elbeeinzugsgebiets angewandt. In *Paulat* (2007) werden darüber hinaus die Ergebnisse weiterer deutscher Flusseinzugsgebiete vorgestellt. In *Hofmann et al.* (2009) wurde eine Liste von exzellenten und besonders schlechten Vorhersagen in den Jahren 2002-2006 mit Hilfe von SAL herausgearbeitet. Des Weiteren erfolgte die Anwendung von SAL auf QNVs in der Schweiz (*Jenkner et al.*, 2008) und Belgien (*Weverberg et al.*, 2010) sowie die systematische Untersuchung der Güte von 20 NWV Modellen (*Zimmer and Wernli*, 2008; *Zimmer et al.*, 2009) täglicher Daten im Sommer 2007 für Südwestdeutschland. Bei *Wittmann et al.* (2010) erfolgte die Anwendung von SAL auf dreistündige QNVs von Modellen mit unterschiedlicher horizontaler Maschenweite (Δx zwischen 2,5 und 25 km) im Sommer 2009 für Österreich. Darüber hinaus wurde SAL angewandt, um die Qualität von Niederschlagsfeldern aus Satellitendaten sowie regionalen Modellsimulationen zu vergleichen (*Früh et al.*, 2007).

Der Datensatz des SVFM ICP Projekts bietet sich hervorragend an, um einige der speziellen Fähigkeiten von SAL aufzuzeigen sowie diejenigen Aspekte darzustellen, die durch SAL nicht bewertet werden. Außerdem können die Ergebnisse im Gegensatz zu den vorherigen Studien mit den Resultaten anderer neuartiger Methoden verglichen werden.

²Schweizer Version des COSMO Modells (siehe Kapitel 1.1.1) mit 45 vertikalen Modellschichten sowie ein leicht gegenüber dem COSMO-EU verschobenes Gitter. Die Anfangs- und Randbedingungen werden bei dieser Version durch das ECMWF Modell zur Verfügung gestellt.

3.2.2 Geometrische Beispiele

Ein Datensatz idealisierter Niederschlagsfelder wurde im Rahmen des Vergleichsprojekts (Ahijevych *et al.*, 2009) erzeugt, die als sogenannte „geom“-Fälle G0-G7 (G0 entspricht geom000, etc.) bezeichnet werden. Alle sieben Fälle sind charakterisiert durch ein einziges Niederschlagsobjekt mit einer einfachen geometrischen Form. Aufgrund der speziellen Form (siehe Gl. 1 in Ahijevych *et al.*, 2009) sind die Strukturkomponente S und die Amplitudenkomponente A immer gleich. Aus diesem Grund werden in der Diskussion ausschließlich die Komponenten A und L betrachtet. In Tabelle 3.2 werden die Werte von SAL und von den traditionellen Fehlermaßen für vier ausgewählte Paare der „geom“-Fälle aufgelistet. Wie in den weiter unter präsentierten Ergebnissen der realen Fälle wurden die Daten auf ein reguläres Länge-Breite Gitter mit einem Abstand von $0,1^\circ$ interpoliert, das nicht den gesamten Bereich der Originaldaten umfasst³. Als Folge davon weichen die Werte der traditionellen Maße in dieser Studie leicht von denen in Ahijevych *et al.* (2009) ab, die die Berechnung auf dem Originalgitter durchführten.

Der erste Vergleich von G0 und G2 (mit G0 als der Wahrheit und G2 als Vorhersage, siehe Abb. 3.3) ist verbunden mit einer Verschiebung des Niederschlagsobjektes in der Vorhersage. Dies wird durch den relativ hohen Wert von $L = 0,39$ und annähernd perfekten Werten von A (und S)⁴ unterstützt. Der ME ist annähernd null, da sich positive und negative Abweichungen ausgleichen, d.h. der ME ist nicht in der Lage einen Verschiebungsfehler zu identifizieren. Ebenso ist der FBI Wert mit 1 nah am perfekten Wert und zeigt damit kein Problem in der Vorhersage auf. Der $RMSE$, obwohl deutlich positiv, weist in diesem Fall den geringsten Wert von den vier betrachteten Paaren auf. Völlig gegensätzlich ist

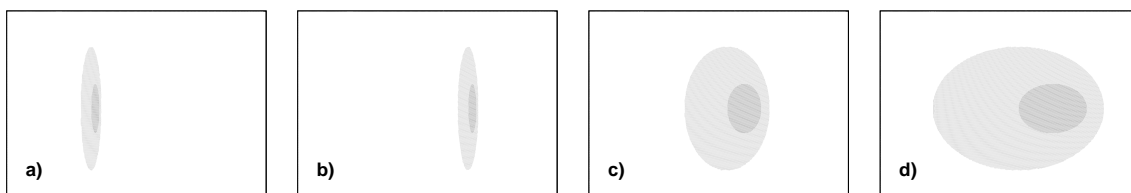


Abbildung 3.3. Synthetische Niederschlagsfelder, die sogenannten „geom“-Fälle: (a) G0, (b) G2, (c) G3 und (d) G5. Die hellere und dunklere Schattierung zeigt schwächeren und intensiveren Niederschlag an.

³Diese Neu-Gitterung wurde durchgeführt, um die Daten für die verwendete Grafiksoftware kompatibel zu machen.

⁴Der Grund weshalb die Werte von A und S nicht genau null sind, ist darin begründet, dass die QNV auf einem diskreten Gitter vorliegt: eine Verschiebung des Niederschlagsobjektes um eine Strecke die nicht einem ganzzahligen Vielfachen des Gitterabstands entspricht, führt zu geringen Interpolationsfehlern.

Vergleich	<i>S</i>	<i>A</i>	<i>L</i>	<i>ME</i>	<i>RMSE</i>	<i>nRMSE</i>	<i>HSS</i>	<i>FBI</i>
G0 vs. G2	-0,03	-0,03	0,39	-0,03	5,51	0,02	-0,07	0,98
G0 vs. G3	1,19	1,19	0,25	2,86	8,77	0,02	-0,12	3,86
G0 vs. G5	1,55	1,55	0,26	6,66	10,77	0,02	0,09	7,68
G3 vs. G5	0,66	0,66	0,01	3,80	7,64	0,01	0,49	1,99

Tabelle 3.2. SAL Werte sowie eine Auswahl an klassischen Bewertungsmaßen (mittlerer Fehler (*ME*, in mm); Wurzel des mittleren quadratischen Fehlers (*RMSE*, in mm); normierter *RMSE* (*nRMSE*); Heidke Skill Score (*HSS*) und der Frequency Bias (*FBI*) für einige ausgewählte Paare der „geom“-Fälle. Für die Berechnung von *HSS* und *FBI* wurde ein Schwellenwert von 0,1 mm verwandt.

der *HSS*, der mit einem Wert nah an null andeutet, dass diese Vorhersage keinen 'Skill' aufweist. Dieses einfache Beispiel zeigt deutlich das unterschiedliche Verhalten bei der Bewertung von QNVs unter Benutzung klassischer Maße. Das Spektrum reicht von annähernd perfekt (*ME* und *FBI*), über den *RMSE* mit nicht wirklich guten, aber mit dem geringsten Wert der vier Beispiele, bis hin zu unbrauchbar (*HSS*). Im Gegensatz dazu liefern die Ergebnisse der drei Komponenten von SAL hilfreiche Informationen, dass *S* und *A* der Vorhersage (fast) perfekt sind aber dass der prognostizierte Niederschlag erheblich zur Beobachtung verschoben ist.

Bei den Paaren G0 vs. G3 und G0 vs. G5 (in beiden Fällen wird G0 als Wahrheit angesehen) ist das Objekt in der Vorhersage verschoben und zu großflächig (siehe Abb. 3.3c,d). Dieses Merkmal wird sehr gut durch SAL (positiver Wert von *L* aber kleiner als für G0 vs. G2; deutlich positive Werte von *A* und *S*) beschrieben. Alle traditionellen Bewertungsmaße weisen schlechte Werte auf, besonders im Fall von G0 vs. G5, wo *ME* und *RMSE* am größten sind, der *FBI* sehr hohe Werte aufweist und der *HSS* nah an null ist. Der Grund für die schlechte Bewertung von G0 vs. G5 durch die klassischen Maße (im Vergleich zu G0 vs. G2) ist darin gegeben, dass Maße wie *ME*, *RMSE* und *FBI* sehr sensitiv auf die räumliche Ausdehnung von Fehlern reagieren. Die Bewertung durch SAL ist dagegen hilfreich: G0 vs. G2 ist schlecht im Sinne von *L*, wohingegen G0 vs. G3 und G0 vs. G5 schlechter sind im Sinne von *A* und *S*. Sicherlich ist die Präferenz von QNVs mit niedrigen Werten bei einer der drei Komponenten abhängig von Benutzer und der Anwendung.

Der Vergleich von G3 vs. G5 ist interessant, da dies der deutlich beste Fall im Sinne von *HSS* und der Ortskomponente von SAL ist. Die beiden anderen Parameter von SAL zeigen, dass G5 die Größe und die Menge des Niederschlags überschätzt, was einhergeht mit moderaten Werten von *ME*, *RMSE* und *FBI*. Es erwächst aus diesen einfachen Beispielen, dass der *HSS* besonders sensitiv auf Verschiebungsfehler anspricht, so dass

eine erhebliche Überlappung der Niederschlagsobjekte für einen deutlich positiven Wert notwendig ist. Diese Resultate sind konsistent mit denen der Analyse eines anderen (variablen) Satzes idealisierter, geometrischer Fallbeispiele aus Wernli *et al.* (2008).

3.2.3 Manipulierte Niederschlagsfelder

Dieser Satz von QNVs wurde erstellt durch künstliche Manipulation der Vorhersage des WRF (Weather Research and Forecasting) Modells (siehe Abschnitt 3.2.4) mit einer horizontalen Gitterweite von 2 km. Im Folgenden werden drei sogenannte „fake“-Fälle mit der nicht manipulierten Vorhersage F0 (F0 entspricht der „fake“-Vorhersage fake000, etc.) verglichen. Die betrachteten Fälle sind F3 (Niederschlagsfeld wurde nach Südosten verschoben), F6 (verschoben wie F3 und zusätzlich multipliziert mit einem Faktor von 1,5: $R(F6) = 1,5 \times R(F3)$) und F7 (verschoben wie F3 und zusätzliche Subtraktion von 1,27 mm: $R(F7) = \max(0, R(F3) - 1.27 \text{ mm})$). Diese drei modifizierten Fälle F3, F6 und F7 sind in Abbildung 3.4 dargestellt, bei dem die rote Kontur dem Schwellenwert R^* zur Identifikation der Niederschlagsobjekte entspricht (siehe Abschnitt 3.1). Es ist klar, dass für F0, F3 und F6 die identifizierten Objekte im Inneren des Untersuchungsgebiets identisch sind, da weder eine Verschiebung des Niederschlagsfeldes (F3) noch eine lineare Skalierung (F6) einen Einfluss auf die Bestimmung des Objektschwellenwertes haben. Allerdings führt die Verschiebung an den Grenzen der Untersuchungsregion des Niederschlagsgebiets dazu, dass einige Niederschlagsobjekte die Untersuchungsregion verlassen und andere wiederum das Gebiet betreten, so dass A und S beim Vergleich von F0 mit F3 nicht gleich null sind (siehe Tabelle 3.3).

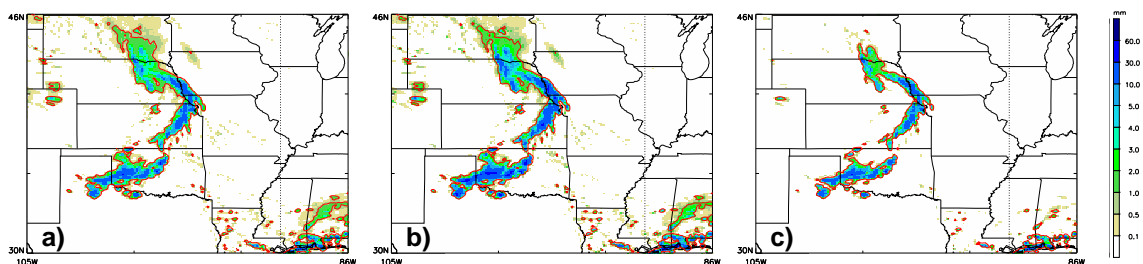


Abbildung 3.4. Künstlich manipulierte Niederschlagsfelder, die sogenannten „fake“-Fälle: (a) F3 (Niederschlagsfeld verschoben in südöstliche Richtung), (b) F6 (zusätzliche Multiplikation eines Faktors 1,5) und (c) F7 (Verschiebung und Verringerung um einen Niederschlagswert von $R_0 = 1,27 \text{ mm}$). Die rote Konturlinie gibt den Schwellenwert R^* der Niederschlagsobjekte an (die Farbskalierung ist identisch für alle Abbildungen in diesem Abschnitt).

Vergleich	S	A	L	ME	$RMSE$	$nRMSE$	HSS	FBI
F0 vs. F3	0,03	-0,01	0,03	0,00	2,91	0,02	0,39	1,02
F0 vs. F6	0,03	0,38	0,04	0,23	3,73	0,02	0,40	1,03
F0 vs. F7	-0,26	-0,31	0,06	-0,13	2,82	0,02	0,27	0,51

Tabelle 3.3. SAL Werte und klassische Bewertungsmaße (siehe Tabelle 3.2) für ausgewählte Paare der „fake“-Fälle.

Besonders interessant ist der Vergleich von F0 mit F7. Wie erwartet wird bei der Subtraktion eines konstanten Wertes wie im Fall von F7 die Niederschlagsmenge im Untersuchungsgebiet unterschätzt und führt zu einem negativen Wert von $A = -0,31$. Darüber hinaus weist auch die Strukturkomponente einen ähnlichen negativen Wert ($S = -0,26$) auf. Diese tritt auf, da die Niederschlagsobjekte kleiner werden wenn die Niederschlagswerte kleiner $R_0 = 1,27$ mm auf null gesetzt werden (vergleiche Abb. 3.4a,c). Dieser Effekt auf S durch die Subtraktion eines konstanten Wertes R_0 ist besonders stark in Fällen variabler Niederschlagsobjekte, z.B. in einer Situation in der große Teile des Niederschlagsobjektes durch schwachen stratiformen Niederschlag und ein geringer Teil durch starken konvektiven Niederschlag geprägt sind. Wenn in derartigen Fällen (Abb. 3.5a) ein konstanter Wert subtrahiert wird (was mit stratiformen Niederschlagsmengen vergleichbar angenommen wird), ändert sich die Größe des Niederschlagsobjekts. Im Gegensatz dazu ist bei einer relativ gleichmäßigen Verteilung des Niederschlags im Objekt (Abb. 3.5b) der Effekt auf die Struktur Komponente durch die Subtraktion eines konstanten Wertes vergleichsweise gering. Für die Verifikation von QNVs bestätigt dieses idealisierte Beispiel, dass die Strukturkomponente die Fähigkeit besitzt zwischen gleichmäßig stratiformen, einer Mischung aus stratiform-konvektiven und reinen konvektiven Niederschlagsfeldern zu unterscheiden. Als Beispiel sei der Fall angeführt bei dem eine

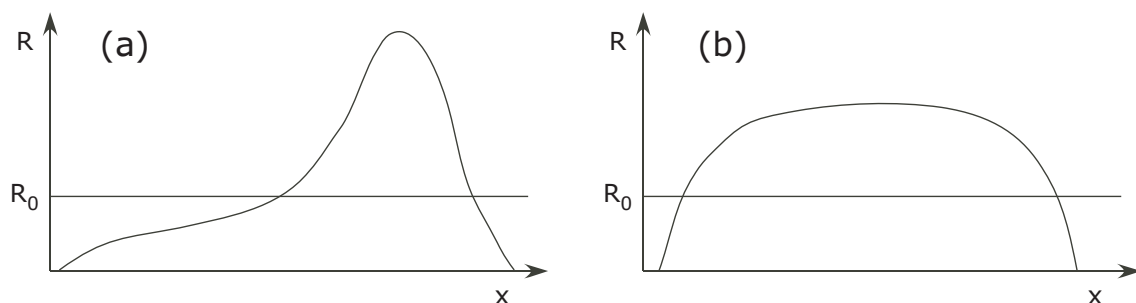


Abbildung 3.5. Schematische Querschnitte die den Einfluss der Subtraktion eines konstanten Wertes R_0 aufzeigen: (a) ein spitzes Niederschlagsobjekt und (b) ein flaches Objekt.

Modellvorhersage lokale konvektive Zellen simuliert allerdings in der Beobachtung der konvektive Niederschlag in großflächigen stratiformen Niederschlag eingebettet ist. Dann weist die Strukturkomponente negative Werte auf.

Die Diskussion und Interpretation der klassischen Bewertungsmaße dieser sogenannten „fake“-Fälle (der Vollständigkeit halber in Tabelle 3.3 aufgelistet) ist in der Publikation von *Ahijevych et al.* (2009) beschrieben.

3.2.4 Reale Situationen von drei Modellen

Das Vergleichsprojekt beinhaltet Vorhersagen von drei Versionen des WRF Modells. Eine mit einer horizontalen Maschenweite von 2 km (bezeichnet als 2CAPS) und zwei mit einem horizontalen Gitter von 4 km (4NCAR und 4NCEP). Alle drei Modellversionen wurden ohne eine Parametrisierung von hoch reichender Konvektion betrieben. Die Diskussion weiterer Details dieser Modellversionen kann aus den Publikationen von *Kain et al.* (2008) und *Ahijevych et al.* (2009) entnommen werden. Stündliche Akkumulationen der drei Modelle⁵ und aus Radarmessungen abgeleitete Niederschlagsbeobachtungen sind verfügbar für 9 Tage im Zeitraum von April - Juni 2005. Diese kleine Stichprobe erlaubt es nicht, eine systematische (statistische) Bewertung der SAL Leistung dieser Modelle vorzunehmen. Daher ist das Ziel dieser Studie nicht das Ermitteln der besten Modellversion, sondern die Diskussion von ausgewählten QNVs, die nicht notwendigerweise repräsentativ für die drei Modellversionen sind. Allerdings kann mit diesen ausgewählten Beispielen aufgezeigt werden, welche Informationen von SAL im Vergleich mit klassischen Bewertungsmaßen erhalten werden können.

a) Ausgewählte Beispiele

Zuerst erfolgt die Analyse der drei Modelle für die QNVs am 1. Juni 2005. Das Hauptziel dieses Beispiels ist das Aufzeigen der Wichtigkeit der Auswahl eines sinnvollen Untersuchungsgebiets für die Verifikation mit SAL. Die Niederschlagsverteilung an diesem Tag war geprägt durch ein langgestrecktes Band (mit deutlichen Unterbrechungen), welches von Nord-Dakota bis nach Nordtexas reichte. Außerdem gab es noch ein schwächeres Niederschlagsgebiet in Alabama und den daran angrenzenden Bundesstaaten (Abb. 3.6a). Als erster Schritt erfolgt die Anwendung von SAL auf das gesamte Modellgebiet mit den horizontalen Ausmaßen von 2000 km × 1600 km (siehe schwarze Rechtecke in den Gra-

⁵Die QNV entspricht der Niederschlagsakkumulation zwischen der 23. und 24. Stunde der Vorhersage, die um 00 UTC gestartet wurde. Der *Gültigkeitstag* wird als Datum in diesem Kapitel verwandt.

Datum	Modelle	Gebiet	S	A	L	ME	$RMSE$	$nRMSE$	HSS	FBI
0601	2CAPS	groß	0,46	0,21	0,06	0,07	2,75	0,03	0,28	0,92
0601	4NCAR	groß	-0,22	0,13	0,07	0,04	2,79	0,03	0,30	0,86
0601	4NCEP	groß	0,68	0,34	0,29	0,12	2,89	0,03	0,19	1,15
0601	2CAPS	Süden	-1,33	0,42	0,08	0,21	2,34	0,06	0,34	0,77
0601	4NCAR	Süden	-1,69	0,40	0,12	0,20	3,01	0,07	0,38	0,67
0601	4NCEP	Süden	-1,57	-0,12	0,26	-0,05	1,34	0,05	0,24	0,67
0603	2CAPS	Süden	-0,79	-0,02	0,26	0,00	0,70	0,12	0,19	0,99
0603	4NCAR	Süden	-0,92	-0,25	0,31	-0,02	0,71	0,12	0,13	0,69
0513	2CAPS	Norden	0,34	0,43	0,24	0,37	4,33	0,03	0,39	1,38
0513	4NCEP	Süden	1,36	1,88	0,16	1,24	4,28	0,83	-0,01	5,89
0513	4NCAR	Norden	0,74	0,35	0,24	0,29	4,13	0,02	0,46	1,25
0513	4NCAR	Süden	-0,22	0,81	0,26	0,05	0,98	0,40	0,01	1,21
0422	2CAPS	Norden	-0,12	0,19	0,18	0,11	2,87	0,03	0,41	1,06
0601	4NCAR	Norden	0,02	0,07	0,22	0,04	3,73	0,03	0,18	0,93

Tabelle 3.4. SAL Werte und klassische Bewertungsmaße (siehe Tabelle 3.2) für ausgewählte QNVs (d.h. Vorhersagen von verschiedenen Modellen in unterschiedlichen Regionen an verschiedenen Tagen). Das Datum (MMDD) bezieht sich auf den Gültigkeitstag der Vorhersage von 00 UTC.

fiken in Abb. 3.6). In diesem großen Gebiet ist bei der Beobachtung das 95 %-Perzentil verbunden mit einem Wert von 6,81 mm/h und der Schwellenwert (Gl. 3.1) für die Bestimmung der Niederschlagsobjekte lautet $R^* = 0,454$ mm/h. Die rote Konturlinie in Abbildung 3.6a entspricht diesem Schwellenwert und zeigt die identifizierten Objekte im beobachteten Niederschlagsfeld (vier große sowie mehrere kleine Objekte). Auf den ersten Blick scheinen alle Vorhersagen das Niederschlagsband im Zentrum des Gebietes ziemlich gut zu erfassen. Die Werte des 95 %-Perzentsils der Vorhersagen und der Beobachtungen sind ähnlich und somit auch der Schwellenwert zur Objektbestimmung. Im Südosten ist es offensichtlich, dass die QNVs durch viele kleine Objekte charakterisiert sind anstatt eines großen Objekts.

Bei der Berechnung der SAL Werte des großen Untersuchungsgebiets (schwarzes Rechteck) folgt, dass das 4NCAR (Abb. 3.6c) für diesen Tag mit den geringsten absoluten Werten bei der Struktur- und der Amplitudenkomponente am besten abschneidet (siehe Tabelle 3.4). Der negative Wert von $S = -0,22$ deutet daraufhin, dass die Niederschlagsobjekte zu klein sind. Die weiteren QNVs besitzen positive Werte für A und S : 2CAPS (Abb. 3.6b) simuliert zu viel Niederschlag bei gleichzeitig zu großen Objekten und 4NCEP (Abb. 3.6d) überschätzt die Niederschlagsmenge mit einem großen Objekt.

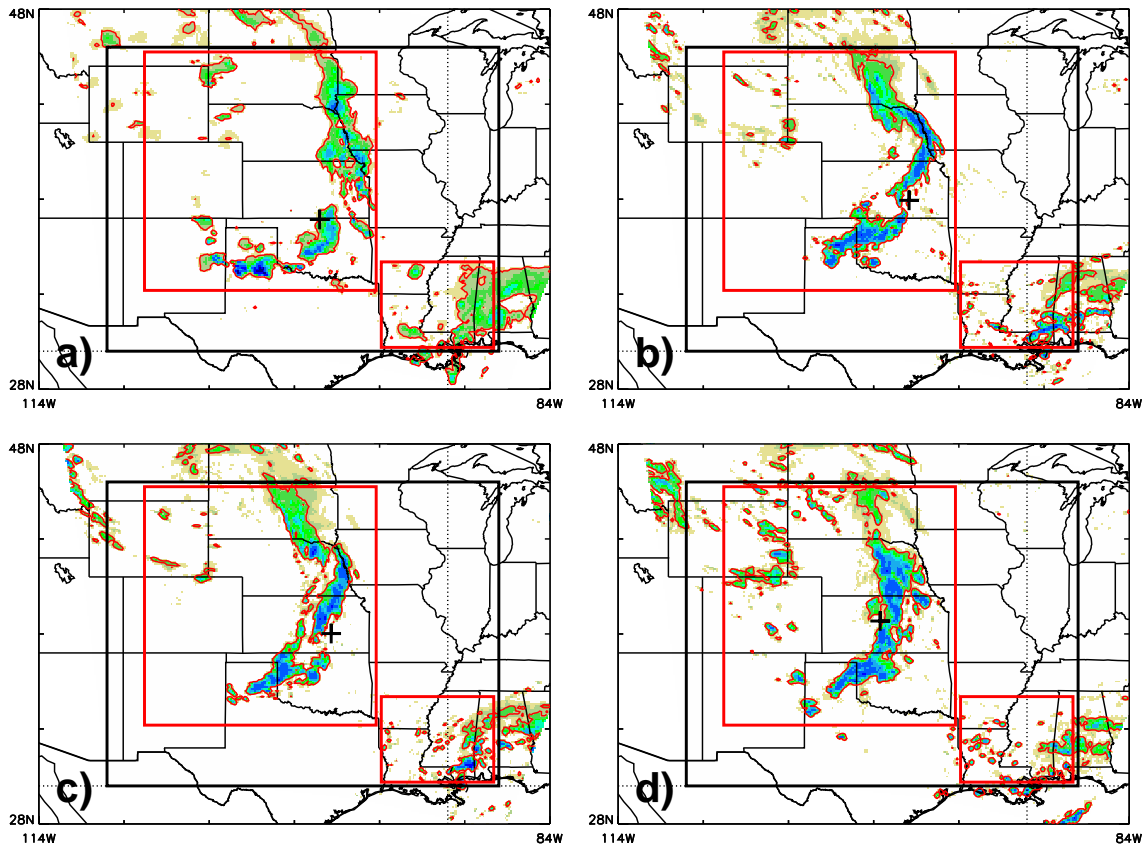


Abbildung 3.6. Niederschlagsverteilung am 1. Juni 2005: (a) Beobachtung und Modellvorhersagen von (b) 2CAPS, (c) 4NCAR und (d) 4NCEP. Die Rechtecke zeigen die drei Untersuchungsregionen für die SAL Analyse der großen (schwarze Linien) sowie nördlichen und südlichen (rote Linien) Region. Das schwarze Pluszeichen markiert den Schwerpunkt des Niederschlagsfeldes des großen Rechtecks.

Im Fall des 4NCEP ergibt sich ein relativ hoher Wert für $L = 0,29$ (im Vergleich mit den anderen Modellen), der durch L_2 begründet ist, dem zweiten Teil des Verschiebungsfehlers, der den gewichteten, mittleren Abstand der Objekte zum Schwerpunkt des gesamten Niederschlagsgebiets beschreibt. Da in dieser Situation für dieses Modell ein großes Objekt, anstatt mehrerer ungefähr gleich großer kleinerer Objekte identifiziert wird, ist der Abstand der Schwerpunkte gering und somit L_2 relativ groß.

Die viel zu kleinen Strukturen der QNV Objekte im südöstlichen Bereich werden durch die Werte von S im großen Gebiet nicht angemessen beurteilt, da die Objekte einen kleinen Beitrag zum Gesamtniederschlag liefern und somit einen geringen Einfluss auf den Wert von S haben (siehe Gl. 3.7 und 3.8), denn an diesem Tag dominieren die Beiträge der

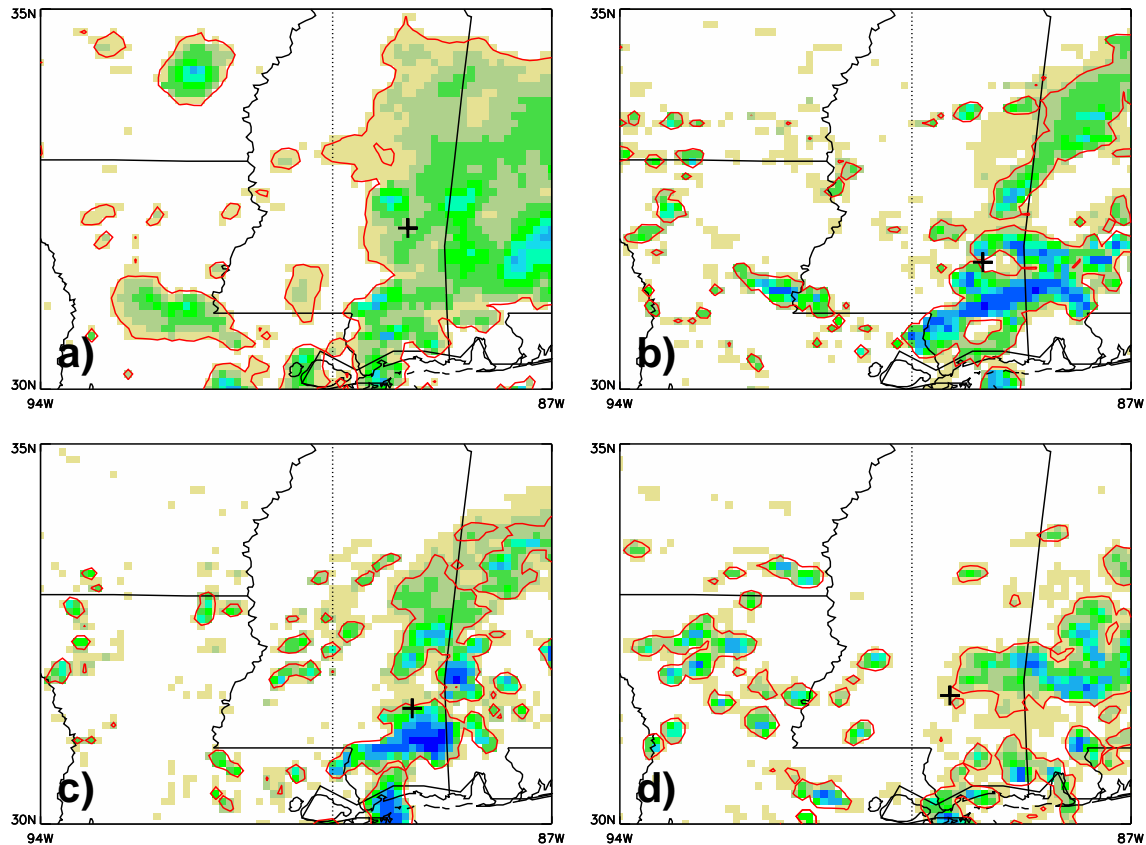


Abbildung 3.7. Niederschlagsverteilung am 1. Juni 2005 im südlichen Untersuchungsgebiet: (a) Beobachtung und Modellvorhersagen von (b) 2CAPS, (c) 4NCAR und (d) 4NCEP. Das schwarze Pluszeichen markiert den Schwerpunkt der Niederschlagsfelder.

Objekte des frontalen Niederschlagsbandes. Daher ist es sinnvoller, SAL auf angepasste Gebiete anzuwenden, um die meteorologisch unterschiedlichen Niederschlagssysteme im Zentrum und im Norden von denen im Süden an der Golfküste zu trennen. Wenn die SAL Analyse nur auf die nördliche Box angewandt wird (rote Rechtecke in den Grafiken in Abb. 3.6), sind die Werte ähnlich (nicht gezeigt) zu denen der großen Box (schwarzes Rechteck), da:

- (i) sich das Niederschlagsmaximum im nördlichen Teilgebiet befindet und deshalb der Schwellenwert zur Bestimmung der Niederschlagsobjekte bei der Betrachtung des nördlichen Teilgebietes unverändert bleibt und
- (ii) sich die stärksten Niederschlagsobjekte im nördlichen Teil befinden.

Der Hauptunterschied zwischen der großen und der nördlichen Box ist, dass die Werte von L größer werden bei der Betrachtung des kleineren Gebiets. Dies ist dadurch zu erklären, dass der Verschiebungsfehler in der Berechnung der L Komponente mit dem Größenparameter d (siehe Abschnitt 3.1) skaliert wird.

Bei der Verifikation der QNVs mit SAL für die südliche Box ergeben sich völlig andere Werte (siehe Tabelle 3.4). Die Abbildung 3.7 zeigt die identifizierten Objekte im Fall des südlichen Teilgebietes. Da die Niederschlagswerte in dieser Region geringer als im nördlichen Teilgebiet sind, ist auch der Schwellenwert R^* geringer, was die Objekte in Abb. 3.7 im Vergleich zu Abb. 3.6 etwas größer werden lässt. Das grundsätzliche Bild ist, dass die Vorhersagen zu viele kleine Niederschlagsobjekte simulieren. Dies wird durch stark negative Werte von S und positive Werte von A (außer 4NCEP wo S leicht negativ ist) beschrieben. Die deutlich unterschiedliche Bewertung der QNV Leistung bei Betrachtung von großen Untersuchungsregionen oder einer Teilregion zeigt, wie die Wahl der Verifikationsregion einen Einfluss auf die SAL Ergebnisse (und anderer klassischer Bewertungsmaße; siehe Tabelle 3.4) hat. Wenn das Untersuchungsgebiet so groß gewählt wird, dass es deutlich unterschiedliche meteorologische Systeme enthält, dann sind die Ergebnisse möglicherweise nicht repräsentativ für das schwächere Niederschlagsystem. Daher sollte im Allgemeinen die Gebietsgröße bei der Anwendung von SAL nicht größer als $500 \times 500 \text{ km}^2$ sein. Mehr zu diesem Punkt folgt in Abschnitt 3.2.5.

b) Gut und schlecht – bezogen auf klassische Bewertungsmaße und SAL

An dieser Stelle liegt der Fokus auf der unterschiedlichen Bewertung der QNV Leistung durch klassische Bewertungsmaße und SAL. Zu diesem Zweck wurden Vorhersagen ausgewählt, die kleine Werte von ME und $RMSE$ (Abb. 3.8) besitzen, sehr große $RMSE$ Werte (Abb. 3.9) und deutlich unterschiedliche Werte beim HSS (Abb. 3.10). Das abschließende Beispiel (Abb. 3.11) zeigt die Situation die bei der Bewertung durch SAL am besten abschneidet.

Entsprechend dem ME und $RMSE$ liefern die QNVs der Modelle 2CAPS und 4NCAR exzellente Vorhersagen in der südlichen Region (siehe Tabelle 3.4). Bei diesen QNVs ist der ME annähernd null und der $RMSE$ ist geringer als 1 mm. Ebenso sind die Werte für den FBI dieser QNVs von ansprechender Qualität, jedoch sind die Werte des HSS gering. Allerdings bietet Abbildung 3.9 einen anderen Eindruck. Obwohl beide Vorhersagen das Niederschlagsereignis am Südrand des Michigansees gut simulieren, weisen sie zu kleine (und ein wenig zu starke) Niederschlagsobjekte in Tennessee und südlicher auf. Dies wird mit deutlich negativen Werten von S ($-0,79$ für 2CAPS und $-0,92$ für 4NCAR) bewertet

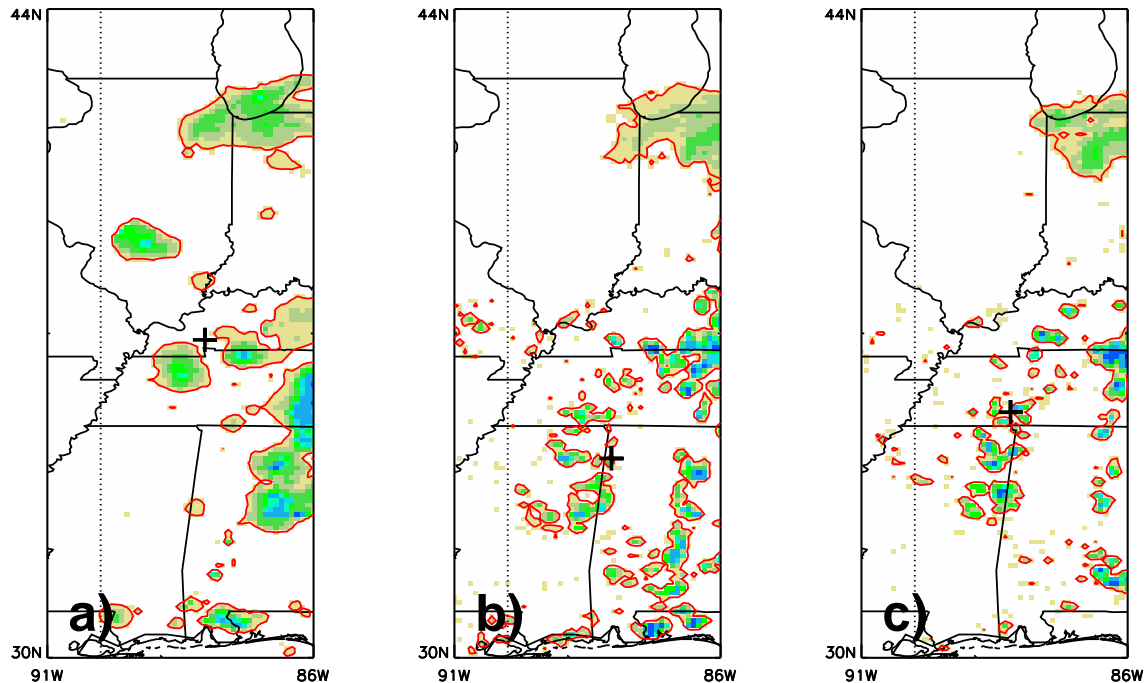


Abbildung 3.8. Niederschlagsverteilung am 3. Juni 2005 im südlichen Untersuchungsgebiet: (a) Beobachtung und Modellvorhersagen von (b) 2CAPS und (c) 4NCAR. Das schwarze Pluszeichen markiert den Schwerpunkt der Niederschlagsfelder.

und zeigt damit, dass die Vorhersagen die Struktur der beobachteten Niederschlagsobjekte nicht gut simulieren. Die relativ großen Werte für L (ungefähr 0,3 für beide QNVs) deuten daraufhin, dass die Vorhersagen weit entfernt vom perfekten Wert sind. Die beiden Werte für A sind annähernd null, was sich mit den Ergebnissen von ME deckt. Ein Grund für die kleinen Werte des $RMSE$ ist die größtenteils schwache Intensität dieses Ereignisses. Dieses spiegelt sich ebenso wieder in den mittelgroßen Werten des $nRMSE$ (0,12 im Vergleich zu $\leq 0,07$ für die QNVs am 1. Juni 2005). SAL bewertet den relativen Fehler von QNVs und kann damit auch Vorhersagemängel von schwachen Niederschlagsereignissen identifizieren.

In Abbildung 3.9 werden die beiden QNVs mit den größten Werten des $RMSEs$ dargestellt. In beiden Fällen, die am selben Tag allerdings in unterschiedlichen Regionen auftraten, erreicht der $RMSE$ einen Wert von ungefähr 4,3 mm. Bei der Betrachtung der anderen Bewertungsmaße sind deutliche Unterschiede in der QNV Leistung zu erkennen (siehe Tabelle 3.4). Außerdem zeigt die Abbildung einen deutlich unterschiedlichen Charakter der Niederschlagsfelder (und Fehler) für die beiden Fälle. Am 13. Mai 2005

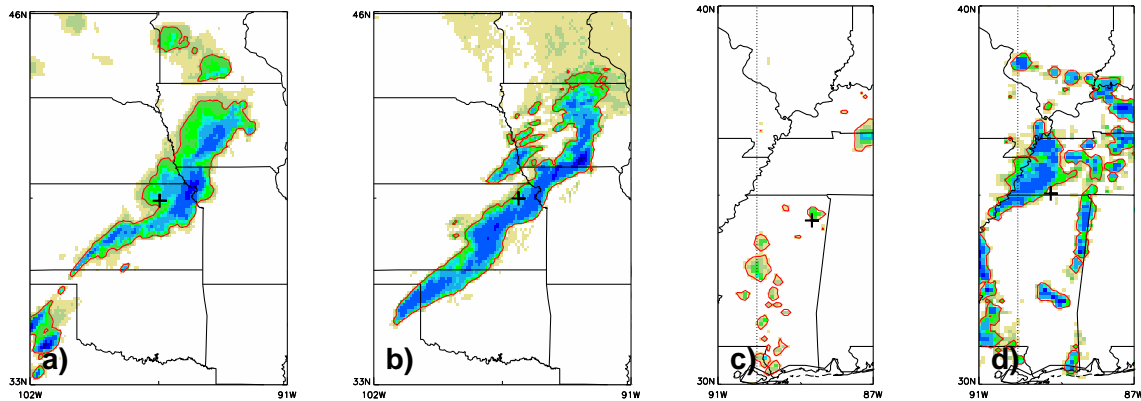


Abbildung 3.9. Niederschlagsverteilung am 13. Mai 2005: Beobachtung im nördlichen (a) und südlichen (c) Untersuchungsgebiet sowie Modellvorhersagen von (b) 2CAPS im nördlichen und (d) 4NCEP im südlichen Untersuchungsgebiet. Das schwarze Pluszeichen markiert den Schwerpunkt der Niederschlagsfelder.

existierte ein starkes Niederschlagsband im nördlichen Teilgebiet, das von Iowa bis nach Kansas reichte (Abb. 3.9a). Die betrachtete QNV (Abb. 3.9b) simulierte die Form und Stärke ziemlich gut, allerdings um 100 km nach Osten verschoben. Beim zweiten Beispiel wird ein schwaches Niederschlagsereignis in Mississippi (Abb. 3.9c) beobachtet, aber die Modellvorhersage produziert zu starken Niederschlag, der hauptsächlich weiter im Norden in Tennessee auftritt (Abb. 3.9d). Für praktische Anwendungen, d.h. für einen operationellen Prognostiker oder für hydrologische Anwendungen, liefert die erste QNV (Abb. 3.9b) nützliche Informationen (Auftreten eines starken Niederschlagsbandes), wohingegen die zweite Vorhersage (Abb. 3.9d) sehr weit entfernt von der Beobachtung ist. Aus diesem Grund ist es nicht angebracht, die Güte der beiden QNVs, bewertet durch den *RMSE*, als gleich anzusehen. Im ersten Beispiel tritt der hohe *RMSE* Wert durch das sogenannte 'Double Penalty'-Problem auf, währenddessen er im zweiten Beispiel durch eine starke Überschätzung der Niederschlagsstärke und -ausdehnung hervorgerufen wird. Die Hinweise durch die SAL Werte erscheinen zielführender. Im ersten Fall sind *S* und *A* geringfügig positiv (das Niederschlagsband ist etwas zu groß und zu stark in der Vorhersage) und mit $L = 0,24$ wird die Verschiebung des Bandes gegenüber der Beobachtung beschrieben. Dagegen sind im zweiten Beispiel die Werte von *S* und *A* deutlich positiv, was verdeutlicht, dass die QNV zu viel Niederschlag in einem viel zu großen Areal simuliert. Die geringe Güte dieser Vorhersage ist außerdem durch die großen Werte beim *ME*, *nRMSE* und *FBI* sowie einem *HSS* von ungefähr null belegt.

Als weiteres Beispiel werden zwei QNVs (Abb. 3.10) diskutiert, die erheblich unterschiedliche Werte beim HSS aufweisen. Wie im vorigen Beispiel ist der 13. Mai der betrachtete Tag für den Norden und den Süden. Im Fall des Niederschlagsbandes (Abb. 3.10a,b) erzielt das 4NCAR die höchsten Werte beim $HSS (= 0,46)$ aller in dieser Untersuchung betrachteten QNVs. Die Begründung hierfür ist die starke Überlappung der Strukturen von QNV und Beobachtung. Die SAL Werte belegen eine mittelmäßige Güte der Vorhersage und $L = 0,24$ ist hervorgerufen durch das Fehlen kleinerer Objekte im Norden und Südwesten des Untersuchungsareals. Der positive Wert von A weist auf eine geringfügige Überschätzung der Niederschlagsmenge hin, wobei der relativ hohe positive Wert von $S = 0,74$ auf den ersten Blick überrascht. Der Grund dafür ist das kleine und relativ starke Niederschlagsobjekt im Norden von Texas (Abb. 3.10a), welches in der Vorhersage fehlt. Das Maximum übersteigt einen Wert von 70 mm/h (während die Maximalwerte des beobachteten und simulierten Niederschlagsbandes weiter im Norden 40 - 50 mm/h betragen). Demzufolge ist das Objekt in Nordtexas extrem stark und führt zu einem kleinen „skalierten Volumen“ (siehe Gl. 3.7) und somit – unter Beachtung des Fehlens dieses starken Objektes in der Vorhersage – zu einem relativ deutlich positiven Wert der Struktur Komponente S .

Im Gegensatz dazu ist der HSS des QNVs des selben Modells im südlichen Untersuchungsgebiet (Abb. 3.10c,d) annähernd null und somit ohne Wert. Ebenso hat der $nRMSE$ einen ziemlich hohen Wert im Vergleich zu den anderen Situationen auf. Völlig anders als die QNV aus Abbildung 3.9d (ähnlich schlechter HSS Wert) scheint diese QNV bei einer subjektiven Verifikation nützlich zu sein, da die Vorhersage den grundlegenden Charakter des Ereignisses mit einigen moderaten Schauern in Mississippi und Tennessee simuliert (vergleiche Abb. 3.10c,d). Die Tatsache, dass der Ort vieler Objekte in Beobachtung und Vorhersage nicht übereinstimmen, ist die Ursache für den niedrigen HSS Wert. Aus den SAL Werten können folgende, hilfreiche Informationen gewonnen werden: S ist leicht negativ in diesem Fall, d.h. die Objekte zu „spitz“ (zu große Maximalwerte in den kleinen Objekten). Die Amplitude A ist deutlich positiv und zeigt damit eine Überschätzung des Gesamtniederschlags an. Auch in diesem Fall, ebenso wie in den Beispielen aus Abbildung 3.8, sind die Werte von ME und $RMSE$ klein, hauptsächlich dadurch bedingt, dass die Menge des Niederschlagsereignisses gering ist.

Abschließend sollen die beiden besten Vorhersagen im Sinne von SAL vorgestellt werden (Abb. 3.11). Die Auswahl erfolgte für diejenigen Vorhersagen des nördlichen und südlichen Untersuchungsgebiets mit den geringsten absoluten Werten aller SAL Komponenten, d.h. für QNVs bei denen $\max(|S|, |A|, L)$ am kleinsten ist⁶. Innerhalb des in dieser

⁶Diese Wahl der Definition einer skalaren Metrik für die drei Komponenten von SAL ist sinnvoll, da die hohe Güte von QNVs durch kleine Werte in *allen* SAL Komponenten wiedergegeben wird. Im Gegensatz

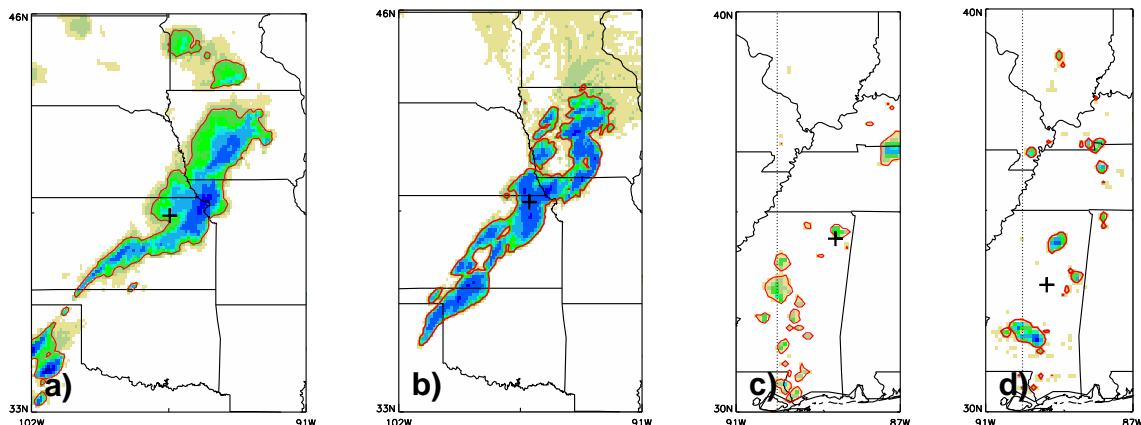


Abbildung 3.10. Niederschlagsverteilung am 13. Mai 2005: Beobachtung im nördlichen (a) und südlichen (c) Untersuchungsgebiet sowie Modellvorhersagen von 4NCAR im nördlichen (b) und südlichen (d) Untersuchungsgebiet. Das schwarze Pluszeichen markiert den Schwerpunkt der Niederschlagsfelder.

Studie genutzten Datensatzes weisen die beiden besten QNVs Werte von $\max(|S|, |A|, L)$ gleich 0,19 (2CAPS im nördlichen Teilgebiet am 22. April 2005; Abb. 3.11a,b) und von 0,22 (4NCAR in nördlichen Teilgebiet am 1. Juni 2005; Abb. 3.11c,d) auf. Diese Werte sind nicht sonderlich klein – bei Hofmann *et al.* (2009) werden einige Beispiel guter QNVs in Deutschland mit Werten von $\max(|S|, |A|, L)$ kleiner als 0,1 angeführt – was darauf hindeutet, dass selbst die besten hier betrachteten QNVs beachtliche Defizite aufweisen. Abbildung 3.11b zeigt, dass die Vorhersage das Niederschlagsband gut in seiner Ausdehnung von der nördlichen und östlichen Grenze Nebraskas simuliert (siehe Abb. 3.11a), aber mit einer merklichen Verschiebung nach Westen. Der Fall der zweitbesten QNV im Sinne von SAL wird in Abbildung 3.11c,d dargestellt. Die Vorhersage umfasst ein ungleichmäßiges Niederschlagsband, welches von Nordtexas bis nach Süd-Dakota reicht, allerdings mit einer deutlichen Verschiebung der Niederschlagsobjekte. Interessant ist die Analyse beider Fälle mit klassischen Bewertungsmaßen (siehe Tabelle 3.4), die ein sehr unterschiedliches Bild aufweisen. In beiden Fällen liegt der *FBI* nah am perfekten Wert von 1, *ME* und *nRMSE* bei sehr kleinen Werten und der *RMSE* bei mittleren Werten. Der Wert des *HSS* ist im ersten Fall relativ gut (0,41) und deutlich geringer im zweiten Fall (0,18). Ein möglicher Grund für den geringeren Wert beim *HSS* im zweiten Fall ist die kleinskalige Verschiebung der Niederschlagsobjekte, was zu zu einer geringeren

dazu ein anderes skalares Maß wie $(S^2 + A^2 + L^2)^{0,5}$, dass möglicherweise eine Kompensation zuließe, z.B. einen größeren Wert von *L* bei kleinen Werten von *S* und *A*.

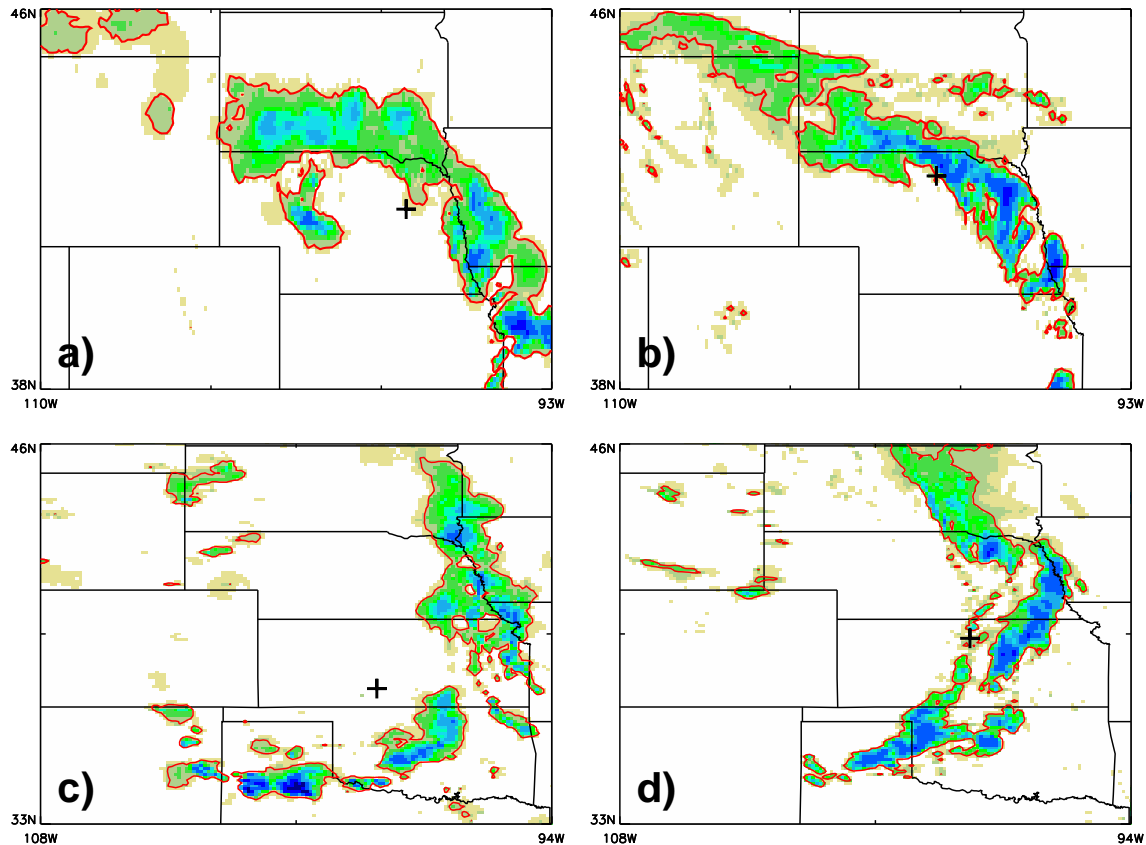


Abbildung 3.11. Niederschlagsverteilung im nördlichen Untersuchungsgebiet am 22. April 2005 für (a) Beobachtung und (b) Modellvorhersage von 2CAPS sowie am 1. Juni 2005 für (b) Beobachtung und (c) Modellvorhersage von 4NCAR. Das schwarze Pluszeichen markiert den Schwerpunkt der Niederschlagsfelder.

Überlappung von simulierter und beobachteter Niederschlagsverteilung führt. Diese beiden letzten Beispiele zeigen deutlich, dass kleine, d.h. gute, SAL Werte einer QNV nicht notwendigerweise mit sehr guten Werten der klassischen Maße übereinstimmen müssen.

3.2.5 Schlussfolgerungen

Die QNV Fälle in den Vereinigten Staaten von Amerika, die für ein Vergleichsprojekt von räumlichen Verifikationsmethoden ausgewählt wurden, liefern einen nützlichen zusätzlichen Einblick in das Verhalten von SAL, besonders beim Vergleich mit klassischen

Bewertungsmaßen. Die wichtigsten Resultate lassen sich folgendermaßen zusammenfassen:

- Die Analyse der „geom“-Fälle untermauert die Ergebnisse von *Wernli et al. (2008)* dahingehend, dass SAL Fehler im Ort und Größe von klar definierten einzelnen Niederschlagsobjekten sehr gut aufzeigt. Allerdings ist SAL nicht in der Lage, die Orientierung von Niederschlagsobjekten zu quantifizieren.
- Vergleichbare Schlussfolgerungen können für die Resultate der „fake“-Fälle gezogen werden. Zusätzlich wurde herausgefunden, dass eine Reduzierung eines Niederschlagsfelds um einen konstanten Faktor dazuführen kann, dass die identifizierten Niederschlagsobjekte kleiner und spitzer sind, was durch einen deutlich negativen Wert der Strukturkomponente S belegt wird. Dies zeigt auf, dass SAL (besonders S) die Fähigkeit besitzt, zwischen Vorhersagen zu unterscheiden die eher durch starke lokale Niederschlagsereignisse geprägt sind und Situationen die in ein großräumiges schwaches Niederschlagsgebiet eingebettet sind.
- Die QNVs des WRF Modells für die Vereinigten Staaten zeigen manchmal ein gegensätzliches Verhalten von SAL im Vergleich zu den Ergebnissen klassischer Bewertungsmaße. Wie erwartet sind die absoluten Maße wie $RMSE$ und ME typischerweise gut bei schwachen Ereignissen und im Gegensatz dazu die relativen Parameter von SAL häufig besser bei starken Ereignissen.
- Sehr gute QNVs im Sinne von SAL sind nicht notwendigerweise mit guten Werten der klassischen Maße verbunden. Besonders der HSS zeigt eine schwache QNV Leistung für Fälle bei denen SAL eine einigermaßen gute Bewertung abgibt an.
- Allgemein zeigt die detaillierte Analyse von idealisierten und realen QNVs im Rahmen des internationalen Vergleichsprojektes, dass SAL nützliche und hilfreiche Hinweise über die Qualität und spezielle Defizite einzelner Vorhersagen liefern kann.

Außerdem konnte herausgearbeitet werden, dass die Anwendung von SAL auf Untersuchungsgebiete mit horizontalen Ausmaßen von $2000 \text{ km} \times 1600 \text{ km}$ (siehe Kapitel 3.2.4a) zu Ergebnissen führen kann, die schwer zu interpretieren sind. Besonders problematisch wird dies, wenn sich innerhalb des Untersuchungsareals mehrere meteorologisch unterschiedliche Niederschlagssysteme befinden, z.B. ein frontales Regenband und eine Region mit Luftmassenkonvektion. In solch einer Situation können die QNV Fehler für verschiedene Niederschlagssysteme unterschiedlich sein und die SAL Resultate beziehen

sich auf ein Mittel, welches nicht sinnvoll ist. Dabei soll klargestellt werden, dass dies kein spezielles Problem der SAL Methode ist, sondern jedes verwendete Bewertungsmaß kann dieses Problem aufweisen, wenn das Verifikationsgebiet sehr viel größer ist als die Skala des Niederschlagssystems. Zum Beispiel kann eine Unterschätzung konvektiver Aktivität in einem Teil des Untersuchungsgebiets möglicherweise durch eine Überschätzung von stratiformen Regen in einem anderen Teil des Untersuchungsgebiets kompensiert werden und zu einem annähernd perfekten Wert im Sinne des MEs führen. Im operationellen Verifikationsbetrieb ist es sehr schwer automatisch zu bestimmen, was eine sinnvolle Größe des Gebiets ist und jede Interpretation von Verifikationsergebnissen sollte unter Beachtung eines möglicherweise zu großen Verifikationsgebietes genau geprüft werden.

Abschließend soll noch angemerkt werden, dass an dieser Stelle im Gegensatz zu der Publikation von *Wernli et al.* (2008) SAL auf stündlich akkumulierte QNVs angewandt wurde. Eine der größten Herausforderungen heutiger QNV Verifikation ist die Bewertung von sehr kurzen Akkumulationsperioden, in denen Niederschlagsfelder häufig durch komplexe Strukturen (z.B. mehr lokale und kleinere Strukturen im Vergleich zu täglichen Akkumulationen) charakterisiert sind. Die Anwendung von SAL und anderer kürzlich entwickelter QNV Qualitätsmaße für größere Datensätze und kurze Akkumulationszeiten stellt einen wichtigen Punkt dar, der im nächsten Abschnitt gezielt untersucht wird.

3.3 Quantifizierung zeitlicher Fehler*

Die hohe räumliche aber auch zeitliche Variabilität von Niederschlagsverteilungen führt zu Problemen bei der Verifikation von QNVs. Sehr kurze Akkumulationszeiträume (eine Stunde oder weniger) sind durch räumliche wie auch zeitliche Fehler (Zuggeschwindigkeit von Fronten, Einsetzen von Konvektion, etc.) charakterisiert. Dabei kann es vorkommen, dass entweder die Beobachtung oder die Vorhersage im betrachteten Untersuchungsgebiet keinen Niederschlag enthält und somit keine Verifikation möglich ist. Das Ziel dieses Kapitels ist die Entwicklung und Beschreibung einer Methode um im Fall leichter zeitlicher Verschiebungen mit SAL dennoch eine aussagekräftige Verifikation durchführen zu können sowie eine quantitative Information über zeitliche Fehler anzugeben. Zuerst wird die Motivation (Abschnitt 3.3.1) zur Entwicklung eines Verfahrens zur Erfassung zeitlicher Fehler bei QNVs mit SAL vorgestellt. Anschließend wird die Methode (Abschnitt 3.3.2) detailliert beschrieben und in Abschnitt 3.3.3 erfolgt die Anwendung der Bestimmung zeitlicher Fehler am Beispiel des COSMO-DE.

*Die Resultate dieses Abschnittes wurden zur Veröffentlichung bei der *Meteorologischen Zeitschrift* eingereicht: Zimmer, M. and H. Wernli (2010). Verification of quantitative precipitation forecasts on short time-scales: A fuzzy approach to handle timing errors.

3.3.1 Motivation für die zeitliche Komponente

Die Entwicklung neuartiger räumlicher Bewertungsmaße (Kapitel 1.2.2) wie SAL zielte darauf ab, das 'Double Penalty'-Problem, welches durch räumlichen Versatz der Vorhersage gegenüber der Beobachtung hervorgerufen wird, zu vermeiden. Ein zeitlicher Versatz kann ebenfalls zu einem 'Double Penalty' führen. Zum Beispiel wenn in einer Stunde kein Niederschlag beobachtet wird, allerdings die entsprechende Vorhersage welchen simuliert. Eine Stunde später aber eine Niederschlagsbeobachtung erfolgt ohne das die Vorhersage Niederschlag enthält. In diesem Fall werden beide stündlichen Vorhersagen überhaupt nicht durch SAL bewertet. Eine Information über den zeitlichen Versatz sowie die Bewertung dieser QNV wäre dementsprechend nützlich. Zwei der häufigsten zeitlichen Fehlercharakteristika bei beobachteten Niederschlagsereignissen sind in Abbildung 3.12 dargestellt. Als erstes Beispiel (Abb. 3.12a) erfolgt die schematische Betrachtung eines frontalen Regenbandes. Beim Vergleich der Vorhersagen ist zu erkennen, dass eine subjektive Bewertung der Situation dazu führt, dass nicht der Zeitpunkt der Beobachtung als der mit der besten Übereinstimmung abschneidet, sondern die Vorhersage eine Stunde später. Das zweite idealisierte Beispiel (Abb. 3.12b) zeigt das Einsetzen von lokaler Konvektion, welche in der Vorhersage zeitlich versetzt ist. Zum Zeitpunkt der Beobachtung ist in der Simulation kein Niederschlag vorhanden. Allerdings ist in der Stunde davor Niederschlag simuliert worden. Eine Erfassung der QNV Qualität einer solchen Situation ist wünschenswert. Im Beispiel sind die Beobachtung und die Simulation zwei Stunden vorher, bei einer subjektiven Bewertung, am ähnlichsten. Ziel ist es, diese subjektive Bewertung mit einem objektiven Verfahren zu beschreiben.

Einen Vorschlag zur Einbeziehung zeitlicher Fehler in die merkmalsbasierte QNV Verifikation stellten *Brown et al.* (2009) beim Verifikationsworkshop in Helsinki im Juni 2009 vor. Dabei wird die Verifikationstechnik MODE (siehe Kapitel 1.2.2) verwandt und es wird davon ausgegangen, dass Niederschlagssysteme in Raum und Zeit zusammenhängend sind. Es werden dreidimensionale Niederschlagsobjekte sowohl in der Beobachtung als auch in der Vorhersage innerhalb eines dreistündigen Zeitfensters ($\pm 1h$) bestimmt und bewertet. Eine direkte Information zum zeitlichen Fehler wird nicht angegeben, allerdings kann die grundsätzliche Fähigkeit der Simulation von Niederschlag des untersuchten Vorhersagesystems besser bewertet werden.

Der für SAL gewählte Ansatz unterscheidet sich von der für MODE gewählten Vorgehensweise, denn die L Komponente enthält indirekt Informationen über einen zeitlichen Fehler von Niederschlagsvorhersagen. Wird zum Beispiel die Zuggeschwindigkeit eines frontalen Regenbandes falsch simuliert (siehe Abb. 3.12a), dann deutet zum Zeitpunkt $\Delta t = 0$ der Wert von L an, dass sich die Front am falschen Ort befindet. Ob also ein vorher-

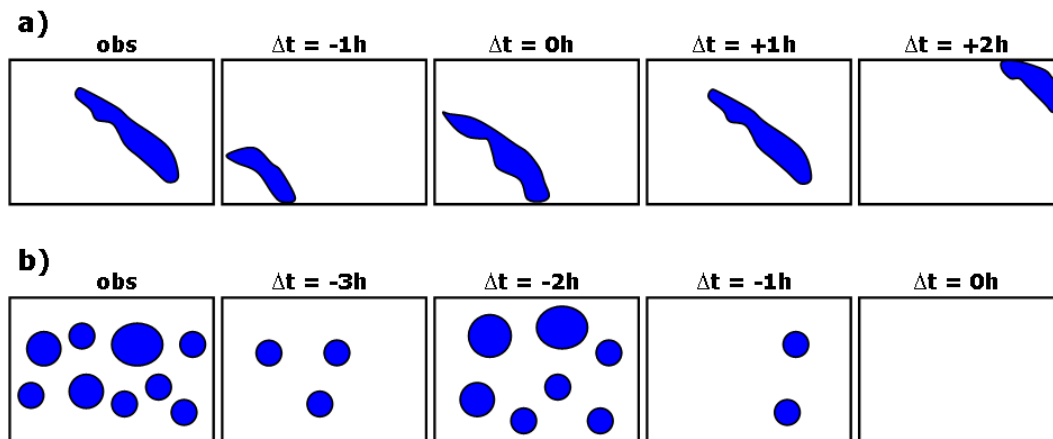


Abbildung 3.12. Schematische Darstellung zeitlicher Fehler bei frontalem Niederschlag a) sowie bei Luftmassenkonvektion b) anhand von Beobachtungen (obs) und zeitversetzter QNVs (Δt).

gesagtes Niederschlagsobjekt am falschen Ort entstanden ist oder ob es zu einem anderen Zeitpunkt am Beobachtungsort auftritt, kann die L Komponente nicht unterscheiden.

3.3.2 Methodische Vorgehensweise

Da die L Komponente von SAL indirekt Informationen über den zeitlichen Fehler von Niederschlagsvorhersagen enthält, ist sie der Ausgangspunkt zur Bestimmung dieser Fehlercharakteristik. Die Idee dabei ist, dass innerhalb eines vorher definierten Zeitfensters um einen Beobachtungszeitpunkt eine Minimierung von L durchgeführt wird. Das Prinzip wird anhand von stündlichen Akkumulationen vorgestellt, welches allerdings keine Beschränkung der Methode darstellt, so dass eine Anwendung auf kürzere Zeiträume ebenfalls möglich wäre.

Die Abbildung 3.13 zeigt eine schematische Darstellung der Vorgehensweise zur Berechnung von SAL. In Abbildung 3.13a ist das Verfahren gezeigt, wie die Berechnung von SAL ohne Betrachtung eines zeitlichen Versatzes erfolgt. Dabei wird die Beobachtung mit der entsprechenden Vorhersage direkt miteinander verifiziert. In Abbildung 3.13b ist die Vorgehensweise zur Bestimmung von SAL bei Einbeziehung einer möglichen zeitlichen Verschiebung dargestellt. Alle Vorhersagen innerhalb eines Zeitfensters, im vorgestellten Fall $\pm 3h$, um den Beobachtungszeitpunkt werden mit SAL verifiziert, d.h. es stehen sieben Vorhersagen für die Bewertung zur Verfügung. Zunächst erfolgt die Verifikation der Vorhersage mit Gültigkeit des Beobachtungstermins. Anschließend wird der

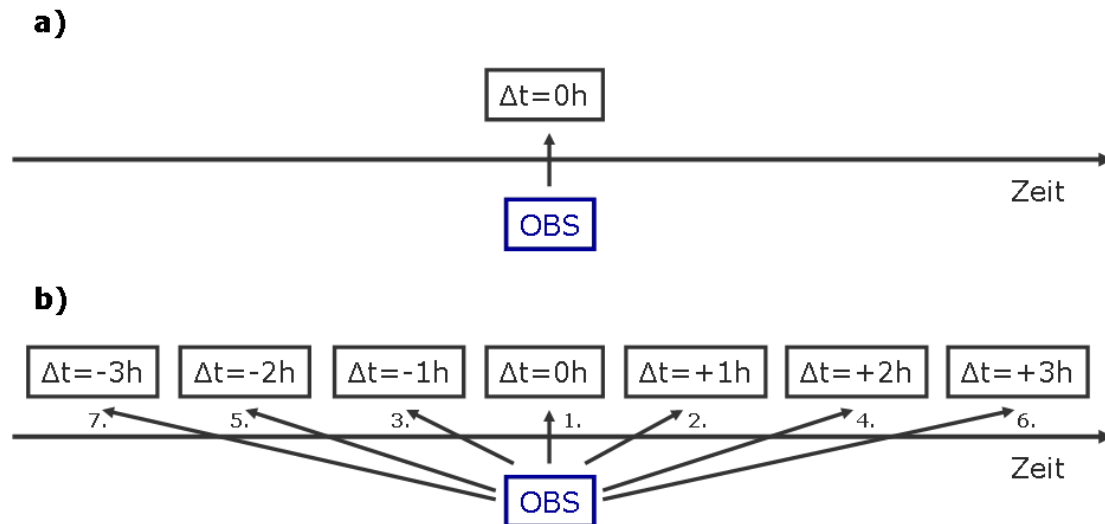


Abbildung 3.13. Schematische Darstellung der Methode zur Berechnung von SAL: ohne (a) und mit (b) Betrachtung eines zeitlichen Versatzes von QNVs.

Zeitpunkt mit $\Delta t = +1h$ verifiziert und der Wert von L mit dem der Vorhersage des Beobachtungstermins verglichen. Für den Zeitpunkt mit dem geringeren L sind die Werte von SAL und der dazugehörige Wert von Δt festzuhalten. Wie Abbildung 3.13b illustriert, erfolgt anschließend der Reihe nach, von Schritt drei bis sieben, die Berechnung von SAL. Bei jedem Schritt wird überprüft, ob der neue Wert von L geringer ist als der bisher kleinste. Am Ende der Auswertung der sieben Zeitpunkte ist derjenige Zeitpunkt mit dem kleinsten Wert für L bestimmt sowie die dazugehörigen Werte von S und A . Eine Ausnahme bilden die Zeitpunkte von Schritt sechs und sieben. Hier wird bei einem geringeren Wert von L nur dann SAL und die Zeitverschiebung Δt in Betracht gezogen, wenn $|S|$ und $|A|$ kleiner 1,5 sind. Damit ist sichergestellt, dass keine Niederschlagsereignisse miteinander verglichen werden, die meteorologisch stark unterschiedlich und deshalb vermutlich voneinander unabhängig sind.

3.3.3 Anwendungsbeispiel

Die im vorigen Abschnitt vorstellte Methode zur Quantifizierung von zeitlichen Fehlern von hochaufgelösten QNVs mit SAL wird an dieser Stelle mit Hilfe von stündlichen Akkumulationen des COSMO-DE veranschaulicht. Zunächst erfolgt die Betrachtung eines realen Beispiels und daran anschließend die Analyse eines Beobachtungszeitpunktes für

den Sommer 2007. Abschließend wird der tageszeitliche Verlaufs von S , A und Δt beschrieben.

a) Reales Beispiel

In Abbildung 3.14 wird ein Beispiel der Vorgehensweise zur Bestimmung der optimalen Zeitverschiebung Δt und den entsprechenden SAL Werten dargestellt. Es werden stündlich akkumulierte Niederschlagssummen im deutschen Abschnitt des Elbeeinzugsgebiets für den Beobachtungszeitpunkt 14 bis 15 UTC am 3. Juli 2007 betrachtet. In der Beobachtung (Abb. 3.14a) ist ein Niederschlagsband im westlichen Bereich des Einzugsgebiets zu erkennen, in die kräftige Schauer eingelagert sind. Der Standardansatz (Vergleich von Beobachtung und Vorhersage gültig zur selben Zeit) liefert Werte von $S = -0,85$, $A = -0,67$, und $L = 0,26$, die andeuten, dass die Vorhersage die Niederschlagsmenge und die räumliche Ausdehnung drastisch unterschätzt (Abb. 3.14e). Außerdem wird der meiste Niederschlag im Osten des Gebiets simuliert, wo keiner beobachtet wird, was zu einem moderaten Wert von L führt. Die Betrachtung der früheren (Abb. 3.14b-d) und der späteren (Abb. 3.14f-h) Stunden zeigt, dass die früheren Vorhersagen durch zu viel Niederschlag im östlichen Bereich (hohe Werte von L) sowie zu wenig im Westen des Einzugsgebiets gekennzeichnet ist. Die späteren Vorhersagen zeigen den Niederschlag im westlichen Bereich mit eingelagerten Schauern. Die Berechnung der SAL Komponenten liefert den kleinsten Wert von $L = 0,03$ für $\Delta t = +2$ h. Die zu dieser Vorhersage gehörenden Werte von $S = -0,57$ und $A = -0,78$ zeigen eine Verbesserung in der Struktur des Niederschlagsfelds im Vergleich zur Simulation ohne Zeitversatz. Alles in allem zeigt dieses Beispiel, dass die Minimierung der L Komponente ein objektives Maß zur sinnvollen Bestimmung eines Zeitversatzes von einigen Stunden zwischen Beobachtung und Vorhersagen ist.

b) Exemplarischer Zeitpunkt

Zur Veranschaulichung der Methode zur Quantifizierung zeitlicher Fehler erfolgt die Anwendung der Methode auf stündliche Akkumulationen des COSMO-DE. An dieser Stelle wird gezeigt wie sich die Einbeziehung zeitlich verschobener QNVs auf S und A beim Zeitpunkt 15 UTC im deutschen Abschnitt des Elbeeinzugsgebietes auswirkt. Die Resultate des Sommers 2007 der Verifikation mit SAL sind in Abbildung 3.15 mit Hilfe von SAL Diagrammen (Kapitel 3.1.2) dargestellt. Zum einen das SAL Diagramm (Abb. 3.15a) der Ergebnisse ohne und zum anderen das SAL Diagramm mit Betrachtung eines zeitlichen Versatzes (Abb. 3.15b).

Das Ergebnis für das COSMO-DE ohne Betrachtung zeitlicher verschobener QNVs liefert negative Medianwerte für $S = -0,3$ und $A = -0,6$, d.h. dieses Modell unterschätzt im Mittel zu diesem Zeitpunkt die Niederschlagsmenge bei gleichzeitig zu kleinen Nieder-

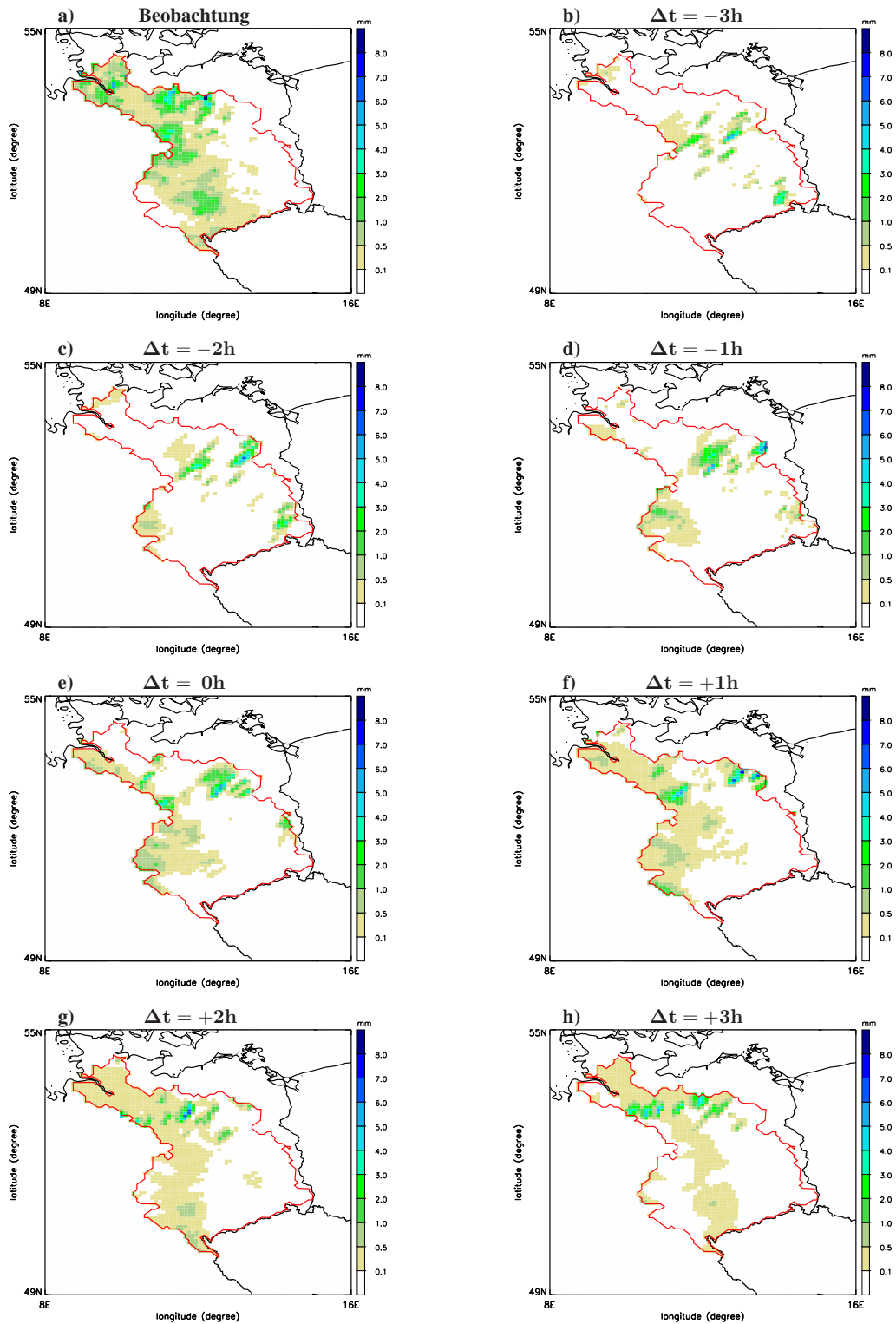


Abbildung 3.14. Stündlich akkumulierte Niederschlagsbeobachtung (a) im deutschen Abschnitt des Elbeeinzugsgebiets von 14 bis 15 UTC am 3. Juli 2007, sowie entsprechende QNVs des COSMO-DE mit einem Zeitversatz Δt zwischen -3 und $+3$ Stunden.

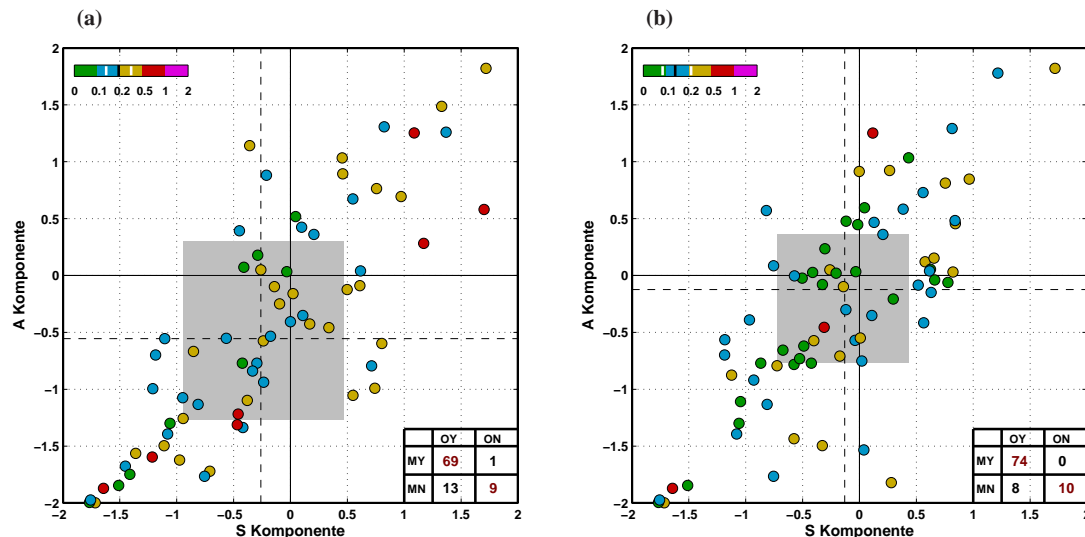


Abbildung 3.15. SAL Diagramme des COSMO-DE ohne (a) und mit (b) Betrachtung der Zeitkomponente für 15 UTC im deutschen Bereich des Elbeeinzugsgebietes für JJA 2007.

schlagsobjekten. Als Maß für die Streuung der Punkte wird der Interquartilsabstand (IQR) betrachtet und der beträgt für $S = 1,5$ und für $A = 1,6$. Der Median und der IQR von L liegt bei $0,2$. Die Betrachtung der Kontingenztabelle ergibt, dass von den 92 Tagen 69 mit SAL verifiziert werden konnten und es wurde an neun Tagen richtigerweise kein Niederschlag vorhergesagt. Dagegen wurde an 13 Tagen kein Niederschlag simuliert obwohl welcher beobachtet wurde und an einem Tag erfolgte die Simulation von Niederschlag obwohl keiner beobachtet wurde.

Im Fall der Einbeziehung eines zeitlichen Versatzes ist das Resultat ein wenig anders. Auch in diesem Fall sind die Medianwerte von S und A negativ, allerdings mit $-0,1$ erheblich näher am perfekten Wert. Die Werte des IQR sind mit $1,1$ für S und $1,2$ für A ebenfalls deutlich kleiner. Die Tatsache, dass sowohl die IQR- als auch die Medianwerte von S und A geringer sind, zeigt, dass das COSMO-DE in der Lage ist, die Struktur und Menge der Niederschlagsverteilung zu simulieren jedoch offensichtlich in einigen Fällen mit geringen zeitlichen Verschiebungen. Dass der Medianwert von L geringer ($0,14$ anstatt $0,2$) ist, liegt an der verwendeten Methode, die auf der Optimierung dieser Komponente beruht und ist somit nicht verwunderlich. Ein weiterer wichtiger Aspekt den die Einbeziehung eines möglichen zeitlichen Versatzes mit sich bringt, ist das dadurch mehr Situationen verifiziert werden können. In der Kontingenztabelle ist abzulesen, dass 74 Situationen verifiziert werden konnten, was ist einem Zuwachs von fünf Situationen gegenüber der Analyse ohne Verwendung der zeitlich verschobenen QNVs entspricht.

ΔT , [h]	-3	-2	-1	0	+1	+2	+3
# Anzahl	7	10	11	18	9	11	9

Tabelle 3.5. Absolute Häufigkeit von Δt für 15 UTC des COSMO-DE im deutschen Abschnitt des Elbeeinzugsgebietes für JJA 2007.

Nachdem die Auswirkungen auf S , A und L bei Betrachtung eines zeitlichen Versatzes von QNVs vorgestellt wurden, folgt jetzt die Beschreibung der Quantifizierung des zeitlichen Fehlers. In Tabelle 3.5 sind die Häufigkeiten von Δt , dem Zeitversatz der Modellvorhersagen für den optimalen Wert von L , für das gewählte Zeitfenster von $\pm 3h$ aufgelistet. Die Tabelle enthält die zeitlichen Verschiebungen Δt der 74 mit SAL verifizierten Situationen sowie einen Fall bei dem das Modell zum Beobachtungszeitpunkt Niederschlag simuliert obwohl keiner beobachtet wurde. Ein positiver Wert von Δt bedeutet, dass das Modell erst nach dem Beobachtungstermin den besten Wert von L aufweist. Analog dazu ist ein negativer Wert mit einem zu frühen Auftreten verbunden. Am häufigsten werden mit 18 Ereignissen Situationen ohne Zeitversatz festgestellt. Der Anteil von positiven und negativen Werten ist für diesen Zeitpunkt annähernd gleich, wobei die positiven Werten für Δt mit 29 ein Ereignis mehr aufweisen als die negativen Werte. Das COSMO-DE simuliert für den betrachteten Beobachtungszeitpunkt den Niederschlag im Mittel weder zu früh noch zu spät.

Damit konnte die grundsätzliche Anwendbarkeit der Methode zur Quantifizierung zeitlicher Fehler für stündliche Akkumulationen für ein Zeitfenster von $\pm 3h$ nachgewiesen werden.

c) Tageszeitlicher Verlauf

An dieser Stelle erfolgt die Untersuchung der zeitlichen Entwicklung der Güte mit fortschreitender Vorhersagezeit für das COSMO-DE im deutschen Abschnitt des Elbeeinzugsgebietes im Sommer 2007. Die Resultate der stündlichen Akkumulationen sind in Abbildung 3.16 in (a) für S und (b) für A dargestellt. Auf der horizontalen Achse ist die Tageszeit in UTC angegeben, die identisch mit der Vorhersagezeit ist, da die 00 UTC Läufe verwendet wurden. Die vertikale Achse gibt den Median an und der IQR ist farblich eingezeichnet mit grün für kleine und rot für große Werte. Der analysierte Tageszeitraum (4-18 UTC) wird durch die Länge der Vorhersageläufe (21h) und dem gewählten Zeitfenster ($\pm 3h$) festgelegt, so dass für jeden Beobachtungszeitpunkt innerhalb des Zeitfensters alle Vorhersagen vorhanden sind.

Die Resultate der S Komponente zeigen, dass der Median ohne Betrachtung von zeitlich verschobenen QNVs bis 10 UTC besser und für beide Methoden positiv ist. Danach ist der Wert des Medianes ohne Zeitbetrachtung erheblich negativ und bei Betrachtung der

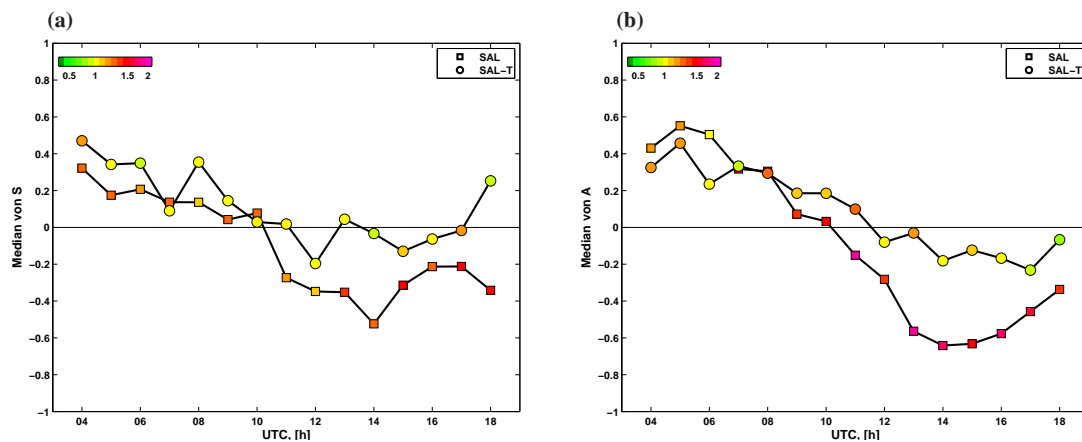


Abbildung 3.16. Darstellung der zeitlichen Entwicklung des Medianes von S (a) und A (b) mit (o) und ohne (□) Betrachtung des zeitlichen Versatzes beim COSMO-DE für den deutschen Bereich des Elbeinzugsgebietes für JJA 2007. Der Interquartilsabstand ist farblich gekennzeichnet mit grün für kleine und rot für große Werte.

Zeitkomponente ist dieser um die null verteilt außer bei 18 UTC wo er deutlich positiv ist. Im Gegensatz zum Median, der zu Beginn ohne Zeitbetrachtung besser ist, ist der IQR für alle betrachteten Zeitpunkte besser bei Einbezug einer möglichen Zeitverschiebung. Im Fall der A Komponente (Abb. 3.16b) sind die Resultate mit denen für S vergleichbar. Bis ungefähr 10 UTC sind die Mediane für beide Methoden vergleichbar und im positiven Bereich. Mit fortschreitender Vorhersagezeit wird der Medianwert negativ, wobei im Fall mit Betrachtung einer möglichen zeitlichen Verschiebung die mittlere Unterschätzung der Niederschlagsmenge geringer ist. Analog zu S ist der IQR bei Einbeziehung der möglichen zeitlichen Verschiebung geringer. Für beide Komponenten kann festgehalten werden, dass vor allem in den Nachmittagsstunden, dem Zeitraum des meisten Niederschlags, die Einbeziehung des Zeitversatzes zu besseren Resultaten der räumlichen Verteilung führt. Damit kann das Potenzial des Modells der Simulation der Niederschlagsverteilung bei optimalem Zeitversatz aufgezeigt werden.

Nach der Beschreibung der Fähigkeit des Modells, die räumlichen Eigenschaften bei optimalem Zeitversatz zu simulieren, erfolgt jetzt die Bewertung der zeitlichen Fehler. Die zeitliche Entwicklung während des Tages von Δt ist in Abbildung 3.17 gezeigt. Die horizontale Achse gibt die Tageszeit in UTC an und auf der vertikalen Achse ist der mittlere zeitliche Fehler aufgetragen. Der mittlere zeitliche Fehler wurde für jeden Zeitpunkt bestimmt und mit Hilfe des 'Bootstrap'-Verfahrens (Efron and Tibshirani, 1986; 1000 Wiederholungen) wurde ein 90 % Konfidenzintervall geschätzt. Der mittlere zeitliche Fehler ist bis 13 UTC deutlich positiv und wird anschließend mit fortlaufender Vorhersagezeit

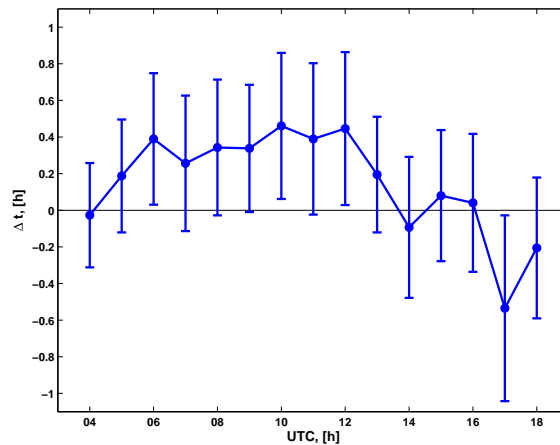


Abbildung 3.17. Darstellung der zeitlichen Entwicklung des mittleren zeitlichen Fehlers Δt für das COSMO-DE im deutschen Bereich des Elbeeinzugsgebietes für JJA 2007. Die Fehlerbalken geben das 90 % Konfidenzintervall der Mittelwertes an, das mit Hilfe des 'Bootstrap'-Verfahrens bestimmt wurde.

negativ. In den Vormittagsstunden in denen das COSMO-DE im Mittel das beste L nach dem Beobachtungstermin hat, ist mit einer Überschätzung der Niederschlagsmenge und zu großflächigen Niederschlagsobjekten (siehe Abb. 3.16) verbunden. Der Zeitraum des Tages (14-18 UTC), in dem im Mittel für den Sommer 2007 im Elbeeinzugsgebiet der meiste Niederschlag beobachtet wird, zeigt zu Beginn Werte von Δt von annähernd null im zeitlichen Fehler und wird dann negativ. Die Konfidenzintervalle der einzelnen Zeitpunkte überschneiden sich meist, so dass kein systematischer Unterschied herauszufinden ist. Somit liefert der Mittelwert von Δt eine Hinweis darauf, dass vor dem Tagesmaximum das COSMO-DE zu spät und danach zu früh den optimalen Wert von L simuliert.

Fazit des Abschnitts

Die Entwicklung und Anwendbarkeit einer Methode zur Quantifizierung zeitlicher Fehler hochaufgelöster QNVs mit SAL wurde in diesem Abschnitt am Beispiel stündlicher Akkumulationen vorgestellt. Die Optimierung der L Komponente in einem Zeitfenster von $\pm 3h$ um den Beobachtungstermin liefert nützliche Informationen zur Qualität der Niederschlagsvorhersage. Am Beispiel des COSMO-DE wurde die grundsätzliche Durchführbarkeit dieser Methode dargelegt. Für das COSMO-DE konnte sowohl für einen Zeitpunkt als auch für den tageszeitlichen Verlauf im Sommer 2007 innerhalb des deutschen Abschnitts des Elbeeinzugsgebiets folgendes aufgezeigt werden:

- Bei Einbeziehung der Zeitkomponente können mehr Situationen mit SAL analysiert werden.
- Der IQR Werte von S und A sind bei Betrachtung eines möglichen zeitlichen Versatzes der QNVs geringer und zeigen damit das Potenzial des COSMO-DE, die räumliche Niederschlagsverteilung zu simulieren.
- Die Medianwerte von S und A sind für die Nachmittagsstunden, im Zeitraum des meisten Niederschlags, mit Betrachtung der Zeitkomponente besser.
- Der zeitliche Versatz ist vor dem Tagesmaximum des Niederschlags positiv, währenddessen annähernd null und danach negativ.

Die Methode kann auch auf Akkumulationszeiträume von weniger als einer Stunde angewandt werden. Das Verwenden eines Zeitfensters von ± 3 Zeitpunkten um die Beobachtung kann ebenfalls geändert werden. Es sollte allerdings darauf geachtet werden, dass innerhalb des Zeitfensters alle Vorhersagen zur Verfügung stehen. Andernfalls werden möglicherweise verhältnismäßig viele positive bzw. negative Werte für Δt bestimmt, wenn vor bzw. nach dem Beobachtungstermin nicht alle Vorhersagen zur Verfügung stehen. Mit Hilfe dieser Methode kann untersucht werden, ob bei verschiedenen Wettersituationen unterschiedliche charakteristische zeitliche Vorhersagefehler auftreten.

Kapitel 4

MAP D-PHASE

Im Zentrum dieses Kapitels steht der Vergleich der Niederschlagsvorhersagen von 19 deterministischen Modellen aus dem MAP D-PHASE Datensatz. Dieser Datensatz bietet die einmalige Möglichkeit, eine große Anzahl an Wettermodellen für das selbe Gebiet bzgl. ihrer Güte der Niederschlagsvorhersage zu vergleichen. In der Vergleichsstudie soll untersucht werden, ob die Modelle der neuesten Generation, die ohne eine Parametrisierung hochreichender Konvektion betrieben werden, bessere Niederschlagsverteilungen simulieren als Modelle mit Konvektionsparametrisierung. Die Verifikation der QNVs wird mit Hilfe des objekt-orientierten Bewertungsmaßes SAL (Kapitel 3.1) für den Südwesten Deutschlands im Sommer 2007 durchgeführt. Die Analyse erfolgt für Akkumulationsperioden von 6 und 12 Stunden. Des Weiteren werden die Vorhersagen separat für zwei typische Kategorien sommerlicher Wettersituationen untersucht.

4.1 Grundlegende Ausführungen

Die Verifikation von Quantitativen Niederschlagsvorhersagen (QNV) ist durch die starke räumliche und zeitliche Variabilität der Niederschlagsverteilung (*Ebert et al., 2003*) eine große Herausforderung für die numerische Wettervorhersage. Die Simulation von Niederschlag ist eine komplexe Aufgabe, da sie atmosphärische Prozesse beschreiben muss, die einen großen Skalenbereich umfassen. Es müssen die Vorgänge auf der synoptischen Skala (~ 1000 km) wie auch die Wolkenmikrophysik (~ 10 μm) simuliert werden (*Richard et al., 2007*). Um diese Herausforderung der numerischen Niederschlagsvorhersage zu meistern, wurden in den letzten Jahren regionale Wettermodelle entwickelt und operationell in Betrieb genommen, die typischerweise einen horizontalen Gitterpunktabstand von 1 - 4 km besitzen. Diese Größenordnung der horizontalen Maschen-

weite erlaubt, es die Vorhersage von sommerlichen Niederschlägen ohne eine Parametrisierung von hochreichender Konvektion durchzuführen, wie es bei Modellen mit größeren horizontalen Gittern notwendig ist (siehe Kapitel 1.1). Für die Verifikation von hochaufgelösten QNVs wurden während der letzten Dekade mehrere neuartige Auswertungsstrategien entwickelt (siehe Kapitel 1.2.2). Mit Einführung dieser neuartigen Modelle in den operationellen Betrieb der Wetterdienste stellt sich die Frage, ob diese Simulationen besser sind, als die der Modelle der vorherigen Generation (*Davis and Carr, 2000; Mass et al., 2002; Kain et al., 2008*). Vergleichende Studien von Modellversionen mit und ohne Konvektionsparametrisierung für konvektive Situationen zeigen ein uneinheitliches Bild. Einerseits scheinen sie subjektiv eine realistischere Niederschlagsverteilung zu simulieren, aber dieses findet sich nicht in den objektiven Ergebnissen der klassischen Bewertungsmaße wieder. Dies zeigen zum Beispiel *Done et al. (2004)* anhand des *ETS* und des *FBI* für mesoskalige konvektive Systeme in den USA für zwei Versionen des WRF Modells für tägliche Niederschlagsakkumulationen. Ein ähnliches Ergebnis finden *Clark et al. (2007)* für tägliche Niederschlagssummen in den USA im Zeitraum April bis Juli 2005 bei Betrachtung des *ETS* und des *FBI*. Dagegen liefert die Studie von *Liu et al. (2006)* anhand von zwei konvektiven Fallstudien, eine bei mäßigen und eine andere bei schwachen synoptischen Antrieb, in den USA einen Hinweis darauf, dass die Modellversion ohne Konvektionsschema bessere Simulationen realisiert. Für Großbritannien konnten *Roberts (2008)* sowie *Lean et al. (2008)* mit einer neuartigen skalen-selektiven Methode zeigen, dass die Modellversion mit 1 km Gitterabstand für konvektive Ereignisse bessere Ergebnisse liefert als diejenige mit 12 km und Konvektionsparametrisierung. In der Publikation von *Amodei and Stein (2009)* konnte unter Verwendung von 'Fuzzy'-Verifikationsverfahren (siehe Abschnitt 1.2.2) bei Analysen für Februar bis Juli 2007 in Frankreich für die höher aufgelöste Version keine bessere Vorhersagequalität aufgezeigt werden. In *Clark et al. (2009)* wurde ein hochaufgelöstes Ensemble ohne Konvektionsparametrisierung gegen eines mit Konvektionsschema verglichen. Im Zeitraum von April bis Juni 2007 in den USA zeigten die meisten hochaufgelösten Ensemble-Mitglieder eine höhere Güte.

Im Jahr 2007 wurden in Zentraleuropa zwei große internationale Projekte mit dem Ziel eines besseren Verständnisses der Niederschlagsprozesse und der -vorhersagen durchgeführt. Zu einem war das die Messkampagne „Convective and Orographically induced Precipitation Study“ (COPS, *Wulfmeyer et al., 2008*) von Juni bis August 2007 in Südwestdeutschland und Nordostfrankreich, die den Zweck hatte, mit Hilfe von Beobachtungen und Messungen konvektive Wettersituationen im Mittelgebirgsraum zu untersuchen. Parallel dazu gab es die Demonstrationsperiode von MAP D-PHASE (Mesoscale Alpine Project Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood

Events in the Alpine region; *Arpagaus et al.*, 2009) von Juni bis November 2007 für den Alpenraum. Deren Ziel war es die Hochwasservorhersagen im Alpenraum zu verbessern und die Informationen von einer großen Anzahl von deterministischen wie probabilistischen Modellvorhersagen zu nutzen.

Aufgrund des Vorhersagedatensatzes des D-PHASE Projektes existiert die einmalige Gelegenheit, QNVs von deterministischen NWV Modellen verschiedenster Konfiguration miteinander zu vergleichen und die Frage zu beantworten, ob systematische Unterschiede in der Vorhersagegüte zwischen Modellen mit und ohne Konvektionsparametrisierung existieren. Außerdem bietet die COPS-Periode die Möglichkeit eine wetterlagenspezifische Untersuchung sommerlicher konvektiver Niederschlagsereignisse durchzuführen.

4.2 Daten und Methodik

Bevor die Verifikation sowie der Vergleich der QNVs von 19 deterministischen Modellen beschrieben wird, erfolgt an dieser Stelle die Beschreibung der benötigten Daten.

Untersuchungsgebiet und -zeitraum

Das in dieser Studie gewählte Untersuchungsgebiet befindet sich im Südwesten Deutschlands (Abb. 4.1) und ist der deutsche Abschnitt des COPS Gebietes. Im Sommer 2007 wurde diese Messkampagne in Südwestdeutschland und Nordostfrankreich durchgeführt. Diese Region ist charakterisiert durch verschiedene topografische Begebenheiten, wie dem Oberrheingraben, dem Schwarzwald sowie der Schwäbischen Alb. Begrenzt wird dieses Gebiet durch die Grenze zu Frankreich im Westen und zur Schweiz im Süden. Im Osten stellt 10°O und im Norden 49°N die Abgrenzung des COPS-Gebietes dar. Die Untersuchungsperiode dieser Analyse, Juni bis August 2007, ist identisch zu der des Feldexperiments COPS.

Beobachtungsdaten

Als Niederschlagsbeobachtungen werden die disaggregierten Daten (Kapitel 1.1.2) verwendet. Dieser mit stündlicher Auflösung für Deutschland vorliegende Datensatz wird für diese Studie zu einer 12 h Summe (06-18 UTC) und zwei 6 h Summen (06-12 und 12-18 UTC) akkumuliert. Die Wahl dieser Akkumulationsperioden erfolgte nach Betrachtung des Niederschlagstagesgangs im COPS-Gebiet 2007 der einzelnen Modelle. Viele

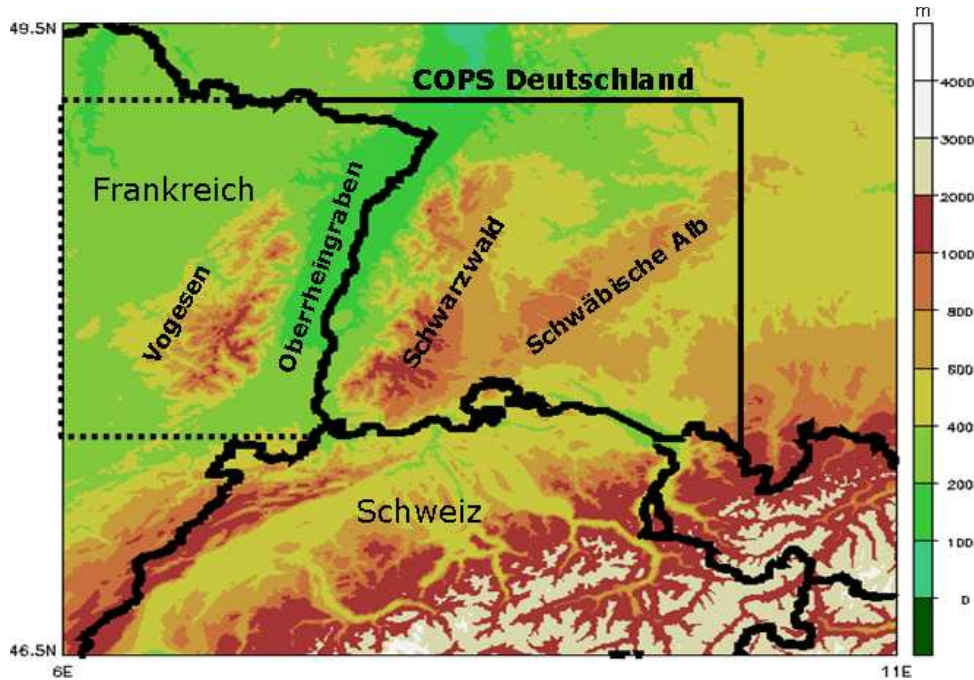


Abbildung 4.1. Topografie der COPS Region mit dem Oberrheingraben, dem Schwarzwald, der Schwäbischen Alb und den Vogesen. Die „COPS Deutschland“ Region wird begrenzt durch die deutsche Grenze (Süden und Westen) sowie durch gerade Linien im Norden bei 49°N und im Osten bei 10°O . Der französische Abschnitt des COPS Gebietes ist mit einer gepunkteten Linie markiert.

der Modelle weisen in den ersten sechs Stunden sogenannte 'Spin-up' Probleme auf, so dass erst nach sechs Stunden alle Modelle einigermaßen die Beobachtung simulieren.

Numerische Niederschlagsvorhersagen

Für diesen Vergleich der Vorhersageleistung verschiedenster Wettermodelle wurden Simulationen von 18 deterministischen regionalen Modellen aus dem MAP D-PHASE Datensatz extrahiert. Die 18 Modelle wurden anhand der Verwendung der Parametrisierung von hochreichender Konvektion in zwei Kategorien unterteilt. Das Ergebnis dieser Unterteilung führt zu einer Kategorie mit zehn Modellen die eine Parametrisierung hochreichender Konvektion besitzen und die im Folgenden als 'Medium'-Modelle bezeichnet werden sollen. Die Kategorie der Modelle ohne eine Parametrisierung von hochreichender Konvektion enthält acht Modelle, die als 'Hoch'-Modelle bezeichnet werden. Außerdem werden die Niederschlagsvorhersagen des globalen deterministischen und hydrosta-

Modell	horizontaler Gitterabstand	Parametrisierung hochreichender Konvektion	Modellkategorie
ECMWF	~30 km	ja	Global
MM5_60	60 km	ja	Medium
MM5_15	15 km	ja	Medium
QBOLAM11	11 km	ja	Medium
ALADAT	9,6 km	ja	Medium
ALADFR	9,5 km	ja	Medium
MESONH8	8 km	ja	Medium
COSMO-EU	7 km	ja	Medium
COSMOCH7	7 km	ja	Medium
LMEURO	7 km	ja	Medium
LAMI7	7 km	ja	Medium
COSMO-DE	2,8 km	nein	Hoch
LMITA	2,8 km	nein	Hoch
AROME	2,5 km	nein	Hoch
COSMOCH2	2,2 km	nein	Hoch
MESONH2	2 km	nein	Hoch
MM5_CT	2 km	nein	Hoch
MM5_4D	2 km	nein	Hoch
ISACMOL2	2 km	nein	Hoch

Tabelle 4.1. Liste der 19 verwendeten numerischen Wettervorhersagemodelle mit der Information über den horizontalen Gitterpunkt Abstand (Δx in km), die Parametrisierung von hochreichender Konvektion und die jeweilige Modellkategorie.

tischen Globalmodells des ECMWF (*engl.: European Centre for Medium Range Weather Forecasts*) mit in die Analyse einbezogen. Damit liegt ein Datensatz mit 19 Modellen vor, der alle aktuell zur numerischen Wettervorhersage verfügbaren Modelltypen enthält. Diese reichen von klassischen, gut bewährten hydrostatischen globalen bis hin zu raffinierten, hochaufgelösten regionalen Modellen. Um die Simulationen der verschiedenen Modelle miteinander vergleichen zu können, ist es aufgrund der unterschiedlichen horizontalen Gitterweiten (2 bis 60 km, siehe Tabelle 4.1) notwendig, die Vorhersagen auf das Beobachtungsgitter mit sieben Kilometer Maschenweite, welches identisch mit dem des COSMO-EU ist, zu transformieren. Die Absicht dieser Studie ist es, die Güte der Vorhersagen der einzelnen Modelle bzgl. der beobachteten Niederschlagsverteilungen zu

bewerten. Im Fall der Modelle mit einer Maschenweite kleiner als 7 km erfolgt eine Aggregation der Vorhersagedaten. Dabei wird aus den Werten aller Gitterpunkte des feineren Gitters, die einen Abstand kleiner als 7 km zum Gitterpunkt des COSMO-EU haben, ein Mittelwert berechnet. Für Modelle mit einem horizontalen Gitterabstand von 7 km oder größer werden die Daten linear auf das Beobachtungsgitter interpoliert.

Synoptische Umgebungsbedingungen für Konvektion

Eines der Hauptziele der COPS Messkampagne im Sommer 2007 (*Wulfmeyer et al.*, 2008) war die Untersuchung von konvektiven Wettersituationen in bergigem Gelände. Während COPS wurden 34 sogenannte 'Intensive Beobachtungsperioden' (IOPs) festgelegt und für die meisten erfolgte eine Einteilung in für Konvektion charakteristische Wettersituationen (Kapitel 14 in *Wulfmeyer et al.*, 2007). Die in Kapitel 4.3 von *Wulfmeyer et al.* (2005) erstellte einfache klimatologische Analyse von sommerlichen konvektiven Tagen in Südwestdeutschland der Jahre 1996-2003 lieferte drei vorherrschende Situationen für Starkregenereignisse. Die beiden häufigsten konvektiven Wettersituationen in der Region „COPS Deutschland“ während des Sommers 2007 werden im Folgenden vorgestellt:

- *Luftmassenkonvektion* (engl.: *High Pressure Convection, HPC*) ist charakterisiert durch einen hohen Luftdruck bei gleichzeitiger Abwesenheit frontaler Zonen in der Umgebung der COPS Region. Die Auslösung von hochreichender Konvektion ist in diesen Situationen durch lokale Prozesse wie den Bodenwärmeflüssen dominiert. Im Sommer 2007 wurden 13 Tage identifiziert die diese Bedingung erfüllten.
- *Erzwungene Konvektion* (engl.: *Strongly Forced Convection, SFC*) tritt auf, wenn eine Region tiefen Drucks oder ein Trog sich über oder westlich der COPS Region befindet. Zusätzlich kann bodennah eine frontale Linie mit eingeschlossener Konvektion beobachtet werden sowie orographische Modifikationen der Strömung. Diese Wettersituation konnte im Sommer 2007 ebenfalls 13 Mal bestimmt werden.

4.3 Verifikationsergebnisse

Der Vergleich der QNV Leistung der drei betrachteten Modellkategorien wird im deutschen Abschnitt der COPS Region für den Sommer 2007 durchgeführt. Die Bewertung

der Vorhersageleistung wird unter Verwendung der objekt-orientierten Verifikationsmethode SAL (Kapitel 3.1) für verschiedene Akkumulationszeiträume durchgeführt. Für alle 19 NWV Modelle werden die 00 UTC Vorhersageläufe betrachtet, da diese für alle vorliegen. Daraus werden akkumulierte Niederschlagsfelder für die jeweiligen Zeiträume 06-18 UTC, 06-12 UTC und 12-18 UTC berechnet. Zunächst erfolgt die Beschreibung der Resultate von SAL der Zwölfstundenperiode und der Sechsstundenzeiträume. Daran anschließend wird für die Modellvorhersagen eine Rangliste der Güte erstellt. Abschließend erfolgt die Analyse der wetterlagenabhängigen Untersuchung.

12 h Akkumulationszeitraum

Als erstes erfolgt die Analyse der Akkumulationsperiode von 06 bis 18 UTC für alle 92 Tage. Dadurch dass die um 00 UTC gestarteten Vorhersageläufe verwendet werden, entspricht dieser Zeitraum dem Vorhersagezeitraum von der siebten bis zur 18. Stunde. Die ersten sechs Stunden der Simulationen werden nicht betrachtet, da diese möglicherweise bei einigen NWV Modellen durch 'Spin-up' Probleme beeinflusst sind. Für jeden Tag und jedes NWV Modell (falls verfügbar) werden die drei Komponenten von SAL berechnet. Die Darstellung der Ergebnisse erfolgt für jedes NWV Modell unter Verwendung eines sogenannten SAL Diagramms (Kapitel 3.1), in dem die Werte für jeden Tag von S , A und L in der Form eines Streudiagramms abgebildet werden. Daraus ergibt sich eine Anzahl von 19 SAL Diagrammen, die hier nicht gezeigt werden sollen.

Zur besseren und kompakten Darstellung der Resultate von allen 19 NWV Modellen wird der Interquartilsabstand (IQR, engl.: *InterQuartile Range*) sowie der Median aus den 92 QNVs für die drei Komponenten von SAL bestimmt. Anschließend ist es möglich, die Ergebnisse in einfacher Form für jede Komponente von SAL einzeln darzustellen (siehe Abb. 4.2 für die 12 h Akkumulationen). Bei dieser Art der Darstellung wird das Ergebnis eines perfekten Modells durch einen Punkt im Ursprung (Median = IQR = 0) abgebildet. Eine Zufallsvorhersage, dessen Realisierung unter Verwendung eines 'Monte-Carlo' Ansatzes bzgl. der Beobachtungsdaten (Wernli *et al.*, 2008) verwirklicht wird, hat für S und A einen Median von ungefähr null sowie Werte von über 2 beim IQR. Bei der L Komponente ergibt sich für den Median und den IQR ein Wert von ungefähr 0,35. Zur Überprüfung wie robust die Ergebnisse des Medianes und IQRs der einzelnen Modelle sind, wurden unter Verwendung der von Efron (1979) eingeführten sogenannten 'Bootstrap'-Methode Konfidenzintervalle geschätzt (Hamill, 1999). Die Bestimmung der Konfidenzintervalle erfolgt durch 1000-maliges Anwenden des 'Bootstrap'-Verfahrens und der Bestimmung von Median und IQR. Aus den so erhaltenen 1000 Werten für den Median und IQR, wird der IQR des Medianes und der IQR des IQRs ermittelt. Die Darstellung des Konfidenzbereichs erfolgt durch Linien, vertikal für den Median und horizontal für den IQR.

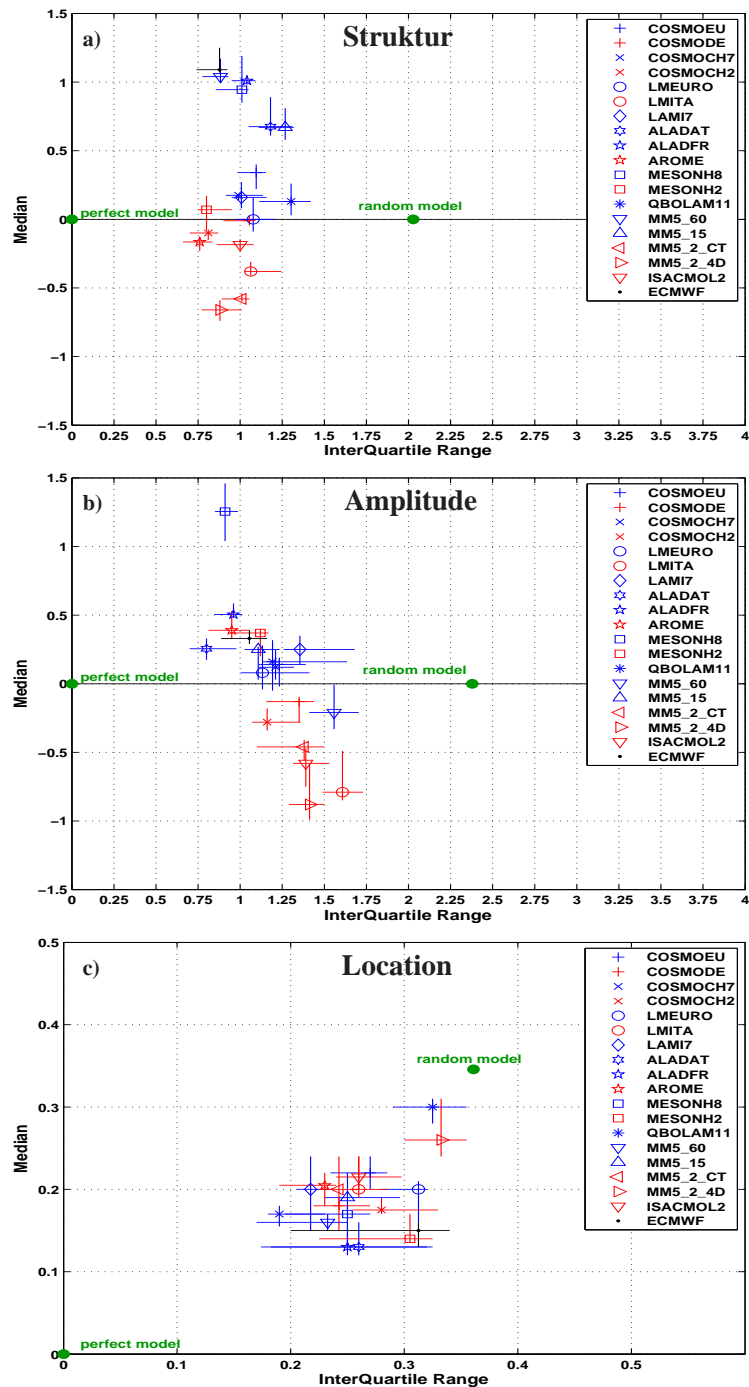


Abbildung 4.2. IQR versus Median für S a), A b) und L c) für die QNVs der 19 NWV Modelle mit einer Akkumulationsperiode von 12 h (06-18 UTC) im Sommer 2007 für den deutschen Abschnitt der COPS Region. Die Ergebnisse der 'Hoch'-Modelle werden als rote, die der 'Medium'-Modelle als blaue und die des globalen Modells als schwarze Symbole dargestellt. Zusätzlich eingezeichnet sind die Ergebnisse einer perfekten und einer zufälligen Vorhersage (grüne Punkte). Die vertikalen Balken geben den mit Hilfe des 'Bootstrap'-Verfahren geschätzten IQR des Medians und die horizontalen Balken den des IQRs an.

Im Fall der Strukturkomponente S (Abb. 4.2a) sind die Ergebnisse der drei Modellkategorien deutlich sichtbar voneinander abgetrennt. Das globale Modell (schwarzer Punkt) hat einen positiven Wert für den Median, in Einklang mit den Resultaten in *Wernli et al.* (2008). Das kann als Hinweis gedeutet werden, dass dieses Modell nicht in der Lage ist, die Struktur der Niederschlagsverteilung der Beobachtung adäquat zu simulieren. Die Eigenschaft dieses Modells ist die Realisierung von zu großflächigen und/oder zu flachen Niederschlägen, was durch den relativ großen horizontalen Gitterpunktabstand erklärt werden kann. Die Modelle mit einer mittleren horizontalen Auflösung (blaue Symbole) weisen eine Tendenz zur Überschätzung der Struktur auf (Median von $S \geq 0$), wobei die COSMO-Modelle Medianwerte relativ nah an der null aufweisen. Die MM5-Modelle, die ALADIN-Modelle und das MESONNH8 besitzen dagegen einen relativ großen positiven Medianwert. Die Ergebnisse der hochaufgelösten Modelle liefern Medianwerte von null oder im negativen Bereich. Die beiden hochaufgelösten MM5-Modelle haben mit Medianwerten von weniger als -0,5 im Mittel eine deutliche Unterschätzung der Struktur, d.h. die Simulation von zu kleinflächigen und/oder zu lokalisierten Niederschlagsereignissen. Beim Vergleich zwischen den Resultaten der Modelle mit mittlerer und hoher Auflösung, deren Median im Bereich von null liegt, kann festgehalten werden, dass die Werte des IQRs bei den hochaufgelösten Modellen geringer sind. Die mittels 'Bootstrap'-Verfahrens geschätzten Konfidenzintervalle für den Median und den IQR bestätigen die deutliche Separation der Ergebnisse zwischen den einzelnen Modellkategorien. Der Vergleich der NWV Modellergebnisse mit der Zufallsvorhersage zeigt, dass alle Modelle einen deutlich geringeren IQR-Wert aufweisen und somit eine gewisse Qualität besitzen. Die Betrachtung der Ergebnisse für die Amplitude (Abb. 4.2b) zeigt ein etwas weniger systematisches Bild im Vergleich zur S Komponente. Allerdings treten bei der A Komponente drei interessante Aspekte auf. Als erster Punkt soll die QNV Leistung der COSMO-Modelle zur Sprache kommen. Die COSMO-Modelle mit Parametrisierung von Feuchtkonvektion haben einen leicht positiven Median und weisen so gut wie keine Unterschiede auf. Dagegen kann bei den hochaufgelösten COSMO-Modellen erkannt werden, dass eine Unterschätzung der Niederschlagsmenge existiert und dass die Unterschiede untereinander doch relativ groß sind, besonders im Fall des LMITA. Als zweiter Punkt kann festgehalten werden, dass mit Ausnahme der beiden französischen Modelle (AROME und MESONH2) alle Modelle der hochaufgelösten Kategorie, ausgedrückt durch den Median, eine Unterschätzung der Menge des Niederschlags im betrachteten Akkumulationszeitraum aufweisen. Der dritte Aspekt der angesprochen werden soll, ist dass bei den betrachteten Modellketten die Version mit der höheren Auflösung eine geringere Niederschlagsmenge simuliert. Das gilt für Modellpaare bei denen das NWV Modell mit einer gröberen Auflösung eine Konvektionsparametrisierung besitzt und das Hochaufgelöste nicht. Im Fall des

MESONH und des ALADIN führt dies zu besseren Resultaten, d.h. zu einer geringeren Überschätzung der Niederschlagsmenge. Dagegen weisen bei den COSMO-Modellen die hochaufgelösten Versionen eine deutliche Unterschätzung der Niederschlagsmenge auf. Die Versionen mit einer Parametrisierung der hochreichenden Konvektion haben Medianwerte um null. Die IQR-Werte der hochaufgelösten COSMO-Modelle sind höher als die der COSMO-Modelle mit Parametrisierung der hochreichenden Konvektion. Im Gegensatz dazu weist das Modellpaar von MM5_60 und MM5_15 (beide mit Konvektionsparametrisierung) eine andere Charakteristik auf. Hier simuliert das Modell mit der gröberen Auflösung geringere Niederschlagsmengen und besitzt einen größeren IQR-Wert. Die A Komponente scheint die größte Variabilität aufzuweisen. Das wird deutlich am größeren Wert des IQRs der Zufallsvorhersage. Die Resultate für den IQR der Modellrealisierungen liegen in etwa bei der Hälfte des Wertes der Zufallsvorhersage und sind damit vergleichbar mit den Ergebnissen der Strukturkomponente.

In Abbildung 4.2c) sind die Ergebnisse für die Ortskomponente L dargestellt. Im Vergleich zu den Ergebnissen von S und A zeigt sich für L , dass viele Realisierungen der NWV Modelle nah an der Zufallsvorhersage sind und das damit L die am schlechtesten vorhergesagte Niederschlagseigenschaft im Sinne von SAL ist. Diese Schwierigkeit den Ort zu simulieren wurde schon in schon in *McBride and Ebert* (2000) bei der QNV Verifikation mit CRA in Australien festgestellt. Zusätzlich kann dies als Hinweis gedeutet werden, dass die Simulation des richtigen Ortes der sommerlichen Niederschlagsereignisse eine große grundsätzliche Herausforderung für die numerische Wettervorhersage darstellt.

Es existiert keine klare Abgrenzung zwischen den einzelnen Modellkategorien. Alle Modelle (außer QBOLAM11 und MM5_2_4D) befinden sich eng gedrängt und überlappen sich bei der Betrachtung der Konfidenzintervalle von Median und IQR. Allerdings scheint es so, dass die meisten Modelle mit Konvektionsparametrisierung (blaue Symbole) einen kleineren Verschiebungsfehler haben. Dies wird durch die geringeren Werte für den Median und den IQR von L deutlich. Beim Vergleich der NWV Modellpaare mit und ohne Parametrisierung von hochreichender Konvektion zeigt sich für einige Modelle (COSMO2/7, MESONH2/8, ALADINFR/AROME), dass die Versionen mit gröberer horizontaler Auflösung eine bessere Ortsdarstellung simulieren können. Im Fall des deutschen und des italienischen COSMO-Modellpaares sind die hochaufgelösten Versionen besser im Sinne von L , ohne dass deutliche Unterschiede existieren.

6 h Akkumulationszeitraum

Um der Frage nachzugehen, ob die für den zwölfstündigen Akkumulationszeitraum gefundenen Ergebnisse auch für kürzere Zeiträume gültig sind, wird der Untersuchungszeitraum in zwei sechsstündige Intervalle (erstes: 06-12 UTC; zweites: 12-18 UTC) unterteilt.

Die Ergebnisse der S Komponente sind ähnlich denen des zwölfstündigen Zeitraums (ist nicht abgebildet). Bei der Amplitude treten allerdings Unterschiede zum Zwölfstundenzeitraum auf. Beim ersten Intervall (Abb. 4.3a) kann eine Unterscheidung zwischen NWV Modellen mit und ohne Konvektionsparametrisierung erkannt werden. Die NWV Modelle

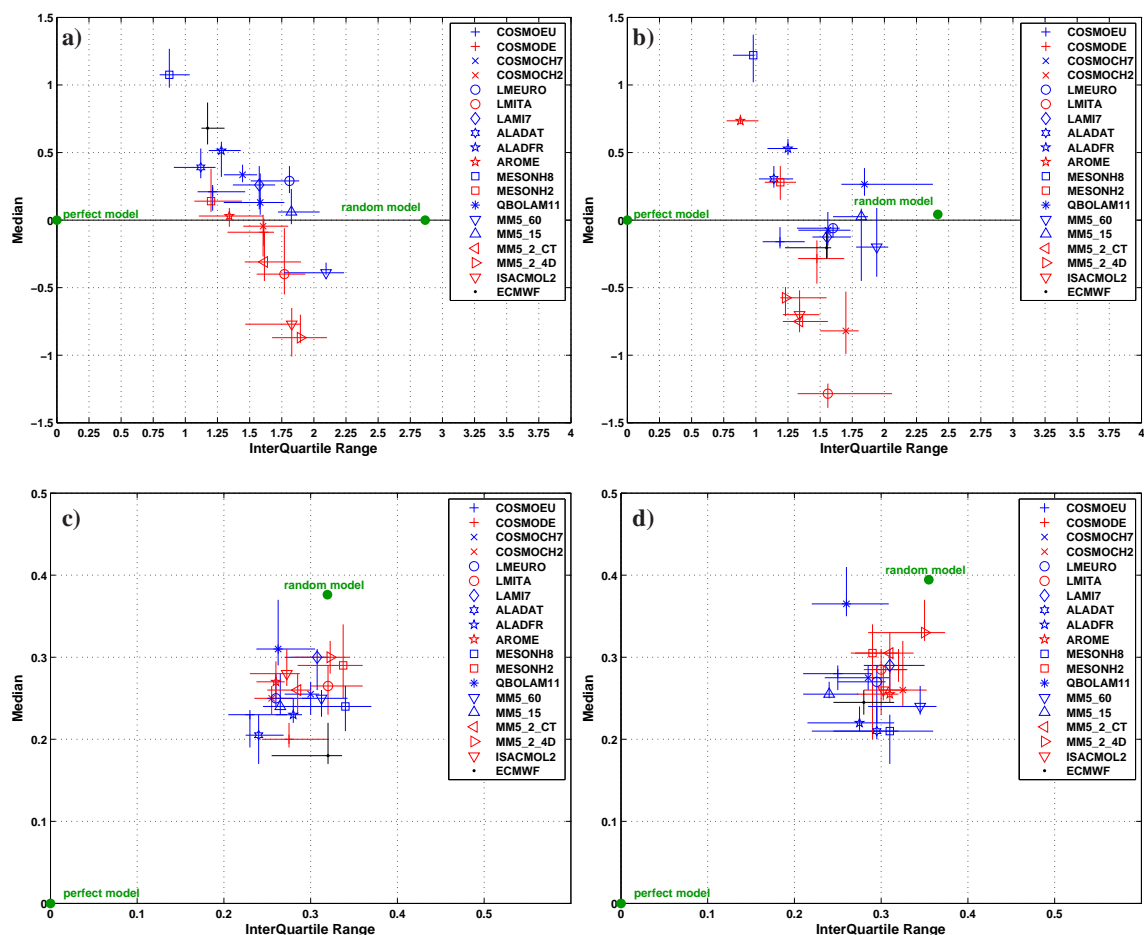


Abbildung 4.3. Wie Abbildung 4.2, aber für sechsstündige Akkumulationen (06-12 UTC (a,c) sowie 12-18 UTC (b,d)). Dargestellt sind die Ergebnisse für den Amplitudenfehler A (a,b) und Verschiebungsfehler L (c,d).

mit Konvektionsparametrisierung (blaue Symbole) haben positive Medianwerte (Ausnahme ist das MM5_60). Die NWV Modelle ohne solch eine Parametrisierung haben negative Medianwerte (Ausnahmen sind MESONH2 und AROME). Außerdem ist festzuhalten, dass die hochaufgelösten NWV Modelle einen kleineren IQR bei vergleichbaren Medianwerten besitzen. Das Bild der Ergebnisse von A des zweiten Intervalls (Abb. 4.3b) zeigt für die hochaufgelösten NWV Modelle (rote Symbole) mit Ausnahme der beiden französischen Modelle (MESONH2 und AROME) deutlich negative Medianwerte. Ein Großteil der NWV Modelle mit einer Konvektionsparametrisierung simulieren die Niederschlagsmenge (Median ≈ 0) gut. Ausnahmen bilden hier ebenfalls die französischen Modelle, die eine deutliche Überschätzung aufweisen. In Abbildung 4.3c,d sind zuletzt die Ergebnisse von L dargestellt. Beim Vergleich der beiden Intervalle fällt auf, dass besonders im zweiten Zeitabschnitt die NWV Modelle ohne Konvektionsparametrisierung die Position der Niederschlagsereignisse schlechter simulieren können. Zwar ist kein großer Unterschied zu erkennen, aber die Tendenz, dass mit fortschreitender Vorhersagezeit die L Komponente für die hochaufgelösten Modelle schlechter wird, ist vorhanden.

Rangliste der Modelle

Nachdem in den beiden vorangegangenen Abschnitten die Eigenschaften der Niederschlagsvorhersagen, die durch die drei Komponenten von SAL verifiziert werden, einzeln vorgestellt wurden, erfolgt an dieser Stelle der Versuch zur Beantwortung der Frage, welches Modell insgesamt die beste Vorhersageleistung abgeliefert hat. Um diese Frage zu beantworten, werden die Ergebnisse der drei Komponenten zusammengefasst. Dies geschieht anhand eines Maßes, das in Anlehnung an das Bewertungsmaß RV (Reduktion der Varianz, Kapitel 1.2.1) berechnet wird. Die RV ist ein sogenannter Skill-Score und benötigt daher eine Referenzvorhersage, die in dieser Studie die Zufallsvorhersage darstellt. Typischerweise wird die RV unter Verwendung des $RMSE$, z.B. für die S Komponente, folgendermaßen definiert:

$$RV_S = \left(1 - \frac{RMSE_S}{RMSE_{S,Zufall}} \right) \cdot 100, \quad (4.1)$$

mit $RMSE_S$ der $RMSE$ aller S Werte der Modellvorhersagen und $RMSE_{S,Zufall}$ der $RMSE$ aller S Werte der Zufallsvorhersagen bzgl. dem perfekten Wert null darstellt. Dabei werden Situationen, in denen kein Objekt in Beobachtung aber in der Vorhersage sowie im umgekehrten Fall identifiziert wird, mit $S,A,L = [2,2,2]$ bzw. $[-2,-2,2]$ behandelt und gehen somit in die Auswertung mit ein. Wird weder in der Beobachtung noch in der Vorhersage

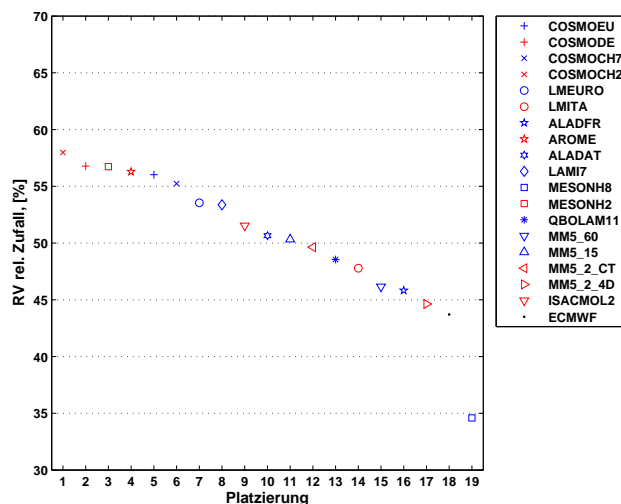


Abbildung 4.4. Rangliste der NWV Modelle für 12 h Akkumulationen (06-18 UTC) der 00 UTC Läufe für COPS Deutschland im Sommer 2007. Die Platzierung der Modelle wurde mit Hilfe von RV bestimmt.

Niederschlag festgestellt, so wird $S = A = L = 0$ für diese Vorhersage notiert. Bei dieser Analyse wird die RV für jede Komponente nach Gleichung 4.1 Art bestimmt. Anschließend erfolgt die Berechnung von $RV_{SAL} = (RV_S + RV_A + RV_L)/3$. Der Wertebereich ist zwischen $-\infty$ und 100, wobei positive Werte angeben das die betrachtete QNV besser ist als die Referenzvorhersage und 100 als bestem Wert. Mit Hilfe dieses Maßes kann gezeigt werden, um welchen Betrag die Modellvorhersage besser ist als die zufällige Referenzvorhersage.

Die Ergebnisse der Bestimmung einer Rangliste mittels RV_{SAL} für zwölfstündige Akkumulationen ist in Abbildung 4.4 dargestellt. Auf der vertikalen Achse ist die RV aufgetragen und die horizontale Achse gibt die Platzierung an. Es ist zu erkennen, dass die ersten vier Plätze durch hochaufgelöste Modelle (rote Symbole) besetzt werden. Die ersten beiden Plätze belegen das COSMO-7 und das COSMO-DE gefolgt von MESONH2 und AROME. Auf den Plätzen fünf bis acht folgen die COSMO-Modelle mit Konvektionsparametrisierung. Einzig das MESONH8 fällt deutlich ab von den restlichen Modellen. Alle Modelle weisen deutlich positive Werte auf, so dass sie gegenüber dem Zufallsmodell eine bessere Vorhersageleistung aufweisen.

Nach der Betrachtung der Rangliste für den zwölfstündigen Akkumulationszeitraum werden an dieser Stelle die beiden sechsstündigen Zeiträume vorgestellt (Abb. 4.5). Im Gegensatz zum Ergebnis des Zwölfstundenzeitraums ist in beiden Fällen das ranglistenbeste Modell eines mit Konvektionsparametrisierung. Im ersten Zeitraum (Abb. 4.5a) ist das

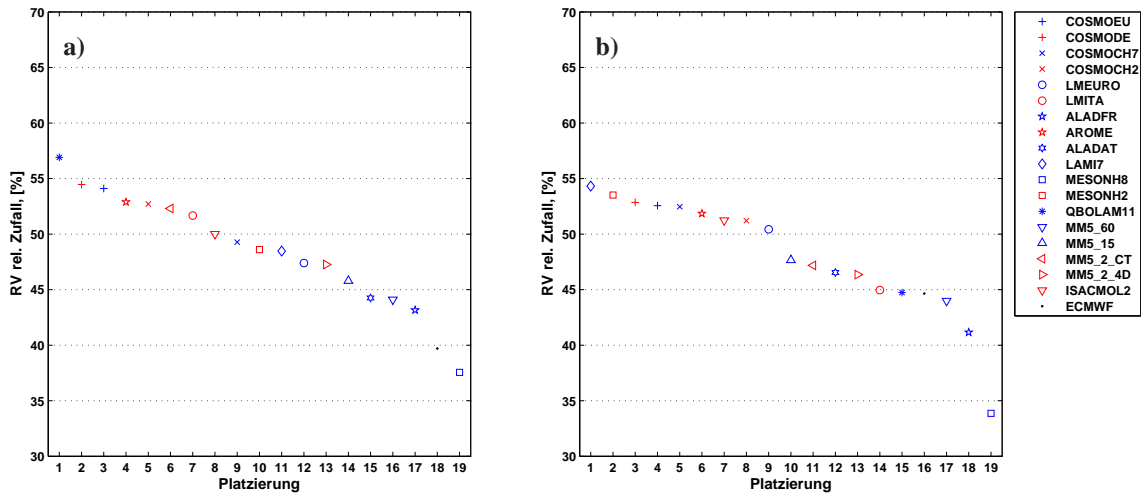


Abbildung 4.5. Wie Abbildung 4.4, aber für sechsstündige Akkumulationen. Links a) 06-12 UTC und rechts b) für 12-18 UTC.

QBOLAM11 am besten, dessen antreibendes Modell immer 12h vorher gestartet wird. Im Gegensatz dazu werden bei allen anderen Modellen die antreibenden Modelle immer gleichzeitig dazu gestartet. Unter den ersten zehn Modellen in diesem Zeitraum befinden sich sieben hochaufgelöste, d.h. alle bis auf eines. Das LAMI7 (Abb. 4.5a) ist im zweiten Zeitraum am besten und das QBOLAM11, welches im ersten Zeitraum am besten abschneidet, ist hier eines der schlechtesten. Bei den Ergebnissen des zweiten Zeitraums sind weniger hochaufgelöste Modelle unter den ersten zehn Rängen (5 anstatt 7). Dies kann als Hinweis darauf gedeutet werden, dass die Vorhersagequalität dieser Kurzzeitfristmodelle schneller mit fortschreitender Vorhersagezeit abfällt als die der 'Medium'-Modelle. Das MESONH8 ist, wie im Zwölfstundenintervall, in beiden Teilintervallen am schlechtesten, fällt aber im späteren Teil deutlicher ab. Mit der Verwendung der RV_{SAL} konnte gezeigt werden, dass die hochaufgelösten Modelle besser sind. Sie profitieren bei diesen Maß von der geringeren Streuung der einzelnen Werte, besonders bei der S Komponente. Im Gegensatz zur Darstellung in den Abbildungen 4.2 und 4.3 werden in diesem Fall alle Situationen bewertet, also auch die Nichtereignisse sowie die Verpassten.

Wetterlagenspezifische Analyse

Aufgrund der Tatsache, dass COPS eine Feldmesskampagne zur Untersuchung konvektiver Niederschlagsereignisse war, werden an dieser Stelle die Vorhersagen für zwei typische Kategorien konvektiver Wettersituationen analysiert. Die Untersuchung wird für

drei Akkumulationsperioden (12 h von 06-18 UTC, 6 h von 06-12 UTC sowie 12-18 UTC) durchgeführt und die Ergebnisse werden in Form von SAL Diagrammen präsentiert. Zur Verifikation erfolgt eine Unterteilung in zwei NWV Modellkategorien, mittlere und hochaufgelöste, d.h. die Kategorien können als eine Art 'Poor-Mans-Ensemble' (Ebert, 2001) mit zehn ('Medium') und acht ('Hoch') Mitgliedern verstanden werden. Die Vorhersagen des deterministischen hydrostatischen Globalmodells des ECMWF werden bei der wetterlagenabhängigen Untersuchung nicht mit einbezogen.

Luftmassenkonvektion (HPC)

Als erstes wird die Analyse für die Tage in der Kategorie Luftmassenkonvektion vorgestellt. Die SAL Diagramme der Ergebnisse werden in Abbildung 4.6 präsentiert. Die Resultate der zwölfstündigen Akkumulationsperiode sind in der oberen, die des ersten sechsstündigen Periode (06-12 UTC) in der mittleren und die der zweite Periode (12-18 UTC) in der untersten Reihe dargestellt. In der linken Spalte befinden sich die Ergebnisse aller Modelle mit und in der rechten die aller Modelle ohne Konvektionsparametrisierung.

Im Fall des zwölfstündigen Akkumulationszeitraums (obere Zeile in Abb. 4.6) zeigen sich deutliche Unterschiede in der Qualität der Niederschlagsvorhersagen der beiden Modellkategorien. Die NWV Modelle mit Konvektionsparametrisierung weisen im Mittel eine leichte Überschätzung der Niederschlagsmenge (Median von $A \approx 0,4$) auf und dabei gleichzeitig viel zu großflächige Niederschlagsereignisse (Median von $S \approx 1$). Das SAL Diagramm der hochaufgelösten Modelle zeigt in diesem Fall eine deutliche Unterschätzung der Niederschlagsmenge (Median von $A \approx -1,6$) und der Struktur (Median von $S \approx -1$). Einige wenige Punkte befinden sich im Bereich des Ursprungs. Diese QNVs stammen von AROME und MESONH2, die allerdings während des gesamten Zeitraums eine massive Überschätzung der Niederschlagsmenge und damit eine völlig andere Charakteristik als die restlichen hochaufgelösten Modelle (siehe Abb. 4.2) aufweisen. Die Ortskomponente L ist für beide Modellkategorien relativ schlecht, was an den vielen gelben und roten Punkten zu erkennen ist.

Nachdem der Zwölfstundenzeitraum untersucht wurde, werden jetzt die beiden sechsstündigen Teilintervalle vorgestellt. Die Ergebnisse der beiden Zeiträume zeigen beim Vergleich der Modellkategorien ein stark unterschiedliches Bild. Für die erste Akkumulationsperiode weisen die Modelle mit Konvektionsparametrisierung im Mittel eine sehr deutliche Überschätzung der Niederschlagsmenge (Median von $A \approx 1,3$) und viel zu großflächige Niederschlagsobjekte (Median von $S \approx 1,5$) auf. Dagegen ist das Bild bei den hochaufgelösten Modellen sehr uneinheitlich. Zwar ist der Median von A annähernd null, allerdings ist der IQR mit einem Wert von 2,6 relativ groß. Der Median von S ist leicht

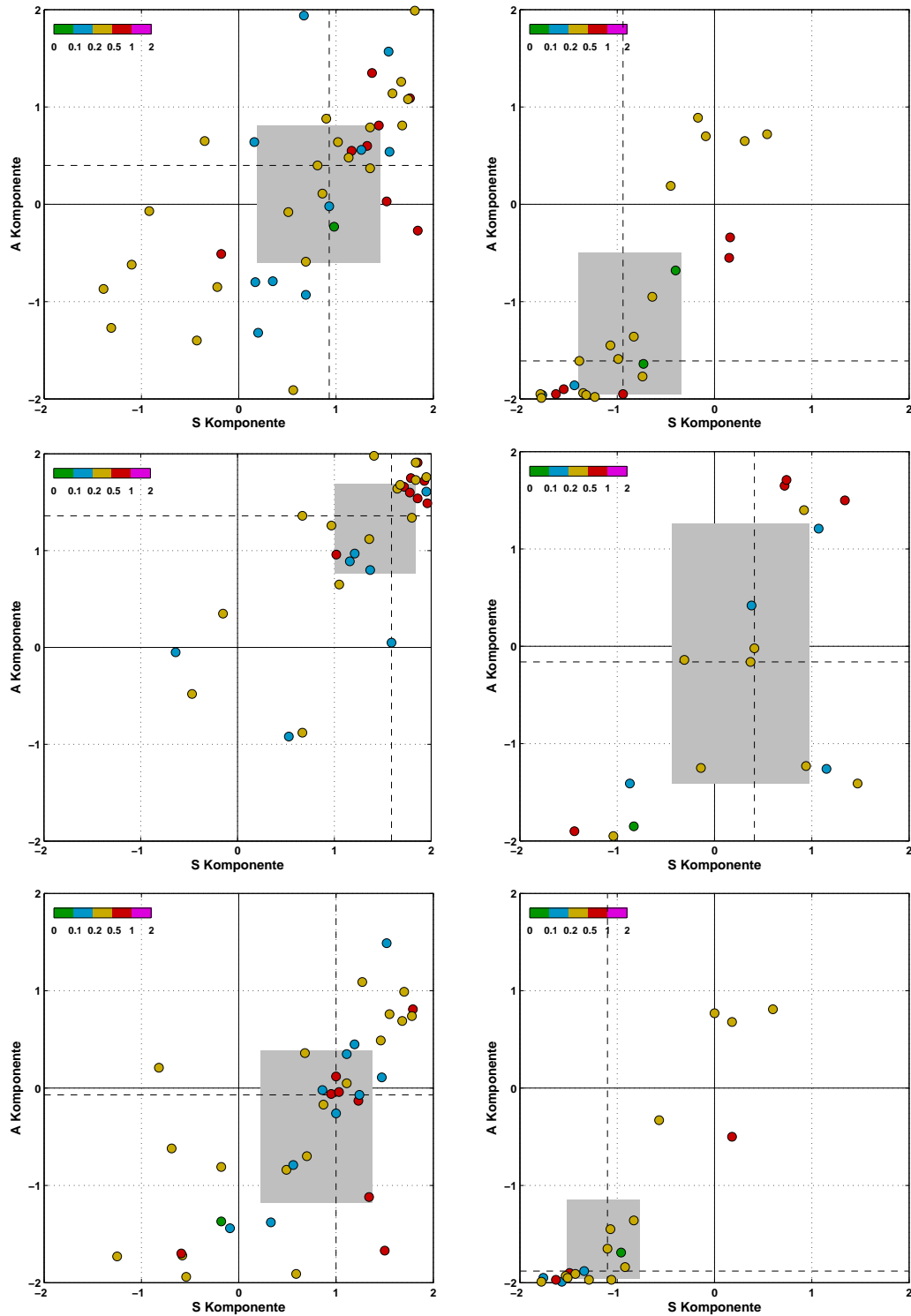


Abbildung 4.6. SAL Diagramme von QNVs für unterschiedliche Akkumulationsperioden (obere Reihe 06 - 18 UTC, mittlere Reihe 06 - 12 UTC und untere Reihe 12 - 18 UTC) bei Luftmassenkonvektion im deutschen Abschnitt der COPS Region im Sommer 2007. Die linke Spalte zeigt die Ergebnisse der Modelle mit Konvektionsparametrisierung und die rechte Spalte der Modelle ohne Konvektionsparametrisierung.

positiv (0,4) und der IQR mit ca. 1,5 nicht so groß wie im Fall der Amplitude. Es können im SAL Diagramm drei getrennte Gruppen erkannt werden. Die erste mit stark negativen Werten von A (< -1) mit einem weitem Spektrum an Werten von S . Eine zweite Gruppe mit Werten von S und A um die null sowie eine dritte mit Werten von S um und A größer +1. Keiner dieser Gruppen kann einem Modelltyp (z.B. COSMO) zugeordnet werden. Im Fall des zweiten Teilzeitraums sind die Resultate analog zu denen im gesamten Zeitraum, d.h. die Modelle mit Konvektionsparametrisierung simulieren im Mittel die richtige Niederschlagsmenge, allerdings viel zu flächig. Dagegen weisen die Modelle ohne Konvektionsparametrisierung eine erhebliche Unterschätzung der Niederschlagsmenge sowie der Struktur auf.

Erzwungene Konvektion (SFC)

Als zweites werden die Ergebnisse für Wettersituationen betrachtet, bei denen die Konvektion durch synoptische Prozesse wie beispielsweise Fronten erzwungen wird. Im Gegensatz zu den Verifikationsergebnissen der Luftmassenkonvektion sind die Unterschiede zwischen beiden Modellkategorien für diese Situation weniger stark ausgeprägt. Grundsätzlich kann festgehalten werden, dass die Resultate für diese konvektive Wettersituation ziemlich gut sind, welches anhand der vielen grünen Punkte in der Nähe des Ursprungs abgeleitet werden kann. Die schwachen erkennbaren Unterschiede sind hauptsächlich an den Werten des IQR für S und A zu erkennen. Im Fall des Zwölfstundenzeitraums ist der Median von S und A bei beiden Kategorien ungefähr null, wobei der Median von S bei den Modellen mit Konvektionsparametrisierung mit +0,4 positiv und bei denen ohne mit -0,1 leicht negativ ist. Der IQR für S und A ist für die Modelle ohne Konvektionsparametrisierung etwas kleiner als 1. Dagegen sind die Werte bei den Modellen mit Konvektionsparametrisierung bei ca. 1,5, wobei der IQR von A etwas größer ist.

Das Bild der beiden Teilintervalle unterscheidet sich kaum von dem des gesamten Zeitraums. Die Ergebnisse von A unterscheiden sich im ersten Zeitraum zwischen beiden Modellkategorien kaum, weder im Median noch im IQR. Im zweiten Teilzeitraum ist der IQR von A der höher aufgelösten Modelle etwas kleiner, allerdings weist hier der Median einen deutlich negativen Wert auf. Die Ergebnisse von S zeigen, dass die höher aufgelösten Modelle dazu neigen, zu kleine und starke Ereignisse zu simulieren. Im Gegensatz dazu in der Vorhersagen der weniger hochauflösten Modelle zu großflächige. Im ersten Zeitraum ist der IQR von S für beide Modellkategorien mit 1,4 ungefähr gleich, aber im zweiten Intervall haben die höher aufgelösten Modelle einen deutlich geringeren IQR im Vergleich zur anderen Modellkategorie. Für beide Teilzeiträume kann bei den hochauflösten Modellen festgestellt werden, dass eine Unterschätzung der Niederschlagsmen-

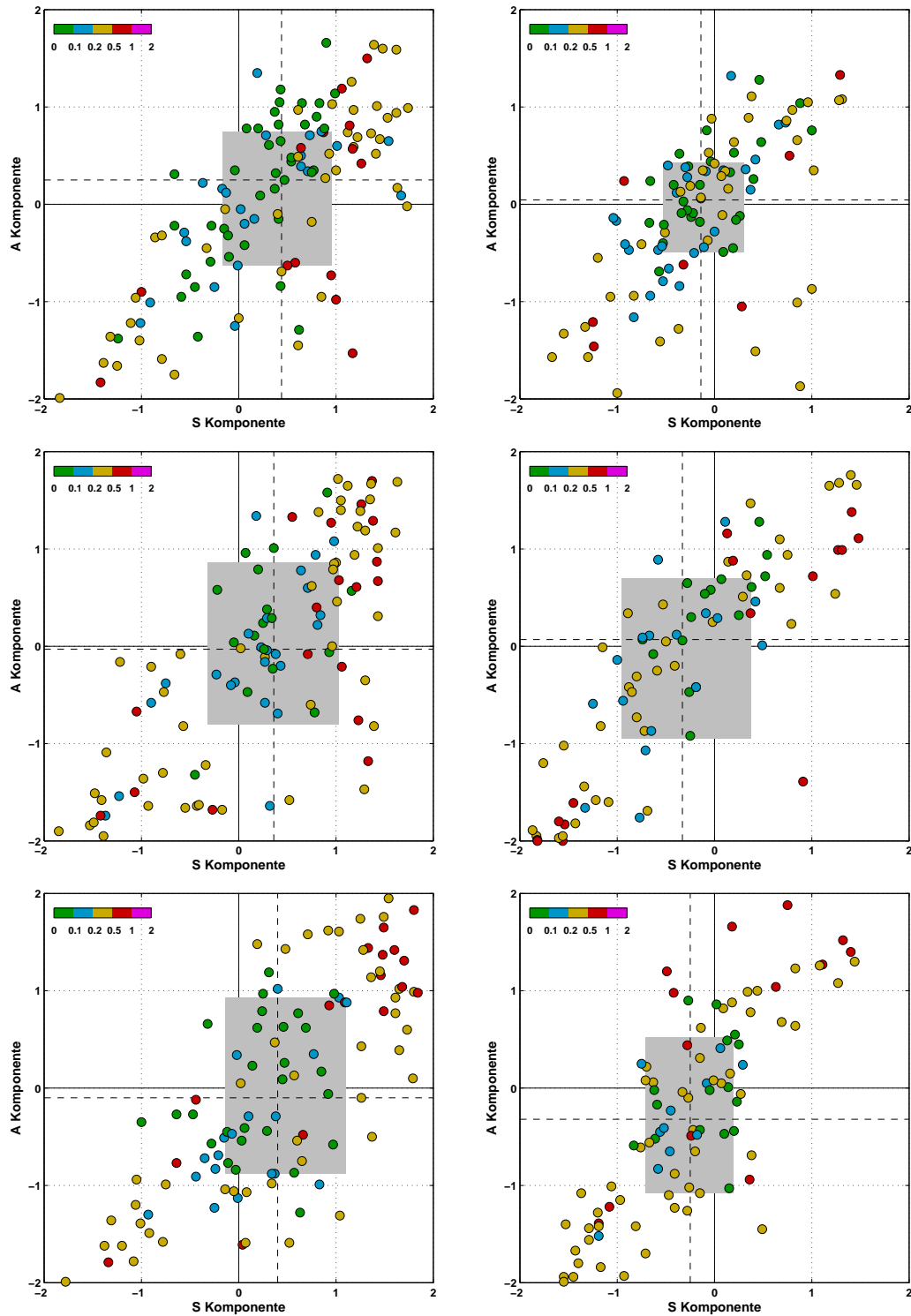


Abbildung 4.7. Wie Abbildung 4.6, jedoch für Wettersituationen in der Kategorie „Erzwungene Konvektion“.

ge ($A < 0$) und gleichzeitig zu großflächige Ereignisse ($S > 0$) weniger häufig auftreten als in den Simulationen der Modelle mit Konvektionsparametrisierung.

4.4 Zusammenfassung der Ergebnisse

Hauptziel dieser Analyse war der Vergleich der QNV Leistung verschiedener, deterministischer NWV Modelle im Rahmen des MAP D-PHASE Projektes. Die QNVs von 19 NWV Modellen wurden für Sommer 2007 in der COPS-Deutschland Region miteinander verglichen. Die Analyse erfolgte für Sechs- und Zwölfstundenakkumulationen der um 00 UTC gestarteten Läufe. Außerdem wurden Untersuchungen für zwei in der Zeitperiode typischen konvektiven Wettersituationen vorgestellt.

Die wichtigsten Ergebnisse der Verifikation mit SAL können folgendermaßen festgehalten werden:

- Modelle ohne Parametrisierung hochreichender Konvektion können die Struktur von Niederschlagsereignissen realistischer simulieren, was durch kleinere Werte des Medianes und besonders des IQRs von S belegt wird.
- Bei Modellpaaren weisen die Simulationen der höher aufgelösten Version, d.h. diejenige ohne Konvektionsparametrisierung, eine geringere Niederschlagsmenge auf.
- Mit zunehmender Vorhersagezeit wird die Unterschätzung der Niederschlagsmenge bei hochaufgelösten Modellen stärker.
- Die L Komponente ist die am schlechtesten durch SAL bewerteten Eigenschaften der Niederschlagsvorhersagen, da die Ergebnisse nah an der Zufallsvorhersage liegen.
- Für die hochaufgelösten Modelle werden die Ergebnisse von L mit fortschreitender Vorhersagezeit schlechter, als die der Modelle mit Konvektionsparametrisierung.
- Die hochaufgelösten Modelle ohne Konvektionsparametrisierung liefern insgesamt die besseren Vorhersagen, was durch höhere Werte beim RV_{SAL} gezeigt wird.
- Im Fall der Luftmassenkonvektion liefern die Simulationen der Modelle mit Konvektionsparametrisierung zu viel und zu großflächigen Niederschlag, wobei für die hochaufgelösten Modelle eine erhebliche Unterschätzung von S und A festzuhalten ist. Die Teilintervalle zeigen ein unterschiedliches Bild, wobei der Abschnitt von

12 - 18 UTC denjenigen von 06 - 18 UTC gleicht. Der erste Abschnitt zeigt eine extreme Überschätzung von S und A der Modelle mit Konvektionsparametrisierung und die ohne eine starke Streuung.

- Für die Situationen der erzwungenen Konvektionen sind geringe Unterschiede zwischen den Modellkategorien zu erkennen, die sich hauptsächlich durch eine kleinere Streuung der Punkte im SAL Diagramm der hochaufgelösten Modelle bemerkbar macht. Diese Ereignisse werden durch beide Modelltypen gut simuliert.

Abschließend bleibt festzuhalten, dass die höher aufgelösten Modelle ohne Parametrisierung hochreichender Konvektion zwar in der Struktur realistischere Niederschlagsverteilungen simulieren können, allerdings nimmt die QNV Leistung betrachtet in Amplitude und Ort mit fortschreitender Vorhersagezeit stärker ab, als bei den Modellen mit Konvektionsparametrisierung. Die Ursache, dass die hochaufgelösten Modelle in der Bewertung der klassischen Bewertungsmaße schlechter abschneiden, obwohl sie subjektiv betrachtet realistischer sind, liegt wahrscheinlich an dieser Eigenschaft. Besonders der Verschiebungsfehler führt zu einem starken 'Double Penalty'. Durch die Verwendung von SAL konnten einzelne Eigenschaften der Niederschlagsverteilungen genauer analysiert werden.

Des Weiteren ist anzumerken, dass beide Modellkategorien bei der Simulation von Niederschlagsereignissen bei Luftmassenkonvektion erhebliche Probleme aufweisen, die allerdings völlig unterschiedlich sind. Die neue Generation ohne Konvektionsparametrisierung ist nicht in der Lage, diese Situationen zu erfassen, sondern verpasst bzw. unterschätzt diese Ereignisse. Diese Resultate sind daher vergleichbar mit denen von *Liu et al.* (2006) für zwei Fallstudien konvektiver Wetterepisoden, die denen der beiden in dieser Studie analysierten Wettersituationen entsprechen. Aus der schwachen Vorhersageleistung von Luftmassenkonvektion lässt sich ableiten, dass eine horizontale Maschenweite von 1 km oder kleiner notwendig ist, wie in *Bryan et al.* (2003) vorgeschlagen, um diese Situationen adäquat zu simulieren. Dabei werden allerdings auch Prozesse explizit simuliert, deren Fehlerwachstum aufgrund ihrer nichtlinearen und chaotischen Natur besonders groß sind (*Walser et al.*, 2004). Aus diesem Grund sollte laut *Mass et al.* (2002) auch mit immer feineren Gittern kaum bzw. keine Verbesserung der Niederschlagsvorhersage im deterministischen Sinne zu erwarten sein. Einen Vorteil der Verfeinerung der Gitter ist die realistischere Simulationen der Niederschlagsstrukturen, wie es durch die besseren Werte von S bei den hoch aufgelösten Modellen bestätigt werden konnte. Um diese Charakteristik auszunutzen schlagen z.B. *Gebhardt et al.* (2008) eine probabilistische Herangehensweise vor. Ein weiteres wichtiges Problem, was bei immer feineren Gitter auftaucht, ist die Bereitstellung von extrem hochaufgelösten Beobachtungen (*Benoit et al.*, 2002),

da die NWV mathematisch betrachtet die Lösung eines Anfangswertproblems darstellt und kleine Unsicherheiten auf dieser Skalen besonders schnell anwachsen. Auch die Versorgung von Randdaten stellt eine Quelle von Unsicherheiten dar. Zusätzlich sind weiterhin viele Prozesse (z.B. Strahlung) parametrisiert. Die Konzeption der meisten dieser Parametrisierungen hatte ihren Ursprung für Modelle mit Gitterweiten im Bereich von $\Delta x \approx 100\text{km}$, so dass an dieser Stelle eine Modifizierung oder völlige Neuentwicklung wünschenswert, wenn nicht notwendig ist.

Daraus kann abgeleitet werden, dass die einfache Verringerung der Gitterweite allein nicht ausreichend ist. Zum einen, weil die Fehler in den Anfangs- und Randdaten auf diesen kleinen Skalen sehr schnell anwachsen. Zum anderen muss die Wechselwirkung von subgitterskaligen Prozessen mit den explizit Simulierten neu beschrieben werden.

Kapitel 5

Langzeitevaluierung von Niederschlagsvorhersagen

In diesem Kapitel wird die Vorhersagegüte von operationellen Atmosphärenmodellen über einen längeren Zeitraum analysiert und bewertet. Das Kapitel besteht aus zwei Abschnitten. Zuerst werden Verifikationsergebnisse von Tagessummen der QNVs von vier Modellen über eine achtjährige Zeitperiode (2001-2008) vorgestellt. Im zweiten Abschnitt erfolgt die Analyse der Vorhersagegüte für zwei charakteristische konvektive sommerliche Wettersituationen am Beispiel der QNVs eines nichthydrostatischen NWV Modells im Zeitraum von Mai bis Oktober der Jahre 2001 bis 2007.

5.1 Zeitliche Entwicklung der Vorhersagequalität

Nachdem im vorangegangenen Kapitel ein Vergleich der Vorhersageleistung von 19 Modellen für den Sommer 2007 in Südwestdeutschland beschrieben wurde, erfolgt in diesem Kapitel die Analyse der zeitlichen Entwicklung der Qualität täglicher QNVs von vier in der operationellen Wettervorhersage zur Anwendung kommenden Atmosphärenmodellen. Die vier Modelle können in drei Kategorien unterteilt werden und umfassen damit das komplette Spektrum aktueller deterministischer numerischer Wettervorhersagemodelle. Im einzelnen sind dies die in Kapitel 1.1.1 vorgestellten Modelle: das deterministische hydrostatische Globalmodell des ECMWF, das COSMO-EU und das COSMO-7¹ als nichthydrostatische Regionalmodelle mit Konvektionsparametrisierung sowie das konvektionserlaubende COSMO-DE. Aus den 00 UTC gestarteten Vorhersageläufen wur-

¹Schweizer Version des COSMO Modells mit 45 vertikalen Modellschichten sowie ein leicht gegenüber dem COSMO-EU verschobenes Gitter. Die Anfangs- und Randbedingungen werden bei dieser Version durch das ECMWF Modell zur Verfügung gestellt.

NWV Modell	Zeitraum	# Fehltage	in %
COSMO-7	01/2001-12/2008	7	0,2
ECMWF	04/2001-12/2008	0	0
COSMO-EU	10/2003-12/2008	3	0,2
COSMO-DE	01/2007-12/2008	1	0,1

Tabelle 5.1. Vorhandene Zeitperioden und Fehltage der Tagessummen von COSMO-7, ECMWF, COSMO-EU sowie COSMO-DE.

den Tagessummen (7. bis 30. Vorhersagestunde) akkumuliert, außer beim COSMO-DE, bei dem eine Kombination der 00 und 12 UTC zur Berechnung der Tagessummen (7. bis 18. Vorhersagestunde) verwendet wurde. Als Basis der Untersuchungsperiode dient die achtjährige Zeitreihe (2001-2008) der gegitterten täglichen Niederschlagsbeobachtungen (Kapitel 1.1.2) für Deutschland. Zur Durchführung der Verifikation wurden die QNVs aller Modelle auf das Beobachtungsgitter ($\Delta x = 7$ km) transformiert (siehe Kapitel 4). Es liegen nicht für alle Modelle QNVs über den gesamten achtjährigen Zeitraum vor. Aus Tabelle 5.1 kann entnommen werden, ab welchem Zeitpunkt die Daten der Modelle vorliegen sowie der absolute und relative Anteil der Fehltage. Allein das COSMO-7 deckt den kompletten Zeitraum ab. Die Anzahl der fehlenden Tage ist für alle Modelle sehr gering, so dass eine sehr gute Datenlage vorhanden ist.

Zur Bewertung der Güte der QNVs werden verschiedene traditionelle Bewertungsmaße sowie SAL für Deutschland angeschaut. Zuerst erfolgt die Beschreibung der Ergebnisse der kontinuierlichen Maße. Anschließend werden die Resultate der kategorischen Maße vorgestellt und zum Abschluss werden die Zeitreihen der SAL Analyse behandelt und mit den Ergebnissen der traditionellen Maße sowie anderer Studien verglichen.

5.1.1 Traditionelle Bewertungsmaße

Um die Entwicklung der Vorhersagegüte zu analysieren, werden an dieser Stelle die Ergebnisse der traditionellen Bewertungsmaße betrachtet. Dabei werden zuerst die Resultate von *RMSE* sowie Reduktion der Varianz (*RV*) als kontinuierliche Maße und daran anschließend die des Häufigkeitsbias (*FBI*) und des Heidke Skill Scores (*HSS*) als kategorische Maße diskutiert.

Kontinuierliche Maße

In Abbildung 5.1 sind die achtjährigen Zeitreihen des *RMSE* (oben) und der *RV* relativ zum Zufall (unten) der vier Modelle dargestellt. Jeder Punkt in dieser Darstellung gibt den mittleren Wert für einen dreimonatigen Zeitraum an.

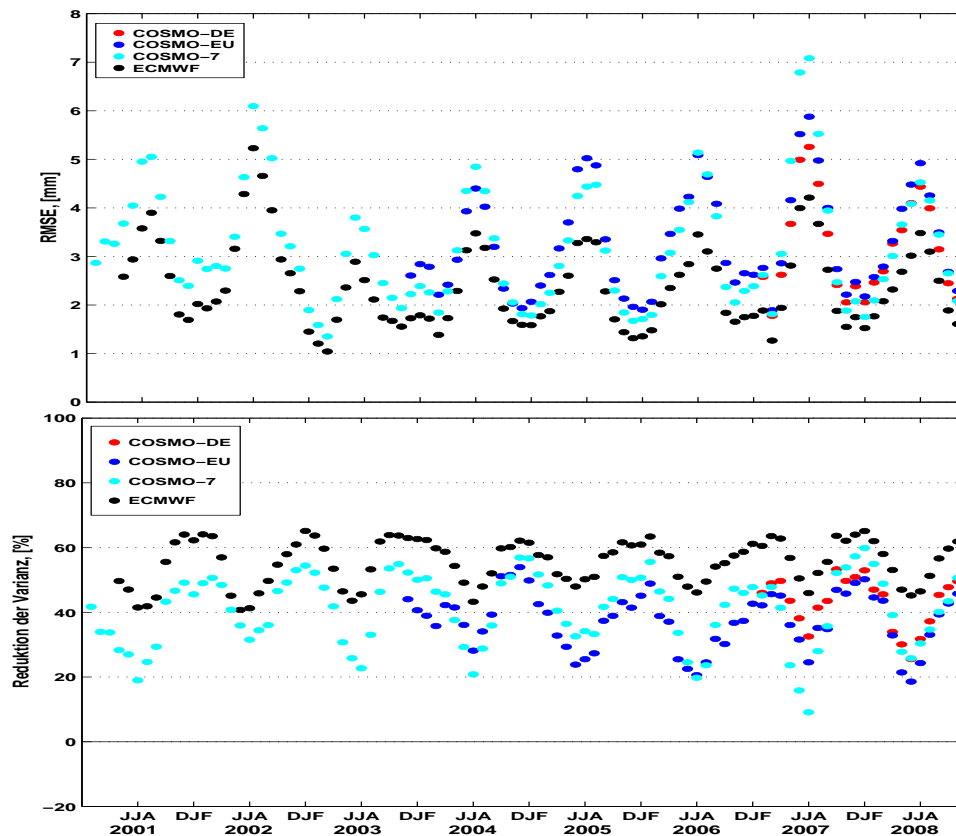


Abbildung 5.1. Zeitreihen (gleitendes Dreimonatsmittel) des RMSE (oben) und der RV relativ zum Zufall (unten) für Tagessummen des COSMO-DE (rot), COSMO-EU (blau), COSMO-7 (cyan) und ECMWF (schwarz) für Deutschland von 2001-2008.

Für alle Modelle kann beim *RMSE* ein deutlicher Jahresgang festgestellt werden, wobei für die Wintermonate die geringsten Werte (ca. 1-2 mm) gefunden werden. Dagegen weisen die Sommermonate bei allen Modellen mit bis zum 7 mm die größten Werte des *RMSE* auf. Allerdings kann eine große und deutliche Variabilität zwischen den einzelnen Jahren, besonders im Sommer, erkannt werden. Die geringsten Werte weist der extrem warme und trockene Sommer 2003 (Schär et al., 2004) auf. Dagegen liefern die Sommer 2002 (Hochwasserereignisse an Donau und Elbe, Rudolf and Rapp, 2002) und 2007 (Wernli et al., 2010) auch aufgrund einer großen positiven Anomalie der Niederschlagsmenge hohe Werte beim *RMSE* mit 6 mm (2002) und 7 mm (2007) für das COSMO-7. Über den gesamten Zeitraum betrachtet ist gemessen am *RMSE*, keine Verbesserung der Vorhersagegüte zu erkennen. Auch ist festzuhalten, dass das Modell mit der größten horizontalen Beschreibung (ECMWF) für annähernd alle Zeitpunkte den geringsten *RMSE* aufweist.

Neben den Ergebnissen des *RMSE* werden die Resultate der Reduktion der Varianz (*RV*) relativ zum Zufall (siehe Kapitel 1.2.1) betrachtet. Als Grundmaß für die *RV* wird der *RMSE* benutzt, so dass die *RV* angibt, um wie viel Prozent der *RMSE* der Modellvorhersagen kleiner ist als der einer Zufallsvorhersage². Ebenso wie beim *RMSE* ist keine eindeutige Verbesserung der Vorhersagegüte innerhalb des Zeitraums zu erkennen. Analog zum *RMSE* ist auch, dass für alle Modelle die Werte im Winter (ca. 60 %) deutlich besser sind als im Sommer (zwischen 20 und 40 %) und das ECMWF Modell die besten Vorhersagen liefert.

Kategorische Maße

Nachdem die zeitliche Entwicklung der Vorhersagegüte anhand von *RMSE* und *RV* beschrieben wurde, werden an dieser Stelle die Ergebnisse von zwei ausgewählten kategorischen Maßen betrachtet. Bei einer kategorischen Betrachtung von Niederschlagsvorhersagen wird überprüft und bewertet, ob der Niederschlag an einem bestimmten Gitterpunkt einen festgelegten Wert überschreitet. In dieser Arbeit werden drei Schwellenwerte (1, 4 und 8 mm pro Tag) bewertet.

Als erstes kategorisches Maß ist in Abbildung 5.2 die Zeitreihe des FBI für Deutschland dargestellt. Ebenso wie im Fall der Darstellung der Ergebnisse der kontinuierlichen Maße stellt jeder Punkt das Mittel für eine dreimonatige Periode dar. Die Ergebnisse der COSMO Modelle zeigen für alle betrachteten Schwellenwerte einen deutlichen jahreszeitlichen Zyklus mit dem Maximum im Winter, was bedeutet, dass im Winter häufiger Ereignisse ($FBI > 1$) vorhergesagt als beobachtet werden. Im Sommer ist das Verhältnis fast ausgeglichen ($FBI \approx 1$). Die Resultate des ECMWF Modells zeigen für die beiden niedrigeren Schwellenwerte (1 und 4 mm/Tag) eine zu häufige Simulation dieser Ereignisse. Für den 1 mm/Tag Schwellenwert ist kein saisonaler Zyklus beim ECMWF Modell zu erkennen, wohingegen für den Schwellenwert 4 mm/Tag ein mit dem der COSMO Modelle vergleichbarer Zyklus festzustellen ist. Im Fall des hohen Schwellenwerts (8 mm/Tag) muss auf jeden Fall beachtet werden, dass die beobachtete Häufigkeit dieser Situationen geringer und deshalb die Interpretation sehr problematisch ist. Die jahreszeitlichen Schwankungen bei den COSMO Modellen haben hier die größte Amplitude im Vergleich mit den beiden anderen Schwellenwerten. Auffällig ist das fehlende Wintermaximum bei COSMO-7 für den Winter 2007/8. Beim ECMWF ist für diesen Schwellenwert eine häufigere Unterschätzung der Ereignisse zu erkennen.

Da der *FBI* keine Aussage über die Vorhersageleistung der Modelle, sondern ausschließlich Informationen über das Verhältnis von vorhergesagten zu beobachteten Ereignissen

²Generiert mit Hilfe des 'Monte-Carlo' Verfahrens basierend auf dem Beobachtungsdatensatz (siehe Wernli et al., 2008).

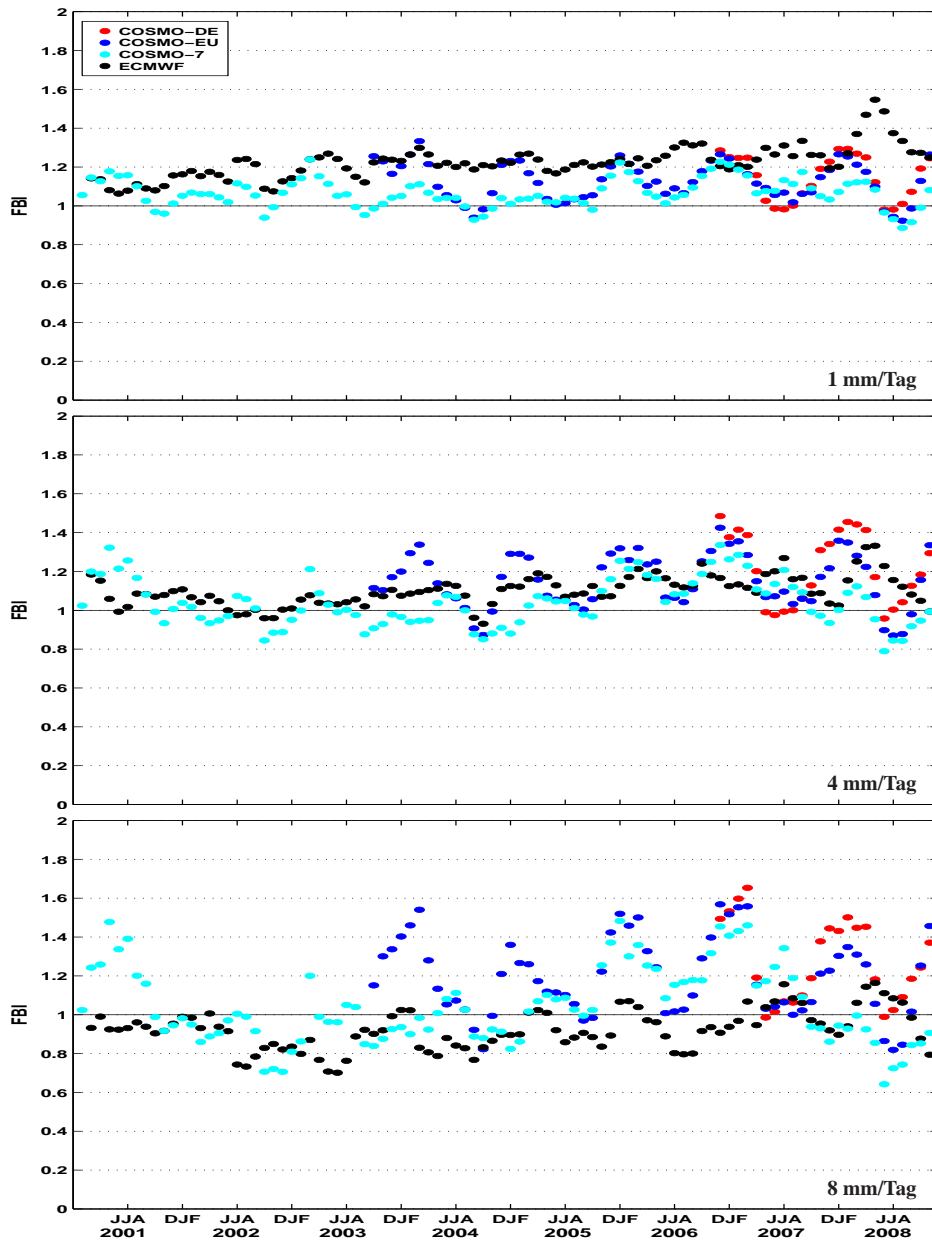


Abbildung 5.2. Zeitreihen (gleitendes Dreimonatsmittel) des Häufigkeitsbias (FBI) für Tagessummen des COSMO-DE (rot), COSMO-EU (blau), COSMO-7 (cyan) und ECMWF (schwarz) für Deutschland im Zeitraum von 2001-2007. Schwellenwerte von oben nach unten: 1 mm/Tag, 4 mm/Tag und 8 mm/Tag.

liefert, werden die Ergebnisse von einem weiteren Maß diskutiert und dies ist der Heidke Skill Score. In Abbildung 5.3 sind die Zeitreihen des *HSS* für die drei Schwellenwerte gezeigt. Je höher der Schwellenwert ist, desto geringer werden die absoluten Werte des *HSS*.

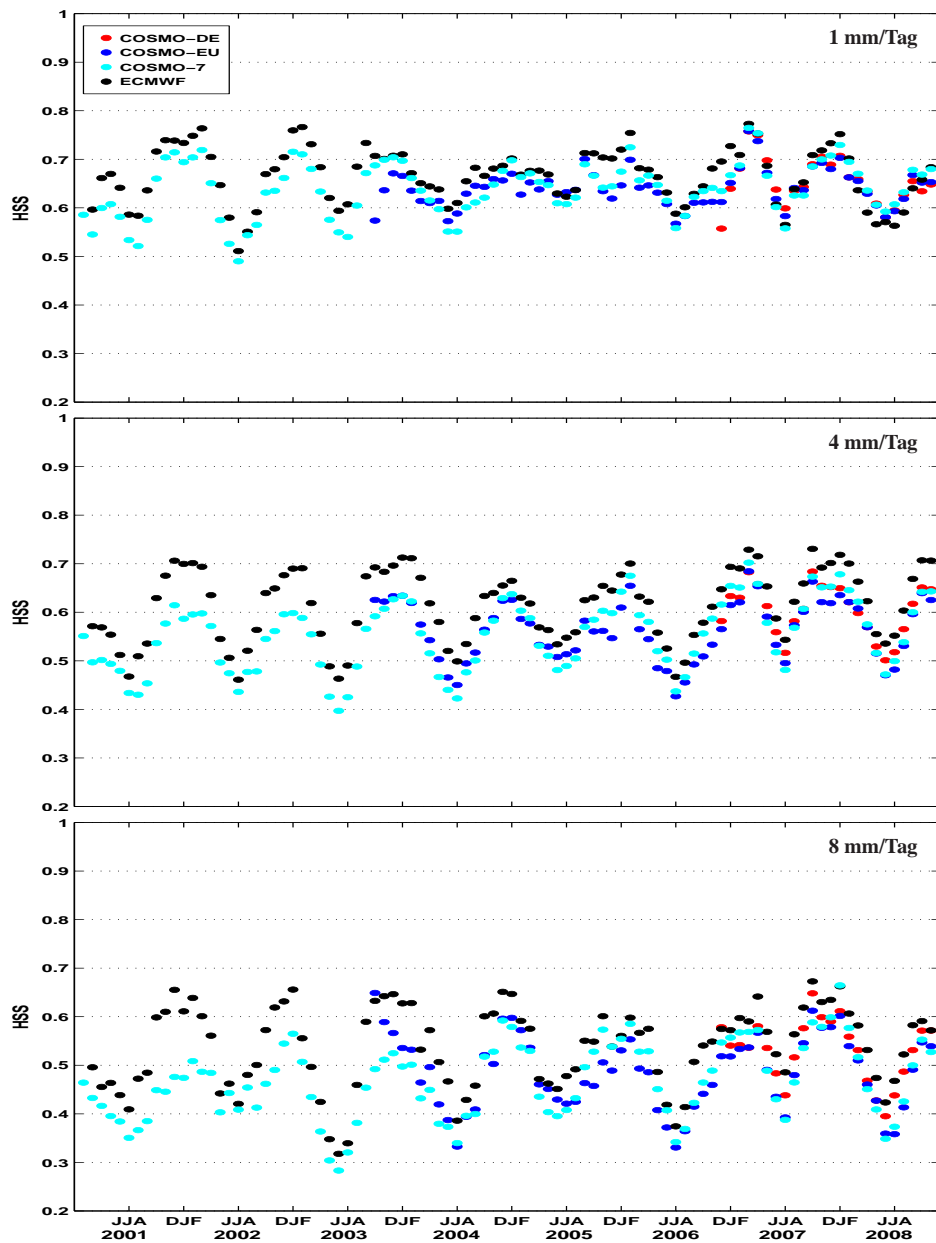


Abbildung 5.3. Wie Abbildung 5.2, aber für den Heidke Skill Score (HSS).

Es kann in der zeitlichen Entwicklung der Vorhersagegüte keine Verbesserung und fast kein Unterschied zwischen den Modellen für alle Schwellenwerte erkannt werden. Eine Ausnahme ist hier der Anfang der Untersuchungsperiode (2001/02), in der das ECMWF im Vergleich zu COSMO-7 deutlich bessere Werte des *HSS* (besonders beim 4 mm/Tag Schwellenwert) aufweist. Analog zu den Ergebnissen beim *FBI* ist ein saisonaler Zyklus

für die COSMO Modelle zu erkennen. Im Gegensatz zum *FBI* zeigt auch das ECMWF bei allen Schwellenwerten diesen Zyklus. Besonders auffällig ist der schwach ausgeprägte Jahresgang zwischen 2003 und 2006 beim geringsten Schwellenwert. Beim geringsten Schwellenwert schneidet der extrem niederschlagsreiche Sommer 2002 am schlechtesten ab, wohingegen beim größten Schwellenwert der trockene Sommer 2003 den kleinsten Wert aufweist. Dies kann allerdings durch den extrem trockenen Sommer und damit einer sehr geringen Anzahl solcher Ereignisse erklärt werden.

Fazit

Als Fazit der Betrachtung der zeitlichen Entwicklung der Vorhersagegüte durch die Ergebnisse einer Auswahl von traditionellen Bewertungsmaßen kann folgendes festgehalten werden:

- Es ist bei keinem der Maße eine allgemeine Verbesserung der Vorhersagegüte festgestellt worden.
- Ein deutlicher saisonaler Zyklus der Güte konnte bei den meisten Maßen erkannt werden und mit den besten Werten im Winter.
- Das ECMWF schneidet bei den verwendeten kontinuierlichen Maßen (*RMSE* und *RV*) am besten ab, wohingegen beim *HSS* dies nur bei höheren Schwellenwerten auftritt.
- Besonders beim *RMSE* ist vor allem im Sommer eine starke Variabilität von Jahr zu Jahr zu beobachten.

Die Resultate der traditionellen Maße für Deutschland könnten den Eindruck erwecken, es gäbe nicht nur keine Verbesserung der Vorhersagegüte in den letzten Jahren, sondern auch, dass das ECMWF Modell mit der größten horizontalen räumlichen Beschreibung die besten Vorhersagen liefert. In einer vergleichbaren Studie von *Oberto and Milelli* (2010) für Italien im Zeitraum beginnend im Winter 2004 bis zum Frühjahr 2009 wurden vier COSMO Modelle (COSMO-I7, COSMO-7, COSMO-EU und COSMO-ME mit einer Gitterweite von 7 km und Konvektionsparametrisierung) bzgl. ihrer Vorhersageleistung untersucht. Es wurden unter anderem die traditionellen Verifikationsmaße wie der *FBI* und der *ETS* für vierundzwanzigstündige Akkumulationen der 00 UTC Läufe benutzt. Der Unterschied dieser Akkumulationen zu denen in der vorliegenden Arbeit verwendeten besteht darin, dass die ersten 24 h anstatt der Periode von 7. bis 30. Stunde verwendet

wurden. Das Ergebnis für den *FBI* zeigt einen starken Jahresgang mit hohen Werten im Sommer und sehr guten Werten im Winter. Er ist damit gegenläufig zu den Ergebnissen in Deutschland. Vermutlich dürfte dies mit dem gegenläufigen Jahresgang des Niederschlags (Italien mit dem Minimum im Sommer und Deutschland mit dem Maximum im Sommer) in beiden Ländern zu erklären sein. Die Werte des $ETS (= HSS / (2 - HSS))$, siehe Kapitel 1.2.1) für Italien zeigen ebenfalls wie die des *HSS* für Deutschland einen starken Jahresgang, der in diesem Fall in Phase ist. Die Tendenz zu besseren Werten des *FBI* bei kleinen Schwellenwerten in Italien, kann für Deutschland für die COSMO Modelle nicht bestätigt werden. Die leichte Verbesserung der Vorhersageleistung für Italien gemessen am *ETS* kann für Deutschland in den letzten Jahren der Untersuchungsperiode (2001-2008) ebenfalls ausgemacht werden, wobei in Italien wie auch in Deutschland der Jahreszyklus deutlich stärker ist als die Verbesserung.

5.1.2 SAL Resultate

Es ist bekannt, dass die höher aufgelösten Modelle bei der Bewertung durch traditionelle Bewertungsverfahren aufgrund des 'Double Penalty'-Problems schlechter abschneiden (siehe Kapitel 1.2.2). Um dies zu lösen, wurden räumliche Bewertungsmaße wie z.B. SAL entwickelt. An dieser Stelle werden die vier Modelle für den selben Zeitraum wie bei den traditionellen Maßen durch Anwendung von SAL analysiert.

Zur besseren Darstellung der SAL Resultate werden bei jedem Modell für Dreimonatsperioden der Median und der Interquartilsabstand (IQR) der drei Komponenten von SAL bestimmt. Dadurch muss nicht eine große Anzahl von SAL Diagrammen abgebildet werden, sondern nur die daraus abgeleiteten statistischen Größen (siehe Abschnitt 3.1.2). In Abbildung 5.4 sind die Zeitreihen für die drei Komponenten Struktur (oben), Amplitude (mitte) und Location (unten) dargestellt. Die vertikale Achse gibt den Median an und der IQR ist farblich in den Symbolen eingezeichnet.

Die Ergebnisse der *S* Komponente zeigen einen deutlichen Unterschied zwischen den COSMO Modellen und dem ECMWF. Das ECMWF Modell weist den höchsten Median (zwischen 0,75 und 1,25) über den gesamten Zeitraum auf. Die hohen positiven Werte des Medians lassen sich damit erklären, dass das Modell im Vergleich zu den Beobachtungen eine zu grobe Auflösung hat und auch nicht die beobachtete Struktur der Niederschlagsverteilung ausreichend gut simulieren kann. Bei den COSMO Modellen kann ein deutlicher Jahresgang beim Median ausgemacht werden mit den besten Werten im Sommer (≈ 0) und den schlechtesten im Winter ($\approx 0,5$). Der IQR von *S* ist im Jahr 2008 bei allen COSMO Modellen am besten, was an den dunkelgrünen Farbtönen zu erkennen

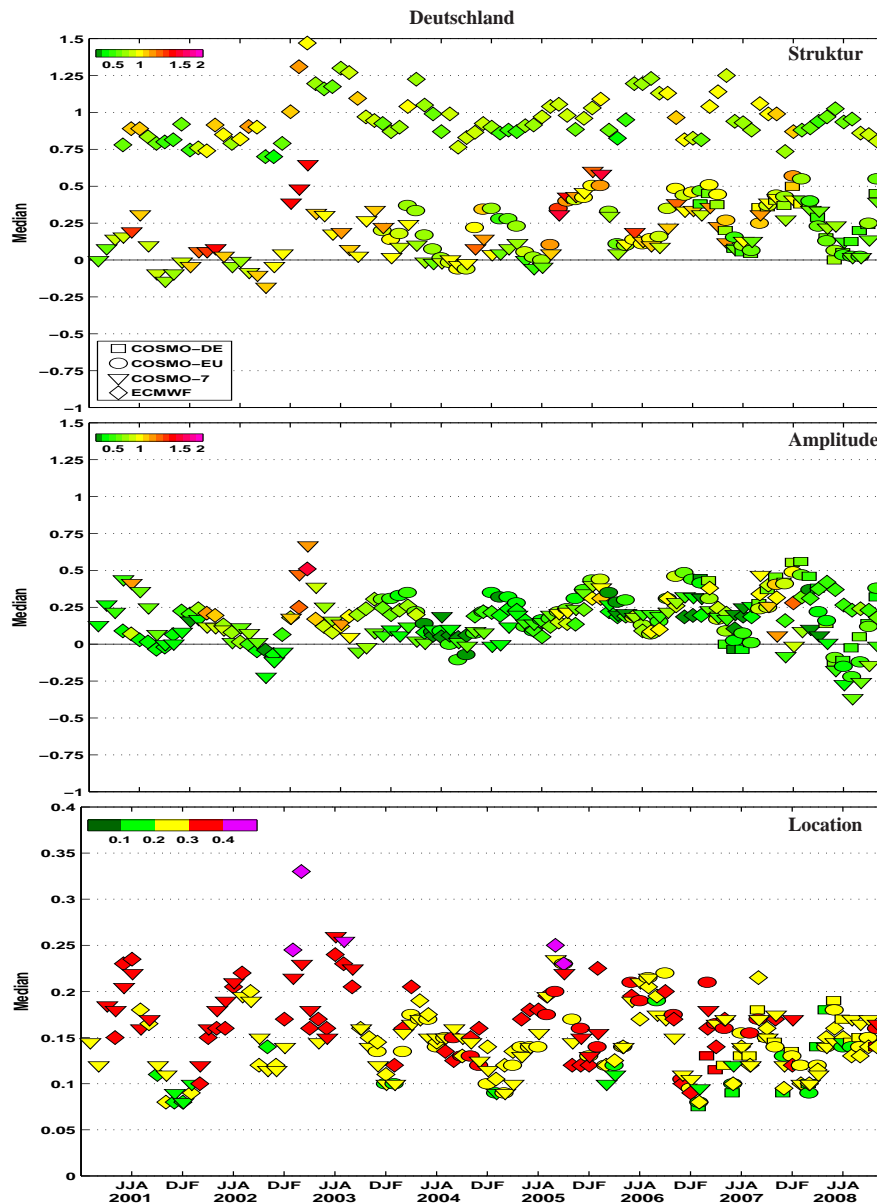


Abbildung 5.4. Zeitreihen (gleitende Dreimonatsmittel) der drei Komponenten von SAL für Tagessummen des COSMO-DE (Quadrate), COSMO-EU (Kreise), COSMO-7 (Dreiecke) und ECMWF (Rauten) für Deutschland im Zeitraum von 2001-2008. Die Komponenten von oben nach unten: Struktur, Amplitude und Location. Auf der vertikalen Achse ist der Median der Komponente aufgetragen und der IQR ist farblich kodiert.

ist. Das häufigere Auftreten der kleineren IQR Werte kann als eine geringe Verbesserung der Qualität der Niederschlagsvorhersage innerhalb des Untersuchungszeitraums im Sinne des Merkmals der Struktur gesehen werden.

Im Gegensatz zur S Komponente ist zwischen allen Modellen bei der A Komponente kein großer Unterschied zu erkennen. Vergleichbar mit den Resultaten der S Komponente ist ein saisonaler Zyklus zu erkennen, der allerdings weniger stark ausgeprägt ist. Insgesamt kann bei der Amplitude, weder im Median noch beim IQR, eine Verbesserung der Vorhersagegenauigkeit festgestellt werden. Die Jahr zu Jahr Variabilität ist sehr deutlich zu sehen. Besonders auffällig ist die deutliche Unterschätzung der Niederschlagsmenge beim COSMO-7 aber auch beim COSMO-EU für den Sommer 2008. Das COSMO-DE zeigt dagegen nur eine leichte Unterschätzung, wohingegen das ECMWF mit einer ausgeprägten Überschätzung zu notieren ist. Die Abweichung des Verlaufs des COSMO-7 und des ECMWF von dem der deutschen COSMO Modelle ist auch schon für den Herbst und Winter 2007/08 zu erkennen.

Als letztes wird die L Komponente betrachtet. Auch für diese Komponente kann ein saisonaler Zyklus erkannt werden, allerdings ist dieser dem von S und A entgegengesetzt. Bei L werden die kleinsten bzw. besten Medianwerte im Winter erzielt und im Sommer die schlechtesten. Die hohen Werte im Sommer können durch das häufige Auftreten konvektiver Situationen erklärt werden, deren korrekte Ortsvorhersage extrem schwierig ist. Am ehesten kann eine merkbare Verbesserung der Vorhersagegüte bei L ausgemacht werden. Hauptsächlich kann dies an den kleineren Werten des IQR festgemacht werden. Vor allem im Sommer kann dies festgestellt werden, wo auch die Medianwerte seit 2003 (JJA 2004 bildet eine Ausnahme) geringer werden. Ebenso wie im Fall der A Komponente ist kaum ein Unterschied zwischen den einzelnen Modellen zu erkennen.

Zum Abschluss der Betrachtung der Ergebnisse der drei SAL Komponenten sei noch darauf hingewiesen, dass das Frühjahr 2003 besonders schlecht für alle drei Komponenten abschneidet. Eine mögliche Erklärung hierfür kann durch die extreme Trockenheit mit sehr geringen absoluten Werten erklärt werden (*Beck et al.*, 2003). Schwache Niederschlagsereignisse werden durch SAL, aufgrund seiner relative Definition, besonders stark bestraft (*Früh et al.*, 2007).

Nachdem die Zeitreihen der Modelle für die drei Komponenten allein beschrieben wurden, wird an dieser Stelle der Versuch unternommen, herauszufinden, wann welches Modell im Sinne von SAL die besten Vorhersagen geliefert hat. Dazu werden die drei SAL Werte zu einem zusammengefasst. Die Vorgehensweise ist identisch mit der in Kapitel 4 für die Rangliste der Modelle bei MAP D-PHASE und ist der Definition der Reduktion der Varianz (RV , siehe Kapitel 1.2.1) angelehnt. Dabei wird für jede Komponente der Gewinn der Modellvorhersage gegenüber der aus Beobachtungen mittels 'Monte-Carlo' Methode generierten Zufallsvorhersage bestimmt. Hierbei wird für jede Komponente zu jeden Zeitpunkt der $RMSE$ bestimmt. Anschließend wird ein Mittelwert für die drei Komponenten

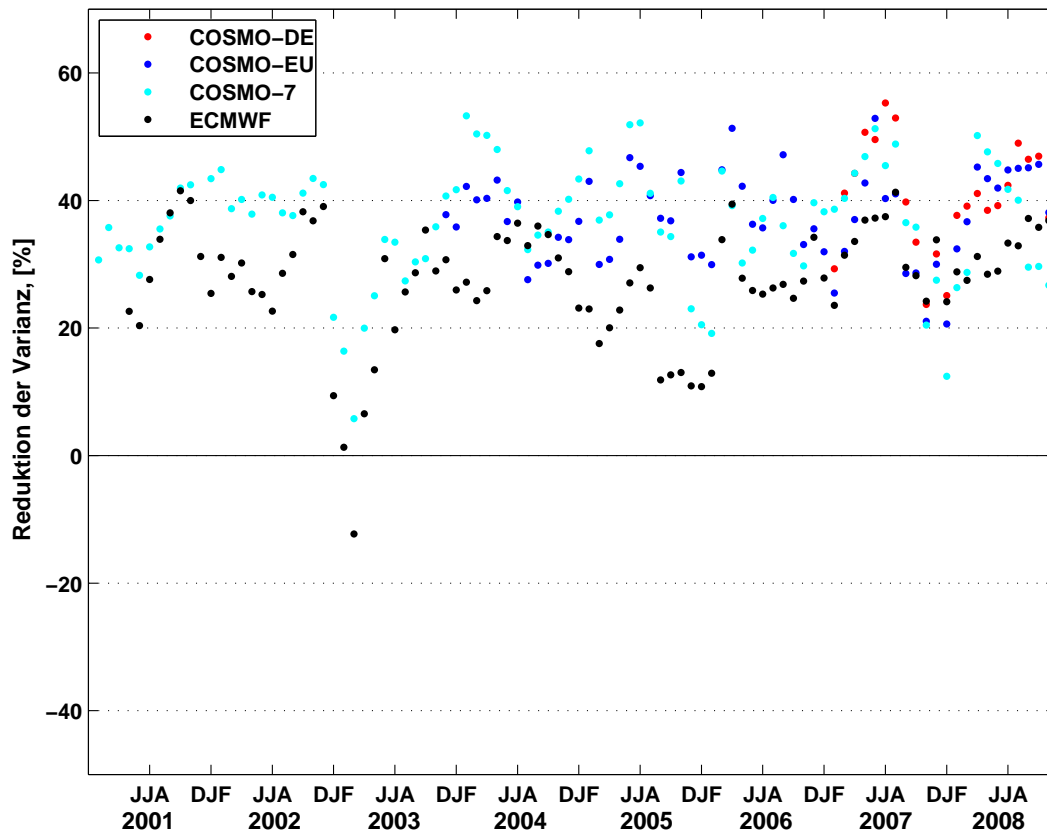


Abbildung 5.5. Zeitreihen der RV_{SAL} (gleitende Dreimonatsmittel) für vier Modelle: COSMO-DE (rot), COSMO-EU (blau), COSMO-7 (cyan) und ECMWF (schwarz) für Deutschland im Zeitraum von 2001-2008.

zusammen berechnet. Im Gegensatz zur Darstellung des IQR gehen in dieses Maß alle verfügbaren Zeitpunkte ein (siehe Kapitel 4).

In Abbildung 5.5 sind die Zeitreihen der RV relativ zum Zufall für SAL (RV_{SAL}) für die vier Modelle gezeigt. Auch für diese Maßzahl ist keine eindeutige Verbesserung der Vorhersagegüte zu erkennen. Es ist eine starke Variabilität zwischen den Modellen aber auch von Jahr zu Jahr zu sehen. Ein saisonaler Zyklus ist kaum bzw. gar nicht zu erkennen. Für die Sommer 2007 und 2008 weist das COSMO-DE die besten Werte auf, allerdings ist der Unterschied zu den anderen Modellen relativ gering. Wie schon bei der Betrachtung der drei einzelnen Komponenten fällt der extrem schlechte Winter/Frühjahr 2002/03 Zeitraum auf, bei dem im Fall des ECMWF (schwarz) es sich zeigt, dass die Zufallsvorhersage bessere Ergebnisse liefert.

5.1.3 Abschließende Bemerkungen

Die Beschreibung der Zeitreihen sowie die Verifikation der QNVs von vier NWV Modellen enthüllen sowohl für traditionellen Maße als auch SAL einige interessante Aspekte. Beim Vergleich von traditionellen und SAL Ergebnissen sind folgende Punkte festzuhalten.

- Anhand der traditionellen Maße konnte keine Verbesserung der Vorhersageleistung festgestellt werden.
- Das Modell (ECMWF) mit der größten Auflösung zeigt die besten Werte ($RMSE$, RV und HSS).
- Mit Hilfe von SAL konnte für die höher aufgelösten COSMO Modelle gezeigt werden, dass sie realistischere Niederschlagsverteilungen simulieren, was an den erheblich besseren Medianwerten von S zu erkennen ist.
- Eine Verbesserung der L Komponente (Median und vor allem bei IQR) im Sommer ist bei allen Modellen zu beobachten.
- Die Jahresgänge von $RMSE$, RV , HSS , S sowie A (Sommer gut Winter schlecht) sind zu L (Winter gut und Sommer schlecht) phasenverschoben.
- Beim RV_{SAL} liefert das COSMO-DE im Sommer die besten Vorhersagen.

Die Tatsache, dass das am höchsten aufgelöste Modell COSMO-DE bei der Betrachtung der Tagessummen keine deutlich bessere Vorhersageleistung im Vergleich zu den größer aufgelösten Modellen liefert, dürfte mit damit begründet werden, dass es als Kurzfristmodell entwickelt wurde. In Kapitel 4 konnte gezeigt werden, dass höher aufgelöste Modelle mit zunehmender Vorhersagezeit (< 12 h) schneller stärker von Null abweichende Werte bei SAL aufweisen.

Das keine klare und deutliche Verbesserung der Vorhersagegüte beim Niederschlag innerhalb dieses Achtjahreszeitraums erzielt wurde, obwohl die Beschreibung der großräumigen Struktur der Atmosphäre (Geopotential) in den acht Jahren immer früher in voraus bestimmt werden kann (siehe Kapitel 1), zeigt die Schwierigkeit der QNV. Deutlich zu erkennen ist in diesen acht Jahren, dass die Leistung der Vorhersagemodelle von Jahr zu Jahr Schwankungen aufweist. Die Bewertung der QNVs mit SAL konnte allerdings Hinweise auf Verbesserungen der Vorhersage während des Achtjahreszeitraums liefern, vor allem bei der Struktur der höher aufgelösten Modelle und in der Simulation des richtigen Ortes der Niederschlagsereignisse im Sommer.

Neben Deutschland wurden auch für die deutschen Abschnitte der Einzugsgebiete von Donau und Elbe die selben Untersuchungen durchgeführt. Diese beiden Regionen stehen exemplarisch für die Flusseinzugsgebiete, da das Einzugsgebiet der Elbe zu den trockensten und das der Donau zu den regenreichsten Gebieten in Deutschland gehören. In beiden Regionen weist die mittlere Niederschlagsmenge im Sommer ein Maximum auf, genauso wie für Deutschland. Das qualitative Bild sieht für beide Gebiete so aus wie für Deutschland auch wenn die Jahr zu Jahr Schwankungen in den Teilgebieten noch ausgeprägter sind.

5.2 Einstufung von konvektiven Niederschlagsereignissen mit Hilfe der konvektiven Zeitskala und deren Vorhersagecharakteristik*

Dieser Abschnitt beschreibt ein Konzept zur Unterscheidung von konvektiven Wettersituationen in zwei Klassen. Neben der Erstellung einer Klimatologie dieser Wettersituationen wird die Güte von QNVs mit Hilfe von SAL für beide Klassen bewertet.

5.2.1 Theoretische Betrachtungen

Viele Eigenschaften einer Kumuluswolke, ihre Vorhersagbarkeit eingeschlossen, ist abhängig von der großskaligen (synoptisch sowie mesoskalig) Umgebung in die sie eingebettet ist. In gewisser Weise kann konvektive Aktivität als durch großskalige Prozesse kontrolliert betrachtet werden und in der Tat ist diese Betrachtung die Grundlage von vielen Konvektionsparametrisierungen in numerischen Atmosphärenmodellen. Allerdings ist diese Kontrolle nur zum Teil gegeben und weist eine hohe zeitliche wie räumliche Variabilität auf (siehe *Arakawa*, 2004 und Referenzen darin).

Die Auslösung und der Lebenszyklus einer Kumuluswolke ist ein unmittelbares Resultat lokaler Prozesse: die bedingte Instabilität im Vertikalprofil, die Abwesenheit einer Inversionsschicht oder anderer konvektionshemmender Faktoren sowie eine variable Grenzschicht in der die Möglichkeit besteht, einen Aufwind zu triggern. Stehen all diese Faktoren zur Verfügung, so wird innerhalb einer halben Stunde eine Wolke entstehen. Sofern

*Die Ergebnisse dieses Abschnitts sind zur Veröffentlichung bei *Geophysical Research Letters* vorbereitet: Zimmer, M., Craig, G., Keil, C. and H.Wernli (2010). Classification of precipitation events with a convective response timescale and their forecasting characteristics.

nichts entgegenwirkt, sorgt die Konvektion dafür, dass die bedingte Instabilität schnell abgebaut wird.

Anders gesagt, es existieren zwei mögliche Szenarien, die Konvektion verhindern können. Zum einem ist dies Abwesenheit von großskaligen Aufsteigen, welches die Troposphäre abkühlen könnte um eine bedingte Instabilität entstehen zu lassen. In diesem Fall dürfte die konvektive verfügbare potentielle Energie (*engl.: Convective Available Potential Energy, CAPE*) schnell erschöpft sein. Zum anderen, wenn die Auslöseprozesse in der Grenzschicht nicht stark genug sind, um die Energiebarriere der Konvektionshemmung (*engl.: Convective Inhibition, CIN*) zu überwinden, so dass selbst bei großem CAPE keine Konvektion auftritt.

Beide Prozesse bedingen möglicherweise das Auftreten von Konvektion und dies führt zu zwei Szenarien des Einflusses großskaliger Prozesse auf die Konvektion. Wenn die Konvektion durch das Vorhandensein von CAPE limitiert ist, so ist ihr Ausmaß (Massenstrom, Niederschlag, etc.) dadurch begrenzt, inwiefern neues CAPE durch die großskalige Strömung erzeugt werden kann. Da einzelne Wolken immer auf lokale Einflüsse reagieren, ist das Ausmaß der Konvektion nur im Mittel eingeschränkt, was zu einem statistischen Gleichgewicht (auch Quasi-Gleichgewicht genannt oder wie in dieser Arbeit ganz einfach Gleichgewicht) führt. Wenn allerdings die Stärke der Konvektion durch eine Wechselwirkung von lokalen Auslöseprozessen und CIN begrenzt ist, dann können unter Umständen große Mengen an CAPE entstehen. In diesem Fall ist kein Grund vorhanden, um von einer direkten Verbindung vom Ausmaß der Konvektion und der großskaligen Strömung auszugehen.

Die beiden Regime, Gleichgewichts- und getriggerte Konvektion, beschreiben in drastischer Form verschiedene Arten der Wechselwirkung zwischen Konvektion und großskaliger Strömung. Von großer Bedeutung ist es, unterscheiden zu können, welche der beiden für eine betrachtete meteorologische Situation die Dominierende ist. Ein möglicher Weg ist, sich CAPE anzuschauen, welches in Gleichgewichtssituationen klein sein sollte (*Emanuel et al., 1994*). Allerdings ist der Wert von CAPE selbst im Gleichgewicht sehr variabel und somit erscheint es unmöglich, einen Schwellenwert zu finden, der beide Regime voneinander trennt. Ein grundlegenderer Ansatz wurde von *Arakawa and Schubert (1974)* vorgestellt, bei dem der Vergleich der Raten der Produktion von bedingter Instabilität durch die großskalige Strömung und der des Abbaus durch Konvektion betrachtet wird. Synoptischskalige Strömungen entwickeln sich auf einer Zeitskala von ungefähr einem Tag, währenddessen Konvektion, bei Fehlen hemmender Faktoren, in der Lage ist, innerhalb einer Stunde CAPE vollständig abzubauen. Ist das der Fall, dann sollte die Rate der Konvektion eng mit dem großskaligen Antrieb verbunden sein. Bei größeren konvektiven Zeitskalen kann davon ausgegangen werden, dass die Konvektion nicht durch groß-

skalige Prozesse bedingt ist und somit auch kein Gleichgewicht existiert. In dieser Arbeit erfolgt die Schätzung dieser Zeitskala für die Beseitigung der bedingten Instabilität durch Konvektion τ_c gemäß folgendem grundsätzlichen Zusammenhang:

$$\tau_c \sim \frac{CAPE}{dCAPE/dt},$$

wobei $CAPE$ definiert ist als

$$CAPE = \int \frac{g}{T_0}(T_a - T)dz,$$

mit T der Umgebungstemperatur, T_a der Temperatur eines aus der Grenzschicht pseudo-adiabatisch gehobenen Luftpakets und T_0 einer konstanten Referenztemperatur.

Entsprechend *Done et al.* (2006) wird angenommen, dass $CAPE$ durch die Hinzugabe von ausreichend Wärme abgebaut werden kann, um so die Differenz zwischen T and T_a innerhalb der Luftsäule zu eliminieren. Die vertikal integrierte freigesetzte latente Wärme kann mit Hilfe der Niederschlagsrate P ($\text{kg s}^{-1} \text{m}^2$) bestimmt werden:

$$L_v P = \int \rho C_p \frac{dT}{dt} dz,$$

so dass

$$\frac{dCAPE}{dt} = \frac{L_v}{C_p} \frac{g}{\rho T_0} P.$$

Die konvektive Zeitskala τ_c kann damit berechnet werden als

$$\tau_c = \frac{1}{2} \frac{C_p \rho T_0}{L_v} \frac{CAPE}{P}. \quad (5.1)$$

Der Faktor 1/2 wird verwendet weil diese Berechnung die konvektive Modifikation der Grenzschicht außer acht lässt und somit die konvektive Zeitskala stark überschätzt wird. Die Verwendung des Wertes von 1/2 entspricht der Annahme, dass troposphärisches Erwärmen und bodennahes Abkühlen (und Austrocknen) zu gleichen Teilen zur Verringerung von $CAPE$ beitragen (z.B. *Betts*, 1986 oder *Betts and Miller*, 1986).

Getriggerte Konvektion ist durch große Werte der konvektiven Zeitskala charakterisiert. Im Fall, der in *Done et al.* (2006) untersucht wurde, konnten Werte im Bereich von 10 bis 100 Stunden bestimmt werden. Außerdem konnte herausgefunden werden, dass die Werte für τ_c kleiner wurden, wenn das Niederschlagsereignis länger andauerte und $CAPE$ verringerte sich sobald Konvektion ausgelöst wurde.

Im Fall von Gleichgewichtskonvektion sollten die Werte von τ_c in der Größenordnung von einer Stunde liegen, also der Rate wie die Konvektion die Umgebung modifiziert. Der Wert sollte nicht mit der Zeit großartig ändern, aber er ist bedingt durch den Einfluss der großskaligen Strömung auf die Konvektion (*Cohen and Craig, 2004*). Für Gleichgewichts- wie auch Nichtgleichgewichtsereignisse ist ein kontinuierlicher Wertebereich für τ_c möglich, so dass irgendein Schwellenwert mit der Größe von einigen Stunden ausreichen sollte, um konvektive Ereignisse in zwei Kategorien zu unterteilen.

Die Anwendung dieser geschätzten konvektiven Zeitskala als Unterscheidungskriterium in zwei Arten der Kontrolle der Konvektion durch die großskalige Strömung konnte in mehreren Studien deutlich aufgezeigt werden. In *Done et al. (2010)* konnte eine unterschiedliche Sensitivität auf Veränderungen der Konvektionsparametrisierung in mesoskaligen NWV Modellen für die beiden Regime herausgefunden werden. Die Publikation von *Craig et al. (2010)* zeigt, dass die Abhängigkeit des zeitlichen Einflusses von assimilierten Radarreflektivitäten proportional zu τ_c ist. Und schließlich konnte in *Molini et al. (2010)* gezeigt werden, dass große bzw. kleine Werte von τ_c , die mit Gleichgewichts- bzw. Nichtgleichgewichtskonvektion verbunden sind, einer unterschiedlichen beobachteten Struktur der Niederschlagsereignisse entsprechen.

Alle aufgezählten Studien nutzen entweder Modell- oder Reanalysedaten zur Bestimmung der konvektiven Zeitskala. Das erste Ziel der vorliegenden Analyse ist die Berechnung von τ_c aus Beobachtungsdaten, um damit einen ersten Anhaltspunkt für die relative Häufigkeit von Gleichgewichts- und Nichtgleichgewichtskonvektion in der realen Atmosphäre zu erhalten. Um dies zu erreichen, wird ein siebenjähriger Datensatz sommerlicher Radiosondenaufstiege über Deutschland verwendet und in Kombination mit einem hoch aufgelösten Niederschlagsdatensatz kann τ_c berechnet werden. Das zweite Ziel dieser Untersuchung ist das Aufzeigen der Nützlichkeit der Unterteilung in Gleichgewichts- und Nichtgleichgewichtsereignisse für die Verifikation von Niederschlagsvorhersagen.

5.2.2 Daten und Methodik

Radiosondendaten

Für die vorliegende Untersuchung wurde eine siebenjährige Periode (2001-2007, Mai bis Oktober) für folgende sieben Radiosondenstandorte in Deutschland (von Nord nach Süd) ausgewertet.

- Bergen (WMO-Nr. 10238, 9,93°O, 52,82°N, 77 m über NN),
- Lindenberg (10393, 14,12°O, 52,22°N, 115 m),

- Essen (10410, 6,97°O, 51,40°N, 153 m),
- Meiningen (10548, 10,38°O, 50,57°N, 453 m),
- Idar Oberstein (10618, 7,33°O, 49,70°N, 377 m),
- Stuttgart (10739, 9,20°O, 48,83°N, 315 m),
- Oberschleissheim (10868, 11,55°O, 48,25°N, 489 m)

An den meisten Standorten wurden vier Aufstiege pro Tag (um 00, 06, 12 und 18 UTC) durchgeführt. Die Daten wurden über das kostenfreie Internetportal der Universität von Wyoming (<http://weather.uwyo.edu/upperair/sounding.html>) heruntergeladen.

CAPE Berechnung

Die Berechnung von CAPE erfolgte aus den Radiosondenaufstiegen sowie der Annahme eines pseudoadiabatisch aus der Grenzschicht gehobenen Luftpakets, wobei nur der Phasenübergang von gasförmigen zu flüssigen Wasser betrachtet wird. Die Anfangswerte des Luftpakets (Temperatur und Luftfeuchte) werden durch Mittlung über die untersten 100 hPa des Radiosondenaufstiegs bestimmt. Zur Berechnung des vertikalen Profils sowie von CAPE wurden die Routinen aus *Früh and Wirth (2007)* verwendet.

Niederschlagsbeobachtungen

Die hoch aufgelösten stündlichen Niederschlagsbeobachtungen für Deutschland (siehe Kapitel 1.1.2) sind eine Kombination aus Niederschlagsmessern und Radarinformationen. Sie liegen auf einem Gitter mit 7 km Maschenweite vor. Für die vorliegende Analyse wurden dreistündige Akkumulationen ($\pm 1,5$ h am Beobachtungszeitpunkt) innerhalb eines Kreises mit einem Radius vom 50 km um die Radiosondenstandorte bestimmt. Für die Berechnung von τ_c ist es erforderlich die Niederschlagssummen in Niederschlagsraten zu transformieren.

Modellbeschreibung

Die verwendeten Modelldaten stammen von der schweizer Version des nichthydrostatischen Modells des COSMO Konsortiums (www.cosmo-model.org) und ist das Gegenstück zum früher „Lokalmodell“ genannten deutschen Modell des DWD (*Steppeler et al.*,

2003). Die horizontale Gitterweite beträgt 7 km und die Konvektion ist durch ein Massenflussschema nach *Tiedtke* (1989) parametrisiert. Für diese Studie wurden dreistündige Akkumulationen um die Zeitpunkte der Radiosondenaufstiege innerhalb eines Kreises (Radius = 50 km) mit den Standort im Zentrum verwandt. Die Untersuchungen wurden mit Hilfe von operationellen Vorhersagen der 00 UTC gestarteten Läufe des COSMO-7 durchgeführt.

5.2.3 Resultate

In diesem Abschnitt werden die Ergebnisse der Analyse beschrieben. Zuerst erfolgt die Vorstellung der Klimatologie der konvektiven Zeitskala und im zweiten Teil die Verwendung dieser Größe zur Kategorisierung zweier Wettersituationen, um die Qualität der QNVs zu quantifizieren.

Klimatologie der konvektiven Zeitskala

Die erste Frage die beantwortet werden soll, ist die nach der relativen Häufigkeit des Auftretens von Gleichgewichtskonvektion, bei der die Erzeugung bedingter Instabilität durch die großskalige Strömung dominiert ist. Wie in der Einleitung ausgeführt sind dies die Fälle bei denen Werte von τ_c erwartet werden, die deutlich kürzer sind als die Skala der Entwicklung großräumiger Dynamik. Diese Skala besitzt eine Größenordnung von einem Tag oder mehr. Bisherige Studien haben Werte von τ_c kleiner als drei Stunden (*Craig et al.*, 2010) oder sechs Stunden (*Molini et al.*, 2010) als Indikatoren für Gleichgewichtskonvektion herausgefunden.

Aus der Tabelle 5.2 kann entnommen werden, dass 48 % der konvektiven Ereignisse ein τ_c kleiner als drei Stunden und 56 % weniger als sechs Stunden besitzen, so dass dies als Hinweis darauf zu verstehen ist, dass ungefähr die Hälfte der sommerlichen konvektiven Niederschlagsituationen im Gleichgewicht stattfinden. In Tabelle 5.2 sind die 5410

τ_c , [h]		> 1	> 3	> 6	> 12	> 24
MJJASO, [%]	n=4790	64,7	51,5	44,2	37,5	30,9
JJA, [%]	n=3264	68,8	55,1	48,0	41,0	33,4
MSO, [%]	n=1526	55,9	43,8	35,9	30,1	25,4

Tabelle 5.2. Anzahl von Ereignissen und deren relative Häufigkeit bei denen τ_c den betrachteten Schwellenwert übertrifft.

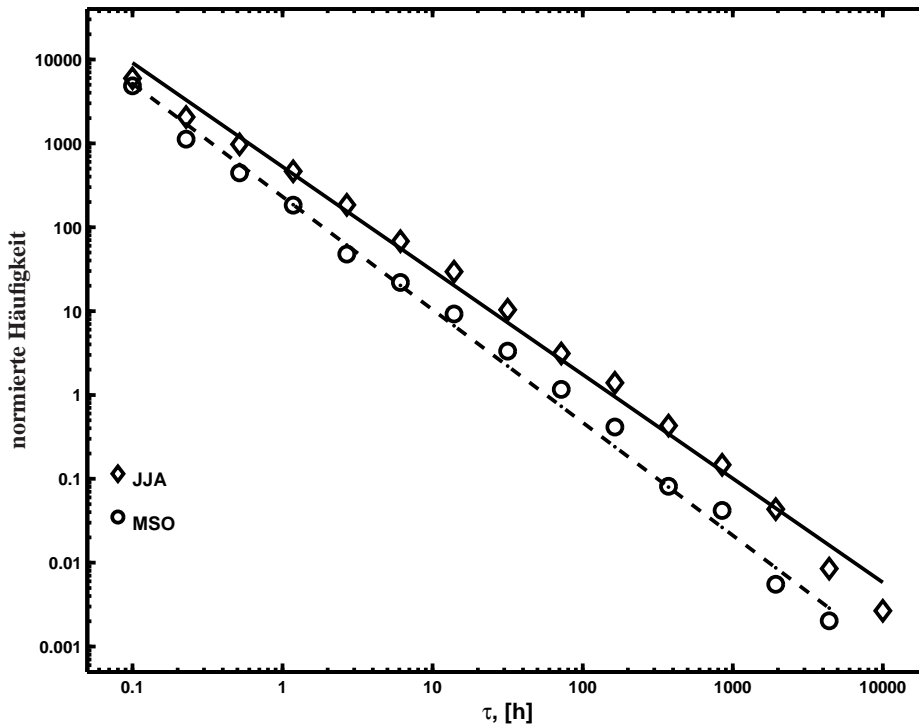


Abbildung 5.6. Doppeltlogarithmisches Diagramm der normierten Häufigkeit von τ_c : Sommer (JJA) als Rauten sowie Mai, September und Oktober (MSO) als offene Kreise. Die Regressionsgerade für JJA (durchgezogen) hat eine Steigung von $-1,24$ und für MSO (strich-punktiert) eine von $-1,35$.

Radiosondenaufstiege nicht enthalten, bei denen die konvektive Zeitskala gleich null ist. Dies ist dann der Fall, wenn CAPE gleich null ist und somit das Niederschlagsereignis vermutlich nicht konvektiver Natur ist.

In Abbildung 5.6 ist die Häufigkeit des Auftretens der Werte von τ_c für den gesamten Datensatz, unterteilt in Sommer (Juni, Juli und August) und den Übergangsmonaten (Mai, September und Oktober), dargestellt. Die Verteilung folgt annähernd einem Potenzgesetz mit einem Exponenten von ungefähr $-1,3$. Es existiert keine klare Grenze zwischen beiden Bereichen. Eine mögliche Grenze von τ_c kann bei ungefähr 10 Stunden ausgemacht werden, mit einer flacheren Steigung bei geringeren und größerer Steigung für größeren Werten von τ_c . Eine scharfe Abgrenzung ist auch nicht zu erwarten gewesen, da viele verschiedene meteorologische Situationen mit unterschiedlichen synoptischen Zeitskalen kombiniert wurden. Außerdem ist es interessant zu spekulieren, ob Regionen mit großem bzw. kleinem τ_c vermischt sind innerhalb des Untersuchungsareal. Dies ist im Bereich des Übergangs von einem ins andere Regime zu erwarten (*Peters and Neelin, 2006*). In

diesem Fall würde der τ_c Wert wiedergeben, welche der beiden Phasen die Dominierende ist.

Häufigkeitsverteilungen von τ_c wurden ebenfalls für jede Station einzeln sowie für die unterschiedlichen Tageszeitpunkte angefertigt und es konnte kein deutlicher Unterschied zu den Ergebnissen die in Abbildung 5.6 dargestellt sind, beobachtet werden. Der Vergleich zwischen Sommer und den Übergangsmonaten ist dagegen interessant. Beide Kurven folgen einem Potenzgesetz, wobei die Steigung im Sommer kleiner ist. Damit verbunden ist im Verhältnis zu den Übergangsmonaten ein Auftreten von mehr konvektiven Situationen mit großen Werten von τ_c und somit mehr Nichtgleichgewichtseignissen.

Dadurch, dass die konvektive Zeitskala als Verhältnis von CAPE und Niederschlagsrate definiert ist, ist es von Interesse den Beitrag der beiden Faktoren zu betrachten. Streudiagramme von τ_c gegen Niederschlagsrate und CAPE weisen eine deutliche Korrelation auf (nicht dargestellt), so dass beide zur Variabilität der Zeitskala beitragen. Dagegen zeigten die Niederschlagsrate und CAPE keine Korrelation und somit kann davon ausgegangen werden, dass beide Faktoren unabhängige Informationen zum Ergebnis beitragen.

Verifikation von COSMO-Vorhersagen

Als ein Anwendungsbeispiel für das Konzept der konvektiven Zeitskala dienen QNVs des COSMO-7 Modells. Die dreistündlich akkumulierten QNVs um 12 UTC der Sommermonate (JJA) wurden mit Hilfe von SAL verifiziert (siehe Abschnitt 3.1.1). Für die statistische Analyse der Vorhersagegüte erfolgte eine Unterteilung der SAL Ergebnisse unter Verwendung von $\tau_c = 12$ h als Schwellenwert. Die erste Klasse enthält alle Ereignisse mit kurzer konvektiver Reaktionszeit ($\tau_c \leq 12$ h) und die andere Klasse alle Situationen mit großen Werten von τ_c . Die Anzahl der Ereignisse in der ersten Klasse ist ungefähr doppelt so groß wie die in der zweiten Klasse.

Die Abbildung 5.7 enthält die SAL Diagramme (siehe Kapitel 3.1) der beiden Regime in den Teilbildern (a) und (b) für alle Standorte außer Stuttgart. Die SAL Ergebnisse für große Werte von τ_c am Standort Stuttgart sind separat dargestellt (Abb. 5.7c). Darüber hinaus sind die Ergebnisse für alle Standorte außer Stuttgart für Situationen bei denen CAPE gleich null ist in Abbildung 5.7d gezeigt. Die gestrichelten Linien geben den Median von S und A an und die graue Box beschreibt den Bereich des 25 %- und 75 %-Perzentsils von beiden Komponenten.

Im Fall der kurzen konvektiven Reaktionszeit (Abb. 5.7a) befinden sich die meisten Punkte im ersten (oben rechts) und im dritten (unten links) Quadranten des Diagramms und liegen darüber hinaus entlang der Diagonalen. Das COSMO-7 überschätzt (unterschätzt) die Niederschlagsmenge bei gleichzeitiger Simulation zu großer (kleiner) und/oder flä-

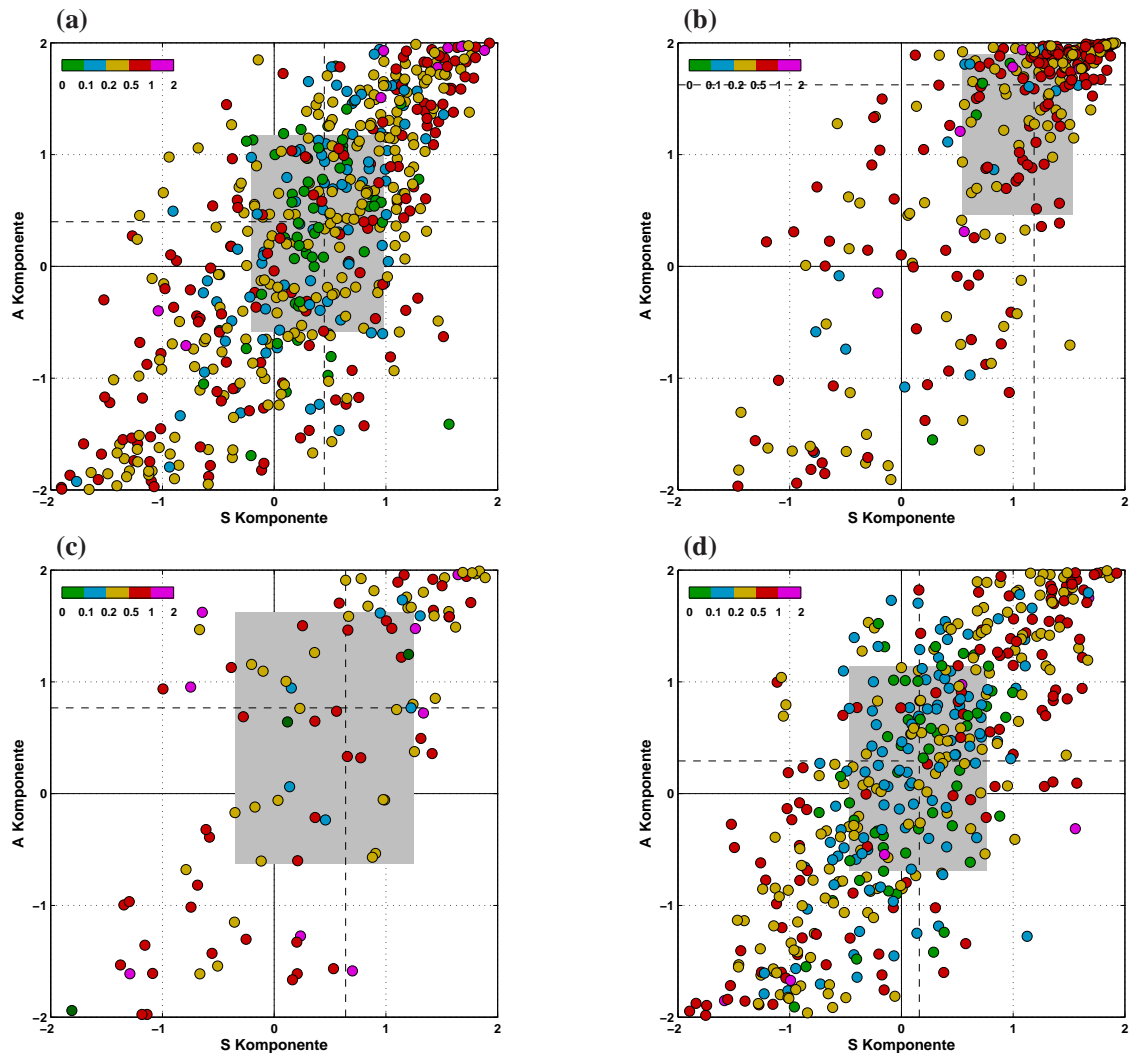


Abbildung 5.7. SAL Diagramme für 3-h akkumulierte QNVs des COSMO-7 für Sommer (JJA) für den Beobachtungszeitpunkt 12 UTC für drei Wetterregime: (a) kleine Werte von τ_c (≤ 12 h), (b) hohe τ_c (> 12 h), und (d) CAPE = 0 für alle Standorte außer Stuttgart. In (c) sind die Ergebnisse für $\tau_c > 12$ h des Standorts Stuttgart gezeigt.

chiger (spitzer) Niederschlagsobjekte. Die Medianwerte von S und A sind positiv (ca. 0,4) und der IQR beträgt 1,8 für A und 1,2 für S . Diese Werte des IQR sind relativ hoch und zeigen damit an, dass die QNVs häufig in mindestens einer der beiden Komponenten schlecht abschneiden. Völlig anders sieht das Bild der QNVs des COSMO-7 im Fall für große Werte von τ_c aus (Abb. 5.7b). Die meisten Vorhersagen sind durch sehr hohe positive Werte von S und A charakterisiert, die durch die hohen Medianwerte von S (1,2) und A (1,8) belegt werden. Der IQR dagegen weist kaum einen Unterschied zwischen

beiden Regimes auf. Mit Hilfe von SAL konnte die unterschiedliche Vorhersagequalität des COSMO-7 für die beiden Regime eindrücklich gezeigt werden. Ein Spezialfall stellt Stuttgart dar (Abb. 5.7c), bei dem der Unterschied zwischen beiden Klassen nicht so deutlich zu Tage tritt, wie im Fall der anderen Stationen. Die Medianwerte von $S(0,7)$ und $A(0,8)$ sind wesentlich geringer für große Werte von τ_c und geringfügig erhöht in der Situationen mit kleinen τ_c Werten gegenüber den Resultaten der anderen Stationen.

Die nichtkonvektiven Niederschlagsereignisse, bei denen CAPE gleich null ist (Abb. 5.7), wurden ebenfalls untersucht. Die Ergebnisse sind vergleichbar mit denen der Situationen mit kleinem τ_c . Die Medianwerte von $S(0,2)$ und $A(0,3)$ sind etwas kleiner und die Werte des IQRs sind für $S(1,2)$ und $A(1,8)$ identisch.

5.2.4 Schlussfolgerungen

Zwei Fragen wurde in dieser Studie nachgegangen. Zum einem wurde zum ersten Mal eine Klimatologie der konvektiven Zeitskala aus Beobachtungen abgeleitet. Zum anderen war dies die Verwendung der konvektiven Zeitskala als Unterteilungskriterium in zwei meteorologische Wetterregime (Gleichgewichts- und Nichtgleichgewichtskonvektion) um die Güte von QNVs in diesen Regimes zu beurteilen. Die wichtigsten Ergebnisse können folgendermaßen zusammengefasst werden:

- Eine physikalisch sinnvolle Unterscheidung zwischen Gleichgewichts- und Nichtgleichgewichtskonvektion kann mit Hilfe der konvektiven Zeitskala τ_c , bestimmt durch das Verhältnis von CAPE und Niederschlagsrate, erstellt werden.
- Die beiden Regime weisen eine deutlich unterschiedliche Wechselwirkung zwischen der großskaligen Strömung (führt zum Aufbau von CAPE) und Konvektion (verbunden mit einer Verminderung von CAPE) auf.
- Dies ist die erste statistische Analyse von τ_c , die ausschließlich auf Beobachtungsdaten (CAPE aus Radiosondenaufstiegen bestimmt, Niederschlag aus einem kombinierten Produkt von Radar- und Stationsbeobachtungen) für sieben Sommer und für sieben deutsche Radiosondenstandorte beruht.
- Es gibt ungefähr ein Verhältnis von 2:1 bei der Anzahl von Ereignissen von Gleichgewichts- zu Nichtgleichgewichtskonvektion (d.h. τ_c kleiner bzw. größer ca. 10 Stunden) innerhalb der untersuchten Gebiete und Jahreszeiten.
- Es konnte mit SAL eine deutlich unterschiedliche Qualität der QNVs für das nicht-hydrostatische COSMO-7 für beide Regime gezeigt werden, mit großer Streuung

im Fall der Gleichgewichtskonvektion und einem großen positiven (zu viel Niederschlag und zu große Objekte) systematischen Fehler in Nichtgleichgewichtssituationen.

Kapitel 6

Zusammenfassung und Ausblick

Die in dieser Arbeit durchgeführten Analysen hoch aufgelöster Niederschlagsvorhersagen mit der merkmalsbezogenen Verifikationsmethode SAL hat einige interessante Ergebnisse zu Tage gefördert. Es konnten verschiedene Eigenschaften bzw. charakteristische Fehler der numerischen Wettervorhersage aufgezeigt werden. Im Folgenden werden die wichtigsten Ergebnisse dieser Arbeit zusammengefasst. Danach werden einige Vorschläge und Ideen zu weiteren Forschungsaktivitäten auf dem Gebiet der merkmalsbezogenen Verifikation vorgestellt.

6.1 Zusammenfassung

Als Niederschlagsbeobachtungen für die meisten Untersuchungen, die in dieser Arbeit vorgestellt wurden, diente der stündliche disaggregierte Datensatz (Kapitel 1.1.2). Im Rahmen dieser Arbeit wurde der Datensatz aus *Paulat* (2007) um vier Jahre auf jetzt acht Jahre (2001-2008) erweitert, so dass damit ein Vergleich der Vorhersagegüte über einen längeren Zeitraum mit dem selben Beobachtungsdatensatz möglich war. Zum einen konnte damit die räumliche Verteilung der saisonalen Niederschlagssummen für die Modelle COSMO-EU und COSMO-DE gegenüber der Beobachtung für einen Zeitraum von zwei Jahren direkt miteinander verglichen werden, mit dem Ergebnis, dass die räumliche Verteilung des Niederschlags beim COSMO-DE besser und die Luv/Lee-Problematik erheblich geringer ist. Zum anderen konnte die Realisierung des mittleren Niederschlagstagesgangs der beiden Modelle mit Beobachtungen verglichen werden. Das Resultat dieser Analyse ist, dass das COSMO-EU ein gegenüber den Beobachtungen zu stark ausgeprägtes Maximum simuliert, welches ca. drei Stunden zu früh auftritt. Dagegen zeigt das COSMO-DE ein zu schwach ausgeprägtes Maximum, dass ca. eine Stunde zu spät im Vergleich zum Beobachteten vorhergesagt wird.

Um quantitative Aussagen über Fehlercharakteristiken zu liefern, wurden die QNVs unterschiedlichster Modelle, die alle momentan operationell verwendeten Typen umschließen, durch die objekt-orientierte Methode SAL, aber auch mit traditionellen Maßen zum Vergleich, verifiziert. Im Rahmen des Vergleichsprojektes ICP (Kapitel 3.2) wurden die speziellen Eigenschaften von SAL anhand von idealisierten, manipulierten und ausgewählten realen Fällen diskutiert. Folgende neue Resultate konnten herausgearbeitet werden:

- Neue Bestimmung des Objektschwellenwerts auf Basis des 95 % Perzentils anstatt des Maximalwerts der Niederschlagsverteilung, liefert robustere Resultate.
- Die Subtraktion eines konstanten Werts kann zu einer völlig anderen Struktur der Niederschlagsverteilung führen, was durch S beschrieben werden kann.
- Der Vergleich mit den Ergebnissen traditioneller Maße führt zu teilweise anderen aber zielführenden Resultaten.

Der im letzten Punkt beschriebene Unterschied der Resultate von traditionellen Maßen zu denen von SAL konnte auch bei der Langzeituntersuchung der Vorhersagegüte (Kapitel 5.1) für die vier Modelle festgestellt werden. Hier schneidet das ECMWF Modell mit der größten Auflösung am besten (besonders beim $RMSE$) ab. Allerdings ist eine erhebliche Variabilität von Jahr zu Jahr zu erkennen. Die Resultate von SAL zeigen ein deutlich differenzierteres Bild für die Entwicklung der Vorhersagegüte auf. Für S konnte gezeigt werden, dass das ECMWF deutlich schlechter abschneidet als die COSMO-Modelle, was an drastisch positiven Medianwerten zu erkennen ist. Bei A ist kaum ein Unterschied zwischen den Modellen festzustellen, außer für das Jahr 2008. Eine Verbesserung über den kompletten Zeitraum ist für S und A nicht erkennbar. Für L sieht das Bild ein wenig anders aus. Hier kann besonders im Sommer eine Verbesserung für diese Eigenschaft anhand von geringeren Werten des Medianes aber auch des IQRs für alle betrachteten Modelle konstatiert werden. Die zusammenfassende Darstellung von SAL durch eine Zahl in Form eines 'Skill Scores' liefert die Information, dass die QNVs des hoch aufgelösten COSMO-DE, besonders im Sommer, am besten sind. Alles in allem bleibt festzuhalten, dass es eine starke Variabilität von Jahr zu Jahr gibt. Alle Maße weisen einen zum Teil recht deutlichen Jahresgang auf, wobei im Winter für die traditionellen Maße $RMSE$, RV , FBI und HSS die Werte am besten sind und im Sommer am schlechtesten. Bei SAL ist ebenfalls ein Jahresgang zu sehen. Hier sind die Ergebnisse für S und A im Sommer am besten, bei L dagegen im Winter. Die schlechten Werte im Sommer von L und der traditionellen Maße weisen deutlich darauf hin, dass in dieser Jahreszeit das 'Double Penalty'-Problem am stärksten ist.

Neben einer Langzeituntersuchung über acht Jahre konnte die Möglichkeit ergriffen werden für Sommer 2007 eine Zahl von 19 Modellen hinsichtlich ihrer Vorhersagegüte miteinander zu vergleichen (Kapitel 4). Dazu diente der Datensatz des MAP D-PHASE Projekts. Das Untersuchungsgebiet befand sich in Südwestdeutschland. Die Resultate lassen sich wie folgt zusammenfassen:

- Realistischere Simulation der Struktur durch die hoch aufgelösten Modelle (kleinerer IQR), wobei die Tendenz zu kleinflächige Objekte zu simulieren vorhanden ist.
- Bei der A Komponente weisen die höher aufgelösten Modelle eines Modellpaares immer geringere Werte auf mit der Konsequenz, dass die Menge unterschätzt wird. Außerdem liegen die meisten Modelle mit Konvektionsparametrisierung relativ dicht beieinander, wohingegen die hoch aufgelösten Modelle eine größere Streuung aufweisen.
- Die Ortskomponente ist die am schlechtesten simulierte Eigenschaft der Niederschlagsverteilung, die durch SAL bewertet wurde.
- Anhand des Maßes RV_{SAL} konnte gezeigt werden, dass zum einen die höher aufgelösten Modelle bessere Vorhersagen liefern und zum anderen, dass sie mit fortschreitender Vorhersagezeit eher schlechtere Werte liefern.
- Bei Hochdrucksituationen zeigen die Modellkategorien eine völlig unterschiedliche Vorhersageleistung: Vormittags sind die hoch aufgelösten Modelle einigermaßen in der Lage die Niederschlagsereignisse zu simulieren aber scheitern nachmittags vollkommen ($S, A \ll 0$). Die Modelle mit Konvektionsparametrisierung weisen vormittags eine erhebliche Überschätzung ($S, A \gg 0$) auf, während sie am Nachmittag die Amplitude sehr gut simulieren, allerdings mit zu flächigen Objekten ($S \approx 1$).
- Im Fall der erzwungenen Konvektion zeigen beide Modelltypen eine gute Vorhersageleistung. Allerdings zeigen die hoch aufgelösten weniger das Problem zu flächiger Objekte bei gleichzeitig zu geringer Niederschlagsmenge.

Hinter den im letzten Punkt beschriebenen sommerlichen Wetterlagen steht das Konzept der 'konvektiven Zeitskala' τ_c , welches in Kapitel 5.2 detailliert beschrieben wurde. Die erste aus Beobachtungen abgeleitete Klimatologie dieser Größe für Deutschland zeigt einen nach einem Potenzgesetz folgenden Abfall zu größeren Werten von τ_c hin. Die Unterteilung von dreistündigen QNVs (00 UTC-Lauf und 12 UTC Analysezeitpunkt) des COSMO-7 für einen siebenjährigen Zeitraum (2001-2007) in zwei Klassen zeigt bei der Verifikation mit SAL, dass für kurze konvektive Zeitskalen (Situationen mit erzwungener

Konvektion) die Vorhersagegüte ziemlich gut ist, für große Werte (Situationen mit Luftmassenkonvektion) dagegen dramatisch schlechte SAL Ergebnisse auftreten. Somit ist die Verwendung der konvektiven Zeitskala eine sehr gute Möglichkeit, um für unterschiedliche Situationen deren Vorhersagbarkeit aufzuzeigen.

Bei der Analyse immer kürzeren Akkumulationszeiten (≤ 1 h) der QNVs wird neben dem Problem der räumlichen Verschiebung der Niederschlagsereignisse auch der Versatz in der Zeit immer bedeutender. Dies kann soweit führen, dass das Ereignis zum Beobachtungszeitpunkt nicht auftritt, sondern kurz davor oder danach. In diesen Situationen kann eine objekt-orientierte Verifikationsmethode, wie z.B. SAL, nicht angewandt werden. Um dieses Problem zu umgehen, wurde ein Ansatz entwickelt, der die QNVs in einem Fenster um den Beobachtungszeitpunkt analysiert. Derjenige Zeitpunkt mit dem kleinsten Verschiebungsfehler L liefert den Zeitversatz sowie die dazugehörigen SAL Werte. Anhand eines Beispiels von stündlichen QNVs des COSMO-DE für Sommer 2007 im Elbeinzugsgebiets, konnte gezeigt werden, dass bei optimalem Zeitversatz die Werte von S und A besser sind. Besonders die Unterschätzung am Nachmittag, dem Zeitpunkt des meisten Niederschlags, tritt nicht auf. Somit konnte die grundsätzliche Fähigkeit dieses Modells aufgezeigt werden, die Struktur und Amplitude der Niederschlagsverteilung korrekt zu simulieren. Die Betrachtung des mittleren optimalen Zeitversatzes zeigt, dass dieses Modell keinen systematischen zeitlichen Fehler aufweist.

Mithilfe von SAL konnten viele nützliche Informationen auf charakteristische Fehler bei der QNV aufgezeigt werden. Aus dem Wissen dieser Defizite können Hinweise zur Verbesserung der NWV Modelle abgeleitet werden und somit liefert diese Arbeit einen wichtigen Beitrag im Rahmen des SPP.

6.2 Ausblick

Die Entwicklung in der numerischen Wettervorhersage geht zu immer feineren Modellgittern sowie zur Generierung hochaufgelöster Ensemble-Systeme. Diese Vorhersageverfahren benötigen spezielle Werkzeuge, um deren Qualität zu bewerten. Im Folgenden werden einige Vorschläge für Untersuchungen und Entwicklungen auf dem Gebiet der merkmalsbezogenen Verifikation in der nahen Zukunft vorgestellt.

Ein erster Aspekt zu weiteren Arbeiten dürfte die Erhöhung der zeitlichen Frequenz des Beobachtungsdatensatz von einer Stunde auf 15 Minuten sein. Auch die Verwendung eines anderen Radarprodukts sollte angedacht werden, da das aktuell verwendete PC-Produkt nur 7 Intensitätsklassen für die Reflektionsrate enthält. Das qualitätsgeprüfte RY-Produkt, welches zur Zeit beim DWD zur Anwendung kommt, besitzt 256 Intensitäts-

klassen. Damit sollte ein zeitliche Auflösung von 15 Minuten ohne weiteres möglich sein. Außerdem sollte überlegt werden, wie mit der geringeren Stationsanzahl zur Bestimmung des gegitterten täglichen Datensatzes umgegangen wird, falls die Anzahl noch erheblich geringer wird. Eine Abschätzung der Unsicherheiten des Beobachtungsdatsatz über die in *Paulat (2007)* gemachten hinaus könnte entsprechend *Ahrens and Beck (2008)* für die gegitterten Stationsdaten durch probabilistisches Hochskalieren ein Ensemble mittels stochastischen Ansatzes erzeugen. Um die Unsicherheiten der Radarmessungen abzuschätzen, kann der Ansatz von *Germann et al. (2009)* angewandt werden, bei dem ebenfalls ein Ensemble generiert wird, welches die singulären Vektoren der Fehlerkovarianzmatrix benutzt.

Die Entwicklung von hochaufgelösten regionalen Ensemble-Vorhersagesystemen wie das COSMO-DE EPS benötigen zur Verifikation von Niederschlagsereignissen merkmalsbezogene Techniken. Dazu müssen die bisher entwickelten Verfahren für die Anwendung auf Ensemble-Vorhersagen abgepasst werden. Vorschläge wie ein solches Verfahren funktionieren könnte, liefert z.B. *Gallus (2010)* bei der Anwendung von CRA und MODE. Hierbei wurden die mittleren Eigenschaften der einzelnen Ensemble-Mitglieder benutzt. Nachdem in dieser Arbeit anhand eines Beispiels gezeigt werden konnte, dass mit Hilfe der Optimierung der L Komponente innerhalb eines Zeitfensters der zeitliche Fehler von QNVs quantifiziert werden kann, sollte dies auf längere Zeitreihen angewandt werden. Auch der Vergleich zwischen verschiedenen Modellen kann sehr lehr- und aufschlussreich sein. Hierfür kann wiederum der D-PHASE Datensatz verwendet werden. Darüber hinaus ist es ebenfalls sehr interessant, die Untersuchung für verschiedene Wetterlagen durchzuführen. In dieser Arbeit wurden zwei konvektive Typen der Niederschlagsereignisse systematisch durch Verwendung der konvektiven Zeitskala separiert und mit SAL verifiziert. Ein weiterer Schritt könnte sein, Niederschlagsereignisse zu unterteilen, die mit Fronten im Zusammenhang stehen. Eine Möglichkeit hierzu stellte *Jenkner et al. (2010)* vor, bei der Fronten als Regionen intensiver horizontaler Gradienten im Feld der potentiellen Äquivalenttemperatur auf 850 hPa (*Hewson, 1998*) identifiziert werden. Die Methode erlaubt es, zwischen quasi-stationären und sich bewegenden Warm- und Kaltfronten zu unterscheiden. Damit können sechs verschiedene Wettersituationen (prä-, post- sowie frontal für Kalt- und Warmfront) klassifiziert und hinsichtlich ihrer Vorhersagegüte mit SAL verifiziert werden.

Als letzter Punkt möglicher zukünftiger Arbeitsgebiete von SAL, sei die Anwendung auf andere hoch variable meteorologische Parameter erwähnt. An dieser Stelle bietet sich der Wind an. Zum einem, weil er im weiteren Fortgang des SVFM ICP als nächster Untersuchungsparameter angedacht ist. Zum anderen, da besonders die Vorhersage von Starkwindereignissen vergleichbar mit dem Niederschlag äußerst problematisch ist. Folgender An-

satz wird beim SVFM ICP gewählt, um das vektorielle Windfeld durch räumliche Methoden zu verifizieren: Das zweidimensionale vektorielle Windfeld wird in einen divergenz- und einen rotationsfreien Anteil zerlegt (Helmholtz-Theorem, siehe z.B. *Salby*, 1996). Für diese erzeugten Felder lassen sich nun Objekte auf charakteristische Eigenschaften hin verifizieren (*Fowler et al.*, 2009). Der divergenzfreie Anteil kann Information über die Simulation von Wirbeln liefern. Dagegen kann mit der Verifikation des rotationsfreien Anteils z.B. die Windscherung untersucht werden. Darüber hinaus kann der Betrag des Windfelds bestimmt werden, um Starkwindereignisse zu bewerten.

Anhang A

Häufigkeiten von Tagessummen

An dieser Stelle erfolgt die Darstellung der geografischen Verteilung der Häufigkeiten von Niederschlagsereignissen mit mehr als 1 mm Niederschlag pro Tag. In Abbildung A.1 ist diese beobachtete Jahreveilung für Deutschland gezeigt.

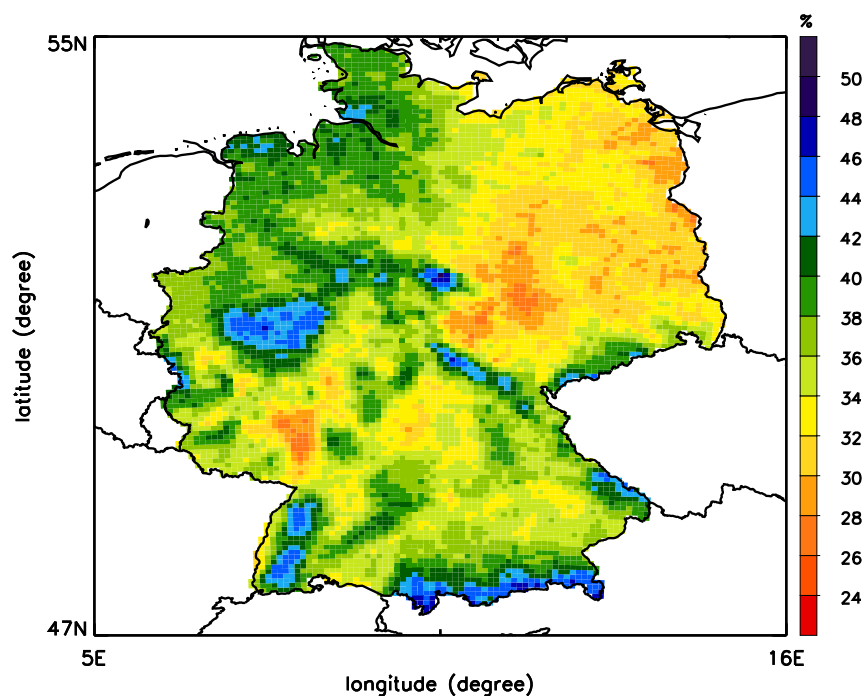


Abbildung A.1. Geografische Verteilung der relativen Häufigkeit (%) von Tagen mit Niederschlagssummen ≥ 1 mm/d aus gerasterten Stationsbeobachtungen für Deutschland im Zeitraum 2001 - 2008.

Zeichentabelle

Abkürzung	Bezeichnung
<i>a</i>	Treffer (Kontingenztabelle)
<i>A</i>	Amplituden-Komponente von SAL
<i>b</i>	Fehlalarm (Kontingenztabelle)
<i>B</i>	Bedeckung
<i>c</i>	Verpassen (Kontingenztabelle)
CAPE	Konvektive verfügbare Energie (<i>engl.: Convective Available Potential Energy</i>)
CIN	Konvektionshemmung (<i>engl.: Convective INhibition</i>)
COPS	<i>engl.: Convective and Orographically induced Precipitation Study</i>
COSMO	<i>engl. CONSORTIUM for Small scale MOdeling</i>
CRA	<i>Contiguous Rain Area</i>
<i>d</i>	korrekte Ablehnung (Kontingenztabelle), maximale Distanz, Tag (<i>engl.: day</i>)
<i>D</i>	gebietsgemittelte Niederschlagssumme
DAS	<i>engl.: Displacement and Amplitude Score</i>
<i>dd</i>	Windrichtung
DFG	Deutsche Forschungsgemeinschaft
D-PHASE	<i>engl.: Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region</i>
DWD	Deutscher Wetterdienst
ECMWF	<i>engl.: European Centre for Medium Range Weather Forecasts</i>
ETS	<i>engl.: Equitable Threat Score</i>
F0-F7	manipulierte Felder beim ICP
<i>ff</i>	Windstärke
FBI	Häufigkeitsbias (<i>engl.: Frequency Bias</i>)
<i>fx</i>	Windböen $> 12\text{ms}^{-1}$
G0-G5	geometrische Felder beim ICP

Abkürzung	Bezeichnung
hPa	HektoPascal
HPC	Lufmassenkonvektion (<i>engl.: High Pressure Convection</i>)
HSS	Heidke Skill Score
ICP	Kurz für SFVM-ICP
IOP	<i>Intensive Observation Period</i>
IQR	Interquartilsabstand (<i>engl.: InterQuartile Range</i>)
JJA	Abkürzung für Juni, Juni, August
L	Orts-Komponente von SAL
L_1	normierte Distanz der Schwerpunkte der Niederschlagsverteilung
L_2	Anteil an L aufgrund von Differenz von r
LHN	<i>engl.: Latent Heat Nudging</i>
M	Anzahl an identifizierten Objekten
MAP	<i>engl.: Mesoscale Alpine Project</i>
MAX	Maximumtemperatur
ME	mittlerer Fehler (<i>engl.: Mean Error</i>)
MIN	Minimumtemperatur
mod	Modell
MODE	<i>engl.: Method for Object-based Diagnostic Evaluation</i>
MSE	mittlerer quadratischer Fehler (<i>engl.: Mean Square Error</i>)
MJJASO	Abkürzung für Mai bis Oktober
MSO	Abkürzung für Mai, September und Oktober
$N.0$	Niederschlag(Ja/Nein)
N	Anzahl der Gitterpunkte
n	Summe aller Einträge in einer Kontingenztabelle
$nRMSE$	mit Beobachtung normierter RMSE
NWV	Numerische WetterVorhersage
obs, OBS	Beobachtung
P	Niederschlagsrate
QNV	Quantitative NiederschlagsVorhersage
QPF	<i>engl.: Quantitative Precipitation Forecasts</i>
r	gewichtete mittlere Distanz
R	Regenrate
R^*	Objektschwellwert bei SAL
R_0	konstanter Niederschlagswert
R^{95}	Wert des 95 % Perzentil des Niederschlags
R^{max}	maximaler Gitterpunktwert des Niederschlags

Abkürzung	Bezeichnung
R^{mod}	Niederschlagsvorhersage
R^{obs}	Niederschlagsbeobachtung
RMSE	Wurzel des mittleren quadratischen Fehlers (<i>engl.: Root Mean Square Error</i>)
RV	Reduktion der Varianz
S	Struktur-Komponente von SAL
SFC	Erzwungene Konvektion (<i>engl.: Strongly Forced Convection</i>)
SKSC	<i>engl.: SKill SCore</i>
SPP	SchwerpunktPunktProgramm
SFVM-ICP	<i>engl.: Spatial Forecast Verification Methods Intercomparison Project</i>
T	Temperatur
T_0	Referenztemperatur
T_a	Umgebungstemperatur
TR	Trefferrate
TS	<i>engl.: Thread Score</i>
USA	Vereinigte Staaten von Amerika <i>engl.: United States of America</i>
UTC	koordinierte Weltzeit <i>engl.: Universal Time Coordinated</i>
V	skaliertes Volumen
VERIPREG	<i>engl.: VERification of PREcipitation over Germany</i>
WRF	<i>engl.: Weather Research and Forecasting</i>
\mathbf{x}	Schwerpunktkoordinaten der Niederschlagsverteilungen
Z	Reflektivitätsrate

Variable	Bezeichnung
α	Irrtumswahrscheinlichkeit
c_p	spezifische Wärmekapazität [$\text{J Kg}^{-1} \text{K}^{-1}$]
Δt	zeitlicher Vorhersagefehler [h]
Δx	horizontale Gitterweite [km]
g	Schwerebeschleunigung [m s^{-2}]
L_v	Latente Verdunstungswärme [J Kg^{-1}]
ρ	Luftdichte [kg cm^{-3}]
τ_c	konvektive Zeitskala (<i>engl.: convective timescale</i>) [h]

Literaturverzeichnis

- Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert (2009), Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts, *Wea. Forecasting*, 24(6), 1485–1497, doi:{10.1175/2009WAF2222298.1}.
- Ahrens, B., and A. Beck (2008), On upscaling of rain-gauge data for evaluating numerical weather forecasts, *Meteorol. Atmos. Physics*, 99(3-4), 155–167, doi:{10.1007/s00703-007-0261-8}.
- Albrecht, B. (1989), Aerosols, Cloud Microphysics, and Fractional Cloudiness, *Science*, 245(4923), 1227–1230.
- Amodei, M., and J. Stein (2009), Deterministic and fuzzy verification methods for a hierarchy of numerical models, *Meteorol. Appl.*, 16(2), 191–203, doi:{10.1002/met.101}.
- Arakawa, A. (2004), The Cumulus Parameterization Problem: Past, Present, and Future, *J. Climate*, 17, 2493–2525.
- Arakawa, A., and W. Schubert (1974), Interaction of a Cumulus Cloud Ensemble with Large-Scale Environment. Part 1., *J. Atmos. Sci.*, 31(3), 674–701.
- Arpagaus, M., et al. (2009), MAP D-PHASE: Demonstrating forecast capabilities for flood events in the Alpine region.
- Baldauf, M., and J.-P. Schulz (2004), Prognostic Precipitation in the Lokalmodell (LM) of DWD, *COSMO Newsletter*, 4, 177–180.
- Baldauf, M., J. Förstner, S. Klink, T. Reinhardt, C. Schraff, A. Seifert, and K. Stephan (2010), Kurze Beschreibung des Lokal-Modells Kurzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD, *Deutscher Wetterdienst, Offenbach, Deutschland*.

- Balzer, K. (2002), Wie gut sind unsere Wettervorhersagen?, *Deutscher Wetterdienst, Basisdienste, Offenbach, Deutschland*.
- Bäumer, D., and B. Vogel (2007), An unexpected pattern of distinct weekly periodicities in climatological variables in Germany, *Geophys. Res. Lett.*, 34(3), doi:{10.1029/2006GL028559}.
- Bäumer, D., R. Rinke, and B. Vogel (2008), Weekly periodicities of Aerosol Optical Thickness over Central Europe - evidence of an anthropogenic direct aerosol effect, *Atmos. Chem. Phys.*, 8(1), 83–90.
- Beany, G., and W. A. Gough (2002), The influence of tropospheric ozone on the air temperature of the city of Toronto, Ontario, Canada, *Atmos. Environ.*, 36(14), 2319–2325.
- Beck, C., J. Grieser, and S. Trömel (2003), Die Trockenperiode des Jahres 2003 in Deutschland im Kontext langzeitlicher Niederschlagsvariabilität, *Klimastatusbericht des Deutschen Wetterdienst (DWD)*, pp. 142–151.
- Bell, T. L., D. Rosenfeld, K.-M. Kim, J.-M. Yoo, M.-I. Lee, and M. Hahnenberger (2008), Midweek increase in US summer rain and storm heights suggests air pollution invigorates rainstorms, *J. Geophys. Res.*, 113(D2), doi:{10.1029/2007JD008623}.
- Benoit, R., et al. (2002), The real-time ultrafinescale forecast support during the special observing period of the MAP, *Bull. Am. Meteorol. Soc.*, 83(1), 85–109.
- Betts, A. (1986), A new Convective Adjustment Scheme Part 1: Observational and Theoretical Basis, *Quart. J. Roy. Meteorol. Soc.*, 112(473), 677–691.
- Betts, A., and M. Miller (1986), A new Convective Adjustment Scheme Part 2: Single Column Tests using GATE WAVE, BOMEX, ATEX AND ARCTIC AIR-MASS Data Sets, *Quart. J. Roy. Meteorol. Soc.*, 112(473), 693–709.
- Binder, P., and C. Schär (1996), *The Mesoscale Alpine Programme MAP: Design Proposal*, second ed., 77 pp., Meteo Swiss, Zürich, Schweiz.
- Bougeault, P., and J. Geleyn (1989), Some Problems of Closure Assumption and Scale Dependency in the Parameterization of Moist Deep Convection for Numerical Weather Prediction, *Meteorol. Atmos. Physics*, 40(1-3), 123–135.
- Briggs, W., and R. Levine (1997), Wavelets and field forecast verification, *Mon. Wea. Rev.*, 125(6), 1329–1341.

- Brown, B., R. Bullock, C. David, J. Gotway, M. Chapman, A. Takacs, E. Gilleland, K. Manning, and J. Mahoney (2004), New verification approaches for convective weather forecasts, Vortrag, 11th Conf. Aviation, Range, and Aerospace Meteorology, 4-8 Oct 2004, Hyannis, MA.
- Brown, D., R. Bullock, and C. Davis (2009), MODE-3D: Incorporation of the time dimension, 4th International Verification Methods Workshop, Helsinki, Finland, 8.-10. Jun. 2009, Vortrag, 4th International Verification Methods Workshop, Helsinki, Finland, 8.-10. Jun. 2009.
- Bryan, G., J. Wyngaard, and J. Fritsch (2003), Resolution requirements for the simulation of deep moist convection, *Mon. Wea. Rev.*, *131*(10), 2394–2416.
- Burt, P., A. Ghelli, and E. Ebert (2008), Special issue on forecast verification, *Meteorol. Appl.*, *15*(1), 1, doi:{10.1002/met.69}.
- Casati, B., and L. J. Wilson (2007), A new spatial-scale decomposition of the Brier score: Application to the verification of lightning probability forecasts, *Mon. Wea. Rev.*, *135*(9), 3052–3069, doi:{10.1175/MWR3442.1}.
- Casati, B., et al. (2008), Forecast verification: current status and future directions, *Meteorol. Appl.*, *15*(1), 3–18, doi:{10.1002/met.52}.
- Casati, B., G. Ross, and D. Stephenson (2004), A new intensity-scale approach for the verification of spatial precipitation forecasts, *Meteorol. Appl.*, *11*(2), 141–154, doi:{10.1017/S1350482704001239}.
- Cerveny, R., and R. Balling (1998), Weekly cycles of air pollutants, precipitation and tropical cyclones in the coastal NW Atlantic region, *Nature*, *394*(6693), 561–563.
- Clark, A. J., W. A. Gallus, Jr., and T.-C. Chen (2007), Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models, *Mon. Wea. Rev.*, *135*(10), 3456–3473, doi:{10.1175/MWR3467.1}.
- Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong (2009), A Comparison of Precipitation Forecast Skill between Small Convection-Allowing and Large Convection-Parameterizing Ensembles, *Wea. Forecasting*, *24*(4), 1121–1140, doi:{10.1175/2009WAF2222222.1}.
- Cohen, B., and G. Craig (2004), The response time of a convective cloud ensemble to a change in forcing, *Quart. J. Roy. Meteorol. Soc.*, *130*(598, Part A), 933–944.

- Craig, G. C., C. Keil, and D. Leuenberger (2010), Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection, *Quart. J. Roy. Meteorol. Soc.*, eingereicht.
- Dai, A., F. Giorgi, and K. Trenberth (1999), Observed and model-simulated diurnal cycles of precipitation over the contiguous United States, *J. Geophys. Res.*, *104*(D6), 6377–6402.
- Davis, C., and F. Carr (2000), Summary of the 1998 workshop on mesoscale model verification, *Bull. Amer. Meteor. Soc.*, *81*, 809–819.
- Davis, C., B. Brown, and R. Bullock (2006), Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas, *Mon. Wea. Rev.*, *134*(7), 1772–1784.
- Davis, C. A., B. G. Brown, R. Bullock, and J. Halley-Gotway (2009), The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program, *Wea. Forecasting*, *24*(5), 1252–1267, doi:{10.1175/2009WAF2222241.1}.
- Done, J., C. A. Davis, and M. Weisman (2004), The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model, *Atmos. Res. Lett.*, *5*, 110–117, doi:{10.1002/asl.72}.
- Done, J., G. Craig, S. Gray, P. Clark, and M. Gray (2006), Mesoscale simulations of organized convection: Importance of convective equilibrium, *Quart. J. Roy. Meteorol. Soc.*, *132*(616, Part A), 737–756, doi:{10.1256/qj.04.84}.
- Done, J. M. a. C. C., S. L. Gray, P. A. Clark, and M. E. B. Gray (2010), Case-to-case variability of predictability of deep convection in a mesoscale model, *Quart. J. Roy. Meteorol. Soc.*, eingereicht.
- Doolittle, M. (1888), Association Ratios, *Bull. Phil Soc., Washington*, *7*, 122–127.
- Ebert, E. (2001), Ability of a poor man’s ensemble to predict the probability and distribution of precipitation, *Mon. Wea. Rev.*, *129*(10), 2461–2480.
- Ebert, E., and J. McBride (2000), Verification of precipitation in weather systems: determination of systematic errors, *J. Hydrol.*, *239*(1-4), 179–202.

- Ebert, E., U. Damrath, W. Wergen, and M. Baldwin (2003), The WGNE assessment of short-term quantitative precipitation forecasts, *Bull. Am. Meteorol. Soc.*, 84(4), 481–492, doi:{10.1175/BAMS-84-4-481}.
- Ebert, E. E. (2008), Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework, *Meteorol. Appl.*, 15(1), 51–64, doi:{10.1002/met.25}.
- Efron, B. (1979), 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife, *Annals of Statistics*, 7(1), 1–26.
- Efron, B., and R. Tibshirani (1986), Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy, *Stat. Sci.*, 1(1), 54–77.
- Elementi, M., C. Marsigli, and T. Paccagnella (2005), High resolution forecast of heavy precipitation with Lokal Modell: analysis of two case studies in the Alpine area, *Nat. Haz. Earth Sys. Sci.*, 5(4), 593–602.
- Emanuel, K., J. Neelin, and C. Bretherton (1994), On large-scale circulations in convecting atmospheres, *Quart. J. Roy. Meteorol. Soc.*, 120(519, Part B), 1111–1143.
- Finke, U., and T. Hauff (1996), The characteristics of lightning occurrence in Southern Germany, *Contrib. Atmos. Phys.*, 69(3), 361–374.
- Finley, J. (1884), Tornado prediction, *Am. Meteorol. Jour.*, 1, 85–88.
- Fowler, T., R. Bullock, B. Brown, and M. Pocerlich (2009), User Relevant Verification for Wind Forecasts, Vortrag, 4th International Verification Methods Workshop, Helsinki, Finland, 8.-10. Jun. 2009.
- Franssen Hendricks, H. J. (2008), An unexpected pattern of distinct weekly periodicities in climatological variables in Germany, *Geophys. Res. Lett.*, 35(5), doi:{10.1029/2007GL031279}.
- Frei, C., and C. Schär (1998), A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Clim.*, 18(8), 873–900.
- Frei, C., J. Christensen, M. Deque, D. Jacob, R. Jones, and P. Vidale (2003), Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps, *J. Geophys. Res.*, 108(D3), doi:{10.1029/2002JD002287}.
- Fritsch, J., and R. Carbone (2004), Improving quantitative precipitation forecasts in the warm season - A USWRP research and development strategy, *Bull. Am. Meteorol. Soc.*, 85(7), 955+, doi:{10.1175/BAMS-85-7-955}.

- Früh, B., and V. Wirth (2007), Convective Available Potential Energy (CAPE) in mixed phase cloud conditions, *Quart. J. Roy. Meteorol. Soc.*, 133(624, Part A), 561–569, doi: {10.1002/qj.39}.
- Früh, B., J. Bendix, T. Nauss, M. Paulat, A. Pfeiffer, J. W. Schipper, B. Thies, and H. Wernli (2007), Verification of precipitation from regional climate simulations and remote-sensing observations with respect to ground-based observations in the upper Danube catchment, *Meteorol. Zeit.*, 16(3), 275–293, doi: {10.1127/0941-2948/2007/0210}.
- Gallus, W. A., Jr. (2010), Application of Object-Based Verification Techniques to Ensemble Precipitation Forecasts, *Wea. Forecasting*, 25(1), 144–158, doi: {10.1175/2009WAF2222274.1}.
- Gebhardt, C., S. Theis, P. Krahe, and V. Renner (2008), Experimental ensemble forecasts of precipitation based on a convection-resolving model, *Atmos. Res. Let.*, 9(2), 67–72, doi: {10.1002/asl.177}.
- Germann, U., M. Berenguer, D. Sempere-Torres, and M. Zappa (2009), REAL - Ensemble radar precipitation estimation for hydrology in a mountainous region, *Quart. J. Roy. Meteorol. Soc.*, 135(639, Part B), 445–456, doi: {10.1002/qj.375}.
- Gilbert, G. (1884), Finley's tornado predictions, *Am. Meteorol. Jour.*, 1, 166–172.
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert (2009), Intercomparison of Spatial Forecast Verification Methods, *Wea. Forecasting*, 24(5), 1416–1430, doi: {10.1175/2009WAF2222269.1}.
- Gordon, A. (1994), Weekdays warmer than Weekends, *Nature*, 367(6461), 325–326.
- Grützun, V., O. Knoth, and M. Simmel (2008), Simulation of the influence of aerosol particle characteristics on clouds and precipitation with LM-SPECS: Model description and first results, *Atmos. Res.*, 90(2-4, Sp. Iss. SI), 233–242, doi: {10.1016/j.atmosres.2008.03.002}.
- Hamill, T. (1999), Hypothesis tests for evaluating numerical precipitation forecasts, *Wea. Forecasting*, 14(2), 155–167.
- Harris, D., E. Foufoula-Georgiou, K. Droegemeier, and J. Levit (2001), Multiscale statistical properties of a high-resolution precipitation forecast, *J. Hydrometeorol.*, 2(4), 406–418.

- Heidke, P. (1926), Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst, *Geografika Annaler, Stockholm*, 8, 301–349.
- Hense, A., G. Adrian, C. Kottmeier, C. Simmer, and V. Wulfmeyer (2003), Quantitative Niederschlagsvorhersage, Proposal zum SPP 1167 der DFG.
- Hewson, T. (1998), Objective fronts, *Meteorol. Appl.*, 5(1), 37–65.
- Hoffman, R., Z. Liu, J. Louis, and C. Grassotti (1995), Distortion Representation of Forecast Errors, *Mon. Wea. Rev.*, 123(9), 2758–2770.
- Hofmann, C., M. Zimmer, and H. Wernli (2009), A brief catalog of poor and excellent COSMO model QPFs in German river catchments., 31 pp.
- Hohenegger, C., P. Brockhaus, and C. Schaer (2008), Towards climate simulations at cloud-resolving scales, *Meteorol. Zeit.*, 17(4, Sp. Iss. SI), 383–394, doi:{10.1127/0941-2948/2008/0303}.
- IPCC (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Solomon, S. and D. Qin and M. Manning and Z. Chen and M. Marquis and K.B. Averyt and M.Tignor and H.L. Miller (Hrsg.)), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Janowiak, J. E., V. J. Dagostaro, V. E. Kousky, and R. J. Joyce (2007), An examination of precipitation in observations and model forecasts during NAME with emphasis on the diurnal cycle, *J. Climate*, 20(9), 1680–1692, doi:{10.1175/JCL14084.1}.
- Jenkner, J., C. Frei, and C. Schwiertz (2008), Quantile-based short-range QPF evaluation over Switzerland, *Meteorol. Zeit.*, 17(6, Sp. Iss. SI), 827–848, doi:{10.1127/0941-2948/2008/0344}.
- Jenkner, J., M. Sprenger, I. Schwenk, C. Schwiertz, S. Dierer, and D. Leuenberger (2010), Detection and climatology of fronts in a high-resolution over model reanalysis the Alps, *Meteorol. Appl.*, 17(1), 1–18, doi:{10.1002/met.142}.
- Jin, M., J. Shepherd, and M. King (2005), Urban aerosols and their variations with clouds and rainfall: A case study for New York and Houston, *J. Geophys. Res.*, 110(D10), doi:{10.1029/2004JD005081}.
- Jolliffe, I. T., and D. B. Stephenson (2003), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 257 pp., John Wiley and Sons, Chichester.

- Jones, C. D., and B. Macpherson (1997), A Latent Heat Nudging Scheme for the Assimilation of Precipitation Data into an Operational Mesoscale Model., *Meteorol. Appl.*, 4, 269–277.
- Kain, J. S., et al. (2008), Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP, *Wea. Forecasting*, 23(5), 931–952, doi:{10.1175/2008WAF2007106.1}.
- Kaufmann, P., F. Schubiger, and P. Binder (2003), Precipitation forecasting by a mesoscale numerical weather prediction (NWP) model: eight years of experience, *Hydrol. Earth Sys. Sci.*, 7(6), 812–832.
- Keil, C., and G. C. Craig (2009), A Displacement and Amplitude Score Employing an Optical Flow Technique, *Wea. Forecasting*, 24(5), 1297–1308, doi:{10.1175/2009WAF2222247.1}.
- Kessler, E. (1969), On the distribution and continuity of water substance in atmospheric circulations, Amer. Meteor. Soc., Boston, 84 pp.
- Laux, P., and H. Kunstmann (2008), Detection of regional weekly weather cycles across Europe, *Environ. Res. Let.*, 3(4), doi:{10.1088/1748-9326/3/4/044005}.
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell (2008), Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom, *Mon. Wea. Rev.*, 136(9), 3408–3424, doi:{10.1175/2008MWR2332.1}.
- Lin, Y., and K. Mitchell (2005), The NCEP Stage II/IV hourly precipitation analyses: Development and applications., *Preprints from the 19th Conf on Hydrology*, Amer. Meteorol. Soc., p. 1.2.
- Liu, C., M. Moncrieff, J. Tuttle, and R. Carbone (2006), Explicit and Parameterized Episodes of Warm-Season Precipitation over the Continental United States, *Adv. Atmos. Sci.*, 23(1), 91–105.
- Lohmann, U., and J. Feichter (2005), Global indirect aerosol effects: a review, *Atmos. Chem. Phys.*, 5, 715–737.
- Lugauer, M., and P. Winkler (2005), Thermal circulation in South Bavaria - climatology and synoptic aspects, *Meteorol. Zeit.*, 14(1), 15–30, doi:{10.1127/0941-2948/2005/0014-0015}.

- Majewski, D., D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen, and J. Baumgardner (2002), The operational global icosahedral-hexagonal grid-point model GME: Description and high-resolution tests, *Mon. Wea. Rev.*, *130*(2), 319–338.
- Mannstein, H., H. Huntrieser, and S. Wimmer (2002), Determination of the mass flux in convective cells over Europe, *Proc. 2002 EUMETSAT Meteorological Satellite Conf., Dublin, Ireland*, pp. 264–269.
- Marzban, C., and S. Sandgathe (2006), Cluster analysis for verification of precipitation fields, *Wea. Forecasting*, *21*(5), 824–838.
- Mass, C., D. Ovens, K. Westrick, and B. Colle (2002), Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific northwest, *Bull. Am. Meteorol. Soc.*, *83*(3), 407+.
- McBride, J., and E. Ebert (2000), Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia, *Wea. Forecasting*, *15*(1), 103–121.
- Mellor, G. L., and T. Yamada (1974), A hierarchy of turbulence closure models for planetary boundary layers., *J. Atmos. Sci.*, *31*, 1791–1806.
- Michalakes, J., S. Chen, J. J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock (2001), Development of a next generation regional weather research and forecast model, 269–276 pp.
- Molinari, J., and M. Dudek (1992), Parameterization of Convective Precipitation in Mesoscale Numerical Models - A Critical Review, *Mon. Wea. Rev.*, *120*(2), 326–344.
- Molini, L., A. Parodi, N. Reboraa, and G. C. Craig (2010), Classifying severe rainfall events over Italy by hydrometeorological and dynamical criteria, *Quart. J. Roy. Meteorol. Soc.*, eingereicht.
- Müller-Westermeier, G. (1995), *Numerisches Verfahren zur Erstellung klimatologischer Karten*, Berichte des Deutschen Wetterdienstes 193, Selbstverlag des Deutschen Wetterdienstes, Offenbach.
- Müller-Westermeier, G., A. Kreis, and E. Dittman (1999), *Klimaatlas Bundesrepublik Deutschland, Teil 1*, Deutscher Wetterdienst, Offenbach, Deutschland.

- Murphy, A. (1993), What is a Good Forecast - An Essay on the Nature of Goodness in Weather Forecasting, *Wea. Forecasting*, 8(2), 281–293.
- Nachamkin, J. (2004), Mesoscale verification using meteorological composites, *Mon. Wea. Rev.*, 132(4), 941–955.
- Oberto, E., and M. Milelli (2010), QPF verification for 2008/2009 of several COSMO-Model versions, *COSMO Newsletter*, 10, 55–63.
- Paulat, M. (2007), Verifikation der Niederschlagsvorhersage für Deutschland von 2001-2004, Ph.D. thesis, Johannes Gutenberg-Universität Mainz.
- Paulat, M., C. Frei, M. Hagen, and H. Wernli (2008), A gridded dataset of hourly precipitation in Germany: Its construction, climatology and application, *Meteorol. Zeit.*, 17(6, Sp. Iss. SI), 719–732, doi:{10.1127/0941-2948/2008/0332}.
- Peters, O., and J. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nature Physics*, 2(6), 393–396, doi:{10.1038/nphys314}.
- Quaas, J., O. Boucher, A. Jones, G. Weedon, J. Kieser, and H. Joos (2009), Exploiting the weekly cycle as observed over Europe to analyse aerosol indirect effects in two climate models, *Atmos. Chem. Phys.*, 9(21), 8493–8501.
- Reutter, P. (2009), Numerical simulations of microphysical processes in pyro-convective clouds, Ph.D. thesis, Johannes Gutenberg-Universität Mainz.
- Richard, E., A. Buzzi, and G. Zaengl (2007), Quantitative precipitation forecasting in the Alps: The advances achieved by the Mesoscale Alpine Programme, *Quart. J. Roy. Meteorol. Soc.*, 133(625, Part B), 831–846, doi:{10.1002/qj.65}.
- Richardson, D. (2010), Landmark in Forecast Performance, *ECMWF Newsletter*, 123, 3–4.
- Ritter, B., and J. Geleyn (1992), A Comprehensive Radiation Scheme for Numerical Weather Prediction Modles with Potential Applications in Climate Simulations, *Mon. Wea. Rev.*, 120(2), 303–325.
- Roberts, N. (2008), Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model, *Meteorol. Appl.*, 15(1), 163–169, doi:{10.1002/met.57}.

- Rosenfeld, D. (2000), Suppression of rain and snow by urban and industrial air pollution, *Science*, 287(5459), 1793–1796.
- Rosenfeld, D., J. Dai, X. Yu, Z. Yao, X. Xu, X. Yang, and C. Du (2007), Inverse relations between amounts of air pollution and orographic precipitation, *Science*, 315(5817), 1396–1398, doi:{10.1126/Science.1137949}.
- Rossa, A., P. Nurmi, and E. Ebert (2008), *Overview of methods for the verification of quantitative precipitation forecasts*, Precipitation: Advances in Measurement, Estimation and Prediction, S.C.Michaelides, Ed. Springer, pp. 418-450.
- Rudolf, B., and J. Rapp (2002), Das Jahrhunderthochwasser der Elbe: Synoptische Wetterentwicklung und klimatologische Aspekte, *Klimastatusbericht des Deutschen Wetterdienst (DWD)*, pp. 172–188.
- Salby, M. (1996), *Fundamentals of Atmospheric Physics*, International Geophysics Series, Academic Press.
- Schär, C., P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, and C. Appenzeller (2004), The role of increasing temperature variability in European summer heatwaves, *Nature*, 427(6972), 332–336, doi:{10.1038/nature02300}.
- Schönwiese, C. D., and R. Janoschitz (2008), Klima-Trendatlas Deutschland 1901-2000, 64 pp.
- Schultz, D. M., S. Mikkonen, A. Laaksonen, and M. B. Richman (2007), Weekly precipitation cycles? Lack of evidence from United States surface stations, *Geophys. Res. Lett.*, 34(22), doi:{10.1029/2007GL031889}.
- Schulz, J., and U. Schättler (2005), Kurze Beschreibung des Lokal-Modells LME und seiner Datenbanken auf dem Datenserver des DWD, *Deutscher Wetterdienst, Offenbach, Deutschland*.
- Seifert, A. (2008), The challenge of convective scale numerical weather prediction, 4th PAN-GCSS meeting on 'Advances in Modelling and observing Clouds and Convection', Météo-France Conference Centre, Toulouse 2-6 June 2008.
- Shutters, S., and R. Balling (2006), Weekly periodicity of environmental variables in Phoenix, Arizona, *Atmos. Environ.*, 40(2), 304–310, doi:{10.1016/j.atmosenv.2005.09.037}.

- Skamarock, W. (2004), Evaluating mesoscale NWP models using kinetic energy spectra, *Mon. Wea. Rev.*, *132*(12), 3019–3032.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows (1989), *Survey of common verification methods in meteorology*, Research report 89-5, Environment Canada. Available from Atmos. Environ., 89 pp.
- Stephan, K., S. Klink, and C. Schraff (2008), Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD, *Quart. J. Roy. Meteorol. Soc.*, *134*(634, Part A), 1315–1326, doi:{10.1002/qj.269}.
- Steppeler, J., G. Doms, U. Schättler, H. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric (2003), Meso-gamma scale forecasts using the nonhydrostatic model LM, *Meteorol. Atmos. Physics*, *82*(1-4), 75–96, doi:{10.1007/s00703-001-0592-9}, 4th SRNWP Workshop on Nonhydrostatic Modelling, BAD ORB, GERMANY, SEP, 2001.
- Theis, S., A. Hense, and U. Damrath (2005), Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach, *Meteorol. Appl.*, *12*(3), 257–268, doi:{10.1017/S1350482705001763}.
- Tiedtke, M. (1989), A Comprehensive Mass Flux Scheme For Cumulus Parameterization in Large-Scale Models, *Mon. Wea. Rev.*, *117*(8), 1779–1800.
- Twomey, S. (1974), Pollution and Planetary Albedo, *Atmos. Environ.*, *8*(12), 1251–1256.
- Walser, A., and C. Schär (2004), Convection-resolving precipitation forecasting and its predictability in Alpine river catchments, *J. Hydrol.*, *288*(1-2), 57–73, doi:{10.1016/j.jhydrol.2003.11.035}, International Conference on Quantitative Precipitation Forecasting (QPF), Reading, England, Sep 02-06, 2002.
- Walser, A., D. Lüthi, and C. Schär (2004), Predictability of precipitation in a cloud-resolving model, *Mon. Wea. Rev.*, *132*(2), 560–577.
- Weisman, M., W. Skamarock, and J. Klemp (1997), The resolution dependence of explicitly modeled convective systems, *Mon. Wea. Rev.*, *125*(4), 527–548.
- Wernli, H., C. Frei, and M. Hagen (2003), Novel Approaches for the verification of precipitation forecasts from numerical models on short time-scales, Antrag innerhalb des SPP 1167 "Quantitative Niederschlagsvorhersage", Mainz, Deutschland.

- Wernli, H., M. Paulat, M. Hagen, and C. Frei (2008), SAL-A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, *Mon. Wea. Rev.*, 136(11), 4470–4487, doi:{10.1175/2008MWR2415.1}.
- Wernli, H., S. Pfahl, J. Trentmann, and M. Zimmer (2010), How representative were the meteorological conditions during the COPS field experiment in summer 2007?, *Meteorologische Zeitschrift*.
- Weverberg, K., N. van Lipzig, L. Delobbe, and D. Lauwaert (2010), Sensitivity of quantitative precipitation forecast to soil moisture initialization and microphysics parametrization, *Quart. J. Roy. Meteorol. Soc.*, (akzeptiert).
- Wilks, D. (2006), *Statistical Methods in the Atmospheric Sciences*, International Geophysics Series, Vol. 59, Academic Press, 627 pp.
- Wilson, J., N. Crook, C. Mueller, J. Sun, and M. Dixon (1998), Nowcasting thunderstorms: A status report, *Bull. Am. Meteorol. Soc.*, 79(10), 2079–2099.
- Wittmann, C., T. Haiden, and A. Kann (2010), Evaluating multi-scale precipitation forecasts using high resolution analysis, *Advances in Science and Research*, 4, 89–98, doi:10.5194/asr-4-89-2010.
- Wulfmeyer, V., Behrendt, C. Kottmeier, and U. Corsmeier (Hrsg.) (2005), Convective and orographically-induced precipitation study (cops): Science overview document, an observation program within the priority program quantitative precipitation forecast (ppp) funded by the german research foundation (erhältlich unter: https://www.uni-hohenheim.de/spp-iop/documents/051109_cops_sod_final.pdf), 148 pp.
- Wulfmeyer, V., A. Behrendt (Hrsg.), and die COPS Forschermannschaft (2007), COPS Field Report (erhältlich unter https://www.uni-hohenheim.de/cops/further_reading/further_reading.htm), 368 pp.
- Wulfmeyer, V., et al. (2008), The Convective and Orographically induced Precipitation Study A Research and Development Project of the World Weather Research Program for Improving Quantitative Precipitation Forecasting in Low-Mountain Regions, *Bull. Am. Meteorol. Soc.*, 89(10), 1477–1486, doi:{10.1175/2008BAMS2367.1}.
- Wulfmeyer, V., et al. (2010), The Convective and Orographically Induced Precipitation Study (COPS): The scientific strategy, the field phase, and first highlights, *Quart. J. Roy. Meteor. Soc.*, eingereicht.

- Zawadzki, I. (1973), Statistical properties of precipitation patterns, *J. Appl. Meteorol.*, 12, 459–472.
- Zimmer, M., and H. Wernli (2008), COPS Atlas – The meteorological situation from June 1 till August 31, 2007., 100 pp.
- Zimmer, M., H. Wernli, C. Frei, and M. Hagen (2009), Feature-based verification of deterministic precipitation forecasts with SAL during COPS, *Proceedings from the MAP D-PHASE Scientific Meeting in Bologna, Italy, May 2008*, pp. 116–121.