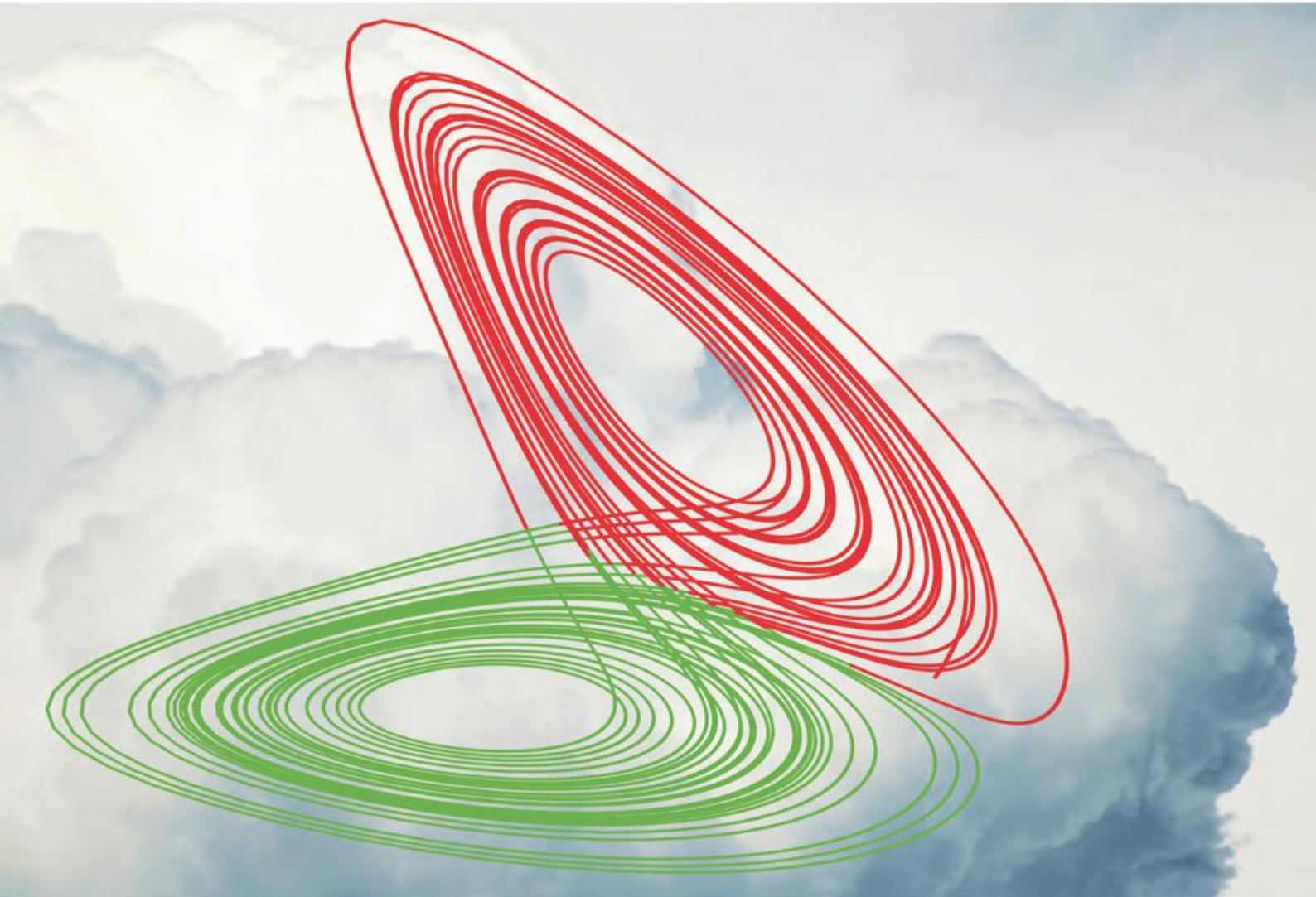


Jahrgang 37 Heft 3/4 2011

promet

meteorologische fortbildung

Probabilistische Wettervorhersage



Herausgeber

Deutscher Wetterdienst

Hauptschriftleiter

Dr. J. Rapp (Offenbach/M.)

Redaktionsausschuss

Prof. Dr. G. Adrian (Offenbach/M.)

Prof. Dr. B. Ahrens (Frankfurt/M.)

PD Dr. F. Berger (Lindenberg)

Prof. Dr. B. Brümmer (Hamburg)

Prof. Dr. G. C. Craig (München)

Prof. Dr. S. Jones (Offenbach/M.)

Prof. Dr. G. Groß (Hannover)

Prof. Dr. A. Macke (Leipzig)

Dr. E. Rudel (Wien)

Prof. Dr. C.-D. Schönwiese (Frankfurt/M.)

Technische Herausgabe (Layout)

S. Stummvoll (DWD, Offenbach/M.)

Internetwww.dwd.de/promet

promet erscheint im Selbstverlag des Deutschen Wetterdienstes – Frankfurter Str. 135, 63067 Offenbach am Main. Bezugspreis pro Jahrgang im Abonnement 24,08 €, Doppelheft 13,91 €, Dreifachheft 20,87 € plus Versandkosten.

Für den Inhalt der Arbeiten sind die Autoren verantwortlich. Alle Rechte bleiben vorbehalten.

Auflage: 3600

Druck:

kraus print u. media GmbH & Co. KG
Am Angertor 11
97618 Wülfershausen

Bezug:

Deutscher Wetterdienst
Fachinformationsstelle und
Deutsche Meteorologische Bibliothek
Frankfurter Str. 135
63073 Offenbach a. M.

ISSN 0340-4552 (Druck)

ISSN 2194-5990 (Internet, pdf)

Redaktionsschluss: 30. November 2012

Meteorologische Fortbildung

37. Jahrgang, Heft 3/4, 2011

Thema des Heftes:

Probabilistische Wettervorhersage

Fachliche Redaktion: R. Hagedorn, Offenbach a.M.

Fachliche Durchsicht: S. Jones, Offenbach a.M., C. Craig, München

Kapitel	Seite
Zu diesem Heft (R. Hagedorn)	1-2
H. WERNLI	
1 Wetter, Chaos und probabilistische Wettervorhersagen	3-11
M. LEUTBECHER	
2 Anfangsbedingungen für Ensemblevorhersagen	12-22
J. BERNER	
3 Verfahren zur Berücksichtigung von Modellfehlern in der numerischen Wettervorhersage	23-30
A. P. WEIGEL	
4 Verifikation von Ensemblevorhersagen	31-42
T. L. THORARINSDOTTIR, M. SCHEUERER, K. FELDMANN	
5 Statistische Nachbearbeitung von Ensemblevorhersagen	43-52
S. THEIS, C. GEBHARDT	
6 COSMO-DE-EPS: das „konvektionserlaubende“ Ensemble am DWD	53-61
L. VON BREMEN, M. LANGE	
7 Ensembleprognosen der Windleistung für Anwendungen in der Energiewirtschaft	62-71
M. DENHARD	
8 Herausforderungen in der Interpretation und Kommunikation von probabilistischen Vorhersagen	72-78
P. BISSOLLI, P. HECHLER, S. RÖSNER, H. NITSCHKE	
Ein Zentrum für europaweite Klimainformationen: Das WMO-Regionalzentrum Europa	79-93
Buchbesprechungen	94-95
Examina in 2010	96-104
Anschriften der Autoren dieses Heftes	105
Redaktionelle Hinweise für Autoren	106-107
Vorschau	107

Probabilistische Wettervorhersage

Zu diesem Heft

Die Idealvorstellung der Nutzer von Wettervorhersagen ist es – so genau wie möglich und so früh wie möglich – eine eindeutige Vorhersage der Zukunft zu bekommen. Die Welt ist jedoch, wie wir alle wissen, nicht ideal, geschweige denn exakt vorhersagbar! Insbesondere die Welt der Wettervorhersage ist durch ihre nichtlineare Natur so komplex, dass man ihr mit „Schwarz-Weiß-Aussagen“ nur in den wenigsten Fällen gerecht werden kann. Diese Tatsache, dass jede Vorhersage eine gewisse Unsicherheit aufweist, ist elementares Wissen in der Meteorologie. Die Kunst besteht nun darin abzuschätzen, wie groß die jeweilige Unsicherheit unserer Vorhersagen ist, oder wie es in der alten konfuzianischen Weisheit heißt:

Zu wissen, was man weiß, und zu wissen, was man nicht weiß, das ist echtes Wissen.

Schon Ende des 19. Jahrhunderts entdeckte Henri Poincaré, dessen Name bei Astronomen hauptsächlich mit seinen Arbeiten zur Himmelsmechanik verbunden ist, in seinen Analysen zur Stabilität des Sonnensystems das „deterministische Chaos“. Obwohl Poincaré prinzipiell noch dem damals vorherrschenden Weltbild des Determinismus folgt, erfasst er jedoch schon früh den Kern der Problematik um Vorhersagbarkeit, indem er schreibt:

„Wenn wir die Gesetze der Natur und den Anfangszustand exakt kennen würden, so könnten wir den Zustand des Universums zu jedem weiteren Zeitpunkt vorhersagen. Aber selbst wenn die Naturgesetze keine Geheimnisse mehr vor uns hätten, so könnten wir die Anfangsbedingungen doch nur genähert bestimmen. Wenn uns dies erlaubt, die folgenden Zustände mit der gleichen Näherung anzugeben, so sagen wir, dass das Verhalten vorhergesagt wurde, dass es Gesetzmäßigkeiten folgt. Aber das ist nicht immer der Fall: Es kann vorkommen, dass kleine Unterschiede in den Anfangsbedingungen große im Ergebnis zur Folge haben (...) Vorhersage wird unmöglich und wir haben ein zufälliges Phänomen.“

Die Erkenntnisse Poincarés gerieten jedoch lange in Vergessenheit, bis in der zweiten Hälfte des 20. Jahrhunderts Edward Lorenz, der heute als einer der Mitbegründer der Chaostheorie gilt, deren Bedeutung für die numerische Wettervorhersage neu entdeckte. Wie so häufig in der Wissenschaft spielte auch bei dieser Entdeckung der Zufall eine Rolle. Der entscheidende Hinweis ergab sich, als Lorenz erste numerische Experimente mit einem vereinfachten Wettermodell durchführte und eine der Simulationen noch einmal für einen längeren Zeitraum wiederholen wollte. Um Rechenzeit zu sparen, startete er die Rechnung nicht vom ursprünglichen Zeitpunkt sondern von einem Zustand in der Mitte des Simulationszeitraums. Dazu gab er den bis dahin berechneten Zustand als Startpunkt für die neue Simulation ein. Für Lorenz sehr überraschend ergaben sich jedoch schon nach nur kurzer Simulationszeit Unterschiede zur ursprünglichen Rechnung, obwohl er exakt das gleiche Modell verwendet hatte. Deshalb vermutete Lorenz zunächst einen Fehler im Rechner, entdeckte aber kurz darauf, dass der Grund in der Rundungsungenauigkeit des von ihm eingegebenen neuen Startpunktes lag. Anstatt alle sechs Dezimalstellen anzugeben, hatte Lorenz den bis dahin berechneten Zustand auf drei Dezimalstellen gerundet, und diese kleine Ungenauigkeit führte dazu, dass die Simulationen nach kurzer Zeit extrem divergieren konnten.

*Die erste – vielzitierte – Publikation zu diesen bahnbrechenden Ergebnissen erschien 1963 in der Zeitschrift *Tellus* mit dem Titel „Deterministic Nonperiodic Flow“, wobei der populäre Begriff des „Schmetterlingseffektes“ jedoch erst später entstand, nachdem Lorenz auf einer wissenschaftlichen Tagung einen Vortrag mit dem Titel „Predictability: Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?“ hielt. Lorenz zeigt in dieser Publikation zum ersten Mal, dass einfache deterministische Systeme prinzipiell nur für einen endlichen Zeitraum vorhersagbar sind, wobei diese Zeitspanne von der Genauigkeit abhängt, mit der die Anfangsbedingungen spezifiziert werden. Dies bedeutet einerseits, dass der mögliche Vorhersagehorizont größer wird, je genauer wir die Anfangsbedingungen kennen. Andererseits müssen wir beachten, dass selbst bei immer kleiner werdenden Fehlern in den Anfangsbedingungen dennoch immer eine Restunsicherheit bestehen bleibt, und es damit immer eine Grenze der Vorhersagbarkeit im Sinne der Anfangsbedingungen geben wird. Eine weitere wesentlich Fehlerquelle sind intrinsische Modellfehler, die sich aus unvermeidbaren Unzulänglichkeiten in den*

Formulierungen des dynamischen Kerns und der physikalischen Parametrisierungen des numerischen Modellsystem ergeben und damit auch einen wesentlichen Aspekt bezüglich Vorhersagbarkeit darstellen.

In Folge dieser ersten Arbeiten entwickelte sich das mittlerweile große Forschungsfeld der probabilistischen oder Ensemblevorhersagen, das mittlerweile nicht mehr aus der numerischen Wettervorhersage wegzudenken ist. So werden inzwischen weltweit mehr als zehn globale Ensemblesysteme operationell betrieben und ständig weiterentwickelt. In diesem Zusammenhang sei zum Beispiel das TIGGE Projekt (THORPEX Interactive Grand Global Ensemble) des World Weather Research Programme der WMO erwähnt, das sich insbesondere auf Fragen zu Vorhersagbarkeit, Ensemblesystemen und Erstellung von Produkten zur Verbesserung der Vorhersagen von Extremereignissen konzentriert (siehe auch <http://tigge.ecmwf.int>).

Das Prinzip von Ensemblevorhersagen ist es also, durch geeignete Variationen in den Anfangsbedingungen und Modellformulierungen abzuschätzen, wie groß die jeweilige Unsicherheit unserer Vorhersagen ist. Das heißt, anstatt nur eine einzige Modellrechnung pro Starttermin durchzuführen, wird ein gesamtes Ensemble von Rechnungen gestartet. Aus der Verteilung der unterschiedlichen Lösungen wird dann versucht, die Unsicherheit der Vorhersage abzuschätzen. Eine weitere wesentliche Erkenntnis in diesem Zusammenhang ist, dass das Wachstum der Fehler nicht konstant ist sondern von Tag zu Tag variieren kann, das heißt von den herrschenden atmosphärischen Bedingungen selbst abhängt.

Neben dieser Aufgabe, das heißt der reinen Bereitstellung von verlässlichen probabilistischen Informationen, ist es von größter Wichtigkeit, dass diese neue Information tatsächlich sinnvoll in die Anwendungen unserer Nutzer integriert wird. Da die letztendlichen Entscheidungs- und Handlungsmechanismen der Nutzer immer deterministischer Natur sind, stellt die Integration der probabilistischen Informationen die eigentliche Herausforderung dar. Das heißt, obwohl theoretische Überlegungen darauf hindeuten, dass die Verwendung probabilistischer Information von grundsätzlichem Vorteil für Nutzer ist, müssen - in enger Kommunikation mit den Nutzern - geeignete Produkte und Entscheidungsprozesse entwickelt werden, um die theoretischen Vorteile in praktischen Nutzen umsetzen zu können.

Alle diese Aspekte, die aktuelle Themen im Forschungsfeld der probabilistischen Wettervorhersagen sind, werden im vorliegenden Heft näher vorgestellt und diskutiert. Nach einer Einführung von Wernli in die theoretischen Hintergründe der Probabilistik, werden in den Beiträgen von Leutbecher und Berner die Herausforderungen im Zusammenhang mit der Simulation von Unsicherheiten in den Anfangsbedingungen bzw. der Modelle selbst diskutiert. Eine wesentliche Voraussetzung zur effizienten Weiterentwicklung und angemessenen Nutzung von Ensemblevorhersagen ist die probabilistische Verifikation, in die der Beitrag von Weigel einführt. Die Ergebnisse solcher Verifikation machen auch deutlich, dass die direkten Modellergebnisse von Ensemblesysteme häufig noch große Defizite in ihrer Verlässlichkeit aufweisen und dementsprechend schwer in die Entscheidungsprozesse der Nutzer einzubinden sind. Um diese Defizite auszugleichen, müssen diese Vorhersagen statistisch nachbearbeitet werden, und im fünften Beitrag beschreiben Thorarinsdottir et al. verschiedene sogenannte „Post-processing“-Techniken und die damit verbundenen Chancen und Herausforderungen. Das noch relativ junge Forschungsfeld der sogenannten „Limited Area“-Ensemblemodelle wird von Theis und Gebhardt am Beispiel des international ersten „konvektionserlaubenden“ Ensembles im operationellen Betrieb, dem COSMO-DE-EPS des DWD, vorgestellt. Von Bremen und Lange führen den Aspekt der Anwendung von Ensemblevorhersagen an praktischen Beispielen aus der Energiewirtschaft ein, während abschließend Denhard auf die Herausforderungen in der Kommunikation und Interpretation von probabilistischen Wettervorhersagen eingeht und hierbei insbesondere den Begriff des „true belief“ definiert und diskutiert.

In freier Interpretation dieses Begriffs möchte ich abschließend meine ureigenste Überzeugung zum Ausdruck bringen, dass die Frage, wie wir mit den unvermeidbaren Unsicherheiten der Wettervorhersage umgehen, in Zukunft noch mehr als bisher ein integraler Teil unserer Arbeit sein wird. Nur durch eine offene und klare Kommunikation der Möglichkeiten – aber auch der Grenzen – der numerischen Wettervorhersage können wir eine optimale Akzeptanz und größtmöglichen Nutzen unserer Arbeit erreichen. Ich hoffe, dass das vorliegende Heft einen Beitrag dazu leisten kann, die Bedeutung der probabilistischen Aspekte im Kontext der Wettervorhersage hervorzuheben und nachfolgende Generationen motivieren kann, sich in diesem spannenden und zukunftssträchtigen Forschungsfeld zu engagieren.

Renate Hagedorn

H. WERNLI

1 Wetter, Chaos und probabilistische Wettervorhersagen

Weather, chaos and probabilistic weather prediction

Zusammenfassung

Die fundamentale Einsicht, dass die Atmosphäre ein chaotisches System darstellt und deshalb der deterministischen Wettervorhersage intrinsische Grenzen gesetzt sind, führte in den letzten Jahrzehnten zum Aufbau sogenannter Ensemblevorhersagesysteme. Diese versuchen, die Unsicherheiten in den aktuellen Anfangsbedingungen zu erfassen und mit Hilfe einer Vielzahl (das heißt eines Ensembles) von Vorhersagen die daraus hervorgehenden Unsicherheiten in der Prognose zu quantifizieren. Zudem bilden sie die Grundlage für probabilistische Vorhersagen und für die Erstellung darauf basierender Vorhersageprodukte, zum Beispiel des EFI („extreme forecast index“). Dieser einführende Artikel gibt einen kurzen Abriss einiger historischer Aspekte und der wichtigsten Konzepte der probabilistischen Wettervorhersage.

Summary

The chaotic nature of atmospheric flows imposes intrinsic limits to deterministic weather forecasting. This led to the establishment of so-called ensemble prediction systems during the last decades. These systems aim at capturing the uncertainty of the current initial conditions and, based upon an ensemble of predictions, to quantify the resulting forecast uncertainty. Furthermore, they are the basis for probabilistic weather predictions and for producing specific forecast products like, for instance, the „extreme forecast index“ EFI. This introductory article recapitulates some historical aspects and introduces the main concepts of probabilistic weather prediction.

1 Einleitung

Mathematisch betrachtet ist die Wettervorhersage ein klassisches Anfangswertproblem. Falls die Anfangsbedingungen (das heißt der dreidimensionale Zustand der Atmosphäre zu einem gewissen Zeitpunkt) und die Gleichungen, welche die physikalischen Gesetze in der Atmosphäre beschreiben, bekannt sind, kann die zukünftige Entwicklung des Wettergeschehens im Prinzip durch Vorwärtsintegration der Gleichungen bestimmt werden. Genau so wird bei der klassischen deterministischen Wettervorhersage mit numerischen Modellen auch verfahren: der Anfangszustand wird mit Hilfe einer Methode der Datenassimilation aus den vorliegenden Messungen und einer früheren Modellvorhersage bestimmt. Die partiellen Differentialgleichungen werden diskretisiert, das heißt mit Hilfe numerischer Methoden in Differenzgleichungen übersetzt und mit Hilfe eines Supercomputers auf einem fein aufgelösten Gitter schrittweise in der Zeit vorwärts integriert. Dieses Verfahren führt zu einer numerischen Wetterprognose, deren Genauigkeit von der Qualität der Erfassung des Anfangszustands, der numerischen Methoden und der mit der Rechenleistung verbundenen Auflösung des Modells abhängt.

Lewis Fry Richardson (1881-1953) hatte bereits 1922 die Vision dieser Methode der Wettervorhersage (RICHARD-

SON 1922) – Jahrzehnte vor der Entwicklung des ersten Supercomputers ENIAC, der 1949 sogleich für numerische Wettervorhersagen mit einem stark vereinfachten Gleichungssystem der atmosphärischen Bewegungen genutzt wurde (CHARNEY et al. 1950). Eine sehr empfehlenswerte ausführliche Darstellung der Anfänge der numerischen Wettervorhersage ist in LYNCH (2006) zu finden. In den folgenden Jahrzehnten wurden die Komponenten dieses deterministischen Vorhersagesystems, das heißt die numerischen Methoden, die Parametrisierungen der kleinräumigen physikalischen Prozesse, sowie die Techniken der Datenassimilation schrittweise verbessert, was – unterstützt durch die rasante Zunahme der verfügbaren Rechenleistung – zu im Durchschnitt immer besseren Wettervorhersagen führte (Abbildung 1-1). Zu Beginn dieser Entwicklung in den 1960er Jahren herrschte die Vorstellung, dass die Qualität der Vorhersagen durch unvermeidbare Modellfehler und durch Fehler in den Anfangsbedingungen beschränkt ist; dass man durch Fortschritte bei diesen Problemen den Zeitraum für den nützliche Vorhersagen möglich sind (das heißt den sogenannten Vorhersagehorizont) grundsätzlich immer weiter in die Zukunft verschieben könne.

Bahnbrechende Arbeiten von Edward N. Lorenz (1917-2008) führten dann bald zur Erkenntnis, dass der Qualität der Wet-

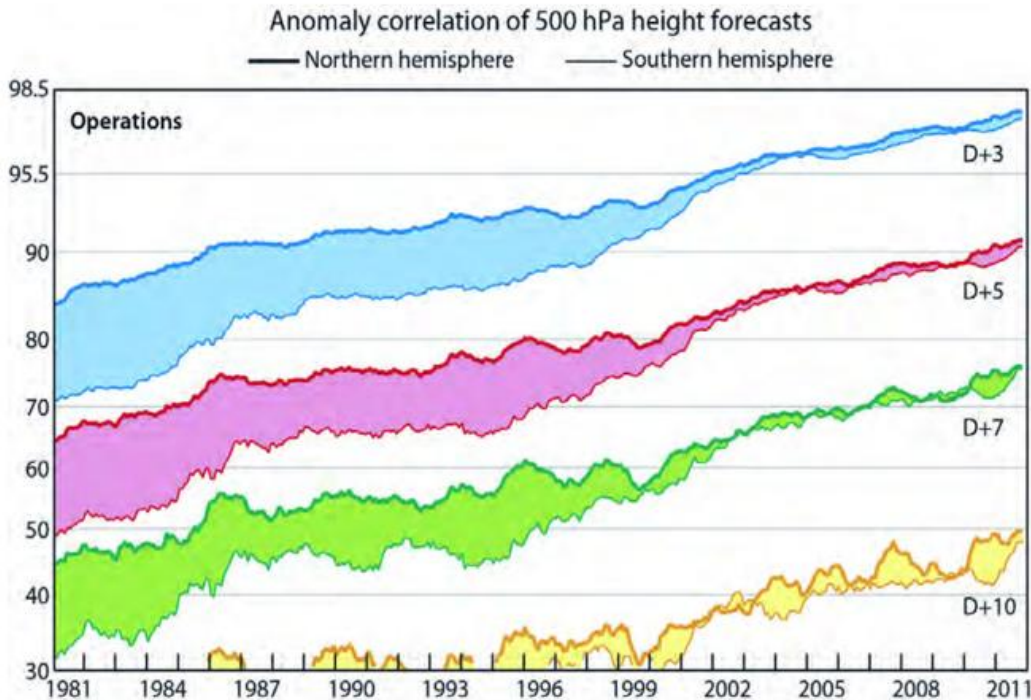


Abb. 1-1: Zeitliche Entwicklung der Qualität der Vorhersage des Modells des EZMW von 1981 bis 2011 (angepasst und erweitert von Simmons & Hollingsworth, EZMW).

tervorhersage durch die nichtlineare Natur der zugrundeliegenden Gleichungen fundamentale Grenzen gesetzt sind (siehe Abschnitt 2.1). Dies führte zu Beginn der 1990er Jahre zu einem eigentlichen Paradigmenwechsel: Die deterministischen Vorhersagen wurden durch ein sogenanntes Ensemblevorhersagesystem ergänzt, mit dem es möglich wurde, probabilistische Prognosen zu erstellen. Das Ziel dieses einführenden Artikels ist es, die theoretischen Hintergründe dieses Paradigmenwechsels anschaulich zu erklären und einige mögliche Anwendungen von Wahrscheinlichkeitsvorhersagen aufzuzeigen. Für einen vertieften und detaillierten Einblick in diese Thematik ist das Buch von PALMER und HAGEDORN (2006) sehr zu empfehlen.

2 Theoretischer Hintergrund

2.1 Die Entdeckung des Chaos

Edward N. Lorenz fand 1963 durch die Untersuchung eines stark vereinfachten Gleichungssystems zur Behandlung der Bénard-Konvektion, dass auch im Falle eines perfekten Modells und beliebig kleiner Fehler in den Anfangsbedingungen eine Grenze der Vorhersagbarkeit besteht. Diese Grenze hat ihre Ursache also nicht in im Prinzip verbesserbaren Unzulänglichkeiten des Systems (Modell- und Messfehler), sondern in der grundlegenden chaotischen Natur des Systems. Dies war der Ursprung einer neuen Wissenschaft, der Chaostheorie, die bei allen naturwissenschaftlichen und technischen Systemen von Bedeutung ist, die durch stark nichtlineare Gleichungen beschrieben werden. Man kann die „Entdeckung“ des Chaos als dritte Revolution in der Physik

des 20. Jahrhunderts betrachten, nach der Relativitätstheorie und der Quantenphysik.

Die Kerneigenschaft von chaotischen Systemen ist die, dass beliebig kleine Unterschiede in den Anfangsbedingungen nach einer gewissen Zeit zu sehr großen Abweichungen führen können. Man spricht von Sensitivität bezüglich der Anfangsbedingungen. Lorenz fand diesen Effekt zufällig, indem er beobachtete, dass Rundungsfehler bei den Anfangsbedingungen seines Modells nach einer gewissen Zeit zu völlig verschiedenen Lösungen führten. Für einen Vortrag im Jahr

1972 wählte Lorenz den Titel „Kann der Flügelschlag eines Schmetterlings in Brasilien in Texas einen Tornado auslösen?“ und prägte damit den populären Begriff des „Schmetterlingseffekts“. Eine der Antworten auf diese provokative Frage von Lorenz selbst war, dass der Schmetterling, wenn er denn einen Tornado auslösen, ihn auch genauso gut verhindern könnte. Damit deutete er an, dass kleinste Störungen die Häufigkeit von extremen Wetterereignissen nicht systematisch erhöhen oder erniedrigen können – aber sie können im Prinzip ihre zeitliche Abfolge modifizieren. Es wird damit aber unmöglich, über einen gewissen Zeithorizont hinaus ein nichtlineares System wie

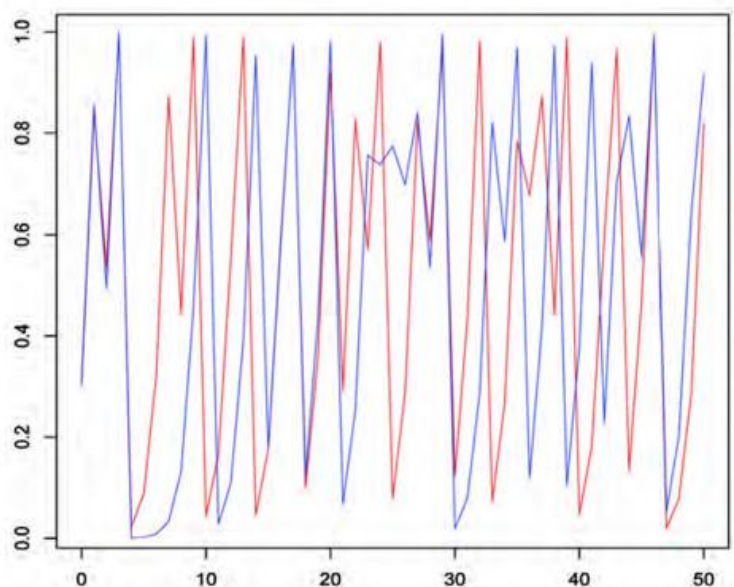


Abb. 1-2: Lösungen der logistischen Gleichung (die ersten 50 Iterationen) für $a = 4$ und den Anfangswerten $x_0 = 0,3$ (rot) und $0,31$ (blau).

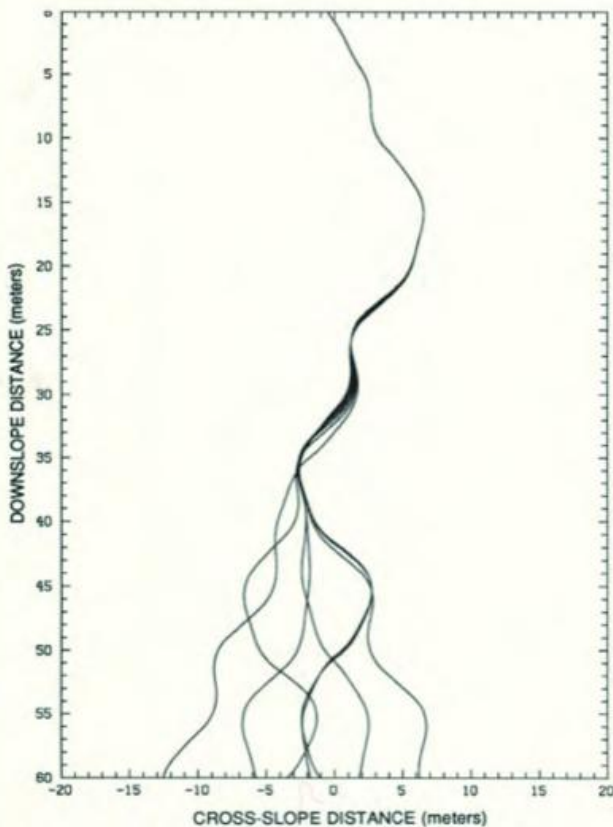


Abb. 1-3: Bahn von 6 Kugeln auf einer Buckelpiste mit leicht unterschiedlichen Anfangspositionen. Aus LORENZ (1993).

die Strömung in der Atmosphäre, das heißt das Wetter, genau vorherzusagen, denn bei der Konstruktion der Anfangsbedingungen „können wir nicht alle Schmetterlinge erfassen“ (siehe Abschnitt 2.3).

2.2 Beispiele chaotischer Systeme

Wir betrachten hier drei einfache Systeme, mit denen das chaotische Verhalten (das heißt die Sensitivität bezüglich der Anfangsbedingungen) gut illustriert werden kann.

Die logistische Gleichung

Diese nichtlineare iterative Gleichung wurde von Pierre Francois Verhulst (1804-1849) eingeführt, um demographische Probleme mathematisch zu behandeln. Die Gleichung für eine Variable x lautet

$$x_{i+1} = a x_i (1 - x_i) \tag{1}$$

wobei a ein freier Parameter des Systems ist. Abhängig von a hat das System ganz andere Eigenschaften: wird zum Beispiel $a = 2,5$ gewählt, so konvergiert die Gleichung zu einem Fixpunkt. Unabhängig von der Anfangsbedingung x_0 ergibt sich nach einigen Iteration ein konstanter Wert von $x_i = 0,6$. Im Falle von $a = 3,25$ hat das System zwei Fixpunkte zwischen denen die Lösung oszilliert, ebenfalls unabhängig von x_0 . Diese Verhaltensmuster sind nicht chaotisch, denn eine kleine Änderung der Anfangsbedingung führt nach kurzer Zeit zur identischen Zeitentwicklung. Für $a = 4$ passieren rätselhafte Dinge, man kommt in den

chaotischen Bereich: Die Zeitentwicklung wird vollständig irregulär, das heißt die Muster wiederholen sich nie. Zudem wird die Zeitentwicklung stark sensitiv bezüglich den Anfangsbedingungen. Die Abbildung 1-2 zeigt die nach wenigen Iterationen stark unterschiedlichen Lösungen, die mit dem selben Wert für a , aber mit leicht unterschiedlichen Anfangswerten $x_0 = 0,3$ beziehungsweise $x_0 = 0,31$ berechnet worden sind. Ein analog chaotisches Verhalten bekommt man auch falls sich die Anfangswerte nur minimal (zum Beispiel um 10^{-4}) unterscheiden. Falls man diese Zeitreihen von x zum Beispiel als Temperaturvorhersagen eines Modells für einen bestimmten Ort betrachtet, so würde man sagen, dass im Modellzustand mit $a = 4$ die Temperaturentwicklung nur für ganz wenige Zeitschritte zuverlässig vorhergesagt werden kann, da wir die Anfangsbedingung unter anderem wegen unvermeidlichen Messfehlern nicht beliebig präzise bestimmen können. Ist jedoch das Modell (beziehungsweise die Atmosphäre) in einem anderen Zustand (zum Beispiel repräsentiert durch $a = 2,5$), dann ist die Anfangsbedingung unwichtig und eine zuverlässige Vorhersage (des stationären Zustands) kann für einen sehr langen Zeitraum gemacht werden. Zu beachten ist, dass hier (wie bei allen chaotischen Systemen) kein „Zufall“ mit im Spiel ist – die logistische Gleichung ist vollständig deterministisch, das heißt wir erhalten für feste Werte von x_0 und a immer die gleiche Zeitreihe – und dennoch ist das System unter gewissen Bedingungen chaotisch.

Die Buckelpiste

Dieses sehr anschauliche Beispiel stammt aus dem sehr lehrreichen Buch „The essence of chaos“ (LORENZ 1993). Das Herunterrollen der Kugeln auf einer 60 m langen Buckelpiste wird numerisch berechnet und die Bahnen von 6 Kugeln, deren Anfangspositionen sich nur um 1 mm unterscheiden, ist in der Abb. 1-3 gezeigt. Am unteren Ende der Buckelpiste sind die Kugeln bis zu 15 m voneinander entfernt (das ist mehr als 10 000 Mal der anfängliche Abstand!). Dieses Beispiel zeigt eindrücklich wie in einem „einfachen“, vollständig deterministischen System kleine Unterschiede in den Anfangsbedingungen in kurzer Zeit zu sehr grossen Differenzen anwachsen können. Das Rollen von Kugeln auf einer Buckelpiste hat natürlich wie das erste Beispiel nichts mit der Physik der atmosphärischen Strömungen zu tun. Dennoch ist es eine sehr nützliche Illustration des Einflusses des Chaos auf die Wettervorhersage: Wir können uns vorstellen, dass eine Kugel, die am Ende der Piste links ankommt eine Prognose für „schönes Wetter“ darstellt, und eine Kugel, die rechts unten ankommt „unfreundliches Wetter“. Da wir das aktuelle Wetter beim Start der Vorhersage nicht beliebig genau kennen (siehe Abschnitt 2.3), wissen wir übertragen auf das Beispiel der Buckelpiste nicht auf den Millimeter genau, wo wir die Kugel am oberen Ende der Piste loslassen sollen. Abbildung 1-3 zeigt uns dann, dass diese Unsicherheit zu Beginn der Vorhersage (das heißt auf den ersten gut 20 m) kaum eine Rolle spielt (die Bahnen der Kugeln mit leicht unterschiedlichen Anfangsbedingungen sind sehr nahe beieinander), nach einer gewissen Zeit weichen die Bahnen jedoch stark voneinander ab und das Ensemble aller Kugeln deutet mit

etwa gleicher Wahrscheinlichkeit auf schönes beziehungsweise unfreundliches Wetter hin. Eine deterministische Vorhersage (das heißt eine einzige Kugel) kann diese Unsicherheit der Vorhersage nicht erfassen. Das Ensemble der Vorhersagen (beziehungsweise Kugeln) macht es hingegen möglich, sowohl Szenarien der Wetterentwicklung zu bestimmen, wie auch den Zeithorizont, ab dem die einzelnen Vorhersagen voneinander abweichen, das heißt die deterministische Vorhersage unsicher wird.

Das Lorenz-Modell

Mit dem folgenden nichtlinearen Gleichungssystem, das in stark vereinfachter Form die Rayleigh-Bénard-Konvektion beschreibt, hat Lorenz einen großen Teil seiner berühmten Untersuchungen durchgeführt:

$$\begin{aligned} dx/dt &= a(y - x) \\ dy/dt &= -xz + cx - y \\ dz/dt &= xy - bz. \end{aligned} \quad (2)$$

Das System hat 3 Variablen (x, y, z) und 3 wählbare Modellparameter (a, b, c). Nichtlinearitäten treten in der zweiten und dritten Gleichung auf (Terme $-xz$ beziehungsweise xy). Dieses System ist chaotisch zum Beispiel für die folgende Wahl der Modellparameter: $a = 10$, $b = 8/3$ und $c = 28$. Dies ist in Abbildung 1-4 für die Variable x illustriert: Zwei leicht verschiedene Anfangsbedingungen führen nach einiger Zeit zu völlig unterschiedlichen Lösungen, analog zum Beispiel der logistischen Gleichung. Es ist aber auch zu beachten, dass sich die beiden Lösungen zuerst für eine bestimmte Zeit sehr ähnlich verhalten. Für diese Zeitperiode scheint es möglich zu sein, eine präzise Vorhersage zu machen, auch wenn der Anfangszustand nicht exakt bekannt ist.

Abbildung 1-5 zeigt die zeitliche Entwicklung von zwei kugelförmigen Anfangsregionen im von x, y und z aufgespannten Phasenraum (die dicken schwarzen Linien) und illustriert dabei zwei weitere wichtige Eigenschaften des Systems:

- Die Sensitivität bezüglich den Anfangsbedingungen ist im Phasenraum nicht homogen (im Bild links bleiben die Punkte kompakt: die Entwicklung ist über einen längeren Zeitraum vorhersagbar; im Bild rechts kommt es zu einer starken Dispersion der Trajektorien im Phasenraum: das System ist bereits nach kurzer Zeit nicht mehr vorhersagbar).
- Im Falle der starken Dispersion im Phasenraum wird

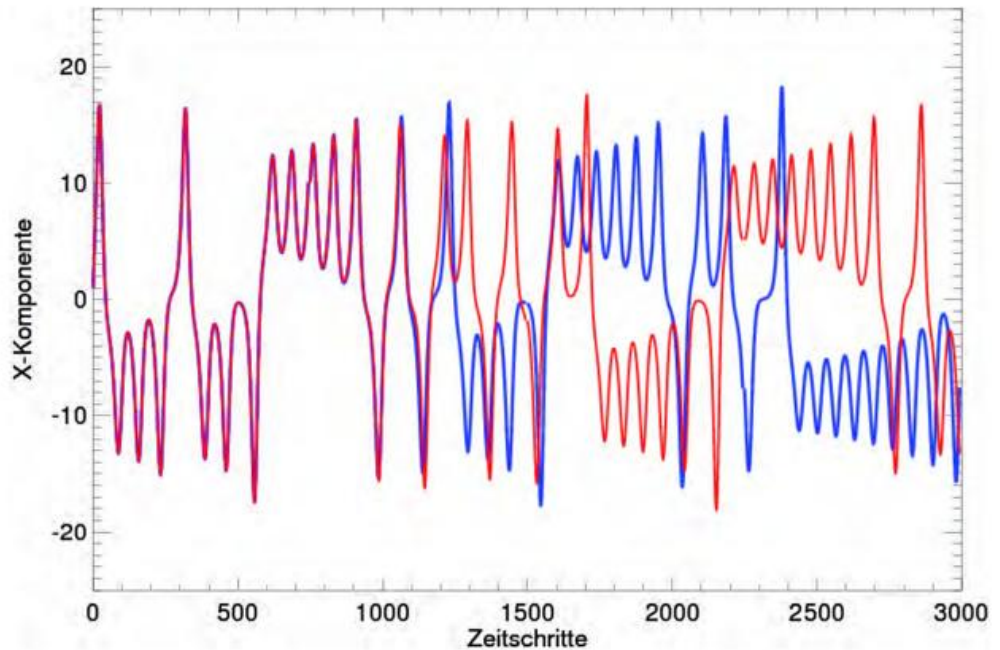


Abb. 1-4: Zeitentwicklung der x -Komponente des Lorenz-Modells für zwei leicht unterschiedliche Anfangsbedingungen (rote und blaue Linien).

die zuerst kugelförmige Struktur zu einem Ellipsoid deformiert und dann, da der Attraktor des Systems räumlich beschränkt ist, stark gekrümmt und zu einem langen Filament verzerrt, das nach langer Zeit den ganzen Attraktor ausfüllt (nicht mehr dargestellt). Die Geschwindigkeit mit welcher die Kugel in der anfänglichen linearen Phase in die Länge gezogen wird, ist bestimmt durch den größten Lyapunov-Exponenten. Ganz allgemein beschreibt der Lyapunov-Exponent eines dynamischen Systems die Geschwindigkeit, mit der sich zwei (nahe beieinanderliegende) Punkte im Phasenraum voneinander entfernen (positives Vorzeichen) oder annähern (negatives Vorzeichen).

2.3 Das Problem der Anfangsbedingungen bei der Wettervorhersage

Es wurde bereits mehrfach erwähnt, dass die Anfangsbedingungen, die wir für eine Wettervorhersage benötigen, nicht beliebig exakt bestimmt werden können. Dafür ist eine Mehrzahl von Gründen verantwortlich, die im Folgenden kurz skizziert werden. Für eine ausführliche Behandlung dieses wichtigen Themas siehe WERGEN (2002).

- Die Anfangsbedingungen bestehen aus einer enormen Anzahl von Zahlenwerten, welche den atmosphärischen Variablen an allen Gitterpunkten des Vorhersagemodells entsprechen. Für ein globales Modell mit einer horizontalen Maschenweite von 100 km und 50 vertikalen Schichten und bei sechs dreidimensionalen prognostischen Variablen (Druck, Temperatur, Feuchte und drei Windkomponenten) ergeben sich beispielsweise global mehr als 10^7 Zahlenwerte. Idealerweise würden diese Werte alle durch direkte Messungen bestimmt werden. Dies ist jedoch unmöglich, da das vorhandene

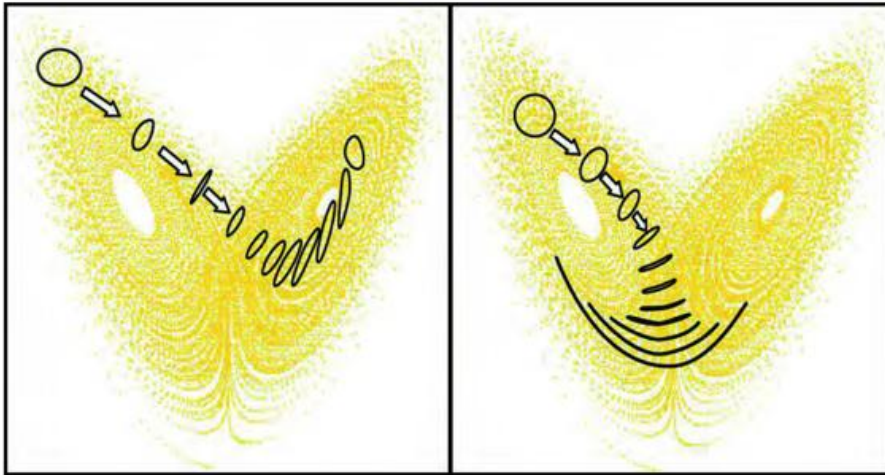


Abb. 1-5: Die Entwicklung von zwei kugelförmigen Anfangsregionen im Phasenraum des Lorenz-Modells für eine Situation mit hoher Vorhersagbarkeit (links) und eine Situation mit sehr niedriger Vorhersagbarkeit (rechts).

Messnetz es nicht erlaubt, zum Beispiel mitten über dem Nordatlantik die Windgeschwindigkeit auf einer Höhe von 10 km zu messen. Solche Messungen sind prinzipiell mit Hilfe von Radiosonden möglich, die jedoch über den Ozeanen nur sehr spärlich vorhanden sind (siehe Abbildung 1-6). Es sei hier nur am Rande vermerkt, dass die Anfangsbedingungen in Regionen, in denen keine Messungen vorhanden sind, in erster Linie mit Hilfe einer kurzfristigen Vorhersage abgeschätzt werden. Die Technik zur Produktion möglichst exakter Anfangsbedingungen heißt Datenassimilation und ist ein zentrales Element der Wettervorhersage, bei dem in den letzten gut 10 Jahren sehr große Fortschritte gemacht worden sind.

- Messungen sind fehlerbehaftet. Auch wenn für jeden Parameter an jedem Gitterpunkt eines Modells ein Messgerät zur Verfügung stünde, könnten die Anfangsbedingungen nicht exakt bestimmt werden. Die Messfehler sind teilweise systematisch (gewisse Feuchtesensoren haben einen Bias bei sehr tiefen Temperaturen) und in erster Linie zufällig. Die Fehler können deshalb in den meisten Fällen nur in einem statistischen Sinne quantifiziert werden.
- Satelliten messen nicht direkt die Größen, die für die Wettervorhersage benötigt werden. Es ist naheliegend zu denken, dass mit Hilfe der Satelliten die Anfangsbedingungen sehr präzise und mit hoher Auflösung bestimmt werden können. Satelliten messen jedoch Strahlungswerte (in Wm^{-2}) und damit keine der Modellvariablen Temperatur, Druck, Feuchte oder Wind. Es braucht deshalb spezielle Techniken, um aus diesen Strahlungswerten Modellparameter ableiten zu können (wiederum mit Hilfe moderner Datenassimilationsverfahren). Zudem erfassen die meisten Satelliten nur ein vertikales Integral entlang einem Profil durch die Atmosphäre. Es müssen vereinfachende Annahmen gemacht werden, um aus diesen integrierten Werten vertikal aufgelöste Profi-

le zum Beispiel von Temperatur und Feuchte erstellen zu können.

2.4 Die Idee der Ensemblevorhersagen

Das nichtlineare Gleichungssystem, das die Bewegungen der Atmosphäre beschreibt und in den Wettervorhersagemodellen numerisch umgesetzt ist, hat wie die im Abschnitt 2.2 beschriebenen einfachen dynamischen Systeme chaotische Eigenschaften. Dies bedeutet, dass eine Vorhersage mit einem numerischen Wettervorhersagemodell stark sensitiv ist bezüglich des Anfangszustands. Dieser ist aber, wie in Abschnitt 2.3 erläutert,

nur näherungsweise bekannt. Die Unsicherheiten in den Anfangsbedingungen können nach wenigen Tagen zu großen Fehlern in der Vorhersage führen, auch im Falle eines perfekten Modells. Es ist deshalb die Idee entstanden, dass man nicht nur eine einzige (sogenannt deterministische) Vorhersage rechnet, sondern ähnlich wie in der Abbildung 1-5 für das Lorenz-Modell gezeigt, ein ganzes Ensemble von Vorhersagen, die von leicht unterschiedlichen, gleichermaßen realistischen, Anfangsbedingungen ausgehen. Die Dispersion dieser Vorhersagen ist dann ein Maß für die Robustheit der Vorhersage: sind alle Ensemblevorhersagen sehr ähnlich, so ist man in der in Abbildung 1-5 links dargestellten Situation und die Wetterentwicklung ist mit einer gewissen Sicherheit über mehrere Tage vorhersagbar. Unterscheiden sich die Ensemblevorhersagen jedoch schon nach wenigen Tagen, so ist dies ein Anzeichen für eine unsichere Vorhersage. Die Unterschiede in den verschiedenen Mitgliedern des Ensembles zeigen aber nicht nur an, ab wann die Vorhersage unsicher wird, sie zeigen auch verschiedene mögliche Entwicklungen (Szenarien) auf. Zudem bilden solche Ensemblevorhersagen die Basis für probabilistische Vorhersagen. Falls sich die einzelnen Vorhersagen zum Beispiel gleichmäßig auf drei Szenarien „aufteilen“, so kann das Eintreten jedes dieser drei Szenarien mit einer Wahrscheinlichkeit von 33 % vorhergesagt werden. Häuft sich jedoch die Mehrheit aller Ensemblevorhersagen bei einem Szenario, so kann dieses als das wahrscheinlichste prognostiziert werden.

2.5 Herausforderungen der Ensemblevorhersagen

Die praktische Konstruktion des Ensembles ist jedoch nicht trivial und erfordert aufwändige Techniken. Beim Lorenz-Modell ist es wegen der geringen Anzahl an Modellvariablen einfach, ein Ensemble zu rechnen, in dem die Anfangsbedingung für jede der drei Variablen leicht variiert wird. Ein Atmosphärenmodell hat jedoch mehr als 10^7 prognostische Variablen, so dass es unmöglich ist, jede dieser Variablen einzeln zu variieren – denn mehr als 100 Ensemblevorhersagen können wegen der beschränkten Rechenzeit nicht gerechnet werden. Falls man

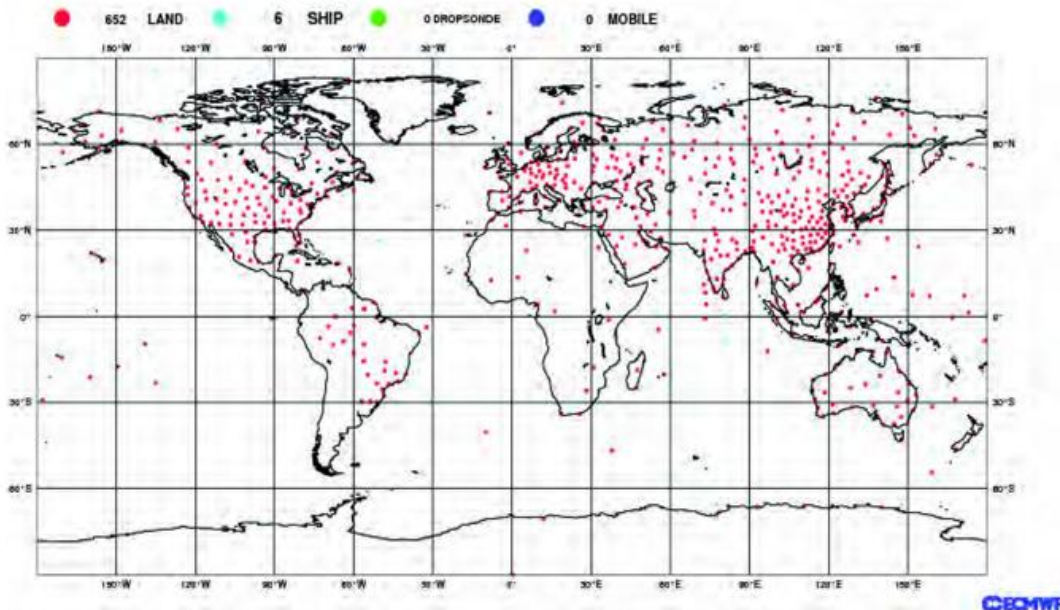


Abb. 1-6: Dichte des globalen Netzes der Radiosonden. An den eingezeichneten Orten wird mindestens zweimal täglich ein Ballon mit einer meteorologischen Sonde steigen gelassen, die die vertikalen Profile von Temperatur, Druck, Wind und Feuchte bis auf eine Höhe von etwa 20 km erfasst.

zufällig von den 10^7 Variablen einige auswählt, die zur Anfangszeit variiert werden, so besteht die Gefahr, dass das Ensemble die Dispersion unterschätzt, was zur irreführenden Annahme einer sehr gut vorhersagbaren Wetersituation führen könnte. Es ist deshalb sehr wichtig, dass man zuerst mit einer objektiven Methode diejenigen Störungen der Anfangsbedingungen bestimmt, die in den nächsten Tagen zu maximal starken Abweichungen von der Kontrollsimulation führen (siehe LEUTBECHER, dieses Heft). Nur so kann mit einem kleinen Ensemble von weniger als 100 Trajektorien im Phasenraum die Vorhersagbarkeit eines Systems mit 10^7 Variablen einigermaßen realistisch abgeschätzt werden. Für die Identifikation dieser schnellstwachsenden Störungen wurden sehr unterschiedliche Techniken entwickelt, zum Beispiel am EZMW die Technik der Singulärvektoren. Es ist bis heute umstritten, mit welcher Technik das beste Ensemble erzeugt werden kann, ganz besonders bei der Produktion von Ensembles mit regionalen Wettervorhersagemodellen (siehe THEIS et al., dieses Heft). Weitere grosse Herausforderungen ergeben sich bei der Verifikation von Ensemblevorhersagen (siehe WEIGEL, dieses Heft) und der Kommunikation von probabilistischen Prognosen (siehe DENHARD, dieses Heft).

3 Probabilistische Vorhersagen: Beispiele und Möglichkeiten

In diesem abschließenden Abschnitt sollen kurz vier Beispiele präsentiert werden, die illustrieren, wie probabilistische Vorhersagen dargestellt und interpretiert werden können. Es handelt sich hier um eine etwas zufällige Auswahl; weitere Beispiele und vertiefte Diskussionen finden sich in den weiteren Artikeln dieses Hefts. Alle hier gezeigten Beispiele stammen vom Ensemblevorhersagesystem des EZMW („Europäisches Zentrum für mittelfristige Wettervorhersage“).

Als erstes zeigt die Abbildung 1-7 eine sogenannten „Briefmarkenplot“. Ein ausgewähltes Feld (hier der Bodendruck über Westeuropa zum Zeitpunkt des Durchzugs des Wintersturms „Lothar“ im Dezember 1999) wird für jede der 50 Ensemblevorhersagen in kleinem Format dargestellt. Diese Visualisierung ist nicht besonders innovativ (das heißt analog zur Visualisierung deterministischer Vorhersagen) und sie überlässt die Synthese

und Interpretation vollständig dem Betrachter. Dieser ordnet „im Kopf“ die ähnlichen Muster und kommt zum Schluss, dass etwa 10 der 50 Vorhersagen ein kleines, aber äußerst intensives Tiefdruckgebiet über Frankreich prognostizieren. Dies bedeutet, dass auf Grund dieses Ensembles ein solches Ereignis (das heißt das Auftreten des Wintersturms „Lothar“) mit einer Wahrscheinlichkeit von 20% vorhergesagt werden kann. Die kleinen Abbildungen zeigen jedoch auch mindestens 20 Vorhersagen ohne Tiefdruckgebiet – was die sehr große Dispersion der möglichen Szenarien in diesem Beispiel illustriert. Interessanterweise hatte die damalige deterministische Vorhersage des EZMW den Sturm „Lothar“ nicht vorhergesagt – die gemäß dem Ensemble prognostizierte Wahrscheinlichkeit von 20% des Eintretens eines starken Sturms wäre demnach ein sehr hilfreicher Anhaltspunkt zur Information und Warnung der Bevölkerung gewesen.

Abbildung 1-8 zeigt eine klassische probabilistische Niederschlagsprognose des EZMW. Ausgehend von der um 00 UTC am 17. Januar 2008 gestarteten Ensembleprognose wurde die Wahrscheinlichkeit berechnet, dass am 21. Januar (das heißt im Vorhersagezeitraum zwischen Tag 4 und 5) an einem Ort mehr als 10 mm Niederschlag fallen. Die Darstellung zeigt sogleich, dass an diesem Tag eine solche Niederschlagsmenge zum Beispiel im Mittelmeerraum sehr unwahrscheinlich ist, während sie über dem Norden der Britischen Inseln mit einer Wahrscheinlichkeit von 35 bis 65% vorhergesagt wird. Diese Wahrscheinlichkeiten wurden berechnet, in dem an jedem Ort gezählt wurde, wie viele der 50 Ensemblevorhersagen mehr als 10 mm Niederschlag prognostizieren. Die möglichen Werte in diesen Karten sind damit 0, 2, 4, ..., 100%. Diese Karten sind recht leicht interpretierbar und fassen die Ergebnisse von 50 Ensemblesimulationen kompakt zusammen. Es muss zur Erstellung jedoch zunächst ein Schwellenwert bestimmt werden (hier 10

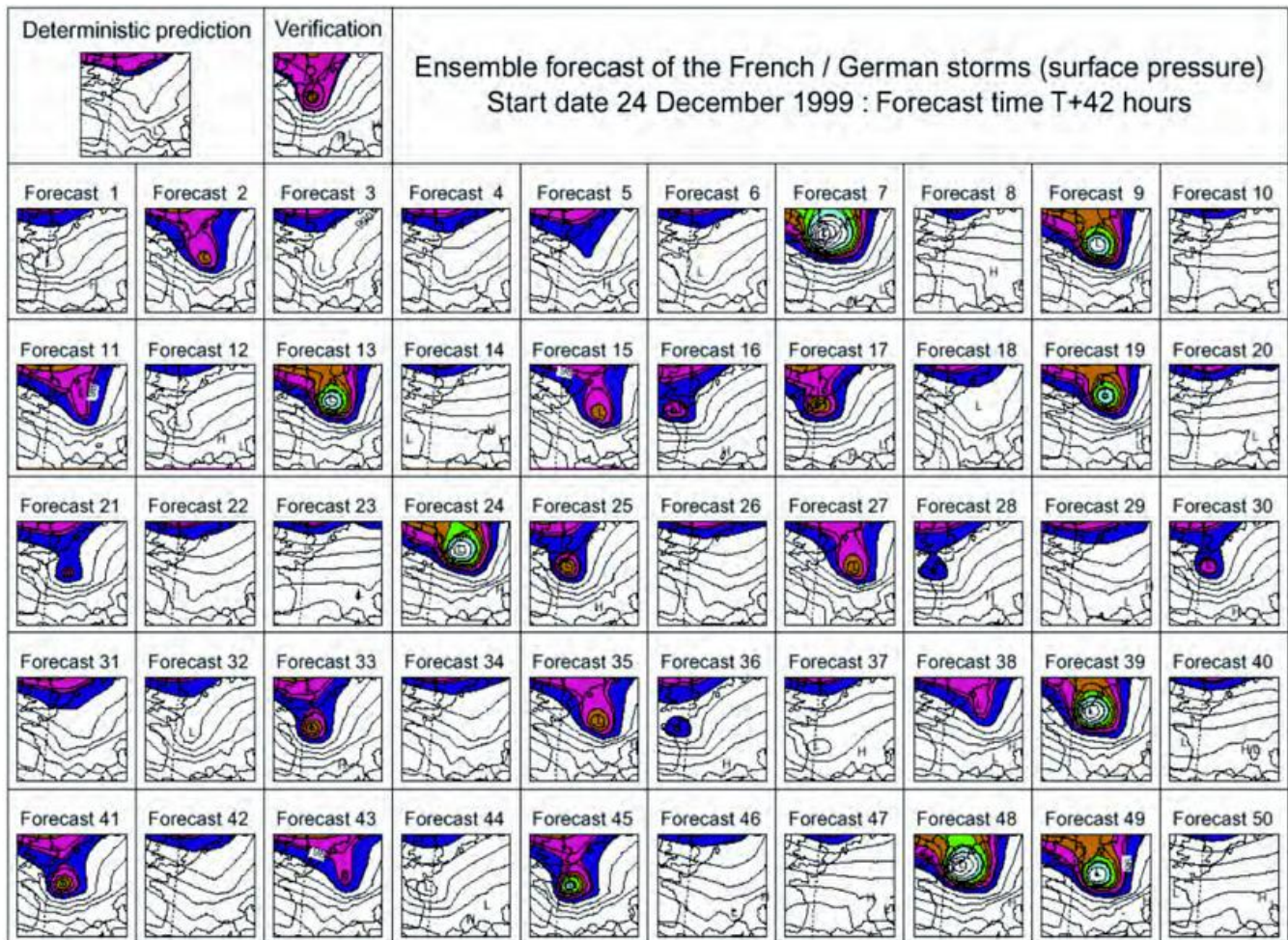


Abb. 1-7: „Briefmarkenplot“ mit der Vorhersage der 50 Ensemblevorhersagen des EZMW für den Bodendruck des Sturms „Lothar“ um 06 UTC am 26. Dezember 1999. Tiefe Druckwerte sind in Farbe dargestellt, extrem tiefe Werte (zum Beispiel Forecast 9) als weiße „Augen“ in den farbigen Flächen.

mm Niederschlag in 24 Stunden), was bedeutet, dass für verschiedene Schwellenwerte mehrere Karten produziert werden müssen. Während ein gewisser Schwellenwert an einem bestimmten Ort bereits als extrem eingestuft werden kann, ist der selbe Schwellenwert an einem anderen Ort noch im normalen Bereich. Deshalb wurden verfeinerte Techniken entwickelt, um in einer Karte die Wahrscheinlichkeit extremer Niederschlägen oder Windböen darstellen zu können.

LALAURETTE (2003) entwickelte eine neue Methode zur Identifikation von sehr seltenen (und damit extremen) Wetter-situationen aus probabilistischen Vorhersagen. Als Grundlage der Methode dient eine Klimatologie des betrachteten Parameters (zum Beispiel

der Böenspitze oder des über einen Tag akkumulierten Niederschlags) an jedem Gitterpunkt des Modells, basierend auf Messdaten der vergangenen Jahre. Diese Klimatologie bestimmt lokal, welche Intensität eines Ereignisses als extrem (da sehr selten) zu betrachten ist.

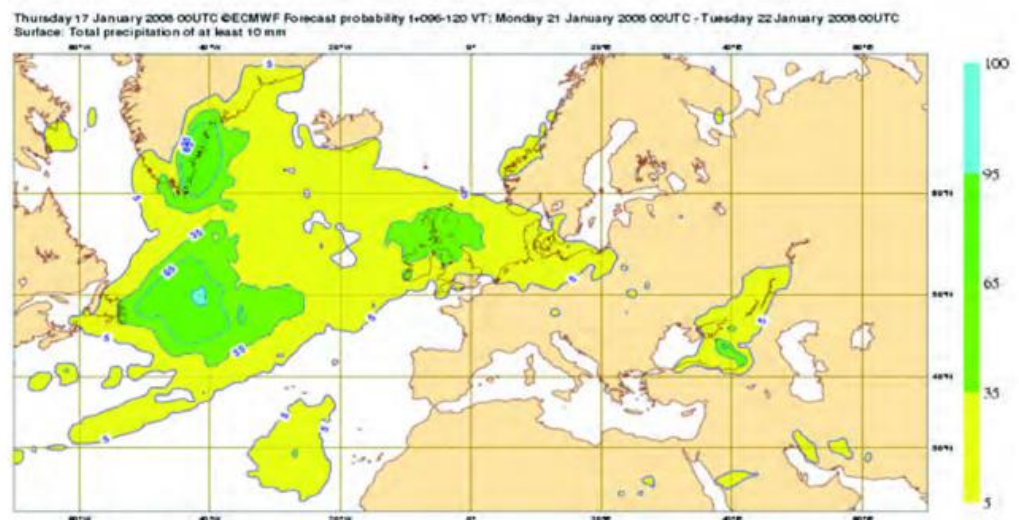


Abb. 1-8: Probabilistische Niederschlagsprognose des EZMW. Dargestellt ist die Wahrscheinlichkeit, dass der akkumulierte Niederschlag am 21. Januar 2008 einen Wert von 10 mm überschreitet – ausgehend von der Ensembleprognose, die um 00 UTC am 17. Januar 2008 gestartet wurde.

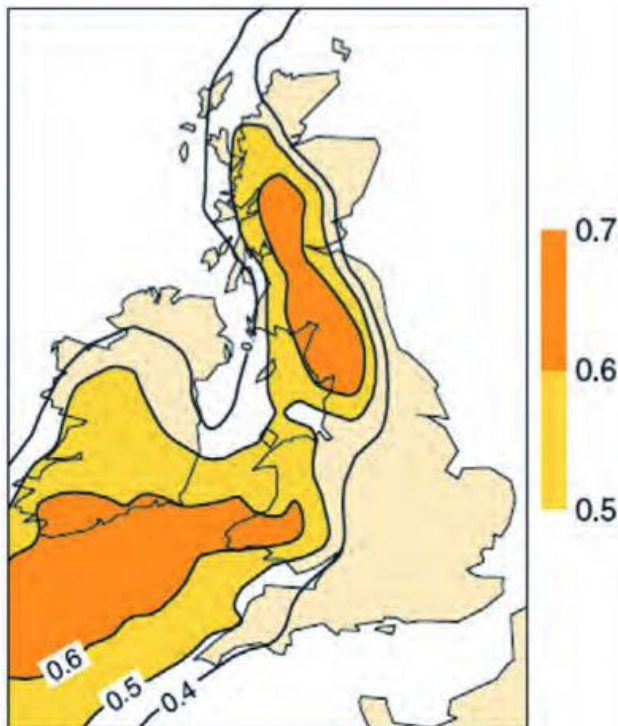


Abb. 1-9: EFI-Vorhersage für die Flut in England am 19. November 2009, basierend auf der um 00 UTC am 15. November 2009 gestarteten Ensemblevorhersage. Aus RICHARDSON (2011).

Der „extreme forecast index“ (EFI) quantifiziert dann an jedem Ort, wie stark die aktuelle Verteilung der 50 Ensemblevorhersagen für den betrachteten Parameter von der klimatologischen Verteilung abweicht. Der Index ist zwischen -1 und $+1$ normiert. Dabei bedeutet ein Wert von $+1$, dass alle Ensemblevorhersagen einen Wert prognostizieren, der grösser ist als der größte Wert in der Klimatologie. Dies bedeutet, dass alle Ensemblemitglieder einvernehmlich das Eintreten eines Rekordmaximums prognostizieren. Die Überlegung ist analog bei einem Wert von -1 für ein Rekordminimum. Ein Wert von 0 bedeutet hingegen, dass die Vorhersage des Ensembles im Vergleich zur Klimatologie in keiner Weise auffällig (das heißt „normal“) ist. In der Anwendung hat sich gezeigt, dass Werte des EFI $> 0,5$ oft auf das Eintreten von außergewöhnlichen und potentiell schadenverursachenden Ereignissen hinweisen. Ein Beispiel dafür ist in Abbildung 1-9 dargestellt, für die Flut im Norden Englands und Teilen Schottlands am 19. November 2009. Hier wurde das Extremereignis vom probabilistischen Vorhersagesystem vier Tage vor dem Eintreten der Flut durch einen hohen EFI korrekt prognostiziert.

Ein letztes Beispiel illustriert die probabilistische Zugbahnvorhersage von tropischen Wirbelstürmen (Abbildung 1-10). Im Gegensatz zu den bisher gezeigten Darstellungen wird hier ein spezifisches Wetterphänomen (Hurrikan „Irene“ im westlichen Nordatlantik) ausgewählt und allein seine Zugbahn betrachtet. Weitere Felder (Niederschlag, Windböen)

werden dabei zunächst außer Acht gelassen. Die Abbildung zeigt die Zugbahn der deterministischen Vorhersage (in schwarz) und der 50 Ensemblevorhersagen (in blau). Zudem ist die Dichte der Ensemble-Zugbahnen in Farbe dargestellt. Solche Vorhersagen von an Land laufenden Wirbelstürmen sind von größter gesellschaftlicher Bedeutung; sie dienen als Grundlage zur Warnung der Bevölkerung und eventuell für die Erstellung von Evakuierungsplänen. Die deterministische Vorhersage liefert nur ein Szenario (das in diesem Fall etwas zu weit östlich verlief) – die probabilistische Vorhersage hingegen liefert ein relativ breites Spektrum an möglichen Entwicklungen (Wirbelsturm geht bereits in North Carolina an Land, oder erst bei New York/Boston, oder sogar erst bei Nova Scotia) mit keiner deutlichen Präferenz für eines der drei Szenarien. Tatsächlich eingetreten ist in etwa das erste dieser Szenarien, so dass hier das probabilistische Vorhersagesystem einige Tage vor dem Eintreten des Ereignisses in der Lage war, die ungefähre Zugbahn des Hurrikans im Osten der USA mit einer gewissen Wahrscheinlichkeit zu prognostizieren.

Diese Beispiele illustrieren die breite Palette an neuen, probabilistischen Vorhersageprodukte, die in den letzten Jahren, nach Einführung der Ensemblevorhersagesysteme entstanden sind. Weitere Anwendungen werden im Detail in weiteren Artikeln dieses Hefts präsentiert (zum Beispiel THORARINSDOTTIR et al., Beitrag 5 in diesem Heft). Diese Produkte stellen eine wissenschaftliche Antwort dar auf Lorenz Erkenntnis, dass der deterministische Vorhersagehorizont der Atmosphäre grundsätzlich begrenzt ist und ermöglichen einen quantitativen Umgang mit den intrinsischen Unsicherheiten nichtlinearer Wetterprognosen. DENHARD (Beitrag 8 in diesem Heft) schildert die Herausforderungen für die Nutzer, die Ergebnisse dieser probabilistischen Vorhersagen zu interpretieren und effektiv zu nutzen.

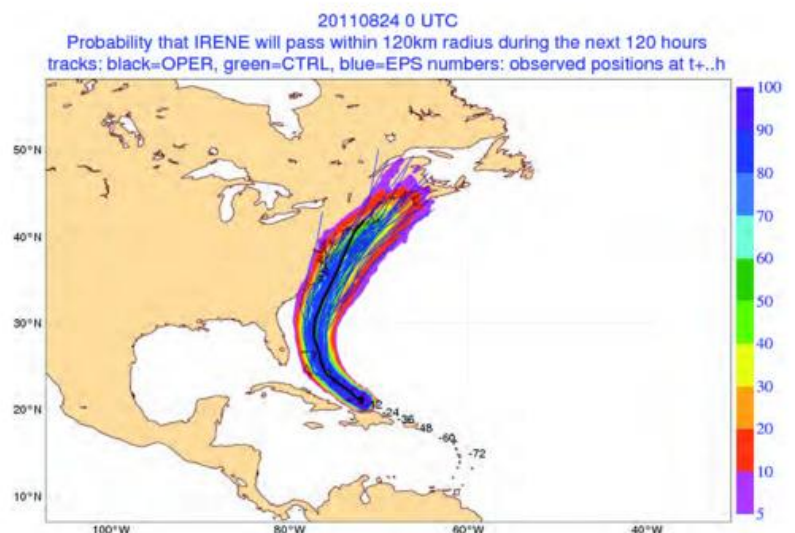


Abb. 1-10: Ensemblevorhersage des EZMW vom 24. August 2011 der Zugbahn des tropischen Wirbelsturms Irene. Details siehe Text.

Literatur

- CHARNEY, J. G., FJORTOFT, R., VON NEUMANN, J., 1950: Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237–254.
- LALAURETTE, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteorol. Soc.* **129**, 3037–3057.
- LORENZ, E. N., 1993: The essence of chaos. University of Washington Press, 227 Seiten.
- LYNCH, P., 2006: The Emergence of Numerical Weather Prediction: Richardson's Dream. *Cambridge University Press*, 290 Seiten.
- PALMER, T., HAGEDORN, R. (Hrsg.), 2006: Predictability of weather and climate. *Cambridge University Press*, 718 Seiten.
- RICHARDSON, D., 2011: From observations to forecasts – Part 11: Ensemble products for weather forecasters. *Weather* **66**, 235-241.
- RICHARDSON, L. F., 1922: Weather Prediction by Numerical Process. *Cambridge University Press*.
- WERGEN, W., 2002: Datenassimilation – ein Überblick. *Promet* **27** (3/4), 142-149.

M. LEUTBECHER

2 Anfangsbedingungen für Ensemblevorhersagen

Initial conditions for ensemble forecasts

Zusammenfassung

Anfangsbedingungen für Wettervorhersagen werden aus Beobachtungsdaten und Kurzfristvorhersagen geschätzt und können deshalb nicht perfekt sein. Ensemblevorhersagen müssen diese Unsicherheit der Anfangsbedingungen repräsentieren. Die Bedeutung quantitativ korrekter Beschreibungen von Unsicherheiten des Anfangszustandes wird zunächst anhand von einfachen niedrigdimensionalen dynamischen Systemen illustriert. Anschließend folgt ein kurzer Überblick über gegenwärtig in operationellen Systemen verwendete Verfahren zur Gewinnung der Anfangsbedingungen von Ensemblevorhersagen.

Summary

Initial conditions for numerical weather prediction are estimated from observational data and short-range predictions. Therefore, they are imperfect. Ensemble forecasts need to represent these uncertainties of the initial conditions. The relevance of quantitatively accurate representations of the initial condition uncertainties is illustrated with simple low-dimensional dynamical systems. Then, a short overview follows on currently operationally used methods to generate initial conditions for ensemble forecasts.

1 Einleitung

Die Vorhersage des Wetters mit einem numerischen Modell kann man als Anfangswertproblem ansehen. Falls man ein regionales Modell verwendet, sind auch die lateralen Randbedingungen entscheidend. Diese sind wiederum von den Anfangsbedingungen des globalen Modelles abhängig, welches die Randbedingungen liefert. Die Bedeutung der Unsicherheit der Anfangsbedingungen wird offensichtlich, wenn man die Dynamik der Störungen von Lösungen der Modellgleichungen untersucht. Hierbei, werden wir als eine Störung die Differenz zwischen einer Referenzvorhersage und einer anderen Vorhersage mit geringfügig abweichenden Anfangsbedingungen bezeichnen. Die nicht-lineare Dynamik der Atmosphäre hat zur Folge, dass die Amplitude von Störungen im Laufe der Vorhersage typischerweise zunimmt. So können geringe Unsicherheiten in den Anfangsbedingungen zu großen Unsicherheiten in der Vorhersage führen (WERNLI 2012). Ensemblevorhersagen sind Mitte der 60er Jahre des 20. Jahrhunderts als eine mögliche Entwicklung vorhergesehen worden und Anfang der 90er Jahre sind die ersten operationellen Ensemble-Vorhersagesysteme mit globalen Modellen in Betrieb genommen worden (LEWIS 2005). Die Darstellung der Anfangsbedingungen ist ein essentieller Bestandteil in der

Ensemblevorhersage und sie ist nach wie vor ein aktives Forschungsgebiet.

Die Datenassimilation beschäftigt sich mit der Schätzung der Anfangsbedingungen aus Messdaten und vorangegangenen Kurzfristvorhersagen. Der so geschätzte Modellzustand wird üblicherweise auch als Analyse bezeichnet. WERGEN (2002) gibt einen kurzen Überblick über die Datenassimilation. Das Buch von DALEY (1991) ist eine ausgezeichnete Einführung. Die Unsicherheit der Anfangsbedingungen ist abhängig von der Menge und Güte der Beobachtungsdaten, der Genauigkeit des Vorhersagemodelles und nicht zuletzt vom Assimilationsverfahren. Verbesserungen in den drei Bereichen ermöglichen die Genauigkeit der Anfangsbedingungen zu steigern. Andererseits ist ein gewisses Maß an Unsicherheit unvermeidbar aufgrund von Lücken im Messnetz, Messfehlern und Fehlerwachstum in der Kurzfristvorhersage, die für die nächste Analyse verwendet wird.

Ziel von Ensemblevorhersagen ist es, die Statistik der Vorhersagefehler quantitativ so gut wie möglich zu beschreiben. Wenn man ein perfektes Vorhersagemodell zur Verfügung hätte, würde es ausreichen, die Statistik der Fehler der Anfangsbedingungen korrekt zu beschreiben,

um zuverlässige Ensemblevorhersagen zu gewinnen. Die genaue Definition des Begriffs der Zuverlässigkeit ("reliability") im Zusammenhang mit Ensemblevorhersagen wird ausführlich beschrieben in dem Beitrag von WEIGEL (dieses Heft). Für realistische Wettervorhersagemodelle ist es notwendig, auch den Modellfehler zu berücksichtigen (BERNER 2012). Wir werden uns im Folgenden auf die Fehler der Anfangsbedingungen konzentrieren. Dies geschieht aus didaktischen Gründen und bedeutet nicht, dass Modellfehler vernachlässigt werden können.

Dieser Artikel soll eine Einführung in die Thematik der Initialisierung von Ensemblevorhersagen geben. Aufgrund der Komplexität der Thematik wird zur Illustration mit einem einfachen dynamischen System begonnen. Anhand eines Beispiels, welches auf einem System von gekoppel-

ten nichtlinearen Differentialgleichungen beruht, wird aufgezeigt, wie mithilfe der Datenassimilation quantitativ die Verteilung der Anfangsfehler geschätzt werden kann. Diese Anfangsverteilung ist abhängig von dem Zustand des Systems und den verwendeten Beobachtungsdaten. Es wird gezeigt, dass die Güte der Ensemblevorhersagen davon abhängt, wie gut diese variable Verteilung von Unsicherheiten der Anfangsbedingungen modelliert werden kann.

Die Simulation von Unsicherheiten der Anfangsbedingungen für Vorhersagen mit realistischen Wettervorhersagemodellen ist aufgrund der Anzahl der Modellvariablen und der nicht genau bekannten Charakteristika der Quellen der Unsicherheit eine große Herausforderung. Der Artikel gibt einen kurzen Überblick über gegenwärtig operatio-

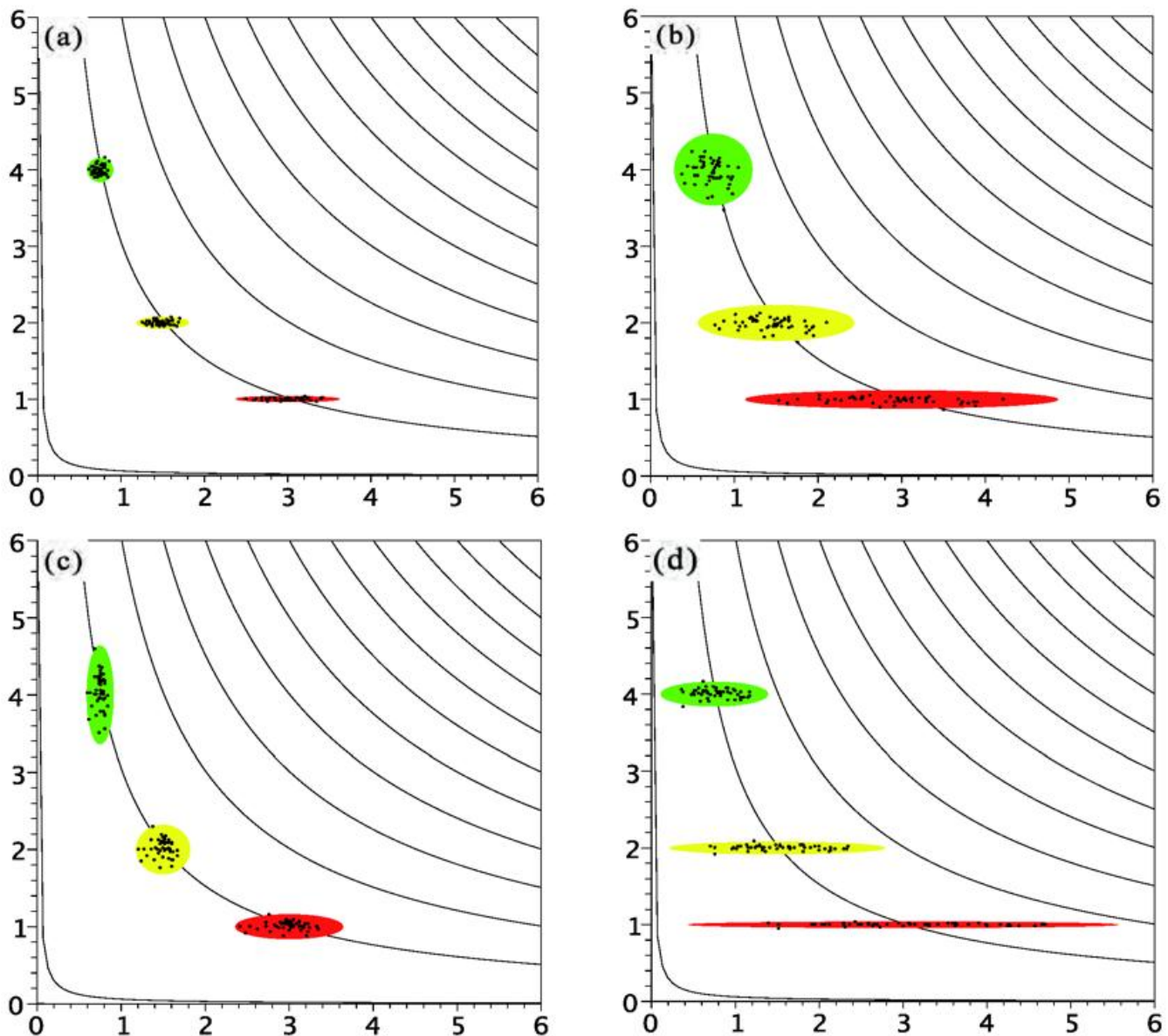


Abb. 2-1: Probabilistische Vorhersage der Position eines Fluidelementes (bzw. des Matrosen Z) in einer stationären zweidimensionalen Strömung (Stromlinien in schwarz) für verschiedene Unsicherheiten der Anfangsposition (grün). Isotrope Wahrscheinlichkeitsdichte zum Anfangszeitpunkt t_0 mit kleiner (a) und großer (b) Varianz; anisotrope Anfangsverteilungen mit dominanter Varianz parallel (c) bzw. transversal (d) zur Strömungsrichtung. Die Gesamtvarianz der Anfangsposition ist identisch in (b-d). Der Bereich von drei Standardabweichungen ist schattiert: $t = t_0$ (grün), $t = t_0 + 0.69$ (gelb) und $t = t_0 + 1.39$ (rot). Aus den Verteilungen zu $t = t_0$ wurden zufällige Stichproben von 50 Positionen gezogen und deren zeitliche Entwicklung berechnet (schwarze Punkte).

nell eingesetzte Verfahren. Zum Abschluss werden die am Europäischen Zentrum für mittelfristige Wettervorhersage verwendeten Störungen der Anfangsbedingungen beschrieben.

2 Niedrigdimensionale Beispiele

In diesem Abschnitt wird anhand von einfachen niedrigdimensionalen Beispielen der Zusammenhang zwischen der anfänglichen Wahrscheinlichkeitsverteilung der Modellvariablen und der vorhergesagten Verteilung veranschaulicht.

2.1 Lineare Dynamik

Wir beginnen mit der Aufgabe, die Position des in einer stationären Strömung treibenden Matrosen Z vorherzusagen, wenn dessen Position (x, y) zum Anfangszeitpunkt t_0 nicht genau bekannt ist. Es wird sich zeigen, dass die Unsicherheit in der vorhergesagten Position, entscheidend von der Unsicherheit in den Anfangsbedingungen abhängt. Es kommt nicht nur auf die Größe des Anfangsfehlers an sondern auch wie dieser relativ zur Strömung orientiert ist (Abb. 2-1). Dieses Beispiel kann als Analogie zur Wettervorhersage interpretiert werden. Die Position des Matrosen entspricht dem Zustand der Atmosphäre. Zuverlässige Wahrscheinlichkeitsvorhersagen der Position zu $t_1 > t_0$ erfordern eine korrekte Beschreibung der anfänglichen Wahrscheinlichkeitsverteilung der Position.

In diesem linearen zweidimensionalen Beispiel gehen wir von einer zeitlich unveränderlichen Strömung $u(x) = UL^{-1}x$, $v(y) = -UL^{-1}y$ in einer L-förmigen Bucht ($x > 0, y > 0$) aus. Hierbei sind U und L Skalen für die Geschwindigkeit der Strömung und Länge. Nach Skalierung der Zeit mit LU^1 und der Länge mit L ergibt sich das dimensionslose System

$$\frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1}$$

Um zu entscheiden, welches Gebiet nach dem Matrosen Z abgesucht werden muss, erbittet die Küstenwache Vorhersagen der Position von Z zur Zeit $t_0 + \ln 2$ und $t_0 + \ln 4$. Die Wahrscheinlichkeitsverteilung p_0 für die anfängliche Position von Z sei eine Gauß-Verteilung. Die Verteilung ist vollständig durch die Angabe des Mittelwertes und der Kovarianz der Verteilung bestimmt. Der Mittelwert von p_0 sei durch $(x, y) = (1, 4)$ gegeben. Abbildung 2-1 zeigt vier Beispiele, die sich durch die Breite der Anfangsverteilung in x - und y -Richtung unterscheiden. Eine höhere Genauigkeit in den Anfangsbedingungen resultiert in genaueren Vorhersagen (vergleiche Abb. 2-1 (a), (b)). Die Beziehung zwischen der vorhergesagten Varianz und der anfänglichen Varianz ist in diesem Beispiel linear; das gleiche gilt für die Standardabweichungen. Aufgrund von Unterschieden im Fehlerwachstum in verschiedenen Richtungen im Zustandsraum hängt die Ungenauigkeit der Vorhersage auch davon ab, wie die Gesamtvarianz auf die Richtungen ver-

teilt ist. In Abbildung 2-1 (b)-(d) werden drei verschiedene Anfangsverteilungen betrachtet, die alle eine Gesamtvarianz von 0,048 aufweisen: isotroper Fehler mit gleicher Varianz in x und y -Richtung (b), Fehler in y -Richtung dominiert (c) und Fehler in x -Richtung dominiert (d). Fehler in x -Richtung, also transversal zur Strömung, führen zu wesentlich größeren Fehlern in der vorhergesagten Position als Fehler entlang der Stromlinien (y -Richtung). In diesem sehr einfachen Beispiel sind die Gleichungen nicht gekoppelt und die Unsicherheit in x -Richtung hat keinen Einfluss auf die Unsicherheit in y -Richtung - und umgekehrt. Tabelle 2-1 enthält die anfängliche und die vorhergesagte Gesamt-Varianz der Verteilungen. Inwieweit eine zuverlässige Wahrscheinlichkeitsvorhersage der Position des Matrosen gemacht werden kann, hängt entscheidend davon ab, ob die angenommene Wahrscheinlichkeitsverteilung der Anfangsposition tatsächlich konsistent die Statistik der Anfangsfehler beschreibt.

Allgemein gilt für beliebige lineare Systeme, dass eine Gauß-Verteilung durch die Dynamik nur in eine andere Gauß-Verteilung überführt werden kann. Ferner gilt für beliebige Anfangsverteilungen mit Kovarianz-Matrix C_0 , dass die Verteilung zu einem späteren Zeitpunkt t_1 die Kovarianz

$$C_1 = MC_0M^T \tag{2}$$

besitzt, wobei die Matrix $M \equiv M(t_1, t_0)$ den Propagator des linearen Systems von t_0 nach t_1 bezeichnet. In der Analysis wird der Propagator auch als Evolutionsoperator bzw. als Hauptfundamentalmatrix des homogenen linearen Differentialgleichungssystems bezeichnet. Die n -mal- n -Matrix $M(t, t_0)$ hat die Eigenschaft, dass für beliebige Spaltenvektoren $a \in \mathbb{R}^n$ die Funktion $t \rightarrow M(t, t_0)a$ eine Lösung des linearen Systems mit Anfangsbedingung a ist.

In dem ersten Beispiel wurden auch zufällige Stichproben von 50 Anfangspositionen aus den Verteilungen der Anfangsbedingungen gezogen und deren zeitliche Entwicklung berechnet (schwarze Punkte in Abbildung 2-1). Diese Ensemblevorhersagen sind hier nur zur Illustration gezeigt, denn die zeitliche Entwicklung der Kovarianzen kann in diesem linearen Beispiel analytisch bestimmt werden.

Experiment				
Zeit	(a)	(b)	(c)	(d)
0	5	48	48	48
0,69	11	103	23	183
1,39	43	388	48	728

Tabelle 2-1: Gesamtvarianz ($\times 1000$) der Unsicherheit des Zustandes eines zweidimensionalen Systems gegeben durch Gleichung (1) für verschiedene Zeiten $t-t_0$. Experimente (a) bis (d) unterscheiden sich durch die Gauß-Verteilungen, welche die Unsicherheit des Anfangszustandes zum Zeitpunkt t_0 beschreibt (siehe Text und Abb. 2-1).

Sobald man jedoch ein nichtlineares höherdimensionales System probabilistisch vorherzusagen will, muss man diskrete Stichproben (ein Ensemble) von Anfangsbedingungen betrachten. Dies liegt daran, dass die direkte Lösung der Liouville-Gleichung, welche die zeitliche Entwicklung der Wahrscheinlichkeitsdichte beschreibt, nicht praktikabel ist. Dies beruht darauf, dass die Liouville-Gleichung für ein System mit n -dimensionalem Zustandsraum eine partielle Differentialgleichung mit n unabhängigen Variablen ist. Wird jede Variable in nur 10 Werte diskretisiert, so ist die diskretisierte Liouville-Gleichung ein 10^n dimensionales System. Deshalb ist eine direkte numerische Lösung für $n > 10$ rechentechnisch zu aufwändig. Ein analytischer Zugang ist im Allgemeinen ebenfalls aussichtslos, da analytische Lösungen von nichtlinearen Systemen nur in Ausnahmefällen bekannt sind.

2.2 Nichtlineare Dynamik

In diesem Abschnitt werden wir zunächst sehen, dass in einem nichtlinearen System das Wachstum von Störungen stark vom Anfangszustand des Systems abhängen kann. Somit kann die Streuung einer Ensemblevorhersage von Fall zu Fall sehr unterschiedlich ausfallen, selbst wenn die Unsicherheit der Anfangsbedingungen immer durch die gleiche Verteilung repräsentiert worden ist. Anschließend

wird illustriert, wie mithilfe eines Datenassimilationsverfahrens die Unsicherheit der Anfangsbedingungen explizit vorhergesagt werden kann und wie sie von Fall zu Fall stark variieren kann. Schließlich wird demonstriert, dass eine solche akkurate und fallabhängige Beschreibung der Unsicherheit in den Anfangsbedingungen zur Qualität der Ensemblevorhersage entscheidend beiträgt.

In den Beispielen dieses Abschnittes wird ein bestimmtes nichtlineares System von gekoppelten gewöhnlichen Differentialgleichungen betrachtet. Dieses eignet sich gut, um grundlegende Aspekte der Vorhersagbarkeit und Datenassimilation zu illustrieren. Das System

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \tag{3}$$

mit $i = 1, 2, \dots, N$ und zyklischen Randbedingungen wurde von LORENZ (1995) eingeführt (siehe auch LORENZ 2005). Für einen hinreichend großen Antriebsterm F treten positive Lyapunov-Exponenten auf, das heißt im Limes für große Zeiten und infinitesimale Störungsamplitude wächst der Abstand zweier Lösungen exponentiell an. Dies hat zur Folge, dass Lösungen von (3) sensibel von den Anfangsbedingungen abhängen. Für die folgenden Beispiele werden nun die allgemein üblichen Werte $N = 40$ und $F = 8$

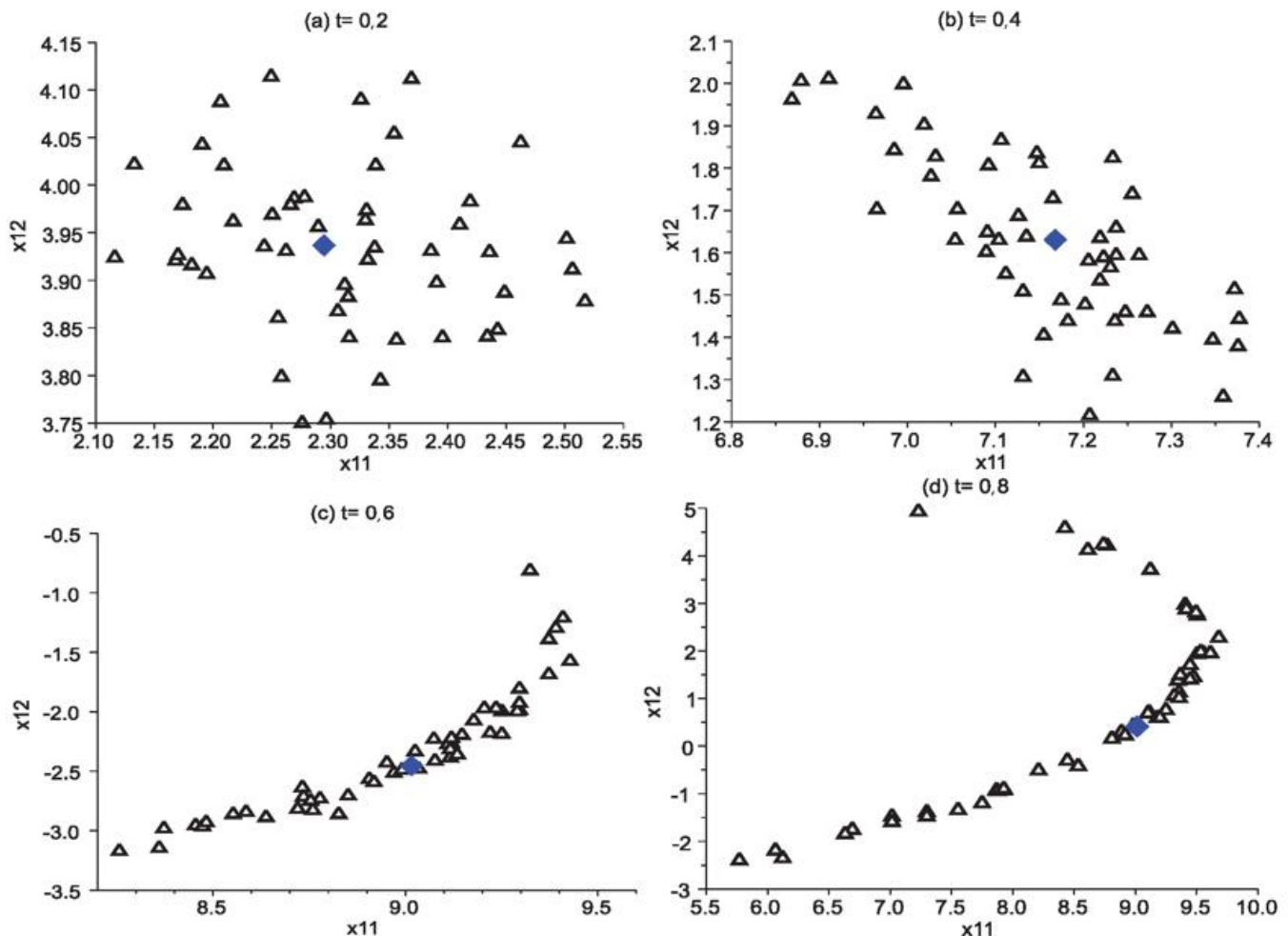


Abb. 2-2: Ensemblevorhersage der Variablen x_{11} und x_{12} für Gauß-verteilte Anfangsstörungen mit Standardabweichung $\sigma = 0,1$ für das Lorenz-95-System. Das blaue Rautensymbol zeigt die Referenzlösung.

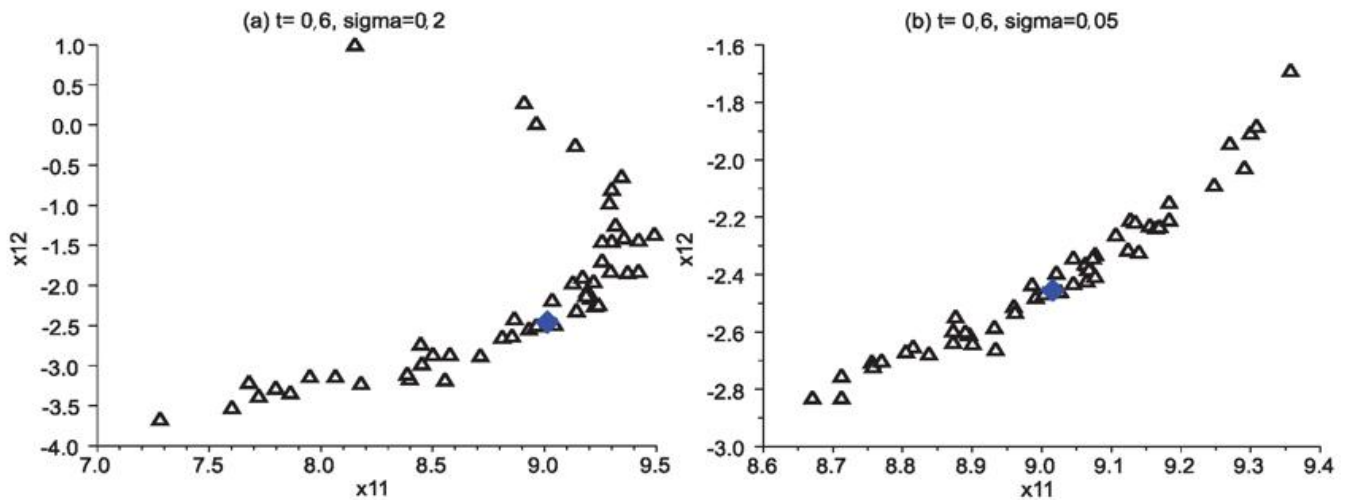


Abbildung 2-3: Wie Abb. 2-2c aber für andere Standardabweichungen σ für die Verteilung der Anfangsstörungen: (a) $\sigma = 0.2$. (b) $\sigma = 0.05$.

verwendet. Lorenz hat vorgeschlagen eine dimensionslose Zeiteinheit des Systems (3) mit 5 Tagen gleichzusetzen, da dann das mittlere Wachstum kleiner Störungen in diesem System dem von synoptischskaligen Störungen in der Atmosphäre in den mittleren Breiten entspricht. Störungen verdoppeln sich im Laufe von etwa 2 Tagen. Das System (3) wird hier mit einem Runge-Kutta-Verfahren vierter Ordnung und einem Zeitschritt von 0,025 gelöst.

2.2.1 Klimatologische Wahrscheinlichkeitsverteilung der Anfangsfehler

Zunächst nehmen wir an, dass die Unsicherheit der Anfangsbedingungen durch eine feste isotrope Gauß-Verteilung gegeben ist

$$p(\mathbf{x}') = [\sigma^2 2\pi]^{-N/2} \exp\left(-\frac{1}{2} \sum_{j=1}^N x_j'^2 \tilde{S}^{-2}\right). \quad (4)$$

Hier bezeichnet $\mathbf{x}' = (x_1', \dots, x_N')$ eine Störung des Zustandes des Systems. Nach einer langen (> 365 d) Integration des Systems mit willkürlich ausgewählten Anfangsbedingungen wird ein Zustand gewonnen, der mit der Dynamik des Systems konsistent ist. Dieser Zustand \mathbf{x}_c wird als ein ungestörter Referenzzustand angesehen. Dann betrach-

ten wir verschiedene Ensembles von Lösungen von (3) mit Anfangsbedingungen $\mathbf{x}_c + \mathbf{x}'_k$, wobei die Störungen der Anfangsbedingungen \mathbf{x}'_k , $k = 1, \dots, L$ eine zufällige Stichprobe der durch (4) gegebenen Verteilung sind. Zur graphischen Darstellung werden nun zwei benachbarte Variablen aus den 40 Variablen ausgewählt. Abbildung 2-2 zeigt ein Ensemble mit $L = 50$ Mitgliedern in der (x_{11}, x_{12}) -Ebene zu vier Vorhersagezeiten. An Tag 1 ist die Verteilung noch isotrop. An Tag 2 hat sich jedoch eine deutliche negative Korrelation zwischen x_{11} und x_{12} entwickelt. An Tag 3 und 4 zeigen sich deutliche Abweichungen von einer Gauß-Verteilung, die auf nichtlineare Effekte rückschließen lassen. Änderungen der Breite der Anfangsverteilung spiegeln sich in der vorhergesagten Unsicherheit deutlich wider, auch wenn die nichtlineare Dynamik bereits merkliche Abweichungen von der Gauß-Verteilung verursacht hat. Abbildung 2-3 (a), (b) und 2-2 (c) zeigen die Verteilung der Variablen x_{11}, x_{12} zur Zeit $t = 0,6$ für drei verschiedene Werte von σ . Der gezeigte Ausschnitt der (x_{11}, x_{12}) -Ebene wächst mit zunehmender Standardabweichung σ der Anfangsverteilung.

Das Störungswachstum ist außerdem erheblich vom ungestörten Referenzzustand \mathbf{x}_c der Anfangsbedingungen abhängig. Abbildung 2-4 (a), (b) und Abbildung 2-2 (c)

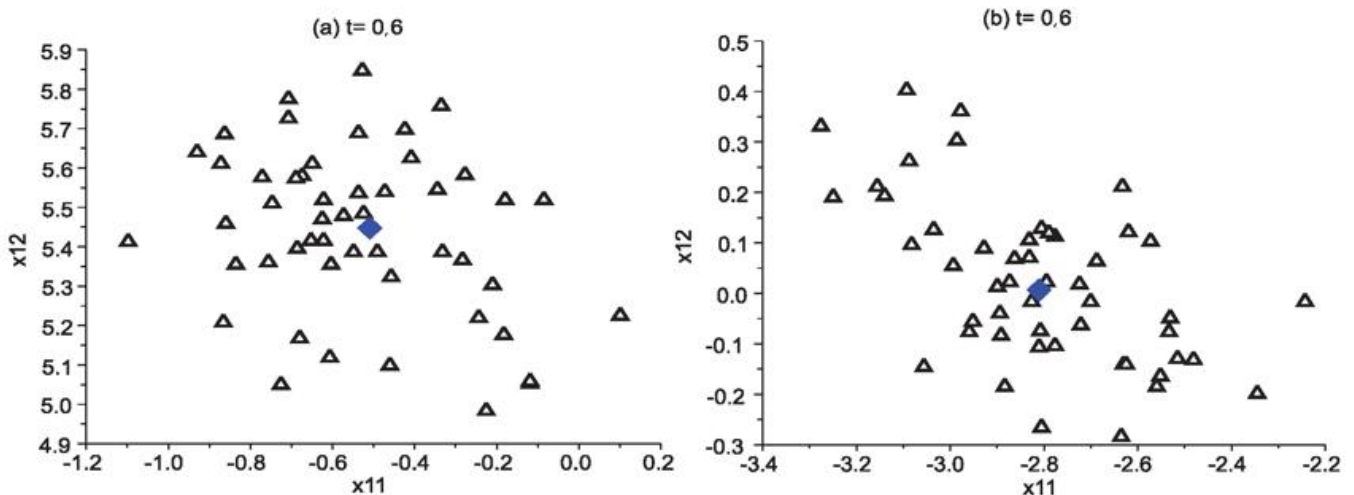


Abbildung 2-4: Wie Abb. 2-2 (c) aber für andere Referenzzustände: (a) 28 Tage, (b) 56 Tage später als Startbedingung in Abb. 2-2.

zeigen Ensemblevorhersagen für drei verschiedene Referenzzustände \mathbf{x}_c . Diese wurden derselben Integration im Abstand von $\Delta t = 5,6$ (28 Tage) entnommen. Die Störungen der Anfangsbedingung sind identisch und entsprechen wiederum einer zufälligen Stichprobe aus der isotropen Gauß-Verteilung. Die resultierenden Ensembleverteilungen an Tag 3 sind sehr verschieden, da das Störungswachstum vom aktuellen Zustand des Systems moduliert wird. Aufgrund dessen ist zu erwarten, dass ein Ensemble, auch wenn es nur eine klimatologische und damit zustandsunabhängige Verteilung für die Generierung der Anfangsstörungen verwendet, trotzdem nützliche strömungsabhängige Unsicherheitsvorhersagen liefern kann.

Ziel von Ensemblevorhersagen ist es, statistisch konsistent zu sein. Das heißt, dass die Verteilung der Anfangsbedingungen so gut wie möglich deren Fehlerstatistik beschreiben sollte. Ist die tatsächliche Varianz der Fehler größer (kleiner) als die Varianz der verwendeten Störungen der Anfangsbedingungen, so unterschätzt (überschätzt) das Ensemble systematisch den Vorhersagefehler.

Die Verteilung der Fehler der Anfangsbedingungen hängt vom Assimilationsverfahren, von den verwendeten Beobachtungen und vom Modellfehler ab. Abhängig vom Zustand des Systems variiert das Fehlerwachstum. Infolgedessen schwankt auch die Genauigkeit der Anfangsbedingungen von Tag zu Tag. Verfahren der Datenassimilation können verwendet werden, um die Statistik der Fehler der Anfangsbedingungen abhängig von den bereits assimilierten Beobachtungen und dem der Analyse vorausgegangenem Fehlerwachstum abzuschätzen. Dies wird im folgenden Abschnitt genauer erläutert.

2.2.2 Variabilität der Unsicherheit der Anfangsbedingungen

Nun betrachteten wir Vorhersageexperimente mit dem Lorenz-95-System, in denen die variable Unsicherheit der Anfangsbedingungen explizit mit einem Datenassimilationsverfahren vorhergesagt wird. Anhand dieses Beispiels wird untersucht, inwiefern die Ensemblevorhersagen davon abhängen, ob die Variabilität von Unsicherheiten der Anfangsbedingungen korrekt repräsentiert wird. Als Assimilationsverfahren wird ein Erweiterter Kalman-Filter verwendet. Für diesen Artikel genügt es zu wissen, dass der Kalman-Filter nicht nur den Anfangszustand aus den Beobachtungen und einer Kurzfristvorhersage - dem Hintergrund - schätzt, sondern auch die Kovarianzmatrix des Fehlers des Anfangszustandes. Wir werden den Anfangszustand, die Analyse, mit \mathbf{x}_a und die zugehörige Fehler-Kovarianzmatrix mit \mathbf{P}_a bezeichnen. Viele momentan operationell verwendete Assimilationsverfahren können als Approximationen des Kalman-Filters angesehen werden. Die Gleichungen des Kalman-Filters sowie der Zusammenhang zwischen Kalman-Filter und der zeitabhängigen Variationsmethode (4D-Var) werden von FISHER et al. (2005) erklärt.

In den folgenden Experimenten werden 8 der 40 Variablen beobachtet: $x_1, x_6, x_{11}, \dots, x_{36}$. Die Standardabweichung σ_0 des Beobachtungsfehlers beträgt 5% der klimatologischen Standardabweichung. Beobachtungen liegen alle 6 Stunden vor. Die synthetischen Beobachtungen stammen aus einer langen Integration. Beobachtungsfehler werden durch Zufallszahlen simuliert, die aus einer Gauß-Verteilung mit der Standardabweichung σ_0 gezogen werden. Die Fehlerkovarianz des Hintergrundes wird im Kalman-Filter durch $\mathbf{P}_f - \mathbf{M}\mathbf{P}_a\mathbf{M}^T$ vorhergesagt. Falls diese vorhergesagte Kovarianz die tatsächlichen Fehlervarianzen signifikant

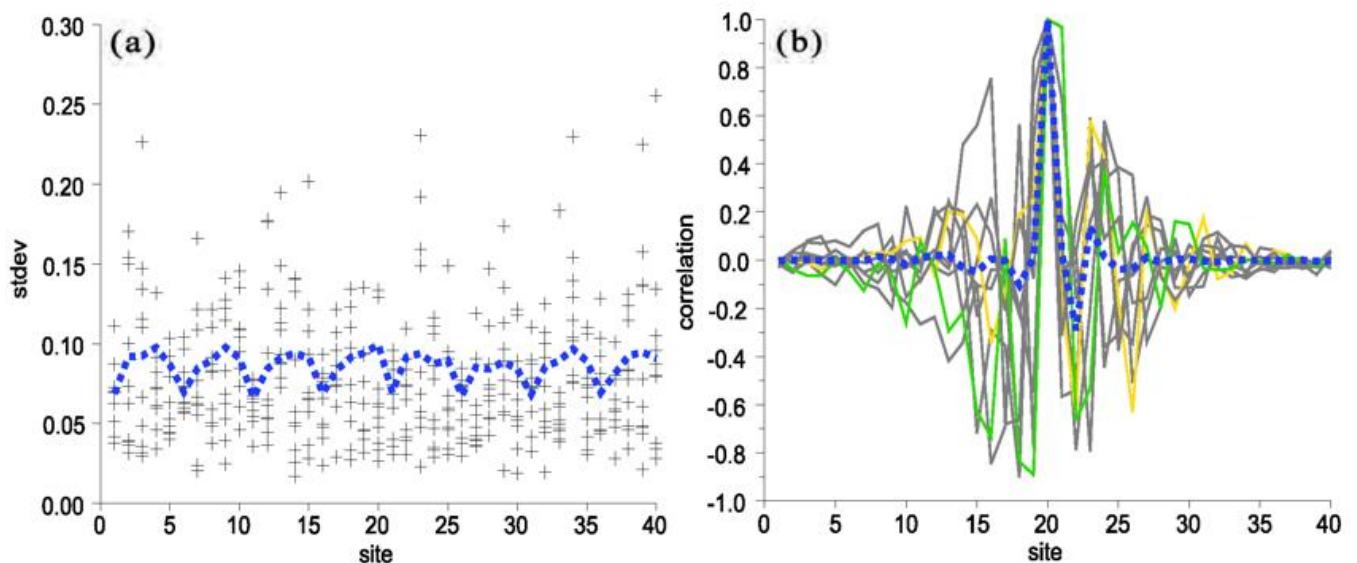


Abbildung 2-5: Zeitliche Variabilität der vom Kalman-Filter für das Lorenz-95-System vorhergesagten Kovarianzmatrix der Analysefehler. (a) Standardabweichung, (b) Korrelation zwischen x_{20} und x_1 . Die x-Koordinate (site) bezieht sich auf den Ort im Lorenz-95-System, das heißt auf den Index i in Gleichung (3). Zehn verschiedene Kovarianzen im Abstand von 5 Tagen sind als Plussymbole (a) beziehungsweise durchgezogene Kurven (b) dargestellt. Die fettgedruckte gepunktete Linie zeigt die über einen Zeitraum von 50 Tagen gemittelte Kovarianzmatrix.

unterschätzt, ignoriert der Kalman-Filter die Information in den Beobachtungen und divergiert von der tatsächlichen Entwicklung des Systems. Um eine solche Divergenz des Kalman-Filters zu vermeiden, wird zu der linearen Vorhersage der Kovarianz $\mathbf{M}\mathbf{P}_a\mathbf{M}^T$ des Hintergrund-Fehlers eine positiv definite symmetrische Matrix \mathbf{Q} addiert. Dies ist notwendig, da die lineare Dynamik die Entwicklung der Unsicherheit im nichtlinearen System nur näherungsweise beschreiben kann. Die Matrix \mathbf{Q} wird üblicherweise als die Kovarianzmatrix des Modellfehlers angesehen. Das nichtlineare Modell ist hier jedoch ein perfektes Modell - es stimmt mit dem System überein. Deshalb hat \mathbf{Q} die Funktion einer Regularisierung von \mathbf{P}_r . Hier wird ein Vielfaches der Einheitsmatrix zur Stabilisierung verwendet: $\mathbf{Q} = \sigma_q^2 \mathbf{I}$. Empirisch wurde ermittelt, dass eine Standardabweichung von $\sigma_q = 0,005$ ausreicht, um für das betrachtete Beobachtungsnetz die Divergenz des Kalman-Filters zu vermeiden.

Die vom Kalman-Filter vorhergesagte Kovarianz-Matrix \mathbf{P}_a der Analysefehler ist zeitlich variabel. Sowohl die Standardabweichung als auch die Korrelationen variieren mit dem Zustand des Systems (Abbildung 2-5). Die Standardabweichungen schwanken in einem Bereich von einem Drittel bis zum zweieinhalbfachen des zeitlichen Mittels. Instantane Korrelationen weichen typischerweise ebenfalls deutlich von den Korrelationen ab, die sich aus der zeitlich gemittelten Kovarianzmatrix ergeben.

In den folgenden Ensembleexperimenten wird die Unsicherheit der Anfangsbedingungen durch eine Gauß-Verteilung repräsentiert. Die Experimente unterscheiden sich ausschließlich durch die Kovarianz dieser Verteilung. Experiment A verwendet die vom Kalman-Filter vorhergesagte zeitlich variable Analysefehler-Kovarianzmatrix. Experiment M verwendet das zeitliche Mittel der vom Kalman-Filter vorhergesagten Analysefehler-Kovarianz. Experiment Z verwendet eine zufällige Permutation der Zeitreihe der vom Kalman-Filter vorhergesagten Kovarianzen. Auf diese Weise hat Experiment Z die gleiche mittlere Kovarianz wie die beiden anderen Experimente, aber die Abweichungen der aktuellen Kovarianz vom zeitlichen Mittel sind rein zufällig und haben nichts mehr mit dem Zustand des Systems zu tun.

Alle Experimente verwenden die gleiche vom Kalman-Filter bestimmte Anfangsbedingung x_a . Jedes Ensemble besteht aus 50 Vorhersagen. Die Anfangsbedingungen dieser Vorhersagen entsprechen einer zufälligen Stichprobe aus der jeweiligen Gaußverteilung. Um einen aussagekräftigen Vergleich machen zu können, werden 200 Starttermine betrachtet.

In einem statistisch konsistenten Ensemble soll die Streuung der Mitglieder ein Maß für den Fehler des Ensemble-Mittelwertes sein. Man kann zeigen, dass die Standardabweichung des Ensembles für hinreichend große Stichproben mit dem quadratischen Mittel des Fehlers des Ensemble-Mittelwertes übereinstimmen muss. Abbildung 6a zeigt, dass für die drei Experimente sowohl die Ensemble-Stan-

dardabweichungen als auch die quadratischen Mittel vom Fehler des Ensemblemittels recht ähnlich sind. Außerdem stimmt die Streuung der Ensemblemitglieder gut mit der Größe des Fehlers des Ensemblemittels überein.

Um die Güte der Ensemblevorhersagen zu vergleichen, werden zwei Bewertungsmaße der Verteilung verwendet: Der Continuous Ranked Probability Score (CRPS) ist der mittlere quadratische Fehler der kumulativen Verteilung. Das hier als Continuous Ignorance Score (CIgnS) bezeichnete Gütemaß ist definiert als

$$\text{CIgnS} = \frac{(\mu_e - y)^2}{2\sigma_e^2} + \log(\sigma_e \sqrt{2\pi}), \quad (5)$$

hierbei stehen μ_e , σ_e und y für das Ensemblemittel, die Ensemble-Standardabweichung und den vom System tatsächlich realisierten Wert. Der CIgnS entspricht dem negativen Logarithmus der vom Ensemble vorhergesagten Gaußschen Wahrscheinlichkeitsdichte ausgewertet am tatsächlichen Wert, den das System annimmt. Beide Gütemaße sind desto kleiner je besser die Ensemble-Vorhersage ist. Die besten Ensemblevorhersagen liefert Experiment A, welches die Variabilität der Kovarianz des Anfangsfehlers korrekt modelliert (Abbildung 2-6). Die Ensemblevorhersagen von Experiment Z sind am schlechtesten aufgrund der falschen zeitlichen Variabilität der Kovarianz. Die zeitlich konstante Kovarianz (Experiment M) fällt zwischen die beiden Experimente A und Z. Die Unterschiede im CRPS sind wesentlich kleiner als im CIgnS. Letzterer reagiert wesentlich sensibler auf die Varianz des Ensembles als Ersterer. Dies ist konsistent mit einem idealisierten Vergleich der beiden Gütemaße (LEUTBECHER 2009).

3 Verfahren für numerische Wettervorhersagen

3.1 Überblick

Eine möglichst genaue Beschreibung von Unsicherheiten der Anfangsbedingungen ist auch für die Ensemblevorhersage mit realistischen numerischen Wettervorhersagemodellen erstrebenswert. Die Güte dieser Vorhersagen ist ebenso wie im Beispiel mit dem Lorenz-95-System abhängig von den Störungen der Anfangsbedingungen. Die Störungen beschreiben die zufälligen Fehler der Anfangsbedingungen. Die Statistik der Störungen soll konsistent mit dem Fehler der Analysen sein. Soweit dies möglich ist, sollte auch die zustandsabhängige Variabilität der Unsicherheit der Anfangsbedingungen repräsentiert werden. Da das Wachstum der Störungen von ihrer Struktur beeinflusst wird, ist es auch wichtig, die räumlichen und multivariaten Korrelationen der Analysefehler zu charakterisieren.

Aufgrund der Komplexität numerischer Wettervorhersagemodelle ist es jedoch wesentlich schwieriger, die Kovarianz der Fehler der Anfangsbedingungen zu bestimmen. Die Anzahl der Variablen in gegenwärtig verwendeten numerischen Modellen ist gewöhnlich größer als 10^7 . Deshalb

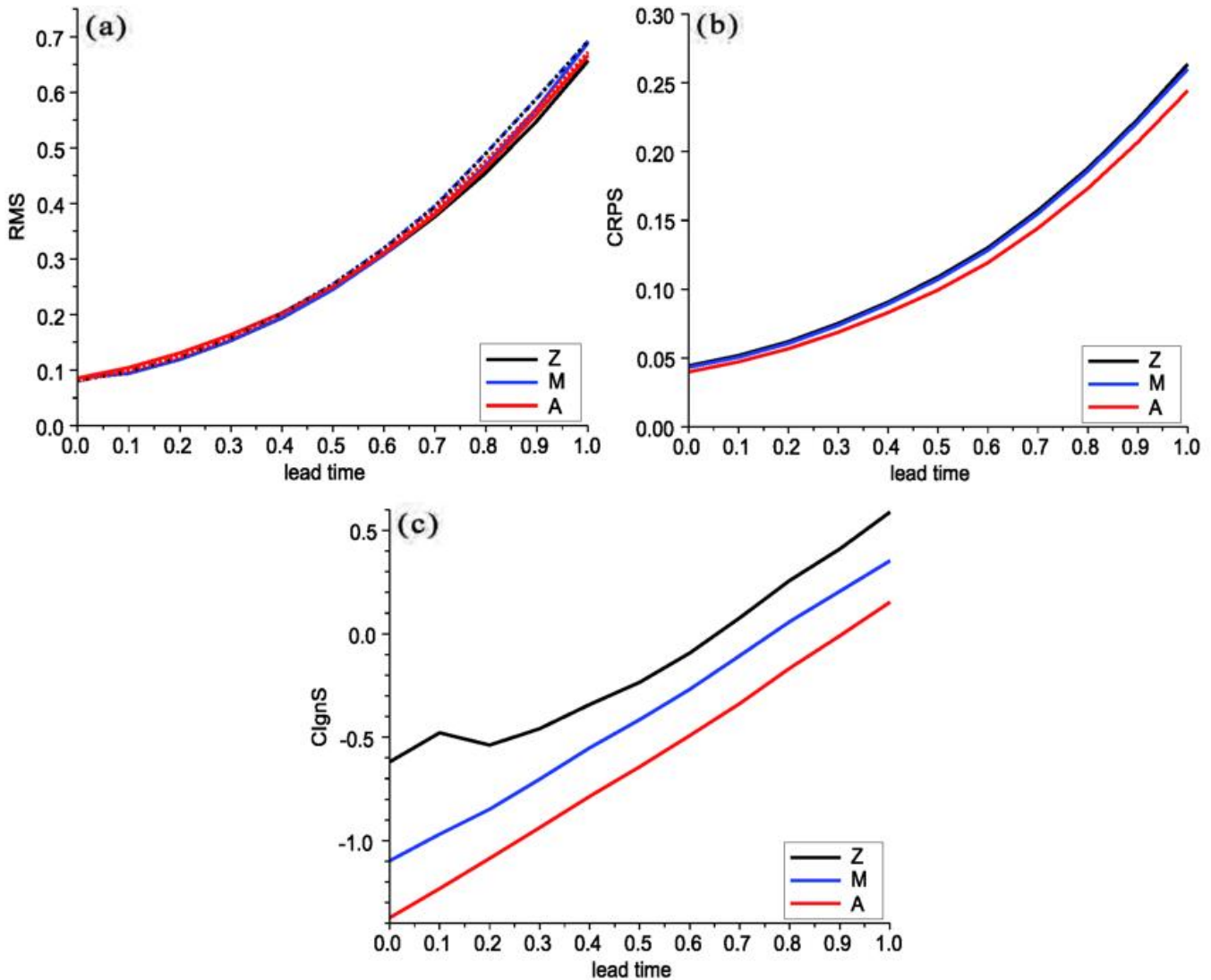


Abbildung 2-6: Verifikation der Ensemblevorhersagen für das Lorenz-95-System mit verschiedenen Kovarianzen für die Verteilung der Anfangsstörungen: Experiment A verwendet die aktuelle vom Kalman-Filter vorhergesagte Kovarianz; Experiment M verwendet das zeitliche Mittel der Kovarianz von A und Experiment Z verwendet eine zufällige Permutation der Zeitreihe der Kovarianzen von A. (a) RMS-Fehler des Ensemblemittelwertes (gepunktet) und Ensemblestandardabweichung (durchgezogen), (b) „Continuous Ranked Probability Score“ und (c) „Continuous Ignorance Score“ (siehe Text für Details).

sind direkte Vorhersagen der Fehlerkovarianzen mit einem erweiterten Kalman-Filter rechentechnisch viel zu aufwendig. Mithilfe von Ensemblemethoden ist es jedoch möglich, zufällige Stichproben aus der Wahrscheinlichkeitsverteilung der Fehler zu erhalten. Viele dieser Methoden sind in gewisser Weise Approximationen der Kovarianzprognose des Erweiterten Kalman-Filters. Der kanadische Wetterdienst verwendet zur Bestimmung der Anfangsbedingungen des globalen Ensembles einen Ensemble-Kalman-Filter (HOUTEKAMER et al. 2005). Der französische Wetterdienst und das Europäische Zentrum für mittelfristige Wettervorhersage verwenden ein Ensemble von 4D-Vars, um Störungen der Anfangsbedingungen zu gewinnen (PEREIRA und BERRE 2006; ISAKSEN et al. 2010). Die einzelnen Mitglieder des Ensembles von Datenassimilationen unterscheiden sich durch unterschiedliche Störungen der Beobachtungswerte. Die Varianzen der aufgebrachten Störungen stimmen mit den geschätzten Varianzen der Beobachtungsfehler überein. Auch bei dem in Kanada verwendeten Ensemble-Kalman-Filter werden die Beobach-

tungen gestört. Es gibt jedoch noch andere Versionen des Ensemble-Kalman-Filters, die keine Störungen der Beobachtungsdaten benötigen.

Um die Kovarianz der Anfangsfehler quantitativ berechnen zu können, ist es notwendig, die Statistik der relevanten Fehlerquellen akkurat beschreiben zu können. Die wesentlichen Elemente sind die Fehlercharakteristika der Beobachtungsdaten sowie die des verwendeten Vorhersagemodells. **Modellfehler haben einen Einfluss auf die Fehler der Anfangsbedingungen**, da Kurzfristvorhersagen zu ihrer Schätzung verwendet werden. ISAKSEN et al. (2010) etwa zeigen, dass die Berücksichtigung der Modellfehler in einem Ensemble von 4D-Vars die vorhergesagte Standardabweichung des 4D-Var-Ensembles um etwa 30% erhöht.

Es ist nicht trivial, die Fehlerstatistik der Modellfehler sowie der Fehler der Beobachtungsdaten genau zu schätzen. Dies kann den anfänglichen Erfolg alternativer Methoden zur Gewinnung der Störungen der Anfangs-

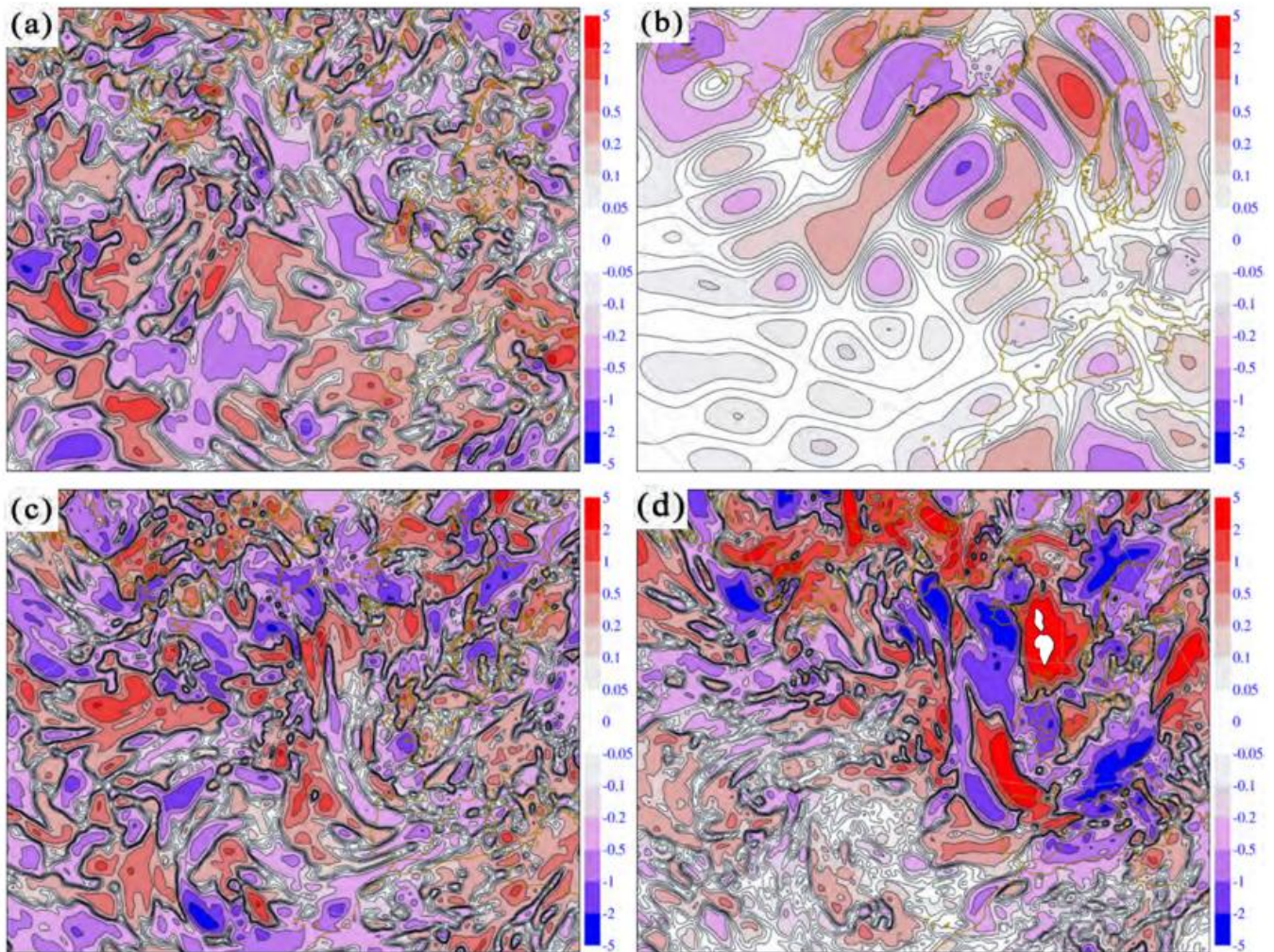


Abbildung 2-7: Störungen der Anfangsbedingungen (a, b) und einer 48-Stunden Vorhersage (c, d) für zwei numerische Experimente mit ECMWF-Ensemble zum Starttermin 1. Oktober 2008, 0 UTC. Gezeigt ist die Differenz zwischen Mitglied 1 und der ungestörten Kontrollvorhersage für Temperatur in Kelvin im 700 hPa-Niveau über dem Nordatlantik und Europa. In (a, c) sind Anfangsbedingungen mit dem Ensemble von Datenassimilationen (EDA) generiert und in (b, d) mit singulären Vektoren.

bedingungen erklären. In der Anfangsphase globaler Ensemblevorhersagen sind insbesondere singuläre Vektoren und „bred vectors“ verwendet worden, um dynamisch relevante Störungen der Anfangsbedingungen zu erzeugen. Letztere werden beim Wetterdienst der USA (NMC, nun NCEP) und Erstere am Europäischen Zentrum für mittelfristige Wettervorhersage (EZMW) verwendet (weiterführende Literaturhinweise finden sich in LEUTBECHER und PALMER 2008).

„Bred vectors“ sind Störungen, die aus finiten Differenzen zweier kurzer nichtlinearer Integrationen gewonnen werden. Eine anfänglich zufällige Störung wird über einen Zeitraum von etwa 6 bis 24 Stunden nichtlinear integriert und dann so skaliert, dass die Amplitude mit der Größe des Analysefehlers übereinstimmt. Nach einer Reihe von Wachstumszyklen entstehen Strukturen, deren Statistik nicht mehr von der ursprünglichen Störung der Anfangsbedingungen abhängt. „Bred vectors“ können auf gewisse Weise als eine Verallgemeinerung von Lyapunov-Vektoren angesehen werden. Sie unterscheiden sich von diesen durch ihre endliche Amplitude und das endliche Zeitfenster der Optimierung.

Singuläre Vektoren sind die führenden Vektoren einer Singulärwertzerlegung des Modellpropagators über einen Optimierungszeitraum. Zur Berechnung der singulären Vektoren ist es notwendig, ein linearisiertes Modell und seine adjungierte Version zur Verfügung zu haben. Dies ist normalerweise der Fall, wenn ein vierdimensionales Variationsverfahren in der Datenassimilation eingesetzt wird. Eine geeignet definierte Singulärwertzerlegung des Propagators liefert Störungen, die sich zu den führenden Eigenvektoren der Kovarianzmatrix des linear geschätzten Vorhersagefehlers entwickeln (siehe LEUTBECHER und PALMER 2008, und darin zitierte Arbeiten).

Der Vorteil von Methoden, die direkt dynamisch relevante Anfangsstörungen erzeugen, ist, dass eine akzeptable mittlere Streuung eines Ensembles erreicht werden kann, ohne detailliert die relevanten Fehlerquellen im verwendeten Assimilationsverfahren modellieren zu müssen. In Ensembles, die singuläre Vektoren oder „bred vectors“ als Anfangsstörungen verwenden, wird die Amplitude dieser Störungen typischerweise empirisch ermittelt. Ziel ist es, für alle Vorhersagezeiten eine Streuung der Ensemblemit-

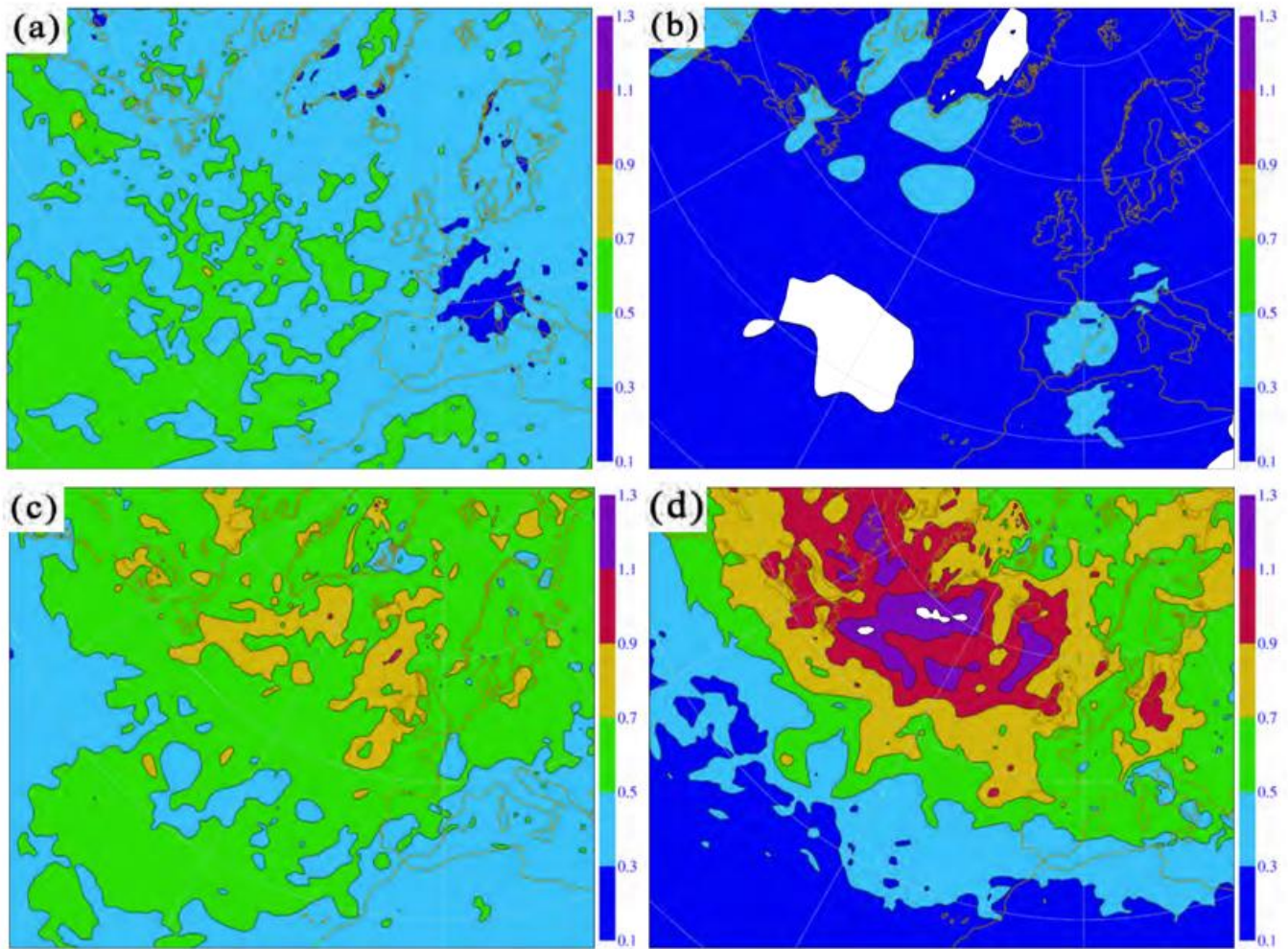


Abbildung 2-8: Standardabweichung der Temperatur in Kelvin im 700 hPa-Niveau zum Anfangszeitpunkt $t = 0$ h (a, b) und in der 48-Stunden-Vorhersage (c, d). Die Anfangsbedingungen des in (a, c) gezeigten Ensembles sind mit Störungen aus einem Ensemble von 4D-Var-Assimilationen erzeugt. Das in (b, d) gezeigte Ensemble verwendet singuläre Vektoren zur Störung der Anfangsbedingungen. Mittel über 27 Termine im Zeitraum August bis Oktober 2008.

glieder zu erreichen, die dem Fehler des Ensembledittelwertes im Mittel über viele Starttermine entspricht.

Die dynamisch konditionierten Störungen können jedoch nicht gut die von den verwendeten Beobachtungsdaten und von der Strömung abhängigen Änderungen der Unsicherheit der Anfangsbedingungen korrekt modellieren, da sie Aspekte der Datenassimilation entweder gar nicht oder nur sehr vereinfacht in der Methodik repräsentieren.

3.2 Anfangsbedingungen für das EZMW-Ensemble

Nun folgt ein kurzer Überblick über die Methodik der Anfangsstörungen, welche im Ensemble des EZMW eingesetzt werden. Zunächst wurden im EZMW-Ensemble die Störungen nur mittels singulärer Vektoren erzeugt. Die Unsicherheit der Anfangsbedingungen wurde durch eine Stichprobe aus einer isotropen Gauß-Verteilung im Unterraum der führenden singulären Vektoren dargestellt. Dabei wurden sowohl anfängliche singuläre Vektoren verwendet, die in den ersten zwei Tagen der Vorhersage das Störungswachstum maximieren, als auch entwickelte singuläre Vektoren. Letztere maximieren das Störungswachstum in den zwei Tagen bis zum Startzeitpunkt der Vorhersage. Das

Wachstum der singulären Vektoren wird mit einer Norm gemessen, die durch die Wurzel der sogenannten Gesamt-Energie definiert ist. Weitere Details werden von LEUTBECHER und PALMER (2008) beschrieben; dort finden sich auch weitergehende Literaturhinweise.

Seit Juni 2010 werden die Anfangsstörungen aus anfänglichen singulären Vektoren und Störungen, die von einem Ensemble von 4D-Vars stammen, zusammengesetzt (BUIZZA et al. 2008). Die entwickelten singulären Vektoren, die das Fehlerwachstum in der Vergangenheit repräsentieren, wurden von den Störungen des Ensembles von Datenassimilationen abgelöst. Theoretisch würde man erwarten, dass die Störungen eines korrekt konstruierten Ensembles von Datenassimilationen ausreichen sollten, um die Unsicherheit der Anfangsbedingungen statistisch konsistent zu modellieren. Die Praxis zeigt jedoch, dass dies mit dem gegenwärtigen Vorhersagesystem noch nicht möglich ist. Deshalb ist es weiterhin erforderlich, die Störungen der anfänglichen singulären Vektoren mit zu verwenden, um statistisch konsistente Vorhersagen zu erzeugen. Außerdem besteht wegen des großen Rechenaufwandes das 4D-Var-Ensemble nur aus 10 Mitgliedern; es werden jedoch 50 verschiedene Anfangsstörungen für die Vorhersagen benötigt.

Abbildung 2-7 zeigt für einen Fall die Temperaturstörungen der Anfangsbedingungen in 700 hPa, welche mit den beiden Methoden gewonnen werden. Es wurde jeweils das erste Mitglied des Ensembles ausgewählt. Statistisch unterscheidet es sich nicht von den anderen Mitgliedern. Die singulären Vektoren sind mit einer Auflösung von T42 berechnet und sind für 48 Stunden optimiert. Das Ensemble von Datenassimilationen verwendet eine Auflösung von T399. Die Störungen des 4D-Var-Ensembles sind überwiegend kleinskaliger als die der singulären Vektoren. Die maximalen Amplituden der beiden Störungstypen haben die Größenordnung von 1 K. Es ist offensichtlich, dass die singulären Vektoren auf bestimmte Gebiete lokalisiert sind, während die Störungen des Ensembles von 4D-Var das gesamte Gebiet gleichmäßiger abdecken. Erwartungsgemäß wachsen die Anfangsstörungen, die auf singulären Vektoren beruhen, wesentlich stärker an als die des Ensembles von Datenassimilationen. In diesem Beispiel, wurden lediglich die Anfangsbedingungen gestört und nicht die Tendenzen des Vorhersagemodelles. Das heißt Modellfehler wurden nicht berücksichtigt (siehe auch BERNER 2012).

Eine quantitative Diagnostik von Ensemblevorhersagen erfordert, die Statistik über hinreichend viele Starttermine zu akkumulieren. Abbildung 2-8 zeigt die Standardabweichung der beiden Ensembles mit den unterschiedlichen Anfangsstörungen für die Periode vom 14. August bis zum 5. Oktober 2008. Hierbei wurde alle 48 Stunden eine Vorhersage gestartet. Die Varianz wurde über die 27 Termine gemittelt und die Standardabweichung aus der gemittelten Varianz bestimmt. Im Allgemeinen ist die Varianz der Anfangsstörungen des Ensembles von Datenassimilationen größer als die Varianz der Anfangsstörungen, die auf singulären Vektoren beruhen. Nach 48 Stunden dominiert jedoch die Varianz in mittleren Breiten in dem Ensemble, das durch singuläre Vektoren gestört wurde. In niedrigeren Breiten dominieren die Störungen des Ensembles von Datenassimilationen.

4 Fazit

Die Darstellung der Anfangsbedingungen ist für Ensemblevorhersagen sehr wichtig. Einerseits ist die Unsicherheit der Vorhersagen vom Zustand der Atmosphäre per se abhängig, andererseits hängt die Vorhersagequalität eines Ensembles entscheidend davon ab, wie gut die - zeitlich variable - Unsicherheit der Anfangsbedingungen berücksichtigt worden ist. In der Praxis der operationellen numerischen Wettervorhersage gibt es zwei verschiedene Paradigmen: Techniken der Datenassimilation, die explizit eine Stichprobe aus der Wahrscheinlichkeitsverteilung der Anfangsbedingungen berechnen sowie empirisch angepasste Verteilungen der Anfangsstörungen. Empirische Verfahren sind nicht oder nur teilweise fähig, die Schwankungen in der Verteilung der Unsicherheit der Anfangsbedingungen vorherzusagen. Deshalb ist zu erwarten, dass

in Zukunft vermehrt Methoden der Datenassimilation eingesetzt werden, um die Anfangsbedingungen von Ensemblevorhersagen mit numerischen Wettervorhersagemodellen zu erzeugen.

Danksagung

Der Autor möchte sich an dieser Stelle bei Simon Lang, Renate Hagedorn und Sarah Jones für nützliche Kommentare bedanken.

Literatur

- BERNER, J., 2012: Verfahren zur Berücksichtigung von Modellfehlern in der numerischen Vorhersage. *Promet*, diese Ausgabe.
- BUIZZA, R., LEUTBECHER, M., ISAKSEN, L., 2008: Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**, 2051-2066.
- DALEY, R., 1991: Atmospheric Data Analysis. *Cambridge Univ. Press*, 457 pp.
- FISHER, M., LEUTBECHER, M., KELLY, G. A., 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **131**, 3235-3246.
- HOUTEKAMER, P. L., MITCHELL, H. L., PELLERIN, G., BUEHNER, M., CHARRON, M., SPACEK, L., HANSEN, B., 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.* **133**, 604-620.
- ISAKSEN, L., BONAVITA, M., BUIZZA, R., FISHER, M., HASELER, J., LEUTBECHER, M., RAYNAUD, L., 2010: Ensemble of data assimilations at ECMWF. Tech. Memo **636**, ECMWF.
- LEUTBECHER, M., PALMER, T. N., 2008: Ensemble forecasting. *J. Comp. Phys.* **227**, 3515-3539.
- LEUTBECHER, M., 2009: Diagnosis of ensemble forecasting systems. In *Seminar on Diagnosis of Forecasting and Data Assimilation Systems*, pages 235-266, Reading, UK. ECMWF.
- LEWIS, J. M., 2005: Roots of ensemble forecasting. *Mon. Wea. Rev.* **133**, 1865-1885.
- LORENZ, E. N., 1995: Predictability: A problem partly solved. In *Seminar on Predictability*, volume Vol. I, pages 1-18, Reading, UK. ECMWF.
- LORENZ, E. N., 2005: Designing chaotic models. *J. Atmos. Sci.* **62**, 1574-1587.
- PEREIRA, M. B., BERRE, L., 2006: The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.* **134**, 2466-2489.
- WEIGEL, A., 2012: Verifikation von Ensemblevorhersagen. *Promet*, diese Ausgabe.
- WERGEN, W., 2002: Datenassimilation - ein Überblick. *Promet* **27**, 141-149.
- WERNLI, H., 2012: Wetter, Chaos und probabilistische Wettervorhersagen. *Promet*, diese Ausgabe.

J. BERNER

3 Verfahren zur Berücksichtigung von Modellfehlern in der numerischen Wettervorhersage

Representing model-error in numerical weather

Zusammenfassung

Dieser Artikel diskutiert die Grundlagen, warum Methoden zur Berücksichtigung von Modellfehlern in der kurzfristigen Wettervorhersage notwendig sind. Wir stellen verschiedene Methoden vor und diskutieren Vor- und Nachteile. Für ein ausgewähltes mesoskaliges Ensemblesystem werden Ergebnisse für die verschiedenen Methoden gezeigt: Klare Verbesserungen in den Wahrscheinlichkeitsvorhersagen ergeben sich für ein stochastisch gestörtes und ein Multiphysik-Ensemblesystem, während der Multiparameteransatz nur wenig Einfluss hat. Am besten sind die Ergebnisse, wenn alle Methoden kombiniert werden. Daraus schließen wir, dass zu diesem Zeitpunkt keine Methode allein den komplexen Charakter von Modellfehlern umfassend beschreiben kann.

Abstract

Focusing on short-range weather ensemble-forecasts the need for model-error representations is established. Various model-error methods are being introduced and their merits and disadvantages discussed. The performance of the different model-error representations is documented for a particular meso-scale ensemble system: for this ensemble system stochastic perturbations and the multi-physics approach resulted in clearly better probabilistic forecasts, while the multi-parameter method showed little impact. Best results were obtained by combining all model-error schemes suggesting that no single representation alone can currently describe the complex nature of model error.

1 Einführung

Die zentrale Aufgabe der numerischen Wettervorhersage ist es, die meso- und synoptischskalige Wetterentwicklung so genau wie möglich vorherzusagen. Dies wird dadurch erschwert, dass selbst kleine Ungenauigkeiten in den Anfangsbedingungen oder in der Modellformulierung zu großskaligen Fehlern in der Vorhersage führen und die Vorhersagbarkeit der exakten Wetterentwicklung begrenzen (siehe auch Beitrag 1 von H. WERNLI in diesem Heft). Eine Möglichkeit, nicht nur eine wahrscheinliche Wetterentwicklung, sondern auch deren Vorhersagbarkeit zu messen, sind Wahrscheinlichkeitsvorhersagen mittels probabilistischer Ensemblesysteme (zum Beispiel MOLTENI und PALMER 1993, TOTH und KALNAY 1993).

Um den Fehler in den Anfangsbedingungen zu berücksichtigen, ist es heutzutage üblich, jedes Ensemblemitglied von einer leicht gestörten Anfangsbedingung zu starten (mehr dazu in Beitrag 2 von M. LEUTBECHER). Oft werden dabei Störungen gewählt, die größtmöglichstes Fehlerwachstum erzielen, zum Beispiel singuläre Vektoren (MOLTENI und PALMER, 1993) oder „bred vectors“ (TOTH und KALNAY 1993, HOUTEKAMER et al. 1996). Selbst mit diesen Anfangsbedingungsstörungen, sind Ensemblevorhersagen generell „unterdispersiv“

und überschätzen die Vorhersagbarkeit der atmosphärischen Entwicklung (BUIZZA et al. 2005). Dies resultiert in unverlässlichen Wahrscheinlichkeitsvorhersagen, insbesondere für seltene Zustände wie zum Beispiel extreme Wettersituationen. Letztere haben jedoch große gesellschaftliche Auswirkungen und ihre Vorhersage ist von zentraler Bedeutung.

Ein weiterer Beitrag zur Vorhersageunsicherheit ist Modellunsicherheit (oder Modellfehler). Numerische Wettermodelle bestehen essentiell aus zwei Komponenten: einem dynamischen Kern und den physikalischen Parametrisierungen. Als dynamischen Kern bezeichnet man die Algorithmen, die die Differenzialgleichungen, welche die atmosphärische Dynamik beschreiben, lösen. Die physikalischen Parametrisierungen sind traditionell deterministische Formulierungen, die den gitterskaligen Einfluss der nicht aufgelösten physikalischen Prozesse (zum Beispiel Konvektion, Strahlung, Grenzschicht und so weiter) beschreiben. Diese Formulierungen haben oft eine Reihe von freien Parametern, die idealerweise von Beobachtungen oder aus theoretischen Erwägungen abgeleitet werden sollten, was leider nicht immer möglich ist. Ein Grund dafür ist, dass die Beobachtungen zu spärlich sind oder aber dass die Formulierungen der Parametrisierungen selbst unzulänglich sind. Dies macht einen gewissen Grad von „Mo-

delluning" unvermeidbar, das heißt verschiedene mögliche Parameter werden getestet und diejenigen gewählt, welche die besten Modellvorhersagen erzielen.

In den letzten Jahren ist mit der Entwicklung von Methoden zur Berücksichtigung von Modellunsicherheit in Ensemblevorhersagen ein neues und wichtiges Gebiet erwachsen (BUIZZA et al. 1999, PALMER 2001, SHUTTS 2005, BERNER et al. 2009, BOWLER et al. 2009, PLANT und CRAIG 2008, PALMER et al. 2009, CHARRON et al. 2010, TENNANT et al. 2011). Diese Methoden lassen sich in zwei Gruppen klassifizieren:

1. Multimodell-Ensemblesysteme (zum Beispiel HOUTEKAMER et al. 1996, HAGEDORN et al. 2005), die unter anderem Multiparameter- (STAINFORTH et al. 2005) und Multi-Physikansätze (MURPHY et al. 2004) umfassen, und
2. stochastische Parametrisierungen oder stochastisch-dynamische Vorhersagen (EPSTEIN und PITCHER 1972, PALMER 2001), bei denen die Unsicherheit in den nichtaufgelösten Prozessen durch stochastische Terme modelliert wird (zum Beispiel BUIZZA et al. 1999).

Bislang gibt es noch keinen eindeutigen wissenschaftlichen Konsensus, welche Methoden für welche Anwendungen am besten geeignet sind.

Alle oben genannten Methoden versuchen auf verschiedene Art und Weise, den Einfluss der nichtaufgelösten Skalen auf die aufgelöste Gitterskala zu berücksichtigen und ändern sowohl die Variabilität (zum Beispiel BERNER et al. 2009, 2011) als auch den mittleren Zustand des Modells (zum Beispiel BERNER et al. 2012). Dabei steht zur Debatte, ob es ausreichend ist, den mittleren Einfluss der subskaligen Fluktuationen über verbesserte deterministische Parametrisierungen zu repräsentieren, oder ob eine explizit stochastische Formulierung der Fluktuationen notwendig ist.

Dieses Manuskript ist folgendermaßen strukturiert: In Abschnitt 2 werden die verschiedenen Methoden zur Berücksichtigung von Modellfehlern vorgestellt: Dabei legen wir vor allem Wert auf die Güte der verschiedenen Methoden bezüglich kurz- und mittelfristiger Wettervorhersagen und diskutieren Aspekte, die für eine operationelle Implementierung relevant sind. Abschnitt 3 enthält ausgewählte Beispiele für die Güte der verschiedenen Methoden und abschließende Bemerkungen sind in Abschnitt 4 zusammengefasst.

2 Methoden zur Berücksichtigung von Modellfehlern in Ensemblesystemen

2.1 Multimodell-Ensemblesysteme

Bei Multimodell-Ensemblesystemen ist jedes Ensemblemitglied ein separates Modell. Im Allgemeinen können Multimodelle verschiedene physikalische Parametrisierungen, verschiedene dynamische Kerne und verschiedene Parameter haben. Ein Ensemblesystem ist ein Multimodell,

wenn verschiedene Ensemblemitglieder unterschiedliche Klimatologien haben.

Die Herausforderung bei der Interpretation von Multimodellen ist, dass alle Ergebnisse vor dem Hintergrund der verschiedenen zugrunde liegenden Klimatologien verstanden werden müssen. Zum Beispiel muss quantifiziert werden, ob ein guter „Gesamt-Spread“ (für eine Definition von „Spread“ siehe Abschnitt 4.2 und Beitrag 2 von LEUTBECHER in diesem Heft) die Folge divergierender Trajektorien ist, oder aber dadurch bedingt ist, dass jedes Ensemblemitglied einen anderen systematischen Fehler („Bias“) hat. Aus diesem Grund erschweren Multimodell-Ensemblesysteme operationelle Nachbehandlung („Post-processing“) wie zum Beispiel statistische Kalibrierung. Im Fall von Multimodellen muss diese für jedes Ensemblemitglied separat vorgenommen werden und benötigt dadurch mehr Ressourcen. In den folgenden Paragraphen werden verschiedene Spezialfälle von Multimodellen im Detail diskutiert.

2.1.1 Multiphysik-Parametrisierungen

Bei der Multiphysik-Methode bekommt jedes Ensemblemitglied seinen eigenen Satz physikalischer Parametrisierungsalgorithmen. Bei der Zusammenstellung werden normalerweise Experten konsultiert, da von bestimmten Kombinationen, zum Beispiel von Grenzschicht- und Konvektionsparametrisierungen, abzuraten ist. Selbst wenn die Wahl umsichtig gemacht wurde, sind die letztendlichen Zusammenstellungen oft etwas heuristisch und eine systematischere Vorgehensweise wäre begrüßenswert. Ein Vorteil dieser Methode ist, dass die verschiedenen Ensemblemitglieder oft große Diversität und hervorragende Divergenzeigenschaften besitzen. Ein Grund mag sein, dass bestimmte Parametrisierungen in ausgewählten Wittersituationen bessere Vorhersagen erzeugen als andere, und dass die Bandbreite der Möglichkeiten durch die Multiphysik-Methode am besten abgedeckt wird.

Aus operationeller Sicht birgt diese Vorgehensweise die Herausforderung, dass viele verschiedene Parametrisierungen entwickelt und unterhalten werden müssen, was erhebliche Ressourcen -vor allem in Expertenwissen für die verschiedenen physikalischen Parametrisierungen- erfordert.

2.1.2 Multiparameter-Parametrisierungen

Beim Multiparameter-Ansatz werden verschiedene Werte für die Parameter in den physikalischen Parametrisierungen gewählt. Ein realistischer Bereich für die möglichen Werte wird meist nach Absprache mit den Parametrisierungsexperten definiert (zum Beispiel HACKER et al. 2011a). Manchmal werden die Parameter innerhalb des realistischen Bereiches zufällig gewählt. Oft jedoch wird die Wahl so vorgenommen, dass der Phasenraum möglichst gleichmäßig abgedeckt wird (zum Beispiel mittels der „Latin hypercube“-Methode). Ein Vorteil der Multiparameter-Methode aus operationeller Perspektive ist, dass nur ein

Satz von physikalischen Parametrisierungen unterhalten werden muss.

In der jahreszeitlichen- und Klimamodellierung hat sich die Multiparameter-Methode sehr bewährt (zum Beispiel MURPHY et al. 2004, STAINFORTH et al. 2005, DOBLAS-REYES et al., 2009) und wird routinemäßig dazu verwendet, Unsicherheiten in der Vorhersage abzuschätzen. Für die Wettervorhersage auf kürzeren Zeitskalen (siehe Beitrag 6 von S. THEIS und C. GEBHARDT in diesem Heft) ergibt sich oft das Problem, dass die Divergenz der Trajektorien nicht so groß wie erwünscht ist (HACKER et al., 2011a).

2.1.3 Abhängigkeit vom dynamischen Kern

Im Allgemeinen wird angenommen, dass die führende Ursache für Modellfehler Unsicherheiten in den physikalischen Parametrisierungen sind. Erst in letzter Zeit werden zunehmend auch Unsicherheiten in der Formulierung des dynamischen Kernes erwogen und systematisch untersucht. Im Zentrum der Diskussion stehen dabei das Diffusionsverhalten verschiedener numerischer Kerne als auch Wechselwirkungen zwischen dynamischem Kern und physikalischen Parametrisierungen, die zum Beispiel bei tropischen Wirbelstürmen eine wichtige Rolle spielen (REED und JABLONOWSKI 2012).

2.1.4 „Poor Man’s“-Ensemblesysteme

Eine spezielle Untergruppe der Multimodelle sind solche, bei der Modellläufe von verschiedenen Forschungszentren als Mitglieder eines einzigen Ensemblesystems interpretiert werden. So basieren zum Beispiel die Mehrzahl der Klimaänderungs-Wahrscheinlichkeitsvorhersagen der AR4 IPCC auf solchen multizentralen Modellen. Ein Beispiel für die Regionalmodellierung ist das SRNWP-PEPS, ein Ensemblesystem, welches Vorhersagen der Short-Range Numerical Weather Prediction, (SRNWP), die an verschiedenen europäischen Wetterzentren produziert werden, kombiniert. Diese Vorgehensweise erhält die Modellvielfalt und spart Ressourcen in dem Sinne, dass die jeweiligen Forschungszentren sich auf die Entwicklung eines einzigen Modelles konzentrieren können und sich nur begrenzt mit Ensembledesign und Initialisierungsproblemen beschäftigen müssen. Deshalb werden sie manchmal auch als „Poor Man’s“-Ensemblesysteme bezeichnet. Generell teilen sie die Vor- und Nachteile von Multimodellen, wie oben diskutiert.

2.2 Stochastische Parametrisierungen

Eine Alternative zu Multimodell-Ensemblesystemen sind stochastisch gestörte Ensemblesysteme, bei denen Modellunsicherheiten mittels Zufallszahlen repräsentiert werden. Obwohl schon länger postuliert (EPSTEIN und PITCHER 1972, PALMER 2001), wurden sie erst in den letzten zwei Jahrzehnten in der numerischen Wettervorhersage operationell eingesetzt (BUIZZA et al. 1999, SHUTTS 2005, BERNER et al. 2009, BOWLER et al. 2009, CHARRON

et al. 2010, TENNANT et al. 2011). Während sie zweifellos die Vorhersagen verbessern, stehen noch viele wissenschaftliche Fragen offen. Ein großer Vorteil von stochastischen Parametrisierungen ist, dass alle Ensemblemitglieder verschiedene Klimatologien haben und somit statisches Postprocessing erleichtern. Jedoch ist die Klimatologie der gestörten Mitglieder in der Regel verschieden von der des ungestörten Modells, da stochastische Störungen in der Regel nicht nur die Variabilität, sondern auch den mittleren Zustand beeinflussen (zum Beispiel BERNER et al. 2012).

2.2.1 Stochastisch gestörte physikalische Tendenzen

Die erste operationell angewandte stochastische Parametrisierung nahm die Tendenzen der physikalischen Parametrisierungen und multiplizierte sie mit einer Zufallszahl zwischen +/- 0,5 (BUIZZA et al. 1999). Diese Vorgehens-

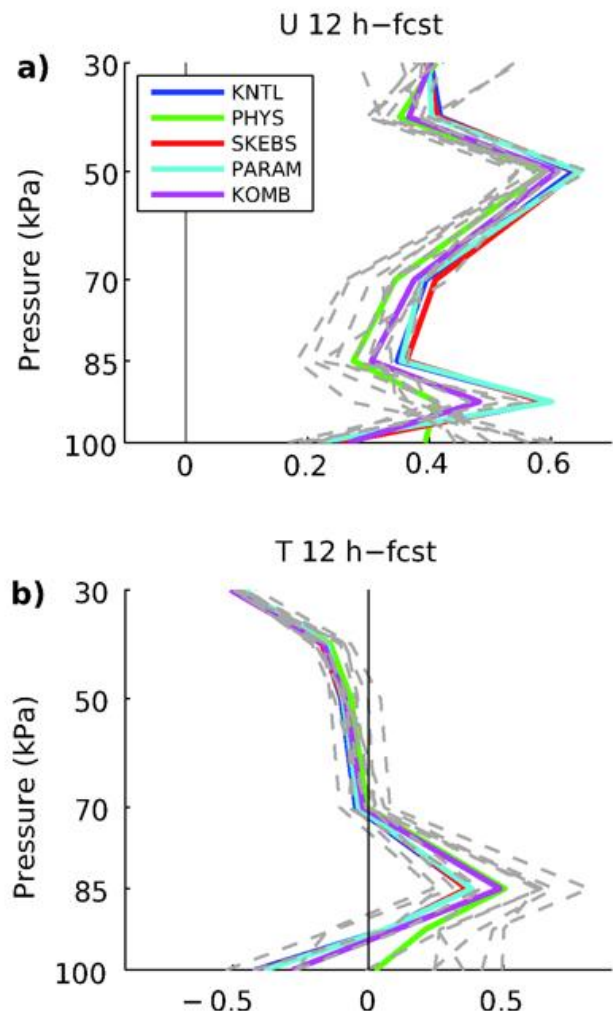


Abb. 3-1: Systematischer Fehler der Ensemble-Mean-Vorhersage für die Periode 21. November bis 21. Dezember 2008 als Funktion des Drucks für den a) zonalen Wind U in m/s und b) die Temperatur T in K. Gezeigt ist der systematische Fehler für die Experimente mit Kontrollphysik (KNTL, blau), Multiphysik (PHYS, grün), stochastische Rückstreuung von kinetischer Energie (SKEBS, rot), Multiparameter-Methode (PARAM, zyanblau) und einer Kombination aller dieser Methoden (KOMB, magentarot). Dünne graue Linien bezeichnen die einzelnen Ensemblemitglieder des Experimentes PHYS.

Experiment	Methode zur Berücksichtigung von Modellfehlern	Farbe	Referenz
KNTL	Kontrollphysik	blau	HACKER et al. (2011b)
PHYS	Multiphysik	grün	HACKER et al. (2011b) Berner et al. (2011)
SKEBS	Stochastische Rückstreuung von kinetischer Energie	rot	BERNER et al. (2011)
PARAM	Multiparameter	zyanblau	HACKER et al. (2011a)
KOMB	Kombination aller Modellfehlermethoden	magentarot	HACKER et al. (2011b)

Tabelle 3-1: Zusammenfassung der Modellexperimente.

weise stellt sicher, dass die Störungen zirkulationsabhängig sind (multiplikatives Rauschen): In Regionen, in denen die Tendenzen klein sind, wird nur wenig gestört. Dabei wird dieselbe Zufallszahl entlang der gesamten vertikalen Säule benutzt, um der Bildung von unphysikalischen Schwellen vorzubeugen. Die Zufallszahlen werden über einen gewissen Zeitraum (etwa 6 Stunden) und geographischen Bereich (etwa 100 km) konstant gehalten, um zeitliche und räumliche Korrelationen einzuführen, die für das Fehlerwachstum zu größeren Skalen während des Vorhersagezeitraums nötig sind.

Kürzlich wurde eine überarbeitete Version dieser Parametrierung am Europäischen Zentrum für mittelfristige Wettervorhersage (EZMW) operationell eingeführt (PALMER et al. 2009). Nun wird das Zufallsfeld spektral erzeugt und variiert die zeitlichen Korrelationen mit der räumlichen Skala. Der zugrunde liegende Gedanke ist, dass kleine räumliche Störungen mit kurzen Zeitskalen und große räumliche Störungen mit längeren Zeitskalen assoziiert sein sollten.

2.2.2 Stochastische Rückstreuung von kinetischer Energie

In den letzten Jahren hat sich eine weitere stochastische Methode zunehmend durchgesetzt: das „stochastic kinetic-energy backscatter“ (SKEBS) oder übersetzt die „Stochastische Rückstreuung von kinetischer Energie“.

In der Atmosphäre wird ständig Energie zwischen kleinen und großen Skalen ausgetauscht. Nach heutigem Verständnis wird auf den Skalen, die von homogener Turbulenz dominiert werden, im Netto mehr Energie auf kleine Skalen übertragen als umgekehrt. Die Energie wird zu zunehmend kleineren Skalen übertragen und letztendlich als Wärme dissipiert. Wettermodelle beschreiben diese Energiekaskade, haben aber unter Umständen zu viel Dissipation, sowohl in Form von zu dissipativen numerischen Algorithmen als auch von zu dissipativen physikalischen Prozessen.

Die stochastische Rückstreuung von kinetischer Energie basiert auf der Annahme, dass ein Bruchteil dieser Energie als kinetische Energie auf größeren Skalen rückgestreut wird und als Energiequelle für wachsende Störungen betrachtet werden kann (SHUTTS 2005). Diese Energiequelle wird dabei als Zufallsfeld modelliert und ihre Amplitude so gewählt, dass in stark dissipativen Regionen (nahe Fronten,

Orographie und in konvektiven Regionen) mehr Energie rückgestreut wird als in schwach dissipativen (SHUTTS 2005, BERNER et al. 2009, 2011).

2.2.3 Stochastische Variation von Parametern

Eine stochastische Version des Multiparameter-Ansatzes (Diskussion siehe oben) variiert einen oder mehrere physikalische Parameter auf stochastische Weise (zum Beispiel BOWLER et al. 2009). Der dahinterstehende Gedanke ist, dass dadurch Ensemblevielfalt entsteht, ohne verschiedene Klimatologien zu erzeugen. Ob dieses Ziel wirklich erreicht wird, hängt von der Länge der Vorhersage und des zeitlichen Gedächtnisses der stochastischen Modulation ab. Für Kurzfristvorhersagen haben die durch diese Methode gestörten Ensemblemitglieder effektiv meist unterschiedliche Klimatologien.

2.2.4 Stochastische Formulierung der physikalischen Parametrisierungen

Idealerweise sollten stochastische Formulierungen nicht separat entwickelt und dann in existierende Ensemblesysteme implementiert werden, sondern sollten direkt im Zusammenhang mit den physikalischen Parametrisierungen entwickelt werden. Dies ist nicht trivial, da zum Beispiel subskalige Fluktuationen in den Konvektionsparametrisierungen mit Schwankungen in der Mikrophysik korreliert und somit die stochastischen Störungen über mehrere Parametrisierungen hinweg konsistent sein sollten. Mehrere Forschungsprojekte mit diesem Ziel sind im Gespräch und sollten zu relevanten Ergebnissen in den nächsten Jahren führen (siehe zum Beispiel EZMW Report on „Representing Model Uncertainty in numerical weather prediction models and in climate models“, erhältlich unter <http://www.ecmwf.int/publications/>).

Eine wichtige Untergruppe sind Wolkenparametrisierungen, die auf Methoden der statistischen Mechanik basieren und Konvektion als makrophysikalisches Zustandsgleichgewicht modellieren (PLANT und CRAIG 2008).

3 Modellexperimente

In diesem Abschnitt zeigen wir, wie sich die verschiedenen Methoden zur Berücksichtigung von Modellfehlern in einem mesoskaligen Ensemblesystem bewähren. Das Ensemblesystem basiert auf dem „Weather Research und Forecasting“-Modell (WRF) und seine Konfiguration ist im Detail in HACKER et al. (2011b) beschrieben. Die Region wurde auf die Vereinigten Staaten beschränkt und alle Experimente haben eine horizontale Auflösung von 45 km.

Das Kontrolleexperiment besteht aus 10 Ensemblemitgliedern, die als Anfangs- und Randbedingungen die inter-

polierten Felder des globalen Ensemblesystems „Global Ensemble Forecast System“ erhalten. Alle Ensemblemitglieder des Kontrollexperimentes haben dieselben physikalischen Parametrisierungen und unterscheiden sich nur in den Anfangs- und Randbedingungen.

Zusätzlich wurden eine Reihe von Experimenten durchgeführt, die die verschiedenen Methoden zur Berücksichtigung von Modellfehlern illustrieren. Alle Experimente benutzen dieselben Anfangs- und Randbedingungen wie das Kontrollexperiment und haben zehn Ensemblemitglieder. Sie sind in Tab. 3-1 zusammen mit den Publikationen, die ihre Implementierungsdetails beschreiben, zusammengefasst.

Das Ensemblesystem wurde zweimal täglich initialisiert (00 UTC und 12 UTC) und während der Periode vom 21. November bis 21. Dezember 2008 jeden zweiten Tag gestartet und für 60 Stunden integriert. Dies resultiert in einer Stichprobenlänge von insgesamt 32 probabilistischen Vorhersagen. Alle Ergebnisse wurden gegen Radiosondenbeobachtungen von 106 Stationen über den kontinentalen Vereinigten Staaten verifiziert. Im Folgenden beschränken wir uns auf die Variablen zonaler Wind U und Temperatur T und eine Vorhersagezeit von 12 Stunden.

4 Ergebnisse

4.1 Systematischer Fehler

Der systematische Fehler oder „Bias“ ist die Differenz zwischen dem einmonatigen Mittel in den Beobachtungen und dem Modell. Das WRF-Modell hat im Mittel zu schwache Zonalwinde (Abb. 3-1a).

In den Experimenten, die verschiedene physikalische Parametrisierungen benutzen (PHYS und KOMB), ändert sich der systematische Fehler grundlegend: in der freien Atmosphäre ist er reduziert und für PHYS am Boden erhöht. Für Temperatur T ist der Bias für PHYS reduziert und in Bodennähe gar vernachlässigbar. Eine genauere Untersuchung der einzelnen Ensemblemitglieder (graue Linien) ergibt, dass jedes Mitglied des Multiphysik-Experimentes einen anderen systematischen Fehler hat, die verschiedenen Fehler sich aber (zufällig) im Mittel zu Null addieren (Abb. 3-1b).

Die Benutzung verschiedener Parameterwerte (PARAM) wirkt sich nur gering auf den mittleren Fehler aus.

4.2 Divergenz der Trajektorien

Ein Maß für die Verlässlichkeit von Wahrscheinlichkeitsvorhersagen ist die Konsistenz zwischen dem mittleren Vorhersagefehler und dem „Spread“. Der mittlere Fehler beschreibt die Differenz zwischen der über die Ensemblemitglieder gemittelten Ensemble-Mean-Vorhersage und den Beobachtungen („root mean square error of the ensemble mean forecast“). Der „Spread“ beschreibt die Spann-

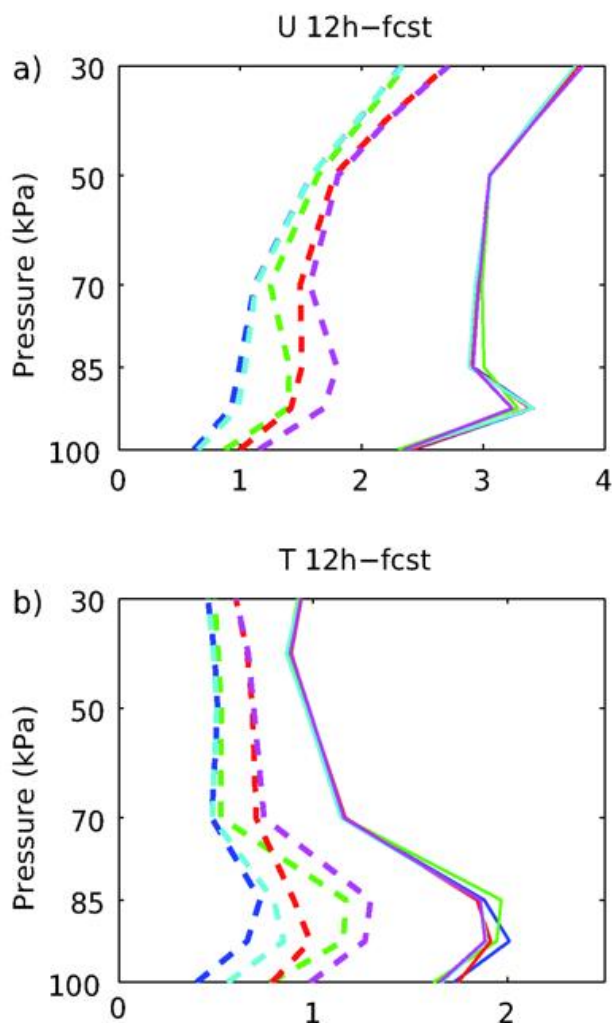


Abb. 3-2: „Spread“ um die Ensemble-Mean-Vorhersage (Strichlinien) und mittlerer Fehler der Ensemble-Mean-Vorhersage (durchgezogene Linien) für eine Vorhersagezeit von 12 Stunden für den a) zonalen Wind U in m/s und b) die Temperatur T in K. Der systematische Fehler wurde vor der Berechnung abgezogen. Für eine Beschreibung der verschiedenen Experimente siehe Tab. 3-1 und Bildunterschrift von Abb. 3-1.

weite der Trajektorien, das heißt, wie stark die Trajektorien der einzelnen Ensemblemitglieder während der Vorhersagefrist divergieren. In einem zuverlässigen Ensemblesystem sollten der mittlere Fehler und der „Spread“ dieselbe Wachstumsrate und Amplitude haben.

Abbildung 3-2 zeigt den mittleren Fehler (durchgezogen) und „Spread“ (gestrichelt) für eine Vorhersagezeit von 12 Stunden. Wir sehen, dass alle Ensemblesysteme unterdispersiv sind, das heißt, dass die Ensemblespanne kleiner als der mittlere Fehler ist. Dies ist ein eindeutiger Hinweis auf die Existenz von Modellfehlern auf dieser Zeitskala. Es muss betont werden, dass hier Messfehler in den Beobachtungen vernachlässigt wurden, was den „Gesamtspread“ vergrößern würde. Des Weiteren würde eine Anzahl von mehr als zehn Ensemblemitgliedern das Divergenz-Fehler-Verhältnis positiv beeinflussen. BERNER et al. (2011) diskutiert dies im Detail und zeigt, dass die Ensemblesysteme auch unterdispersiv sind, wenn eine Schätzung des Messfehlers mit einbezogen wird.

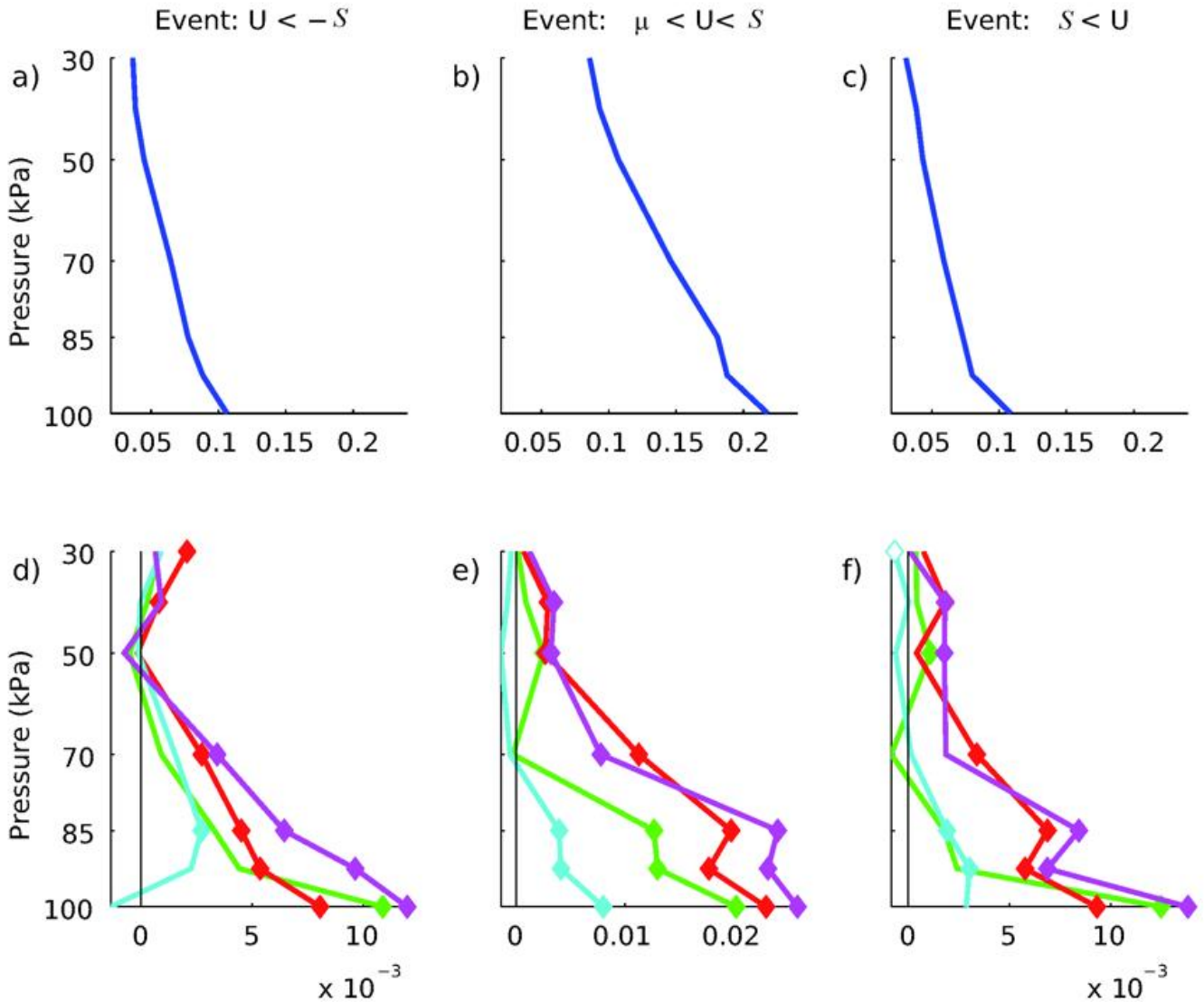


Abb. 3-3: „Brierscore“ für den zonalen Wind U und eine Vorhersagezeit von 12 Stunden als Funktion des Drucks. Gezeigt werden die Werte für das Experiment KNTL und drei Verifikationsereignisse: a) $u(\lambda, \varphi, p) < -\sigma_u(\lambda, \varphi, p)$, b) $\mu_u < u(\lambda, \varphi, p) < \sigma_u(\lambda, \varphi, p)$ und c) $\sigma_u(\lambda, \varphi, p) < u$, wobei μ_u das klimatologische Mittel und $\sigma_u(\lambda, \varphi, p)$ die klimatologische Standardabweichung von U als Funktion der geographischen Länge λ , Breite φ und des Drucks p ist. Das Profil für den Fall $-\sigma_u(\lambda, \varphi, p) < u(\lambda, \varphi, p) < \mu_u$ ähnelt dem Profil in b) und ist deshalb nicht gezeigt. Kleinere Werte für den „Brierscore“ bedeutet bessere Vorhersagequalität. d-f) „Brier skillscore“ der verschiedenen Experimente für dieselben Verifikationsereignisse, wobei als Referenz die Vorhersagegüte des Experimentes KNTL gewählt wurde. Positive Werte bedeuten eine Verbesserung der Vorhersagbarkeit gegenüber der des Experimentes KNTL. Gefüllte (offene) Symbole geben die statistische Signifikanz der Verbesserung (Verschlechterung) mit einem Konfidenzniveau von 95 % an. Für eine Beschreibung der verschiedenen Experimente siehe Tab. 3-1 und Bildunterschrift von Abb. 3-1.

Des weiteren demonstriert Abbildung 3-2, dass die Methoden zur Berücksichtigung von Modellfehlern den „spread“ vergrößern, ohne aber den Fehler der mittleren Vorhersage zu verschlechtern. Parametervariationen (PARAM) haben dabei den kleinsten Einfluss und stochastische Parametrisierungen (SKEBS) und die Multiphysik-Methode den größten. Die Kombination der verschiedenen Methoden (KOMB) erzielt die besten Ergebnisse, vor allen in Bodennähe und in der Grenzschicht. Insgesamt ist der Einfluss der verschiedenen Methoden auf U und T qualitativ vergleichbar, bis auf den Wert in 850 hPa, wo für Temperatur das Multiphysik-Experiment PHYS eindeutig mehr Dispersion als SKEBS erzielt. Der Einfluss auf den mittleren Fehler ist klein, aber interessanterweise weist

das Experiment KOMB den kleinsten mittleren Fehler und den größten „spread“ auf, was in der besten Konsistenz resultiert.

4.3 „Brierskillscore“

Während eine gute „Spread-Error-Konsistenz“ wünschenswert ist, ist das ultimative Kriterium für eine gute Wahrscheinlichkeitsvorhersage die probabilistische Güte oder „Skill“. Hier zeigen wir Ergebnisse für den „Brierscore“. Der „Brierscore“ ist so orientiert, dass eine Wahrscheinlichkeitsvorhersage umso besser ist, je kleiner der „Brierscore“ ist. Für eine Definition und Details der Berechnung verweisen wir auf zum Beispiel auf den Beitrag

4 von A. P. Weigel in diesem Heft. Die obere Reihe von Abbildung 3-3 zeigt „Brierscore“-Profile für drei verschiedene Verifikationsereignisse (siehe Bildunterschrift) für das Kontrollensemble KNTL. Sowohl für „häufige“ Verifikationsereignisse innerhalb einer Standardabweichung (Abb. 3-3b) als auch für „extreme“ Verifikationsereignisse von mehr als einer Standardabweichung (Abb. 3-3a,c) sind die Wahrscheinlichkeitsvorhersagen am besten in der freien Atmosphäre und schlechter in der Nähe des Bodens. Der „Brier skill score“ (BSS) der verschiedenen Modellexperimente unter Berücksichtigung von Modellfehlern ist in der unteren Reihe gezeigt, wobei der „Brierscore“ des Kontrollensembles, BS_{ctr} , als Referenz benutzt wurde:

$$BSS = 1 - (BS_{exp} / BS_{ctr}).$$

Ein positiver „Skill score“ bedeutet eine Verbesserung der Vorhersage im Vergleich zum Kontrollensemble.

Das Ergebnis ist eindeutig: Alle Methoden zur Berücksichtigung von Modellfehlern geben eine bessere Wahrscheinlichkeitsvorhersage als das Kontrollensemble, und die Verbesserung ist in den meisten Druckniveaus statistisch signifikant mit 95%-iger Konfidenz (gefüllte Symbole). Das beste Ergebnis liefert das Experiment KOMB, gefolgt von SKEBS, PHYS und PARAM, wobei in Bodennähe und für große Anomalien PHYS besser als SKEBS ist. Wie gezeigt ist dieses Ergebnis signifikant für den gewählten Zeit- und Regionalraum, muss aber noch für andere Regionen und Jahreszeiten separat bestätigt werden.

5 Zusammenfassung

Verschiedene Methoden zur Berücksichtigung von Modellfehlern in der numerischen Wettervorhersage wurden vorgestellt und ihre Vor- und Nachteile diskutiert. Mehrere Methoden wurden in ein mesoskaliges Ensemblesystem implementiert und ihr Einfluss auf die Verbesserung kurzfristiger Wahrscheinlichkeitsvorhersagen analysiert. Für das vorliegende Modell erzielen Ensemblevorhersagen mit einer stochastischen Störung oder aber basierend auf der Multiphysik-Methode die besten Ergebnisse, während die Multiparameter-Methode nur wenig Einfluss hat. Am allerbesten jedoch sind Wahrscheinlichkeitsvorhersagen von einem Ensemblesystem, das alle Methoden für die Berücksichtigung von Modellfehlern kombiniert.

Während nicht ausgeschlossen werden kann, dass die Details der hier vorgestellten Ergebnisse modell- und implementierungsabhängig sind, ist das Grundergebnis wahrscheinlich allgemein gültig: Modellfehler sind so komplex, dass sie mittels mehr als einer einzigen Methode allein repräsentiert werden müssen. Dieses Ergebnis wurde in mehreren Studien bestätigt, zum Beispiel PALMER et al. (2009), CHARRON et al. (2010), HACKER et al. (2011b) und BERNER et al. (2011).

Danksagung

Ich danke Dr. Josh Hacker, dass er mir die Daten für die Analyse der Experimente PARAM und KOMB zur Verfügung gestellt hat. Dank geht an Dr. Soyoung Ha, Dr. Josh Hacker und Dr. Chris Snyder für viele ergiebige Diskussionen. Die Kommentare von Dr. Renate Hagedorn zu einer ersten Version dieses Artikels waren sehr hilfreich, vor allem der Vorschlag den „Brier skill score“ anstatt Brierscoredifferenzen zu benutzen. Die Vorschläge von Prof. George Craig haben diesen Artikel verständlicher und umfassender gemacht. Frau Elfriede Berner sei für umsichtiges Korrekturlesen gedankt.

Literatur

- BERNER, J., SHUTTS G., LEUTBECHER M., T. PALMER, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.* **66**, 603–626.
- BERNER, J., HA S.-Y., HACKER J. P., FOURNIER A., SNYDER C., 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multi-physics representations. *Mon. Wea. Rev.* **139**, 1972–1995.
- BERNER, J., JUNG T., PALMER T. N., 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate* **25**, 4946–4962.
- BOWLER, N. E., ARRIBAS A., BEARE S. E., MYLNE K. R., SHUTTS G. J., 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.* **135**, 767–776.
- BUIZZA, R., MILLER M., PALMER T. N., 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.* **125**, 2887–2908.
- BUIZZA, R., HOUTEKAMER P. L., TOTH Z., PELLERIN G., WEI M., ZHU Y., 2005: A comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.* **133**, 1076–1097.
- CHARRON, M., PELLERIN G., SPACEK L., HOUTEKAMER P. L., GAGNON N., MITCHELL H. L., MICHELIN L., 2010: Toward random sampling of model error in the canadian ensemble prediction system. *Mon. Wea. Rev.* **138**, 1877–1901.
- DOBLAS-REYES, F., WEISHEIMER A., DÉQUÉ M., KEENLYSIDE N., MCVEAN M., MURPHY J.M., ROGEL P., SMITH D., PALMER T.N., 2009: Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Quart. J. Roy. Meteor. Soc.* **135**, 1538–1559.
- EPSTEIN, E. S., PITCHER E. J., 1972: Stochastic analysis of meteorological fields. *J. Atmos. Sci.* **29**, 244–257.
- HACKER, J. P., SNYDER C., HA S.-Y., POCERNICH M., 2011A: Linear and nonlinear response to parameter variations in a mesoscale model. *Tellus A*, **63**, 429–444.

- HACKER, J. P., ET AL., 2011B: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus A*, 63, 625–641, DOI: 10.1111/j.1600-0870.2010.00497.x.
- HAGEDORN, R., DOBLAS-REYES F. J., PALMER T. N., 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting – i. basic concept. *Tellus* **57A**, 219–233.
- HOUTEKAMER, P. L., LEFAIVRE L., DEROME J., RITCHIE H., MITCHELL H. L., 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**, 1225–1242.
- MOLTENI, F., PALMER T. N., 1993: Predictability finite-time instability of the Northern winter circulation. *Quart. J. Roy. Meteor. Soc.* **119**, 269–298.
- MURPHY, J., SEXTON D., BARNETT D., JONES G., WEBB M., COLLINS M., STAINFORTH D., 2004: Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772.
- PALMER, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction. *Quart. J. Roy. Meteor. Soc.* **127**, 279–304.
- PALMER, T. N., BUIZZA R., DOBLAS-REYES F., JUNG T., LEUTBECHER M., SHUTTS G., STEINHEIMER M., WEISHEIMER A., 2009: Stochastic Parameterization and Model Uncertainty. *ECMWF Technical Memorandum*, **598**, available at <http://www.ecmwf.int/publications/>.
- PLANT, R. S., CRAIG G. C., 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.* **64**, 87–105.
- Proceedings of ECMWF Workshop on Model-Error Representations in weather and climate models, June 24–28, 2011, available at <http://www.ecmwf.int/publications/>.
- REED, K. A., JABLONOWSKI C., 2012: Idealized tropical cyclone simulations of intermediate complexity: A test case for AGCMs *James* **4**, M04001, 25 pp.
- SHUTTS, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.* **612**, 3079–3102.
- STAINFORTH, D., AINA T., CHRISTENSEN C., COLLINS M., FAULL N., FRAME D. J., KETTLEBOROUGH J. A., KNIGHT S., MARTIN A., MURPHY J. M., PIANI C., SEXTON D., SMITH L. A., SPICER R. A., THORPE A. J., ALLEN M. R., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406.
- TENNANT, W. J., SHUTTS G. J., ARRIBAS S. A. T. A., 2011: Using a stochastic kinetic energy backscatter scheme to improve MOCREPS probabilistic forecast skill. *Mon. Wea. Rev.* **139**, 1190–1206.
- TOTH, Z., KALNAY E., 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.* **74**, 2317–2330.

A.P. WEIGEL

4 Verifikation von Ensemblevorhersagen¹

Verification of ensemble forecasts

Zusammenfassung

Ensemblevorhersagen werden in der Regel als Wahrscheinlichkeitsvorhersagen interpretiert, stellen aber a priori „lediglich“ eine endliche Menge deterministischer Einzelvorhersagen dar. Entsprechend dieser beiden Interpretationsniveaus gibt es zwei Kategorien von Ansätzen zur Verifikation von Ensemblevorhersagen: Zum einen Tests, welche Ensembles als Stichproben einer Verteilung interpretieren und die kollektiven statistischen Eigenschaften der Einzelvorhersagen untersuchen, zum anderen probabilistische Gütemaße, welche erfordern, dass bereits auf irgendeine Weise Wahrscheinlichkeitsvorhersagen aus den Ensembles abgeleitet worden sind. Für beide Kategorien wird hier ein kurzer Überblick über gängige Verifikationsansätze gegeben, und die zu Grunde liegenden Fragestellungen und Annahmen werden diskutiert.

Summary

While ensemble forecasts are usually interpreted as probabilistic forecasts, they represent a priori “only” finite sets of deterministic forecast realizations. These two levels of interpretation are reflected in the skill metrics commonly applied for ensembles, which can be categorized into two groups: on the one hand are tests that consider the individual ensemble members as discrete samples from a probability distribution and analyze their collective statistical properties, on the other hand are truly probabilistic skill metrics that require that some form of probability distribution has been derived from the ensemble members prior to verification. For both verification categories, common skill metrics are here described, and the underlying questions and assumptions are discussed.

1 Einleitung

Die Dynamik der Atmosphäre lässt sich mit ein paar wenigen Differentialgleichungen vollständig beschreiben. Diese Gleichungen sind seit langem bekannt, und ihre numerische Integration bildet das Fundament der klassischen Wettervorhersage. Auf den ersten Blick mag die numerische Wettervorhersage folglich als ein vollständig deterministisches und gelöstes Problem erscheinen. Die Wirklichkeit sieht jedoch ein bisschen anders aus, und zwar aus folgendem Grund: Für die numerische Integration der Gleichungen muss man den Zustand der Atmosphäre zum Anfangszeitpunkt der Rechnung kennen. Diese sogenannten Anfangsbedingungen lassen sich aber nicht mit unbegrenzter Genauigkeit bestimmen, sind also immer mit Unsicherheiten behaftet. Diese Unsicherheiten, so klein sie auch sein mögen, haben die Tendenz, sich mit der Zeit aufzuaddieren und das großskalige Strömungsmuster in einer nicht vorhersagbaren Weise zu beeinflussen. Aus diesem Grund spielen die Begriffe Zufall, Unsicherheit und Wahrscheinlichkeit eine große Rolle bei numerischen Wetterprognosen; eine Rolle, die umso größer wird, je weiter man in die Zukunft blicken will. Ensemble-

vorhersagen stellen einen pragmatischen Ansatz dar, diese Vorhersageunsicherheiten zu berücksichtigen und zu quantifizieren (LEITH 1974). Grob gesagt besteht die Idee darin, ein Prognosemodell mehrfach von leicht variierten Anfangsbedingungen ausgehend anzuwenden, um damit die Bandbreite möglicher Ereignisse abzuschätzen, die sich aus der nur ungenauen Kenntnis der Anfangszustände ergibt. Man betrachtet also nicht mehr nur das Ergebnis einer einzelnen Vorhersage, sondern analysiert die zeitliche Entwicklung eines ganzen Ensembles an Prognosen. Bei der Betrachtung von längeren Vorhersagezeiträumen (Wochen bis Monate) wird die Vorhersageunsicherheit zunehmend durch Unsicherheiten in relevanten Randbedingungen wie beispielsweise der Meeresoberflächentemperatur dominiert. Durch die Berücksichtigung dieser Unsicherheiten bei der Ensembleerzeugung und -modellierung kann das Ensemblekonzept auch auf Klimaskalen ausgeweitet werden, was zum Beispiel die Berechnung saisonaler Vorhersagen ermöglicht (STOCKDALE et al. 1998).

Wie kann man die Qualität (Fachbegriff: *Skill*) von Ensemblevorhersagen evaluieren? Zunächst einmal muss

¹ Dieser Artikel ist eine gekürzte und ins Deutsche übersetzte Version eines vom selben Autor verfassten Buchkapitels (Weigel 2011).

man festhalten, dass eine Ensemblevorhersage als solche noch keine Wahrscheinlichkeitsprognose darstellt, sondern „lediglich“ eine endliche Menge deterministischer Einzelvorhersagen. Um aus einem solchen Ensemble eine Vorhersageunsicherheit abzuschätzen, oder eine Eintrittswahrscheinlichkeit für ein spezifisches Ereignis anzugeben, müssen Annahmen über die statistischen Eigenschaften der Einzelvorhersagen getroffen werden. Die Frage, wie man Ensemblevorhersagen verifizieren soll, hängt also stark davon ab, wie man sie interpretiert. Je nachdem, ob man die Einzelvorhersagen eines Ensembles als Stichproben einer a priori unbekanntem Wahrscheinlichkeitsverteilung betrachtet, oder ob man aus dem Ensemble bereits in irgendeiner Form konkrete Wahrscheinlichkeiten abgeschätzt hat, müssen unterschiedliche Gütemaße (Fachbegriff: *Skill Scores*) angewandt werden.

Diejenigen *Skill Scores*, welche für die „Stichprobeninterpretation“ konzipiert sind, charakterisieren in der Regel die kollektiven statistischen Eigenschaften der Ensembles. Sie erfordern nicht, dass die Ensemblevorhersagen bereits vor der Verifikation in Wahrscheinlichkeitsverteilungen umgewandelt worden sind, sondern können im Prinzip auf die „rohen“ Einzelvorhersagen angewendet werden. Oftmals stellt diese Art der Verifikation den ersten Schritt zu einer später folgenden Wahrscheinlichkeitsabschätzung dar. Einige der für diesen Zweck gebräuchlichsten Tests werden in Kapitel 3 vorgestellt.

Sobald man Wahrscheinlichkeiten aus den Ensemblevorhersagen abgeleitet hat, auf welche Weise auch immer, kann eine andere Klasse von Gütemaßen verwendet werden, nämlich die sogenannten probabilistischen *Skill Scores*. Bei der Anwendung dieser Scores muss man allerdings beachten, dass sie nicht nur die Qualität des Wettermodells an sich messen, sondern auch, inwieweit die Annahmen zur probabilistischen Interpretation der Ensembles realistisch waren. Kapitel 4 wird diesen und andere Aspekte der probabilistischen Ensembleverifikation diskutieren und einige der geläufigsten probabilistischen *Skill Scores* vorstellen. Viele der in Kapitel 3 und 4 diskutierten Gütemaße werden mit Beispielen veranschaulicht. Die hierfür verwendeten Daten werden in Kapitel 2 beschrieben. Der Artikel schließt mit einer Zusammenfassung in Kapitel 5.

2 Datengrundlage

Die in diesem Text gezeigten Beispiele basieren auf drei Datenquellen: Erstens mittelfristigen Ensemblevorhersagen des EZMW-Ensemblevorhersagesystems „EPS“ (PALMER et al. 2007); zweitens saisonalen Ensemblevorhersagen des EZMW-Jahreszeitenvorhersagemodells „System 3“ (ANDERSON et al. 2007); und drittens synthetischen Ensemblevorhersagen, welche mit einem auf Gauß-Verteilungen basierenden stochastischen Modell generiert wurden („Toymodell“).

Bei den EPS-Beispielen wurden Vorhersagen der 500 hPa geopotenziellen Höhe in den Außertropen der nördlichen Hemisphäre (20-90°N) für Vorhersagezeiträume von 1 bis 15 Tagen ausgewertet. Die Vorhersagen waren während des Zeitraums Dezember 2009 bis Februar 2011 in Zwölfstundenintervallen berechnet worden. Die Vorhersageensembles hatten 51 Mitglieder und wurden gegen die operationelle Analyse des EZMW verifiziert.

Bei den System-3-Beispielen wurden Vorhersagen des Dreimonatsmittels (Dezember-Februar) der 2m-Temperatur für die Niño3.4-Region im äquatorialen Pazifik (5°S-5°N, 120-170°W) ausgewertet. Verwendet wurden Modellläufe für die Jahre 1960-2001, wobei das Modell jeweils mit den Anfangsbedingungen des 1. November initialisiert worden war. Die Ensembles hatten 9 Mitglieder. Falls systematische Modellfehler (Fachbegriff: *Bias*) vor der Verifikation entfernt wurden, werden die Vorhersagen als „biaskorrigiert“ bezeichnet. Falls zusätzliche Rechenschritte durchgeführt wurden, um die Breite der Ensembleverteilungen (Fachbegriff: *Spread*) a posteriori zu korrigieren, werden die Vorhersagen als „rekalibriert“ bezeichnet. Andernfalls sprechen wir von „rohen“ Vorhersagen. Für die Rekalibrierung wurde ein statistischer Algorithmus namens „Climate Conserving Recalibration (CCR)“ von WEIGEL et al. (2009) verwendet. Die Vorhersagen wurden gegen die vierzigjährigen Reanalyse des EZMW („ERA40“) verifiziert.

Für systematischere Analysen des Zusammenhangs zwischen *Skill* und spezifischen Ensembleeigenschaften eignet sich das Toymodell. Mit ihm können beliebig viele Vorhersage-Beobachtungspaare synthetisch erzeugt werden, und zwar so, dass vorgegebene Bedingungen bezüglich *Skill* und *Spread* erfüllt werden. Ein Vorhersage-Beobachtungspaar wird dabei in einem zweistufigen Prozess erzeugt: Im ersten Schritt wird das *vorhersagbare Signal* s mit einem Zufallsgenerator aus einer Normalverteilung mit Mittel 0 und Varianz 1 gezogen. Darauf aufbauend wird im zweiten Schritt eine Ensemblevorhersage mit m Mitgliedern, $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m)$, und eine zugehörige verifizierende Beobachtung, x , gemäß folgender Vorschrift erzeugt:

$$x = s + \varepsilon_x$$

$$\hat{x}_i = s + \varepsilon_i + \beta$$

$$\text{wobei } \varepsilon_x \sim N(0, \alpha)$$

$$\varepsilon_i \sim N(0, \gamma\alpha)$$

$\sim N(\mu, \sigma)$ steht hierbei für eine Zufallszahl, welche aus einer Normalverteilung mit Mittel μ und Varianz σ^2 gezogen wird. Der Term ε_x repräsentiert die unvorhersagbaren und chaotischen Komponenten der Beobachtungen, während ε_i die Störung des i -ten Ensemblemitglieds angibt. Das Toymodell hat drei freie Parameter α , β und γ . Der Parameter α kontrolliert das Signal-Rausch-Verhältnis. Ohne Einschränkung der Allgemeinheit ist

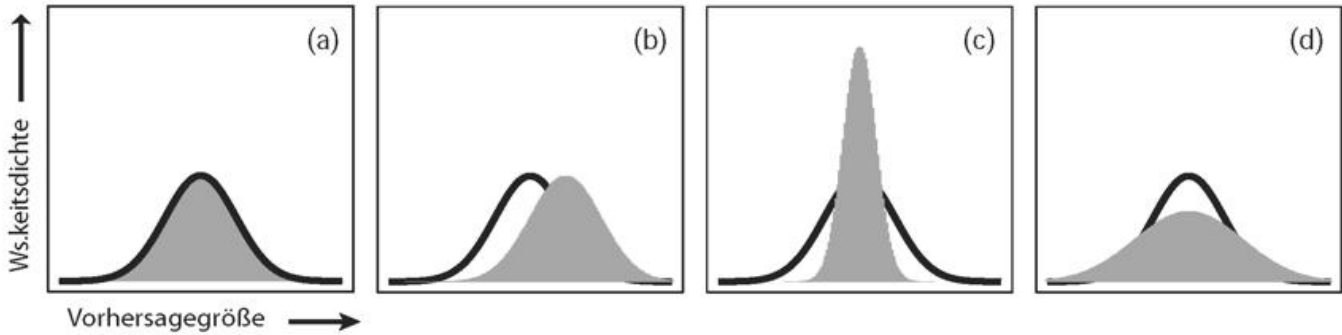


Abbildung 4-1: Dispersionseigenschaften von Ensemblevorhersagen. Für ein gegebenes Vorhersageproblem sei die Wahrscheinlichkeitsverteilung möglicher Beobachtungswerte bekannt (dicke schwarze Linie). Ein Ensemblevorhersagesystem werde verwendet, um diese Wahrscheinlichkeitsverteilung abzuschätzen (grau schattierte Fläche). Ist die Vorhersage *reliable* (a), so sollten sich die Ensemblemitglieder statistisch genau wie der Beobachtungswert verhalten, das heißt sie sollten derselben Wahrscheinlichkeitsverteilung folgen wie die Realität. Für das Toymodell aus Kapitel 2 würde dies einer Parameterwahl von $\beta=0$ und $\gamma=1$ entsprechen. Falls die Vorhersagen einen systematischen Bias aufweisen (b), so ist die Ensembleverteilung bezüglich der tatsächlichen Verteilung verschoben (Toymodell: $\beta>0$ und $\gamma=1$). Unterdispersive Ensemblevorhersagen (c) schaffen es nicht, den gesamten Unsicherheitsbereich abzudecken und sind zu schmal ($\beta=0$ und $\gamma<1$), wohingegen überdispersive Vorhersagen (d) den tatsächlichen Unsicherheitsbereich überschätzen und zu breit sind ($\beta=0$ und $\gamma>1$).

hier α auf 1 gesetzt. Der Parameter β ist der systematische Bias der Vorhersagen, und γ kontrolliert die Dispersionseigenschaften des Ensembles und damit den Ensemble-Spread. Ist $\beta=0$ und $\gamma=1$, so werden \hat{X}_i und x aus derselben zu Grunde liegenden Verteilung gezogen, d.h. der Ensemble-Spread deckt den tatsächlichen Unsicherheitsbereich adäquat ab. Solche Vorhersagen werden als *reliable* bezeichnet (illustriert in Abbildung 4-1a). Ist $\beta>0$ (beziehungsweise $\beta<0$), so haben die Vorhersagen einen positiven (beziehungsweise negativen) Bias (Abbildung 4-1b). Ist $\gamma<1$, so sind die Vorhersagen zu schmal und damit unterdispersiv (Abbildung 4-1c). Ist hingegen $\gamma>1$, so ist der Spread zu breit und die Vorhersagen damit überdispersiv (Abbildung 4-1d). Eine tiefer gehende Diskussion von Toymodellen dieser Art findet sich zum Beispiel bei WEIGEL und BOWLER (2009).

3 Ensembles als Stichproben einer Verteilung

Der Begriff der Vorhersagegüte, oder des *Skill*, ist sehr vielschichtig und beinhaltet eine ganze Reihe von Aspekten, die sogenannten *Skill-Attribute*. Im Folgenden stellen wir die wichtigsten dieser Attribute vor und diskutieren, wie sie bei Ensemblevorhersagen zu verstehen sind und gemessen werden können. Es sei angemerkt, dass die hier beschriebenen *Skill Scores* ausnahmslos für eindimensionale Punktvorhersagen konzipiert sind. Viele Entscheidungsprozesse basieren allerdings auf der simultanen Auswertung mehrerer Variabler an mehreren Orten. Die Verifikation derartiger Vorhersagen erfordert spezielle multidimensionale Techniken, auf die hier nicht eingegangen wird, die aber zum Beispiel bei WEIGEL (2011) beschrieben werden.

3.1 Skill-Attribute

Eines der wichtigsten *Skill-Attribute* von Ensemblevorhersagen ist die sogenannte *Reliability*. Diese bezeichnet, inwieweit die für ein Ereignis vorhergesagten Eintrittswahr-

scheinlichkeiten mit den beobachteten Eintrittshäufigkeiten konsistent sind. Bei Ensemblevorhersagen, welche *reliable* sind, sind die Ensemblemitglieder und die tatsächlichen Beobachtungswerte Stichproben derselben zu Grunde liegenden Wahrscheinlichkeitsverteilungen, das heißt sie sind *statistisch nicht unterscheidbar* (siehe auch Abbildung 4-1a). Kapitel 3.2 stellt einige der gebräuchlichsten Tests von Ensemble-*Reliability* vor.

Zwei weitere zentrale *Skill-Attribute* sind die miteinander verwandten Attribute *Resolution* und *Discrimination*. Beide beschreiben, grob gesagt, wie stark Vorhersagen und Beobachtungen im Sinne einer Korrelation zusammenhängen. *Resolution* betrachtet diese Fragestellung aus dem Blickwinkel der Vorhersagen. *Resolution* misst, inwieweit sich Beobachtungen voneinander unterscheiden, wenn die zugehörigen Vorhersagen unterschiedlich sind. *Resolution* testet also zum Beispiel, ob heute mehr Regen als gestern beobachtet wird, wenn die heutigen Vorhersagen mehr Regen als die gestrigen prognostiziert haben. *Discrimination* betrachtet dieselbe Fragestellung aus dem Blickwinkel der Beobachtungen und beschreibt, inwieweit sich Vorhersagen unterscheiden, wenn die zugehörigen Beobachtungswerte unterschiedlich sind. Wenn also heute zum Beispiel mehr Regen beobachtet wurde als gestern, dann testet *Discrimination*, ob für heute auch mehr Regen vorhergesagt war als für gestern. Ein Beispiel für einen *Skill Score* zur Messung der *Discrimination* wird in Kapitel 3.3 vorgestellt.

3.2 Reliability von Ensemblevorhersagen

3.2.1 Der Zusammenhang zwischen mittlerem Spread und Fehler

Ein sehr einfach implementierbarer und häufig verwendeter Test von Ensemble-*Reliability* vergleicht den mittleren quadratischen Fehler (Abkürzung MSE - von **m**ean **s**quared **e**rror) der Ensemblemittelwerte mit dem durchschnittlichen *Spread*. Konkret besagt dieser Test:

„Eine notwendige Bedingung für Reliability ist, dass der MSE der Ensemblemittelwerte und die durchschnittliche Varianz der Ensemblemitglieder identisch sind.“

Dieses Kriterium gilt in dieser Form allerdings nur für große Ensembles. Wie weiter unten erläutert, muss man bei kleinen Ensembles der Größe m die Varianz noch mit einem Korrekturfaktor von $(m+1)/m$ multiplizieren.

Wie lässt sich dieses Kriterium verstehen? Gegeben sei ein Menge von n Ensemblevorhersagen X_1, \dots, X_n und zugehörigen Beobachtungen x_1, \dots, x_n . Jedes Ensemble habe m Mitglieder, wobei $\hat{x}_{t,i}$ das i -te Mitglied der t -ten

Ensemblevorhersage bezeichne. $\bar{\hat{x}}_t = \frac{1}{m} \sum_{i=1}^m \hat{x}_{t,i}$ sei der Mittelwert der Mitglieder der t -ten Ensemblevorhersage, und μ sei der Erwartungswert der zu Grunde liegenden und normalerweise unbekanntes Wahrscheinlichkeitsverteilung, aus welcher \hat{x}_t „gezogen“ wurde. Mit S_t^2 bezeichnen wir die Stichprobenvarianz der Ensemblemitglieder von \hat{x}_t , das heißt $S_t^2 = \frac{1}{m-1} \sum_{i=1}^m (\hat{x}_{t,i} - \bar{\hat{x}}_t)^2$. Falls das Ensemblevorhersagesystem *reliable* ist, dann sollten die Beobachtungen und die m Ensemblemitglieder statistisch voneinander nicht unterscheidbar sein. Dies bedeutet, dass Beobachtung x_t und alle Ensemblemitglieder von \hat{x}_t Stichproben derselben zu Grunde liegenden Verteilung sein sollten. Daraus folgt für jedes Ensemblemitglied $i=1,2,\dots,m$:

$$E_t \left[(x_t - \mu_t)^2 \right] = E_t \left[(\hat{x}_{t,i} - \mu_t)^2 \right]$$

und damit

$$E_t \left[(x_t - \mu_t)^2 \right] = E_t \left[\frac{1}{m} \sum_{i=1}^m (\hat{x}_{t,i} - \mu_t)^2 \right]$$

wobei E_t der Erwartungswert über t ist. Ausgehend von dieser Gleichung kann man nach ein paar Umformungen das oben erwähnte *Reliability*-Kriterium für Ensemblevorhersagen herleiten (siehe WEIGEL (2011) für Details):

$$E_t \left[\left(x_t - \bar{\hat{x}}_t \right)^2 \right] = \frac{m+1}{m} E_t \left[S_t^2 \right] \quad (1)$$

Die linke Seite von Gleichung (1) ist der MSE der Ensemblemittelwerte, und die rechte Seite ist die durchschnittliche Ensemblevarianz, multipliziert mit einem Korrekturfaktor $(m+1)/m$, welcher bei großen Ensembles ignoriert werden kann. Abbildung 4-2 zeigt ein Beispiel für die Anwendung dieses Kriteriums. Hierfür wurden mittelfristige EPS-Vorhersagen des 500 hPa Geopotenzials für die nördlichen Außertropen ausgewertet, wie in Kapitel 2 beschrieben. Die Abbildung zeigt den MSE der Ensemblemittelwerte und die durchschnittliche Ensemblevarianz als Funktion des Vorhersagezeitraums. Die Varianzen wurden mit einem Korrekturfaktor von 1,02 multipliziert (das Ensemble hat 51 Mitglieder). Das Beispiel zeigt, dass bis zu Vorhersagezeiten von etwa 6 Tagen

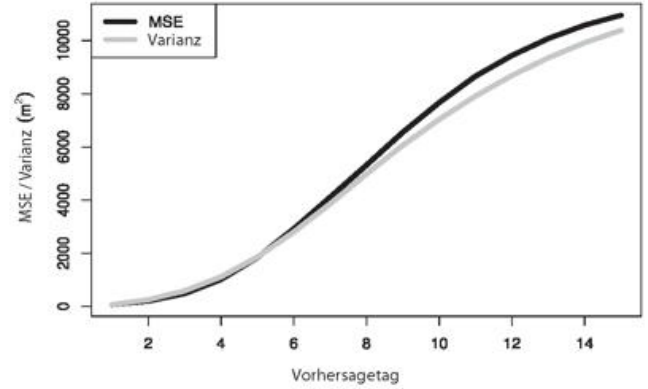


Abbildung 4-2: Mittlerer quadratischer Fehler des Ensemblemittelwertes (schwarz) und durchschnittliche Ensemblevarianz (grau). Betrachtet wurden EPS-Vorhersagen des 500 hPa-Geopotenzials wie in Kapitel 2 beschrieben. Die dargestellten Werte basieren auf der Evaluierung von Vorhersagen für die nördlichen Außertropen (20-90°N) und den Zeitraum Dezember 2009 bis Februar 2010. Die Varianzen wurden mit einem Faktor 1,02 skaliert, um den Effekt der limitierten Ensemblegröße (51 Mitglieder) zu berücksichtigen. Dem EZMW wird für dieses Datenbeispiel gedankt.

die Identität von Gleichung (1) gut erfüllt ist. Für längere Vorhersagehorizonte wächst hingegen der MSE schneller als die mittlere Ensemblevarianz, das heißt die Vorhersagen werden zunehmend *unreliable*.

3.2.2 Ranghistogramme

Das Ranghistogramm wurde unabhängig voneinander von ANDERSON (1996), HAMILL und COLUCCI (1997), und TALAGRAND et al. (1997) eingeführt und stellt einen der am häufigsten verwendeten Tests von Ensemble-*Reliability* dar. Ranghistogramme analysieren, wie sich die Beobachtungswerte bezüglich der zugehörigen und nach ihrer Größe sortierten Ensemblemitglieder einordnen lassen. Daraus kann man dann Einblicke in die Dispersionseigenschaften der Vorhersagen gewinnen, wie im Folgenden gezeigt wird.

Betrachten wir eine Ensemblevorhersage mit m Mitgliedern $\hat{x}_t = (\hat{x}_{t,1}, \dots, \hat{x}_{t,m})$, und eine zugehörige Beobachtung x_t . Nehmen wir an, dass x_t größer als M der m Ensemblemitglieder ist ($M \leq m$) und kleiner als die verbleibenden $(m-M)$ Mitglieder. Der Rang der Beobachtung x_t (bezüglich der sortierten Ensemblemitglieder) ist dann $r_t = M+1$. Falls x_t kleiner als alle Ensemblemitglieder ist, dann hat die Beobachtung Rang $r_t = 1$, und falls x_t größer als alle Ensemblemitglieder ist, dann hat sie Rang $r_t = m+1$. Aus einem Verifikationsdatensatz von n Vorhersage-Beobachtungspaaren lässt sich das Ranghistogramm nun folgendermaßen berechnen: Zunächst bestimmt man die Ränge r_t aller n Beobachtungen, dann zählt man, wie oft jeder der $m+1$ möglichen Rangwerte angenommen wird, und schließlich stellt man diese Werte in Form eines Histogramms dar.

Ist das Ensemblevorhersagesystem *reliable*, so sind die Ensemblemitglieder und Beobachtungen statistisch nicht voneinander unterscheidbar, das heißt für jede Beobachtung

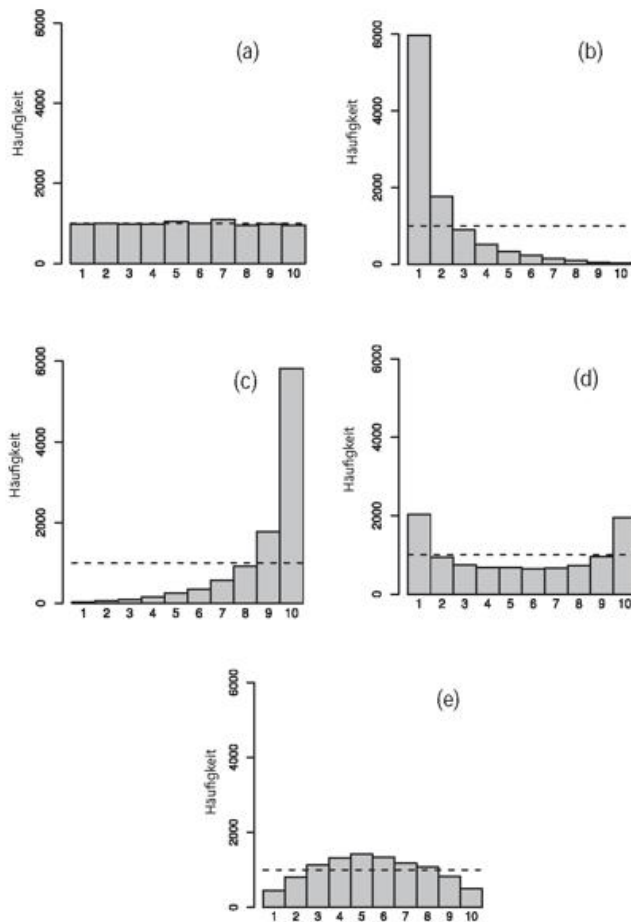


Abbildung 4-3: Ranghistogramme für Ensemblevorhersagen mit unterschiedlichen Dispersionseigenschaften. Jedes Histogramm basiert auf 10.000 mit dem Toymodell erzeugten Vorhersage-Beobachtungspaaren (Details in Kapitel 2) mit Ensemblegröße 9. Abbildung (a) zeigt das Histogramm für Vorhersagen, welche *reliable* sind, so wie in Abb. 4-1a illustriert (Toymodellparameter: $\alpha=1, \beta=0, \gamma=1$); (b) basiert auf Vorhersagen mit einem systematischen positiven Bias ($\alpha=1, \beta=1,5, \gamma=1$) wie in Abb 4-1b, und (c) basiert auf Vorhersagen mit einem systematischen negativen Bias ($\alpha=1, \beta=-1,5, \gamma=1$). Abbildung (d) basiert auf unterdispersiven Vorhersagen ($\alpha=1, \beta=0, \gamma=0,6$) wie in Abb. 4-1c illustriert, und (e) basiert auf überdispersiven Vorhersagen ($\alpha=1, \beta=0, \gamma=1,5$) wie in Abb. 4-1d. Die gestrichelten waagrechten Linien geben an, wie oft jede Rangkategorie im Mittel besetzt wäre, wenn die Vorhersagen *reliable* wären.

wäre jeder der möglichen Rangwerte $r_i=1,2,\dots,m+1$ gleich wahrscheinlich. Folglich wäre das zugehörige Ranghistogramm *flach* - abgesehen von Fluktuationen, die von der Stichprobenunsicherheit herrühren. Die erwartete Zahl an Beobachtungen je Rangkategorie wäre in diesem Fall $\frac{n}{m+1}$. Dies ist in Abbildung 4-3a illustriert. Das dort gezeigte Ranghistogramm wurde auf der Basis von künstlich erzeugten Toymodellvorhersagen (siehe Kapitel 2) berechnet, wobei die Vorhersagen so konstruiert wurden, dass sie *reliable* sind. Im Gegensatz zu diesem Beispiel implizieren Ranghistogramme, welche (signifikant) nicht flach sind, dass die Vorhersagen *unreliable* sind. Die Form des Histogramms kann dann Hinweise auf die Art der Vorhersagemängel geben. Wenn die Vorhersagen beispielsweise einen systematischen positiven Bias aufweisen, so ist die Wahrscheinlichkeit erhöht, dass eine Beobachtung kleiner ist als die Mehrheit der Ensemblemitglieder. Dies führt zu einer überproportio-

nal häufigen Belegung der niedrigeren Ränge und damit zu einem Histogramm mit abfallender Neigung (Abbildung 4-3b). Ist der Bias nur groß genug, so würden die Beobachtungen ab einem gewissen Punkt sogar von allen Ensemblemitgliedern überschritten werden, so dass im zugehörigen Ranghistogramm alle Kategorien mit Ausnahme der niedrigsten leer wären. Analog führt ein Vorhersagesystem mit systematisch negativem Bias zu Histogrammen mit ansteigender Neigung (Abbildung 4-3c). Falls sowohl die obersten als auch die untersten Ränge „überbevölkert“ und die mittleren Ränge „unterbevölkert“ sind (Abbildung 4-3d), so deutet dies auf unterdispersive Ensembles hin: Da der Ensemble-Spread im Mittel zu schmal ist, kommt es überproportional häufig vor, dass eine Beobachtung nicht vom Ensemble erfasst wird, das heißt dass sie Rang 1 oder $m+1$ hat. Umgekehrt haben überdispersive Ensembles die Tendenz, die Beobachtung zu oft zu enthalten, was dann in konvexen Histogrammen mit einer Häufung der mittleren Ränge resultiert (Abbildung 4-3e).

Im Folgenden noch ein wichtiger Hinweis zur Interpretation von Ranghistogrammen: Wenn Ranghistogramme nicht flach sind, so heißt das nicht automatisch, dass die Vorhersagen *unreliable* sind. Wenn der Datensatz klein und folglich die Stichprobenunsicherheiten groß sind, kann es auch zufällige Abweichungen von der Gleichverteilung geben. Ein einfacher Ansatz zur Prüfung dieser Frage ist der sogenannte χ^2 -Anpassungstest mit der Nullhypothese eines flachen Ranghistogramms: Es sei n die Anzahl an Vorhersage-Beobachtungspaaren, m die Ensemblegröße, n_i die Anzahl an Belegungen in der i -ten Rangkategorie des Histogramms, und $e = \frac{n}{m+1}$ die Anzahl an Belegungen, welche man für jede Rangkategorie im Falle einer Gleichverteilung erwarten würde. Die Teststatistik T sei folgendermaßen definiert:

$$T = \sum_{i=1}^{m+1} \frac{(n_i - e)^2}{e}$$

Man kann zeigen, dass diese Teststatistik näherungsweise einer χ^2 -Verteilung mit m Freiheitsgraden entspricht, falls die Nullhypothese wahr ist. T kann deshalb relativ leicht zur Untersuchung der Signifikanz einer Histogrammstruktur eingesetzt werden. Als Beispiel betrachte man die Ranghistogramme, welche mit den in Kapitel 2 beschriebenen saisonalen Vorhersagen (Ensemblegröße 9) berechnet wurden. Die Histogramme sind einmal mit den „rohen“ (Abbildung 4-4a), einmal mit „bias-korrigierten“ (Abbildung 4-4b) und einmal mit rekalierten (Abbildung 4-4c) Vorhersagen konstruiert worden. Für das Ranghistogramm in Abbildung 4-4a erhält man $T=117,5$. Die Form deutet sowohl auf Unterdispersion als auch auf einen systematischen negativen Bias hin - Unterdispersion weil die äußeren Ränge überbelegt sind; und ein negativer systematischer Bias, weil der oberste Rang häufiger belegt ist als der unterste Rang. Das Histogramm in Abbildung 4-4b hat $T=51,8$. Hier ist zwar der systematische Bias aus den Vorhersagen entfernt, aber die U-förmige Verteilung deutet immer noch auf

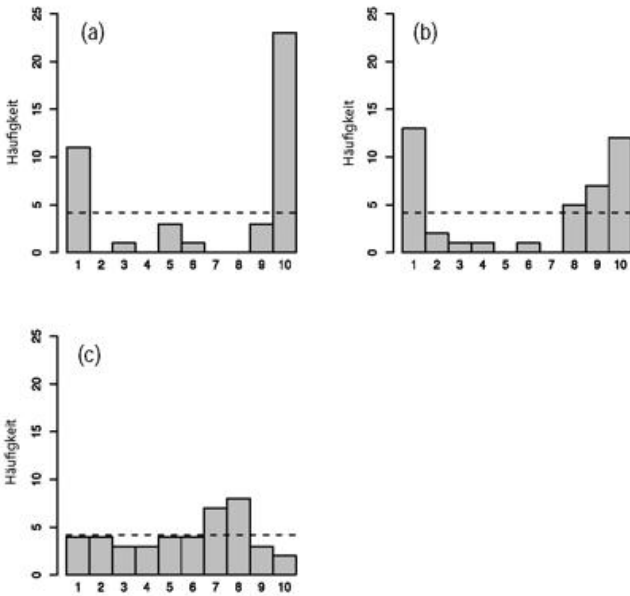


Abbildung 4.4: Ranghistogramme für saisonale Vorhersagen (EZMW-System 3) der 2m-Temperatur in der Niño3.4-Region (Details in Kapitel 2). Das Ranghistogramm in (a) basiert auf „rohen“ Ensemblevorhersagen, (b) basiert auf bias-korrigierten Vorhersagen, und (c) auf rekalierten Vorhersagen. Die gestrichelten waagrechten Linien geben an, wie oft jede Rangkategorie im Mittel besetzt wäre, wenn die Vorhersagen *reliable* wären.

Unterdispersion hin. Das Histogramm der rekalierten Vorhersagen in Abbildung 4-4c schließlich hat $T=7,5$ und sieht vergleichsweise flach aus. Da das 99 %-Quantil der Nullverteilung (einer χ^2 -Verteilung mit 9 Freiheitsgraden) 21,67 ist, kann die Nullhypothese eines flachen Ranghistogramms folglich sowohl für die rohen als auch für die bias-korrigierten Vorhersagen auf dem 1 %-Niveau verworfen werden, nicht aber für die rekalierten Vorhersagen. Eine weiter gehende Diskussion derartiger Signifikanztests findet sich zum Beispiel im Buch von WILKS (2006) und bei WEIGEL (2011).

3.3 Discrimination

Wie bereits in Kapitel 3.1 erwähnt ist *Discrimination* neben *Reliability* ein weiteres wichtiges Skill-Attribut. *Discrimination* beschreibt, inwieweit sich Vorhersagen voneinander unterscheiden, wenn die zugehörigen Beobachtungen unterschiedlich sind. Obwohl *Discrimination* eines der wichtigsten Skill-Attribute ist, gibt es nur wenige Scores zur Messung der *Discrimination* von Ensemblevorhersagen, welche als Stichproben interpretiert werden (übrigens ebenso wenig wie zur Messung des verwandten Skill-Attributs der *Resolution*). Einer der wenigen derartigen Scores ist der *Verallgemeinerte Discrimination Score D*.

Dieser *D-Score* wurde von MASON und WEIGEL (2009) im Sinne eines generischen Verifikationskonzeptes eingeführt, welches für alle denkbaren Vorhersagetypen angewend-

bar ist, seien es binäre Vorhersagen, die sich mit ja oder nein beantworten lassen (zum Beispiel Regen > 5mm), seien es kategorische Vorhersagen für mehr als zwei mögliche Ereigniskategorien (zum Beispiel Kategorie 1: Regen > 5 mm; Kategorie 2: Regen ≤ 5 mm und > 1 mm; Kategorie 3: Regen ≤ 1mm), seien es Vorhersagen auf einer kontinuierlichen Werteskala (zum Beispiel Niederschlag in mm), oder seien es probabilistische Vorhersagen. Unabhängig vom Vorhersagetyp basiert die Berechnung von D immer auf demselben Prinzip und versucht, die folgende Frage zu beantworten: Falls sich zwei Beobachtungen voneinander unterscheiden, wie wahrscheinlich ist es dann, dass diese Beobachtungen mit Hilfe der zugehörigen Vorhersagen korrekt ihrer Größe nach sortiert werden können? Falls die Vorhersagen keinerlei nützliche Information enthalten und damit in statistischem Sinne unabhängig von den Beobachtungen sind, so würde eine vorhersagebasierte Sortierung zweier Beobachtungen im Mittel nur in 50% der Fälle erfolgreich sein, das heißt man erhielte $D=0,5$. Dies entspricht derselben Trefferquote, die man auch durch willkürliches Raten erzielen würde. Je besser die Vorhersagen die zugehörigen Beobachtungen unterscheiden können, desto mehr nähert sich der Score dem Wert 1 an.

Die Prozedur zur Berechnung von D spezifisch für Ensemblevorhersagen wurde von WEIGEL und MASON (2011) formuliert und ist in Abbildung 4-5 illustriert: Zunächst bildet man aus den Verifikationsdaten alle möglichen Kombinationen von je zwei Vorhersage-Beobachtungspaaren. Für jedes dieser Zweier-Sets testet man dann, ob die beiden Beobachtungen und anhand der beiden zugehörigen Vorhersagen \hat{X}_s und \hat{X}_t korrekt sortiert werden können. Die relative Häufigkeit, mit der diese Tests positiv verlaufen, ergibt den Score D. Diese Prozedur erfordert allerdings eine Definition, wie man Ensembles überhaupt sortieren kann. Wie entscheidet man, ob $\hat{X}_s > \hat{X}_t$ oder $\hat{X}_s < \hat{X}_t$? WEIGEL und MASON (2011) schlagen hierfür die folgende Definition vor: \hat{X}_s

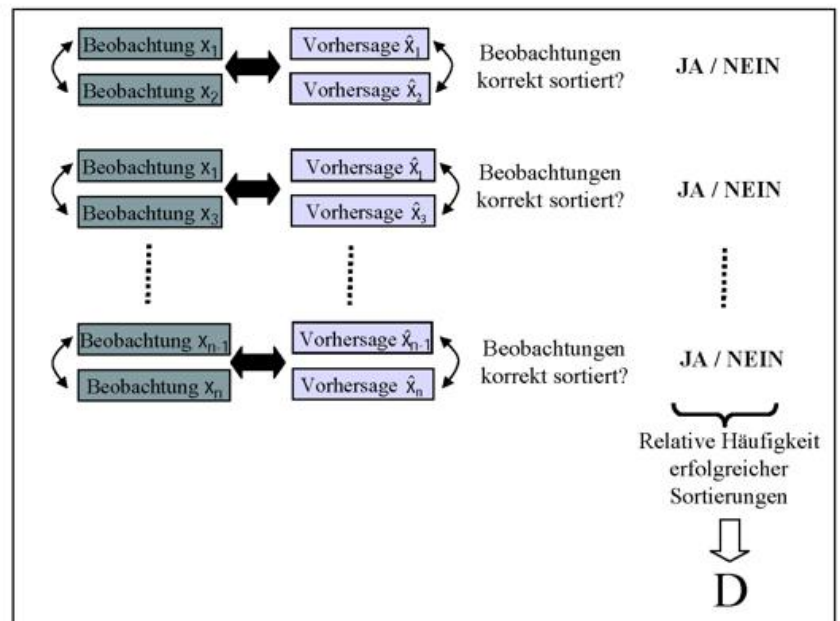


Abbildung 4-5: Illustration des Konzepts des *Verallgemeinerten Discrimination Scores D*.

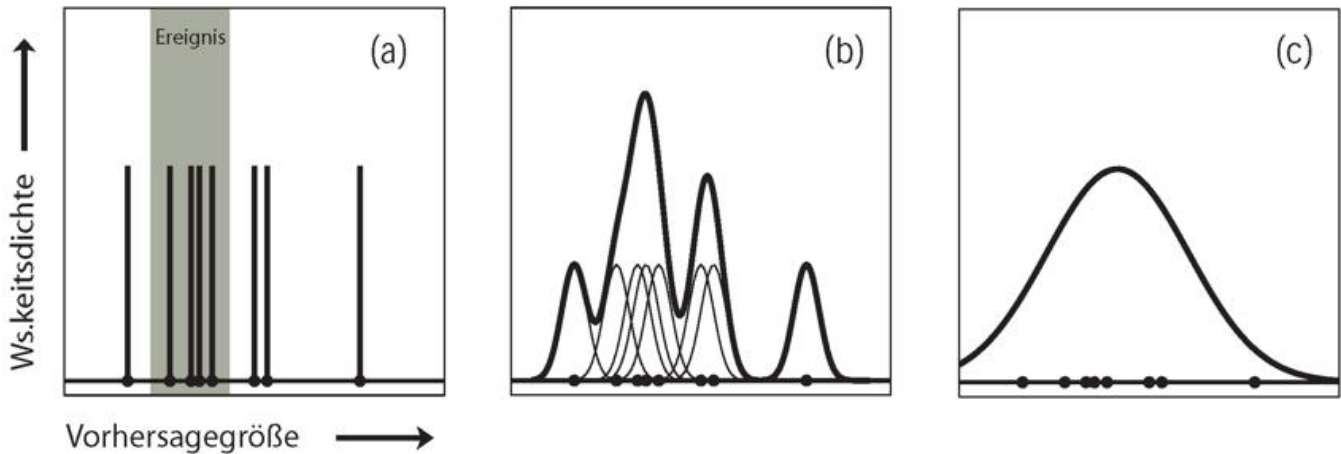


Abbildung 4.6: Ansätze zur probabilistischen Ensembleinterpretation. (a) illustriert die Häufigkeitsinterpretation. In dieser ergibt sich die Eintrittswahrscheinlichkeit für eine spezifische Ereigniskategorie (grau schattiert) aus dem Bruchteil an Ensemblemitgliedern (schwarze Punkte), welche diese Kategorie vorhersagen. Im dargestellten Beispiel hätte man 50%. (b) illustriert die Methode des Kernel-Dressing, bei der die einzelnen Ensemblemitglieder mit einer geeigneten Kernel-Funktion „eingekleidet“ werden (dünne Linien). Die Wahrscheinlichkeitsvorhersage ergibt sich dann aus der Summe dieser Kernel-Funktionen (dicke schwarze Linie). (c) schließlich illustriert den Fit einer geeigneten parametrischen Verteilungsfunktion an die Ensemblemitglieder.

ist „größer“ als \hat{X}_t , falls die Wahrscheinlichkeit, dass ein beliebiges Ensemblemitglied von \hat{X}_s größer als ein beliebiges Ensemblemitglied von \hat{X}_t ist, 50 % übersteigt. Wendet man dieses Kriterium auf alle Kombinationen von je zwei Vorhersage-Beobachtungspaaren an, so lässt sich eine relativ einfach implementierbare Formulierung von D herleiten, die im wesentlichen eine Verallgemeinerung der Kendallschen Rangkorrelation (SHESKIN 2007) darstellt. Für die technischen Details sei auf den Artikel von WEIGEL und MASON (2011) verwiesen.

Während die Definition von D auf ersten Blick recht unverständlich wirken mag, so hat sie den entscheidenden Vorteil einer intuitiven Interpretation. Für das bereits diskutierte Beispiel von saisonalen Temperaturvorhersagen für die Niño3.4-Region erhält man $D=0.92$. Das heißt, in 92 % der Fälle wären zwei beliebige beobachtete saisonale Mitteltemperaturen korrekt mit Hilfe der zugehörigen Ensemblevorhersagen unterschieden worden.

4 Ensembles als Wahrscheinlichkeitsprognosen

Die in Kapitel 3 beschriebenen Tests haben alle gemeinsam, dass sie Ensembles formal als Stichproben einer zu Grunde liegenden Wahrscheinlichkeitsverteilung interpretieren. Sobald man jedoch – auf welche Weise auch immer – wirkliche Wahrscheinlichkeitsvorhersagen aus den Ensembles abgeleitet hat, kommt eine andere Kategorie von *Skill Scores* zur Anwendung, die sogenannten probabilistischen *Skill Scores*. Bei diesen *Scores* muss man generell beachten, dass sie nicht nur die Modellqualität per se beschreiben, sondern gleichzeitig auch davon beeinflusst werden, auf welche Weise man die Wahrscheinlichkeitsvorhersagen abgeleitet hat, und wie realistisch die zu Grunde liegenden Annahmen sind. In diesem Kapitel geben wir zunächst einen kurzen Überblick über gängige Ansätze der probabilistischen Interpretation von Ensemblevorhersagen (Kapitel 4.1).

Danach, in Kapitel 4.2, stellen wir ein paar der geläufigsten probabilistischen Skillmetriken vor – wohl wissend, dass es sich hierbei nur um eine kleine und unvollständige Auswahl aus der großen Menge existierender *Scores* handeln kann. Kapitel 4.3 kehrt schließlich zum Problem der probabilistischen Ensembleinterpretation zurück und diskutiert den Einfluss der Ensemblegröße auf probabilistische *Skill Scores*.

4.1 Probabilistische Interpretation von Ensembles

Ensemblevorhersagen werden sehr oft in Form von Wahrscheinlichkeiten für vorher definierte Ereigniskategorien angegeben. Für eine erste einfache Abschätzung dieser Wahrscheinlichkeiten wird häufig der Bruchteil an Ensemblemitgliedern bestimmt, welche das zu quantifizierende Ereignis vorhersagen („Häufigkeitsinterpretation“, siehe Abbildung 4-6a). Wenn beispielsweise 40 von 50 Ensemblemitgliedern vorhersagen, dass am nächsten Tag mehr als 5 mm Regen fallen, so würde man in der Häufigkeitsinterpretation diesem Ereignis eine Wahrscheinlichkeit von 80% zuordnen. Es ist offensichtlich, dass derartig abgeschätzte Wahrscheinlichkeiten selbst bei einem perfekten Modell nur dann der Wirklichkeit nahe kommen können, wenn sehr viele Ensemblemitglieder vorhanden sind (siehe auch Kapitel 4.3).

Bessere Wahrscheinlichkeitsabschätzungen kann man erhalten, wenn man die Ensemblemitglieder mit sogenannten Kernel-Funktionen „einkleidet“ (Fachbegriff: *Kernel dressing*), das heißt wenn man anstelle der diskreten Einzelvorhersagen geeignete stetige Wahrscheinlichkeitsverteilungen setzt, welche den Unsicherheitsbereich eines jeden Ensemblemitglieds charakterisieren. Die Vorhersageverteilung ist dann durch die Summe der „eingekleideten“ Ensemblemitglieder gegeben (Abbildung 4-6b). Mehr Informationen zu Kernel-Dressing-Methoden finden sich zum Beispiel bei ROULSTON und SMITH (2003) und bei WANG und BISHOP (2005).

Ein dritter häufig verwendeter Ansatz ist der Fit einer geeigneten parametrischen Verteilung an das Vorhersageensemble (Abbildung 4-6c). Falls sich die Ensemblemitglieder tatsächlich als unabhängige Stichproben einer zu Grunde liegenden Vorhersageverteilung interpretieren lassen, und falls man realistische Annahmen über die Art der Verteilung machen kann, so kann ein solcher Ansatz deutlich bessere Ergebnisse liefern als die oben erwähnte Häufigkeitsinterpretation. Mehr Informationen hierzu finden sich zum Beispiel bei WILSON et al. (1999) und bei WILKS (2002).

Unabhängig davon, wie man nun Wahrscheinlichkeiten aus den Ensembles abgeschätzt hat, muss man Folgendes beachten: Selbst perfekt abgeschätzte Wahrscheinlichkeitsverteilungen („perfekt“ im Sinne einer korrekten Wiedergabe der dem Ensemble zu Grunde liegenden Verteilungen) können auf Grund von Modellfehlern *unreliable* sein. Dieses Problem kann man zumindest teilweise beheben und die Vorhersagegüte entsprechend anheben, indem man die Fehlerstatistik vergangener Vorhersagen in die Ensembleinterpretation mit einbezieht. Solche Ansätze werden als statistisches Postprocessing, Rekalibrierung, oder **Model Output Statistics (MOS)** bezeichnet. Mehr Details hierzu finden sich zum Beispiel bei STEPHENSON et al. (2005), BRÖCKER und SMITH (2008) und WEIGEL et al. (2009).

4.2 Probabilistische *Skill Scores* für Ensembles

Sobald man Wahrscheinlichkeitsvorhersagen aus den Ensembles abgeleitet hat - egal mit welcher Methode - können diese mit probabilistischen *Skill Scores* verifiziert werden. Hierfür steht ein ganzer „Zoo“ an *Scores* zur Verfügung. Generell hängt die Wahl des *Scores* zum einen von der Art der Vorhersage ab, also davon ob zum Beispiel Wahrscheinlichkeiten für zwei oder mehrere Ereigniskategorien oder für eine kontinuierliche Werteskala gegeben sind, und zum anderen natürlich vom Skill-Attribut, für das man sich interessiert, also zum Beispiel *Reliability* oder *Discrimination*. Im Falle von Wahrscheinlichkeitsvorhersagen für binäre Ereignisse werden *Reliability* und *Resolution* oftmals mit sogenannten *Reliability-Diagrammen* (WILKS 2006) oder den *Reliability- und Resolutionkomponenten* des *Brier-Scores* bestimmt (WILKS 2006), während *Discrimination* durch die „Fläche unter der ROC-Kurve“ gemessen werden kann (MASON 1982). Bei letzterer handelt es sich übrigens um einen Spezialfall des Verallgemeinerten *Discrimination Scores D* (MASON und WEIGEL, 2009). Häufig werden auch „Gesamtmaße“ verwendet, welche gleichzeitig *Reliability* und *Resolution* messen. In diese Kategorie fallen zum Beispiel der *Brier Score* („BS“, BRIER 1950) für binäre Ereignisse, der *Ranked Probability Score* („RPS“, EPSTEIN 1969) und die *Ignoranz* („IGN“, ROULSTON und SMITH 2002) im Falle mehrerer Ereigniskategorien, und der *Continuous Ranked Probability Score* („CRPS“, HERSBACH 2000) für Vorhersagen auf einer kontinuierlichen Werteskala. All diese Gesamtmaße haben die vorteilhafte Eigenschaft, dass sie *strictly proper* sind, das heißt, die Vorhersagegüte kann nicht durch künstliche Manipula-

tion der Vorhersagen erhöht werden (BRÖCKER 2011). Im Folgenden sollen nur drei der geläufigsten *Scores* exemplarisch kurz vorgestellt werden, nämlich der BS, der RPS, und die Ignoranz. Für eine tiefer gehende Diskussion ihrer Eigenschaften, und für eine Beschreibung anderer probabilistischer *Skill Scores*, sei der Leser an die entsprechende Fachliteratur verwiesen.

Brier Score (BS)

Wie oben erwähnt wird der BS zur Verifikation von Wahrscheinlichkeitsvorhersagen eines spezifischen binären Ereignisses (zum Beispiel: Regen > 5mm) verwendet, welches entweder eintreten kann oder nicht. Es bezeichne \hat{p}_t die Wahrscheinlichkeit, mit der das Ereignis gemäß der t -ten Vorhersage eintreten sollte. Mit y_t werde die zugehörige Beobachtung bezeichnet. Es sei $y_t=1$, falls das Ereignis tatsächlich eintritt, und $y_t=0$, falls es nicht eintritt. Der BS ist dann gegeben durch:

$$BS = \frac{1}{n} \sum_{t=1}^n (\hat{p}_t - y_t)^2,$$

wobei n die Zahl aller Vorhersage-Beobachtungspaare ist, die in die Verifikation mit eingehen. Oftmals ist man aber nicht so sehr am direkten (und etwas abstrakten) Zahlenwert des BS interessiert, sondern möchte wissen, um wie viel die Vorhersagen besser sind als eine bereits verfügbare und in der Regel einfachere alternative Vorhersagestrategie, wie beispielsweise der Klimatologie. Dies kann man quantifizieren, indem man den BS der zu evaluierenden Vorhersagen mit dem BS der alternativen Referenzvorhersagen ins Verhältnis setzt. Man spricht dann vom sogenannten *Brier Skill Score (BSS)*. Nimmt man die Klimatologie als Referenzstrategie, und bezeichnet man mit c die klimatologische Wahrscheinlichkeit für das Ereignis, so ist der BSS gegeben durch:

$$BSS = 1 - \frac{BS}{BS_{cl}} \quad \text{wobei } BS_{cl} = \frac{1}{n} \sum_{t=1}^n (c - y_t)^2.$$

Positive BSS-Werte bedeuten dann, dass die Vorhersagen gegenüber der klimatologischen Referenz einen Informationsgewinn bringen. Ein perfektes Vorhersagesystem ergäbe BSS=1. Wie bereits weiter oben angedeutet lassen sich der BS und der BSS so zerlegen, dass man die Beiträge von *Reliability* und *Resolution* zum Gesamt-Score separat quantifizieren kann. Diese sogenannte *Brier-Score-Zerlegung* wird häufig angewandt, und ist zum Beispiel im Buch von WILKS (2006) ausführlich beschrieben.

Ranked Probability Score (RPS)

Der RPS ist eine Verallgemeinerung des BS für Wahrscheinlichkeitsvorhersagen für mehr als zwei Ereigniskategorien. Angenommen, es werden K Ereigniskategorien betrachtet, und c_k ist die klimatologische Wahrscheinlich-

keit dafür, dass die Beobachtung in Kategorie k fällt. Es seien n Vorhersage-Beobachtungspaare gegeben, und es sei $\hat{p}_{t,k}$ die Wahrscheinlichkeit, welche von der t -ten Vorhersage für das Eintreten der k -ten Kategorie angegeben wird. Ferner sei $y_{t,k}=1$, falls die t -te Beobachtung in Kategorie k fällt, und $y_{t,k}=0$ falls sie in eine andere Kategorie fällt. Mit $\hat{P}_{t,k} = \sum_{l=1}^k \hat{p}_{t,l}$ und $Y_{t,k} = \sum_{l=1}^k y_{t,l}$ lässt sich der RPS dann folgendermaßen formulieren:

$$RPS = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^K \left(\hat{P}_{t,k} - Y_{t,k} \right)^2.$$

Wie der BS, so wird auch der RPS oft als relatives Gütemaß berechnet, um den Informationsgewinn bezüglich einer geeigneten Referenzstrategie zu quantifizieren. Man spricht dann vom *Ranked Probability Skill Score* (RPSS). Mit der

Klimatologie als Referenz und mit $C_k = \sum_{l=1}^k c_l$ ist der RPSS gegeben durch:

$$RPSS = 1 - \frac{RPS}{RPS_{cl}} \text{ wobei } RPS_{cl} = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^K \left(C_k - Y_{t,k} \right)^2.$$

Ignoranz (IGN)

Die Ignoranz, oftmals auch als logarithmischer Score bezeichnet, lässt sich wie der RPS auf Wahrscheinlichkeitsvorhersagen für multiple Ereigniskategorien anwenden. Wie oben sei K die Anzahl an Kategorien, und $\hat{p}_{t,k}$ sei die Wahrscheinlichkeit, mit der gemäß der t -ten Vorhersage die Beobachtung in die k -te Kategorie fallen sollte. Ferner bezeichne $k^*(t)$ diejenige Kategorie, in welche die t -te Beobachtung schlussendlich gefallen ist. Die Ignoranz ist dann gegeben durch

$$IGN = -\frac{1}{n} \sum_{t=1}^n \log_2 \left(\hat{p}_{t,k^*(t)} \right).$$

Der Ignoranz-Score besitzt ein paar interessante Eigenschaften, welche ihn konzeptionell vom RPS deutlich unterscheiden: Zunächst einmal hat die Ignoranz eine klare informationstheoretische Interpretation. Sie misst nämlich – in Bits – das mittlere Informationsdefizit von jemandem, der die Vorhersage zwar kennt, aber noch nicht weiß, welche Kategorie schlussendlich beobachtet werden wird (ROULSTON und SMITH, 2002). Zweitens ist die Ignoranz lokal, d.h. sie wertet nur die Wahrscheinlichkeit aus, welche der beobachteten Kategorie zugeordnet worden ist ($\hat{p}_{t,k^*(t)}$), und hängt nicht von den Wahrscheinlichkeiten der anderen Kategorien ab. Ob dieses Attribut der Lokalität eine wünschenswerte Eigenschaft ist oder nicht ist Gegenstand anhaltender Diskussionen (zum Beispiel BENEDETTI, 2010) und hängt unter anderem vom Anwendungskontext ab. Drittens nimmt die Ignoranz unendlich große Werte an, wenn für die tatsächlich beobachtete Kategorie eine Eintrittswahrscheinlichkeit von 0 % vorhersagt war. Dies wird oftmals als eine unerwünschte Eigenschaft

gesehen, vor allem im Zusammenhang mit Ensemblevorhersagen, bei denen Wahrscheinlichkeitsvorhersagen von 0 % durchaus einmal auftreten können, wenn die Ensembles klein sind und die Wahrscheinlichkeiten durch eine einfache Häufigkeitsinterpretation abgeschätzt werden (Kapitel 4.1). Andererseits wird auch argumentiert (zum Beispiel von WEIJS et al. 2010), dass es durchaus sinnvoll sein kann, eine Vorhersage als unendlich schlecht zu klassifizieren, wenn sie mit absoluter Sicherheit „Entwarnung“ für ein Ereignis gegeben hat, welches dann letztendlich doch eintritt. Denn der Geldwert, den ein Anwender auf eine als 100 % sicher kommunizierte Vorhersage setzen würde, kann ebenfalls beliebig groß sein. Um unendlich große Ignoranzwerte zu vermeiden, sollte man deshalb in erster Linie die angewandte Methode der Ensembleinterpretation überdenken. So treten bei Kernelansätzen oder angepassten parametrischen Verteilungsfunktionen in der Regel keine 0 %-Wahrscheinlichkeiten auf. Will man nichtsdestotrotz an einer Häufigkeitsinterpretation der Ensembles festhalten, so kann man, sozusagen als „Notlösung“, einfach ein zusätzliches virtuelles Ensemblemitglied auf alle Kategorien verteilen und damit 0 %-Wahrscheinlichkeiten vermeiden (ROULSTON und SMITH 2002).

4.3 Ensemblegröße und Skill

Im Abschnitt 4.1 wurden zwei Faktoren diskutiert, welche den Skill probabilistisch interpretierter Ensemblevorhersagen beeinflussen: Zum einen die Qualität des Vorhersagemodells, und zum anderen die Methode, mit der die Ensembles in Wahrscheinlichkeiten umgewandelt wurden. Daneben gibt es aber noch einen dritten Faktor, nämlich die Ensemblegröße. So wird die Genauigkeit, mit der sich eine Wahrscheinlichkeitsverteilung aus einer Stichprobe ableiten lässt, umso kleiner, je weniger Daten die Stichprobe umfasst. Wenn man beispielsweise den Erwartungswert einer Vorhersageverteilung (sie habe Standardabweichung σ) anhand des Mittels eines Ensembles mit m Mitgliedern abschätzen will, so beträgt der Standardfehler dieser Schätzung $\frac{\sigma}{\sqrt{m}}$. Selbst in einem perfekten Ensemblevorhersagesystem führen kleine Ensemblegrößen zu einer Schätzungsgenauigkeit, welche die *Reliability* der Vorhersagen implizit etwas reduziert. Diese sogenannte *intrinsic Unreliability* ist umso größer, je kleiner die Ensembles sind (WEIGEL et al. 2007). Aus diesem Grund sind probabilistische *Skill Scores*, insbesondere wenn sie *Reliability* messen, stark von der Ensemblegröße abhängig. Dies kann man beispielsweise in Abbildung 4-7 sehen, welche den *Ranked Probability Skill Score* (RPSS) für synthetisch generierte Toymodellvorhersagen in Abhängigkeit von der Ensemblegröße zeigt (durchgezogene schwarze Linie). Wie man sehen kann nimmt der *Skill* mit kleiner werdender Ensemblegröße ab, und zwar unabhängig davon, ob die Vorhersagen *reliable*, überdispersiv oder unterdispersiv sind.

Im Allgemeinen ist diese Abhängigkeit des *Skills* von der Ensemblegröße eine gewollte Eigenschaft, da sie die reduzierte Vorhersagequalität angesichts erhöhter Stichprobenunsicherheit widerspiegelt. Es gibt aber auch Fragestellun-

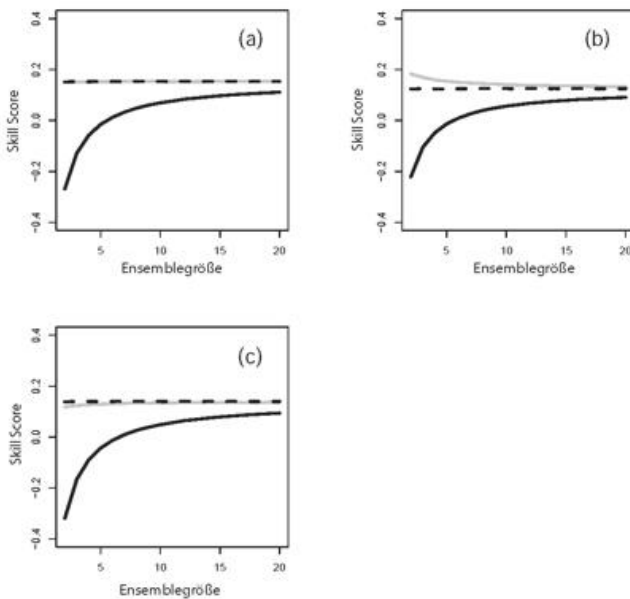


Abbildung 4-7: RPSS (durchgezogene schwarze Linie), RPSSd (durchgezogene graue Linie) und der mit der Methode des „Unbiased RPS-Estimator“ korrigierte RPSS (gestrichelte schwarze Linie) in Abhängigkeit von der Ensemblegröße. Die Skill-Werte basieren auf synthetischen Toymodellvorhersagen (Details in Kapitel 2) für drei gleichwahrscheinliche Ereigniskategorien. (a) basiert auf Vorhersagen, die *reliable* sind (Toymodellparameter: $\alpha=1$, $\beta=0$, $\gamma=1$). (b) basiert auf unterdispersiven Vorhersagen ($\alpha=1$, $\beta=0$, $\gamma=0,6$). Und (c) basiert auf überdispersiven Vorhersagen ($\alpha=1$, $\beta=0$, $\gamma=1,5$).

gen, bei denen diese Abhängigkeit unerwünscht ist, zum Beispiel wenn man mehrere Modelle unterschiedlicher Ensemblegröße miteinander vergleichen möchte; oder wenn man die potenzielle Vorhersagegüte abschätzen möchte, die man bei unendlicher Ensemblegröße erreichen würde. Für die meisten probabilistischen *Skill Scores* ist der Einfluss der Ensemblegröße allerdings noch nicht explizit quantifiziert worden. Eine Ausnahme bildet der *Ranked Probability Skill Score* RPSS (und damit automatisch auch der *Brier Skill Score* BSS), für den es zwei Ansätze einer Ensemblegrößenkorrektur gibt. Der erste Ansatz („RPSSd“, graue Linie in Abbildung 4-7) reduziert die Abhängigkeit des RPSS von der Ensemblegröße durch eine Anpassung der klimatologischen Referenzstrategie (MÜLLER et al. 2005, WEIGEL et al. 2007). Der zweite Ansatz („Unbiased RPS-Estimator“, gestrichelte Linie) ändert nichts an der Referenzstrategie, sondern schätzt direkt die Auswirkung der Ensemblegröße auf den RPS ab (FERRO et al. 2008). Wie man aus Abbildung 4-7 sehen kann, reduzieren beide Ansätze die Ensemblegrößenabhängigkeit deutlich. Eine technische Beschreibung und ein Vergleich der beiden Ansätze finden sich bei WEIGEL (2011).

5 Zusammenfassung und Ausblick

Ziel dieses Artikels war, einen kurzen Überblick über Fragestellungen und gebräuchliche Ansätze der Verifikation von Ensemblevorhersagen zu geben. Eine der wichtigsten Aussagen ist, dass eine Ensemblevorhersage a priori noch keine Wahrscheinlichkeitsprognose ist, sondern „lediglich“ ein

Set von deterministischen Einzelvorhersagen. Im Idealfall stellen die Ensemblemitglieder unabhängige Stichproben der zu Grunde liegenden Wahrscheinlichkeitsverteilung der Vorhersage dar. In diesem Fall lassen sich aus den Ensembles dann Vorhersageunsicherheiten oder Eintrittswahrscheinlichkeiten für spezifische Ereignisse ableiten. Diese beiden Interpretationsniveaus – Stichproben und Wahrscheinlichkeiten – spiegeln sich auch in den für Ensemblevorhersagen verwendeten Gütemaßen wider. So gibt es einerseits *Skill Scores*, welche die einzelnen Ensemblemitglieder als Stichproben einer Wahrscheinlichkeitsverteilung interpretieren. Sie charakterisieren in der Regel die kollektiven statistischen Eigenschaften der Ensembles. Sie erfordern nicht, dass die Ensemblevorhersagen bereits vor der Verifikation in Wahrscheinlichkeitsverteilungen umgewandelt worden sind, sondern können im Prinzip auf die „rohen“ Einzelvorhersagen angewendet werden. Diese Art der Verifikation gibt oftmals sehr nützliche Informationen darüber, welche Annahmen notwendig sind, um aus den Ensembleverteilungen dann Wahrscheinlichkeiten abzuleiten.

Auf der anderen Seite gibt es probabilistische *Skill Scores*, welche sich dann anwenden lassen, wenn die Ensembles bereits auf irgendeine Weise in eine Wahrscheinlichkeit oder eine Unsicherheit „umgewandelt“ worden sind. Sie messen also die aus Anwendersicht besonders interessante „konkrete“ Güte der gemachten Wahrscheinlichkeitsprognosen.

Für die erste Kategorie, die Stichprobeninterpretation, gibt es eine Vielzahl von Tests, die in der Fachliteratur genau beschrieben sind. Die meisten dieser Tests charakterisieren vor allem die *Reliability* der Ensembles. Das heißt, sie untersuchen, ob sich die einzelnen Ensemblemitglieder statistisch identisch wie die Beobachtungswerte verhalten, und falls nicht, ob dies bestimmten charakteristischen Fehlermustern wie beispielsweise Unterdispersion oder systematischem Bias zugeordnet werden kann. Die *Reliability* von Ensembles ist dabei immer als eine mittlere statistische Eigenschaft des gesamten Verifikationsdatensatzes zu verstehen. Entsprechend bewerten die meisten Tests die durchschnittliche Konsistenz zwischen den Ensemblemitgliedern und den Beobachtungen, machen aber keine Aussagen über die Situationsabhängigkeit (zum Beispiel Wetterlagen) der *Reliability*. In diesem Zusammenhang besteht sicherlich noch Forschungsbedarf. Des Weiteren gibt es noch Potenzial in der Formulierung und Anwendung von multidimensionalen Ensembleverifikationsansätzen. Die meisten existierenden Skillmaße sind für statistisch unabhängige eindimensionale Punktvorhersagen konzipiert und ignorieren dabei, dass in vielen Anwendungen Vorhersagen mehrerer Variabler an mehr als einem Ort simultan verwendet werden, was dann auch eine entsprechende mehrdimensionale Verifikation erfordert. Zwar gibt es hierzu bereits ein paar viel versprechende Ansätze; diese werfen ihrerseits aber wieder neue Fragen auf, welche weiterer Erforschung bedürfen.

Für die zweite Verifikationsklasse, die der probabilistisch interpretierten Ensemblevorhersagen, können im Prinzip

alle gängigen probabilistischen Gütemaße verwendet werden. Die Wahl des Scores hängt dabei vor allem vom Vorhersagekontext ab, also davon zum Beispiel, ob die Vorhersagen für diskrete Ereigniskategorien oder für ein kontinuierliches Wertespektrum berechnet werden, sowie vom Skill-Attribut, welches man quantifizieren will, also zum Beispiel Reliability oder Resolution. Hierbei ist es wichtig zu betonen, dass probabilistische Skill Scores nicht nur die Qualität des Vorhersagemodells per se testen, sondern auch, inwieweit die zur Abschätzung von Wahrscheinlichkeiten und Unsicherheiten angewandte Methodik auf realistischen und gerechtfertigten Annahmen beruht. Außerdem hängen viele probabilistische Skill Scores auch von der Ensemblegröße ab. In der Regel nimmt der Skill ab, wenn die Ensembles kleiner werden. Für bestimmte Anwendungen kann dies eine unerwünschte Eigenschaft darstellen, zum Beispiel dann, wenn Modelle unterschiedlicher Ensemblegröße miteinander verglichen werden sollen. In den vergangenen Jahren hat die Forschung damit begonnen, diesen Aspekt näher zu untersuchen.

Danksagung

Diese Arbeit wurde mit Mitteln des Schweizerischen Nationalfonds im Rahmen des Nationalen Forschungsschwerpunkts Klima (NCCR Climate) unterstützt.

Literatur

- ANDERSON, D.L.T., STOCKDALE, T., BALMASEDA, M., et al., 2007: Development of the ECMWF seasonal forecast System 3. *ECMWF Technical Memorandum* **503**, 56 pp.
- ANDERSON, J.L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate* **9**, 1518-1530.
- BENEDETTI, R., 2010: Scoring rules for forecast verification. *Monthly Weather Review* **138**, 203-211.
- BRIER, G.W., 1950: Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**, 1-3.
- BRÖCKER, J., 2011: Probability forecasts. In: Forecast Verification: A Practitioner's Guide in Atmospheric Science, Zweite Auflage. Herausgegeben von I.T. Jolliffe and D.B. Stephenson. Veröffentlicht 2011 von John Wiley & Sons, Ltd.
- BRÖCKER, J., SMITH, L.A., 2008: From ensemble forecasts to predictive distribution functions. *Tellus* **60A**, 663-678.
- EPSTEIN, E.S., 1969: A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology* **8**, 985-987.
- FERRO, C.A.T., RICHARDSON, D.S., WEIGEL, A.P., 2008: On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications* **15**, 19-24.
- HAMILL, T.M., COLUCCI, S.J., 1997: Verification of Eta-RSM short-range forecasts. *Monthly Weather Review* **125**, 1312-1327.
- HERSBACH, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**, 559-570.
- LEITH, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review* **102**, 409-418.
- MASON, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine* **30**, 291-303.
- MASON, S.J., WEIGEL, A.P., 2009: A generic forecast verification framework for administrative purposes. *Monthly Weather Review* **137**, 331-349.
- MÜLLER, W.A., APPENZELLER, C., DOBLAS-REYES, F.J., LINIGER, M.A., 2005: A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal of Climate* **18**, 1513-1523.
- PALMER, T.N., BUIZZA, R., LEUTBECHER, M., et al., 2007: The ECMWF Ensemble Prediction System: recent and on-going developments. *ECMWF Technical Memorandum* **540**, 53 pp.
- ROULSTON, M.S., SMITH, L.A., 2002: Evaluating probabilistic forecasts using information theory. *Monthly Weather Review* **130**, 1653-1660.
- ROULSTON, M.S., SMITH, L.A., 2003: Combining dynamical and statistical ensembles. *Tellus* **55A**, 16-30.
- SHEKIN, D.J., 2007: Handbook of parametric and non-parametric statistical procedures. *Chapman and Hall/CRC*, 1776 Seiten.
- STEPHENSON, D.B., COELHO, C.A.S., DOBLAS-REYES, F.J., BALMASEDA, M., 2005: Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus* **57A**, 253-264.
- STOCKDALE, T.N., ANDERSON, D.L.T., ALVES, J.O.S., BALMASEDA, M.A., 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature* **392**, 370-373.
- TALAGRAND, O., VAUTARD, R., STRAUSS, B., 1997: Evaluation of probabilistic prediction systems. In: Proceedings of *ECMWF Workshop on Predictability*, 20-22 October 1997, 1-25.
- WANG, X., BISHOP, C.H., 2005: Improvement of ensemble reliability with a new dressing kernel. *Quarterly Journal of the Royal Meteorological Society* **131**, 965-986.
- WEIGEL, A.P., 2011: Ensemble verification. In: Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2. Auflage. Herausgegeben von I.T. Jolliffe und D.B. Stephenson. Veröffentlicht von John Wiley & Sons, Ltd.
- WEIGEL, A.P., BOWLER, N.E., 2009: Comment on "Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?" *Quarterly Journal of the Royal Meteorological Society* **135**, 535-539.
- WEIGEL, A.P., MASON, S.J., 2011: The Generalized Discrimination Score for Ensemble Forecasts. *Monthly Weather Review* **139**, 3069-3074
- WEIGEL, A.P., LINIGER, M.A., Appenzeller, C., 2007: The discrete Brier and ranked probability skill scores. *Monthly Weather Review* **135**, 118-124.

- WEIGEL, A.P., LINIGER, M.A., Appenzeller, C., 2009: Seasonal ensemble forecasts: are recalibrated single models better than multimodels? *Monthly Weather Review* **137**, 1460-1479.
- WEIJS, S.V., VAN NOOIJEN, R., VAN DE GIESEN, N., 2010: Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. *Monthly Weather Review* **138**, 3387-3399.
- WILKS, D.S., 2002: Smoothing forecast ensembles with fitted probability distributions. *Quarterly Journal of the Royal Meteorological Society* **128**, 2821-2836.
- WILKS, D.S., 2006: *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, Vol. 91, Academic Press, 627 Seiten.
- WILSON, L.J., BURROWS, W.R., LANZINGER, A., 1999: A strategy for verification of weather element forecasts from an ensemble prediction system. *Monthly Weather Review* **127**, 956-970.

T. L. THORARINSDOTTIR, M. SCHEUERER, K. FELDMANN

5 Statistische Nachbearbeitung von Ensemblevorhersagen

Statistical post-processing of ensemble forecasts

Zusammenfassung

Statistische Nachbearbeitung von Ensemblevorhersagen ist ein wichtiger Bestandteil der modernen Wettervorhersage. In diesem Artikel geben wir einen Überblick über jene Nachbearbeitungsverfahren, die eine vollständige Vorhersageverteilung liefern. Wir diskutieren insbesondere BMA („Bayesian model averaging“) und EMOS („Ensemble model output statistics“) Verfahren sowie deren räumliche Erweiterungen. Die Leistungsfähigkeit dieser Methoden wird an einem Datenbeispiel mit COSMO-DE-Ensemblevorhersagen für Temperatur über Deutschland demonstriert, und die Notwendigkeit räumlicher Modellierung wird an einem Beispiel veranschaulicht. Zum Schluss diskutieren wir aktuelle Herausforderungen auf dem Gebiet der statistischen Nachbearbeitung.

Summary

Statistical post-processing of ensemble forecasts is an integral component of modern weather forecasting. In this paper, we review post-processing procedures that return full predictive distributions with a focus on Bayesian model averaging (BMA), ensemble model output statistics (EMOS), and spatial extensions thereof. We demonstrate the predictive performance of BMA and EMOS by applying them to COSMO-DE temperature forecasts over Germany and illustrate the need for spatial modelling in the post-processing procedure. Finally, we outline some current challenges in the development of statistical post-processing methods for ensemble forecasts.

1 Einführung

Eine der bedeutendsten Entwicklungen bei der Wettervorhersage in den letzten Jahrzehnten ist die Einführung von Ensemblevorhersagen. Traditionell wurde ein numerisches Modell mit einer Randbedingung verwendet, um die bestmögliche Vorhersage zu erstellen. Heutzutage werden stattdessen häufig mehrere numerische Modelle und/oder unterschiedliche Randbedingungen verwendet, um eine Vielzahl von unterschiedlichen Vorhersagen (ein „Ensemble“) zu erzeugen, die verschiedene Szenarien der Wetterentwicklung erfassen und die Vorhersageunsicherheit abbilden sollen. Um das maximale Leistungsvermögen einer Ensemblevorhersage zu erreichen, ist es wichtig, eine statistische Nachbearbeitung der Ensemblevorhersage durchzuführen. Obwohl Ensemblevorhersagen sich in

den letzten Jahren stark verbessert haben, können einfache statistische Verfahren – wie zum Beispiel Biaskorrektur und Nachkalibrierung der Unsicherheit – eine positive Auswirkung auf die Güte der Vorhersagen haben.

In diesem Beitrag geben wir einen Überblick über statistische Nachbearbeitungsverfahren von Ensemblevorhersa-

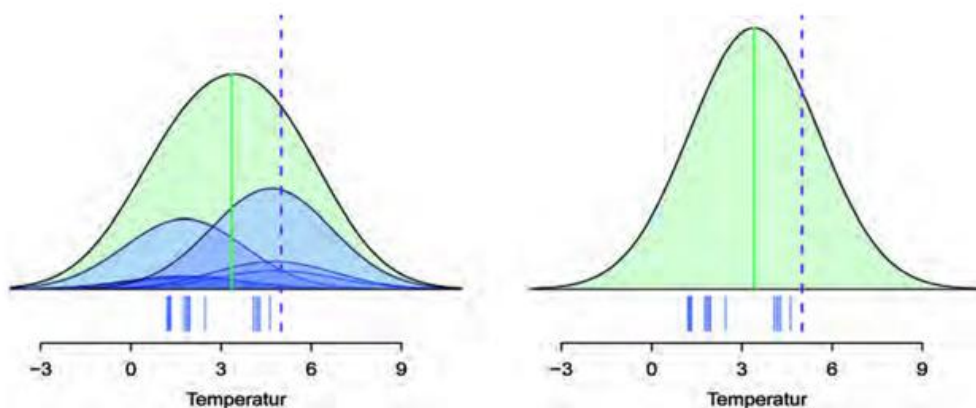


Abbildung 5-1: Vorhersagedichten für die Temperatur an der Station Berlin-Tegel am 17. Januar 2011, 21 UTC, berechnet mittels BMA (links) beziehungsweise EMOS+ (rechts), siehe Abschnitt 2.2. Die einzelnen Kerndichten für BMA sind blau eingefärbt. Die beobachtete Temperatur ist 5°C (violette Linien), die COSMO-DE-Ensemblevorhersagen liegen zwischen 1,2°C bis 4,6°C (blaue Striche), und die Mediane der BMA und EMOS+ Vorhersagedichte sind hier beide gleich 3,4°C (grüne Linien).

gen auf der Mesoskala, die eine Vorhersage in Form einer Wahrscheinlichkeitsverteilung (vergleiche Abbildung 5-1) liefern. Die probabilistischen Vorhersagen, die dadurch erzeugt werden, ermöglichen die Quantifizierung von wetterabhängigen Risiken und bieten den Nutzern von Wettervorhersagen zusätzliche qualitative Information an. Dies ist in vielen Fachgebieten – wie zum Beispiel in der Landwirtschaft, in der Luftfahrt und in der Schifffahrt – von großer Bedeutung. Wir erheben nicht den Anspruch mit diesem Artikel eine vollständige Übersicht über die Literatur zu diesem Thema zu geben; vielmehr haben wir uns Arbeiten ausgesucht, die für uns wichtig und von Interesse sind. Weitere Darstellungen des Themas gibt es zum Beispiel in HAMILL et al. (2000).

Abschnitt 2 beginnt mit einer Übersicht über univariate (das heißt für einzelne Variablen und an einzelnen Orten, ohne Berücksichtigung von räumlichen Zusammenhängen) Vorhersagen. Wir beschreiben, wie die Güte einer univariaten probabilistischen Vorhersage bewertet werden kann, und besprechen die statistische Nachbearbeitung für verschiedene Wettervariablen. Im darauf folgenden Abschnitt 3 diskutieren wir die Bewertung von multivariaten probabilistischen Vorhersagen und beschreiben einige Verfahren zur Erzeugung von räumlich konsistenten, probabilistischen Vorhersagen. Der letzte Abschnitt 4 enthält eine Diskussion, in der wir vor allem die aktuellen Herausforderungen zum Thema statistische Nachbearbeitung von Ensemblevorhersagen darstellen.

2 Univariate Vorhersagen

Bezeichne mit y_s die vorherzusagende Größe am Ort s und $\hat{x}_{1s}, \dots, \hat{x}_{ms}$ die entsprechenden m Vorhersagen der Mitglieder des dynamischen Ensembles. Diese können als Stichprobe einer Vorhersageverteilung interpretiert, und probabilistische Aussagen daraus abgeleitet werden. Beispielsweise kann man für einen Schwellwert τ die relative Häufigkeit der Ensemble Mitglieder, die einen Wert größer als τ vorhersagen, als Überschreitungswahrscheinlichkeit für diesen Schwellwert am Ort s interpretieren. Der Ensemblemittelwert $\bar{x}_{1s} = \frac{1}{m} \sum_{k=1}^m \hat{x}_{ks}$ kann als Punktvorhersage, und die Ensemblevarianz $S_s^2 := \frac{1}{m-1} \sum_{k=1}^m (\hat{x}_{ks} - \bar{x}_{1s})^2$ als Maß für die Vorhersageunsicherheit verwendet werden. Hier zeigt sich oft bereits eine Schwäche der „rohen“ Ensembles: in vielen Fällen ist die Ensemblevarianz zu gering, das heißt die tatsächliche Unsicherheit wird unterschätzt (siehe Abschnitt 2.2.3), weil nicht alle Quellen von Unsicherheit explizit im numerischen Modell berücksichtigt werden können. Ebenso entsprechen die oben angesprochenen relativen Häufigkeiten oft nicht den wahren Überschreitungswahrscheinlichkeiten. Anstatt diese Werte einzeln zu korrigieren, werden zunehmend statistische Nachbearbeitungsmethoden verwendet, die die Ensemblevorhersagen in eine vollständige Vorhersageverteilung überführen, aus der dann sämtliche probabilistische Aussagen abgeleitet werden können. Die zwei vielleicht bekanntesten Verfahren sind BMA („Bayesian model averaging“) und EMOS („ensemble model output statistics“), letzteres wird in der Literatur häufig auch als NGR („nonhomogeneous Gaussian regression“) bezeichnet. Beide sind parametrisch und basieren auf der Annahme, dass sich der Vorhersagefehler der aktuellen Vorhersage ähnlich wie der Vorhersagefehler der n vorausgegangenen Tage verhält.

Bei BMA wird jedem Mitglied des Ensembles zunächst eine bedingte Verteilung („Kern“) zugeordnet, die zugehörige Kerndichte bezeichnen wir mit $p_k(y_s | \hat{x}_{ks})$. Die Notation deutet dabei an, dass diese Dichtefunktion für y_s von der Vorhersage \hat{x}_{ks} abhängt. Die Vorhersageverteilung entsteht dann als gewichtetes Mittel der Kerne

$$p(y_s | \hat{x}_{1s}, \dots, \hat{x}_{ms}) = \sum_{k=1}^m w_k p_k(y_s | \hat{x}_{ks}) \quad (1)$$

wobei die Gewichte w_1, \dots, w_m die Güte der einzelnen Mitglieder in der Trainingsperiode widerspiegeln. Die Vorhersageverteilung beim EMOS-Verfahren besteht dagegen aus einer einzigen Verteilung, wobei die Parameter der Verteilung (Mittelwert, Varianz, etc.) von den Ensemblevorhersagen abhängen. Welche Verteilungen dabei verwendet werden, hängt von den Eigenschaften der zugrunde liegenden Wettervariable ab.

Wir geben zunächst einen kurzen Überblick über die Bewertungsmethoden von univariaten Vorhersagen, und beschreiben dann einige Verfahren zur Nachbearbeitung der Wettervariablen Temperatur, Druck, Niederschlag und Wind. Weitere Verfahren werden zum Beispiel in BRÖCKER und SMITH (2008), WILKS (2006) und WANG und BISHOP (2004) beschrieben. Ferner gibt es Arbeiten zu anderen Wettervariablen, wie etwa BMA-Verfahren für Nebel (ROQUELAURE und BERGOT 2008) und Sichtweite (CHMIELECKI und RAFTERY 2010).

2.1 Bewertung von univariaten Vorhersagen

Das Ziel probabilistischer Vorhersagen ist, die Schärfe der Vorhersagedichte unter der Nebenbedingung, dass die Vorhersageverteilung kalibriert ist, zu maximieren, das heißt die Dichte soll eine möglichst kleine Standardabweichung haben (GNEITING und RAFTERY 2007). Die Schärfe ist ausschließlich eine Eigenschaft der Dichte, während die Kalibrierungseigenschaft sowohl von der Vorhersage als auch von den Beobachtungen abhängt. Ob ein Ensemble kalibriert ist, kann mittels des Ranghistogramms überprüft werden: Berechne für jede Beobachtung ihren Rang innerhalb der gemeinsamen Stichprobe aus Beobachtung und den m Ensemblevorhersagen. Falls die Beobachtungen und ihre dazugehörigen Vorhersagen jeweils aus denselben Verteilungen stammen, sollte das Histogramm der Ränge aller Beobachtungen gleichverteilt auf $1, \dots, m+1$ sein. Ein U-förmiges Histogramm deutet auf eine zu geringe Streuung, ein in der Mitte

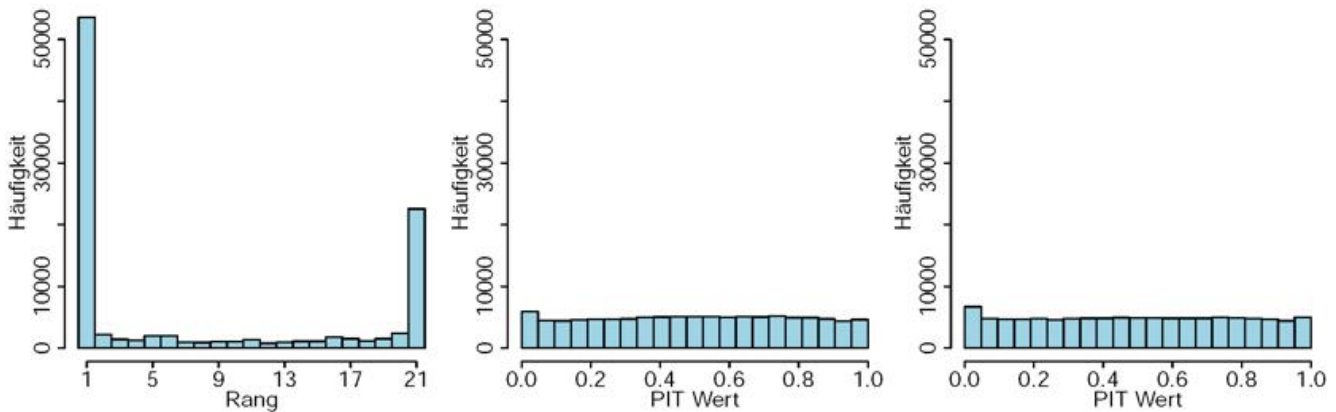


Abbildung 5-2: Ranghistogramm des COSMO-DE-Vorhersageensembles (links) und PIT-Histogramme für BMA- (Mitte) und EMOS*-Vorhersageverteilung (rechts). Die Ergebnisse sind für den Zeitraum von 05.01.2011 bis 30.06.2011 und über alle Standorte aggregiert.

nach oben gewölbtes Histogramm auf eine zu große Streuung des Ensembles hin (HAMILL und COLUCCI 1997). Im Beitrag 4 dieser PROMET Ausgabe von A. P. Weigel werden Ranghistogramme ausführlich erklärt und mit Beispielen illustriert. Der Gegenpart zum Rang im kontinuierlichen Fall, das heißt wenn statt eines Ensembles eine Vorhersagedichte gegeben ist, ist der PIT („probability integral transform“) Wert; hier wird für jede Beobachtung der Wert der Vorhersageverteilung im Beobachtungswert ausgerechnet. Falls die Vorhersagen kalibriert sind, dann sollte das PIT-Histogramm ebenfalls gleichverteilt aussehen (vergleiche GNEITING et al. 2007, und dort angegebene Referenzen). Beispiele für Rang- und PIT-Histogramme sind in Abbildung 5-2 gegeben, Details zu den dort verglichenen Vorhersagen werden in Abschnitt 2.2.3 gegeben.

Die Güte einer Vorhersage kann auch durch Bewertungsregeln gemessen werden. Für einzelne Vorhersagen werden dafür häufig der mittlere absolute Fehler (MAF) und der mittlere quadratische Fehler (MQF) verwendet. Um diese Bewertungsregel auf Vorhersageverteilungen anzuwenden, muss die Verteilung auf einen Wert reduziert werden. Dabei ist es wichtig, den theoretisch optimalen Wert bezüglich der Bewertungsregel zu verwenden. Für den MAF ist dies der Median, für den MQF der Mittelwert der Verteilung (GNEITING 2011). Die dritte Bewertungsregel, die wir hier betrachten, ist der CRPS („continuous ranked probability score“). Für eine Vorhersageverteilung F und eine Beobachtung y ist er gegeben durch

$$CRPS(F, y) = \int_{-\infty}^{\infty} (F(t) - 1\{y \leq t\})^2 dt = E|X - y| - \frac{1}{2} E|X - X'|$$

Dabei bezeichnet $1\{y \leq t\}$ die Indikatorfunktion, die für $y \leq t$ den Wert 1, anderenfalls den Wert 0 annimmt, E ist der Erwartungswert, X und X' sind unabhängige Zufallsvariablen mit Verteilungsfunktion F (vergleiche GNEITING et al. 2007, und dort angegebene Referenzen). Im Allgemeinen ist es üblich, nur korrekte

Bewertungsregeln zu verwenden; eine Bewertungsregel wird als „korrekt“ bezeichnet, falls die wahre Verteilung die bestmögliche erwartete Bewertung erzielt. Alle drei oben genannten Bewertungsregeln sind in diesem Sinne korrekt. Weitere Informationen zu Bewertungsregeln gibt es zum Beispiel in GNEITING und RAFTERY (2007) und HELD (2008).

2.2 Vorhersagen für Temperatur und Druck

2.2.1 BMA

Der einfachste Fall für die statistische Nachbearbeitung liegt vor, wenn die zu untersuchende Wettervariable durch eine Normalverteilung $N(\cdot, \cdot)$ gut beschrieben werden kann, beispielsweise Temperatur oder Druck (vergleiche RAFTERY et al. 2005). Der der Vorhersage des k -ten Ensemblemitglieds entsprechende Kern $p_k(y_{st} | \hat{x}_{kst})$ ist dann die Dichte der Verteilung $N(a_k + b_k f_{ks}, \sigma^2)$ mit Parametern $a_k, b_k, k = 1, \dots, m$, die eine mitgliederspezifische Bias-Korrektur der Vorhersagen \hat{x}_{ks} bewirken, und einem gemeinsamen Varianzparameter σ^2 . Die Gesamtvarianz der BMA-Vorhersagedichte ist dann

$$V(y_s | \hat{x}_{1s}, \dots, \hat{x}_{ms}) = \sum_{k=1}^m w_k \left((a_k + b_k \hat{x}_{ks}) - \left(-\sum_{k=1}^m w_k (a_k + b_k \hat{x}_{ks}) \right)^2 \right) + \sigma^2$$

der erste Teil davon wird durch die Streuung des Ensembles generiert, der zweite modelliert die zusätzliche, nicht durch das Ensemble erfasste Unsicherheit.

Gegeben Beobachtungen und Ensemblevorhersagen an den Trainingstagen \mathcal{T} , erhält man die Modellparameter wie folgt. Zunächst bestimmt man $a_k, b_k, k = 1, \dots, m$ für jedes Mitglied durch einfache lineare Regression. Diese Parameter werden danach als bekannt betrachtet, und im zweiten Schritt werden w_1, \dots, w_m und σ^2 so bestimmt, dass die Log-Likelihood

$$l(w_1, \dots, w_m, \sigma^2) = \sum_{t \in \mathcal{T}} \log \left(\sum_{k=1}^m w_k p_k(y_{st} | \hat{x}_{kst}) \right) \quad (2)$$

maximiert wird. Die Trainingstage werden zur Vereinfachung als unabhängig angenommen. Selbst unter Normalverteilungsannahme kann das Maximum von (2) nicht analytisch berechnet werden. RAFTERY et al. (2005) verwenden dazu einen EM-Algorithmus, der abwechselnd die Erwartungswerte einiger Hilfsgrößen berechnet („E-Schritt“) und die Log-Likelihood unter Verwendung dieser Hilfsgrößen maximiert („M-Schritt“). Die Erwartungswerte hängen dabei selbst wieder von den unbekanntem Parametern ab, daher werden jeweils die aktuellen Schätzwerte verwendet und solange zwischen beiden Schritten iteriert, bis sich der Zielwert praktisch nicht mehr ändert (Details siehe RAFTERY et al. 2005). In einem dritten, optionalen Zusatzschritt kann der Parameter σ^2 durch numerische Minimierung des CRPS in der Trainingsperiode erneut geschätzt werden. Da CRPS eine der wichtigen Bewertungsregeln ist, kann dadurch die Qualität der BMA-Vorhersage gegebenenfalls noch weiter verbessert werden.

Ein Anwendungsbeispiel von BMA in der Praxis findet sich auf der Website www.probcast.com, in der BMA-Vorhersagen für Temperatur und Niederschlag bis zu 48 Stunden über dem Pazifischen Nordwesten der USA und Kanada angegeben werden. WILSON et al. (2007) verwenden BMA für Temperaturvorhersagen in Kanada, SOLTANZADEH et al. (2011) benutzen die Methode für Vorhersagen im Iran. Einige kritische Kommentare und Verbesserungsvorschläge zu BMA findet man bei HAMILL (2007). BISHOP und SHANLEY (2008) argumentieren, dass BMA unter Umständen eine falsche Vorhersageklimatologie und insbesondere zu viele Extremwetterereignisse produziert. Sie schlagen ein Bayesianisches Verfahren vor, in dem BMA modifiziert und durch klimatologische Information ergänzt wird. Alternative Ansätze zur Parameterschätzung gibt es zum Beispiel in VRUGT et al. (2006) und VRUGT et al. (2008), ein vollständiger Bayesianischer Ansatz wird von DI NARZO und COCCHI (2010) diskutiert. Das hier angesprochene BMA-Verfahren ist in einem R-Paket (FRALEY et al. 2011) implementiert, welches auch Erweiterungen für austauschbare Ensemblemitglieder und fehlende Daten enthält (FRALEY et al. 2010).

2.2.2 EMOS

Bei der EMOS-Methode wird, wie oben schon erwähnt, eine einzelne Vorhersageverteilung angepasst, deren Parameter von den Ensemblevorhersagen abhängen. Für Wettervariablen wie Temperatur und Druck schlagen GNEITING et al. (2005) das folgende Modell vor:

$$y_s = a + b_1 f_{1s} + \dots + b_K f_{Ks} + \varepsilon_s, \quad \varepsilon_s \sim N(0, c + dS_s^2) \quad (3)$$

wobei S_s^2 die Ensemble-Varianz und c, d nichtnegative Koeffizienten sind. Dies entspricht einem Regressionsmodell mit den Ensemblevorhersagen als Prädiktoren und normalverteiltem Fehler, bei dem die Varianz der Fehlerterme in linearer Beziehung zur Ensemblevarianz steht.

Zur Schätzung der Modellparameter für ein bestimmtes Vorhersagedatum werden wieder die Daten der zugehörigen Trainingsperiode \mathcal{T} verwendet. Von den beiden von GNEITING et al. (2005) getesteten Schätzmethoden – Maximum Likelihood und Minimierung des CRPS – liefert Letztere für den dort untersuchten Datensatz deutlich bessere probabilistische Vorhersagen. Der CRPS kann im Normalverteilungsfall in geschlossener Form angegeben werden (GNEITING et al. 2005), was dessen Berechnung deutlich vereinfacht und den Rechenaufwand der EMOS Modellanpassung am Computer vergleichsweise gering hält. Die Regressionskoeffizienten b_1, \dots, b_K in (3) können im Prinzip beliebige Werte annehmen. Da sie jedoch als Kombination einer multiplikativen Bias-Korrektur und einer Gewichtung der Ensemblemitglieder interpretiert werden können, schlagen GNEITING et al. (2005) vor, die Modellanpassung unter der zusätzlichen Annahme der Nichtnegativität für diese Parameter durchzuführen. In ihrem Datenbeispiel erzielen sie mit diesem EMOS-Verfahren vergleichbare Verifikationsergebnisse wie mit dem Modell ohne die zusätzliche Einschränkung.

Auch zu EMOS (beziehungsweise NGR) gibt es zahlreiche Verweise in der Literatur. So vergleichen etwa WILKS und HAMILL (2007) das EMOS-Verfahren mit verschiedenen anderen statistischen Nachbearbeitungsmethoden anhand von aktualisierten Vorhersagen („reforecasts“) über mehrere Jahrzehnte. EMOS erzielt dort sehr gute Ergebnisse, vor allem für 6 bis 14 Tage-Temperaturvorhersagen, ebenso in der Vergleichsstudie von WILKS (2006). In einer weiteren solchen Studie kommen VANNITSEM und HAGEDORN (2011) auch zum Schluss, dass EMOS im Allgemeinen sehr gute Vorhersagen liefert, sich die Normalverteilungsannahme in einigen Fällen jedoch als ungeeignet zeigt. Das EMOS-Verfahren wird von KANN et al. (2009) für Temperaturvorhersagen in Österreich verwendet.

2.2.3 Datenbeispiel

In unserer bisherigen Notation für BMA und EMOS haben wir unterstellt, dass Beobachtungsdaten und Vorhersagen nur an einem Ort s gegeben sind. In der Praxis betrachtet man jedoch typischerweise ein größeres Gebiet, in dem Ensemblevorhersagen auf einem regulären Gitter und Beobachtungsdaten an einer Anzahl unregelmäßig verteilter SYNOP-Stationen vorliegen. Am einfachsten geht man dann wie folgt vor:

1. Berechne Ensemblevorhersagen für alle Standorte $S \in \mathcal{S}$ der SYNOP-Stationen mittels bilinearer Interpolation der Vorhersagen auf dem Gitter.
2. Passe ein BMA/EMOS-Modell mit den Trainingsdaten an diesen Standorten an,
 - mit identischen Modellparametern für alle $S \in \mathcal{S}$,
 - für die Parameterschätzung werden die Daten an allen Standorten S und allen Trainingstagen berücksichtigt.
3. Um eine probabilistische Vorhersage an einem beliebigen Ort zu erhalten, setze die entsprechenden Ensemblevorhersagen in die angepassten Modelle (1) bzw. (3) ein.

	MAF (°C)	\sqrt{MQF} (°C)	CRPS (°C)	Prädiktionsintervalle	
				Abdeckung (%)	Mittlere Weite (°C)
Ensemble	1,757	2,243	1,559	25,9	0,65
BMA	1,436	1,840	1,026	89,8	6,01
EMOS ⁺	1,438	1,845	1,027	88,6	5,80

Tabelle 5-1: Bewertungsergebnisse des unbearbeiteten COSMO-DE-Ensembles, der BMA- und der EMOS⁺-Methode, sowie Abdeckung und mittlere Weite von 90,5%-Prädiktionsintervallen. Die Ergebnisse sind für den Zeitraum von 05.01.2011 bis 30.06.2011 und über alle Standorte aggregiert.

Eine Verallgemeinerung dieses Vorgehens wird in Abschnitt 3.1 dargestellt, zunächst sollen aber die beiden beschriebenen Nachbearbeitungsverfahren an einem Datenbeispiel veranschaulicht werden.

Wir verwenden Temperaturvorhersagen des COSMO-DE-Ensembles, welches der Deutsche Wetterdienst seit 10. Dezember 2010 in einer präoperationellen Phase mit 20 Mitgliedern betreibt und ausführlich im Beitrag 6 dieser PROMET-Ausgabe von S. Theis und C. Gebhardt beschrieben wird. Betrachtet werden die Vorhersagen für jeweils 21 UTC mit Initialisierungszeit 00 UTC. Verifiziert wird mit SYNOP-Daten von über 500 Standorten in Deutschland. Wir vergleichen die probabilistischen Vorhersagen, die direkt aus dem unbearbeiteten Ensemble abgeleitet werden können, mit denen der BMA/EMOS⁺-Modelle, welche wie oben beschrieben mit Trainingsperioden von jeweils 25 Tagen angepasst wurden. Der Verifikationszeitraum beginnt somit am 05.01.2011 und endet am 30.06.2011. In Abbildung 5-1 werden BMA und EMOS⁺-Vorhersageverteilungen für einen ausgewählten Verifikationstag veranschaulicht. Die Dichtefunktion der BMV-Vorhersage (grün) entsteht durch Überlagerung der einzelnen, gewichteten Kerndichten (blau). Die EMOS-Vorhersage entspricht hingegen einer einzelnen Normalverteilungsdichte, deren Parameter von den Ensemblevorhersagen (blaue Striche) abhängen. Beide Vorhersagedichten sind deutlich breiter als die Spannweite des Ensembles, und korrigieren damit dessen zu geringe Streuung.

Abbildung 5-2 zeigt Rang- beziehungsweise PIT-Histogramme für die probabilistischen Vorhersagen des unbearbeiteten Ensembles, BMA und EMOS⁺. Die ausgeprägte U-Form im ersten Bild zeigt, dass die Vorhersageunsicherheit vom Ensemble deutlich unterschätzt wird. Sowohl BMA als auch EMOS⁺ korrigieren diese Unterschätzung und liefern gut kalibrierte Vorhersagen. Die erzielten Bewertungsergebnisse (siehe Tabelle 5-1) bestätigen dies. Bei MAF und MQF schneiden BMA und EMOS⁺ bereits wesentlich besser ab, beim CRPS ist der Unterschied sogar noch deutlicher, weil hier neben der Vorhersagegenauigkeit auch die realitätsgetreue Wiedergabe der Vorhersageunsicherheit bewertet wird. Die zu geringe Streuung des Ensembles macht sich ebenfalls bei den Prädiktionsintervallen bemerkbar. Theoretisch

sollten $19/21 \approx 90,5\%$ der Beobachtungen zwischen jeweils der kleinsten und größten Ensemblevorhersage liegen, tatsächlich sind es nur 26 %. Die zum selben Niveau berechneten BMA- und EMOS⁺-Prädiktionsintervalle halten die versprochene Abdeckung gut ein.

2.3 Vorhersagen für Niederschlag

Die Normalverteilung als grundlegende Form der Vorhersagedichte ist nicht für alle Wettervariablen geeignet. Zum Beispiel können Windgeschwindigkeiten oder Niederschlagsmengen nicht negativ sein, was sich durch eine Normalverteilung nicht abbilden lässt. Für Niederschlag hat man die zusätzliche Schwierigkeit, dass der Null-Millimeter-Wert eine Sonderrolle einnimmt, da er mit positiver Wahrscheinlichkeit auftritt. Eine rein kontinuierliche Vorhersagedichte kann dies nicht abbilden.

Ein BMA-Ansatz, der dieser Herausforderung Rechnung trägt, wurde von SLOUGHTER al. (2007) vorgestellt. Die BMA-Vorhersageverteilung ist wieder eine Mischverteilung im Sinne von (1), die dem k -ten Ensemblemitglied entsprechende Komponente ist nun jedoch zusammengesetzt aus einer Punktmasse $s \in S$ auf der Null und einer Gamma-Verteilung $\Gamma(\alpha_{ks}, \beta_{ks})$:

$$\tilde{y}_s | \hat{x}_{ks} \sim \pi_0(\hat{x}_{ks}) \cdot \delta_0 + \pi_+(\hat{x}_{ks}) \cdot \Gamma(\alpha_{ks}, \beta_{ks})$$

mit Parametern α_{ks}, β_{ks} die von \hat{x}_{ks} abhängen. Dabei bezeichnen $\tilde{y}_s | \hat{x}_{ks}$ die bedingte Verteilung von \tilde{y}_s gegeben \hat{x}_{ks} und $\pi_0(\hat{x}_{ks}), \pi_+(\hat{x}_{ks})$ die Wahrscheinlichkeiten für Null- beziehungsweise positiven Niederschlag. Modelliert wird damit die dritte Wurzel \tilde{y}_s der Niederschlagsmenge y_s , für die dieses Verteilungsmodell eine gute Approximation darstellt.

Für die Wahrscheinlichkeit $\pi_0(\hat{x}_{ks})$, dass kein Niederschlag auftritt, wird – separat für jedes Ensemblemitglied – ein logistisches Regressionsmodell angepasst. Der Mittelwert der Gammaverteilungskomponente wird durch einfache lineare Regression auf die dritte Wurzel \tilde{f}_{ks} der vorhergesagten Niederschlagsmenge bestimmt. Die Parameter für die Varianz dieser Komponente, hier ebenfalls abhängig von den Ensemblevorhersagen, sind für alle Mitglieder einheitlich und werden wie in Abschnitt 2.2 zusammen mit den Gewichten durch einen EM-Algorithmus geschätzt.

Ein ganz anderer Ansatz zur statistischen Nachbearbeitung von Niederschlagsvorhersagen, der ebenfalls eine vollständige, probabilistische Vorhersage liefert, wurde von WILKS (2009) entwickelt. Ein logistisches Regressionsmodell, welches zunächst nur für einzelne Nicht-Überschreitungswahrscheinlichkeiten $P(y_s \leq q)$ mit festem q definiert ist, wird dabei so erweitert, dass alle Wahrscheinlichkeiten $P(y_s \leq q), q \geq 0$, simultan in einem einzigen logistischen

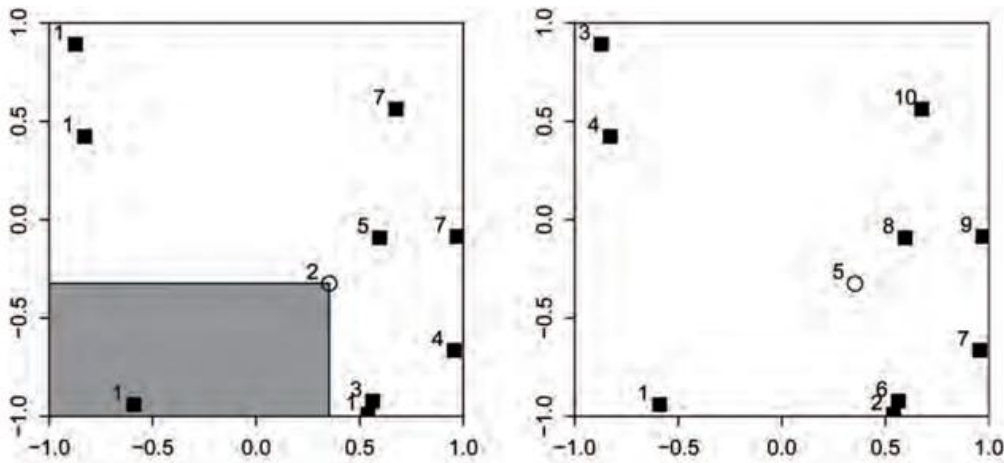


Abbildung 5-3: Ein Beispiel für die Berechnung multivariater Ränge eines Ensembles mit 9 Mitgliedern (schwarze Quadrate) und der dazugehörigen Beobachtung (Kreis) auf $[-1, 1] \times [-1, 1]$. Links: die Prä-ränge. Die Beobachtung hat Prä-rang 2, da ein Ensemblemitglied links unterhalb der Beobachtung liegt. Rechts: die multivariaten Ränge. Vier Ensemblemitglieder haben Prä-rang 1, während nur die Beobachtung Prä-rang 2 hat. Daher bekommt die Beobachtung Rang 5. Den vier Ensemblemitgliedern mit Prä-rang 1 werden zufällig die Ränge 1,2,3,4 zugeteilt.

Regressionsmodell modelliert werden. Erreicht wird dies durch die Annahme einer festen Beziehung zwischen den verschiedenen Schwellenwerten q in der Regressionsgleichung, was implizit der Annahme einer parametrischen Verteilungsdichte entspricht. Anders als bei einem logistischen Regressionsmodell, das die verschiedenen Schwellenwerte unabhängig voneinander behandelt, garantiert das erweiterte logistische Regressionsmodell von Wilks, dass alle daraus abgeleiteten probabilistischen Aussagen in sich konsistent sind. Nach unserem Wissen gibt es bislang noch keine direkte Verallgemeinerung des in Abschnitt 2.2 vorgestellten EMOS-Modells für Niederschlag, das die besonderen Herausforderungen dieser Wettervariablen an die Form der Vorhersageverteilung angemessen berücksichtigt.

SCHMEITS und KOK (2010) vergleichen die beiden oben beschriebenen Methoden in einer Studie mit aktualisierten EZMW-Vorhersagen für Niederschlag über den Niederlanden für einen Zeitraum von 20 Jahren. Sie schlagen eine Modifikation der ursprünglichen BMA-Methode vor, die insbesondere für längere Vorhersagezeiten zu besseren Vorhersagen führt.

2.4 Vorhersagen für Windvariablen

Windgeschwindigkeit nimmt, ähnlich wie Niederschlag, nur nichtnegative Werte an. Da Windgeschwindigkeitsmessungen aber selten genau gleich null sind, ist eine Punktmasse in Null wie bei Niederschlag nicht notwendig. SLOUGHTER u. a. (2010) schlagen ein BMA Modell für Windgeschwindigkeit vor, in dem die Kerndichten p_k aus dem allgemeinen BMA Modell in (1) durch Gamma-Verteilungen modelliert werden. Mittelwert und Varianz dieser Verteilung werden als lineare Funktionen von der jeweiligen Ensemblevorhersage \hat{x}_{ks} dargestellt. Ein BMA Ansatz für die Windrichtung wird von (2010) beschrieben. Die statistische Nachverarbeitung wird in zwei Schritten vorgenommen: Der erste besteht aus der Biaskorrektur der

einzelnen Vorhersagen, im zweiten Schritt wird das BMA Modell in (1) auf das biaskorrigierte Ensemble mit Kerndichten p_k , die durch von Mises-Verteilungen modelliert werden, angewandt.

THORARINSDOTTIR und GNEITING (2010) entwickeln einen EMOS-Ansatz für Windgeschwindigkeit, bei dem die Vorhersageverteilung durch trunkierte (unterhalb von Null abgeschnittene) Normalverteilung gegeben ist. Der Lageparameter der Verteilung hängt dabei über ein lineares Modell

von den Ensemblevorhersagen ab, der Skalenparameter steht in linearer Beziehung zur Varianz der Ensemblevorhersagen. THORARINSDOTTIR und GNEITING (2010) vergleichen ihr EMOS-Verfahren mit dem BMA-Verfahren von SLOUGHTER et al. (2010) und erhalten vergleichbare prädiktive Leistungen der beiden Methoden auf dem betrachteten Datensatz. Der EMOS-Ansatz wird in THORARINSDOTTIR und JOHNSON (2011) auf Windböen erweitert. Dort wird angenommen, dass keine Ensemblevorhersage für Windböen vorhanden ist, und die Vorhersageverteilung wird von der Vorhersageverteilung der Windgeschwindigkeit durch Böenfaktoren hergeleitet. Eine ähnliche Erweiterung ist für die BMA-Methode wegen der komplizierten Mischungsdichten in diesem Ansatz nicht möglich.

3 Statistische Nachbearbeitung im Raum

3.1 Multivariate Bewertungsmethoden

Eine Verallgemeinerung des univariaten Ranghistogramms ist das multivariate Ranghistogramm (GNEITING et al. 2008). Der Einfachheit halber beschreiben wir die Methode in zwei Dimensionen. Für reelle Variablen mit mehr als einer Dimension gibt es keine strikte Ordnung der Werte, wir können zum Beispiel nicht sagen, dass $(2, 4)$ größer oder kleiner als $(4, 2)$ ist. Hier verwenden wir die Ordnungsrelation

$$(x_1, x_2) \preceq (y_1, y_2) \text{ genau dann, wenn } x_1 \leq y_1 \text{ und } x_2 \leq y_2$$

für die zwei Vektoren $\mathbf{x} = (x_1, x_2)$ und $\mathbf{y} = (y_1, y_2)$. Der Prä-rang der Beobachtung \mathbf{y} innerhalb der Stichprobe S aus der Beobachtung und den m Ensemblevorhersagen ist gegeben durch

$$\rho(\mathbf{y}) = \sum_{\mathbf{x}' \in S} \mathbb{1}\{\mathbf{x}' \preceq \mathbf{y}\}$$

wobei $\mathbb{1}$ die Indikatorfunktion bezeichnet. Der Rang der Beobachtung wird dann wie folgt berechnet: Bestimme die Präränge aller Vektoren aus S und setze den Rang von y gleich dem Rang von $\rho(y)$ in der Menge

$$\{\rho(x)\}_{x \in S}$$

Falls mehrere Vektoren aus S den gleichen Prärang wie y haben, wird der Rang zufällig aus den möglichen Rängen gezogen. Das heißt, falls man zwei Vektoren mit demselben Prärang die Ränge 3 und 4 zuordnen müsste, wird einer davon auf 3 und der andere auf 4 gesetzt, jeweils mit 50 % Wahrscheinlichkeit. Abbildung 5-3 zeigt ein Beispiel zur Bestimmung von multivariaten Rängen in zwei Dimension, siehe auch Abbildung 2 in GNEITING et al. (2008).

Daneben gibt es auch multivariate Erweiterungen des MAF und des CRPS. Die Erweiterung des absoluten Fehlers ist der euklidische Abstand der Vektoren und die entsprechende Bewertungsregel wird der euklidische Fehler (EF) genannt. Das ES („energy score“) ist die multivariate Erweiterung des CRPS und wird durch

$$ES(F, y) = E\|X - y\| - \frac{1}{2} E\|X - X'\|$$

definiert, wobei X und X' unabhängige Zufallsvariablen mit Verteilungsfunktion F sind und $\|\cdot\|$ den euklidischen Abstand bezeichnet (GNEITING und RAFTERY, 2007). Weitere Information zur multivariaten Bewertung gibt es in GNEITING et al. (2008). Multivariate Bewertungsmethoden werden benötigt, wenn man entweder mehrere Wettervariablen, mehrere Vorhersagezeiten oder mehrere Verifikationsstandorte simultan betrachtet. Letzteres wird in den beiden folgenden Abschnitten im Zusammenhang mit statistischer Nachbearbeitung weiter diskutiert.

3.2 Lokal-adaptive Modellanpassung

In Abschnitt 2.3 haben wir einen einfachen Ansatz beschrieben, wie ein Modell für ein größeres Gebiet mit mehreren SYNOP-Stationen angepasst werden kann, indem man für alle Stationen dieselben Modellparameter wählt. Neben der stabileren Parameterschätzung (größere Datenmenge) hat dies den Vorteil, dass das Modell an jedem beliebigen Punkt im Raum definiert ist, und man nur die entsprechenden Ensemblevorhersagen einsetzen muss, um eine probabilistische Vorhersage zu bekommen.

Ein Nachteil bei diesem Vorgehen ist jedoch, dass das Nachbearbeitungsmodell auf lokale Charakteristiken keine Rücksicht nimmt, und damit auch nur die Kalibrierung im Mittel über alle Verifikationsstandorte S garantieren kann. KLEIBER et al. (2011a) und KLEIBER et al. (2011b) entwickeln eine flexible Erweiterung zum BMA, das sogenannte GMA („geostatistical model averaging“) Verfahren, in dem die Modellparameter im Raum variieren dürfen:

1. Berechne Ensemblevorhersagen für alle Standorte S mittels bilinearer Interpolation.
2. Passe für jedes $s \in S$ eine individuelle Kerndichte $p_{ks}(y_s | \hat{x}_{ks})$ an.
3. Bestimme mittels EM-Algorithmus die (für alle s identischen) Gewichte w_p, \dots, w_k .
4. Passe für jeden Modellparameter der Kerndichten aus 3. ein geostatistisches Modell an, welches neben den geographischen Koordinaten auch von weiteren Größen wie Meereshöhe oder Landnutzung abhängen kann.
5. Benutze diese Modelle um durch „Kriging“ (räumliche Interpolation) die Modellparameter aus 3. an einem beliebigen Ort s_o zu interpolieren.
6. Durch 5. ist in s_o wieder ein vollständiges BMA-Modell definiert, das die Ensemblevorhersagen in eine Vorhersageverteilung überführt.

3.3 Räumlich konsistente Vorhersagefelder

Die in Abschnitt 3.2 beschriebene Vorgehensweise stellt sicher, dass die berechneten probabilistischen Vorhersagen lokal kalibriert sind, und liefert Parameter, die einerseits räumlich variieren, andererseits aber für benachbarte Regionen annähernd gleich sind. Keines der bisher angesprochenen Verfahren modelliert jedoch explizit die Zusammenhänge der Vorhersageunsicherheit zwischen benachbarten Orten – vergleichbare Parameter der Vorhersageverteilung garantieren nicht, dass die Realisierungen der zugehörigen Zufallsvariablen ähnliche Werte annehmen. Solche Zusammenhänge können von großer Bedeutung sein, wenn man sich für räumlich aggregierte Größen (Mittelwerte, Maxima, etc.) interessiert. Beispielsweise im Hochwassermanagement möchte man die Gesamt-Niederschlagsmenge in einem bestimmten Flusseinzugsgebiet vorhersagen. Wird nun die Niederschlagsmenge an einem bestimmten Ort unterschätzt, dann mit hoher Wahrscheinlichkeit auch an einem benachbarten Ort. Die lokalen Vorhersageunsicherheiten hängen also voneinander ab, und dies beeinflusst die Gesamtunsicherheit.

Nun kann man, basierend auf den gegebenen Vorhersageverteilungen, ein statistisches Ensemble simulieren, das diese Verteilungen repräsentiert. Aus diesem kann man die räumlich aggregierten Größen von Interesse berechnen, und in geeignete Statistiken (Überschreitungswahrscheinlichkeiten, Quantile) zurückübersetzen. Führt man die Simulation jedoch – mangels Beschreibung der räumlichen Zusammenhänge – für jeden Ort unabhängig voneinander durch, ergibt sich ein statistisches Ensemble ähnlich dem in Abbildung 5-4. Die grundlegende Struktur, die den ursprünglichen Ensemblevorhersagen (hier nur vom ersten Ensemblemitglied) folgt, ist noch deutlich erkennbar, doch die unabhängig voneinander simulierten Abweichungen von der Vorhersage führen zu einem physikalisch unrealistischen Verhalten. Werden aus einem solchen statistischen Ensemble nun probabilistische Vorhersagen für räumlich aggregierte Größen abgeleitet, kann die Vorhersageunsicherheit unter Umständen völlig falsch eingeschätzt werden.

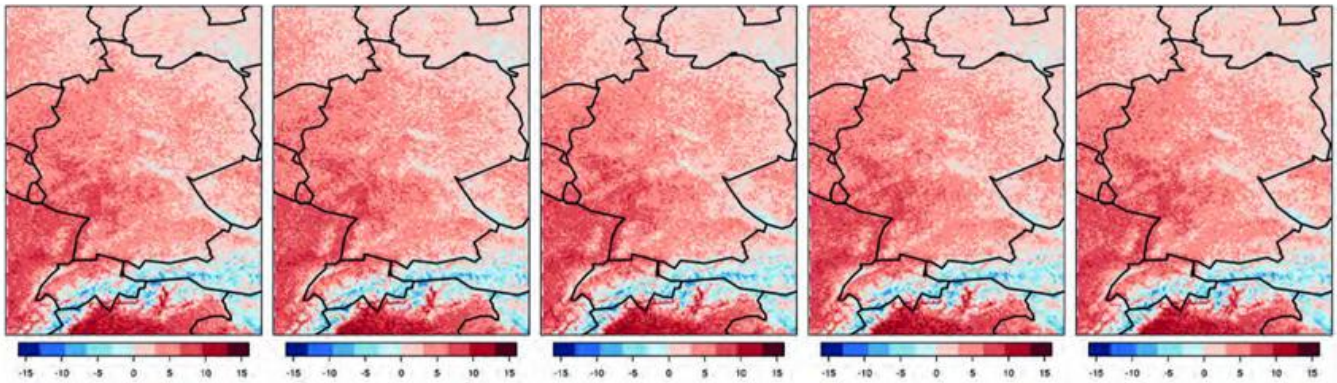


Abbildung 5-4: Statistisches Ensemble für die Temperatur in Deutschland am 01.03.2011, 15 UTC. Ohne ein Modell für räumliche Zusammenhänge wird die Vorhersageunsicherheit an allen Orten unabhängig voneinander simuliert und zu der Vorhersage des ersten COSMO-DE-Mitglieds addiert.

Auch hier können Methoden der räumlichen Statistik verwendet werden, um physikalisch sinnvollere, statistische Ensembles zu generieren. GEL et al. (2004) beschreiben, wie man für eine einzelne numerische Temperaturvorhersage ein geostatistisches Modell für die Vorhersageunsicherheiten anpassen, und damit räumlich konsistente Temperaturfelder simulieren kann. Räumlich konsistent bedeutet hier, dass benachbarte Orte tendenziell ähnliche Werte annehmen, auch wenn diese Werte selbst zufällig sind. Abbildung 5-5 zeigt fünf mit diesem GOP („geostatistical output perturbation“) Verfahren erzeugte Simulationen. Die Abweichungen von der ursprünglichen Vorhersage sind hier physikalisch plausibler, und Unsicherheiten von räumlich aggregierten Größen können deutlich besser **quantifiziert werden**. Letzteres wird in einem Datenbeispiel von BERROCAL et al. (2007) nachgewiesen, die GOP mit BMA zu einer neuen Methode „Spatial BMA“ kombinieren, die eine räumliche Modellierung von probabilistischen Temperaturvorhersagen ausgehend von einem Vorhersageensemble ermöglicht.

Ein Gegenpart zu GOP für Niederschlag ist das TSS („two stage spatial“)-Modell von BERROCAL et al. (2008). Der komplizierteren Natur der Niederschlagsverteilung, dort ebenfalls als Mischung von Punktmasse in Null und Gamma-Verteilung modelliert, wird durch ein Zwei-Schritt-Modell Rechnung getragen, das Auftreten und Intensität von Niederschlag durch zwei getrennte - „versteckte“- räumliche Prozesse beschreibt. Eine Kombination mit BMA wurde nach unserem Wissen bislang noch nicht implementiert. Konzeptionell einfacher und für alle Wettervariablen **gleichermaßen anwendbar**, ist das **Schaake-Shuffle** von CLARK et al. (2004). Bei dieser Methode wird die räumliche Struktur basierend auf historischen Daten modelliert, was allerdings voraussetzt, dass ein hinreichend langes Datenarchiv vorhanden ist.

4 Diskussion

Statistische Nachbearbeitung von Ensemblevorhersagen ist ein wichtiger Bestandteil der modernen Wettervorhersage. In diesem Artikel haben wir einen Überblick über die

Entwicklung solcher Verfahren über ungefähr die letzten zehn Jahre gegeben. Angesichts des enormen Wachstums dieses Gebietes in der nahen Vergangenheit haben wir nicht versucht eine vollständige Übersicht über die Literatur zu geben, sondern haben uns auf Arbeiten konzentriert, die für uns wichtig und interessant sind. Wir haben uns dabei auf Verfahren beschränkt, die eine vollständige Vorhersageverteilung liefern, und insbesondere die Methoden BMA und EMOS für verschiedene Wettervariablen genauer beschrieben. Auch hier konnten wir nicht alle Themen abdecken, und einige wichtige Teilgebiete der Meteorologie, wie etwa die Hydrologie, wurden nicht angesprochen. Dafür wurden neben den Methoden zur Erzeugung von Vorhersageverteilungen auch die Methoden zur Bewertung von (probabilistischen) Vorhersagen kurz angesprochen, da diese nicht nur zur Beurteilung der Leistungsfähigkeit eines Verfahrens benötigt werden, sondern oft selbst Bestandteil der Verfahren sind.

Wir haben die Themen näherungsweise chronologisch geordnet. Zunächst wurden Methoden entwickelt, die Vorhersageverteilungen für einzelne Variablen an einzelnen Orten, das heißt ohne die räumliche Struktur zu berücksichtigen, generieren. Verfahren für Temperatur und Druck wurden dabei zuerst entwickelt, da man diese Variablen gut durch Normalverteilung beschreiben kann, und diese vergleichsweise einfach handhabbar ist. In den letzten Jahren sind dann auf den univariaten Methoden aufbauende Verfahren entstanden, die zusätzlich die räumliche Struktur der Daten berücksichtigen.

Weiterer Entwicklungsbedarf bei statistischen Nachbearbeitungsverfahren für Ensemblevorhersagen besteht unserer Ansicht nach hauptsächlich in folgenden Bereichen: Bei Verfahren speziell für Extremwetterlagen, und allgemeinen Verfahren, die sowohl räumliche und zeitliche (mehrere Vorhersagezeiten) Zusammenhänge, als auch Zusammenhänge zwischen verschiedenen Wettervariablen berücksichtigen. Die Herausforderungen bei der Modellierung von Extremereignissen bestehen nicht nur in der Entwicklung geeigneter Modelle für diese Ereignisse, sondern auch in der Bewertung der entsprechenden Vorhersagen, die aufgrund des seltenen Auftre-

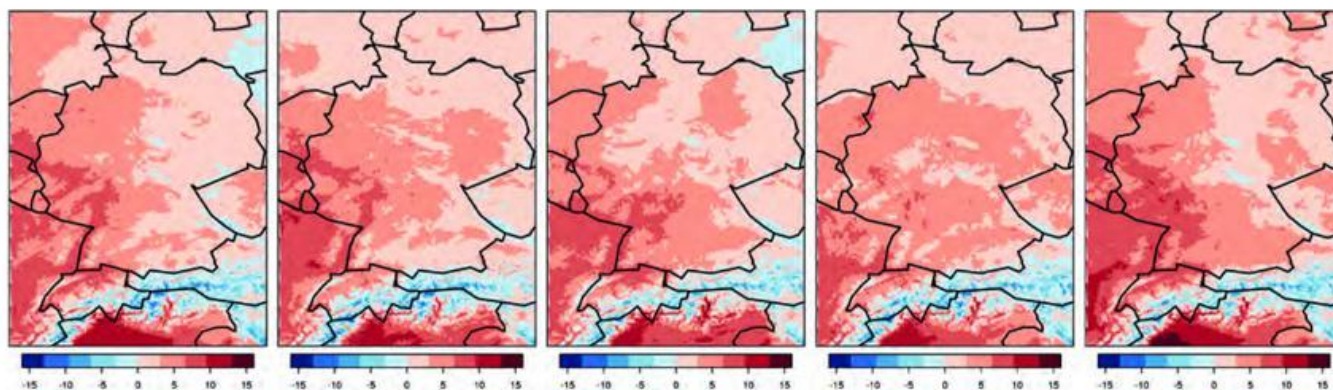


Abbildung 5-5: Statistisches Ensemble für die Temperatur in Deutschland am 01.03.2011, 15 UTC, diesmal mit der GOP-Methode simuliert. Auch hier wird nur die Vorhersage des ersten COSMO-DE-Mitglieds zugrunde gelegt, anhand der Trainingsdaten aber zusätzlich ein räumliches Modell angepasst und für die Simulation verwendet.

tens der Ereignisse besondere Sorgfalt erfordert. Daher werden sowohl theoretische als auch angewandte Fortschritte in der Entwicklung von Bewertungsmethoden benötigt, die speziell auf Extremereignisse zugeschnitten sind.

Die schon entwickelten multivariaten Verfahren im Abschnitt 3 können in zwei Gruppen eingeteilt werden. Parametrische Methoden wie jene in BERROCAL et al. (2008), KLEIBER et al. (2011a) und KLEIBER et al. (2011b), und nichtparametrische Methoden wie das Schaake-Shuffle von CLARK et al. (2004). Ein nichtparametrisches Verfahren braucht im Allgemeinen einen größeren Datensatz, ist dafür aber oft flexibler und kann in vielen Fällen mit nur geringem Rechenaufwand auf größere Gebiete oder mehrere Variablen erweitert werden. Die Erweiterung von parametrischen Modellen dagegen ist oft viel komplizierter, insbesondere kann es schwierig sein, die multivariate Korrelationsstruktur mittels eines parametrischen Modells zu beschreiben. Der benötigte Rechenaufwand, um die Parameter eines solchen Modells zu schätzen, kann mit der Größe des Datensatzes und der Anzahl der Dimensionen sehr schnell wachsen. Parametrische Modelle sind dafür meist besser geeignet um bestimmte Wetterlagen oder die Wetterlage an bestimmten Orten zu beschreiben. Für Extremereignisse sind häufig nicht genügend Daten vorhanden, und nicht parametrische Methoden dadurch praktisch nicht anwendbar. Parametrische Modelle sind auch für die Anwendung auf großen Flughäfen sehr vielversprechend, wo eine sehr genaue Vorhersage gebraucht wird, die sowohl in der Zeit als auch zwischen mehreren Wettervariablen konsistent ist. Auch hier besteht jedoch noch weiterer Entwicklungsbedarf.

Danksagung

Wir danken Renate Hagedorn für die Einladung, diesen Artikel zu schreiben, sowie Petra Friederichs, Tilmann Gneiting und Nina Schuhen für hilfreiche Kommentare und Anregungen. Die Arbeit von Michael Scheuerer und Kira Feldmann wird vom Deutschen Wetterdienst im Rahmen

des extramuralen Forschungsprojekts (Kennzeichen 33) unterstützt.

Literatur

BAO, L., GNEITING, T., GRIMIT, E. P., GUTTORP, P., RAFTERY, A. E., 2010: Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Mon. Weather Rev.* **138**, 1811–1821.

BERROCAL, V. J., RAFTERY, A. E., GNEITING, T., 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Weather Rev.* **135**, 1386–1402.

BERROCAL, V. J., RAFTERY, A. E., GNEITING, T., 2008: Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann. Appl. Stat.* **2**, 1170–1193.

BISHOP, C. H., SHANLEY, K. T., 2008: Bayesian model averaging’s problematic treatment of extreme weather and a paradigm shift that fixes it. *Mon. Weather Rev.* **136**, 4641–4652.

BRÖCKER, J., SMITH, L. A., 2008: From ensemble forecasts to predictive distribution functions. *Tellus A* **60**, 663–678.

CHMIELECKI, R. M., RAFTERY, A. E., 2010: Probabilistic visibility forecasting using Bayesian model averaging. *Mon. Weather Rev.* **139**, 1626–1636.

CLARK, M. P., GANGOPADHYAY, S., HAY, L. E., RAJAGOPALAN, B., WILBY, R. L., 2004: The Schaake Shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.* **5**, 243–262.

DI NARZO, A. F., COCCHI, D., 2010: A Bayesian hierarchical approach to ensemble weather forecasting. *J. Roy. Stat. Soc. Ser. C* **59**, 405–422.

FRALEY, C., RAFTERY, A. E., GNEITING, T., 2010: Calibrating multimodal forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Weather Rev.* **138**, 190–202.

FRALEY, C., RAFTERY, A. E., GNEITING, T., SLOUGHTER, J. M., BERROCAL, V. J., 2011: Probabilistic weather forecasting in R. *The R Journal* **3**, 55–63.

- GEL, Y., RAFTERY, A. E., GNEITING, T., 2004: Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation (GOP) method (with discussion and rejoinder). *J. Am. Stat. Assoc.* **99**, 575–590.
- GNEITING, T., 2011: Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **106**, 746–762.
- GNEITING, T., BALABDAOUI, F., RAFTERY, A. E., 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. B* **69**, 243–268.
- GNEITING, T., RAFTERY, A. E., 2005: Weather forecasting with ensemble methods. *Science* **310**, 248–249.
- GNEITING, T., RAFTERY, A. E., 2007: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**, 359–378.
- GNEITING, T., RAFTERY, A. E., WESTVELD, A. H., GOLDMAN, T., 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **133**, 1098–1118.
- GNEITING, T., STANBERRY, L. I., GRIMIT, E. P., HELD, L., JOHNSON, N. A., 2008: Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). *Test* **17**, 211–264.
- HAMILL, T. M., 2007: Comments on “Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging”. *Mon. Weather Rev.* **135**, 4226–4230.
- HAMILL, T. M., COLUCCI, S. J., 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Weather Rev.* **125**, 1312–1327.
- HAMILL, T. M., MULLEN, S. L., SNYDER, C., TOTH, Z., BAUMHEFNER, D. P., 2000: Ensemble forecasting in the short to medium range: report from a workshop. *B. Am. Meteorol. Soc.* **81**, 2653–2664.
- HELD, L., 2008: Methoden der statistischen Inferenz. Spektrum Akademischer Verlag, 304 Seiten.
- KANN, A., WITTMANN, C., WANG, Y., MA, X., 2009: Calibrating 2-m temperature of limited-area ensemble forecasts using high-resolution analysis. *Mon. Weather Rev.* **137**, 3373–3387.
- KLEIBER, W., RAFTERY, A. E., BAARS, J., GNEITING, T., MASS, C., GRIMIT, E. P., 2011: Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Mon. Weather Rev.* **139**, 2630–2649.
- KLEIBER, W., RAFTERY, A. E., GNEITING, T., 2011: Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *J. Am. Stat. Assoc.*, to appear.
- RAFTERY, A. E., GNEITING, T., BALABDAOUI, F., POLAKOWSKI, M., 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **133**, 1155–1174.
- ROQUELAURE, S., BERGOT, T., 2008: A local ensemble prediction system for fog and low clouds: construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteorol.* **47**, 3072–3088.
- SCHMEITS, M. J., KOK, K. J., 2010: A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Mon. Weather Rev.* **138**, 4199–4211.
- SLOUGHTER, J. M., GNEITING, T., RAFTERY, A. E., 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* **105**, 25–35.
- SLOUGHTER, J. M., RAFTERY, A. E., GNEITING, T., FRALEY, C., 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Weather Rev.* **135**, 3209–3220.
- SOLTANZADEH, I., AZADI, M., VAKILI, G. A., 2011: Using Bayesian model averaging (BMA) to calibrate probabilistic surface temperature forecasts over Iran. *Ann. Geophys.* **29**, 1295–1303.
- THORARINSDOTTIR, T. L., GNEITING, T., 2010: Probabilistic forecasts of wind speed: ensemble model output statistics using heteroskedastic censored regression. *J. Roy. Stat. Soc. A* **173**, 371–388.
- THORARINSDOTTIR, T. L., JOHNSON, M. S., 2011: Probabilistic wind gust forecasting using nonhomogeneous Gaussian regression. *Mon. Weather Rev.*, in press.
- VANNITSEM, S., HAGEDORN, R., 2011: Ensemble forecast post-processing over Belgium: comparison of deterministic-like and ensemble regression methods. *Meteorol. Appl.* **18**, 94–104.
- VRUGT, J. A., CLARK, M. P., DIKS, C. G. H., DUAN, Q., ROBINSON, B. A., 2006: Multiobjective calibration of forecast ensembles using Bayesian model averaging. *Geophys. Res. Lett.* **33**, L19817.
- VRUGT, J. A., DIKS, C. G. H., CLARK, M. P., 2008: Ensemble Bayesian model averaging using Markov Chain Monte Carlo sampling. *Environ. Fluid Mech.* **8**, 579–595.
- WANG, X., BISHOP, C. H., 2004: Improvement of ensemble reliability with a new dressing kernel. *Q. J. Roy. Meteor. Soc.* **131**, 965–986.
- WILKS, D. S., 2006: Statistical Methods in the Atmospheric Sciences, 2. Auflage. Elsevier Academic Press, 627 Seiten.
- WILKS, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorol. Appl.* **16**, 361–368.
- WILKS, D. S., HAMILL, T. M., 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Weather Rev.* **135**, 2379–2390.
- WILSON, L. J., BEAUREGARD, S., RAFTERY, A. E., VERRET, R., 2007: Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging. *Mon. Weather Rev.* **135**, 1364–1385.

S. THEIS, C. GEBHARDT

6 COSMO-DE-EPS: Das „konvektionserlaubende“ Ensemble am DWD

COSMO-DE-EPS: The convection-permitting ensemble at DWD

Zusammenfassung

COSMO-DE-EPS ist ein Ensemblevorhersagesystem basierend auf dem numerischen Wettervorhersagemodell COSMO-DE. Seine besonders feine Gittermaschenweite von 2,8 km erlaubt die explizite Simulation hochreichender Konvektion ohne die vereinfachende Parametrisierung. Konvektive Prozesse sind in ihrer deterministischen Vorhersagbarkeit stark begrenzt, so dass eine probabilistische Betrachtungsweise der Vorhersage erforderlich ist. Dazu dient das kürzlich entwickelte COSMO-DE-EPS. Es fällt in die Kategorie der „konvektionserlaubenden“ Ensembles. Seit Aufnahme des präoperationellen Betriebes im Dezember 2010 produziert das COSMO-DE-EPS Ensemblevorhersagen im zeitkritischen Modus. Nach erfolgreicher Evaluierung und Verifikation der Vorhersageprodukte während dieser präoperationellen Phase wurde das COSMO-DE-EPS am 22. Mai 2012 für den operationellen Betrieb freigegeben. Es ist damit europaweit das erste konvektionserlaubende Ensemble mit operationellem Status. Der Artikel führt in grundsätzliche Aspekte von „limited-area“ Ensembles ein, stellt den Aufbau von COSMO-DE-EPS vor, und präsentiert Verifikationsergebnisse sowie Feedback von Forecastern.

Summary

COSMO-DE-EPS is an ensemble prediction system based on the numerical weather prediction model COSMO-DE. Its very fine grid spacing of 2.8 km allows for the explicit simulation of deep convection without the simplifying parametrisation. Deterministic predictability of convective-scale processes is very limited, so the forecast must be treated in a probabilistic way. Therefore, COSMO-DE-EPS has recently been developed. It belongs to the category of „convection-permitting“ ensembles. Since December 2010, COSMO-DE-EPS has been providing ensemble forecasts in a time-critical manner. After successful evaluation of the ensemble products, the COSMO-DE-EPS has been switched to operational mode on 22nd May 2012. It is the first convection-permitting ensemble in operational state in Europe. This paper explains some general aspects of limited-area ensembles; it introduces the setup of COSMO-DE-EPS; and it presents verification results and feedback by forecasters.

1 Einleitung

Im Rahmen des Innovationsprogramms entwickelt und implementiert der DWD das Ensemblevorhersagesystem COSMO-DE-EPS. Es basiert auf dem numerischen Wettervorhersagemodell COSMO-DE (BALDAUF et al. 2011) und umfasst zurzeit 20 Ensemblemember. Die Mitglieder des Ensembles werden mit der gleichen horizontalen Gittermaschenweite gerechnet wie die operationelle Konfiguration des COSMO-DE (2,8 km). Das Ensemble produziert Vorhersagen für die nächsten 0 bis 21 Stunden und für ein Modellgebiet, das Deutschland und umliegende Gebiete umfasst.

Das COSMO-DE-EPS ist ein „Limited-area“-Ensemble, das heißt, das Modellgebiet umfasst nicht den ganzen Globus, sondern ein Teilgebiet. Durch die Reduzierung auf ein Teilgebiet kann die Vorhersage mit feinerer räumlicher Maschengitterweite gerechnet werden, ohne den Rahmen verfügbarer Computerressourcen zu sprengen. Im Falle von COSMO-DE-EPS ist die Maschengitterweite so fein, dass hochreichende Konvektion sogar explizit gerechnet werden kann und die vereinfachende Parame-

trisierung wegfällt. Das COSMO-DE-EPS fällt somit in die Kategorie der „konvektionserlaubenden“ Ensembles. Es ist europaweit das erste seiner Art im operationellen Betrieb.

Zur besseren Einordnung von COSMO-DE-EPS erläutert Abschnitt 2 zunächst einige grundsätzliche Aspekte von „Limited-area“-Ensembles im Vergleich zu globalen Ensembles. Abschnitt 3 erläutert den Aufbau von COSMO-DE-EPS. Abschnitt 4 beleuchtet erste Erfahrungen seit der Einführung des COSMO-DE-EPS in die Vorhersagepraxis am DWD. Dieser Abschnitt beschreibt die Verfügbarkeit von COSMO-DE-EPS Vorhersagen und eine erste Qualitätseinschätzung durch Verifikation und Forecaster. Abschnitt 5 schließt mit einer Zusammenfassung.

2 Grundsätzliche Aspekte von „Limited-area“-Ensembles

In der globalen Modellierung sind Ensemblesysteme seit 15 Jahren gut etabliert. Es entstand die Notwendigkeit, die Vorteile von Ensembles auch für die „Limited-area“-Modellierung zu erschließen. Zurzeit existieren etwa zehn

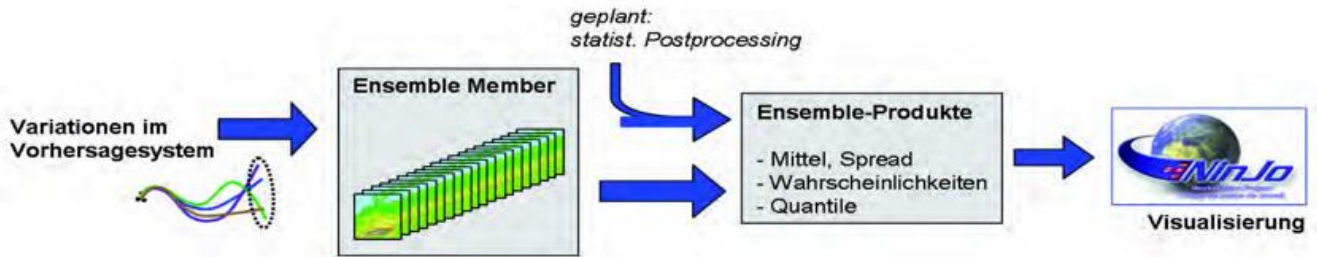


Abb. 6-1: Produktionsschritte im COSMO-DE-EPS: Generierung von Ensemblemitgliedern durch Variationen im Vorhersagesystem, Ableitung von Ensembleprodukten, statistisches Postprocessing durch logistische Regression, Visualisierung in NinJo.

operationelle und präoperationelle „Limited-area“-Ensemblesysteme in Europa. Sie beschränken sich mehrheitlich auf kurzfristige Vorhersagezeiten bis zu drei Tagen.

„Limited-area“-Modelle werden mit einer feineren räumlichen Maschenweite gerechnet und können auch kleinräumigere Prozesse explizit darstellen. Dies ist besonders wichtig für eine realistischere Darstellung der Topographie sowie von kleinräumigen atmosphärischen Prozessen wie zum Beispiel der Konvektion. Aufgrund der neuen Fokussierung auf bodennahe Variablen, konvektive Prozesse und kürzere Vorhersagezeiten wird die Nutzung von Vorhersageensembles für immer mehr Anwendungsgebiete relevant, zum Beispiel für die klassischen Unwetterwarnungen und Hochwasserwarnungen, für Vorhersagen für Luftfahrt und Windenergie.

Der Trend geht zu immer feineren Maschenweiten. Zurzeit werden „Limited-area“-Ensembles mit Gittermaschenweiten von 5 bis 25 km gerechnet. Gegenstand der Forschung und Entwicklung sind die konvektionserlaubenden Ensembles mit einer Gittermaschenweite von 1-3 km. In konvektionserlaubenden Modellen ist sogar der Lebenszyklus einzelner konvektiver Zellen verfolgbar, während in gröbermaschigen Modellen lediglich der Nettoeffekt parametrisiert wird. Allein durch die explizite Simulation dieser Lebenszyklen entsteht ein realistischeres Bild in der Vorhersage. Es ist zum Beispiel zu erwarten, dass warnrelevante Extremniederschläge besser vorhergesagt werden.

Es besteht jedoch nicht der Anspruch, die Lebenszyklen im Einzelnen deterministisch richtig vorherzusagen. Viele kleinräumige Prozesse sind schnelllebig und stark nichtlinear. Das Fehlerwachstum ist noch schneller als bei großräumigeren Prozessen. Somit ist die Vorhersagbarkeit stärker begrenzt. Insbesondere für die konvektionserlaubenden Modelle ist oftmals eine probabilistische Herangehensweise notwendig, um den Vorteil dieser Modelle überhaupt sichtbar zu machen. Genau diese Herangehensweise geschieht über die Ensembletechnik.

Aufgrund der unterschiedlichen Charakteristiken und Zielsetzungen der konvektionserlaubenden Modelle und Vorhersagen müssen für eine erfolgreiche Generierung entsprechender Ensemblesysteme andere Anforderungen erfüllt werden als bei Ensembles, die auf globalen Modellen basieren.

- Beispielsweise ist die „tangent-linear approximation“ nicht mehr gültig, die in der globalen Ensemblegenerierung ein Schlüssel zum Erfolg war. Die Näherung

ist nicht mehr erfüllt, weil nun schnelllebigere, nicht-lineare Prozesse explizit beteiligt sind.

- Auch mag die Unsicherheit in der „Modellphysik“ neue Ansätze benötigen, weil jetzt andere Zielgrößen, Prozesse und Skalen im Vordergrund stehen.
- Weiterhin muss das „Limited-area“-Ensemble berücksichtigen, dass die Vorhersage am Rand des Modellgebietes ebenfalls unsicher ist. Dies erfordert die Kopplung an ein globales Ensemblesystem.
- Die Unsicherheiten im Anfangszustand nehmen in „Limited-area“-Ensembles eine andere Rolle ein. Variationen der Anfangsbedingungen verlieren üblicherweise nach einigen Stunden ihren Effekt, weil sie aus dem Modellgebiet hinausgetragen werden und der Einfluss der seitlichen Randbedingungen das gesamte Modellgebiet erfasst.
- Weil das „Limited-area“-Ensemble auf eher kurze Vorhersagezeiten fokussiert, muss das Fehlerwachstum schon in den ersten Vorhersagestunden repräsentiert sein. Dies gilt auch für das antreibende globale Ensemblesystem.

Auch die Produkterzeugung steht vor neuen Herausforderungen. So ist es zwar erstrebenswert, das Ensemble mit einer möglichst feinen räumlichen Auflösung zu rechnen, um die relevanten Prozesse überhaupt darstellen zu können. Jedoch ist es für manche Anwendungen nicht erstrebenswert, die resultierenden Wahrscheinlichkeiten für eine möglichst genaue räumlich-zeitliche Auflösung anzubieten. Aussagekräftiger mögen Wahrscheinlichkeiten sein, dass es irgendwann in einem bestimmten Gebiet regnet und irgendwann innerhalb eines bestimmten Zeitraumes. Diese Darstellung filtert unvermeidbare Vorhersageunsicherheiten in der exakten Position und im exakten Zeitpunkt. Sie kann insbesondere bei Niederschlag zu einer verbesserten Übereinstimmung mit Beobachtungen und somit zu einer fundierteren Vorhersage führen. Die Vorteile des feinen Modellgitters bleiben dennoch erhalten, weil das Modell dadurch relevante Prozesse simulieren kann und einige warnwürdige Ereignisse in der Vorhersage überhaupt erst erfasst sind.

3 Der Aufbau von COSMO-DE-EPS

Im Folgenden wird der Aufbau von COSMO-DE-EPS erläutert. Beleuchtet werden die drei Produktionsschritte (vergleiche Abb. 6-1):

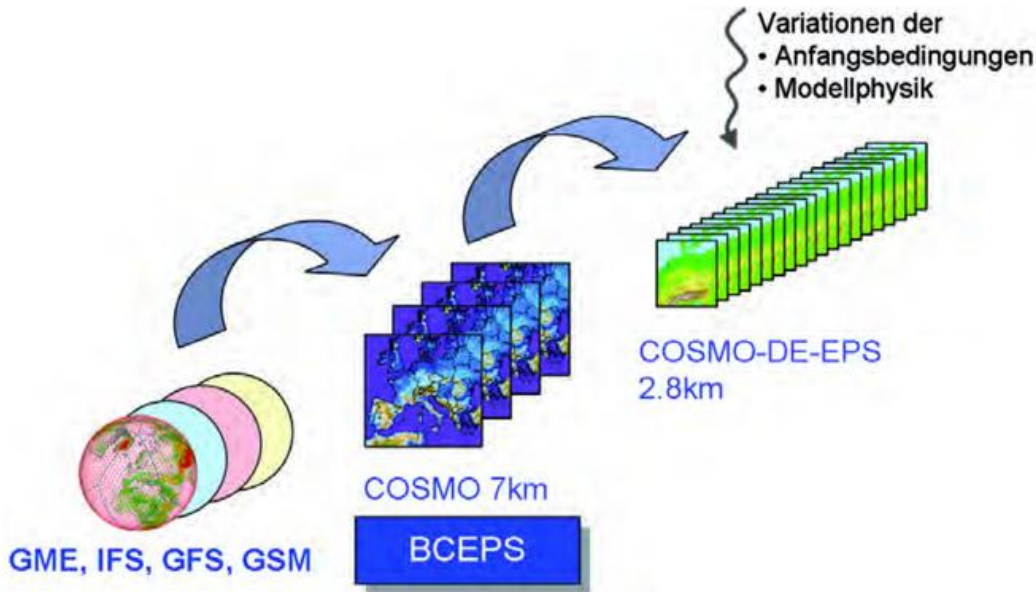


Abb. 6-2: Vorhersagekette, die dem COSMO-DE-EPS vorgeschaltet ist. Die Member des COSMO-DE-EPS werden in Vorhersagen des BICEPS („boundary condition ensemble prediction system“) genestet. Das BICEPS umfasst derzeit vier Member, die sich durch ihr jeweils antreibendes globales Modell unterscheiden: (a) $\sigma = 0,2$, (b) $\sigma = 0,05$.

- Generierung von Ensemblemember durch Variationen im Vorhersagesystem,
- Ableitung von Ensembleprodukten (zum Beispiel Ensemblemittel, Spread, Wahrscheinlichkeiten, Quantile),
- statistisches Postprocessing.

3.1 Generierung der Ensemblemember

Die derzeit implementierte Methode zur Generierung von Ensemble-Membern umfasst Variationen von

- Anfangsbedingungen,
- seitlichen Randbedingungen,
- und Modellphysik.

Zurzeit werden Methoden eingesetzt, die dem Ansatz des „Multi-Modell“-Ensembles ähneln. Für die Variation der Anfangsbedingungen und der seitlichen Ränder werden Vorhersagen verschiedener Globalmodelle verwendet. Für die Variation der Modellphysik werden verschiedene Konfigurationen des COSMO-DE eingesetzt.

Die verschiedenen Globalmodelle werden über eine Kette von Ensembles eingebunden (Abb. 6-2).

Die Member des COSMO-DE-EPS werden in verschiedene Vorhersagen des COSMO Modells mit 7 km Gittermaschenweite genestet. Diese COSMO Vorhersagen bilden zusammen das Randdatenensemble BICEPS („boundary condition ensemble prediction system“). Das BICEPS umfasst derzeit vier Member, die sich durch ihr jeweils antreibendes globales Modell unterscheiden. Folgende vier globale Modelle sind beteiligt: das globale Modell GME vom DWD, das GFS vom NCEP, das IFS vom ECMWF und das GSM von der JMA.

Dieses Multi-Modell und Multi-Analysis-Ensemble hat vor allem den Vorteil, dass sich die einzelnen Vorhersagen bereits in der Kurzzeit voneinander unterscheiden.

Bei der Variation der Anfangsbedingungen ist es wünschenswert, Informationen der originalen COSMO-DE-Anfangsbedingungen teilweise zu erhalten, weil die COSMO-DE Analyse das wertvolle „Latent Heat Nudging“ (STEPHAN et al. 2008) mit Hilfe der Radarmessungen enthält. Daher wird die Information, die

von den verschiedenen COSMO 7 km Vorhersagen kommt, nicht einfach übernommen und auf das COSMO-DE Gitter interpoliert (PERALTA et al. 2012). Die ursprüngliche Anfangsbedingung des COSMO-DE wird in den Ensemble Membern lediglich modifiziert mit Hilfe der Differenzen zwischen den verschiedenen COSMO 7 km Vorhersagen im BICEPS und dem operationellen COSMO-EU.

Für die Variation der Modellphysik wurden verschiedene Modellkonfigurationen des COSMO-DE konzipiert und getestet (GEBHARDT et al. 2011). In den verschiedenen Konfigurationen wird der Wert jeweils eines Parameters geändert, der im Zusammenhang mit einer Parametrisierung steht, zum Beispiel Turbulenz, Mikrophysik, flache Konvektion. Der geänderte Wert wird während der gesamten Simulation konstant gehalten. Somit verfolgt die Variationsmethode eher den Ansatz der Multi-Konfiguration und nicht den stochastischen Ansatz. Zur Auswahl einer neuen Modellkonfiguration wurde vorab in Experimenten geprüft,

- ob das Verfahren in der Lage ist, andere Vorhersagen zu erzeugen als die Standardkonfiguration,

Nr.	geänderter Parameter	mit Bezug zur Parametrisierung	Standardwert	geänderter Wert
1	entr_sc	flache Konvektion	0,0003 m ⁻¹	0,002 m ⁻¹
2	q_crit	Wolkenmikrophysik	1,60	4,0
3	rlam_heat	Grenzschicht	1,0	0,1
4	rlam_heat	Grenzschicht	1,0	10,0
5	tur_len	Turbulenz	150,0 m	500,0 m

Tab. 6-1: Die fünf verschiedenen COSMO-DE-Modellkonfigurationen im COSMO-DE-EPS. In den verschiedenen Konfigurationen wird der Wert jeweils eines Parameters geändert, der im Zusammenhang mit einer Parametrisierung steht.

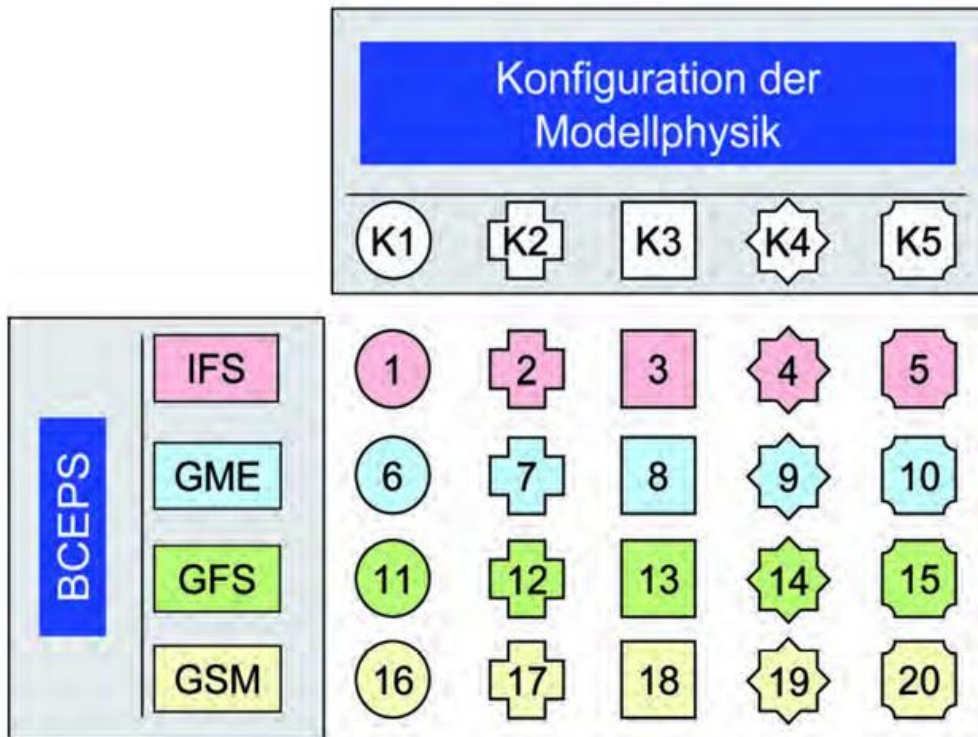


Abb. 6-3: Konstruktion der 20 Member im COSMO-DE-EPS. Die fünf Symbole (Spalten) stehen für die fünf verschiedenen COSMO-DE-Modellkonfigurationen (vergleiche Tab. 6-1). Die vier Farben (Zeilen) stehen für die vier Member des BCEPS und sind jeweils durch ihr entsprechendes Globalmodell gekennzeichnet (vergleiche Abb. 6-2). Die 20 Member im COSMO-DE-EPS entstehen durch die Kombination der fünf Modellkonfigurationen mit den vier verschiedenen Randantrieben durch BCEPS. Sie sind gemäß der Zahlen in den farbigen Symbolen nummeriert. Die Variation der Anfangsbedingungen ist ebenfalls enthalten. In einem bestimmten COSMO-DE-EPS-Member wird immer genau das BCEPS-Member zur Modifizierung des Anfangszustands herangezogen, welches auch die seitlichen Randbedingungen bereitstellt.

- und ob der Qualitätsunterschied zwischen den jeweiligen Vorhersagen gering ist, betrachtet über eine Vielzahl von Fällen.

Zurzeit bestehen fünf verschiedene Modellkonfigurationen (Tab. 6-1). Kombiniert mit den vier verschiedenen Membern im Randdatenensemble BCEPS ergibt dies eine Gesamtanzahl von $4 \times 5 = 20$ Ensemblemember im COSMO-DE-EPS (Abb. 6-3). Die Variation der Anfangsbedingungen ist in diesem Ensembleaufbau ebenfalls enthalten. In einem bestimmten COSMO-DE-EPS-Member wird immer genau das BCEPS-Member zur Modifizierung des Anfangszustands herangezogen, welches auch die seitlichen Randbedingungen bereitstellt.

Für die kommenden Jahre ist geplant, den Ensembleaufbau zu überarbeiten. In Zukunft wird geprüft, ob sich der Ensemble Kalman Filter auch zur Variation der Anfangsbedingungen eignet. Derzeit ist der Filter noch für das COSMO-DE-Assimilationssystem in Entwicklung. Weiterhin bestehen Pläne zu Erweiterungen in der Ensemblegenerierung, zum Beispiel neue Variationen in der Modellphysik und das Hinzufügen von Ensemble-Variationen in der Bodenfeuchte. Außerdem ist geplant, die Bedingungen und Variationen am seitlichen Modellrand des COSMO-DE durch ein globales Ensemble vorzugeben, welches auf dem Modell ICON basiert. Das Global-

modell ICON ist derzeit in Entwicklung.

3.2 Generierung von Ensembleprodukten

In der Produktgenerierung sind verschiedene Methoden denkbar. Im einfachsten Falle werden die Produkte mit der simplen „Democratic voting“-Methode geschätzt, das heißt jedes Ensemble Member erhält gleiche Gewichtung und es fließen auch keine sonstigen Informationen oder Erfahrungen ein. Diese Methode wird derzeit operationell angewendet.

Verbesserungen sind jedoch denkbar, zum Beispiel über den Einbezug aufeinanderfolgender Produktionszyklen. Ein voller Satz von 20 Ensemble Membern wird alle drei Stunden gestartet für einen Vorhersagezeitraum von jeweils 21 Stunden. Dadurch überlappen die Vorhersagezeiträume, die von aufeinanderfolgenden Pro-

duktionszyklen erzeugt sind. Das schafft die Möglichkeit, Ensemble-Member aus verschiedenen Produktionszyklen zu mischen und in die Generierung von Ensembleprodukten einzubeziehen. Somit kann die statistische Stichprobe von 20 Ensemble-Membern zum Beispiel auf die doppelte oder dreifache Anzahl erhöht werden. Jüngste Untersuchungen mit dem COSMO-DE-EPS zeigen, dass die Vermischung von Membern aus verschiedenen Produktionszyklen tatsächlich von Vorteil ist.

Eine weitere einfache Methode zur Erhöhung der Stichprobe ist der Einbezug einer räumlich-zeitlichen Umgebung. Hier besteht die Annahme, dass sich Vorhersagen innerhalb einer bestimmten Umgebung nur zufällig unterscheiden und somit gegeneinander austauschbar seien. Selbstverständlich kann diese Annahme problematisch sein, zum Beispiel wenn wichtige, topografisch bedingte Strukturen vernachlässigt werden. Somit ist diese Methode vor allem für die 2m-Temperatur nicht ohne weiteres anwendbar. Jüngste Untersuchungen zum COSMO-DE-EPS zeigen jedoch, dass sie für Niederschlag durchaus sinnvoll ist.

Weiterhin können auch Erfahrungen aus vergangenen Vorhersagefehlern für die Produktgenerierung ausgenutzt werden. Hier ist die Methode des statistischen Post-

processing zu nennen, die im Projekt COSMO-DE-EPS ebenfalls untersucht wird (nächster Abschnitt).

3.3 Statistisches Postprocessing

Zum statistischen Postprocessing der COSMO-DE-Ensemblevorhersagen entwickelt das Projekt COSMO-DE-EPS eine logistische Regression. Sie kann zur Berechnung von Überschreitungswahrscheinlichkeiten spezifischer Niederschlagsmengen angewendet werden. Erste Verifikationsergebnisse zeigen, dass die Wahrscheinlichkeiten des COSMO-DE-EPS vom Postprocessing profitieren (siehe Abschnitt 4.2 „Verifikation“). Eine Implementierung des Verfahrens ist geplant.

Weiterhin finanziert der DWD zwei Forschungsprojekte zum Postprocessing an Universitäten. Die Projekte sind jeweils an der Universität Bonn (P. Friederichs) und der Universität Heidelberg (T. Gneiting) angesiedelt. Verschiedene Ansätze werden angewendet wie zum Beispiel Quantilregression, Extremwertstatistik und „Bayesian Model Averaging“ inklusive Berücksichtigung räumlicher Korrelationen.

4 Einführung des COSMO-DE-EPS in den operativen Betrieb

4.1 Bereitstellung

Seit Dezember 2010 produziert das COSMO-DE-EPS am DWD Vorhersagen rund um die Uhr im zeitkritischen Modus, zunächst im präoperationellen und seit dem 22. Mai 2012 im operationellen Betrieb. Der volle Satz von 20 Mitgliedern wird alle drei Stunden mit aktualisiertem Anfangszustand gestartet (00 UTC, 03 UTC, 06 UTC, ...) und berechnet jeweils eine Ensemblevorhersage für die folgenden 21 Stunden. Etwa zwei Stunden nach Starttermin sind die Ensemblevorhersage und die Ensembleprodukte fertiggestellt und stehen den Forecastern am DWD zur Verfügung.

Die Bereitstellung der Ensembleprodukte erfolgt über das Visualisierungssystem NinJo, welches der Standard für die DWD Vorhersage- und Beratungszentralen ist. Somit kann das COSMO-DE-EPS jetzt unmittelbar in die Praxis einfließen. Hierfür hat das Projekt

COSMO-DE-EPS im NinJo-System eigens das neue „Ensemble Layer“ entworfen und implementiert.

Folgende Ensembleprodukte stehen zur Verfügung:

- „Ensemble Mean“ und „Spread“,
- Quantile (10%, 25%, 50%, 75%, 90%),
- Wahrscheinlichkeiten für eine Vielzahl verschiedener Ereignisse, zum Beispiel Niederschlag in einer Stunde > 10 mm.

Wahrscheinlichkeiten werden im Wesentlichen für warnwürdige Ereignisse ausgegeben, passend zu den standardisierten Warnkriterien am DWD.

Die Variablenauswahl umfasst zunächst:

- Niederschlag (mehrere Akkumulationszeiträume),
- 2m-Temperatur (Momentanwert, zeitliches Minimum und Maximum),
- 10m-Windböen,
- Schneehöhe (Angabe als Wasseräquivalent für mehrere Akkumulationszeiträume).

Die Ensembleprodukte liegen auf dem Modellgitter des COSMO-DE vor (2,8 km). Zusätzlich wird die Wahrscheinlichkeit ausgegeben, dass ein Niederschlagsereignis irgendwo innerhalb einer Fläche von 10 x 10 Gitterboxen auftritt.

4.2 Verifikation

Die Verifikation gegen Beobachtungen liefert eine Einschätzung von derzeitigen Qualitätseigenschaften des COSMO-DE-EPS. Bei COSMO-DE-EPS liegt das Hauptaugenmerk auf der Niederschlagsvorhersage. Es wird untersucht,

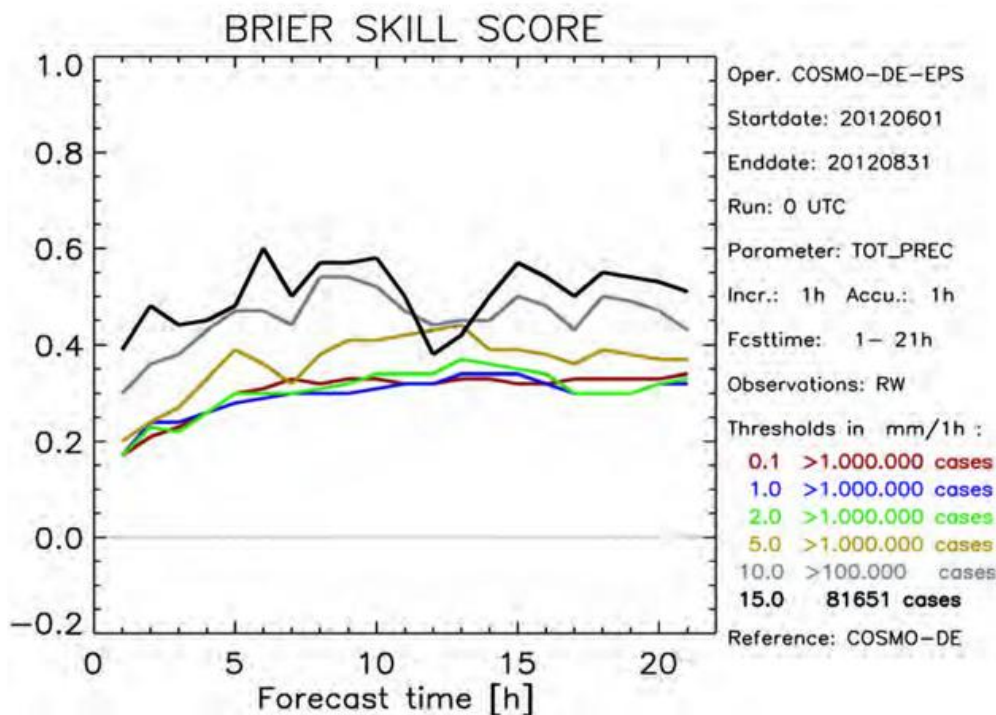


Abb. 6-4: „Brier Skill Score“ des COSMO-DE-EPS für einstündigen Niederschlag als Funktion der Vorhersagezeit („forecast time“ auf x-Achse) und des Schwellenwertes („threshold“ in verschiedenen Farben) als Mittel über alle Vorhersagen mit Startzeit 00 UTC aus den Monaten Juni bis August 2012. Die Referenzvorhersage für die Bewertung des Skills ist der deterministische COSMO-DE-Lauf.

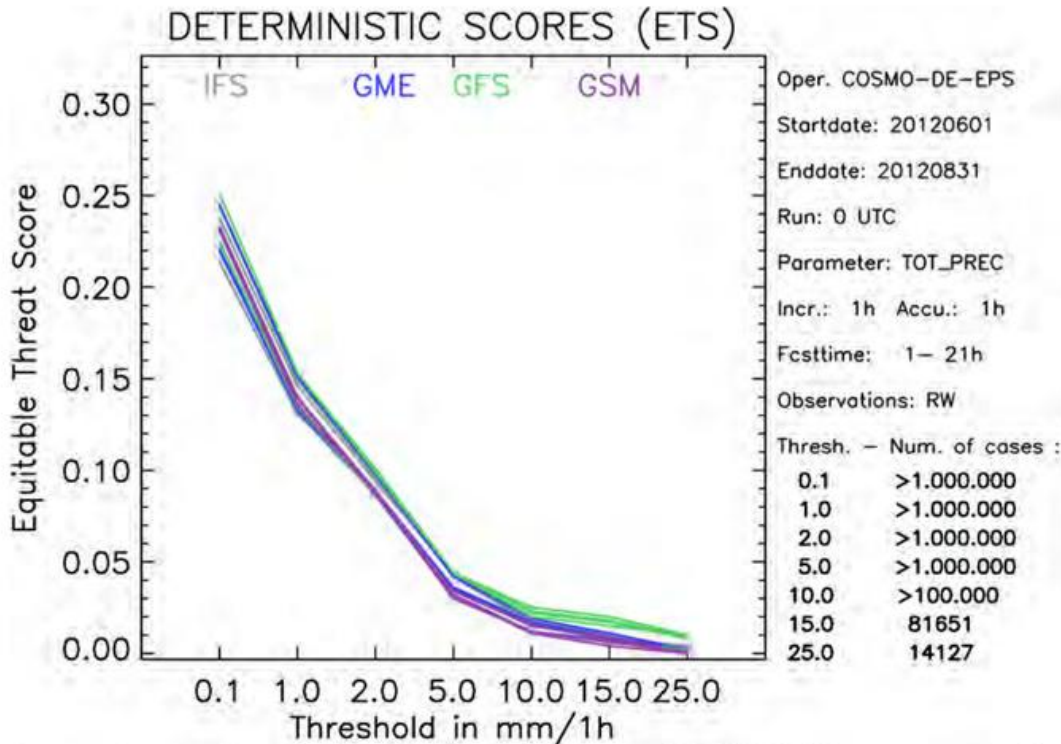


Abb. 6-5: Equitable Threat Score der 20 einzelnen Member des COSMO-DE-EPS für einstündigen Niederschlag als Funktion des Schwellenwertes („threshold“) als Mittel über alle Vorhersagen mit Startzeit 00UTC aus den Monaten Juni bis August 2012. Die unterschiedlichen Farben kennzeichnen das jeweilige Globalmodell, das der Rand- und Anfangsstörung eines Members zugrunde liegt.

ob das Ensemblesystem fundamentale Qualitätsanforderungen erfüllt und einen Zusatznutzen und Qualitätsgewinn gegenüber der rein deterministischen COSMO-DE-Vorhersage liefert. Hierfür werden sowohl probabilistische Gütemaße eingesetzt als auch eine deterministische Qualitätsprüfung der einzelnen Member. Die Beobachtungsgrundlage für die im Folgenden gezeigten Ergebnisse bilden Radardaten, die an Beobachtungen aus Regentöpfen angeeicht wurden. Diese Daten haben den Vorteil, dass sie über die Radarbeobachtungen die räumliche Struktur des Niederschlags auf den relevanten Skalen erfassen und die Niederschlagsmenge über die direkten Niederschlagsmessungen kontrolliert wird. Die Verifikationsergebnisse werden mit dem Programm PACprove berechnet, das ebenfalls innerhalb des Projektes COSMO-DE-EPS entwickelt wurde.

Eine Zusammenfassung der bisherigen Verifikationsergebnisse für Niederschlag:

- Bezüglich des klassischen Gütemaßes „Brier Score“ weisen die ensemblebasierten Wahrscheinlichkeiten eine

deutlich bessere Qualität auf als die deterministischen Ja/Nein-Vorhersagen des operationellen COSMO-DE. Dies gilt für alle untersuchten Akkumulationszeiträume, Vorhersagezeiten und Niederschlagsintensitäten. Der Vergleich des Brier Scores aus dem COSMO-DE-EPS mit dem Brier Score der deterministischen Vorhersage lässt sich über den Brier Skill Score (BSS) quantifizieren. Beispielhaft zeigt Abb. 6-4 den BSS für einstündigen Niederschlag als Funktion der Vorhersagezeit („forecast time“) und des Schwellenwertes („threshold“) für das betrachtete Ereignis

(zum Beispiel Niederschlag über 1 mm/h) als Mittel über alle Vorhersagen mit Startzeit 00 UTC aus den Monaten Juni bis August 2012. Positive Werte des BSS bedeuten hier, dass das COSMO-DE-EPS Vorteile gegenüber der deterministischen Vorhersage aufweist. Es zeigt sich die generelle Tendenz einer Zunahme des BSS mit der Vorhersagezeit und mit der Höhe des Schwellenwertes. Dies spiegelt die abnehmende deterministische

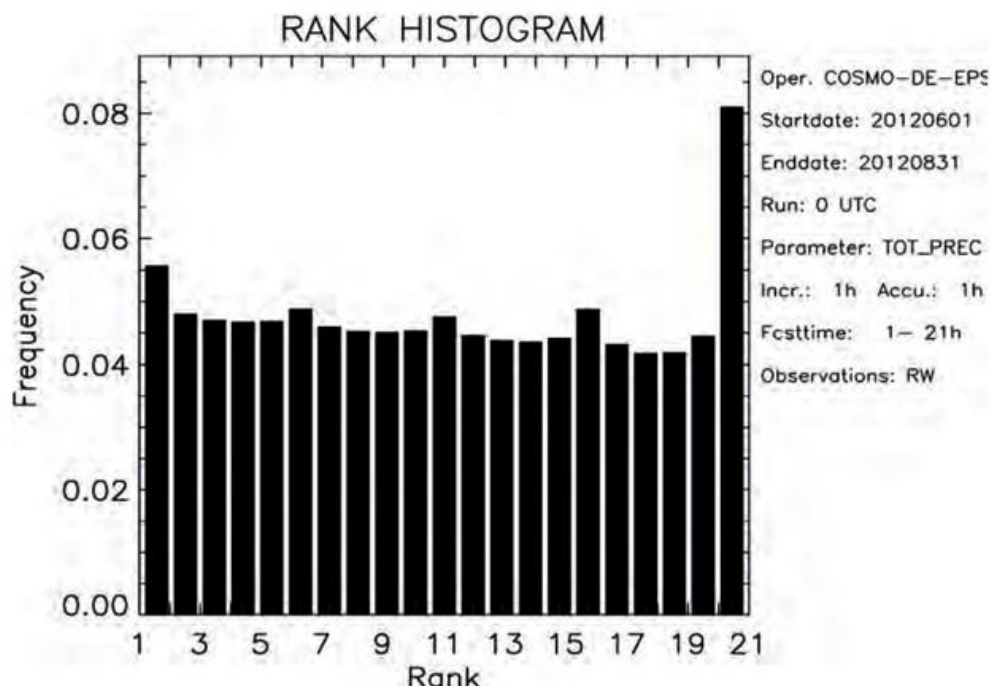


Abb. 6-6: Rank-Histogramm des COSMO-DE-EPS für einstündigen Niederschlag bestimmt über alle Vorhersagen mit Startzeit 00 UTC aus den Monaten Juni bis August 2012.

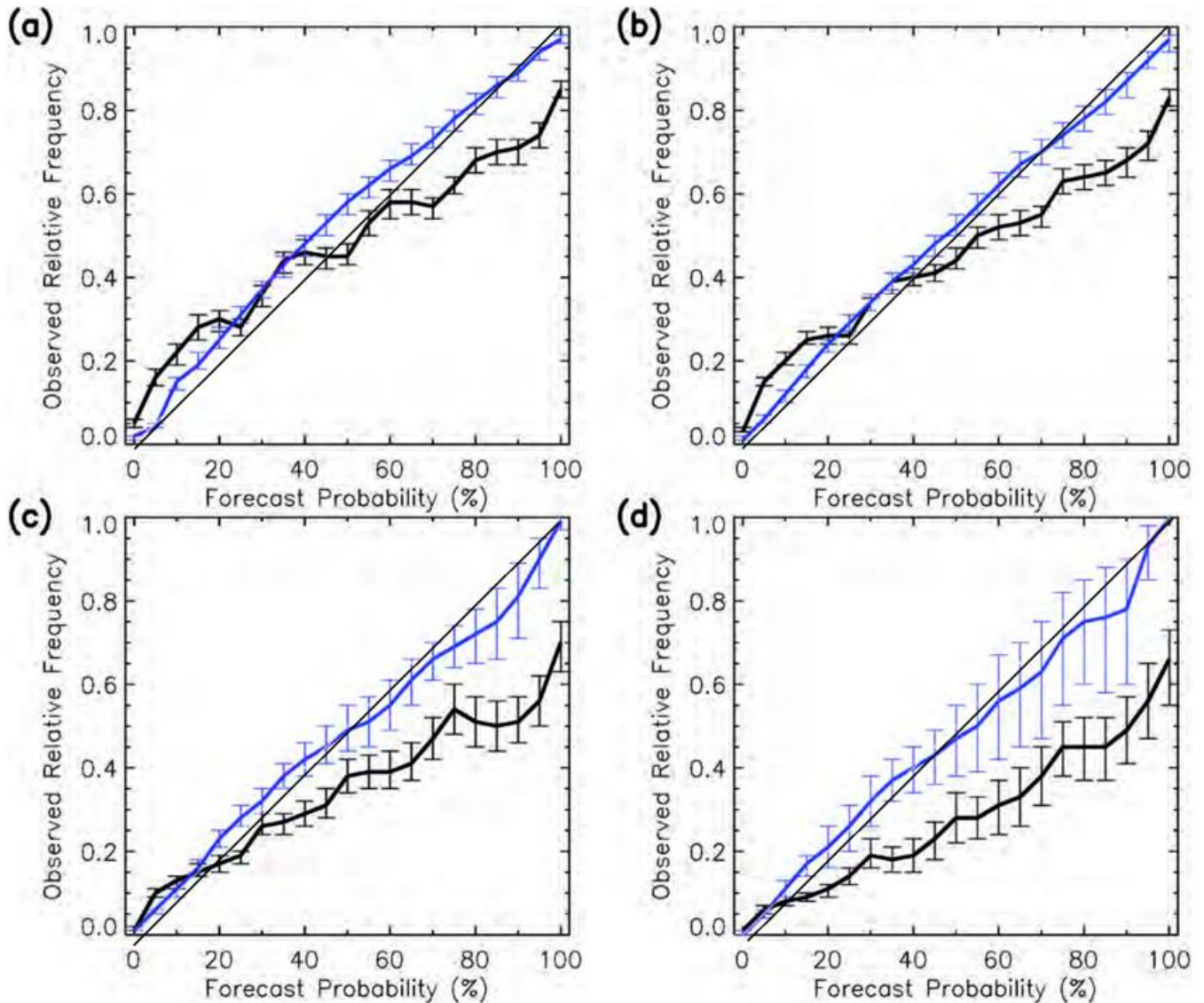


Abb. 6-7: Reliability-Diagramme für sechsstündigen Niederschlag und die Schwellenwerte a) 0,1mm/6h, b) 1mm/6h, c) 5mm/6h und d) 10mm/6h für das unprozessierte (schwarz) und das kalibrierte (blau) COSMO-DE-EPS. Die x-Achsen kennzeichnen die vorhergesagten Überschreitungswahrscheinlichkeiten, gegen die die bedingten Häufigkeiten aus den Beobachtungen aufgetragen werden (y-Achsen). Die Fehlerbalken sind Ergebnis eines Bootstrapping-Ansatzes. Die Datengrundlage stammt aus dem Sommer 2011.

Vorhersagbarkeit und damit den zunehmenden Nutzen des Ensembles mit zunehmendem Vorhersagehorizont und zunehmender Intensität des Ereignisses wider. Gleichzeitig lässt sich feststellen, dass sich mit steigendem Schwellenwert auch die Schwankungen der dargestellten Kurven mit der Vorhersagezeit verstärken. Dies ist korreliert mit einer Abnahme der Fälle („cases“), die in die Statistik eingehen. Daran lässt sich gut erkennen, dass robuste statistische Aussagen über seltener auftretende Ereignisse längere Perioden als Datengrundlage benötigen.

- Über eine Vielzahl von Fällen gesehen haben die einzelnen Member im Durchschnitt eine vergleichbare Qualität. Somit können die Member des COSMO-DE-EPS als gleichwahrscheinlich behandelt werden. Dies ist eine wichtige Feststellung im Hinblick auf die derzeitige Generierungsmethodik des COSMO-DE-EPS, die sich auf „Multi-Modell“ und „Multi-Konfiguration“ stützt. Die vergleichbare Qualität der einzelnen Member wird durch die Be-

trachtung verschiedener deterministischer Scores überprüft. Exemplarisch ist in Abb. 6-5 der Equitable Threat Score (ETS) aller 20 einzelnen Member für einständigen Niederschlag des 00 UTC-Laufes gemittelt über alle Vorhersagezeiten und die Monate Juni bis August 2012 als Funktion des Schwellenwertes dargestellt. Der ETS bewertet die Trefferquote einer Vorhersage für bestimmte Ereignisse unter Berücksichtigung falscher Vorhersagen und zufälliger Treffer (siehe DAMRATH 2002). Höhere Werte des ETS bedeuten einen höheren Skill. Außer der Erkenntnis, dass alle ETS-Werte positiv sind und die einzelnen Member damit grundsätzlich Skill aufweisen, soll hier nicht näher auf die absoluten Werte des ETS eingegangen werden. Entscheidend ist hier, dass alle Member eine ähnliche Qualität besitzen und in diesem Sinne als über viele Fälle gleichwahrscheinlich betrachtet werden können. Oder mit anderen Worten: Es gibt im COSMO-DE-EPS kein Member, das als systematischer Ausreißer aus der Gesamtheit

der Member betrachtet werden muss. Lediglich eine leichte Gruppierung der Member nach den in BCEPS genutzten Globalmodellen lässt sich erkennen. Eine eventuelle systematische, alle Member betreffende Verbesserung des ETS wäre im Rahmen einer Verbesserung des zugrundeliegenden Modells COSMO-DE durchzuführen.

- Das COSMO-DE-EPS bildet bereits einen großen Anteil des gesamten Vorhersagefehlers ab. Das Ensemble steht mit seinem vergleichsweise hohen Spread im internationalen Vergleich sehr gut da. Eine umfassende Erfassung des Vorhersagefehlers wäre natürlich wünschenswert, ist aber derzeit nicht zu erwarten, weil auch der internationale Forschungsstand diesem Anspruch bei bodennahen Variablen bisher nicht gerecht wird. Dies muss bei der Nutzung der COSMO-DE-EPS Vorhersagen berücksichtigt werden. Abb. 6-6 zeigt das Rank-Histogramm für den einstündigen Niederschlag aller 00 UTC-Läufe des COSMO-DE-EPS aus den Monaten Juni bis August 2012 unter Berücksichtigung aller Vorhersagestunden. Ein Rank-Histogramm beschreibt, wie gut die Vorhersagevariabilität eines Ensemblesystems die tatsächliche beobachtete Variabilität der betrachteten Größe erfasst (siehe zum Beispiel HAMILL 2001). Wenn die vorhergesagte Variabilität die beobachtete korrekt wiedergibt, dann sind **alle Balken des Histogramms gleich hoch** („flaches Histogramm“). In der Abb. 6-6 sind die beiden äußersten Balken höher als die anderen, was sich aus einer Unterschätzung der Variabilität durch das COSMO-DE-EPS ergibt („U-förmiges Histogramm“). Gleichzeitig wird durch die Asymmetrie der Balken ein Bias im System deutlich. Die leichte Erhöhung der Balken 6, 11 und 16 ist wiederum auf die Tendenz zur Gruppierung der Niederschlagsvorhersagen der einzelnen Member gemäß der in BCEPS benutzten Globalmodelle zurückzuführen. Wenn für jede Vorhersagestunde das Rank-Histogramm einzeln untersucht wird, dann wird das Histogramm ungefähr ab Vorhersagestunde 13 **flach, abgesehen von der Wirkung des Bias auf den äußersten rechten Balken** (ohne Abbildung). Das heißt, dass mit zunehmender Vorhersagezeit der Vorhersagefehler immer besser erfasst wird.
- In Fällen mit niedrig vorhergesagten Wahrscheinlichkeiten tritt das Ereignis tatsächlich seltener auf als bei hohen vorhergesagten Wahrscheinlichkeiten. Somit weist das COSMO-DE-EPS bereits gute Ergebnisse in „Reliability“ und „Resolution“ auf. Abb. 6-7 zeigt Reliability-Diagramme des COSMO-DE-EPS für die Wahrscheinlichkeit, dass sechsstündige Niederschlagssummen bestimmte Schwellenwerte überschreiten. In solch einem Diagramm werden vorhergesagte Überschreitungswahrscheinlichkeiten mit der beobachteten **Häufigkeit des Ereignisses verglichen**. Im Detail wird dabei überprüft, wie oft in genau den Fällen, in denen eine bestimmte Überschreitungswahrscheinlichkeit A vorhergesagt wurde, das Ereignis tatsächlich beobachtet wurde („bedingte Häufigkeit“). Diese Betrachtung wird über eine ausreichende Anzahl von Vorhersagen

und für verschiedene Werte von A durchgeführt und in das Diagramm eingetragen. Idealerweise liegen die Ergebnisse auf der Diagonalen, weil dann vorhergesagte **Wahrscheinlichkeiten und bedingte Häufigkeiten** übereinstimmen. In Abb. 6-7 sind als schwarze Linien die Ergebnisse des nicht weiter prozessierten COSMO-DE-EPS für den Sommer 2011 dargestellt. Hier zeigt sich eine Abweichung von der Diagonalen, die mit dem untersuchten Schwellenwert zunimmt und besonders für hohe Schwellenwerte eine **Überschätzung der Häufigkeit des Ereignisses aufzeigt**.

- Die in Abb. 6-7 erkennbaren Defizite in der Reliability lassen sich durch eine Nachprozessierung („postprocessing“) oder auch Kalibrierung der vorhergesagten Wahrscheinlichkeiten deutlich mildern. Hierbei wird die Vorhersagestatistik des COSMO-DE-EPS auf Basis eines systematischen Vergleichs mit Beobachtungen der Statistik dieser Beobachtungen angepasst. Die blauen Linien in Abb. 6-7 beruhen auf aktuellen Entwicklungsarbeiten im DWD zur Kalibrierung des COSMO-DE-EPS und zeigen das Potential einer solchen Methode. Die kalibrierten Wahrscheinlichkeiten liegen deutlich näher an der Diagonalen. Die Fehlerbalken an den Linien sind Ergebnisse eines Bootstrapping-Ansatzes zur Abschätzung der Robustheit der Resultate. Dass sich diese Fehlerbalken in den meisten Fällen nicht überlappen, **verdeutlicht die signifikante Verbesserung, die sich durch Kalibrierung erzielen lässt**. Die hier benutzte Kalibrierung beruht auf einer DWD-internen Weiterentwicklung eines Verfahrens von WILKS (2009) durch Z. Ben Bouallègue. Diese Entwicklungsarbeit wird mit dem Ziel der Implementierung eines Kalibrierungsverfahrens für COSMO-DE-EPS fortgeführt.

Zusammenfassend lässt sich für Niederschlag feststellen, dass das operationelle COSMO-DE-EPS die fundamentalen Qualitätsanforderungen gut erfüllt und gegenüber der Beschränkung auf den deterministischen Lauf des COSMO-DE eine Verbesserung darstellt. Weiterhin zeigt sich in der Kalibrierung der Ensemblevorhersagen ein deutliches Verbesserungspotential.

Für 2 m-Temperatur und 10 m-Windböen zeigt die Verifikation **andere Ergebnisse**. Bei diesen Variablen ist der Ensemble Spread viel zu gering. Es zeigt sich deutlich, dass das COSMO-DE-EPS zunächst im Hinblick auf die Niederschlagsvorhersage entwickelt worden ist. Zukünftige Weiterentwicklungen des COSMO-DE-EPS werden eine breitere Spanne von Variablen berücksichtigen.

4.3 Evaluierung durch Forecaster

In der Qualitätsbewertung ist die Evaluierung ein wichtiges Gegenstück zur Verifikation. Hier bewerten Forecaster, wie hilfreich das COSMO-DE-EPS in der Vorhersagepraxis ist. Ein Vorteil der Evaluierung besteht in der Auswahl von interessanten und relevanten Fällen.

Dadurch verliert die Untersuchung zwar an statistischer Belastbarkeit, aber gewinnt an Praxisrelevanz. Weiterhin stützt sich die Evaluierung nicht auf mathematische Gütemaße, sondern stellt die Frage, ob COSMO-DE-EPS dem Forecaster weiterhilft. Dabei fließt zum Beispiel der wichtige Aspekt ein, ob der Forecaster sich die Zusatzinformation des Ensembles genauso gut auch subjektiv hätte erschließen können. Außerdem bewertet die Evaluierung die Güte des COSMO-DE-EPS nicht isoliert, sondern in Zusammenwirken mit dem Forecaster, dem Visualisierungssystem, dem Produktangebot und allen anderen vorhandenen Informationen, zum Beispiel Nowcasting und andere Modellvorhersagen. Selbstverständlich beinhaltet die Evaluierung auch einen Lernprozess im Umgang mit den COSMO-DE-EPS-Vorhersagen. Die erfolgreiche Evaluierung während der präoperationalen Phase war ein wichtiges Kriterium auf dem Weg zur „operationellen Reife“ und für die Überführung des COSMO-DE-EPS in den operationellen Betrieb des DWD.

In der Evaluierung hat sich herausgestellt:

- Forecaster schauen bei konvektiven Lagen bisher meist auf das 90%-Quantil für Niederschlag sowie auf die Wahrscheinlichkeit, dass irgendwo innerhalb einer Fläche von 10x10-Gitterboxen eine Niederschlagsmenge überschritten wird.
- Forecaster schätzen die frühzeitigen Anzeichen für Starkniederschlag im COSMO-DE-EPS. Auch wird das COSMO-DE-EPS gerne als Indiz verwendet, ob das deterministische COSMO-DE möglicherweise einen falschen Alarm enthält.
- Forecaster kritisieren die sogenannte „Jumpiness“ zwischen den Niederschlagsvorhersagen aus aufeinanderfolgenden Produktionszyklen. Es tritt immer wieder der Fall auf, dass sich die Wahrscheinlichkeitsfelder für den gleichen Vorhersagezeitpunkt stark unterscheiden, je nach Startzeitpunkt der jeweiligen Ensemblevorhersage (00 UTC, 03 UTC, ...). **Jüngere Vorhersageläufe sind nicht zwangsläufig verlässlicher als ältere.** Ein weiterer Kritikpunkt ist der mangelhafte Ensemble Spread in 2m-Temperatur und 10m-Windböen.
- Forecaster lernen, mit geringen Wahrscheinlichkeiten umzugehen. Dabei geht es um die Frage, ob eine COSMO-DE-EPS Wahrscheinlichkeit von beispielsweise 10 % ausreichen würde, um eine Warnung auszusprechen. Weiterhin werden zahlreiche allgemeine Verständnisfragen zur probabilistischen Vorhersage aufgebracht. Nicht zuletzt dadurch wird klar, dass die Einführung von En-

semblevorhersagen in die Vorhersagepraxis durch Schulungen begleitet werden sollte.

5 Zusammenfassung

Mit dem COSMO-DE-EPS verfügt der DWD jetzt über ein Ensemblevorhersagesystem basierend auf dem Modell COSMO-DE. Europaweit ist es das erste konvektionserlaubende Ensemble im operationellen Betrieb. Sein Aufbau folgt dem Ansatz der Multi-Modell/Analysis und Multi-Konfiguration. Das Hauptaugenmerk von COSMO-DE-EPS liegt auf der Niederschlagsvorhersage. Verifikationsergebnisse zeigen, dass fundamentale Qualitätsanforderungen für Niederschlag erfüllt sind und das Ensemblesystem in der Vorhersagepraxis hilfreich ist. Die Vorhersagen des COSMO-DE-EPS stehen den DWD-Forecastern über das Visualisierungssystem NinJo zeitkritisch zur Verfügung.

Literatur

BALDAUF, M., A. SEIFERT, J. FÖRSTNER, D. MAJEWSKI, M. RASCHENDORFER, T. REINHARDT, 2011: Operational convective-scale numerical weather prediction with the COSMO model. *Mon. Weather Rev.* **139**, 3887-3905 doi:10.1175/MWR-D-10-05013.1.

DAMRATH, U., 2002: Verifikation von Wettervorhersagen. In: Die neue Modellkette des DWD II. *Promet* **28**, Nr. 1/2.

GEBHARDT, C., S. E. THEIS, M. PAULAT, Z. BEN BOUALLÈGUE, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research* **100**, 168-177.

HAMILL, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.* **129**, 550-560.

PERALTA, C., Z. BEN BOUALLÈGUE., S.E. THEIS, C. GEBHARDT, M. BUCHHOLD, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research* **117**, D07108, doi:10.1029/2011JD016581.

STEPHAN, K., S. KLINK, C. SCHRAFF, 2008: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quart. J. Roy. Met. Soc.* **134**, 1315-1326.

WILKS, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorol. Appl.* **16**, 361-368.

L. VON BREMEN, M. LANGE

7 Ensembleprognosen der Windleistung für Anwendungen in der Energiewirtschaft

On the use of ensemble wind power forecasts for the energy sector

Zusammenfassung

Seit Jahren sind Vorhersagen der zu erwartenden Windleistung für Netzbetreiber unverzichtbar geworden, um die rasant gestiegenen Anteile an Windstrom sicher in das Stromnetz zu integrieren. Die Kenntnis der Unsicherheit der Vorhersage wird als sehr wichtige zusätzliche Information betrachtet und kann in Zukunft aus Ensemble Vorhersagen abgeleitet werden. Die Bereitstellung von Windgeschwindigkeiten der Ensemblemodelle in 100 m Höhe steigert die Qualität von probabilistischen Windleistungsvorhersagen außerordentlich im Vergleich zu Windgeschwindigkeiten in 10 m Höhe. Die Einsatzmöglichkeiten der Ensemblevorhersage von Windleistung wird im vorliegenden Artikel besprochen.

Summary

Short-term wind power forecasts became an essential tool for grid operators to integrate increasing shares of wind power into the power supply system while keeping current security standards. The knowledge of forecast uncertainty is regarded as very useful information and will be provided by Ensemble forecasts in the near future. The introduction of ensemble winds in 100 m height increased the skill of probabilistic wind power forecasts considerably compared to 10 m wind speeds. The multiple use of ensemble wind power forecast for grid integration is discussed in this article.

1 Einführung

Erneuerbaren Energien (EE) kommen bezüglich der bereits eingeleiteten Klimaveränderung eine sehr wichtige Doppelrolle zu. Einerseits ist die rasche Verringerung des CO₂-Ausstoßes möglich und erstrebenswert und andererseits wird ein Weg eingeschlagen, um auf lange Sicht Energie und insbesondere zusätzliche Energie vor dem Hintergrund der Klimaanpassung zu geringeren Kosten erzeugen zu können. Die im Jahre 2008 verabschiedeten „20-20-20“-Ziele (20 % Reduktion der CO₂-Emissionen gegenüber 1990, 20 % Verringerung des Primärenergieverbrauchs und 20 % der EE am Bruttoendenergieverbrauch) im sogenannten „EU Climate and Energy Package“ für das Jahr 2020 sind allenfalls als Anfang einer europäischen Energiewende zu sehen.

Die notwendigen Veränderungen in der Energieerzeugung und in der Integration von variabler Wind- und Solarenergie in die Stromversorgungssysteme stellen große Herausforderungen dar. Der numerischen Wettervorhersage erwächst hier ein neues und sehr umfängliches Anwendungsgebiet, mit großer Bedeutung für Ökonomie und Versorgungssicherheit.

Für die Stabilität des Stromversorgungssystems muss gewährleistet sein, dass Stromproduktion und Stromverbrauch jederzeit balanciert sind. Deutschland ist an das Europäische Verbundsystem (ENTSO-E) angeschlossen, dessen primäre Aufgabe darin besteht, auftretende regionale oder nationale Ungleichgewichte zwischen Erzeugung und Verbrauch durch räumlichen Ausgleich auszubalancieren. Die Stromproduktion konventioneller Kraftwerke folgt einem festgelegten Fahrplan, um den prognostizierten Strombedarf zu decken. Verständlicherweise folgt die Stromerzeugung regenerativer Quellen wie Wind und Sonne keinem Fahrplan.

Vorhersagen der Windeinspeisung über die nächsten Stunden und Tage sind daher aus dem deutschen Energiemarkt nicht mehr wegzudenken. Bei mittlerweile 29,7 GW installierter Leistung (Stand Juni 2012 für On- und Off-shore) hat die wetterabhängige Windenergie einen massiven Einfluss auf die Zusammensetzung des Strommixes, auf die Preise an der Strombörse und die Steuerung der Stromnetze. Daher muss der Beitrag der Windenergie im Voraus so präzise wie möglich durch Vorhersagen bekannt sein, vor allem für den Intraday- und Day-Ahead-Bereich,

also bis 48 Stunden im Voraus. Windleistungsvorhersagen werden von mehreren Dienstleistungsunternehmen auf der Basis numerischer Wettermodelle erstellt und kommerziell angeboten.

Die deutschen Übertragungsnetzbetreiber, die bisher den größten Anteil der Windenergie im Rahmen des Erneuerbare-Energien-Gesetzes (EEG) am Strommarkt platziert haben, stellen die Windleistungsvorhersagen als Fahrplan an der Leipziger Strombörse (EEX) ein. Seit Anfang 2012 ist durch Einführung des Marktprämienmodells eine Vielzahl von Stromhändlern hinzugekommen, die ebenfalls Strom aus Windparks auf den Energiemärkten anbieten und daher ebenso auf präzise Windleistungsvorhersagen angewiesen sind. Die Vorhersagen sind damit der Schlüssel zur effektiven Marktintegration der Windenergie.

In den weitaus meisten Fällen kommen für die Fahrplananwendung im Rahmen des Energiehandels deterministische Vorhersagen zum Einsatz, die pro Zeitstempel genau einen Wert liefern. Ensemblevorhersagen sind hier nicht sehr verbreitet, da sie nicht unmittelbar als Fahrplan verwendet werden können. Dabei ist den Nutzern der Vorhersagen natürlich bekannt, dass Ensemblevorhersagen durch ihren probabilistischen Charakter mehr Informationen bieten. Aber bisher wird diese Zusatzinformation nicht in den operationellen Prozessen der meisten Händler und Übertragungsnetzbetreiber zur Bewirtschaftung der Bilanzkreise abgebildet und kann daher nicht adäquat verwendet werden.

2 Windleistungsvorhersage

Windleistungsprognosemodelle sind Mitte der 1990er Jahre in Dänemark (LANDBERG 1998) und Deutschland (BEYER et al. 1988) aus den Verfahren zur Berechnung lokaler Windgeschwindigkeiten zur Standortbeurteilung für Windkraftanlagen entstanden. Getragen wurde die Entwicklung zunächst von Universitäten und Forschungseinrichtungen. Die zunehmende Professionalisierung wurde in den letzten Jahren von Unternehmen geleistet. Im Laufe der letzten zehn Jahre hat sich eine Vielzahl von Vorhersagesystemen am Markt etabliert (GIEBEL et al. 2011).

In diesem Beitrag sollen die Methoden von Windleistungsvorhersagemodellen jedoch nicht im Vordergrund stehen, denn den Details der sich entwickelten Vielfalt wird ein eigener Beitrag in der geplanten Promet-Ausgabe „Energiemeteorologie“ gewidmet sein.

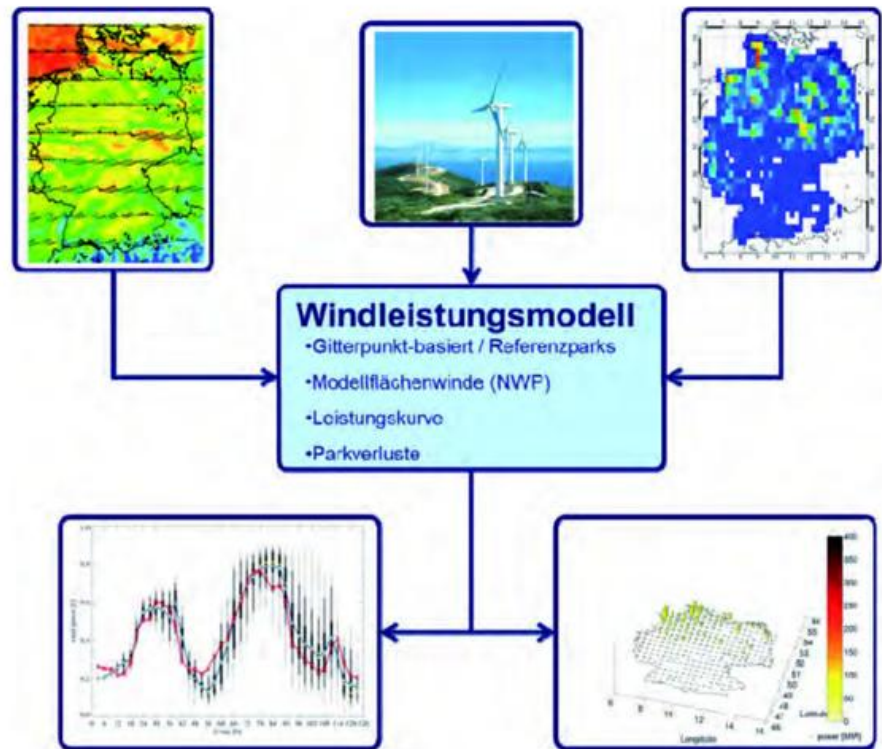


Abbildung 7-1: In das Windleistungsvorhersagemodell gehen Felder der vorhergesagten Windgeschwindigkeit (oben links) und Detailinformationen über die geographische Verteilung (oben rechts) der Windkraftanlagen und deren Charakteristiken (oben Mitte) ein. Die Ausgabe des Windleistungsvorhersagemodells ist eine aggregierte Ensemblevorhersage für Deutschland (unten links) und einzelne Gitterflächen (unten rechts).

Allen erfolgreichen Vorhersagesystemen für Windleistung ist gemeinsam, dass sie Daten aus numerischen Wettermodellen als Eingangsgrößen verwenden, um die zukünftigen Windbedingungen zu kennen. Meistens werden die Wettervorhersagen von etablierten Wetterdiensten bezogen und nur in wenigen Fällen betreiben die Anbieter der Windleistungsvorhersagen eigene Wettermodelle. Es hat sich als Stand der Technik durchgesetzt, aus mehreren verschiedenen Wettermodellen eine Windleistungsvorhersage zu berechnen, beispielsweise durch wetterlagenabhängige Gewichtung der numerischen Wettermodelle (LANGE et al. 2006). Das grundsätzliche Schema eines Windleistungsvorhersagesystems, das auf einem physikalischen Modellierungsansatz beruht, ist in Abbildung 7-1 zu sehen. Für eine regionale Vorhersage wie beispielsweise die Deutschlandvorhersage sind neben den Eingangsdaten der numerischen Wettermodelle Standortdaten (geographische Koordinaten, Typen der Windkraftanlagen, Nabenhöhen, Kennlinien, und so weiter) aller Windparks in Deutschland notwendig. Die meteorologischen Daten werden mit dem Windleistungsmodell in die zu erwartende Einspeisung der Windparks umgerechnet. Dabei besteht die Kunst vor allem darin, aus den geeigneten Variablen des numerischen Wettermodells die richtige Windgeschwindigkeit auf Nabenhöhe zu bestimmen.

In den Anfängen der Windleistungsvorhersage stand die Entwicklung von Vorhersageverfahren im Vordergrund, die als meteorologische Eingabe (das heißt Vorhersage) die

Windgeschwindigkeit in 10 m Höhe verwendeten. Rückblickend gesehen, gibt es zwei Erklärungen, wieso die Windgeschwindigkeit in solch einer geringen Höhe (10 m) verwendet wurde. Einerseits waren die Nabenhöhen der Windkraftanlagen zu Anfang der 1990er Jahre im Mittel nur 30 m und andererseits standen die Windgeschwindigkeiten auf den Modellflächen der Wettervorhersagemodelle den Entwicklern von Windleistungsprognosen nicht zur Verfügung. Mit zunehmenden Nabenhöhen jenseits von 30 m war klar, dass es nicht mehr ausreichend war, 10 m-Winde mit dem logarithmischen Windprofil unter Einbezug der Oberflächenrauigkeit z_0 und Annahme einer neutralen Schichtungsstabilität auf Nabenhöhe zu extrapolieren. In dieser Zeit wurden mit gutem Erfolg die Monin-Obukov-Erweiterung des vertikalen Windprofils bei nicht-neutralen Schichtung adaptiert, um so die vorhergesagten 10 m-Winde auf Nabenhöhe zu extrapolieren (LANGE und FOCKEN 2005). Auf diese Weise ist es möglich geworden, die sehr oft beobachtete abendliche Erhöhung der Windleistungsproduktion (durch Ausbildung einer stabilen thermischen Schichtung nach Sonnenuntergang) zu reproduzieren und vorherzusagen. Weiterhin ist es möglich, die, unter gewissen Wettersituationen einsetzende mittägliche Labilisierung der Atmosphäre durch Turbulenzproduktion und die damit verbundene Abschwächung der Windgeschwindigkeit auf Nabenhöhe korrekt vorherzusagen.

Heutzutage stehen die Windgeschwindigkeiten aller Modellflächen eines Wettervorhersagemodells zur Windleistungsvorhersage zur Verfügung, so dass die Windgeschwindigkeit in Nabenhöhe jeder Windkraftanlage direkt aus der Windgeschwindigkeit auf den benachbarten Modellflächen berechnet werden kann. Durch Wegfall der Unsicherheit in der Vorhersage der atmosphärischen Stabilität und somit der Windgeschwindigkeit auf Nabenhöhe hat sich die Genauigkeit der Windleistungsprognose erhöht.

Die im nächsten Abschnitt diskutierten Ergebnisse zeigen die Windleistungsvorhersage für Deutschland und die 50 Hertz-Regelzone mit den neuen 100 m-Winden (siehe Hauptabschnitt 3.1) des Ensemble Prediction Systems (EPS) des Europäischen Zentrums für Mittelfristige Wettervorhersage EZMW für den Zeitraum Februar 2010 bis April 2011. An jedem Modellpunkt ($0,25^\circ$ -Auflösung) wurde aus der 100 m- und 10 m-Windgeschwindigkeit mittels linearer Interpolation die Windgeschwindigkeit in der Höhe bestimmt, die der mittleren Nabenhöhe, der dort installierten Windkraftanlagen entspricht. Die mittlere (leistungsgewichtete) Nabenhöhe in Deutschland beträgt 88,2 m (Stand Dezember 2010), so dass weiterhin vereinfachend von 100 m-Winden gesprochen wird. Die installierte Windkraftleistung ist ebenfalls für jedem Modellpunkt bzw. jede Gitterfläche bekannt. Die Umrechnung von Windgeschwindigkeit in Windleistung erfolgt mit einer regionalen Leistungskurve (Mc LEAN 2008), die berücksichtigt, dass

- es sich um ein Portfolio von unterschiedlichen Anlagen handelt,
- Verluste durch Abschattungseffekte in Windparks und Netzverluste auftreten und

- Windkraftanlagen im Falle des Überschreitens der Abschaltwindgeschwindigkeit nicht alle gleichzeitig vom Netz gehen.

3 Analyse von Windenergie-Ensembleprognosen

3.1 100 m-EPS-Winde

Das EZMW archiviert nur die „Control Forecast“ mit dem gleichen Detaillierungsgrad wie die deterministische Vorhersage. Bei den Ensemblemitgliedern werden nur Einflächensfelder und Vorhersagen auf Druckflächen gespeichert, so dass Windgeschwindigkeiten in Nabenhöhe nicht errechnet werden konnten. Die Gründe sind im enormen Speichervolumen, die das Archivieren von Modellflächen aller Ensemblemitglieder darstellt, zu sehen. Am 26. Januar 2010 hat das EZMW ein 10 m-Windprodukt operationell eingeführt. Es geht zurück auf die Beteiligung des EZMWs im europäischen Forschungsprojekt „SafeWind“ (www.safewind.eu) und dem Betreiben der beteiligten Forscher, dass probabilistische Windleistungsprognosen deutlich verbessert werden können, sobald Windgeschwindigkeiten auf Nabenhöhe nicht mehr aus 10-m Winden (ohne verfügbare Stabilitätsinformation) berechnet werden müssen.

Abbildung 7-2 verdeutlicht, dass die Verwendung von 10 m-Winden zur deutschen Windleistungsvorhersage zu sehr großen systematischen Fehlern führt (grüne Linie), solange nicht eine nachgestellte Bias-Korrektur der Windleistungsvorhersage (rote durchgezogene Linie) durchgeführt wird. Allerdings kommen die Ergebnisse nach Bias-Korrektur bei weitem nicht an die wesentlich geringeren Vorhersagefehler bei Verwendung von 100 m-Winden

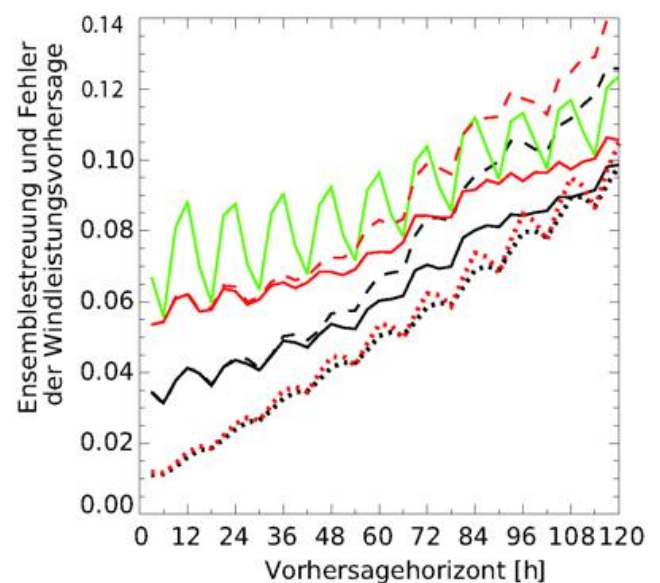


Abbildung 7-2: RMSE des Vorhersagefehlers der Windleistungsvorhersage für Deutschland (durchgezogen für Ensemblemittel und gestrichelt für deterministische Vorhersage) und Ensemble-Spread (gepunktet) gegen Vorhersagehorizont für 10 m-Winde (grün), 10 m-Winde einschließlich Bias-Korrektur (rot) und 100 m-Winde (schwarz).

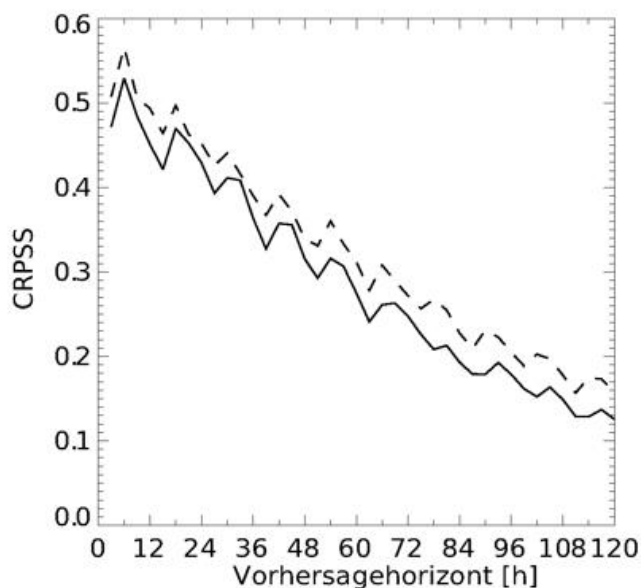


Abbildung 7-3: CRPSS der 100 m-Windleistungsvorhersage für Deutschland (durchgezogen) und die 50 Hertz-Regelzone (gestrichelt). Die Referenzvorhersage ist gerechnet mit 10 m-Winden und beinhaltet eine Bias-Korrektur, um den ausgeprägten Tagesgang zu eliminieren.

(schwarze durchgezogene Linie) heran. Die Folgetagsprognose, die auf dem Ensemble-Mittel mit 100 m-Winden beruht, besitzt einen um etwa 2 % geringeren quadratischen Fehler (RMSE), bezogen auf die installierte Windleistung von 27 GW (Stand Ende 2010). Im Zeitbereich der Folgetagsprognose ist die Qualität der „Control Forecast“ und des Ensemblemittels in etwa gleich. Mit zunehmendem Vorhersagehorizont werden die Vorteile des Ensemblemittels gegenüber einem einzelnen Vorhersagelauf („Control Forecast“) augenscheinlich.

Hinsichtlich der probabilistischen Eigenschaften beziehungsweise Unterschiede von 10 m- und 100 m-Ensemblewinden wird bereits deutlich, dass die Spread-Skill-Beziehung, das heißt die Beziehung zwischen der Streuung des Ensembles und dem mittleren quadratischen Vorhersagefehler (RMSE) des Ensemblemittels bei 100 m-Winden, deutlich günstiger ausfällt. Im Idealfall würde die Linie der Ensemblestreuung (gepunktet) auf der Linie des RMSE-Vorhersagefehlers liegen. Je weiter die Linien voneinander entfernt sind und wenn die RMSE-Linie oberhalb liegt, desto schwerwiegender ist die Tatsache, dass in vielen Situationen die Streuung des Ensembles nicht groß genug ist. Dies bedeutet, dass die Beobachtung außerhalb des Bereiches liegt, der von den Ensemblemitgliedern abgedeckt wird.

Bei höheren Vorhersagehorizonten zeigt sich, dass Streuung und mittlerer Fehler dichter beieinander liegen; dies bedeutet im Umkehrschluss, dass das Ensemble bei kurzen Vorhersagehorizonten eindeutig nicht dispersiv genug ist. Dies ist aus dem Grund verständlich, da das EZMW-EPS für die mittelfristige Vorhersage entwickelt wurde. Der Erprobung von Ensembles, die speziell für die Kurz- und Kurzzeitprognose optimiert sind und von mehreren

nationalen Wetterdiensten entwickelt wurden, steht kurz bevor. Zu nennen sind hier das COSMO-DE-EPS vom DWD, MOGREPS vom UK MetOffice und PEARP von Meteo France.

Der Vorteil von 100 m-EPS-Winden gegenüber 10 m-EPS-Winden für die Windleistungsvorhersage wird nicht nur bei der Betrachtung des Fehlers des Ensemblemittels (Abbildung 7-3) deutlich, sondern zeigt sich auch in den probabilistischen Eigenschaften. Der „Continuous Rank Probability Skill Score“ (CRPSS) gibt an, wie groß die Verbesserung bei Verwendung von 100 m-Winden bezogen auf die 10 m-Winde ist. Im Kurzfristbereich ist die Überlegenheit der 100 m-Winde sehr deutlich (CRPSS > 0,4) und mit zunehmendem Vorhersagehorizont wird der Vorteil der 100 m-Winde kleiner. Für das Gebiet der 50 Hertz-Regelzone (neue deutsche Bundesländer) ist der Vorteil von 100 m-Winden etwas größer als für Gesamtdeutschland.

3.2 Eigenschaften von probabilistischen Windleistungsvorhersagen

An Hand der Spread-Skill-Beziehung (Abbildung 7-2) wurde schon erläutert, dass die Streuung der Ensemblemitglieder eine wesentliche Eigenschaft eines Ensemblesystems darstellt. Das Talagrand-Ranghistogramm (siehe Beitrag 4 in diesem Heft) in Abbildung 7-4 zeigt auf anschauliche Weise, ob die Ensemblestreuung angemessen, zu groß oder zu klein ist. Im EZMW-EPS ist es vorgesehen, dass jeder Ensemblelauf die gleiche Wahrscheinlichkeit besitzt ($= 1/51 = 0,0196$), am dichtesten an der Beobachtung zu liegen. Damit ergibt sich, dass im Falle eines optimalen Ensembles alle Intervalle des Talagrand-Ranghistogramms die gleiche Häufigkeit besitzen. Tritt dies nicht ein, sondern zeigt sich wie im Falle der 100 m-Wind-EPS-Vorhersage für Deutschland (Abbildung 7-4, links), dass die Ränge am linken und rechten Rand wesentlich häufiger besetzt sind, ist zu schließen, dass das Ensemble eine zu geringe Streuung aufweist. In diesem Fall ist die beobachtete Windleistung häufiger als erwartet kleiner (größer) als das kleinste (größte) Ensemblemitglied, so dass sich eine typische U-Form ausbildet. Bei den Vorhersagen für Tag 1 und 2 ist die zu geringe Streuung des Ensembles noch deutlicher ausgeprägt als an Vorhersagetag 3, der in Abbildung 7-4 (links) dargestellt ist.

Die Korrektur der Streuung (Dispersion) des Ensembles ist mit entsprechenden Kalibrierungsmethoden möglich, die im Beitrag 5 in diesem Heft vorgestellt werden, nichtsdestotrotz ist es wissenswert, ob es Erklärungen für die zu geringe beobachtete Streuung gibt.

Wenn bei der Verifikation die beobachtete Windleistung mit der für Deutschland simulierten Windleistung ausgetauscht wird, ist das 100 m-Windensemble nicht mehr unterdispersiv (Abbildung 7-4, rechts), sondern zeigt eine annähernde Gleichverteilung auf die Intervalle im Talagrand-Ranghistogramm. Aus diesem Ergebnis kann geschlossen werden, dass die Umrechnung von Windge-

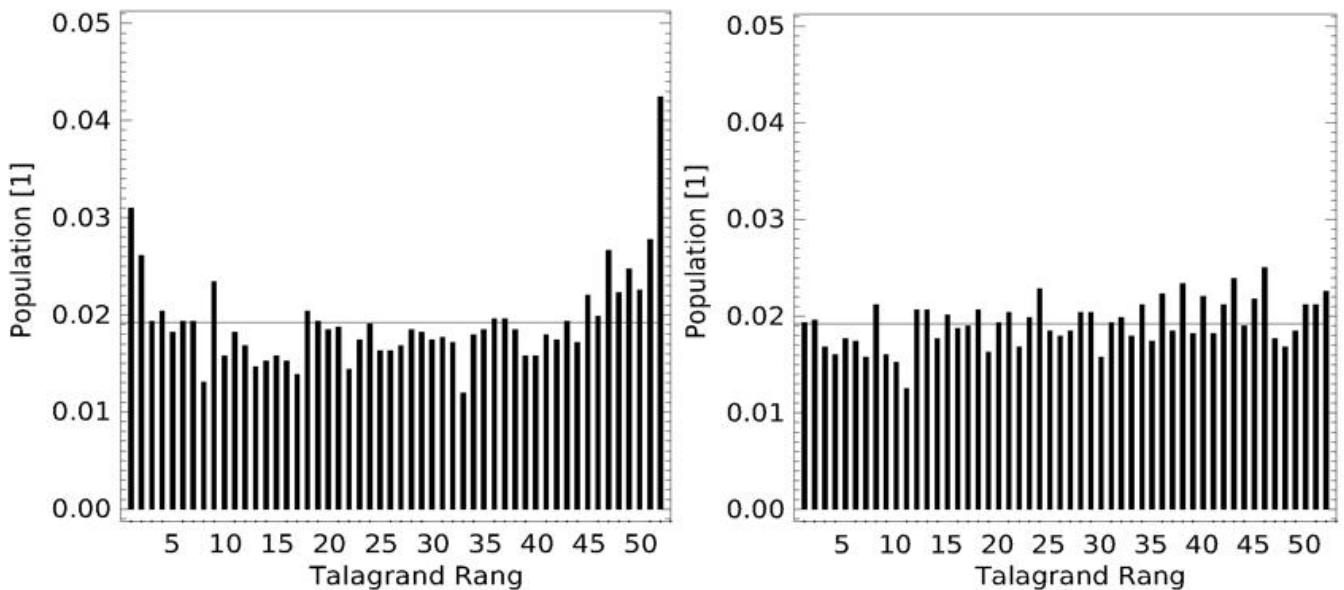


Abbildung 7-4: Talagrand-Diagramme der 100 m-Winde-Vorhersage (+72 h) validiert mit beobachteter Windeinspeisung in Deutschland (links) und mit simulierter Windleistung (rechts), die aus 100 m-Wind-Analysedaten berechnet wurde.

schwindigkeit in Windleistung mittels der nicht-linearen Leistungskurve nicht für die Entstehung der zu geringen Streuung verantwortlich ist. Darüber hinaus führt die Summation über alle deutschen Windparks, die räumlich ungleichmäßig verteilt sind, bei der Berechnung der gesamtdeutschen Windleistung, nicht zu einem negativen Einfluss auf die Eigenschaften des Windleistungsensembles. Es lässt sich zudem feststellen, dass dieses vielversprechende Ergebnis auch bedeutet, dass die neuen 100 m-Winde des EZMW-EPS im Bereich von 72 bis 120 Stunden gegenüber der Analyse ausreichend Streuung besitzen.

4 Anwendungen von probabilistischen Windleistungsvorhersagen

An der Anwendung von meteorologischen Ensembles im Bereich der Windleistungsprognose wird seit Jahren geforscht. Die Methodik zur Erstellung von Unsicherheitsintervallen und von Risikoindices aus der Ensemblestreuung wurde in vielen Studien erarbeitet und auch an diversen Beispielen demonstriert (PINSON und MADSEN 2009 sowie PINSON et al. 2009). Den (ökonomischen) Wert von probabilistischen Windleistungsvorhersagen gegenüber einer deterministischen Windleistungsvorhersage konnten ROULSTON et al. (2003) sehr früh mit einem sehr simplifizierten Marktmodell zeigen. Wie bereits angedeutet, basiert die heutige Netzzintegration von erneuerbaren Energien nicht auf probabilistisch beschreibbaren Entscheidungsprozessen. Vielmehr beruht die Netzzintegration auf der Maxime der Versorgungssicherheit, das heißt die Netzstabilität besitzt allerhöchste Priorität. Daher werden Ensembleprognosen zwar zur visuellen Abschätzung von Prognosesicherheit und Prognoseunsicherheit herangezogen, aber eine quantitative Bewertung des Nutzens von Windleistungsensembles in einem operationellen Umfeld ist bislang nicht

möglich. In Hauptabschnitt 4.1 wird die heutige Praxis der Nutzung von Unsicherheitsintervallen besprochen und in den Hauptabschnitten 4.2 und 4.3 wird ein Blick in die Zukunft gegeben, welche Anwendung räumlich differenzierte probabilistische Windleistungsvorhersagen haben können.

4.1 Wettersituationsangepaßte Unsicherheitsintervalle

Die Erfahrung in der Praxis aber auch wissenschaftliche Untersuchungen haben gezeigt, dass die Fehler der Windleistungsvorhersage von der Wetterlage abhängen (LANGE und FOCKEN 2005). Dies liegt zum einen an den zugrundeliegenden numerischen Wettermodellen und der Datenassimilation, aber auch an der allgemeinen Vorhersagbarkeit der Atmosphäre. Letztlich trägt die Transformation von meteorologischen Größen in Windleistung dazu bei, dass kleine Fehler in der Windgeschwindigkeit mal größere und mal weniger starke Fehler in der Windleistungsvorhersage verursachen.

Generell sind dynamische Wetterlagen, beispielsweise der Durchzug von Fronten, mit größeren Vorhersagefehlern im numerischen Wettermodell, beispielsweise der Windgeschwindigkeit, verbunden, als stabile Hochdruckwetterlagen. So können zum Beispiel beim Durchzug von Fronten große Abweichungen auftreten, wenn sich die Front schneller oder langsamer bewegt, als vorhergesagt. Diese sogenannten Phasenfehler können in der Deutschlandvorhersage, also der Summe aller deutschen Windparks, mehrere Gigawatt betragen und mehrere Stunden andauern.

Die nichtlineare Beziehung zwischen Wind und Windleistung führt dazu, dass bei Windgeschwindigkeiten im steilen Teil der Kennlinie zwischen 6 und 12 m/s kleine Fehler in der Vorhersage der Windgeschwindigkeit in relativ große Leistungsfehler umgewandelt werden, während bei kleinen

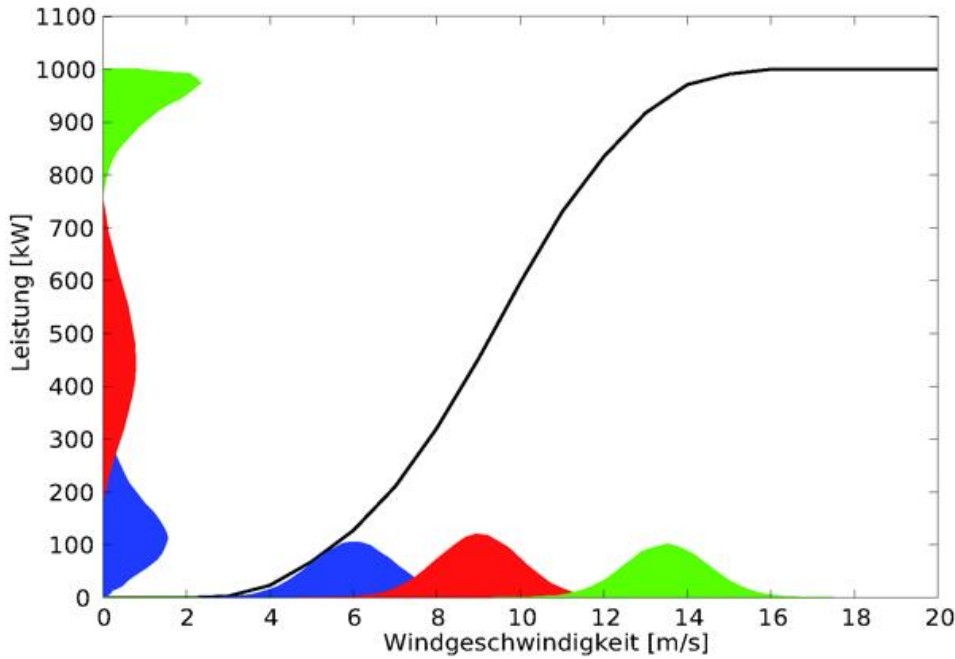


Abbildung 7-5: Die nichtlineare Kennlinie einer Windkraftanlage verstärkt Fehler in der Windgeschwindigkeitsvorhersage entsprechend ihrer Steigung, das heißt im mittleren Bereich der Kennlinie am stärksten. Betrachtet man gleiche gaußförmige Fehlerverteilungen der Windgeschwindigkeit zu verschiedenen Windgeschwindigkeitsniveaus (blau, rot, grün), so werden diese durch die Kennlinie stark verzerrt. Diese Eigenschaft muss bei der Bestimmung der Konfidenzintervalle berücksichtigt werden.

(0–4 m/s) und sehr großen Windgeschwindigkeiten (>13 m/s) die Kennlinie sehr flach ist und sich kleine Fehler in der Windgeschwindigkeitsvorhersage nur in geringem Umfang in der Leistungsvorhersage niederschlagen (Abbildung 7-5).

Die Anwender von Windleistungsvorhersagen favorisieren zwar eine deterministische Vorhersage, um beispielsweise eindeutige Fahrpläne für den Energiemarkt erstellen zu können. Darüber hinaus sind die Kunden jedoch an einer Einschätzung der Vorhersageunsicherheit interessiert, um das Risiko für Abweichungen einschätzen zu können und um ggf. die Absicherung mit Regelleistung zu erhöhen. In vielen Fällen erhalten die Kunden daher Abschätzungen der möglichen Abweichungen in Form von Unsicherheitsgrenzen, die die deterministischen Vorhersagewerte einschließen (Abbildung 7-6). In der Regel werden Konfidenzintervalle mit einem festgelegten Konfidenzniveau verwendet, die aus historischen Vorhersage- und Messdaten bestimmt werden, zum Beispiel indem Wahrscheinlichkeitsverteilungen historischer Prognosefehler bedingt auf aussagekräftige Variablen, die

typische Vorhersagesituationen und Wetterlagen beschreiben, verwendet werden. Auf diese Weise kann jeder Wert der deterministischen Leistungsvorhersage stundengenau mit einem passenden Unsicherheitsbereich versehen werden.

Die so bestimmten Konfidenzintervalle für das 70 %-Vertrauensniveau sind in Abbildung 7-6 als „klassisches“ Konfidenzintervall zu sehen und sind nach gleichlautender Rückmeldung der Kunden sehr gut geeignet, die tatsächlich auftretenden Abweichungen richtig einzuschätzen. In diesem Beispiel ist das aus der Streuung des EZMW-EPS berechnete Konfidenzintervall (ebenfalls 70 %) deutlich zu klein. Typischerweise in Wettersituationen mit mittlerer Einspeisung ist die Ensemblestreuung zu gering.

Vielen Nutzern kommt es im täglichen Umgang mit den Vorhersageunsicherheiten sehr stark auf den visuellen Eindruck der Konfidenzintervalle an, das heißt, ob diese weit auseinander liegen („unsicher“) oder dicht beisammen sind („sicher“). Die mathematische Bedeutung tritt dabei gegenüber der eigenen operationellen Erfahrung

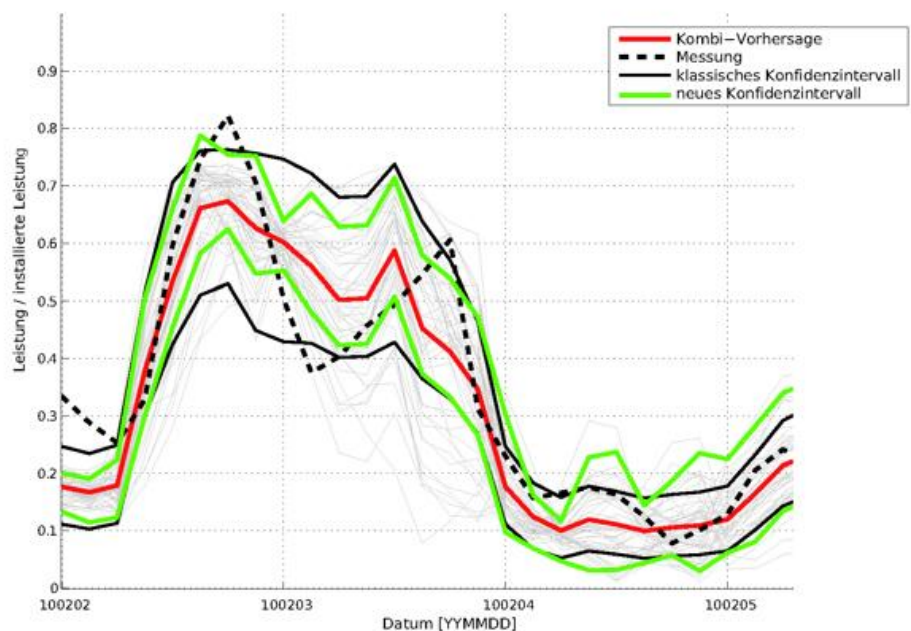


Abbildung 7-6: Bei dem EZMW-Ensemble sind die wetterlagenabhängigen Konfidenzintervalle (grün) kleiner als die „klassischen“ Konfidenzintervalle (schwarz). Der Hauptgrund scheint die zu geringe Streuung der Ensemblemitglieder zu sein (graue Linien). Es ist die „Day-Ahead-Vorhersage“ für den Zeitraum vom 2.2.2010 bis zum 5.2.2010 in der 50 Hertz-Regelzone dargestellt.

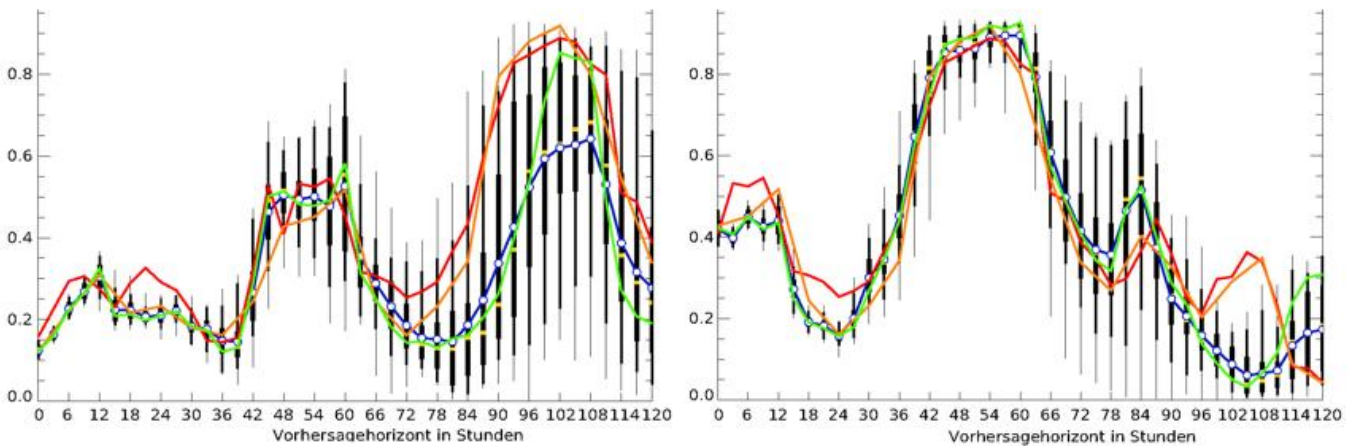


Abbildung 7-7: „Powergram“ der probabilistischen Windleistungsvorhersage, normiert mit der installierten Windleistung für Sturm „Xynthia“ in der 50 Hertz-Regelzone am 25. (links) und 27. (rechts) März 2010. Die gemessene Windleistung ist in rot und die aus der Analyse simulierte Windleistung ist in orange dargestellt. Die deterministische Vorhersage und das Ensemblemittel der vorhergesagten Windleistung ist in grün beziehungsweise blau gekennzeichnet. Die 50 %- beziehungsweise 90 %-Vorhersageintervalle sind als vertikale Boxen und die minimalen und maximalen Prognosen als Linien dargestellt.

in den Hintergrund. Die Erfahrung hat gezeigt, dass 90 %-Konfidenzintervalle zu groß sein können, weil sie keine große Variation und Dynamik aufweisen. Zu weite Konfidenzintervalle können beispielsweise dazu führen, dass ein Netzbetreiber sich aufgrund seiner Einschätzung, es am Folgetag mit unsicheren Vorhersagen zu tun zu haben, mit zusätzlichen Regelenergieserven eindeckt oder bestimmte Wartungsarbeiten im Netz verschiebt, um seine Flexibilität im Netzbetrieb nicht einzuschränken. Es zeigt sich deutlich, dass die Kommunikation wie Nutzer einfache probabilistische Informationen zu bewerten haben, extrem wichtig ist. Im Rahmen von Forschungsprojekten ist es notwendig, exemplarische Anwendungen von Ensembleprognosen für die Energiewirtschaft zu entwickeln. Der Entwurf von einfachen probabilistischen Kosten-Nutzen-Rechnungen sollte dabei die Ausbildung von probabilistischen Entscheidungsprozessen stimulieren.

4.2 Räumlich differenzierte Ensembleprognose

Nachdem zuvor Prognoseunsicherheitsintervalle und deren Anwendung für eine gesamte Regelzone (50 Hertz) diskutiert wurden, werden in diesem Abschnitt neue Entwicklungen für eine räumlich differenzierte Windleistungsprognose mit Angabe des Unsicherheitsbereiches vorgestellt. Räumlich differenzierte oder sogenannte netzknotenscharfe Vorhersagen sind bislang nur mit deterministischen Modellen üblich. Zunehmend haben Übertragungs- sowie Verteilnetzbetreiber ein sehr großes Interesse, zukünftige Einspeisungen an jedem Ort im Netz quantitativ (mit der deterministischen Vorhersage) aber auch probabilistisch beurteilen zu können, das heißt mit welcher Wahrscheinlichkeit beispielsweise das Gros der Einspeisung regional konzentriert eintritt oder ob sich die Einspeisung homogen über die Regelzone verteilt.

Zunehmend wird die Nord-Süd-Übertragungsnetzkapazität zum Nadelöhr der Windenergieeinspeisung. Schon heute ist die Abregelung von Windparks vielfach eingesetzte Praxis

der Netzbetreiber, die Überschreitung von Netzkapazitäten im Mittelspannungsnetz (110 kV) und zunehmend im Übertragungsnetz zu vermeiden (ECOFYS 2011). Auch wenn die entgangenen Einspeiseentgelte den Windparkbetreibern erstattet werden, sind diese Situationen nur vertretbar, wenn kein Strom aus konventionellen Kraftwerken durch diese Leitungen fließt. Die flächendeckende regionale Windleistungsprognose mit Konfidenzintervallen eignet sich für Planungen, in denen regelbare (konventionelle) Kraftwerke regional derart flexible eingesetzt werden, um Schwankungen von Wind und Sonne regional auszugleichen. Selbstverständlich sind hier die Regelungszeiten des jeweiligen Kraftwerkstyps zu berücksichtigen. Die trägsten Kraftwerke richten sich beispielsweise nach der „2 Day-ahead-Vorhersage“ (48 bis 72 Stunden) und „Day-ahead-Vorhersage“ (24 bis 48 Stunden) und die flexibelsten Kraftwerke (Gaskraftwerke und Kraftwärmekopplung) werden eingesetzt, die mittels Kurzfristvorhersagen (< 24 h) ermittelten Fehler auszugleichen. Längere Prognosehorizonte als 72 Stunden können für Windleistung auch interessant werden, wenn vorhersehbar überschüssige Windenergie im Sinne eines intelligenten Speichermanagements gespeichert werden soll.

Wie oben bereits erwähnt ist der Status Quo in Deutschland, dass die gesamte deutsche Windenergie an der Stromhandelsbörse für den Folgetag verkauft wird. Durch diese Praxis wird am Markt den (regionalen) Engpässen bei der Netzintegration leider nicht Rechnung getragen. Die zukünftige Entwicklung des europäischen Strommarktes ist größtenteils eine politische Herausforderung, um Bedingungen zu schaffen, dass erneuerbare Energien vorrangig und ökonomisch sinnvoll integriert werden.

Exemplarisch ist in Abbildung 7-8 eine räumlich differenzierte Windleistungsvorhersage dargestellt, die am 1. März 2010 (0 UTC) gültig ist. Zu diesem Zeitpunkt hat der Sturm „Xynthia“ die 50 Hertz-Regelzone erreicht. Das Risiko sehr hoher Windeinspeisung zeigt sich bereits am „Powergram“ der Vorhersage am 25. Februar (Ab-

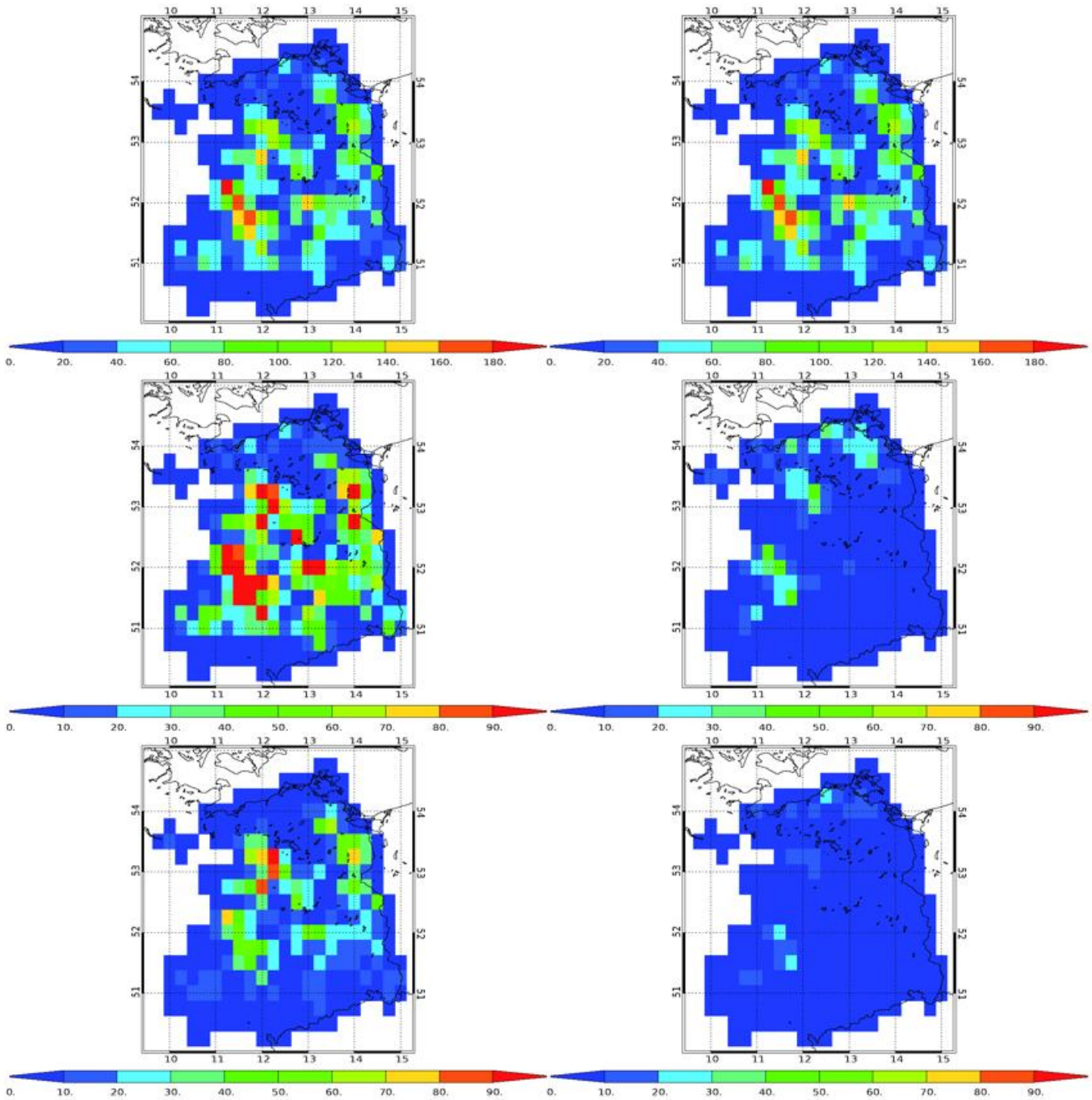


Abbildung 7-8: 96 h- (links) und 48 h-Windleistungsvorhersage (rechts) für Sturm „Xynthia“ in der 50 Hertz-Regelzone gültig am 1. März 2010, 0 UTC. Ensemblemittel (oben), 70 % Prognoseintervall (Mitte) und absoluter Prognosefehler (unten) in Megawatt.

bildung 7-7, links); allerdings weist die deterministische Vorhersage und das Ensemblemittel einen Phasenfehler von etwa 12 Stunden auf. Die deterministische und Ensemblevorhersage am 27. Februar (Abbildung 7-7, rechts) kann das Eintreffen des Sturmes ohne Phasenverschiebung sehr gut vorhersagen. Die räumlich dargestellten Ensemblemittel sind am 25. Februar 2010 (0 UTC) deutlich zu klein (Abbildung 7-8, oben links), so dass sich Fehler von über 80 MW für manche Gitterflächen ergeben (Abbildung 7-8, unten links). In Abbildung 7-8, Mitte, ist die Größe des 70 %-Konfidenzintervalls dargestellt, das aus den Ensemblemitgliedern bestimmt wurde. In den Regionen mit hoher Leistungsdichte (Magdeburger Börde) ist die Größe des Konfidenzintervalls sehr groß

(> 90 MW), das heißt das Ensemble zeigt an, dass eine sehr große Unsicherheit in der Vorhersage besteht. Die 48 h-Vorhersage von „Xynthia“ ist auch in ihrer räumlichen Verteilung sehr gut. Die Fehler des Ensemblemittels betragen weniger als 30 MW pro Gitterflächen (Abbildung 7-8, unten rechts). Die hohe Sicherheit dieser Prognose wird durch die deutlich kleineren Prognoseintervalle (< 50 MW) angezeigt (Abbildung 7-8, Mitte rechts).

4.3 Szenarien für Netzsicherheitsrechnung

Sturm „Xynthia“ eignet sich sehr gut, ein weiteres Anwendungsbeispiel der räumlich differenzierten probabilistischen Windleistungsvorhersage zu erläutern.

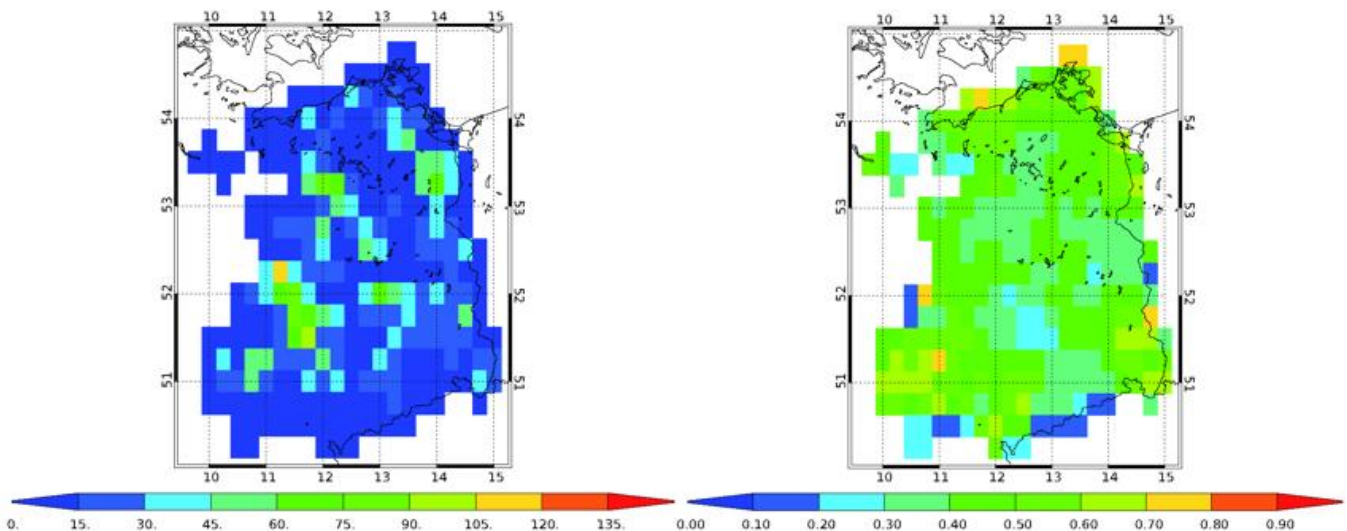


Abbildung 7-9: Maximal zu erwartende Windenergieeinspeisung (99%-Quantil) in Megawatt- (links) und normierter (rechts) Leistung pro Gitterfläche am 28. Februar 2010, 0 UTC (+72 h-Vorhersage) in der 50 Hertz-Regelzone basierend auf dem EZMW-EPS-Vorhersagegelauf am 25 Februar 2010 (0 UTC).

tern, das in Zukunft von den Übertragungsnetzbetreibern eingesetzt werden könnte. Um die Netzsicherheit zu gewährleisten, führen Übertragungsnetzbetreiber Lastflussrechnungen für verschiedene Einspeise- und Verbrauchsszenarien unter Berücksichtigung des sogenannten N-1-Kriteriums durch. N-1-Kriterium bedeutet hier, dass das System bei Ausfall einer wichtigen Komponente stabil bleibt. Es gilt festzustellen, ob Szenarien eintreten können, bei denen festgeschriebene Kapazitätsgrenzen von Leitungen und Umspannwerken überschritten werden. Dies gilt insbesondere, wenn im Rahmen von Wartungsarbeiten von der regulären Netztopologie abgewichen werden muss.

Bezüglich Wind und Solarenergie wird gegenwärtig von einem Volllast-Einspeiseszenario zu jedem Zeitpunkt ausgegangen. Die Situation, dass alle Erzeugernennleistung einspeisen ist zwar nicht unmöglich, jedoch relativ unwahrscheinlich. Hier können Ensembleprognosen eingesetzt werden, um die maximal erwartete regional differenzierte Windeinspeisung vorherzusagen. Im einfachsten Fall wird für jeden Gitterpunkt der Ensemblelauf mit maximaler vorhergesagter Windleistung herangezogen. Das in Abbildung 7-9 gezeigte Beispiel für den 28. Februar 2010 (0 UTC) ist so gewählt, dass es zwischen zwei Einspeisemaxima liegt (siehe Abbildung 7-7). In Abbildung 7-9 (links) ist zu erkennen, dass lediglich für einige Gitterflächen ein Risiko von 1% besteht, dass die Einspeisung 80% der Nennleistung erreicht. Für weite Teile der 50 Hertz-Regelzone beträgt die vorhergesagte maximale Einspeisung nur die Hälfte der installierten Windleistung. Das heißt, dass trotz des herannahenden Sturmes und des zu erwartenden hohen Leistungsgradienten eine Lastflussrechnung nicht mit maximaler Einspeisung für jede Gitterfläche gerechnet werden müsste. Während dieser Zeitspanne ist also die Netzbelastung durch Windenergie vergleichsweise gering, so dass zum Beispiel Wartungsarbeiten an Umspannwerken und Leitungsabschnitten möglich wären.

Das Risiko von Volllasteinspeisungen wird erst wenige Stunden später relevant.

5 Zusammenfassung und Ausblick

Die sichere Netzintegration der nicht-steuerbaren aber vorhersagbaren Energieerzeugung aus Wind (und Sonne) ist nötig für den eingeleiteten Umbau des Stromversorgungssystems zu mehr Nachhaltigkeit. Der Güte von Windleistungs- (aber auch Solarleistungs-) Prognosen kommt hier eine sehr große Bedeutung zu, den Marktwert von Wind- und Solarenergie zu erhöhen. Auch wenn gegenwärtig der Fokus auf Vorhersagen für den Folgetag liegt, ist absehbar, dass sich die nachgefragten Vorhersagehorizonte deutlich vergrößern werden. Hier ist beispielsweise die Planung von Wartungsarbeiten an Windkraftanlagen oder am Stromnetz zu nennen, aber auch das Speichermanagement und der Brennstoffeinsatz in fossilen Kraftwerken.

Ensembleprognosen können hier die notwendige Vorhersagegüte als Ensemblemittel liefern einschließlich der Abschätzung der Prognoseunsicherheit. Die Entwicklung von probabilistisch ausgelegten Entscheidungsmodellen in der Energiewirtschaft wird entscheiden, wie erfolgreich der volle Informationsgehalt von Ensembles letztendlich verwertet werden wird. Bei der Entwicklung von Ensembleanwendungen ist die Kommunikation mit den Nutzern in der Energiewirtschaft sehr wichtig, um die Akzeptanz und den Einsatz von Ensembles zu unterstützen.

In den letzten Jahren haben Forscher aus der Meteorologie und aus der Windenergiebranche gemeinsam erfolgreiche Forschungsprojekte durchgeführt. Die Prognosegüte wurde durch weiterentwickelte Prognoseverfahren und bessere Daten der Wettervorhersagemodelle deutlich verbessert. Dies bezieht sich auch auf die Nutzung von En-

sembledaten, die beispielsweise am EZMW nun auch in Nabhöhe von Windkraftanlagen verfügbar sind. Der Einsatz lokaler Ensemblemodelle der nationalen Wetterdienste verspricht in der Kurz- und Kurzfristvorhersage sehr wichtige Impulse bei der Netzbetriebsplanung und der Netzbetriebsführung hinsichtlich weiter stark steigendem Wind- und Solarenergiezubau.

Literatur

- BEYER, H.G., HEINEMANN, D., MELLINGHOFF, H., MÖNNICH, K., WALDL, H.P., 1998: Vorhersage der regionalen Leistungsabgabe von Windkraftanlagen. *Tagungsband der 4. Deutschen Windenergiekonferenz 1998*, Wilhelmshaven.
- ECOFYS, 2011: Abschätzung der Bedeutung des Einspeisemanagements nach EEG 2009 - Auswirkungen auf die Windenergieerzeugung in den Jahren 2009 und 2010. Projektbericht für *Bundesverband Windenergie e.V.*, Berlin.
- GIEBEL, G., BROWNSWORD, R., KARINIOTAKIS, G., DENHARD, M., DRAXL, C., 2011: The State-Of-The-Art in Short-Term Prediction of Wind Power A Literature Overview. Project report for the Anemos.plus and Safe-Wind projects. 2nd Edition, Risø, Roskilde, Denmark.
- LANDBERG, L., 1998: Short-term Prediction of the power production from wind farms. *J. of Wind Engineering and Industrial Aerodynamics* **80**, 207-220.
- LANGE, M., FOCKEN, U., 2005: Physical Approach to Short-Term Wind Power Prediction. *Springer* Berlin, Heidelberg, New York, 2005.
- LANGE, M., FOCKEN, U., MEYER, R., DENHARD, M., ERNST, B., BERSTER, F., 2006: Optimal Combination of Different Numerical Weather Models for Improved Wind Power Predictions. In: 6th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms, Delft.
- MCLEAN, J. R., 2008: Equivalent Wind Power Curves. *TradeWind-Project*, Deliverable D2.4.
- PINSON, P., MADSEN H., 2009: Ensemble-based probabilistic forecasting at Horns. *Rev. Wind Energy* **12**, 137-155.
- PINSON, P., NIELSEN, H.A., MADSEN, H., KARINIOTAKIS, G., 2009: Skill forecasting from ensemble predictions of wind power. *Applied Energy* **86**, 1326-1334.
- ROULSTON, M.S., KAPLAN, D.T., HARDENBERG, J., SMITH, L.A., 2003: Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy* **28**, 585-602.

M. DENHARD

8 Herausforderungen in der Interpretation und Kommunikation von probabilistischen Vorhersagen

Challenges in the interpretation and communication of probabilistic forecasts

Zusammenfassung

Da eine Wetterentwicklung unsicher ist, hat die Verwendung von Wahrscheinlichkeiten in der Wettervorhersage eine lange Tradition. Mit den heutigen Möglichkeiten der numerischen Ensemblevorhersage entsteht allerdings eine Informationsmenge von bisher nicht gekanntem Ausmaß und die Verfahren in der Kommunikation von Unsicherheit müssen erweitert oder sogar neu definiert werden. Von besonderer Wichtigkeit ist dabei die Entscheidungsfindung mit Wahrscheinlichkeitsvorhersagen aus Ensemblesystemen.

Abstract

Because a weather forecast is uncertain, using probabilities in weather forecasting has a long tradition. Today's ensemble techniques provide a huge amount of forecast information never seen before and the communication of uncertainty requires new terms and definitions. Especially important is the decision making process based on probabilistic forecast information from ensembles.

1 Einleitung

Erfolgreiche Kommunikation benötigt ein gemeinsames Vokabular aus Begriffen und Definitionen, die als Transportmittel für den Austausch von Informationen dienen. Im Bereich Wettervorhersage wird ein reiches Vokabular aus relativierenden Begriffen verwendet, um die Unsicherheiten aber auch die Gefahren einer Wetterlage richtig zu kommunizieren. Für das Wettertextbeispiel im gelben Kasten sind dies unter anderem die Begriffe „vielerorts“ oder „im Tagesverlauf“, hinter denen sich nicht genauer quantifizierte Wahrscheinlichkeitsaussagen verbergen.

Eine Wettervorhersage beginnt immer mit einer Wetterdiagnose unter Verwendung der verfügbaren Daten und numerischen Modellvorhersagen. Die Meteorologen erstellen eine Beschreibung der wesentlichen für eine Wetterlage charakteristischen Prozesse in einem konzeptionellen Modell (KRENOVSKI et al. 2010). Daraus entstehen Vorhersagekonzepte für Kürzest-, Kurz- und Mittelfrist, die neben der wahrscheinlichsten Entwicklung auch eine Bewertung der Vorhersageunsicherheit und möglicher Alternativen enthalten.

Der Meteorologe kommuniziert diese Guidance und ihre Unsicherheit in seinen Vorhersagetexten durch relativierende Terminologie oder verwendet großzügige Symbolik mit Piktogrammen bei der Wetterpräsentation. Das Wettertextbeispiel im gelben Kasten zeigt, wie unsicher eine „sichere“

Prognose sein kann, denn der Vorhersagetext ist keinesfalls eindeutig. Ort und Zeit des Auftretens der Wolkenbildung und der Niederschläge bleiben in einem gewissen Rahmen offen. Der Terminus „sicher“ meint hier, dass der prinzipielle Charakter des Wetters mit seinen möglichen Begleiterscheinungen für den genannten Tag angegeben werden konnte. Die Unsicherheit im Wettertext bezieht sich auf die zum Ausgabezeitpunkt des Textes noch nicht vorhersagbaren Teile der Wetterdynamik.

Eine seriöse Wettervorhersage ist deshalb immer probabilistisch, auch wenn die Unsicherheit manchmal gering ist, und beschreibt alle möglichen Wetterentwicklungen mit ihren Eintrittswahrscheinlichkeiten. Dieses Informationspaket wird in der Literatur als der true belief einer Wetterentwicklung bezeichnet (siehe zum Beispiel JOLLIFFE 2008). Der true belief steht oft im Gegensatz zur Erwartung an den Meteorologen, doch genau zu sagen „wie das Wetter wird“. Wohl wissend, dass er diesem Anspruch nicht gerecht werden kann, wählt der Meteorologe aus dem Bündel an Möglichkeiten die seiner Ansicht nach wahrscheinlichste Entwicklung aus und hofft, dass er mit dieser deterministischen Aussage richtig liegt.

Über die einem Kundenwunsch nachgebende Intention hinaus ist in vielen Situationen eine Entscheidung unumgänglich. Neben Steuerungsprozessen in Industrie, Landwirtschaft oder Katastrophenschutz ist auch die alltägliche Frage nach der Mitnahme eines Regenschirms betroffen.

Wettertext

„48 Stunden vor dem Verifikationstermin war dann eine sichere Prognose möglich, die dann auch eingetroffen ist. Sie lautete: ...Auch am Ostersonntag scheint weiterhin vielerorts die Sonne und es bleibt meist trocken. Lediglich im Südwesten und Westen entwickeln sich im Tagesverlauf einige Quellwolken, die zu einzelnen Schauern oder Gewittern führen können...“

Aus KRENOVSKI et al. (2010).

Dieser Artikel soll zeigen, wie die Notwendigkeit einer Entscheidung mit der Unsicherheitsinformation aus Ensemblesystemen verknüpft werden kann.

Der Abschnitt 2 soll das Verständnis für den Begriff des *true belief* einer Wettervorhersage und seiner Eigenschaften vertiefen. Daraus ergeben sich Regeln im Umgang mit Wahrscheinlichkeitsvorhersagen, die wichtig für ihre korrekte Interpretation und Kommunikation sind. Der Abschnitt 3 erläutert den Bezug von Wahrscheinlichkeitsaussagen zu Raum und Zeit und der Abschnitt 4 diskutiert, wie Entscheidungen auf der Basis probabilistischer Vorhersagen entstehen und welche Annahmen dafür notwendig sind.

2 Dynamik der Unsicherheit

In der mathematischen Abstraktion ist der Zustand eines dynamischen Systems zu einer bestimmten Zeit durch einen Punkt im Raum seiner Zustandsgrößen definiert. Zu den Zustandsgrößen der Atmosphäre gehören alle meteorologisch relevanten Parameter, wie Temperatur, Druck oder Feuchte, im Raum. Eine Änderung des Atmosphärenzustands manifestiert sich durch die Bewegung des Zustandspunktes im Zustandsraum. Solche Trajektorien im deterministischen Chaos verlaufen eindeutig, obwohl sie nur begrenzt vorhersagbar sind (WERNLI in diesem Heft).

Diese Art der Beschreibung erfordert ein wenig Vorstellungskraft, da die Atmosphäre sehr viele Zustandsgrößen besitzt. Sie ist aber sehr nützlich, denn die Art des Auseinanderlaufens benachbarter Zustandspunkte gibt Aufschluss über die vorherrschende Dynamik und damit die Vorhersageunsicherheit. Je schneller sich die Trajektorien voneinander entfernen, desto sensibler ist die Dynamik gegenüber Störungen und desto größer die Vorhersageunsicherheit.

Der *true belief* über eine Vorhersage ist durch alle simulierten Trajektorien im Zustandsraum gegeben, die ausgehend von den möglichen Anfangszuständen eine Chance auf Realisierung haben. Dabei werden alle Anfangszustände einbezogen, die durch die vorhandenen Messungen nicht ausgeschlossen werden können. Der Zustandsraum einer Vorhersage ist aber nicht mit dem wahren Zustandsraum

der Atmosphäre identisch, sondern wird vom verwendeten Modellsystem definiert. Der *true belief* beschreibt deshalb nicht die Wahrheit, sondern umfasst nur das vorhandene Wissen über dieselbe.

Eine wichtige Eigenschaft des *true belief* ist seine Verlässlichkeit. Dazu muss die tatsächliche Entwicklung eines Systems immer im Bereich des Bündels der Vorhersage-trajektorien verlaufen, wobei die Dichte der Trajektorien im Zustandsraum ein guter Schätzer für die Wahrscheinlichkeit sein soll, dass die Wetterentwicklung diesen ausgewählten Trajektorien folgen wird (WEIGEL in diesem Heft). Konkret bedeutet dies, dass die tatsächliche Wettertrajektorie immer mindestens einen Begleiter, ihren sogenannten Schatten (im Englischen *shadow*) hat. Nur wenn probabilistische Vorhersagen verlässlich (im Englischen *reliable*) sind können sie in wetterabhängigen Entscheidungsprozessen genutzt werden.

3 Vom Ensemble zu Wahrscheinlichkeiten

Wahrscheinlichkeiten aus numerischen Ensemblesystemen werden durch Abzählen der Mitglieder im Ensemble gewonnen, die ein Ereignis vorhersagen. Je kleiner das Ensemble ist, desto ungenauer sind die Wahrscheinlichkeitsangaben. Da in kleinen Ensemblesystemen nur wenige Vorhersagetrajektorien durch den sehr komplexen „Zustandsraum der Atmosphäre“ laufen, kann es leicht passieren, dass die tatsächliche Wettertrajektorie aus dem Ensemble der Vorhersagetrajektorien ausschert. Dann muss entweder das Ensemble vergrößert oder ein statistisches Postprocessing durchgeführt werden, dass die fehlende Unsicherheit ergänzt (THORARINSDOTTIR in diesem Heft). Wahrscheinlichkeiten sind allerdings nicht so einfach zu interpretieren wie eine deterministische Vorhersage, bei der es nur eine einzige Vorhersagetrajektorie gibt und der Vorhersagewert einer meteorologischen Größe an einem Ort eindeutig bestimmt ist. Egal wie groß ein Gebiet um diesen Ort herum auch sein mag, die Frage, ob das Ereignis in diesem Gebiet auftreten wird, muss bei einer deterministischen Vorhersage immer mit „ja“ beantwortet werden, denn über das Auftreten des Ereignisses an diesem Ort wurde bereits bei der Erstellung der Vorhersage entschieden. Bei Wahrscheinlichkeiten ist das nicht so.

Gefragt sei die Eintrittswahrscheinlichkeit $P(E \text{ in } G)$ für ein Ereignis E in einem Gebiet G , das sich aus mehreren Teilgebieten G_1, \dots, G_n , zum Beispiel den Boxen eines Modellgitters, zusammensetzt. Die Gebietswahrscheinlichkeit ergibt sich für drei Teilgebiete zu:

$$P(E \text{ in } G) = P(E \text{ in } G_1) + P(E \text{ in } G_2) + P(E \text{ in } G_3) + P(E \text{ in } G_1 \text{ und } G_2 \text{ und } G_3) - P(E \text{ in } G_1 \text{ und } G_2) - P(E \text{ in } G_1 \text{ und } G_3) - P(E \text{ in } G_2 \text{ und } G_3) \tag{1}$$

Tritt das Ereignis in den Teilgebieten näherungsweise unabhängig auf, zum Beispiel Gewitter in Afghanistan, Argentinien und Australien, bleiben für die bedingten Terme,

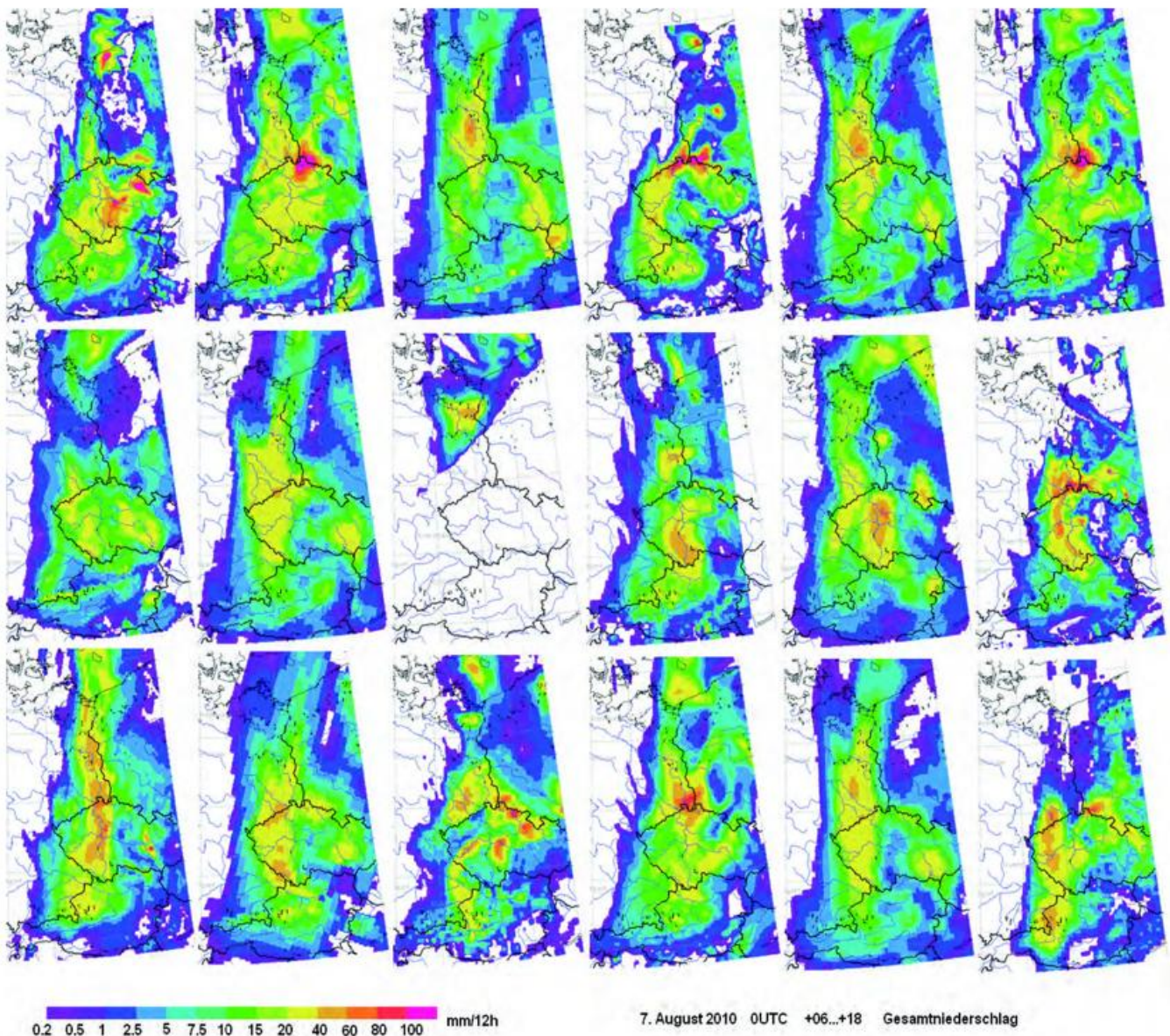


Abb. 8-1: Verschiedene Lokalmodellvorhersagen der nationalen Wetterdienste in Europa im Rahmen des SRNWP-PEPS (Poor Man's Ensemble Prediction System). Dargestellt ist die 12-stündige Niederschlagssumme von 6 UTC bis 18 UTC im Bereich Deutschland, Polen, Tschechien und Österreich am 7. August 2010. Fast alle Vorhersagen zeigen ein relevantes Niederschlagsereignis in diesen Gebieten an, aber die Orte mit der erwarteten maximalen Niederschlagsmenge unterscheiden sich deutlich.

das sind die mit einem „und“, nur die zufälligen Koinzidenzen. Diese sind bei seltenen Ereignissen äußerst gering und die Wahrscheinlichkeit im Gesamtgebiet ergibt sich im Wesentlichen aus der Summe der Teilgebietswahrscheinlichkeiten. Sind die Teilgebiete in Bezug auf das Ereignis dagegen vollständig miteinander verbunden, zum Beispiel wenn alle Teilgebiete innerhalb eines großräumigen Wetterphänomens liegen, dann sind alle Terme in Gleichung (1) identisch und es bleibt nur ein einziger Term übrig. Das ist dann die Wahrscheinlichkeit dafür, dass ein Wetterphänomen überhaupt auftritt.

Zwischen diesen Extremen variiert die Wahrscheinlichkeit mit der Gebietsgröße und der räumlichen Ausdehnung des Wetterphänomens, das heißt seiner dynamischen Struktur. So kann es passieren, dass in keinem der Teilgebiete die vorhergesagte Wahrscheinlichkeit ausreicht, um eine Hand-

lung auszulösen, dagegen für das Gesamtgebiet die Handlungsschwelle überschritten wird. Die Gleichung (1) zeigt uns, dass dies umso häufiger passieren kann, je kleiner der dynamische Zusammenhang zwischen den Teilgebieten ist, da sich deren Wahrscheinlichkeiten dann addieren.

Die Abbildung 8-1 zeigt ein Beispiel für das SRNWP-PEPS, das die in Europa betriebenen Lokalmodellvorhersagen der nationalen Wetterdienste in ein Ensemblesystem kombiniert. Dargestellt sind die Niederschlagsvorhersagen der einzelnen Ensemblemitglieder, hier 18 Stück. Betrachten wir zunächst das Ereignis „Niederschlagsmenge > 5 mm in 12 Stunden“, das überall dort vorhergesagt wird, wo mindestens die Farbe hellblau gilt, also auch in Gebieten mit Grün-, Gelb- und Rottönen. Stellen wir uns weiterhin vor, dass all diese Farben nach der Zuordnung zum Ereignis einheitlich grau eingefärbt werden. Dann sind die ent-

stehenden grauen Muster in den einzelnen Vorhersagen sehr ähnlich, das heißt das genannte Ereignis konnte in seiner Ausdehnung vom Ensemble gut eingegrenzt werden. Da die meisten Ensemblemitglieder in etwa dem gleichen Gebiet „ja“ zum Ereignis sagen, entsteht dort eine hohe Eintrittswahrscheinlichkeit.

Dagegen sind die Niederschlagsmuster im Ensemble bei einer Ereignisschwelle von „40 mm in 12 Stunden“, hier mindestens die Farbe Gelb, sehr verschieden und kleinräumig. Die Muster überlappen nur selten, so dass die lokal vorhergesagten Wahrscheinlichkeiten aus dem Ensemble für dieses Ereignis klein bleiben. Ist, wie im vorliegenden Fall, die Vorhersagbarkeit für ein Phänomen gering, kann man sich mit einem Trick behelfen. Dazu unterteilt man das Vorhersagegebiet in relativ große Teile. Dabei gilt: Je größer ein solches Gebiet ist, desto mehr lokale Muster aus dem Ensemble werden entsprechend Gleichung (1) über ihre Zugehörigkeit zum gleichen Bezugsgebiet miteinander verbunden, so dass dort die Eintrittswahrscheinlichkeit für das Ereignis steigt.

Wie dieses Beispiel zeigt, benötigt man zur Bestimmung von Gebietswahrscheinlichkeiten die raumzeitliche Ausdehnung von Wetterphänomenen und ihre diesbezügliche Unsicherheit, die am besten durch ein numerisches Modellensemble beschrieben wird. Damit kann aus der Konjunktion beziehungsweise Disjunktion der Vorhersagemuster im Augenblick der Vorhersage die Eintrittswahrscheinlichkeit im Gesamtgebiet bestimmt werden (vergleiche dazu auch das Zahlenbeispiel im blauen Kasten). Diese muss für jedes Gebiet und jede Vorhersagesituation neu aus dem Ensemble ermittelt werden, denn einerseits lässt sich aus der Wahrscheinlichkeit für das Gesamtgebiet nicht auf die Wahrscheinlichkeiten in seinen Teilen schließen (Gleichung (1) ist nicht eindeutig umkehrbar), und andererseits benötigt ein „Upscaling“ auf ein größeres Gebiet wiederum die dynamische Information aus dem Ensemble. Die Kenntnis über die räumliche Korrelation reicht dazu nicht aus, da diese als statistisches Mittel über viele verschiedene Wetterzustände keine gute Beschreibung des dynamischen Einzelfalls bietet.

Entsprechendes wie für Gebiete gilt auch für Zeitintervalle in der Vorhersage, zum Beispiel 3-, 6-, 12- oder 24-stündige Niederschlagssummen. Je länger das Zeitintervall der Auswertung, desto wahrscheinlicher ist es, dass ein Ereignis im Zeitintervall registriert wird. Die Länge des Zeitintervalls spielt besonders dann eine relevante Rolle, wenn die Dauer des Wetterphänomens kürzer als die Länge des Zeitintervalls ist.

4 Entscheidung mit unsicherer Information

Die Kommunikation von Wetterinformation hat vor allem deren optimale Nutzung in Entscheidungsprozessen zum Ziel, egal, ob es um die Frage der Mitnahme eines Regen-

schirms oder um die Regelung von Kraftwerken bei der Stromproduktion geht. Da eine Wettervorhersage unsicher ist, gilt dies natürlich auch für die getroffene Entscheidung. Sie kann falsch sein. Im Gegensatz zur richtigen Entscheidung gibt es aber zwei Arten von falschen Entscheidungen: man hat den Regenschirm umsonst mitgenommen oder man wird nass. Im ersten Fall ist man auf einen falschen Alarm hereingefallen und im zweiten Fall wurde das Ereignis nicht vorhergesagt, also vom Vorhersagesystem verpasst.

Die beiden Arten einer falschen Entscheidung können für einen Nutzer von unterschiedlicher Qualität sein. Dem Einen ist es sehr lästig einen Regenschirm mit sich zu tragen, während ein Anderer ihn auch als Spazierstock benutzt und kein Problem damit hat, ihn mitzunehmen. Der zweite Nutzer ist damit toleranter gegenüber einem falschen Alarm und hat ein anderes Kosten/Nutzen-Verhältnis als der Erste.

Im deterministischen System, das nur eine eindeutige *Guidance* kommuniziert, ohne diese in den Kontext mit den Alternativen zu stellen, hat der Nutzer keine Wahlmöglichkeit. Er muss der *Guidance* folgen, da er keine weiteren Informationen besitzt. Im probabilistischen Vorhersagesystem dagegen erhält der Nutzer eine aktive Rolle bei der Entscheidungsfindung. Beispielsweise könnte sich ein „regentoleranter Nutzer“ erst bei einer Wahrscheinlichkeit von wenigstens 80 % für Regen entschließen, den Schirm mitzunehmen, wohingegen ein eher hydrophob orientierter Nutzer schon bei einer Regenwahrscheinlichkeit von 30 % seinen Schirm dabei hat. Das heißt aber nicht, dass er ihn auch brauchen wird. Er trifft lediglich Vorsorge entsprechend seiner Sensitivität gegenüber dem Nasswerden und handelt damit aus seiner Sicht ökonomisch sinnvoll.

Gebietswahrscheinlichkeit

Ein Ensemble habe 100 Mitglieder. Im Gebiet G_1 wird das Ereignis von einem Mitglied, in G_2 von 3 und in G_3 von 5 Mitgliedern vorhergesagt. Die Wahrscheinlichkeit, das Ereignis im Gesamtgebiet $G=G_1+G_2+G_3$ zu finden, entspricht dem Anteil der Ensemblemitglieder, die das Ereignis in G vorhergesagen. Diese Zahl ergibt sich nicht unbedingt aus der Summe der Zahlen in den Teilgebieten. Es gilt:

- Nur wenn alle Mitglieder in allen Teilgebieten verschieden sind, ist $P(E \text{ in } G)=9\%$.
- Sind aber z. B. nur alle Mitglieder in G_1 und G_2 verschieden und sagen diese Mitglieder das Ereignis auch in G_3 vorher, so wird das Ereignis im Gesamtgebiet nur von fünf Mitgliedern gestützt und es gilt: $P(E \text{ in } G)=5\%$.

Die konsistente Schätzung einer Gebietswahrscheinlichkeit setzt die Kenntnis über die Dynamik einer Wettersituation in Raum und Zeit im Augenblick der Vorhersage voraus und gilt nur für das zugrunde liegende Gebiet.

Die konsistente Entscheidungsfindung auf Nutzerseite setzt die Kommunikation des true belief, das heißt einer verlässlichen probabilistischen Vorhersage, voraus. Ist dies nicht der Fall oder wird nur die wahrscheinlichste Entwicklung kommuniziert, übernimmt der Provider der Wetterinformation die Verantwortung für die wetterabhängige Entscheidung. Kann der Provider dagegen den true belief erzeugen, stehen ihm zwei Optionen offen:

1. Der Provider verwendet ein aus seiner Sicht angemessenes oder in Absprache mit dem Nutzer erstelltes Kosten-Verlust-Verhältnis und behält die Entscheidungskompetenz.
2. Der Provider übermittelt dem Nutzer den true belief über seine Vorhersage und versetzt ihn damit in die Lage, die Entscheidung selbst zu treffen.

Damit erweitert die probabilistische Vorhersage die Kommunikation von Wetterinformation um die Möglichkeit, Verantwortung für wetterabhängige Entscheidungen dort zu treffen, wo sie am besten wahrgenommen werden kann. Dies kann auch der Nutzer sein, der die Wetterinformation mit anderen, für seine Entscheidung relevanten Kriterien, verknüpft.

4.1 Kontinuierliche Ereignisvariable

Betrachten wir zunächst den allgemeinen Fall einer kontinuierlichen Observablen y mit einer kontinuierlichen Ereignisvariablen x . SMITH et al. (2001) geben als Beispiel die Zahl x von einem Bäcker an einem Tag produzierten Bagel (ringförmiges Gebäck aus Hefeteig) an, wobei die Nachfrage y auch vom Wetter abhängig ist. Die Kostenstruktur $U(x, y)$ verrechnet die Kosten der Produktion (PK) mit dem Verkaufsgewinn (VG).

$$\begin{aligned} U(x, y) &= yVG - xPK & Y \leq X \\ U(x, y) &= x(VG - PK) & Y > X \end{aligned} \tag{2}$$

wobei X die Zahl der produzierten und Y die Zahl der tatsächlich verkauften Bagels ist. Der erwartete Nutzen einer Entscheidung $X \in x$ aufgrund einer wetterabhängigen probabilistischen Vorhersage $p(y)$, zum Beispiel für die Zahl der an einem Tag verkauften Bagels, ergibt sich durch die erwarteten Kosten $E_x[U(x, y)]$ der Entscheidung X zu

$$E_x[U(x, y)] = \int_y U(x, y) p(y) dy \tag{3}$$

Einsetzen von (2) in (3) ergibt

$$E_x[U(x, y)] = VG \cdot \left[\int_{y \leq x} y p(y) dy + x \int_{y > x} p(y) dy \right] - PK \cdot x \tag{4}$$

und zeigt, dass eine optimale Entscheidung hier in komplexer Weise von der Form der Wahrscheinlichkeitsvorhersage abhängig ist. Daher brauchen wetterabhängige Entscheidungen verlässliche probabilistische Vorhersagen, um optimal sein zu können.

	binär	kontinuierlich
Treffer:	$p > P_T \rightarrow Y \in]X_1, X_2]$	$Y = X$
Verpasstes Ereignis:	$p \leq P_T \rightarrow Y \in]X_1, X_2]$	$Y > X$
Falscher Alarm:	$p > P_T \rightarrow Y \in]X_1, X_2]$	$Y < X$

Tab. 8-1: Definition von Treffern, falschen Alarmen und verpassten Ereignissen für binäre und kontinuierliche Ereignisvariablen x , wobei das binäre Ereignis auf der kontinuierlichen Observablen y durch das Intervall $]X_1, X_2]$ definiert ist. X ist die Entscheidung und Y die Beobachtung. Eine „Ja“-Vorhersage im binären Fall erfordert das Überschreiten der Wahrscheinlichkeitsschwelle P_T durch die vorhergesagte Wahrscheinlichkeit p , das Ereignis zu beobachten.

Für die Gewinnoptimierung im Bagel-Problem ist die Marge bestimmend, das heißt das Verhältnis aus Produktionskosten und Verkaufsgewinn. Sind die Produktionskosten niedrig, kann eine gewisse Überproduktion in Kauf genommen werden und die Produktion X sollte dem Erwartungswert y (Ensemblemittelwert) entsprechen. Ist dagegen die Marge klein, SMITH et al. (2003) verwenden 3 %, sollte die Produktion deutlich unterhalb des Erwartungswerts liegen, um Überproduktionen möglichst zu vermeiden.

Unter der Voraussetzung, dass es sich bei $p(y)$ um den true belief einer Vorhersage handelt, liefert die Gleichung (4) eine optimale Entscheidung über die Produktionsrate X . Mögliche Unterschiede in der Vorhersagbarkeit des Bagel-Bedarfs an warmen und kalten Tagen werden im true belief berücksichtigt.

4.2 Binäre Ereignisse

Binäre Ereignisse sind durch ein festes Intervall $]X_1, X_2]$ auf einer kontinuierlichen Observablen y definiert. Ein Beispiel ist das Überschreiten einer Warnschwelle, zum Beispiel Niederschlagsmenge in 24 h > 20 mm. Für diese Ereignisse ist nicht die Nähe zu einem bestimmten optimalen Wert entscheidend, sondern das Ereignisintervall unterteilt die Wahrscheinlichkeit in Bereiche für und gegen das Ereignis. Damit sind die Kosten nicht mehr von den Ereignisvariablen x und y abhängig sondern für alle möglichen Ausgänge der Vorhersage fest vorgegeben und die Kostenfunktion ergibt sich zu

$$\begin{aligned} U(x, y) &= a_{V,T} & Y \in]X_1, X_2] \\ U(x, y) &= a_{F,N} & Y \notin]X_1, X_2] \end{aligned} \tag{5}$$

wobei $a_{F,N}$ entweder die Kosten eines falschen Alarms oder einer korrekten „Nein“-Vorhersage und $a_{V,T}$ die Kosten eines Treffers bzw. eines verpassten Ereignisses sind. Im Gegensatz zur kontinuierlichen Ereignisvariablen greifen die Ereignisse im binären Fall paarweise auf die gleichen Bereiche der Wahrscheinlichkeitsdichte zu (siehe Tabelle 8-1). Liegt die Beobachtung Y zum Beispiel im Intervall $]X_1, X_2]$, so kann die Wahrscheinlichkeit

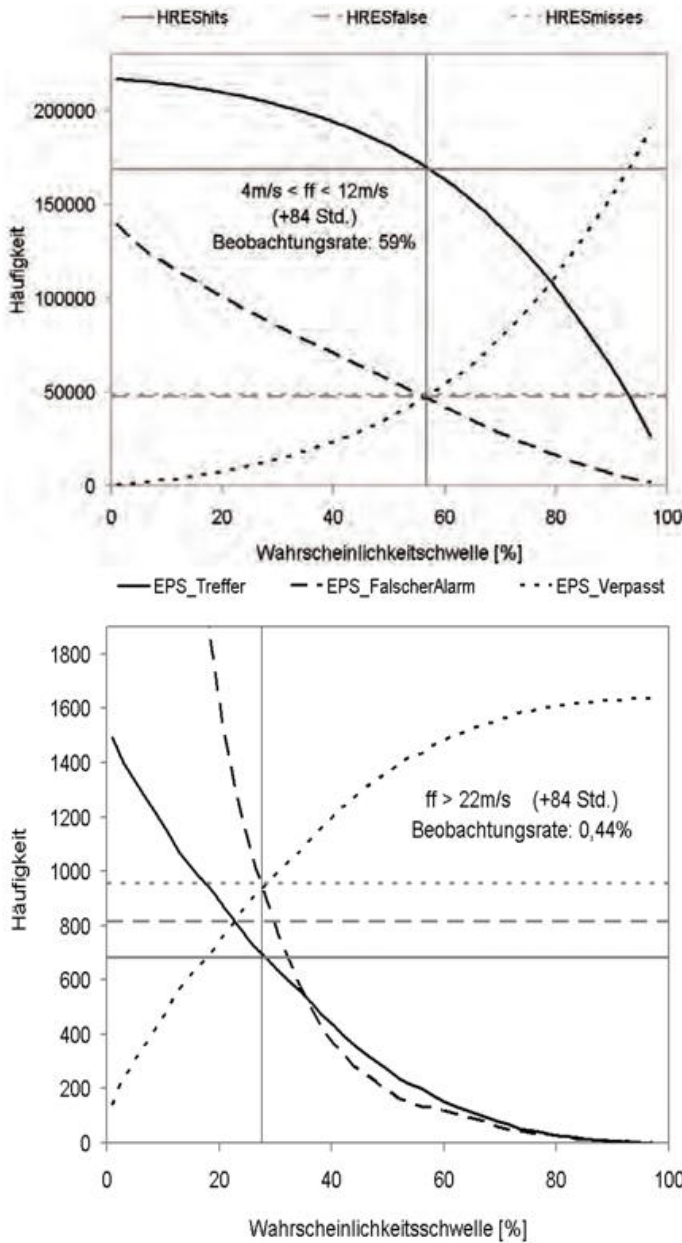


Abbildung 8-2: Darstellung der Abhängigkeit von Treffern, falschen Alarmen und verpassten Ereignissen von der gewählten Wahrscheinlichkeitsschwelle für die Ereignisse „10 m-Windgeschwindigkeit zwischen 4 und 12 m/s“ (obere Grafik) und „10 m-Windgeschwindigkeit größer 22 m/s“ (untere Grafik), jeweils für eine Vorhersagezeit von +84 Stunden. Als Referenz werden die Analysen des EZMW verwendet und gitterpunktweise verifiziert. Die vertikalen Linien zeigen die Wahrscheinlichkeitsschwelle bei biasfreier Vorhersage (Frequency Bias =1).

$$p = P(Y \in]X_1, X_2]) = \int_{y>X_1}^{X_2} p(y) dy \tag{6}$$

die Beobachtung im Ereignisintervall zu finden, sowohl einen Treffer als auch ein verpasstes Ereignis auslösen. Zur Unterscheidung wird eine Wahrscheinlichkeitsschwelle P_T benötigt, die überschritten werden muss, damit die probabilistische Vorhersage ein „ja“ für das Eintreten des Ereignisses liefert und ein Treffer entstehen kann. Überschreitet p die Wahrscheinlichkeitsschwelle nicht, wird das beobachtete Ereignis verpasst.

Die Abbildung 8-2 zeigt als Beispiel, die Anzahl der Treffer (T), der falschen Alarme (F) sowie der verpassten Ereignisse (V) in Abhängigkeit von der Wahrscheinlichkeitsschwelle P_T . Dargestellt sind zwei Ereignisse, $4m/s < ff < 12m/s$ (obere Grafik) und $ff > 22m/s$ (untere Grafik), für eine 84-Stunden-Vorhersage mit einem deterministischen (hochauflösender Lauf des EZMW, HRES) und einem probabilistischen Vorhersagesystem (EZMW-EPS). Die Werte der deterministischen Vorhersage sind als gestrichelte Linien parallel zur Ordinate dargestellt.

Für das gut vorhersagbare Ereignis $4m/s < ff < 12m/s$ (obere Grafik der Abbildung 8-2) liefert das deterministische Modell etwa gleich viele falsche Alarme (rot) und verpasste Ereignisse (gelb) mit einem deutlichen Vorteil der Treffer (blau) im Vergleich zu den Fehlvorhersagen. Das Verhalten von HRES wird vom Ensemble bei einer Wahrscheinlichkeitsschwelle (schwarze senkrechte Linie) von etwa 58% reproduziert. Dort haben die Vorhersagen keinen sogenannten frequency bias, da es gleich viele falsche Alarme und verpasste Ereignisse gibt. Für einen Nutzer ist eine biasfreie Vorhersage dann sinnvoll, wenn falsche Alarme und verpasste Ereignisse gleichwertig sind.

Da ein Ereignis zu verpassen aber oft stärkere Konsequenzen nach sich zieht, als ein falscher Alarm, sind die in Abbildung 8-2 eingezeichneten Schwellen nicht unbedingt optimal. So könnte man zum Beispiel für das Ereignis $4m/s < ff < 12m/s$ (obere Grafik in Abbildung 8-2) eine im Vergleich zur eingezeichneten Schwelle (schwarze senkrechte Linie) kleinere Wahrscheinlichkeitsschwelle verwenden. Dies würde zu einer größeren Zahl der Treffer und zu einer Verringerung der verpassten Ereignisse führen. Gleichzeitig erhöht sich aber die Zahl der falschen Alarme (Überwarnung). Ein falscher Alarm ist zwar nicht unbedingt willkommen, wird aber eher akzeptiert, da nur die Vorsorgemaßnahme umsonst war, mögliche schwerwiegende Folgen des Ereignisses aber ausgeblieben sind.

Das untere Bild in Abbildung 8-2 zeigt, dass die Optimierung der Wahrscheinlichkeitsschwelle für seltene Ereignisse, die weniger gut vorhersagbar sind, durchaus schwierig werden kann. Für das Ereignis „Mittlere Windgeschwindigkeit größer 22 m/s“ wird eine biasfreie Vorhersage nur durch das Ensemble bei einer Wahrscheinlichkeitsschwelle von 27% erreicht. Es könnte hier vorteilhaft sein, die Schwelle für die Wahrscheinlichkeit etwas über die 30%-Marke hinaus zu erhöhen, da sich dabei die Rate der falschen Alarme extrem stark reduziert (rote Kurve). Voraussetzung ist allerdings, dass die dabei steigende Zahl der verpassten Ereignisse und die damit einhergehende Reduzierung der Treffer akzeptiert werden kann.

Die richtige Schwelle P_T für eine Entscheidung hängt daher von verschiedenen Kriterien ab, wobei die Kostenstruktur des Problems eine entscheidende Rolle spielt. Diese wird durch das sogenannte „Cost-Loss ratio“ $\alpha=C/L$

bestimmt, wobei C die Kosten der Vorsorgemaßnahme und L der durch Vorsorge vermeidbare Verlust ist und $0 < \alpha < 1$ gilt. Bei näherer Betrachtung stellt sich heraus, dass das Kosten-Nutzen-Verhältnis im Wesentlichen die relative Gewichtung der falschen Alarme und der verpassten Ereignisse regelt, wie im Regenschirmbeispiel aus dem vorigen Abschnitt beschrieben. Es kann deshalb auch qualitativ einfach durch den Vergleich der Wichtigkeit beider Fehlvorhersagen für den Entscheidungsprozess bestimmt werden, ohne konkrete Kosten angeben zu müssen.

Treffer und korrekte „Nein“-Entscheidungen sind richtige Vorhersagen, die in jedem Fall nützlich sind, obwohl zum Beispiel im „Economic Value“-Ansatz (RICHARDSON 2000) Treffer mit den Kosten der Vorsorge und den möglicherweise nicht kompensierbaren Verlust durch das Ereignis belastet werden. Die Treffer geraten dadurch in die Rolle eines unerwünschten Ausgangs der Vorhersage, was nicht immer Sinn macht. Dies wird insbesondere dann wichtig, wenn eine Vorhersage oder Warnentscheidung durch ihre Qualität, das heißt vor allem durch eine hohe Treffsicherheit, die Aufmerksamkeit des Nutzers gewinnen muss.

5 Fazit

Eine Wettervorhersage ist nur dann vollständig, wenn sie den *true belief* über eine zukünftige Entwicklung wiedergibt. Dieser beinhaltet alle möglichen Szenarien der Wetterentwicklung mit ihren Eintrittswahrscheinlichkeiten und ist elementar probabilistisch. Die Bestrebungen des Meteorologen, den *true belief* einer Wettersituation möglichst umfassend und genau zu beschreiben, wird in Zukunft immer mehr durch numerische Ensemblesysteme unterstützt. In Anbetracht der Tatsache, dass von den Wetterdiensten schon heute einige hundert numerische Modellvorhersagen pro Tag produziert werden, stellt diese Aufgabe Mensch und Maschine vor neue Herausforderungen und wird notwendigerweise die Automation des Vorhersageprozesses weiter vorantreiben. Die Entwicklung wird dabei weg vom Mensch-Maschine-Konflikt (BALZER 2002) hin zur Mensch-Maschine-Kooperation führen, um den *true belief* gemeinsam möglichst nahe an die Wahrheit zu bringen. Denn wie die numerische Vorhersage selbst ist auch die Schätzung der Vorhersageunsicherheit mit Ensemblesystemen keineswegs perfekt.

Den Nutzern von Wetterinformation wird in Zukunft mehr Verantwortung für die Entscheidungsfindung bei

wetterabhängigen Produktionsprozessen und im Katastrophenschutz zukommen. Auf der Basis eines vom Provider der Wetterinformation kommunizierten *true belief* können Entscheidungsprozesse dort optimiert werden, wo sie stattfinden, so dass Wetter- und Klimainformation eine Steigerung ihres Nutzungsgrads erfahren. Die Verwendung probabilistischer Vorhersagen in der Anwendung wird zusätzliche Abstimmungsprozesse zwischen den Nutzern erfordern. Beispielsweise sollten Polizei und Feuerwehr für die gemeinsamen Einsatzfelder eine identische Interpretation der probabilistischen Vorhersagen verwenden, damit der jeweils andere nicht allein am Einsatzort erscheint (pers. Kommentar U. Ulbrich).

Numerische Ensemblesysteme ermöglichen eine sehr genaue raumzeitliche Darstellung der Vorhersageunsicherheit. Wie und über welche Medien lassen sich diese Informationen kommunizieren? Wie ist die Akzeptanz für Unsicherheit in der Öffentlichkeit? Wird ein Mehr an Unsicherheit nicht eher zu einer ablehnenden Haltung bei den Nutzern führen? Für diese und andere in der Zukunft wichtige Fragestellungen wurde 2011 im Rahmen des Hans-Ertel-Zentrums für Wetterforschung das Schwerpunktthema „Optimale Nutzung von Informationen aus Wettervorhersage und Klimamonitoring für die Gesellschaft“ an der Freien Universität Berlin eingerichtet (www.dwd.de/ertel-zentrum).

Literatur

- BALZER, K., 2002: Zum Mensch – Maschine-Konflikt in der Wettervorhersage. *Promet* **28**, 46-54.
- JOLLIFFE, I. T., 2008: The impenetrable hedge: a note on propriety, equitability and consistency. *Meteorol. Appl.* **15**, 25-29
- KRENOVSKI, S., ANGER, B., BÄHNKE, K., HOFFMANN, J., ENGEL, A., 2010: Struktur und Methodik der operationellen Wettervorhersage im Deutschen Wetterdienst (DWD). *Promet* **35**, 12-22.
- RICHARDSON, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **126**, 649-667.
- SMITH, L. A., ROULSTON, M. S., VON HARDENBERG, J., 2001: End to end ensemble forecasting: Towards evaluating the economic value of the Ensemble Prediction System. *Technical Memorandum* **336**, ECMWF, Shinfield Park, Reading UK.

P. BISSOLLI, P. HECHLER, S. RÖSNER, H. NITSCHKE

Ein Zentrum für europaweite Klimainformationen: Das WMO-Regionalzentrum Europa

A centre for European climate information: The WMO Regional Climate Centre (RCC) for Europe

Zusammenfassung

Regionale Klimazentren (engl. **R**egional **C**limate **C**entres, **RCCs**) sind Institutionen, die gemäß Mandat der Weltorganisation für Meteorologie (engl. **W**orld **M**eteorological **O**rganization, **WMO**, www.wmo.int) qualitativ hochwertige regionale Klimainformationen entwickeln und operationell bereitstellen. Sie stellen eine wichtige regionale Infrastruktur in dem 2009 durch die Weltklimakonferenz 3 beschlossenen Globalen Rahmenwerk für Klimadienstleistungen (engl. **G**lobal **F**ramework for **C**limate **S**ervices, **GFCS**) dar (http://www.wmo.int/pages/gfcs/index_en.php). Im Jahr 2009 wurde ein RCC-Netzwerk aus drei Konsortien für Europa und den Nahen Osten zunächst für eine Pilotphase eingerichtet. Dabei übernahm Deutschland neben der Funktion des Netzwerkkoordinators auch die Verantwortung für den Bereich Klimamonitoring. Der Beitrag skizziert das WMO-RCC-Konzept, stellt das europäische RCC-Netzwerk vor und beschreibt ausführlich die wichtigsten Klimaüberwachungsprodukte.

Abstract

Regional Climate Centres are institutions under mandate of the World Meteorological Organization (www.wmo.int, **WMO**), which develop and provide operational high-quality regional climate information. They represent an important regional infrastructure in the Global Framework for Climate Services (**GFCS**), adopted by the World Climate Conference 3 in 2009. In the same year, an RCC network of three consortia for Europe and the Middle East was first established for a pilot phase. Germany took over in addition to the function of the network coordinator also the lead of the RCC node of climate monitoring. This paper outlines the WMO RCC concept, presents the European RCC network and describes in some detail the most important climate monitoring products

1 Einführung

Meteorologische Prozesse sind durch Wechselwirkungen über alle Skalen und Ländergrenzen hinweg gekennzeichnet. Die Weltorganisation für Meteorologie trägt diesem Umstand Rechnung, in dem sie eine dreigliedrige Infra- und Servicestruktur mit globalen, regionalen und nationalen Elementen betreibt. Die globale Skala wird hierbei von einigen designierten Zentren sowie großen Wetterdiensten

abgedeckt; die nationale Skala ist in der Regel durch die nationalen Wetterdienste repräsentiert. Auf der regionalen Skala wurde die Erde im WMO-Kontext in sechs Regionen unterteilt: Afrika, Asien, Südamerika, Mittel- und Nordamerika, Australien und Ozeanien sowie Europa (einschließlich Naher Osten); die Antarktis ist mit Blick auf ihren völkerrechtlichen Sonderstatus nicht in diese Regionalstruktur eingebunden (siehe Abb. 1).

Innerhalb dieser Regionen betreiben die Staaten diverse Netzwerke und Institutionen; für Europa seien exemplarisch EUMETSAT, die europäische Organisation zum Betreiben meteorologischer Satelliten in Darmstadt, Deutschland, und das EZMW, Europas Zentrum für mittelfristige Wettervorhersage in Reading, Großbritannien, genannt. Ein weiteres klimarelevantes Infrastruktur- und Serviceelement stellen die kürzlich definierten Regionalen Klimazentren der WMO dar, die im Folgenden am Beispiel des europäischen RCC-Netzwerks vorgestellt werden. Solche regionalen Institutionen und Netzwerke dienen den Zielen, möglichst vielen Mitgliedsländern kontinentweite konsistente Informationen bereitzustellen, die nationale Bereitstellung meteorolo-



Abb. 1: Regionalstruktur der WMO.

gischer Services zu unterstützen und Doppelarbeit langfristig zu verhindern.

1.1 Historie

Spätestens nach dem starken El Niño-Ereignis 1997/98 mit seinen weltweit beobachteten Auswirkungen wurde deutlich, dass es für klimarelevante Prozesse und Phänomene jenseits der Wettervorhersage keine ausreichende operationelle Infra- und Servicestruktur zur Unterstützung der Mitgliedsländer gab. In Analogie zur äußerst erfolgreichen Weltwetterwacht – einem WMO-System zum weltweiten Austausch von meteorologischen Informationen mit dem Schwerpunkt Wetteranalyse und -vorhersage – beschlossen die WMO-Mitgliedsstaaten die Einrichtung von Infrastruktur- und Serviceelementen für klimarelevante Prozesse. Schwerpunkte bildeten zunächst die Einrichtung globaler Produktionszentren für Jahreszeitenvorhersagen (engl. **Global Producing Centres for Long-range Forecasting, GPCs**) sowie von Regionalen Klimazentren in den oben angegebenen WMO-Regionen. Beide Elemente sollen in enger fachlicher und technischer Zusammenarbeit alle WMO-Mitgliedsländer und Programme mit einem Satz definierter klimarelevanter Grundprodukte versorgen und gleichzeitig in die Lage versetzen, diese nutzbringend anzuwenden. Mit heutigem Stand existieren weltweit 12 GPCs und zwei asiatische RCCs in Peking, China, und Tokio, Japan. Weitere knapp 20 RCCs befinden sich in verschiedenen Identifizierungs- und Planungsstadien. Von diesen stehen derzeit zwei RCCs kurz vor ihrer offiziellen Designierung: Das europäische RCC-Netzwerk sowie das nordasiatische RCC.

1.2 Basisfunktionen

Verschiedene WMO-Arbeitsgruppen erarbeiteten unter führender Mitarbeit des DWD Richtlinien zur Einrichtung und zum Betrieb der RCCs. Im Sinne der Verpflichtung, alle WMO-Programme und damit alle Mitgliedsländer weltweit zu unterstützen, wurden auch operationelle Basisfunktionen und regelmäßig zu erstellende und zu verbreitende Pflichtprodukte definiert. Darüber hinaus wurde eine Liste besonders empfohlener Produkte und Services erarbeitet; die RCC-Kandidaten werden ermuntert, so viele dieser Leistungen wie möglich anzubieten. Mit diesem Ansatz wird den unterschiedlichen Möglichkeiten und Anforderungen in den oben genann-

ten Regionen Rechnung getragen und sogleich ein weltweit konsistenter Grundstandard realisiert.

Pflichtfunktionen der WMO-RCCs umfassen die Themenkomplexe Klimadaten, Klimaüberwachung, Jahreszeitenvorhersage sowie Training. Zu den regelmäßig zu erstellenden Pflichtprodukten gehören interpretierte saisonale Vorhersagen und zugehörige Verifikationsdatensätze, Klimaanalysen, Referenzklimatologien, ein Informationssystem zu Witterungsanomalien sowie geeignete Klimadaten-sätze. Die Pflichtservices umfassen die Bereitstellung der Produkte über ein Internetportal, die Verfügbarmachung von Methoden- und Prozessbeschreibungen zu den Produkten, Klimadatenmanagementunterstützung sowie die Organisation und Durchführung von Training.

Besonders empfohlene Funktionen schließen unter anderem auch die Bereiche Klimaprojektionen und regionale Koordinierung ein.

Diesen Funktionen ist zu entnehmen, dass RCC-Produkte und Services in allererster Linie für die nationalen Wetterdienste erbracht werden. Darüber hinaus ist die direkte Belieferung regionaler Endnutzer möglich, sofern alle Wetterdienste der Region zustimmen.

1.3 Das Europäische RCC-Netzwerk

Die Vorbereitungen für die Einrichtung von RCCs in Europa begannen vor gut einem Jahrzehnt. Aufgrund der relativ hohen Anzahl etablierter Wetterdienste in Europa wurde von vornherein ein Netzwerk favorisiert. Die Jahre des Aufbaus beinhalteten eine umfassende Analyse der

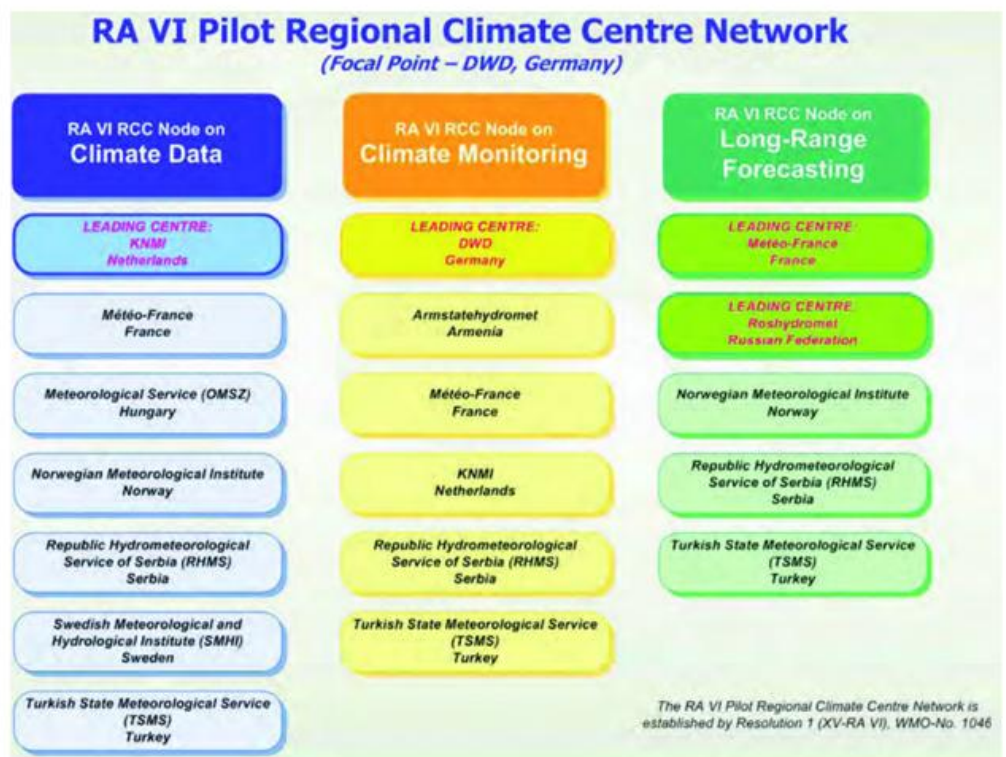


Abb. 2: Organisationsstruktur des europäischen RCC-Netzwerks.

Pflichtleistung	Produkt/Service	Produzent	Gebiet
Klimaüberwachung auf regionaler und subregionaler Ebene	Annual Bulletin on the Climate in RA VI	Koordinator: Deutscher Wetterdienst	Europa und Naher Osten
	Annual BAMS 'State of the Climate' Report; Europäisches Kapitel (http://www.ncdc.noaa.gov/bams-state-of-the-climate)		
	Monthly Bulletin on the Climate in RA VI	Deutscher Wetterdienst	Europa und Naher Osten; Wasserdampf: 60W-60E, 30N-80N; Wolkenbedeckung: 60W-60E, 30N-60N
	Monatskarten (Mittel/Summen und Anomalien) für Temperatur, Niederschlag, Luftdruck, Sonnenscheindauer, Schneebedeckung, Wasserdampf und Wolkenbedeckung		
	Modellbasierte Monatskarten für Temperatur, Niederschlag, Luftdruck, Wind und Luftfeuchte	Météo-France	Global, Europa und Naher Osten
	Monatskarten, -Diagramme und Trends verschiedener Klimaindizes für Temperatur und Niederschlag auf Grundlage des stationsbezogenen ECA/D (European Climate Assessment and Dataset) Datensatzes	KNMI, Niederlande	Europa und Naher Osten
	Modellbasierte Monatskarten für Temperatur und Niederschlag	Armstatehydromet, Armenien	Südkaucasus
	Stations- und modellbasierte Monatskarten für Temperatur und Niederschlag	TSMS, Türkei	Östliche Mittelmeerregion
Erstellung von Referenzklimatologien auf regionaler und subregionaler Ebene	Referenzklimatologien für Temperatur, Niederschlag, Luftdruck, Sonnenscheindauer, Schneebedeckung, Wasserdampf und Wolkenbedeckung	Deutscher Wetterdienst	Europa und Naher Osten
	Referenzklimatologien für Temperatur und Niederschlag sowie verschiedene Klimaindizes auf Grundlage des stationsbezogenen ECA/D Datensatzes	KNMI, Niederlande	
Einrichtung eines regionalen Informationssystems für Witterungsanomalien	Hinweise auf beobachtete und erwartete Witterungsanomalien (Climate Watch)	Koordinator: Deutscher Wetterdienst	Südosteuropa
		Koordinator: RHMS Serbien	
Produkt und Servicedokumentation sowie Nutzungshinweise	Wird von allen Produzenten für alle Produkte und Leistungen erstellt (z. Zt. im Aufbau); Hinweise über Newsletter des RCC-CM		
Koordinierung von Trainingsaktivitäten	Wird von allen Produzenten via Workshops, ELearning und Showcases realisiert (im Aufbau); Hinweise über Newsletter des RCC-CM		
Bereitstellung eines Online-Zugangs zu den RCC-Leistungen	Zugang zu allen Produkten und Leistungen via http://www.rccra6.org oder http://www.dwd.de/rcc-cm (Hinweis: Einige Produkte und Leistungen sind auf die exklusive Nutzung durch Nationale Wetterdienste beschränkt und daher passwortgeschützt; Interessenten sollten sich an ihren Nationalen Wetterdienst wenden.)		
Nutzerfeedback	Realisierung via Internet und Umfragen, unter anderem auch über den Newsletter des RCC-CM Knotens		

Tab. 1: Produktpalette des RCC-CM.

Anforderungen an solch ein europäisches Netzwerk, eine Analyse der Fähigkeiten der europäischen Wetterdienste, an solch einem Netzwerk teilzunehmen sowie schließlich die detaillierte Verhandlung der Netzwerkstruktur mit all seinen technischen, politischen und administrativen Facetten. Dieser vom Deutschen Wetterdienst maßgeblich angeleitete Prozess führte im Jahr 2009 zum Beginn der **Pilotphase des RCC-Netzwerks**. Die offizielle Designierung durch den Exekutivrat der WMO ist für 2013 vorgesehen.

In Anlehnung an die oben beschriebenen RCC-Basisfunktionen besteht das europäische RCC-Netzwerk aus drei funktionsbezogenen Knoten, welche zur Zeit von Wetterdienstkonsortien betrieben werden (Abb. 2).

Das Netzwerk wird zur Zeit vom Deutschen Wetterdienst koordiniert und ist via www.rccra6.org im Internet erreichbar.

1.4 RCC-Netzwerkknotten für Klimaüberwachung

Am Beispiel des vom Deutschen Wetterdienst geleiteten RCC-Knotens für Klimaüberwachung (engl. RCC Node on Climate Monitoring, RCC-CM) soll die Rollenverteilung im Konsortium exemplarisch beschrieben werden; ähnlich sind auch die anderen beiden Knoten organisiert. Grundsätzlich stellt der DWD als Leitinstitution des RCC-CM **alle RCC-Pflichtfunktionen für die gesamte Region** bereit. Der französische und der niederländische Wetterdienst (Météo-France und KNMI) vervollständigen dieses Grundangebot mit europaweiten spezialisierten Monitoringprodukten auf der Basis von Modellanalysen und speziellen europäischen Datensätzen. Die Wetterdienste von Armenien (Armstatehydromet), Serbien (RHMS) und der Türkei (TSMS) bieten verfeinerte Produkte und Dienste für die Subregionen Südkaukasus, Südosteuropa und Östliches Mittelmeer an. Tabelle 1 gibt einen Überblick über die Produktpalette des RCC-CM.

Über die Pflichtprodukte hinaus bieten die beteiligten Konsortialpartner des RCC-CM weitere Produkte und Leistungen an. Diese entsprechen zunächst den Kompetenzschwerpunkten der Partner. So bietet der DWD beispielsweise ein Dürremonitoring auf der Basis eines modifizierten **Standardized Precipitation Index (SPI)** an sowie Karten und Berichte zu extremen Ereignissen.

Das RCC-CM bietet aber nicht nur zentral erstellte Produkte sondern auch eine Plattform für nationale Karten und Berichte aus den 50 europäischen Wetterdiensten (einschließlich Naher Osten). Diese seit gut 10 Jahren existierende Plattform stellt neben dem oben genannten „Annual Bulletin on the Climate in RA VI“ (Publikation seit 1994) die historische Basis für die Entwicklung des RCC-CM ebenso dar wie für das entsprechende Engagement des Deutschen Wetterdienstes. Beide Aktivitäten können als hervorragende und erfolgreiche Beispiele enger internationaler Zusammenarbeit über alle Grenzen hinweg – von Israel im Süden bis zur Arktis im Norden und von Grön-

land und den Azoren im Westen bis nach Kasachstan im Osten – genannt werden. Diese Aktivität verdeutlicht aber auch einmal mehr die Notwendigkeit einheitlicher nationaler Standards und Methoden im Klimamonitoring, um die Vergleichbarkeit dieser Routineinformationen im Rahmen einer übernationalen Zusammenschau zu gewährleisten. Hier besteht nach wie vor großer Handlungsbedarf wobei unter anderem zu beachten ist, dass es den verschiedenen Ländern und Wetterdiensten vielfach nicht leicht fällt, ihre vieljährige Praxis anzupassen.

Im Rahmen seiner Leitfunktion koordiniert der Deutsche Wetterdienst auch die Abstimmung im Konsortium des RCC-CM und treibt dabei unter anderem die Vereinheitlichung der Standardprodukte voran, gibt einen Newsletter heraus, koordiniert die Weiterentwicklung der Produkt- und Servicepalette und agiert als Ansprechpartner für Fragen der Klimaüberwachung innerhalb des europäischen RCC-Netzwerks.

2 Produkte

Im Folgenden sollen nun exemplarisch die einzelnen Produkte des RCC-CM beschrieben werden, wobei die Darstellung auf diejenigen Produkte beschränkt wird, die im DWD hergestellt werden. Unter Klimaüberwachung (auch als Klimamonitoring bezeichnet) versteht man die Verarbeitung von aktuellen Klimadaten, mit dem Ziel, aus diesen Daten die wesentliche Information zur räumlichen und zeitlichen Klimavariabilität zu extrahieren und darzustellen, und dies kontinuierlich, also im Rahmen eines Routinebetriebs. Die Klimaüberwachung eines RCC ist also kein Forschungsprojekt, sondern eine Daueraufgabe und wird daher von nationalen Wetterdiensten wahrgenommen. Dabei stützen sich die Produkte natürlich auf die Erkenntnisse der Klimaforschung.

Demzufolge umfassen die RCC-CM-Produkte zwei wesentliche Produktgruppen: zum einen Karten und Diagramme, mit denen die wesentliche Information möglichst anschaulich analysiert und visualisiert werden soll, und zum anderen auch Texte, in denen diese Information beschrieben, interpretiert und diagnostiziert werden soll, um aus diesen **Karten die Information herauszufiltern und soweit möglich zu erklären**. Eine Sonderstellung nimmt die sogenannte Klimawacht (engl. Climate Watch) in der Produktpalette ein. Hier geht es speziell um Frühwarnungen vor zukünftigen Ereignissen, wobei aber die übrigen Klimaüberwachungsprodukte dazu als Grundlage dienen sollen. Alle RCC-CM-Produkte können im Internet über www.dwd.de/rcc-cm abgerufen werden, mit Ausnahme der Klimawacht-Texte, die nur an die Wetterdienste gesendet werden.

2.1 Karten für Europa

Karten sind ein geeignetes Mittel, um speziell die räumliche Variabilität des Klimas darzustellen. In einem klimatisch sehr heterogenen Gebiet wie Europa beziehungsweise

se der RA VI-Region, die sich von Grönland bis hin zum Nahen Osten erstreckt (vergleiche Abb. 1), ist die räumliche Klimavariabilität verhältnismäßig groß. Um die Besonderheiten eines einzelnen Monats für jeden Ort auf der Karte herausheben zu können, werden daher die Werte eines aktuellen Monats mit einer festen Referenzklimatologie verglichen. Die WMO empfiehlt dazu die 30jährige Standardnormalperiode, derzeit 1961-1990. Da nicht alle Daten bis 1961 zurückreichen (zum Beispiel Satellitendaten), werden stattdessen auch andere Referenzperioden verwendet.

Die zeitliche Variabilität wird meist in Form von Zeitreihendiagrammen (auf der x-Achse ist die Zeit, auf der y-Achse sind die Werte des jeweiligen Klimaelements aufgetragen) dargestellt oder auch zum Beispiel als Trendkarten (räumliche Variabilität der zeitlichen Trends). Derzeit konzentrieren sich die meisten Darstellungen des RCC-CM auf die räumliche Variabilität aktueller Monate; zusätzliche zeitliche Darstellungen sind jedoch in Entwicklung.

Der DWD stellt für verschiedene Klimaelemente monatliche, jahreszeitliche und jährliche Karten bereit, für die Maximum- und Minimumtemperatur sogar auf Tagesbasis. Als Datengrundlage werden zum einen Stationsdaten verwendet, die in die Fläche interpoliert werden, zum anderen Satellitendaten und Analysen von numerischen Modellen. Welche Datenbasis verwendet wird, hängt vom jeweiligen Klimaelement ab. Für einige Klimaelemente wie zum Beispiel die Lufttemperatur in 2 Meter Höhe über Grund liefern nach wie vor die Messdaten, die weltweit an Wetter- und Klimastationen sowie auf Schiffen erhoben werden, die genauesten Ergebnisse. Für andere Elemente, wie zum Beispiel die Bewölkung erscheinen die Satelliten als die beste Wahl, zumal sie flächendeckend beobachtet und nicht interpoliert werden müssen. Da der Satellit nicht direkt die Klimagrößen misst, sondern diese aus den Messdaten durch geeignete Algorithmen abgeleitet werden müssen, sind naturgemäß Unsicherheiten bei den Satellitendaten vorhanden. Diese lassen sich aber noch minimieren, indem Satellitendaten mit Stationsdaten verknüpft bzw. an diese angeeicht werden. Numerische Modellanalysen beinhalten ebenfalls sowohl Stations- als auch Satellitendaten, aber zusätzlich noch die Modellphysik, die zwar ein konsistentes physikalisches Abbild der Natur liefert, jedoch auch weitere Unsicherheiten durch numerische Vereinfachungen aufweist.

2.1.1 Karten aus Stationsdaten

2.1.1.1 Temperatur

Datenbasis für die Temperatur über den Landoberflächen sind die sogenannten CLIMAT-Stationsdaten, die weltweit über das Globale Fernmeldenetz verbreitet werden und somit allen angeschlossenen Wetterdiensten zur Verfügung stehen. Der CLIMAT Schlüssel enthält fertig berechnete Monatswerte von verschiedenen Klimagrößen sowie Anomalien von der Referenzperiode 1961-1990. Da

es sich bei CLIMATs ausschließlich um Monatswerte von einer überschaubaren Anzahl von Stationen handelt (etwa 800 Stationen werden für die RA VI-Region verwendet), liegen bereits 10 Tage nach Abschluss des Folgemonats die Stationsdaten einschließlich einer ersten Datenprüfung vor. Bis Ende des Folgemonats ist eine erweiterte Datenprüfung verfügbar. Die Karten werden sofort nach Eintreffen der Daten erstellt, so dass in der Regel nach 10 Tagen eine erste Kartenversion vorliegt. Zum Ende des Monats wird dann auf der Basis der zweiten Datenprüfung eine revidierte Karte erstellt, nachdem noch vorhandene Datenfehler korrigiert worden sind. Vor der Publikation der Karten erfolgt eine weitere visuelle Kontrolle, wobei gegebenenfalls noch verbleibende offensichtlich fehlerhafte Stationswerte entfernt werden.

Für die Wasserflächen, die bei der RA VI-Region einen erheblichen Anteil ausmachen, liegen Beobachtungen von fahrenden Handelsschiffen vor. Da die Schiffsbeobachtungen im Gegensatz zu den Landstationen nicht ortsfest sind, werden die jeweiligen Schiffsbeobachtungsorte (das heißt ihre Koordinaten zum Zeitpunkt der Beobachtung) in feste Gitterpunktsboxen der Größe 2,5° geografische Breite x 2,5° geografische Länge gemittelt. Damit kommen weitere rund 100 Stützstellen hinzu.

Für die Interpolation in eine Karte mit einer räumlichen Auflösung von 0,1° geografischer Breite und Länge hat es sich als sinnvoll erwiesen, die Stationspunkte nicht direkt in die Fläche zu interpolieren, sondern zunächst die geografischen Einflüsse aus den Stationsdaten herauszufiltern (zum Beispiel DOBESCH et al. 2007). Dies geschieht durch eine lineare Regression der Temperatur T_{reg} mit den Einflussgrößen (Prediktoren) x_1, x_2, \dots, x_n :

$$T_{reg} = a_1x_1 + a_2x_2 + \dots + a_nx_n + b \tag{1}$$

Als Prediktoren werden für die Temperatur die Größen geografische Breite, die Höhe über dem Meeresspiegel, die Kontinentalität und die mittlere Jahresamplitude verwendet, also alles einmalig feststehende Größen. Breite und Höhe sind als Stationskoordinaten bekannt, die Jahresamplitude wird vereinfacht aus der Differenz der Normaltemperatur Juli minus Januar 1961-1990 berechnet. Für die Kontinentalität wird ein Index verwendet, der seinerseits nur von den Größen geografische Breite und Jahresamplitude (nichtlinear) abhängt und somit aus diesen berechnet werden kann. Die geografische Breite und die Jahresamplitude gehen damit sowohl explizit in den linearen Regressionsansatz also auch implizit als gekoppelte Größe in die Kontinentalität (beim Kontinentalitätsindex steht die Jahresamplitude im Zähler, die geografische Breite im Nenner) mit ein. Dies führt zu einer besseren Erfassung der Abhängigkeit von den Prediktoren als eine reine lineare Abhängigkeit. Im Sommer wird ein relativ hoher Anteil der räumlichen Temperaturvarianz durch die Kontinentalität erklärt, aber ein geringer Anteil durch die Jahresamplitude allein, während es im Winter umgekehrt ist (HOGEWIND und BISSOLLI

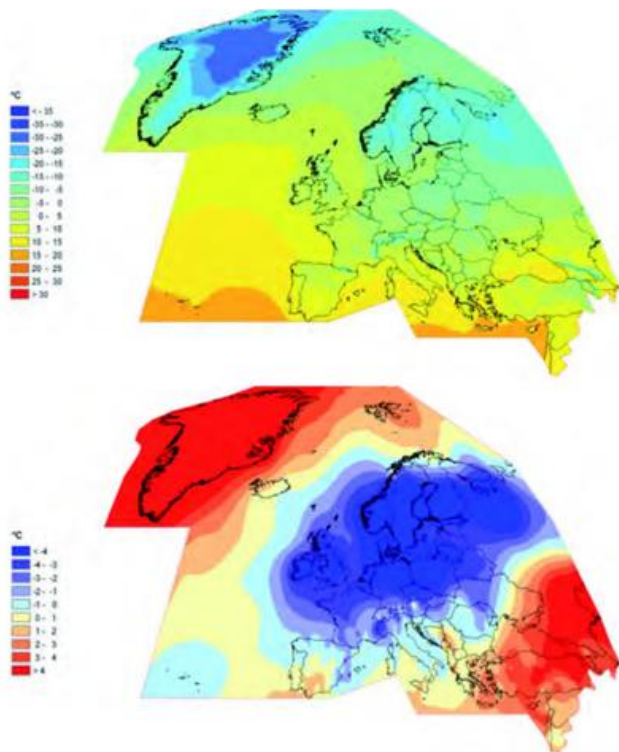


Abb. 3: Monatliche Mitteltemperatur für die RA VI-Region für den Monat Dezember 2010 (oben) und Anomalien vom Referenzzeitraum 1961-1990 (unten).

2011). Dies bedeutet, dass im Sommer die Abhängigkeit der Temperatur von der geografischen Breite besonders stark mit ihrer Jahresamplitude gekoppelt ist, im Winter dagegen nur wenig.

Damit liegen alle Werte sowohl von der Temperatur als auch von allen Prediktoren vor, und die Regressionskoeffizienten a_1, a_2, \dots, a_n, b können bestimmt werden. Da das Regressionsmodell in der Regel die Temperatur nicht perfekt beschreibt, unterscheidet sich die wahre Temperatur T von der durch Regression berechneten T_{regr} durch ein Restglied R , das auch als Residuum bezeichnet wird:

$$T = T_{\text{regr}} + R \quad (2)$$

Die Werte R an jeder Station stellen sozusagen die von den Prediktoren bereinigten (reduzierten) Temperaturen dar. Die räumlichen Unterschiede zwischen den einzelnen Stationen sind bei R in der Regel viel kleiner als bei T , so dass man bei der anschließenden räumlichen Interpolation glattere Interpolationsfelder erhält.

Für die Interpolation selbst existieren mehrere geeignete Interpolationsverfahren. International wird häufig ein Kriging-Verfahren empfohlen (DOBESCH et al. 2007). HOGEWIND und BISSOLLI (2011) haben jedoch gezeigt, dass ein anderes, weniger rechenaufwändiges Verfahren (Radial-Basisfunktionen) ähnliche, teilweise sogar noch etwas bessere Interpolationsergebnisse liefert.

Abb. 3 zeigt ein Beispiel für die Temperaturkarten für den Monat DEZEMBER 2010. Dieser Monat hatte sich durch

eine ausgeprägte Kälteperiode in weiten Teilen von Nord- und Mitteleuropa herausgehoben. Auf der Karte der Mitteltemperatur sind erwartungsgemäß einige grundlegende klimatische Eigenschaften zu erkennen, wie zum Beispiel eine Temperaturabnahme von Süd nach Nord, höhere Lufttemperaturen über Wasser im Vergleich zu nahegelegenen Landflächen, Kontrast der kälteren Gebirge im Vergleich zum Flachland. Daneben zeigen sich aber auch Merkmale, die nur für den speziellen Monat zutreffen, wie etwa die ausgeprägten niedrigen Temperaturwerte in Nord- und Mitteleuropa. Die entsprechende Anomaliekarte zeigt dies noch wesentlich deutlicher: die blaugefärbten Gebiete, die von der Kältewelle betroffen waren, lassen sich deutlich von den Erwärmungsgebieten über Grönland und über dem Nahen Osten abgrenzen. Die Ergebnisse weisen auch darauf hin, dass bereits eine relativ geringe Stationsdichte genügt,

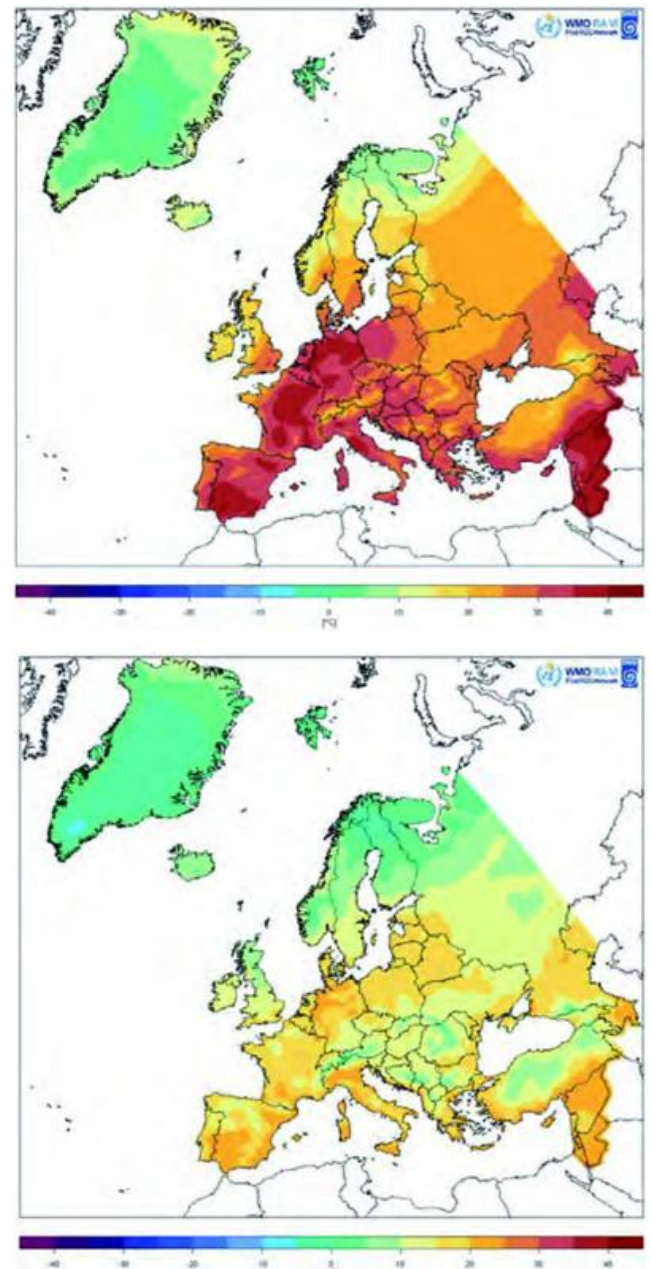


Abb. 4: Tägliche Maximumtemperatur für die RA VI-Region für den 19. August 2012 (oben) und Minimumtemperatur der darauffolgenden Nacht vom 19. auf den 20. August 2012 (unten).

um sowohl die wesentlichen klimatologischen Charakteristika als auch die monatlichen Besonderheiten zu erfassen, da die Temperatur im Vergleich zu vielen anderen Größen eine relativ hohe räumliche Repräsentanz aufweist.

Zeitlich gesehen schlagen stärkere Hitze- und Kältewellen durchaus auf Monatsmittelwerte durch, so dass diese in vielen Fällen einen Einblick in die Besonderheiten des jeweiligen Monats geben können. Für kurzzeitige Temperaturschwankungen oder auch, wenn man an täglichen Extrema interessiert ist, reicht die monatliche Auflösung nicht mehr aus, so dass tägliche Karten erforderlich werden. Daher wurde ein ähnliches Interpolationsverfahren auf tägliche Maximum- und Minimumtemperaturen von synoptischen Stationen angewendet (KRÄHENMANN et al. 2011). Für den 19. August 2012, einen der wärmsten Tage des Jahres in Mitteleuropa, zeigt Abb. 4 deutlich eine warme Zunge über Spanien, Frankreich bis nach Mitteleuropa mit Temperaturmaxima vielfach zwischen 35 und 40°C. Im gleichen Bereich lagen an dem Tag die Temperaturen über Nord- und Mittelitalien sowie im Nahen Osten. Dagegen waren die Temperaturen über der Balkan-Halbinsel, wo eine vorher über mehrere Wochen andauernde Hitze-welle allmählich am Abklingen war, nun etwas niedriger. Schön zu sehen ist außerdem die scharfe Temperaturgrenze zur Polarluft über Nordrussland und dem nördlichen Skandinavien. Die Minimumtemperaturen der darauffolgenden Nacht zeigen unter anderem Werte über 20°C (sogenannte tropische Nächte) vor allem über West- und Norddeutschland, wo nach dem vergangenen heißen Tag Bewölkung aufzog.

2.1.1.2 Sonnenscheindauer

Für die Monatssumme der Sonnenscheindauer, ebenfalls eine Standard-Klimagröße, die für viele praktische Anwendungen nach wie vor sehr gefragt ist, wird ein ähnliches Verfahren wie bei der monatlichen Temperatur angewendet. Auch hier liegen CLIMAT-Stationsmeldungen und Anomalien vor, jedoch keine Schiffsmeldungen, so dass diese Größe nur über Land berechnet wird. Auch das Regressionsverfahren ist gegenüber der Temperatur etwas vereinfacht, denn es werden nur die drei Prediktoren geografische Breite, Länge und Höhe verwendet. Ansonsten ist der Ablauf des Verfahrens der gleiche wie bei der Temperatur.

Als Beispiel sind in Abb. 5 die Sonnenscheinkarten für den Juli 2012 gezeigt. Charakteristisch ist für diesen Monat ein

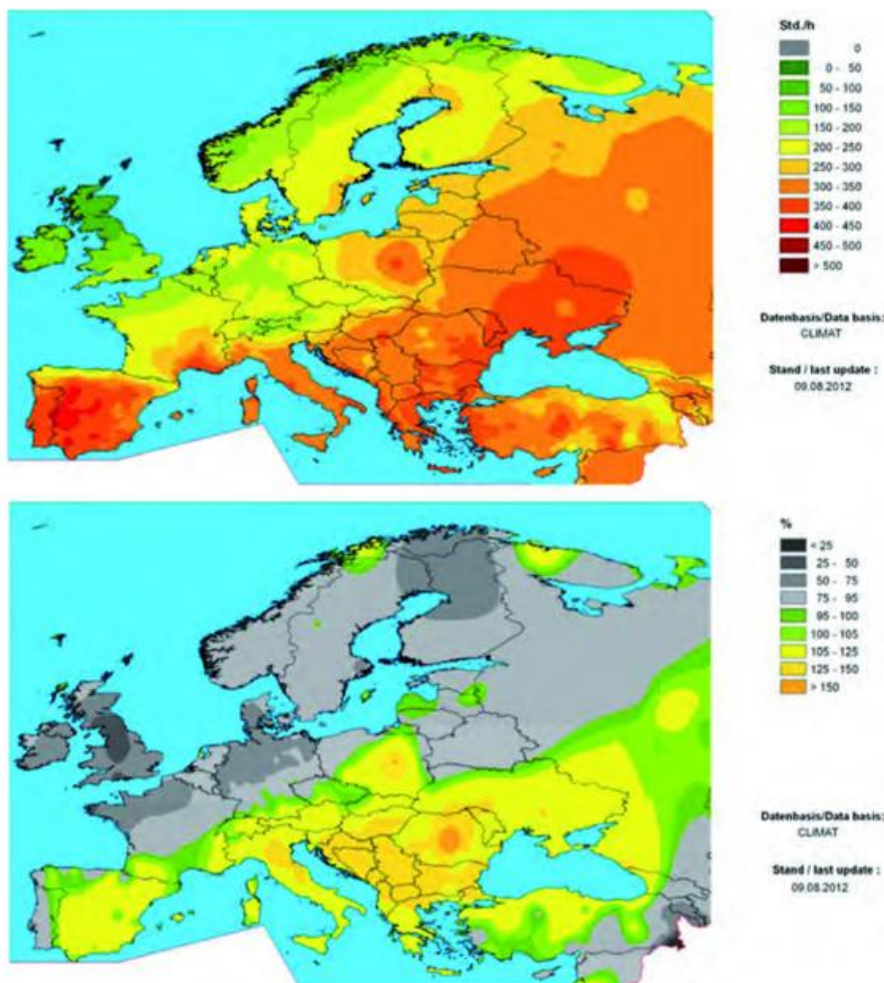


Abb. 5: Monatliche Sonnenscheindauer für die RA VI-Region für den Juli 2012 (oben) und Anomalien vom Referenzzeitraum 1961-1990 (unten).

Anstieg der Sonnenscheindauer von Nordwest nach Südost. Die dazugehörigen Anomaliekarten zeigen, dass die Sonnenscheindauer im Nordwesten deutlich unternormal, im Südosten deutlich überrnormal war, wobei ein Schwerpunkt außergewöhnlich hoher Sonnenscheindauer über dem Balkan lag, wo zu dieser Zeit eine längere Hitzewelle auftrat.

2.1.1.3 Niederschlag

Im Gegensatz zur Temperatur ist der Niederschlag eine Größe, die nur eine sehr geringe räumliche Repräsentanz aufweist, das bedeutet, die räumliche Niederschlagsvariabilität kann auch über kurze Entfernungen hinweg sehr groß sein, auch bei Monatssummen. Es werden daher möglichst viele Stationsdaten benötigt. Die wahrscheinlich beste Stationsdatenbasis über Land dürfte beim Weltzentrum für Niederschlagsklimatologie (WZN) vorliegen (<http://gpcc.dwd.de>). Sie beinhaltet nicht nur Stationswerte, die über das globale Fernmeldenetz verbreitet werden, sondern darüber hinaus auch noch Stationsdaten, die von den nationalen Wetterdiensten direkt an das WZN gemeldet werden. Das WZN verarbeitet die Daten, führt eine Datenprüfung durch und interpoliert die Stationsdaten in Rastern mit verschiedenen räumlichen Auflösungen. Nach etwa fünf Tagen sind bereits erste Raster auf einer Basis von mehr als 6 000 synoptischen, mit einem automatischen Verfahren

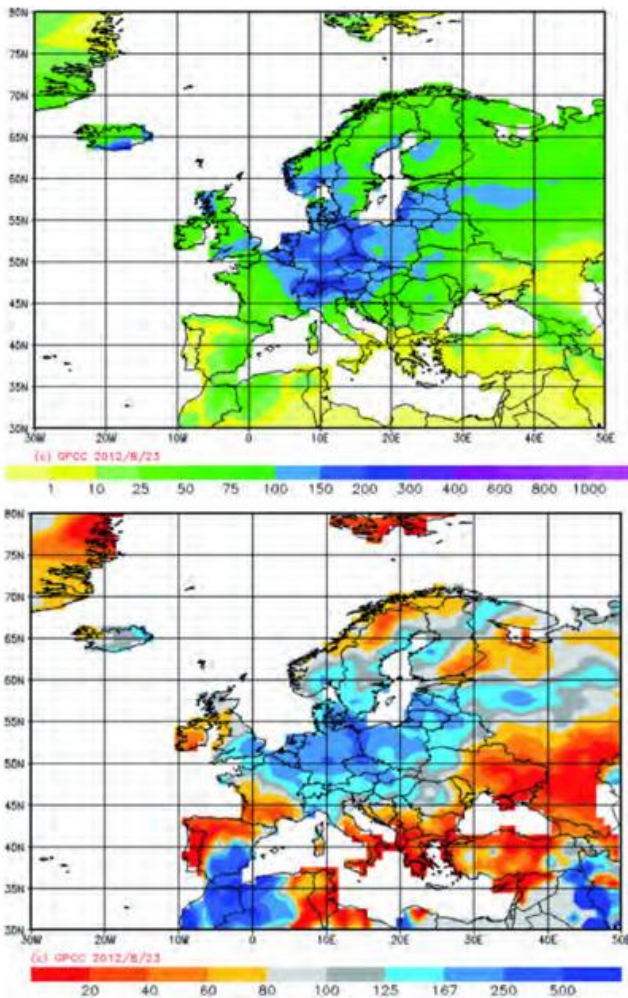


Abb. 6: Monatliche Niederschlagssumme für die RA VI-Region für den August 2010 (oben) und als Prozent vom Mittelwert August 1951-2000 (unten).

qualitätsgeprüften Stationen verfügbar (so genannte „First Guess-Analyse“). Nach etwa zwei Monaten liegt die sog. Monitoring-Analyse vor, die auf etwa 7 000 bis 8 000 Stationen und einer intensiveren manuellen Qualitätskontrolle basiert (SCHNEIDER et al. 2011). Beide Versionen werden mit der räumlichen Auflösung von 1° geografischer Breite und Länge direkt vom RCC-CM übernommen.

Im August 2010 fielen innerhalb weniger Tage Starkniederschläge im Dreiländereck Deutschland – Polen – Tschechien, die zu einer markanten Hochwassersituation besonders in den Einzugsbereichen der Oberläufe der Flüsse Elbe, Spree und Neiße führten (BISSOLLI et al. 2010). Die monatlichen WZN-Analysen (Abb. 6) zeigen deutlich den betroffenen Bereich mit Monatssummen über 200 mm beziehungsweise 250% vom vieljährigen Mittelwert, während gleichzeitig große Trockenheit im Südosten der RA VI-Region herrschte, verbunden mit einer ausgeprägten Hitzewelle mit Schwerpunkt über dem Süden des europäischen Teils von Russland (FRIEDRICH und BISSOLLI 2011).

2.1.1.4 Dürre

Wenn der Niederschlag über längere Zeit weit unter dem Normalwert liegt, spricht man von einer Dürreperiode. In

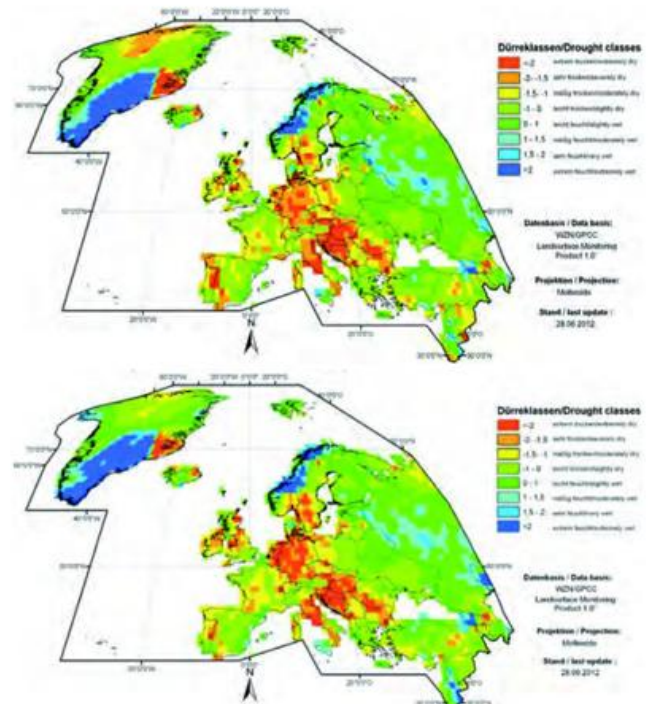


Abb. 7: Monatswerte für den Dürreindex SPI-DWD nach PIETZSCH und BISSOLLI (2011) für den Monat März 2012 (oben) und Anomalien dazu (unten).

der Literatur gibt es eine Reihe von Indizes, die versuchen, dieses Phänomen Dürre nach ihrer Stärke zu klassifizieren. Ein häufig verwendeter und international sehr empfohlener Dürreindex ist der so genannte SPI (Standardized Precipitation Index), der ausschließlich vom Niederschlag abhängt. Hierfür werden die Werte einer Zeitreihe an einem Datenpunkt durch eine Gamma-Verteilung beschrieben, die dann in eine Standardnormalverteilung transferiert wird. Die z-Werte (Abszissenwerte) der Standardnormalverteilung stellen dann die Dürreklassen dar. Entsprechend der Glockenform der Standardnormalverteilung werden nur seltene Ereignisse mit hohen Abweichungen vom Mittelwert als extreme Dürren beziehungsweise extrem niederschlagsreiche Ereignisse bezeichnet.

Dieses Verfahren bewährt sich auch für Europa und die RA VI-Region mit Ausnahme der ariden Gebiete im Mittelmeerraum im Sommer, wenn die Niederschlagsdaten nicht der Gamma-Verteilung folgen. Daher wurde im DWD eine modifizierte Version des SPI entwickelt (PIETZSCH und BISSOLLI 2011). Die wesentliche Modifikation besteht aus einer sog. Nullkorrektur, wobei Monate, in denen kein oder kaum Niederschlag gefallen ist, generell mit einem Indexwert für extreme Dürre belegt werden. Mit Abnahme der Anzahl der extrem niederschlagsarmen Monate geht der Index stetig in den ursprünglichen SPI über, so dass für die meisten Teile des RA VI-Gebiets die ursprüngliche und die modifizierte Version des SPI ganz oder nahezu identisch sind.

Abb. 7 zeigt ein Beispiel für die Dürre im Frühjahr 2012, die weite Teile von Südwest-, West- und Mitteleuropa erfasst hatte und im März 2012 vor allem über Deutschland und Südosteuropa einen Höhepunkt erreichte. In einigen Gebieten wird sowohl für die Monatswerte selbst als auch für die

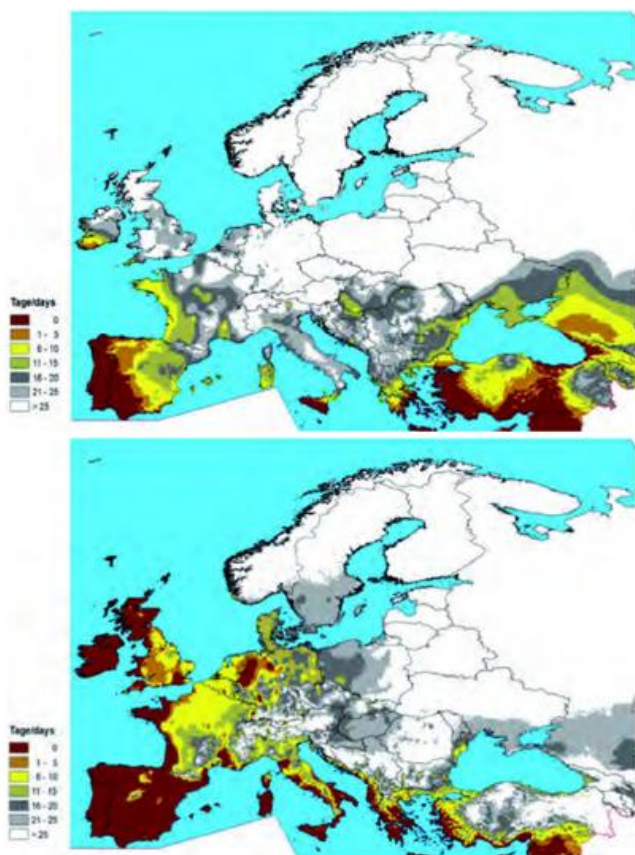


Abb. 8: Monatliche Anzahl der Tage mit einer Schneehöhe von mindestens 1 cm für die Monate Dezember 2010 (oben) und Februar 2012 (unten) .

Anomalien diese Dürre als extrem eingestuft. Weitere Einzelheiten zu diesem Dürreereignis wurden von BISSOLLI et al. (2012) an anderer Stelle beschrieben.

2.1.1.5 Schneebedeckung

An vielen synoptischen Stationen wird täglich die Gesamtschneehöhe gemeldet. Diese Daten können für ein Schneemonitoring genutzt werden. Angeboten werden hierzu vom RCC-CM Karten für die monatliche mittlere und maximale Schneehöhe, die Anzahl der Tage im Monat mit mindestens 1 cm Gesamtschnee und den ersten und letzten Schneetag einer Saison vom 1. September bis 31. August des Folgejahres. Neben den Monatskarten wird derzeit auch ein tägliches Schneemonitoring aufgebaut. Da dem DWD erst ab 1981 eine ausreichende Anzahl an Schneedaten von synoptischen Stationen aus dem Ausland vorliegen, konnten keine vieljährigen Mittelwerte für den Zeitraum 1961-1990 berechnet werden und damit auch keine Anomalien. Es ist jedoch eine Klimatologie für den neueren Zeitraum 1981-2010 in Arbeit.

Die oben genannten Größen werden für jede einzelne Station berechnet; anschließend wird darauf das gleiche Regressions- und Interpolationsverfahren wie für die Sonnenscheindauer angewendet (vergleiche oben). Da die Schneegrößen je nach Höhenlage und Wettersituation fast ebenso wie der Niederschlag räumlich sehr stark variieren können, ist eine möglichst große Stationsanzahl erforder-

lich. Die synoptischen Stationen sind wesentlich zahlreicher als die CLIMAT-Stationen; dies erfordert allerdings auch einen höheren Datenprüfungsaufwand, der gerade beim Schnee nicht unproblematisch ist. Neben zahlreichen Fehl- und Lückenwerten besteht das generelle Problem, dass in Fällen, in denen kein Schnee liegt oft auch keine Meldung abgesetzt wird, so dass von den Werten her nicht a priori entschieden werden kann, ob kein Schnee gefallen oder die Meldung ausgefallen ist. Daher wurde im DWD ein spezielles Datenprüfungsverfahren entwickelt, das jeden einzelnen Wert prüft, gegebenenfalls verwirft und auch Lückenwerte beziehungsweise verworfene Daten durch andere plausible Werte auffüllt. Dazu werden noch weitere Hilfsgrößen herangezogen, sofern vorhanden, wie die Schneehöhe vom Vortag, die Niederschlagshöhe und die Maximum- und Minimumtemperaturen. Insgesamt liegen somit jeden Monat meist zwischen etwa 2000 und 3000 Werte für die RA VI-Region vor. Beim Auffüllen lässt sich natürlich keine Zentimetergenauigkeit erreichen, allerdings übertrifft dieser Fehler nicht wesentlich die anderen Unsicherheiten wie zum Beispiel bedingt durch die Rasterung und die Interpolation. Es zeigt sich jedoch, dass die großräumigen Strukturen sinnvoll wiedergegeben werden.

Als ein interessantes Beispiel soll hier die Anzahl der Schneetage der beiden Monate Dezember 2010 und Februar 2012 gegenübergestellt werden (Abb. 8). Beide Monate waren durch eine ausgeprägte Kälteanomalie im überwiegenden Teil Europas gekennzeichnet, wenn auch mit unterschiedlichen geografischen Schwerpunkten. Deutliche Unterschiede gab es jedoch im Schneesverhalten: Im Dezember 2010 lag im gesamten Nord-, Mittel- und Osteuropa fast an allen Tagen oder sogar durchweg Schnee. Der Februar 2012 hatte dagegen nur an einigen, zum Teil nur wenigen Tagen in Mitteleuropa eine Schneedecke. Dies liegt daran, dass es im Februar 2012 zwar auch eine Kältewelle gab, die jedoch in Mitteleuropa nur die erste Hälfte des Monats andauerte und teilweise mit trockener kontinentaler Kaltluft aus Osten ohne Schneesiederschlag verbunden war. Dagegen blieb der Schnee in diesem Monat lange im Südosten der RA VI-Region liegen und zog sich dort auch sehr weit nach Süden; dies wiederum war im Dezember 2010 in diesem Ausmaß nicht der Fall.

2.1.2 Karten aus Satellitendaten

2.1.2.1 Wolkenbedeckung

Die Wolkenbedeckung ist ein klassisches Beispiel für eine Größe, die sich besonders gut durch Satelliten erfassen lässt. Satelliten messen die an Wolken reflektierte oder von ihnen emittierte Strahlung in verschiedenen Spektralkanälen. Da sich die Strahlungseigenschaften der Wolken von denen des Erdbodens meist gut unterscheiden lassen, lassen sich die Wolken lokalisieren. Die Satelliten haben zudem den Vorteil, dass sie in den nicht sichtbaren Spektralkanälen auch nachts gut und flächendeckend messen können, während ein Bodenbeobachter nachts die Wolken nur relativ schwierig erfassen kann.

Um von der eigentlichen Satellitenmessung auf die Wolkenbedeckung zu kommen, existieren verschiedene Algorithmen, mit denen die Datenhersteller die Wolkenbedeckung (Maßeinheit Achtel) berechnen. Das RCC-CM verwendet derzeit einen Datensatz des sogenannten „Satellitenwetters“ von ROSENOW et al. (2001), der routinemäßig im DWD erstellt wird. Dabei wird ein Cluster-Algorithmus auf die Satellitenmessungen angewendet. Die Clusterwerte werden wiederum an beobachtete Synop-Stationsmeldungen, die in Achtel angegeben werden, angeeicht. In das Endprodukt gehen somit sowohl Satelliten- als auch Stationsmessungen ein. Über Wasser, wo keine Stationsmeldungen vorliegen, kommt nur der Cluster-Algorithmus zum Tragen. Dort sind daher die Daten unsicherer als über Land. Für die Zukunft ist geplant, einen anderen Datensatz zu verwenden, der ausschließlich auf Satellitenmessungen basiert, nämlich die Daten der EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF, KASPAR et al. 2009), die ebenso wie das RCC-Netzwerk vom DWD geleitet wird.

Da die Daten des Satellitenwetters nur wenige Jahre zurückreichen, stellt sich die Frage nach einem geeigneten Referenzdatensatz. Aus dem International Satellite Cloud Climatology Project (ISCCP) der NASA (<http://isccp.giss.nasa.gov>) stehen Wolkendaten aus operationellen Wettersatelliten bereits seit 1983 zur Verfügung. Diese wurden im DWD aufbereitet, wiederum an Bodenbeobachtungen angeeicht und in ein $0,25^\circ \times 0,25^\circ$ -Gitter transformiert (BISSOLLI 2003).

Aus diesen Daten hat das RCC-CM geeignete Monitoringprodukte entwickelt. Neben den Anomalien, die analog zu den anderen Klimavariablen berechnet werden, kommen als weitere interessante Größen die Anzahl der heiteren und trüben Tage (Tage mit einer Wolkenbedeckung von weniger als 2/8 beziehungsweise mehr als 6/8) sowie die maximale Andauer von heiteren und trüben Tagen pro Monat hinzu. Daraus lässt sich ersehen, ob eine Wolkenanomalie durch eine generell höhere oder geringere Bewölkung zustande kam oder durch eine einzelne, begrenzte Periode mit sehr hoher bzw. sehr geringer Wolkenbedeckung.

Als Beispiel ist hier die Wolkenverteilung für den Februar 2012

dargestellt (Abb. 9). In diesem Monat entwickelte sich eine Dürreperiode, die zunächst die Iberische Halbinsel erfasst hatte und sich im März weiter nach Mitteleuropa ausbreitete (vergleiche oben im Abschnitt „Dürre“). Entsprechend war über der Iberischen Halbinsel und vor allem über den umgebenden Seegebieten im Mittel nur geringe Bewölkung vorhanden, während über dem übrigen Teil Europas größtenteils stärkere Bewölkung vorherrschte. Ein wesentlich kontrastreicherer und differenzierterer Bild zeigt dagegen die Anzahl der heiteren Tage: Während es im größten Teil Europas nur sehr wenige heitere Tage gab, präsentierten sich die Seegebiete um die Iberische Halbinsel als fast durchweg heiter. Die Iberische Halbinsel selbst hingegen weist punktweise teils mehr, teils weniger heitere Tage auf, also eine höhere räumliche Variabilität, somit Tage mit isolierter Bewölkung, die über den Landflächen teils auch orografisch bedingt entstanden war.

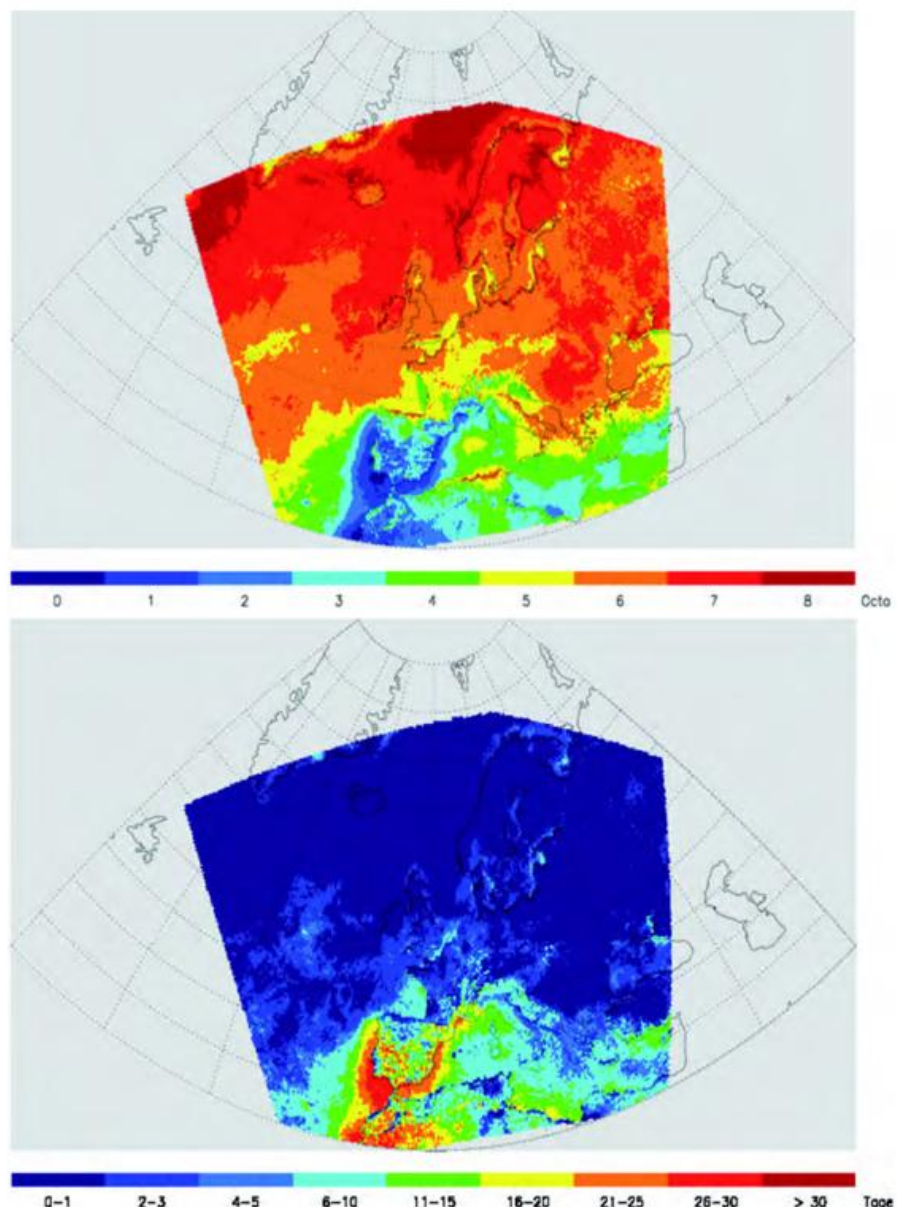


Abb. 9: Anzahl der heiteren Tage (oben) und mittlere Wolkenbedeckung (unten) im Februar 2012.

2.1.2.2 Wasserdampf

Um neben der Bewölkung auch die räumliche Luftfeuchteverteilung einschätzen zu können, bietet es sich an, zusätzlich den Wasserdampfgehalt in der Atmosphäre zu betrachten. Außerdem ist der Wasserdampf auch als Treibhausgas von klimatologischer Bedeutung.

Die Wasserdampfgröße, die im RCC-CM betrachtet wird, ist das gesamte niederschlagbare Wasser (engl.: **total precipitable water** oder kurz TPW). Es ist definiert als die Wassermenge, die man erhalten würde, wenn in einer Säule vom Erdboden bis zum Oberrand der Atmosphäre der gesamte dort enthaltene Wasserdampf auskondensieren würde. Das niederschlagbare Wasser ist somit ein Maß für die absolute Luftfeuchte, wird aber wie der Niederschlag in mm angegeben und bezieht sich wie gesagt auf eine vertikale Luftsäule integriert über die gesamte Atmosphäre und nicht nur auf ein einzelnes Luftvolumen in einer bestimmten Höhe.

Die Daten werden aus Satellitenmessungen des ATOVS-Instruments, das sich auf den amerikanischen NOAA-Satelliten befindet, abgeleitet und von der NOAA als sogenannte SATEM-Code-Meldung international verbreitet; sie stehen damit auch dem DWD zur Verfügung. Im RCC-CM werden Monatsmittel gebildet, die in einem 5°x5°-Raster zusammengefasst werden.

Anschließend werden diese Rasterdaten auf ein feineres 1°x1°-Gitter interpoliert. Als Referenzdatensatz (Zeitraum 1980-2005) werden ebenfalls ATOVS-Daten verwendet, für die frühen Jahre jedoch noch Daten des Vorgängerinstruments TOVS. Dadurch sind die Referenzdaten nicht ganz frei von Inhomogenitäten, so dass eine gewisse Unsicherheit in Kauf genommen werden muss. Wie bei den anderen Größen werden jedoch auch Anomalien berechnet. Um die Qualität der Daten bewerten zu können, werden zusätzlich Punktvalidierungen mit Radiosondenstationen routinemäßig durchgeführt.

Zur Veranschaulichung soll hier wieder analog zur Bewölkung der Monat Februar 2012 betrachtet werden (Abb. 10). Da das niederschlagbare Wasser ein Maß für die absolute Luftfeuchte ist, weist es eine hohe Abhängigkeit von der Temperatur auf (wärmere Luft kann mehr Feuchtigkeit aufnehmen als kältere). Daher zeigt sich auch bei den absoluten Monatswerten eine deutliche Abnahme von Süd nach

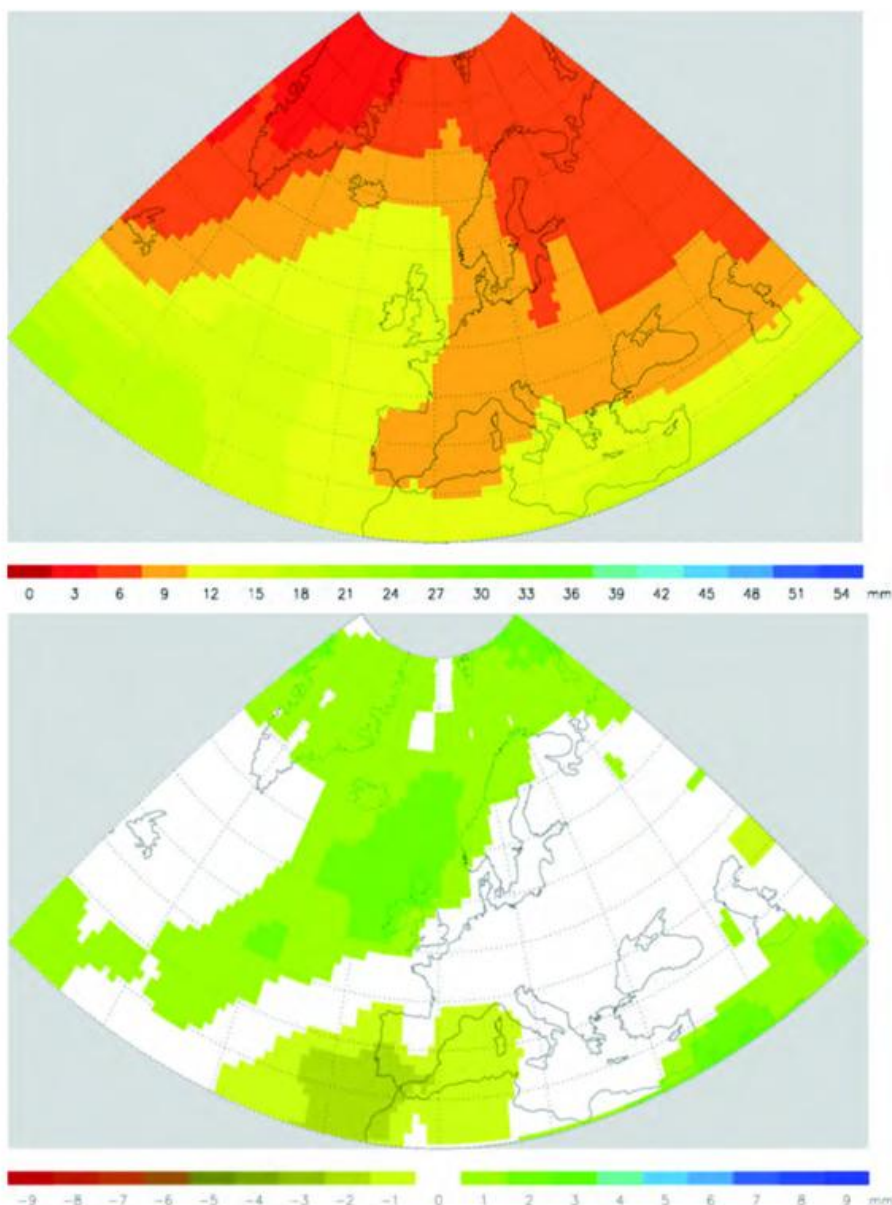


Abb. 10: Monatsmittel des niederschlagbaren Wassers für den Monat Februar 2012 (oben) und Anomalien vom Referenzzeitraum 1980-2005 (unten) .

Nord. Im Februar 2012 unterschied sich jedoch der Wasserdampfgehalt über Spanien und dem westlichen Mittelmeer kaum von dem über Mitteleuropa und dem westlichen Skandinavien, obwohl dort die Temperatur deutlich niedriger war. Entsprechend zeigen die Anomalien über Westskandinavien höhere und um die Iberische Halbinsel wie bei der Bewölkung niedrigere Werte als der Referenzwert. Damit lässt sich ganz konsistent von einem niedrigen Wasserdampfgehalt auf die geringe Bewölkung über dem südwesteuropäischen Dürregebiet schließen.

2.1.2.3 Strahlung

Eine weitere Größe, für die sich Satellitendaten gut eignen, sind Größen der Strahlungsbilanz, da auch hier wieder die Absorptions- und Reflexionseigenschaften verschiedener Oberflächen genutzt werden können, die der Satellit erfassen kann. Das Monitoring der Strahlungsbilanzgrößen soll dazu dienen, die Anomalien anderer Klimavariablen

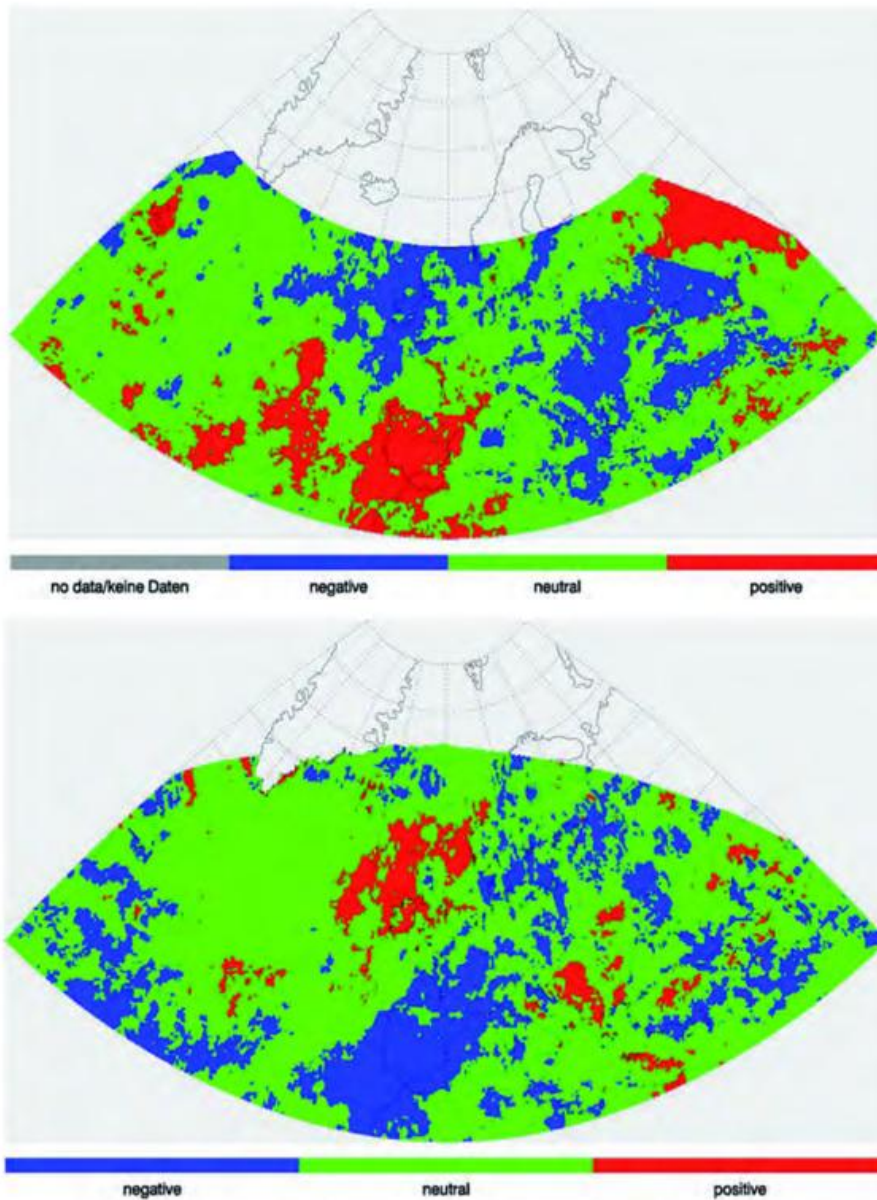


Abb. 11: Anomalien der kurzwelligen (oben) und langwelligen (unten) Strahlungsbilanz für den Monat Februar 2012 als Terzile von einer Wahrscheinlichkeitsverteilung über den Referenzzeitraum 1983-2012.

physikalisch, das heißt mit Hilfe der Energiebilanz erklären zu können. Damit wird dem Anspruch des RCC-CM Rechnung getragen, dass nicht nur ein rein beschreibendes, sondern auch ein physikalisch erklärendes Klimamonitoring im Sinne einer Klimadiagnose durchgeführt werden soll.

Im RCC-CM werden vier Größen der Strahlungsbilanz am Erdboden betrachtet: die einfallende kurzwellige Strahlung (Globalstrahlung), die kurzwellige Gesamtbilanz (Globalstrahlung minus Albedo), die vom Erdboden ausgehende langwellige Strahlung und die langwellige Strahlungsbilanz (Gegenstrahlung minus vom Erdboden ausgehende Strahlung). Datensätze werden über das oben genannte CM-SAF bezogen und sind näher in den einschlägigen Produktdokumentationen (Product User manuals) des CM-SAF beschrieben (www.cmsaf.eu). Als Referenzdaten existiert ein Datensatz aus dem GEWEX-

SRB-Projekt der NASA (<http://gewex-srb.larc.nasa.gov/>), der die Jahre 1983-2007 abdeckt. Diese Daten wurden an die CM-SAF-Daten für die neueren Jahre angeeicht, so dass der Referenzzeitraum bis hin zum aktuellen Jahr verlängert werden kann. Wegen der Verfügbarkeit beziehungsweise zu hoher Ungenauigkeiten der Satellitendaten sind insbesondere die polaren Breiten nicht abgedeckt.

Bei den Anomalien wird hier ein etwas anderer Weg gegangen als bei den übrigen Variablen. Anstelle von Differenzen zum Referenzdatensatz werden hier Terzile betrachtet, das heißt in den Karten wird nur festgehalten, ob sich ein Strahlungswert im oberen, mittleren oder unteren Terzil einer Wahrscheinlichkeitsdichteverteilung befindet, welche an die Daten des Referenzzeitraumes angepasst worden ist.

Ein Beispiel dazu ist in Abb. 11 zu sehen. Der Einfachheit halber ist hier wiederum der Februar 2012 ausgewählt worden. Die kurzwellige Strahlungsbilanz lag speziell für das Dürregebiet auf der Iberischen Halbinsel und die umgebenden Seegebiete im obersten Terzil, die langwellige Strahlung im untersten. Dies ist konsistent zur oben genannten geringen Bewölkung über diesem Gebiet. Kurzwellige Strahlung konnte weitgehend ungehindert den Erdboden erreichen, die Globalstrahlung war also sehr hoch, während der Erdboden ebenso

ungehindert langwellige Strahlung emittieren konnte. Da ein Strahlungsgenuss des Erdbodens positiv und ein Strahlungsverlust negativ gezählt werden, ergeben sich für die langwellige Strahlung hohe negative Werte und daher eine Klassifizierung für das unterste Terzil.

2.1.3 Luftdruckkarten aus numerischen Modellanalysen

Neben der Strahlungsbilanz können monatliche Klima-anomalien oft auch durch länger andauernde oder häufige Advektion großräumiger Luftmassen erklärt werden. Hierfür sind Kartendarstellungen der mittleren räumlichen Bodenluftdruckverteilung sehr hilfreich, aus denen vor allem die wesentlichsten steuernden Druckgebilde ersichtlich sind. Hierzu werden die täglichen numerischen 00 UTC-Luftdruckanalysen des globalen Wettervorhersagemodells GME des DWD monatlich gemittelt und ein Europa-Ausschnitt bereitgestellt.

Auch in dieser Hinsicht war der Februar 2012 ein interessanter Monat (Abb. 12). Die mittlere Luftdruckverteilung zeigt hier unter anderem, dass das Islandtief sehr weit nach Südwesten und im Gegenzug der Kern des Azorenhochs weit nach Nordosten verschoben war, so dass große Teile von West- und Mitteleuropa im Mittel unter Hochdruckeinfluss lagen. Speziell für die Iberische Halbinsel erklärt dies auch die geringe Bewölkung und die oben beschriebene Strahlungsbilanz. Die Anomalien sind in Nordwest- und Westeuropa deutlich positiv, das bedeutet, dass eine derartige Ausdehnung des Hochs in keinsten Weise dem klimatologischen Mittel für den Februar entspricht.

2.2 Berichte

Neben der Erstellung von Kartenmaterial erhebt das RCC-CM auch den Anspruch, dieses zu beschreiben und zu interpretieren und die Ergebnisse in Form von bebilderten Textberichten zu verbreiten. Diese Berichte sollen keine neuen Forschungsarbeiten sein, sondern lediglich bereits vorhandene Erkenntnisse und Material auf aktuelle Situationen anwenden.

Es werden zwei Formen von Berichten herausgegeben: zum einen regelmäßige monatliche und jährliche Klimarückblicke und zum anderen unregelmäßige Analysen von aktuellen besonderen Wetterereignissen, die bei Bedarf erstellt werden.

2.2.1 Monatlicher und jährlicher Klimarückblick

Zu jedem aktuellen Monat wird im Lauf des Folge-monats ein Bericht herausgegeben, der die wesentlichen Anomalien, einige charakteristische Größen der atmosphärischen Zirkulation sowie Extremwerte von Klimavariablen von jedem Land und die herausragenden Witterungsereignisse (sowohl als Text als auch als Kartendarstellung) des abgelaufenen Monats enthält. Als Datenbasis dienen neben den oben genannten RCC-CM-Kartenprodukten auch weitere Daten aus der Klimadatenbank des DWD und vor allem Informationen aus dem Internet von anderen nationalen Wetterdiensten und anderen namhaften meteorologischen Institutionen. Das Ziel ist, möglichst nach kurzer Zeit aber dennoch möglichst umfassend einen Überblick über den betreffenden Monat bereitzustellen. Dies erfordert allerdings auch, dass auf ungeprüfte oder nur schnellgeprüfte Daten zurückgegriffen werden muss, die gegebenenfalls zu einem späteren Zeitpunkt noch aktualisiert werden müssen.

Nach Ablauf eines Jahres wird zusätzlich noch ein ausführlicherer jährlicher Klimarückblick erstellt, das

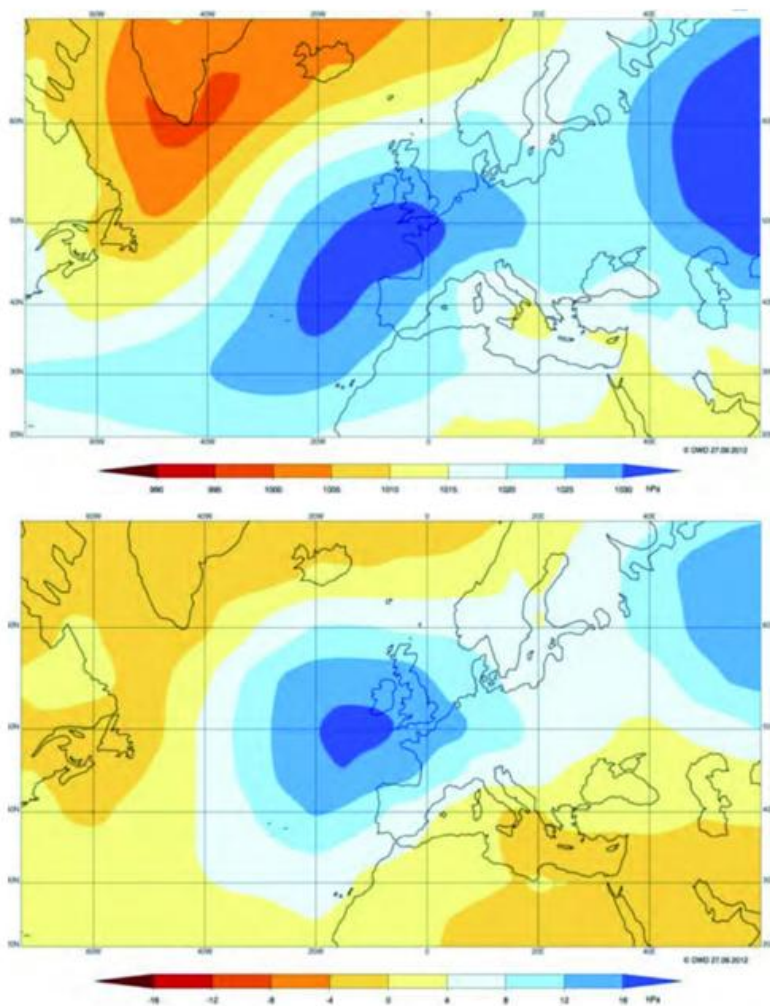


Abb. 12: Mittlere räumliche Verteilung des Bodenluftdrucks für den Monat Februar 2012 (oben) und Anomalien vom Referenzzeitraum 1961-1990 (unten).

Annual Bulletin. Die wesentlichste Informationsquelle hierfür sind jährliche Zusammenfassungen von anderen nationalen Wetterdiensten, welche direkt an den DWD zugemeldet werden; also Informationen, die teilweise im Internet oder auf sonstigem Weg nicht oder nur sehr umständlich öffentlich zugänglich sind. Dies macht den Mehrwert gegenüber vielen anderen Klimarückblicken aus. Kurzfassungen bzw. vorläufige Fassungen dieses Jahresrückblicks werden auch an anderer Stelle veröffentlicht, nämlich im „State of the Climate“-Bericht im Bulletin of the American Meteorological Society (BAMS, <http://www.ncdc.noaa.gov/bams-state-of-the-climate>), im Klimastatusbericht des DWD (<http://www.dwd.de>) sowie im „WMO Statement on the Status of the Global Climate“ der WMO (http://www.wmo.int/pages/themes/climate/status_global_climate.php).

2.2.2 Besondere Wetterereignisse

Im Fall eines länger andauernden signifikanten Wetterereignisses (zum Beispiel einer Hitze- oder Kältewelle, Starkregen- oder Dürreperiode) innerhalb der RA VI-Region, welches mehrere Länder betrifft, wird dieses Ereignis, sofern es von allgemeinem Interesse ist, im Rahmen eines

Spezialberichts ausführlicher beleuchtet. Im Wesentlichen enthält ein solcher Bericht eine detaillierte Beschreibung des Ereignisses, Informationen über Auswirkungen und aufgetretene Schäden oder Verlust von Menschenleben, eine klimatologische Einordnung im Vergleich zu früheren Ereignissen und eine Diskussion möglicher Ursachen, die zu diesem Ereignis geführt hatten. Der Bericht wird in der Regel direkt nach Eintreten des Ereignisses begonnen und wenige Wochen danach im Internet verbreitet um diese Information möglichst zeitnah, aber dennoch ausführlich genug verfügbar zu machen. Adressaten sind hier die Medien und die interessierte allgemeine Öffentlichkeit; die Berichte sollen also einen populärwissenschaftlichen Charakter haben.

2.3 Warnungen vor aufkommenden Klimaereignissen (Klimawacht, Climate Watch)

Eine weitere wichtige Aufgabe des RCC-CM ist der Aufbau eines Warnmanagements vor aufkommenden Witterungsereignissen. Hierfür ist nicht nur ein Blick in die Vergangenheit und Gegenwart notwendig, sondern auch in die Zukunft, also eine Interpretation von Witterungs- und Klimavorhersagen. Obwohl Vorhersagen in der Zeitskala von Jahreszeiten oder gar noch darüber hinaus gerade für Europa nach wie vor noch sehr unsicher sind, liegt inzwischen dazu sehr viel interpretierfähiges Material vor, das von mehreren Institutionen erstellt wird und aus dem zumindest einige Hinweise auf sich abzeichnende extreme Anomalien entnommen werden können.

Es wird in der Regel so vorgegangen, dass zunächst der Ist-Zustand betrachtet wird, also die oben genannte Analyse der Klimamonitoring-Produkte, und anschließend anhand der Vorhersagen in verschiedenen Zeitbereichen (nächste Tage, nächste Wochen, Monate, Jahreszeit) geprüft, ob eine bereits bestehende extreme Anomalie sich möglicherweise in der Zukunft fortsetzen wird oder ob Umstellungen der großräumigen Zirkulation zu erwarten sind, die auf ein neues Ereignis schließen lassen oder ein bestehendes Ereignis zum Erliegen bringen, also eine diagnostische Arbeitsweise soweit möglich.

Ergibt die Analyse, dass ein markantes Ereignis zu erwarten ist, so wird vom DWD ein Warnhinweis formuliert, der zunächst mit den anderen RCC-Partnern abgestimmt und anschließend an alle betreffenden nationalen Wetterdienste gesendet wird. Danach steht es im Ermessen der Wetterdienste, inwieweit und in welcher Form diese ihrerseits konkrete detailliertere Warnungen an ihre Endkunden weitergeben.

3 Ausblick

Die formelle Designierung (Einrichtung) des europäischen RA VI RCC-Netzwerks ist ein entscheidender Meilenstein aus organisatorischer und formaler Sicht. Diese

Designierung wird auf Grundlage der RCC-Pflichtprodukte durchgeführt. Im September 2012 hatte die WMO-Kommission für Basissysteme (Commission for Basic Systems, CBS) das RCC-Netzwerk aus fachlich-technischer Sicht angenommen. Formal abgeschlossen ist die Designierung jedoch erst mit der Annahme durch den Exekutivrat der WMO (voraussichtlich 2013). Beim weiteren Ausbau der Produkt- und Servicepalette ist das RCC-Netzwerk frei, diejenigen Themen anzupacken, die von den europäischen Mitgliedsländern – in diesem Fall vertreten durch die nationalen Wetterdienste – als prioritär eingestuft werden. Ganz generell gesprochen wird es sich dabei

- um die Weiterentwicklung qualitativ hochwertiger Datensätze handeln – möglichst gerastert und per moderner Webservices nutzerorientiert visualisierbar,
- um die Aufnahme weiterer relevanter Parameter und Indizes in das Klimamonitoring einschließlich Ursachenzuordnung beobachteter Anomalien und Trends und schließlich
- die Einbeziehung relevanter Zeitskalen wie Monat (Monatsvorhersage), Jahrzehnt (Dekadenvorhersage) und Jahrhundert (Klimaprojektionen) in die Vorhersagefunktion des RCC-Netzwerks. Unter dem Stichwort ‚Nutzerorientierung‘ wird dieser weitere Entwicklungsprozess in engster Abstimmung mit den Wetterdiensten Europas und des Nahen Ostens betrieben werden.

Danksagung

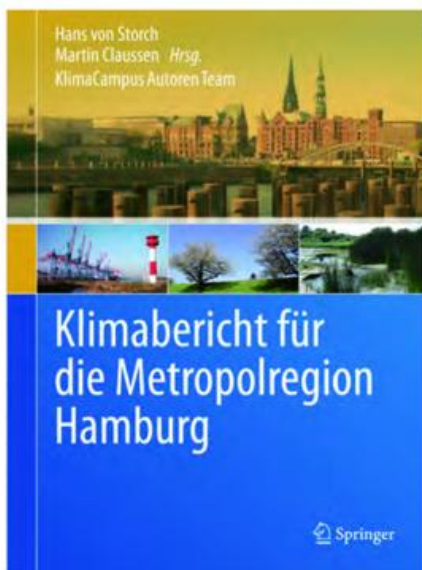
Diese Publikation wäre nicht möglich gewesen ohne den engagierten Einsatz unserer Kolleginnen und Kollegen aus unserem Referat „Regionale Klimaüberwachung“ im DWD, die uns bei der Entwicklung und operationellen Bereitstellung der Produkte sehr unterstützt haben. Namentlich danken wir hier insbesondere K. Friedrich, A. Obregón, S. Brinckmann, A. Kreis, M. Körber, J. Bollenbach, H.-J. Müller, J.M. Rocek und V. Zins. Ebenso danken wir weiteren Kolleginnen und Kollegen aus dem Referat „Satellitengestütztes Klimamonitoring (CM SAF)“ für die Bereitstellung des Strahlungsprodukts, aus dem Referat „Maritime Klimaüberwachung“ für das Luftdruckprodukt, aus dem Referat „Niederschlagsüberwachung“ (WZN) für das Niederschlagsprodukt und W. Rosenow für die Daten des „Satellitenwetters“ als Grundlage für das Wolkenprodukt. F. Hogewind (früher als Doktorand in unserem Referat, inzwischen am KIT Karlsruhe tätig) entwickelte die grundlegende Methodik für das Interpolationsverfahren für die monatliche Temperatur, Frau S. Pietzsch (ebenfalls früher in unserem Referat, inzwischen in der Abteilung „Agrarmeteorologie“) für das Dürreprodukt.

Darüber hinaus gilt unser Dank J. Rapp, der diese Publikation angeregt hatte, sowie C.-D. Schönwiese für die Durchsicht des Manuskripts.

Literatur

- BISSOLLI, P., 2003: Construction of a cloud cover reference climatology using satellite and in situ data for the climate monitoring programme SAT-KLIM, Proceedings of the 2003 EUMETSAT Meteorological Satellite Conference, Weimar, Germany, 29.09. - 03.10.2003.
- BISSOLLI, P., RAPP, J., FRIEDRICH, K., ZIESE, M., WEIGL, E., NITSCHKE, H., MALITZ, G., BECKER, A., 2010: Zur Rolle des Starkniederschlages am 7.-9. August 2010 im Dreiländereck Polen, Tschechien, Deutschland bei der Entstehung der Hochwasser von Neiße, Spree und Elbe. <http://www.dwd.de/rcc-cm>.
- BISSOLLI, P., ZIESE, M., PIETZSCH, S., FINGER, P., FRIEDRICH, K., NITSCHKE, H., OBREGÓN, A., 2012: Trockenheit in Europa im Frühjahr 2012. <http://www.dwd.de/rcc-cm>.
- DOBESCH, H., DUMOLARD, P., DYRAS, I., 2007: Spatial interpolation for climate data. *Iste*, London.
- FRIEDRICH, K., BISSOLLI, P., 2011: Analysis of temperatures and precipitation recorded at stations in Eastern Europe during the heat wave in summer 2010. <http://www.dwd.de/rcc-cm>.
- HOGEWIND, F., BISSOLLI, P., 2011: Operational maps of monthly mean temperature for WMO Region VI (Europe and Middle East). *Idöjaras* **115**, 31-49.
- KASPAR, F., HOLLMANN, R., LOCKHOFF, M., KARLSSON, K.-G., DYBBROE, A., FUCHS, P., SELBACH, N., STEIN, D., SCHULZ, J., 2009: Operational generation of AVHRR-based cloud products for Europe and the Arctic at EUMETSAT's Satellite Application Facility on Climate Monitoring (CM-SAF). *Adv. Sci. Res.* **3**, 45-51.
- KRÄHENMANN, S., BISSOLLI, P., RAPP, J., AHRENS, B., 2011: Spatial gridding of daily maximum and minimum temperatures in Europe. *Meteorology and Atmospheric Physics* **114**, 151-161.
- PIETZSCH, S., BISSOLLI, P., 2011: A modified drought index for WMO RA VI. *Adv. Sci. Res.* **6**, 275-279.
- ROSENOW, W., GÜLDNER, J., SPÄNKUCH, D., 2001: The "Satellite Weather" of the German Weather Service - an assimilation procedure with a spectral component. - Proceedings of the 2001 EUMETSAT Met. Data User's Conf., EUMETSAT Publ. EUM P **33**, 541-545.
- SCHNEIDER, U., BECKER A., MEYER-CHRISTOFFER A., ZIESE M., RUDOLF B., 2011: Global precipitation Analysis Products of the GPCC. ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf.

Buchbesprechungen



HANS VON STORCH, MARTIN CLAUSSEN (HRSG.), *Klimacampus Autoren-Team: Klimabericht für die Metropolregion Hamburg*. Springer, Heidelberg, 2011, 99,95 Euro, 321 Seiten, ISBN 978-3-642-16034-9.

<http://www.klimabericht-hamburg.de/>

Für jeden, der sich aktuell und umfassend über Klima und Klimawandel im Hamburger Raum informieren möchte, ist das im letzten Jahr im Springer-Verlag erschienene Werk ein „Muss“. Der „Klimabericht für die Metropolregion Hamburg“ ist ein Sachstandsbericht und beschreibt nicht nur ausführlich das Klima der Region und seine möglichen künftigen Änderungen, sondern geht auch auf den klimabedingten Wandel in Ökosystemen und in bestimmten Wirtschaftssektoren ein. Schließlich zeigt er am Ende auch planerische und technische Anpassungspotenziale auf.

Der Klimabericht wurde im Rahmen des Exzellenzclusters „CliSAP“ am Klima-Campus der Hamburger Universität erarbeitet. Das Autorenteam setzt sich aus einer Vielzahl von Wissenschaftlern aus den verschiedensten Disziplinen und Institutionen zusammen. Als Herausgeber fungieren Hans von Storch vom GKSS und Martin Claussen von der Universität

Hamburg. Für die einzelnen Beiträge wurden Frank Chmielewski, Franciscus Colijn, Achim Daschkeit, Hans-Ulrich Fanger, Harald Heinrichs, Kai Jensen, Jörg Knieling, Gudrun Rosenhagen, Michael Schatzmann, Nicole von Lieberman, Ralf Weisse und Norbert Winkel als Leitautoren gewonnen.

Das Buch gliedert sich in vier Teile:

- Teil I: Klima der Region bisher und mögliche Änderungen in der Zukunft (mit Deutscher Bucht und Tidelbe).
- Teil II: Klimabedingte Änderungen in Ökosystemen (terrestrisch, aquatisch).
- Teil III: Klimabedingte Änderungen in Wirtschaftssektoren (Beispiele Tourismus, Landwirtschaft).
- Teil IV: Planerisch-organisatorische und technische Anpassungspotenziale.

Besonders lesenswert ist das von den beiden Herausgebern verfasste Einleitungskapitel, das gleichzeitig auch eine solide Zusammenfassung aller Beiträge des über 300 Seiten umfassenden Bandes beinhaltet. Es bietet eine motivierende Einführung, indem es zunächst dafür sensibilisieren will, zwischen kulturellem und wissenschaftlichem „Klimakonstrukt“ zu unterscheiden. Mit diesen Vorüberlegungen lesen sich die nachfolgenden 300 Seiten grundsätzlich etwas kritischer, auch wenn sie eher selten Querbezug zur Bedeutung von Klima und Klimawandel für die kulturelle Sphäre nehmen. Im Lichte solcher Überlegungen bietet das Buch aber immerhin eine fundierte wissenschaftliche Grundlage für die sicher notwendige Diskussion zur kulturellen Bedeutung der Klimaänderungen für die Gesellschaft. Ob es dann direkt das Klima und den möglichen Klimawandel beschreibt oder „nur das Wissen darüber“ erscheint in diesem Kontext als wissenschaftsphilosophische Petitesse.

Von besonderem Reiz des Buches ist grundsätzlich, dass nicht nur sich

streng auf die Metropolregion beziehenden Texte und Abbildungen abgedruckt werden, sondern immer wieder auch der Blick auf größere Regionen (Deutschland, Mitteleuropa, Europa) gelenkt wird. Dadurch wird eine Einordnung des Klimas in einen größeren Rahmen ermöglicht.

Im Einzelnen gibt es aber doch noch einige Bemerkungen und Fragen, die an dieser Stelle nicht alle aufgelistet werden können, wovon aber doch Einige genannt seien und die vielleicht bei einer Neuauflage berücksichtigt werden können. Und eine Neuauflage ist dem neuen Hamburger „Standardwerk“ nur zu wünschen, möglichst nicht allzu lange nach der Publikation der neuen Ergebnisse der regionalen Klimamodelle im kommenden Jahr. Alle diese Hinweise können aber den ausgezeichneten Gesamteindruck des Werkes nur wenig mindern.

- Der Beitrag zu den meteorologischen Messungen und Beobachtungen beinhaltet eine allzu vorsichtige Einschätzung der Datenqualität, die im Prinzip korrekt ist, aber dennoch Gefahr läuft, einen unvoreingenommenen Leser stark zu verunsichern. Die überaus informative Darstellung kann daher zu einer Abwertung des Werts der Klimabeobachtungsstatistik (insbesondere im Falle der Lufttemperatur) führen, was im Vergleich zu der deutlich geringeren Genauigkeit von Klimamodellergebnissen „ungerecht“ erscheint.
- Einige Trendanalysen (Abb. 2.22, 2.25, 2.28) des vergangenen Klimas berücksichtigen leider nicht mehr die letzten zehn Jahre, in denen die Lufttemperaturen ein besonders hohes Niveau erreichten. Es ist zu wünschen, dass derartige Trendkarten regelmäßig aktualisiert publiziert werden.
- Es ist empfehlenswert, immer die Referenzperiode von Mittelwertgrafiken anzugeben. Die Karte der Eisgrenzen (Abb. 4.5) dürfte sich nach meiner Recherche wahrscheinlich auf die Zeit vor 1990 beziehen und

- damit die beiden letzten warmen Jahrzehnte nicht beinhalten.
- Die Rolle des überaus interessanten „invers-barometrischen Effekts“ bei der Änderung der Meeresspiegelhöhe wird im Text, gerade auch hinsichtlich der Klimaänderungen, etwas zu wenig verdeutlicht.
 - Das Kapitel zu den terrestrischen und semiterrestrischen Ökosystemen hätte noch einige ausdrucksstarke Fotos vertragen.
- Es ist sicher wünschenswert, wenn in den nächsten Jahren für möglichst viele deutsche Metropolregionen (Berlin, Ruhrgebiet, Köln, Rhein-Main, Rhein-Neckar, München, ...) derart detaillierte und umfassende Klimaberichte, die deutlich auch die potentiellen Klimaänderungen mit einschließen, erscheinen würden.
- Jörg Rapp

Habilitationen, Dissertationen Master-, Diplom- und Bachelorarbeiten im Jahr 2010

Universität Bayreuth

Dissertationen

LEHMANN, Volker: Improved intermittent clutter filtering for wind profiler radar.

STAUDT, Katharina: Modeling the exchange of energy and matter within and above a spruce forest with the higher order closure model ACASA.

Diplomarbeiten

HAASE, Max: Kohlenstoffflüsse über einer Wiesenfläche im Fichtelgebirge.

HÜBNER, Jörg: Einsatzvorbereitung einer REA-Anlage über Wiesenflächen.

PAUSCHER, Lukas: Scintillometer Measurements above the Urban Area of London.

RÜTZ, Friederike: Parameterization of the roughness sublayer above low and tall vegetation.

SCHÄFER, Andreas: Modellierung des Kohlenstoff- und Energieaustausches am Waldstein/Weidenbrunnen im Jahr 2003.

THIEME, Christoph: WRF-Modellierung meteorologischer Größen über Wald- und Gras-Ökosystemen.

Bachelorarbeiten

Teil Mikrometeorologie

DIESING, Sophie Charlotte: Voruntersuchungen zur Kalibrierung des Schiffsregenschmessers.

KITTLER, Fanny: Untersuchung der Periodenlängen für künstliche Beschneidung im gegenwärtigen und künftigen Klima.

KOHNERT, Kathrin: Untersuchung des Bowen-Verhältnisses über verschiedenen Landnutzungstypen während des COPS-Experimentes.

SOLDNER, Monika: Homogenisierung der Bayreuther Klimareihe.

Bachelorarbeit

Teil Atmosphärische Chemie

WALZ, Josefine: Volatile fraction of

Arctic aerosol particles in the summertime.

Freie Universität Berlin

Dissertationen

CLAUSSNITZER, Antje: Statistisch-dynamische Analyse skalenabhängiger Niederschlagsprozesse: Vergleich zwischen Beobachtungen und Modell.

DONAT, Markus: European wind storms, related loss potentials and changes in multi-model climate simulations.

LEINWEBER, Ronny: Radar-based precipitation classification in the Baltic Sea area.

SOMMER, Matthias: Numerische und geometrische Aspekte der Energie-Wirbel-Theorie.

RUHTZ, Thomas: Beiträge zur Messung des in der Atmosphäre gestreuten Sonnenlichtes.

KÖHLER, Philipp: Bestimmung und Validierung der Höhe von Mehrschichtbewölkung mit MERIS und AATSR.

MAHNKOPF, Tobias: Zeitliche und räumliche Analyse der Blitzhäufigkeit in Europa.

MÖLLER, Lisa: Skalenabhängige Analyse von hochaufgelösten Niederschlagsmessungen im Berliner Stadtgebiet.

PEDERSEN, Philipp: Verifikation der COSMO-DE Niederschlagsvorhersage für das Berliner Stadtmessnetz.

SLABY, Maria: Analyse der Abhängigkeiten zwischen externen Klimaantrieben und Dürren für Ägypten von 4000-0 v. Chr.: eine Modellstudie.

Masterarbeit

KRUSCHKE, Tim: Dekadische Variabilität von Sturmrisiko.

Diplomarbeiten

BEFORT, Daniel: Untersuchung der Charakteristika des australischen Monsuns mit Hilfe des regionalen Klimamodells COSMO-CLM.

HANNIG, Katrin: Einfluss der Randbedingungen aus dem hemisphärischen Chemie-Transport-Modell TM5 auf das regionale Aerosol-Chemie-Transport-Modell REM-CALGRID.

GRIEWANK, Philipp: Implementierung der Energie- und Wirbelerhaltung in einem Mehrschicht-Flachwassermodell.

KIRSCH, Annekatrin: Untersuchung des Indischen Monsuns im Mittleren Holozän und Maunder-Minimum im Vergleich zum heutigen Klima mit dem gekoppelten Modell ECHO-G.

NEUBER, Jana: Analyse von mesoskaligen Wolkenkomplexen über Westafrika mit Hilfe von bodengebundener und satellitengetragener Fernerkundung.

SCHMIEDEL, Diana: Untersuchung zum großräumigen Transport von sechs allergierelevanten Pollenarten in Ostdeutschland (2003-2006).

STEGE, Sara: Rolle der obertroposphärischen Divergenz für die Zugbahnen extremer außertropischer Zyklogen im rezenten und zukünftigen Klima.

WERSING, Claudia: Einfluss des Ozons der frühwinterlichen Stratosphäre auf das Totalozon im Frühjahr in hohen Breiten anhand von Daten des ECHAM5/Messy-Modells.

Bachelorarbeiten

BIEBER, Anne: Optimierung und Evaluierung von MOS-Vorhersagen für minutliche Pegelstandsvorhersagen Cuxhaven.

DREFENSTEDT, Sarah: Optimierung und Evaluierung von MOS-Vorhersagen für den Stau. Cuxhaven

EICHHORN, Katharina: Analyse von simulierten Windgeschwindigkeiten über Deutschland aus zwei verschiedenen Modellversionen des CLM.

FENGLER, Carolin: Der Einfluss der Änderung der solaren Aktivität auf das Klima der Mittelmeerregion am Beispiel der Medieval Warm Period und des Maunder-Minimums.

- FISCHER, Madlen: Vergleich von hochaufgelösten Niederschlagsmessungen mit verschiedenen Messgeräten.
 FREUND, Mandy: Klima/Wachstumsanalyse an Hand von Jahrringbreiten der Pinus sylvestris aus dem Tegeler Forst.
 GUTTZEIT, Martin: Die Bewertung der Umweltzone Berlins.

Universität Bonn

Diplomarbeiten

- BAUM, Julia: Die Einwirkungen von Aerosolen auf den Strahlungshaushalt der Atmosphäre.
 MAAHN, Maximilian: Measuring precipitation with Micro Rain Radars in Svalbard.
 MEYER, Gesa: Modellierung phänologischer Daten für Nordwestdeutschland.
 MÜLLER, Jennifer: Determination of radiative fluxes from satellite data and effects of scale.
 SCHALGE, Bernd: Blocking in einem einfachen globalen Zirkulationsmodell.
 UEBACHS, Annika: Multivariate analysis and regionalization of climate variability in Germany in the second half of the 20th century.

Bachelorarbeiten

- BIERDEL, Lotte Beata: Räumliche Varianzspektren in dem konvektionsauflösenden Wettervorhersagemodell COSMO-DE.
 HOFFMANN, Silvia: Vergleich verschiedener Konvektionsparametrisierungen in Bezug auf Niederschlagsintensität und -verteilung.
 KELBCH, Alexander: Mesoskalige konvektive Systeme – squall-lines.
 KEUNE, Jessica: Probabilistisches Post-Processing von GFS-Ensemble-Simulationen.
 KNIST, Sebastian: Alpine Leezyklone.
 WÜRTZ, Andreas: Ein lokales Windsystem: Die Bora.

Universität Bremen

Dissertation

- GOLCHERT, Sven: Stratospheric Water Vapour in the Tropics – Observations by Ground-Based Microwave Radiometry.

Universität Dresden

Dissertationen

- BARFUS, Klemens: On the reconstruction of three-dimensional cloud fields by synergistic use of different remote sensing data.
 SPANK, Uwe: Site Water Budget: Influences of Measurement Uncertainties on Measurement Results and Model Results.
 TITTEBRAND, Antje: Analysis of the spatial heterogeneity of land surface parameters and energy flux densities.

Diplomarbeiten

- BERNSTEIN, Christina: Analyse des Saftflussverhaltens von Fichte mit besonderer Berücksichtigung des Einflusses atmosphärischer Bedingungen auf seine zeitliche Dynamik.
 KRONENBERG, Rico: Erstellung einer radarbasierten Niederschlagsklimatologie für Deutschland mit dem Schwerpunkt im erweiterten REGKLAM – Modellgebiet.
 STEINBORN, Alexander: Untersuchungen zur Definition eines robusten Trockenheitsindex unter Verwendung der klimatischen Wasserbilanz.

Universität Frankfurt

Dissertationen

- HOMAN, Carine Dorianne: Transport and Mixing in the Tropical Tropopause Layer.
 REGNERY, Julia: Organophosphates in precipitation, lake water, and groundwater from urban and remote areas.

Diplomarbeiten

- BORCHERT, Sebastian: Ein Finite-Modell des rotierenden Annulus.
 EBACH, Friedemann: Simulation der Größenverteilung von organischen Aerosolen während der Alken-Ozon-Reaktion unter dem Einfluss von NO.
 HUMMEL, Matthias: Laborstudie zum Beitrag organischer Peroxyradikale (RO₂) bei der Partikelneubildung während der Ethen-Ozon-Reaktion.
 KEBER, Timo: Systematische Untersuchungen der Aufnahme von atmosphärischem Wasserstoff im Boden.

- KEESS, Ulrike: Anomalie-Experimente mit einem reduzierten barotropen Klimamodell.
 SCHALLENBERG, Stephan: Bestimmung der Eiskeimkonzentration mit der Vakuumdiffusionskammer FRIDGE.

Bachelorarbeit

- LANG, Thorsten: Entstehungsmechanismen starker Konvektion und die Möglichkeit der Abschätzung von Konvektiven Erscheinungen anhand von Labilitätsindizes.

Technische Universität Bergakademie Freiberg

Diplomarbeiten

- ADOLPHI, Florian: Holocene temperature reconstruction in Baltic Sea sediments for the last 2000 years, using the biomarker TEX86.
 GERBER, Bianca: Veränderung der Zusammensetzung von Niederschlägen und interzeptierten Nebelwässern bei der Kronenraumpassage eines Fichtenbestandes.
 STÄRZ, Michael: Marine Dynamik der Arktis im letzten Hochglazial.

Bachelorarbeiten

- BUERMANN, Christoph: Aerosolverteilung in Zeit und Raum Sachsens – Aufbau Sächsische Feinstaubdatenbank (PM 10, PM 2,5). (REGKLAM, mit LFULG).
 POHL, Eric: Variability of NO, NO₂ and O₃ in Saxony.
 WEYMANN, Wiebke: Atmospheric aerosols: Comparison of analytical methods for their applicability in the REGKLAM project at TU Bergakademie Freiberg.
 WOLF, Marlen: Analysis of soil enzyme activity, active microbial biomass and the CO₂ emissions from soil under different fertilizers.

Universität Freiburg

Dissertation

- ENDLER, Christina: Analyse von hochaufgelösten Klimasimulationen für die Schwarzwaldregion - Eine tourismus-klimatische Perspektive.

Masterarbeiten

- RICHERT, Clemens: GIS-gestützte Analyse des klimatischen Potenzials der Windenergie in der Region Freiburg im Breisgau auf der Grundlage von Messungen und Klimasimulationen.
- MUTHERS, Stefan: Untersuchung des Zusammenhangs von thermischem Bioklima in Österreich auf der Grundlage von Messdaten und regionalen Klimamodellen.

Bachelorarbeiten

- FRÖHLICH, Dominik: Modifikation des thermischen Bioklimas durch die Umgestaltung des Platzes der alten Synagoge in Freiburg.
- HERRMANN, Jan: Beeinflussung des thermischen Bioklimas in urbanen Strukturen in Freiburg.
- REUTER, Markus: Ermittlung von Windverhältnissen und möglichen Standorten für Windenergieanlagen in der Region Cascais, Portugal.
- HÄMMERLE, Martin: Comparison and validation of models calculating the sky view factor.
- SCHMIDT, Philipp: Ermittlung des Schneepotenzials für die Region Feldberg auf der Grundlage von Messungen und Klimasimulationen - Künstliche Beschneigung: ja oder nein?
- KNECHT, Swantje: Klima und klimatisches Tourismuspotenzial von Fehmarn.

Universität Hamburg

Dissertationen

- BELLAIRE, Sascha: Spatial variability of the snow cover and its effect on avalanche formation.
- BOTHE, Oliver: Teleconnections for precipitation.
- GEHLOT, Swati: Feedbacks between convection and climate: analyse with global modelling and satellite observations.
- HARMS, Frank: Systematische Windkanaluntersuchungen zur Charakterisierung instationärer Ausbreitungsprozesse einzelner Gaswolken in urbanen Rauigkeitsstrukturen.
- LASSLOP, Gitta: Model data fusion for terrestrial biosphere models with carbon and water cycle observations.

- OTTO, Juliane: Interactions between climate and vegetation at high northern latitudes during the mid-Holocene.
- RIEMANN-CAMPE, Kathrin: Can convective precipitation variability be deduced from the variability in CAPE and CIN? An Analysis of global CAPE and CIN variability in present and future climates.
- RIES, Hinnerk: Influences of surface layer modelling approaches on model performance.
- ROSS, Jens Ole: Simulation of atmospheric krypton 85 transport to assess the detectability of clandestine nuclear reprocessing.
- SIENZ, Frank: Meteorological extremes and their variability on different time scales.
- SCHNEIDERREIT, Andrea: Lebenszyklen außertropischer Zyklonen im heutigen und künftigen Klima.
- WOHLFAHRT, Jens: Validation of climate system models with mid-Holocene palaeobotanic data for the Northern Extratropics.

Diplomarbeiten

- BEITSCH, Alexander: Decadal to multi-centennial arctic climate variation during the last Millennium.
- FEIERTAG, Nicole: Lagrangesche Niederschlagsverifikation: Analyse von konvektiven Niederschlagszellen mit Hilfe des Radartrackers RAD-TRAM.
- KIEPERT, Kristin: Systematische Analyse des bodennahen mittleren Windfeldes in einer komplexen Stadtstruktur.
- MEI Mei: Systematische Untersuchung des Ausbreitungsverhaltens von Schadstoffwolken in einer idealisierten Stadtstruktur.
- PELTRET, Dennis: A Rossby wave breaking climatology in ECHAM 5: present day climate and greenhouse warming.
- RÖHRS, Johannes: Detection of sea ice leads with passive microwave imagery and their formation in response to sea ice motion.
- RUDDAT, Angela: Flugzeugmessungen der turbulenten, vertikalen Flüsse in der Grenzschicht über arktischem Meereis während ABSIS 2003 und Vergleich mit Parametrisierungen.

- SPENSBERGER, Clemens: Albedoänderungen in der Arktis durch Russmissionen von Schiffen und Flugzeugen,
- SCHOETTER, Robert: Einflüsse von Quellkonfiguration und Modelleigenschaften auf simulierte Konzentrationsfelder.
- STOFFELS, Nicole: Improving an Aerosol Climatology with Geostatistical Methods.
- ZYGMUNTOWSKA, Marta: Clouds and the Radiation Budget over the Arctic Ocean: results from CloudSat and CALIPSO Satellites.

Bachelorarbeiten

- BITTNER, Matthias: Analyse der Auswirkungen des Pinatubo 1991 auf die mittleren Breiten basierend auf den ERA40 Re-Analysen .
- BOYSEN, Lena: Beurteilung numerisch simulierter Grenzschichtprofile im Vergleich zu Profilen nach TA Luft.
- BRÜCK, Heiner Matthias: Analyse der Lorenz Energetik im globalen Zirkulationsmodell PUMA.
- BURDANOWITZ, Jörg: Lebenszyklus konvektiven Zellen in Modellen und Radarmessungen.
- CONRADY, Christina: Typische Werte statistischer Maßzahlen zur Evaluierung mesoskalischer Modelle.
- FISCHER, Mirka: Analyse des Hamburger Stadtklimas anhand von Schulwetterstationen.
- HANSEN, Thor: Häufigkeiten und Eigenschaften niedriger Inversionen über Hamburg.
- KEIDERLING, Stefan: Analyse räumlich hochaufgelöster Niederschlagsmessungen des Schneefalls im Winter 2009/2010 mit drei neuartigen optischen Disdrometern auf dem Dach des Geomatikums.
- KILIAN, Markus: Klimaeffekt der Kuawe-Eruption 1453.
- NESKE, Stefanie: Statistische Untersuchung von Windböen in Abhängigkeit von der Windstärke, Höhe und Schichtung am Wettermast Hamburg.
- PHILLIP, Anne: Weiterentwicklung eines effizienten Auswertungsprogramms für im Windkanal gemessene Schadstoffwolkenausbreitung.
- POPKE, Dagmar: Untersuchung der Skalenabhängigkeit einer statistischen Wolkenparametrisierung

im Globalen Zirkulationsmodell ECHAM6.
 TATARINOVA, Anastasia: Strahlungs- und Energiebilanz über arktischem Meereis in Abhängigkeit von der Wolken- und Eisbedeckung anhand von Flugzeugmessungen.

Universität Hannover

Diplomarbeiten

ADLER, Dominik: Entwicklung eines Kurzfristprognoseverfahrens der Windenergie und dessen Anwendung auf den Windpark Meyenburg.
 BARRE, Stefanie: Auswirkungen von Klimaänderungen auf das Pflanzenwachstum und die Ertragsentwicklung ausgewählter regionaler Kulturpflanzen mittels numerischer Simulation.
 BERGER, Rabea: Moore im Klimasystem am Beispiel Chinas.
 GERKEN, Julia: Verifizierung von Vereisungsvorhersagen des britischen Wetterdienstmodells „Unified Model“.
 HELD, Benjamin: Regionalisierung von Klimamodellen.
 HILGER, Viola: Numerische Simulationen zur Schadstoffausbreitung im Einflussbereich einer Windenergieanlage.
 HUCHZERMEYER, Hauke: Optimierung eines Spektralradiometers.
 KAMPMEYER, Jens: Grobstruktursimulationen zur Entwicklung von freier und erzwungener Rollenkonvektion in Kaltluftausbrüchen.
 KANANI, Farah: Footprintbestimmung mittels LES für das LITFASS-2003 Experiment.
 MARONGA, Björn: The Effect of Secondary Circulations on the Boundary Layer Turbulence during LITFASS-2003 – an LES Study.
 SIEBENAND, Henning: Untersuchung des Stadt-Umland-Effektes von Hannover anhand von Temperatur- und Feuchtedaten.
 STRAUSS, Annett: Erstellung eines wetterbezogenen Sicherheitsmaßes im Flugverkehr.
 SÜHRING, Matthias: Entwicklung und Optimierung numerischer Verfahren höherer Ordnung und ihre Implementierung im LES-Modell PALM.

VOSS, Linda: Der Einfluss kleinskaliger Heterogenitäten auf die bodennahe Turbulenzstruktur.

Bachelorarbeiten

FUHRMANN, Markus: Extremwertstatistische Auswertungen zu ausgewählten meteorologische Parametern in Hannover und Umgebung.
 GÜNTHER, Robert: „Ein mikroskaliges Modell: Aufbau – Eingangsdaten – Anwendung“.
 KUHLMANN, Gerrit: Using Anemometer Stations for Wind Shear Alerts at the Hong Kong International Airport.
 KURTZ, Anna: Kurzfristprognose postfrontaler Schauer.
 LAUBE, Natalie: Das Klima in China – Auswertung von Modell- und Beobachtungsdaten.
 ROLOFF, Katharina: Winterwetter in einem Luftverkehrsmodell.

Universität Karlsruhe

Dissertationen

LUNDGREN, Kristina: Direct Radiative Effects of Sea Salt on the Regional Scale.
 SCHWENDIKE, Juliane: Convection in an African Easterly Wave over West Africa and the Eastern Atlantic: a Model Case Study of Hurricane Helene (2006) and its Interaction with the Saharan Air Layer.
 TRÄUMNER, Katja: Einmischprozesse am Oberrand der konvektiven atmosphärischen Grenzschicht.
 VERSICK, Stefan: Ableitung von H₂O aus MIPAS/ENVISAT-Beobachtungen und Untersuchung der Wirkung von energetischen Teilchen auf den chemischen Zustand der mittleren Atmosphäre.

Diplomarbeiten

ADLER, Bianca: Der Einfluss von Landoberflächeninhomogenitäten auf die Auslösung und Entwicklung eines mesoskaligen konvektiven Systems: Eine budgetbasierte Modellanalyse.
 ASCHENBRENNER, Thomas: Untersuchung der Struktur der Konvektion über komplexem Terrain anhand von Messungen mit dem Forschungsflugzeug DO 128 während COPS.

BÜCKLE, Jennifer: Korrektur von Second Trip Echos in Radardaten.
 FISCHER, Lucas: Vergleich von Modellrechnungen mit COSMO-CLM/TERRA und COSMO-CLM/VEG3D.
 FÜTTERER, Daniel: Analyse der Datenqualität der neuen Multi-Sensor Fallsonde KITsonde.
 JANDT, Simon: Numerische Untersuchungen zur Niederschlagsbeeinflussung im östlichen Mittelmeerraum: Fallstudie für eine reale Land-See-Wind-Zirkulation.
 KLÜPFEL, Vera: Ensemblesimulationen konvektiven Niederschlags für Westafrika durch Variation des Anfangszustandes mit verschiedenen Bodenfeuchtedaten.
 KOCH, Ulrike Angelika: Zusammenhang zwischen Starkniederschlagsereignissen und Großwetterlagen nach Re-Analysedaten.
 KUNTZ, Miriam: Vektorinterpolation auf unstrukturierten Gittern.
 PLIENINGER, Johannes: Vergleich von Spektren aus Breitband – Strahlungsübertragungsrechnungen mit MIPAS-ENVISAT Messungen.
 SCHMIDBERGER, Manuel: Der Transport von Mineralstaub und die Wechselwirkung zwischen Strahlung und Dynamik - eine Modellstudie über Westafrika im Juni 2006.
 WOLFF, Sarah Isabelle: Parametrisierung und Simulation des Auswaschens von Aerosolpartikeln.

Universität Kiel

Dissertationen

DING, Hui: Annual to Interannual Equatorial Atlantic Variability: Mechanisms and Tropical Impacts.
 FRAUEN, Claudia: ENSO mechanisms and interactions in a hybrid coupled recharge oscillator model.

Masterarbeit

BANCALA, Severin: Analysing Major Sudden Stratospheric Warmings in the coupled middle atmosphere ocean model MAECHA5/MPI-OM.

Diplomarbeiten

BAYR, Tobias: Die Veränderungen des Luftdrucks im IPCC A1B Szenario: Mittlerer Zustand und Muster der Variabilität.

- CESKO, Thomas: Klimaeinfluss langperiodischer Schwankungen der Solarkonstante im Kiel Climate Modell.
- GEIB, Anika: Transportprozesse in der Tropischen Tropopausenschicht im Nordhemisphärischen Winter mit ERA-Interim Assimilationsdaten.
- RENKOSIK, Niko: Energiebilanz arktischer Bewölkung aus Modell und Beobachtung.
- RICKERT, Tilman: Eiszeiten in einem einfachen Klimamodell.
- TESSENDORF, Alrun: Strahlungsbilanz arktischer Bewölkung aus Modell und Beobachtung.
- WACHE, Sebastian: Transport of trace gases measured in the marine boundary layer in the tropical West Pacific during the TransBrom-Sonne campaign.

Bachelorarbeiten

- KINZEL, Julian: Die Bedeutung der Ozean-Atmosphäre-Wechselwirkung für die natürliche Klimavariabilität in der Arktis.
- SCHAFFER, Janine: Die Arktische Oszillation im Kieler Klimamodell.
- SCHEEF, Helga: Der Einfluss von sehr starken Vulkaneruptionen auf die Nordatlantische und Arktische Oszillation im MAECHAM5 Modell.
- ZANDER, Stefanie: Die Auswirkung von Vulkaneruptionen auf die atmosphärische Zirkulation der Südhemisphäre unter besonderer Berücksichtigung der Ausbruchstärke

Universität Köln

Habilitation

- KIENDLER-SCHARR, Astrid: Formation of secondary organic aerosols from biogenic emissions of volatile organic compounds.

Dissertationen

- ANDERT, Thomas: Masses of Small Bodies: Mass estimation of small solar system bodies using Radio Science data from close flybys.
- EBELL, Kerstin: Characterization of clouds and their radiative effects using ground-based instrumentation at a low-mountain site.
- FIELITZ, Daniel: 2D cross-hole MMR: Survey design and sensitivity analy-

- sis for cross-hole applications of the magnetometric resistivity method.
- POHLE, Susan: Synoptische und dynamische Aspekte tropisch-extratropischer Wechselwirkungen: Drei Fallstudien von Hitzetiefentwicklungen über Westafrika während des AMMA-Experiments 2006.

Diplomarbeiten

- HARLAND, Melanie: Radiomagnetotellurische Messungen zur Erdfallgefährdung in Münsterdorf (Kreis Steinburg, Schleswig-Holstein).
- HOPPE, Charlotte: Entwicklung und Anwendung eines variationellen Datenassimilationssystems für ein Soil-Vegetation-Atmosphäre-Transfer-Modell.
- KARREMAN, Melanie: Auswirkung der Klimaänderung auf die Wiederkehrperiode von Winterstürmen über Europa.
- KLOSE, Martina: Development and implementation of a dust emission scheme for convective atmospheric conditions.
- OSCHLISNIOK, Janusz: Absorption und Ausbreitungseffekte von Radiowellen in der Venusatmosphäre beobachtet von Venus Express Radio Science Experiment.
- PANKE, Patrick: Evaluation des COSMO-DE mittels passiven Satellitenbeobachtungen im Mikrowellenspektrum.
- PASCHARAT, Laura: Time Series Analysis of Magnetic Field Data Recorded in the Inner Solar Wind with Regard to Turbulence and a Particular Focus on the Anisotropy of Turbulent Fluctuations.

- RÖDDER, Annika: Interpretation von SHOTEM Daten mit mehrdimensionalen Leitfähigkeitsmodellen am Beispiel der Araba Verwerfung, Jordanien.

- SEKULOW, Ina: Volatilitätsanalyse der Windstromerzeugung in Europa als Beitrag einer optimierten Netzintegration.

- YOGESHWAR, Pritam: Grundwasserkontamination bei Roorkee/ Indien: 2D Joint Inversion von Radiomagnetotellurik und Gleichstromgeoelektrikdaten.

Bachelorarbeiten

- DULING, Stefan: Die Polarlichter des Jupitermondes Ganymed.

- MATTHIAE, Sabrina: Die Variabilität des Harmattan in Westafrika in den Jahren 2002/03 bis 2009/10.

- SCHLEICHER, Christina: Niederschlagsstatistik im Rheinland.

- SCHREINER, Anne: Frequenz-Wellenvektor-Korrelation anisotroper Sonnenwindturbulenz.

- SEIDEL, Marc: Auswirkungen der Magnetfeldrichtung in stellaren Winden bei der Plasma- Wechselwirkung mit Exoplaneten.

- SEREGINA, Larisa: Variabilität von **Wolkenwasserprofilen**.

- TUCHSCHERER, Yvonne: Trendanalyse von monatlichen Stationsniederschlägen in Westafrika zwischen 1921 und 2008.

Universität Leipzig

Dissertationen

- ARRAS, Christina: A Global Survey of Sporadic E Layers based on GPS Radio Occultations by CHAMP, GRACE and FORMOSAT-3/COSMIC.

- VOIGTLÄNDER, Jens: Hygroscopic growth and CCN activation of slightly soluble organic and inorganic compounds - Evaluation of experimental LACIS data with FLUENT/FPM.

Diplomarbeiten

- BARTHEL, Stefan: Modellierung des Lebenszyklus und der Chemie von Seesalzaerosolen in der maritimen Grenzschicht.

- BRECHT, Benedict: Akustische Anemometrie und Thermometrie.

- BRÜCKNER, Marlen: Depolarisationsmessungen in Cirren mit einem konventionellen Rückstreulidar.

- FRICKE, Clemens: Messung und Simulation polarisierter Strahldichten: **Einfluss des solaren Zenitwinkels** und der Aerosol optischen Dicke.

- FRIEDRICH, Anett: Messung und Modellierung der kumulierten thermischen Belastung in Leipzig.

- GEIMECKE, Christina: Regionale Modellierung direkter und indirekter Effekte des Saharastaubs über Europa.

- GÖBEL, Tina: Hygroskopizität anthropogener Aerosole bei hohen Feuchten in der nordchinesischen Tiefebene.

GROSS, Peter: Global Distribution of Submicrometer Aerosol Particles: Comparison between In Situ Measurements and Global Model Outputs.

KAMILLI, Katharina: Hygroskopische Eigenschaften des urbanen Aerosols einer europäischen „Megacity“.

KINDER, Friederike: Hygroscopic properties of atmospheric aerosol particles in central Europe.

LONITZ, Katrin: Comparison of MISR and Meteosat-9 Cloud Motion Winds.

LOUCA, Kyriakos Yiannakis: Untersuchung der Eigenschaften von Windprofilen in der atmosphärischen Grenzschicht mit verschiedenen Messmethoden am Standort Lindenberg.

OPELT, Claudia: Räumliche Variabilität atmosphärischer Partikelgrößenverteilung im städtischen Ballungsraum Dresden.

RADDATZ, Michael: Messungen von Impaktwinkeln und Auslösewinkeln von Sandkörnern mittels Hochgeschwindigkeitskamera.

RÖSCH, Carolin: Indoor air and health effects.

ROST, Florian: Untersuchung und Simulation des Einflusses von Turbulenz auf die Schallausbreitung in der Atmosphäre

SCHLIMPER, Susan: Meteorologische und statistische Auswertung von Langzeitmessungen zur Charakterisierung des Aerosols an der Forschungsstation Melpitz.

SCHMEISSNER, Tina: Tropical free tropospheric aerosol microphysical properties measured at a high altitude research station in the Venezuelan Andes.

SCHÖNE, Luisa: Laboruntersuchungen zur möglichen SOA-Bildung infolge der troposphärischen Isoprenoxidation.

SEELIG, Torsten: Vergleich barokliner Moden in Modell und Experiment.

SMALLA, Alexander: Hoch aufgelöste Chemietransport-Modellierung für den Großraum Dresden mit dem WRF/Chem-Modell.

STADLER, Andreas: Vergleich von Windmessungen mit einem SODAR und anderen Messmethoden am Standort Lindenberg.

STOLL, Jens: Kopplung des Wettermodell WRF (ARW) mit dem

Chemietransportmodell MUSCAT: Vergleich mit weiteren meteorologischen Antrieben.

THIEM, Elisabeth: Physikalische Eigenschaften des atmosphärischen Aerosols am Polarmeer in Nordalaska.

UNGLAUB, Claudia: EUV-TEC: Ein Index zur Beschreibung der ionosphärischen Variabilität.

Universität Mainz

Dissertationen

ASENSIO, Hermann: Diagnose der Residuumszirkulation und ihre Bedeutung für die Klimavariabilität der Stratosphäre.

BÖTTCHER, Maxi: Dynamik diabatischer Rossby-Wellen.

BOZEM, Heiko: Photochemie der Troposphäre in niedrigen und mittleren Breiten: Die Rolle von Konvektion.

KALESSE, Heike: Influence of Ice Crystal Habit and Cirrus Spatial Inhomogeneities on the Retrieval of Cirrus Optical Thickness and Effective Radius.

REINERT, Daniel: Dynamik orographischer Bannerwolken.

Diplomarbeiten

ANGER, Johanna: Analyse hochaufgelöster Time-Lagged Ensemble Kurzfristprognosen anhand von Messdaten erhoben am Flughafen Frankfurt.

BACHMANN, Julia: Untersuchung der Vorhersage von Zugbahn und Struktur außertropischer Tiefdruckgebiete.

BERKES, Florian: The Climatology and Dynamics of Intense Squalls off Western Equatorial Africa.

HOFMANN, Christiane: Fallstudie einer diabatischen Rossby-Welle über dem Nordpazifik.

TRAUD, Sebastian: Emissionen von Spurenstoffen und deren Transport über Fernost.

Universität München

Dissertationen

DAHL, Johanns: The development of a new lightning-frequency parameterization and its implementation in a weather prediction model.

GARNY, Hella: Causes and impacts of changes in the stratospheric meridional circulation in a chemistry-climate model.

HAMANN, Ulrich: Die Erwärmungsrate in der tropischen Tropopause-region.

WELKER, Christoph: Vorhersagbarkeit tropischer Zyklonen und ihrer versicherungsrelevanten Schäden im indopazifischen Raum.

Diplomarbeiten

BRETL, Sebastian: Untersuchung des Lebenszyklus von Gewittern in Mitteleuropa mit Hilfe von Fernerkundungs- und Modelldaten.

BRÜSER, Paul: Beobachtung, Verfolgung und Kurzfristvorhersage von Starkniederschlag mit Hilfe des hochaufgelösten Deutschland Radar – Kompositos.

EICHINGER, Roland: Analyse thermohaliner Ozean Dynamik – Eine konzeptionelle Modellstudie mit Meereis.

FEISTL, Thomas: Schneehöhenverteilung im bayerischen Alpenraum.

GRÖBEL, Anna: Zeitreihenanalyse von Temperatur und Niederschlag ausgewählter Stationen im Bayerischen Wald.

KAPOSVARI, Michaela: Numerische Modellierung von Gewittern in Anwesenheit von Orographie.

KEIS, Felix: Die Temperatur- und Niederschlagsmessungen in München.

KILJUS, Niels: Bodengestützte passive Wolkenseitenfernerkundung der Wolkenphase.

KOSMALE, Miriam: Fehlerquantifizierung des Inversionsalgorithmus für ein Raman-Lidar durch theoretische Simulation.

LEUTHOLD, Eva: Über den Einfluss der Variabilität tropischer Meeresoberflächentemperaturen auf die Dynamik der unteren Stratosphäre.

MAGET, Christoph: Untersuchung des Wettereinflusses auf den Verkehr.

SCHEUFELE, Katrin: Probabilistic forecasting of precipitation by combining nowcasts and the time-lagged COSMO-DE ensemble.

SIEBLER, Dominik: Bestimmung von Größen- und Formfaktoren von Schnee anhand von Messungen am Schneefernhaus.

THAMMER, Markus: Evapotranspiration im Kontext weltweiter Nah-

runnungsmittelproduktion: Modellgestützte Quantifizierung blauen und grünen virtuellen Wasserhandels.

WILCKE, Renate: Madden-Julian Oscillation and its correlation with observation data in Northern Australia.

ZINK, Matthias: Verbesserte Flüssigwasseralgorithmien durch Kombination von passiver Zentimeter- und Millimeterwellenfernerkundung.

Technische Universität München

Masterarbeit

ACACIA, Mathado: The real costs of tourism to climate change.

Diplomarbeit

BÖCK, Andreas: Kartierung sowie höhen- und expositionsabhängige Analyse der Daten aus den Jahren 2008 und 2009.

Universität Basel

Dissertation

FREY, Corinne M.: On the determination of the spatial energy balance of a megacity on the example of Cairo, Egypt (Vortrag).

Masterarbeiten

SÜTTERLIN, Melanie: Analysis of the spatial and temporal change of late summer snow patches

WERNER, Ambros: Flussmessungen mittels Eddykovarianz auf der Mittelmeerinsel Pianosa.

MAASS, Ueli: Bestimmung der Energie- und Wärmebilanzen der Stadt Istanbul mit Gernerkundungsmethoden.

BÜHRER, Michael: Solarstrom – Vorhersage und Markteinbindung

DIETHELM, Samuel: Open- or closed-path gas analyser: which one delivers the „better“ CO₂ Fluxes?

Universität Graz

Dissertationen

LACKNER Bettinia Christina: Exploring Trend Indicators of Climate Change From Radio Occultation and Optimal Trend Detection.

PIRSCHER Barbara: Multi-Satellite Climatologies of Fundamental Atmospheric Variables From Radio Occultation and Their Validation.

SCHWEITZER Susanne: Accurate monitoring greenhouse gases and primary greenhouses gas isotopes form space.

SUKLITSCH Martin: Non-Hydrostatic Climate Modelling: Sensitivity Studies and Error Characterization.

TRUHETZ Heimo Franz: High resolution wind field modelling over complex topography: analysis and future scenarios.

Masterarbeiten

FUCHS, Reinhard: Empirical-Statistical Post-Processing of Radiation and Humidity Parameters in Regional Climate Simulations.

GÖLLES, Thomas: Modeling Oxygen Isotopes in Polythermal Ice Sheets.

LANDSCHÜTZER, Peter: Air-Sea Gas Exchange of CO₂ and O₂ under Climate Change.

ZOTTER, Peter: Einfluss von Lärmschutzwänden auf die Schadstoffbelastung neben Straßen.

ZWANZIGER, Christof: ACCURATE: Performance Analysis and Absorption Simulations for the ESA-IRDAS-EXP Demonstration Experiment.

Diplomarbeiten

BIEDERMANN, Andreas: Extreme Klima- und Wetterereignisse in der Region Südoststeiermark.

GÖGER, Michael: Der Vergleich von Atmosphärenprofilen mit einem Inter-Validierungssystem und dessen Anwendung auf Radiookkultations- und Radiosondendaten.

OSWALD, Angela: Monitoring Active Layer Development and Freeze-Back using DC Resistivity Tomography.

SUPPAN, Claus: Untersuchung der Langzeitstabilität des Abtastfehlers von CHAMP Radiookkultationsmessungen 2001 – 2008.

Bachelorarbeiten

BEHAM, Michael: The Supercell Thunderstorm: History, Life Cycle, Ingredients and a Brief Comparison between Austria and the United States.

DEFNER, Beppino: Methoden der quantitativen Klimamodellierung.

TRUMMER, Michaela: Atmosphärische Optik.

WEBER, Julia Sarah: Immissionsmessungen Ozon.

Universität Innsbruck

Diplomarbeiten

DAUTZ, Erika: The Interruption of Alpine Foehn by a Cold Front: A Modeling Case Study.

DOBESBERGER, Paul: Meteorologische und schneephysikalische Untersuchung von Gleitrissen und Gleitschneelawinen.

FIGWER, Florian: Rückstreusignaturen in der Akkumulationszone Grönlands aus Mehr-Frequenz-Radarmessungen.

FILIPITSCH, Florian: Simulationen zur Datenrückgewinnung von Aerosoleigenschaften aus Schichten unter dünner Cirrus-Bewölkung für satellitengetragene Lidar-Messungen.

GAITAN, Emma Fernandez: Modellierung des Wasserkreislaufs im Einzugsgebiet von Obergurgl 1961-90 mit dem hydrometeorologischen Modell OEZ 2.1.

GASSNER, Andreas: Kritische Vergleiche des Klima-Grid HISTALP mit der empirischen Reihe Rudolfs-hütte.

HARTL, Lea: The Gepatschferner from 1850 - 2006: Changes in Length, Area and Volume in Relation to Climate.

HÖLZL, Simon: Quality control and analysis of meteorological mesonet data with special emphasis on disdrometer measurements.

JUEN, Martin: Laserscanning zur Bestimmung der Ablation im schuttbedeckten Teil des Hintereisferners.

KLEE, Alexander: Statistische Windvorhersage mit Ensemble Dressing.

KOCH, Roland: Key Analysis Errors and Airborne Wind Lidar Observations.

MANZL, Rainer: Ein Energiebilanzmodell des Hintereisferners.

MARKSTALER, Gerhard: Meteorologisches Windenergiepotential am Nob in Vorarlberg.

NIKOLAI, Eva: Novemberniederschlagsverteilung in Südtirol von 1982-2008: Eine klimatologische und synoptische Untersuchung.

ORTNER, Andreas: Detaillierte Statistik des Südföhns in Innsbruck: **Einfluss des Südföhns auf die Temperatur.**

PROKSCH, Martin: Comparing CT experiments and field observations with the SNOWPACK microstructure.

REUTER, Benjamin: The Effect of Surface Warming on Snowpack Stability.

SANDNER, Roman: Ice retreat after disintegration of the northern Larsen Ice Shelf.

SCHREITER, Martin: Auswertungen von Ceilometerdaten hinsichtlich Grenzschicht- und Wolkhöhe im Raum Innsbruck und Wien.

SEISER, Bernd: Gletscherinventar 2006 der Stubai Alpen.

STANDTEINER, Alexander: Analyse des **Oberflächengeschwindigkeitsfeldes** des Larsen-C-Schelfeises, Antarktis; basierend auf Fernerkundungsdaten.

STOCKER-WALDHUBER, Martin: Untersuchung des Massenhaushalts am Mullwitzkees im Nationalpark Hohe Tauern.

STRAUSS, Susanne: Flugzeuggetragene Wasserdampf-Lidarmessungen im Subtropenjet im Vergleich mit ECMWF-Modelldaten.

STÜTZ, Elisabeth: Dynamically and Thermally Driven Flows Over and Around Svalbard: a Case Study Based on Numerical Simulations and Airborne Measurements.

SUMMER, André: Spaceborne Measurements of Aerosol Optical Depth and Derivation of Particulate Matter Concentration.

WALTER, Dagmar: Analyse meteorologischer Verhältnisse im hinteren Wattental während der Winter 2006 bis 2009.

WELZENBACH, Felix: Reverse valley winds in the Stanzer Valley.

Bachelorarbeiten

DABERNIG, Markus: Stormsurgemodelle.

EMPRECHTINGER, Markus: Vertikale Temperaturgradienten im hinteren Ötztal.

GOLLER, Markus: Gletscherinventar von Vorarlberg und Westtirol von 2006.

GRUBER, Stefanie: Analyse der Flächen- und Volumensänderung des Hintereisferners von 1855 - 2006 und des Lewis-Gletschers von 1958 - 2004.

KÄGI, Markus: Studie zur Gültigkeit parametrischer Windfeldmodelle.

LANG, Moritz Nikolaus: Feinskalige Struktur von Kaltfronten im Inn- und Wipptal während MAP.

MERANER, Katharina: Validation of statistical downscaling with the weather generator LARS-WG for South Tyrol.

PRESSER, Manuel: Eine statistische Auswertung von Kaltfrontdurchgängen im Inntal.

STAUDACHER, Jutta: Wetterlagenabhängigkeit längerer Episoden hoher PM10-Belastung im oberen Unterinntal (Wintermonate 2006 - 2008/09).

TIEFENGRABER, Martin: Global Dimming und Brightening in Innsbruck, eine Trendanalyse der Globalstrahlung in Innsbruck von 1958 bis 2007.

TILG, Anna-Maria: Vergleich verschiedener meteorologischer Datensätze im Raum Ny-Alesund, Spitzbergen.

ZIMMERMANN, Nikolas: Orographisch initiierte Gewitter an einer Konvergenzlinie während COPS.

Universität Wien

Diplomarbeiten

KUMPFMÜLLER, Thomas: Freie Wetterinformationen im Internet.

MAYER, Michael: Corrected energy fluxes as evaluated from reanalysis data.

RADANOVICS, Sabine: Homogenisierung von Radiosondentemperaturzeitreihen in der Antarktis mit ERA-Interim-Daten.

SPRINGER, Claudia: Berechnung von Zukunftsszenarien für Gletschermassenbilanzen.

Bachelorarbeit

LEUKAUF, Daniel: Vergleich von automatischen Homogenisierungsverfahren anhand von synthetischen Benchmarkdaten aus dem COST

Projekt HOME.

Universität Zürich

Dissertationen

CHIACCHIO, Marc; Decadal variations of surface solar radiation and its connection to climate and atmospheric processes.

ISOTTA, Francesco; Shallow cumulus clouds parameterization in the climate model ECHAM5-HAM.

JESS, Stephanie; Impact of subgrid variability on large-scale precipitation formation in the climate model ECHAM5.

KAMMERMANN, Lukas; Aerosol Hygroscopicity and CCN properties at remote sites.

MAHLSTEIN, Irina; Improving climate model projections by model evaluation and regional aggregation.

SIEGENTHALER-LE DRIAN, Colombe; Stratocumulus clouds in ECHAM5-HAM.

VACHARAPORN, Soonsin; The influence of physical state on hygroscopicity and vapor pressure of single, levitated aerosol particles containing organics.

YIPIN, Zhou; Optimization and application of satellite observations for tropospheric NO₂ Monitoring in Europe.

ZUBLER, Elias; Representation of aerosol and cloud microphysics in a regional climate model - Implementation, evaluation and long-term simulations.

Nachträge für das Jahr 2009:

Universität Potsdam

Dissertationen

FRANCKE, Till: Measurement and modelling of water and sediment fluxes in meso-scale dryland catchments.

Universität Bremen

Promotionen

KAZEMINEJAD, Shahin: Untersuchungen zum Einfluß von solaren Teilchen (Elektronen, Protonen) auf die mittlere Atmosphäre.
MACATANGAY, Ronald: Entwick-

lung und Erprobung neuartiger Meßmethoden zur Fernerkundung von CO₂.

PETERSEN, Anna Katinka: Atmospheric trace gas measurements in the tropics by ground-based FTIR-spectroscopy.

Veröffentlichungshinweis

Das Lehrbuch "Die Physik des Wetters und des Klimas" frei verfügbar

Das oben genannte Lehrbuch von H.-J. Lange (mit dem Untertitel "Ein Grundkurs zur Theorie der Atmosphäre") ist im Jahre 2002 im Dietrich Reimer Verlag Berlin erschienen und schon im Jahr 2006 ausverkauft. 2007 sind die Rechte an den Autor zurückgefallen. Inzwischen hat der Autor eine PDF-Version des gesamten Buches zum kostenlosen Download auf seiner Homepage www.hajolange.de zur Verfügung gestellt, einschließlich sämtlicher ihm bekannt gewordener Rezensionen (Klick auf [Rezensionen\(pdf\)](#)) und einschließlich weiterer Kommentare (Klick auf [Kommentare\(html\)](#)). Der Zugang zum Buch selbst erfolgt kapitelweise durch die folgenden Klicks:

[Vorwort und Inhalt.pdf](#)

[Kap.01 Thermodynamische Grundlagen.pdf](#)
[Kap.02 Nablarechnung und Hydrodynamik.pdf](#)
[Kap.03 Atmosphärische Gleichungen.pdf](#)
[Kap.04 Wirbeldynamik.pdf](#)
[Kap.05 Die feuchte Atmosphäre.pdf](#)
[Kap.06 Atmosphärische Grenzschicht.pdf](#)
[Kap.07 Strahlung und atmosphärische Energetik.pdf](#)

[Kap.08 Klima und Klimatheorien.pdf](#)
[Kap.09 Der Einfluss der Chaostheorie.pdf](#)
[Kap.10 Die Energie-Wirbel-Theorie.pdf](#)
[Anhang A Deformationen Spannungen und Fluide.pdf](#)
[Anhang B Kugelkoordinaten und Kugelfunktionen.pdf](#)
[Anhang C Wellen und Spektraltheorie.pdf](#)
[Anhang D Funktionale und Funktionalableitungen.pdf](#)
[Symbole, Literatur und Index.pdf](#)

Aus der obigen Aufstellung der Links ist die Grundgliederung des Buches ersichtlich. Das ausführliche neunseitige Inhaltsverzeichnis in dreifacher Untergliederung befindet sich in [Vorwort und Inhalt.pdf](#).

Das Buch widmet sich der theoretisch ausgerichteten Beschreibung der Atmosphäre in Lehre und Forschung. Die mathematischen Grundlagen werden vollständig entwickelt. Auf exakte Beweisführungen wurde jedoch verzichtet, denn diese Verrichtung haben Mathematiker im Sinne einer wissenschaftlichen Arbeitsteilung bereits getan. Somit beschränkt sich der Autor auf die Darstellung der Plausibilität und der Anschaulichkeit der „bewiesenen“ Mathematik, verzichtet aber nicht auf die Vermittlung von Anwendungssicherheit. Somit wurden die technischen Hürden für die Lektüre niedrig gehalten.

Adressen der Autoren dieses Heftes

DR. JUDITH BERNER
 Mesoscale and Microscale Meteorology Division and
 Climate and Global Dynamics Division
 National Center for Atmospheric Research (NCAR)
 P.O. Box 3000
 Boulder CO 80307-3000
 USA
 berner@ucar.edu

DR. PETER BISSOLLI
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 Peter.Bissolli@dwd.de

DR. LÜDER VON BREMEN
 ForWind – Zentrum für Windenergieforschung
 Universität Oldenburg
 Ammerländer Heerstrasse 136
 26129 Oldenburg
 lueder.von.bremen@uni-oldenburg.de

DR. MICHAEL DENHARD
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 michael.denhard@dwd.de

KIRA FELDMANN
 Institut für Angewandte Mathematik
 Universität Heidelberg
 Im Neuenheimer Feld 294
 69120 Heidelberg
 kirafeldmann@gmx.ne

DR. CHRISTOPH GEBHARDT
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 christoph.gebhardt@dwd.de

DIPL.-MET. PEER HECHLER
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 Peer.Hechler@dwd.de

DR. MATTHIAS LANGE
 energy&meteo systems GmbH
 Marie-Curie-Straße 1
 26129 Oldenburg
 Matthias.lange@energymeteo.de

DR. MARTIN LEUTBECHER
 ECMWF
 Shinfield Park
 Reading, Berkshire RG2 9AX
 United Kingdom
 martin.leutbecher@ecmwf.int

DIPL.-MET. HELGA NITSCHKE
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 Helga.Nitsche@dwd.de

DIPL.-MET. STEFAN RÖSNER
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 Stefan.Roesner@dwd.de

DR. MICHAEL SCHEUERER
 Institut für Angewandte Mathematik
 Universität Heidelberg
 Im Neuenheimer Feld 294
 69120 Heidelberg
 michael.scheuerer@uni-heidelberg.de

DR. SUSANNE THEIS
 Deutscher Wetterdienst
 Frankfurter Straße 135
 63067 Offenbach am Main
 susanne.theis@dwd.de

DR. THORDIS L. THORARINSDOTTIR
 Norwegian Computing Center
 P.O. Box 114 Blindern
 NO-0314 Oslo
 Norway
 thordis@nr.no

DR. ANDREAS P. WEIGEL
 Bundesamt für Meteorologie und Klimatologie
 MeteoSchweiz
 Krähbühlstrasse 58
 8044 Zürich
 Schweiz
 Andreas.weigel@alumni.ethz.ch

PROF. DR. HEINI WERNLI
 Institut für Atmosphäre und Klima
 ETH Zürich
 Universitätstrasse 16
 8092 Zürich
 Schweiz
 heini.wernli@env.ethz.ch

Redaktionelle Hinweise für Autoren

(Kurzfassung)

1 Allgemeines

Promet dient vornehmlich der Fortbildung von Meteorologen und Wetterberatern, aber auch von Studenten. Die Beiträge zum Thema des Heftes sollen den neuesten Stand des zu behandelnden Spezialgebietes auf wissenschaftlicher Basis in einer verständlichen und didaktisch anschaulichen Weise darstellen.

• Fachredaktion

Er gewinnt die Autoren für die einzelnen Kapitel, legt die Gliederung an Hand der einzelnen Manuskriptzusagen fest und stimmt die ihm eingereichten Manuskripte fachlich aufeinander ab. Die Struktur des Heftes stimmt der Fachredakteur in einem möglichst frühen Stadium des Heftes mit dem Gutachter ab, um für die einzelnen Autoren Doppelarbeit zu vermeiden. Die einzelnen Manuskripte einschl. Abbildungen werden vom Fachredakteur gesammelt und zunächst ausschließlich an den Schriftleiter geschickt, der sie nach Prüfung anschließend dem Gutachter weiterleitet.

• Fachliche Durchsicht (Gutachter)

Eine Persönlichkeit, die nicht Autor oder Fachredakteur des betreffenden Heftes ist, führt einen Review-Prozess durch, er begutachtet die Einzelbeiträge. Dabei sollte sowohl auf wissenschaftliche Aktualität wie auch auf Verständlichkeit geachtet werden. Der Reviewer sollte bereits in einem sehr frühen Planungsstadium des Heftes durch den Fachredakteur um seinen fachlichen Rat gebeten werden.

• Schriftleitung

- (a) Er pflegt den Kontakt zwischen dem Fachredakteur, dem Gutachter sowie den einzelnen Autoren, um in vorgegebener Zeit die Druckvorlage eines Heftes an die Druckerei übergeben zu können.
- (b) Er spricht selber oder auf Hinweis mögliche Autoren mit der Bitte um einen Sonderbeitrag an. Das sind Beiträge, die mit dem Hauptthema des Heftes in keinem Zusammenhang stehen.

2 Text

Als Textsoftware bitte MS-WORD verwenden mit einer Schriftgröße von Punkt 12, einem Zeilenabstand von 1,5 und einem Seitenrand von jeweils 2,5 cm.

Der Umfang einer gedruckten Promet-Seite, die nur Text enthält, lässt sich wie folgt ermitteln: etwa 2,5 DIN A4-Seiten Text wie oben festgelegt entsprechen 1 Druckseite. Werden Abbildungen gebracht, so ist der dafür benötigte Platz derart abzuschätzen, dass die Abbildungen im

Druck in der Regel eine Breite von 85 mm (einspaltig) bzw. 170 mm (zweispaltig) haben werden.

Text bitte in Deutsch nach den neuen Rechtschreibregeln.

Gliederung des Textes: Die Nummerierung der Kapitel eines Beitrages erfolgt stringent nach dem Dezimalsystem (1, 1.1, 1.2, 1.2.1, ..., 2, 2.1, 2.2, usw.).

Dem Beitrag sind voranzustellen:

- die Namen der Autoren,
- die deutschsprachige Überschrift,
- die englischsprachige Überschrift,
- die deutschsprachige Zusammenfassung (etwa 5 - 8 Zeilen),
- die englischsprachige Zusammenfassung (etwa 5 - 8 Zeilen).

Dabei ist darauf zu achten, dass die deutschen und englischen Texte möglichst inhaltsgleich sind.

Am Ende des Beitrages sind zu nennen:

- Vor- und Zuname des/der Autors/Autoren,
- akademischer Grad,
- Anschrift der Dienststelle/Institut, unter der der/die Autor(en) zu erreichen ist/sind,
- E-Mail-Adresse(n),
- Die Angabe einer Telefonnummer ist vom Schriftleiter für kurzfristige Rückfragen erwünscht. Sie wird in Promet nicht veröffentlicht.

3 Literaturhinweise

Die Anzahl der Literaturhinweise ist deutlich zu begrenzen. Die Literatur ist so auszuwählen, dass mit ihrer Hilfe eine Vertiefung der Thematik leicht möglich ist. Es ist wünschenswert, wenn neben einigen Fachbüchern, die die Thematik des Kapitels grundlegend behandeln, nur wenige (nicht mehr als 10) Zeitschriftenartikel genannt werden. Auf keinen Fall sollte die Literaturliste einen Umfang von mehr als einer Druckseite umfassen.

Die Literatur ist wie folgt zu zitieren:

- *im Text: MÜLLER (1966) oder (MÜLLER 1966).*
- *Im Literaturverzeichnis: Autoren in alphabetischer Reihenfolge. Herausgeber durch den Zusatz „Hrsg.“ kennzeichnen.*

Beispiele:

Zeitschriftenaufsatz: KURZ, M., 1982: Zum Einfluß diabetischer Prozesse auf die Frontogenese in Bodennähe. *Meteorol. Rdsch.* **35**, 21-30.

Buch: SCHÖNWIESE, C.-D., 1980: Klimaschwankungen, Berlin: Springer-Verlag, 181 Seiten.

4 Nummerierung der Tabellen und Abbildungen im Text

Tab. [Beitrags-Nr.] – [laufende Nr.], zum Beispiel: Tab. 2-3,
Abb. [Beitrags-Nr.] – [laufende Nr.], zum Beispiel: Abb. 1-4.

5 Abbildungen

Abbildungen bitte nicht in den Text einbinden, sondern als getrennte Dateien an den Schriftleiter. Die Abbildungslegenden bitte am Ende des Textes zusammen auflisten.

Die Abbildungen sowie Fotografien werden als EDV-Datei erbeten. Bevorzugte Dateiformate: *.tiff oder *.eps. Bei einer Breite von 85 mm bzw. 170 mm sollte die Abbildung eine Auflösung von mindestens 300 dpi haben (mindestens 980 Pixel Kantenlänge).

Abbildungen und Fotos mit einer zu geringen Druckauflösung können in der Regel nicht akzeptiert werden! Bitte nehmen Sie ggf. Kontakt mit der Schriftleitung auf.

Bei der Beschriftung der Achsen bezüglich der Dimension wie folgt verfahren:

Beispiele: „Höhe in km“, „Geschwindigkeit in m/s“, „spektrale Bestrahlungsstärke in $\text{mW cm}^{-2} \mu\text{m}^{-1}$ “. Bitte die Dimension nicht in runden oder eckigen Klammern setzen.

Alle Beschriftungen (Achsen wie Liniensätze) sollten so groß gewählt werden, dass sie bei einer Verkleinerung auf eine Breite von 85 bzw. 170 mm noch gut lesbar sind.

Falls eine Abbildung nur in englischsprachiger Fassung vorliegt, bitte eine tabellarische Übersicht englisch ↔ deutsch beifügen, mit deren Hilfe die deutschen Bezeichnungen an den Achsen und innerhalb der Abbildung zweifelsfrei eingefügt werden können.

6 Korrekturen

Autoren, die das Thema eines Kapitels behandeln, senden ihre Beiträge an den Fachredakteur, der die einzelnen Kapitel fachlich aufeinander abstimmt. Nach Begutachtung der Beiträge und Einarbeitung der Korrekturen und Hinweise durch den Autor erhält der Schriftleiter die abschließende Version von den Autoren, veranlasst daraufhin die Umbruchkorrekturen. Vom Schriftleiter erhält jeder erstgenannte Autor eines Beitrages Fahnenabzüge zur Korrektur. Nach Einarbeitung dieser Korrekturen, die nur noch offensichtliche Fehler, aber keine inhaltliche Änderungen betreffen, wird das Heft endgültig vom Schriftleiter zum Druck freigegeben.

7 Belegexemplare

Autoren erhalten in der Regel 3 bis 5 Belegexemplare des betreffenden Heftes. Zusätzliche Exemplare können gegen Erstattung der Fortdruckkosten bei der Rücksendung der Korrekturfahnen beim Schriftleiter bestellt werden

Auch die Abgabe eines Beitrages an den betreffenden Autor als pdf-Datei ist möglich. Bei der Veröffentlichung dieser Datei, beispielsweise auf der Homepage des Autors oder des Instituts, ist stets die Literaturquelle zu vermerken: „aus *promet* – Fortbildungszeitschrift des DWD, Jahrg. XX, (20YY), Heft ZZ, Seiten aa bis bb.“ Eine Veröffentlichung kommt allerdings erst nach einem halben Jahr in Absprache mit der Schriftleitung in Frage

promet

Vorschau auf die nächsten Hefte

- Agrar- und Forstmeteorologie (Fachredaktion: Dr. K.-P. Wittich)
- Meteorologische Aspekte regenerativer Energien (Fachredaktion: Dr. D. Heinemann)
- Flugmeteorologie (Fachredaktion: Prof. Dr. T. Hauf)