



universität  
wien

# Diplomarbeit

Titel der Diplomarbeit

Vorhersageverifikation hochauflösender Modelle mit MET und  
VERA

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Verfasser: Kiesenhofer, Stefan  
Matrikelnummer: 0301040  
Studienrichtung: A415 Meteorologie  
Betreuer: O. Univ. Prof. Dr. Reinhold Steinacker

Wien, im Juni 2011

# Zusammenfassung

Die getroffene Auswahl an Vorhersagemodellen zur Erstellung einer umfassenden vergleichenden Verifikationsstudie beinhaltet sowohl hydrostatische als auch nicht-hydrostatische Modelle mit zum Teil sehr unterschiedlichen Auflösungen. Im Mittelpunkt steht dabei die Auswertung der damit verbundenen Unterschiede in Bezug auf Auflösung, Modellformulierung und dem Zusammenschluss mehrerer Modelle zu Modellketten. Als Referenzdatensatz dient der modellunabhängige VERA Re-Analyse-Datensatz (Vienna Enhanced Resolution Analysis) mit einer räumlichen Auflösung von 8 km und einer zeitlichen von 1h, dessen Basis vom JDC (Joint D-PHASE COPS) Datensatz mit ca. 11000 Stationen über Zentraleuropa gebildet wird. Zur Berechnung der Verifikationsmaße wurde das am NCAR entwickelte Softwarepaket MET (Model Evaluation Tool) verwendet, das neben der Berechnung einer Vielzahl klassischer Maßzahlen auch eine neuere Methode zur objekt-orientierten Verifikation beinhaltet. Der Fokus in Hinblick auf den Modellvergleich liegt auf der Vereinheitlichung der Modellfelder durch eine räumliche Interpolation der Parameter auf VERA- Gitterpunkte, sowie der Reduktion der Modelldomänen auf einen gemeinsamen geographischen Ausschnitt. Durch die Sicherstellung einer optimalen Vergleichbarkeit sind qualifizierte Aussagen über die unterschiedlichen Arbeitsweisen und möglichen Defizite der Modelle bei der Prognose meteorologischer Parameter, in erster Linie aber bei der Untersuchung von Niederschlagsereignissen, möglich. Vor allem in Bezug auf Konvektion zeigen sich dabei teilweise deutliche Differenzen. Die räumliche und zeitliche Überschneidung mit COPS ermöglicht zudem ein weiteres Anwendungsgebiet zur Untersuchung lokaler Einflussfaktoren wie etwa Luv-Lee Effekte, hervorgerufen durch die Mittelgebirge von Vogesen und Schwarzwald. Zur statistischen Analyse der mit MET berechneten Maßzahlen zur qualitativen Auswertung kategorischer Variablen, wie etwa Niederschlag, wurden beispielsweise ROC- Diagramme verwendet. Im Gegensatz dazu erfolgten die Untersuchungen zur quantitativen Auswertung etwa durch die Analyse des Bias.

# Abstract

For the purposes and the development of a comparative verification study a selection of hydrostatic and non-hydrostatic forecast models with different spatial resolutions has been taken into account. The main focus of the composition is based on differences according to resolution, the design of the model and the aggregation of several models into model chains. As a reference database the model independent VERA- reanalyses dataset (Vienna Enhanced Resolution Analysis) with a spatial resolution of 8 km and temporal resolution of 1h is used, based on the JDC (Joint D-PHASE COPS) dataset including about 11000 stations over central europe. All verification scores have been calculated with the software environment of MET (Model Evaluation Tool), developed and distributed at NCAR. In addition to supplying a wide variety of traditional verification scores a new object oriented method, MODE, has been worked out and is incorporated to the software package of MET. The main focus to accomplish the model comparison is on using uniform model fields concerning the spatial interpolation of the parameters to VERA- gridpoints as well as the restriction of model domains to a common geographic operations area. Due to optimal comparisons requirements qualitative conclusions across the differences in operation and possible lacks in forecasting meteorological parameters can be determined, primarily concerning investigations of precipitation events. Regarding convection, explicit differences come into notice. The spatial and temporal intersection with COPS furthermore permits an additional area of applications to investigate the influence of a windward/lee effect induced by mountain ranges of Vosges mountains and Black Forest. Statistical analysis of the scores computed in MET with respect to a qualitative verification access of categorical variables are accomplished with a ROC- diagram. In contrast quantitative evaluations are achieved by analyses of the forecast Bias.



# Inhaltsverzeichnis

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Einleitung</b>  | <b>1</b> |
| 1.1      | Zielsetzung und Überblick über die Kapitelinhalte . . . . .    | 2        |
| <b>2</b> | <b>Aufbereitung der Datenbasis</b>                             | <b>5</b> |
| 2.1      | Beschreibung der verwendeten Modelle . . . . .                 | 5        |
| 2.1.1    | ALADIN - France . . . . .                                      | 6        |
| 2.1.2    | AROME . . . . .  | 7        |
| 2.1.3    | COSMO-EU . . . . .   | 7        |
| 2.1.4    | COSMO-DE . . . . .   | 7        |
| 2.1.5    | COSMO 7 . . . . .  | 8        |
| 2.1.6    | COSMO 2 . . . . .  | 8        |
| 2.1.7    | CMC-GEM-L . . . . .  | 9        |
| 2.2      | Modelldaten . . . . .  | 9        |
| 2.2.1    | Datenformat der Modelle in MAP D-PHASE . . . . .               | 10       |
| 2.3      | Modellgitternetz und Struktur der VERA-Gitters . . . . .       | 10       |
| 2.3.1    | Gitterstruktur von VERA und dessen Projektion . . . . .        | 10       |
| 2.3.2    | Gitterstrukturen und Projektionen der Modelle . . . . .        | 13       |
| 2.3.2.1  | Rotiertes sphärisches Gitter . . . . .                         | 14       |
| 2.3.3    | Modellgebiete . . . . .  | 15       |
| 2.3.4    | Verwendete Modellläufe . . . . .                               | 17       |
| 2.4      | Anpassung und Ergänzung der Modellvariablen an VERA . . . . .  | 18       |
| 2.4.1    | Potentielle und äquivalentpotentielle Temperatur . . . . .     | 18       |
| 2.4.2    | Reduzierter Druck . . . . .                                    | 19       |
| 2.4.3    | Mischungsverhältnis . . . . .                                  | 21       |
| 2.4.4    | Höhenkorrektur für $\Theta$ und $\Theta_e$ . . . . .           | 22       |
| 2.5      | Interpolation der Modellfelder auf VERA-Gitterpunkte . . . . . | 22       |
| 2.5.1    | Gemeinsamer Bereich der jeweils verwendeten Modelle . . . . .  | 24       |

|          |   |           |
|----------|---|-----------|
| <b>3</b> | <b>Model Evaluation Tools (MET)</b>   | <b>27</b> |
| 3.1      | Überblick über die Komponenten von MET . . . . .                                    | 28        |
| 3.1.1    | Kurze Charakterisierung der einzelnen Programmkomponenten . . . . .                 | 28        |
| 3.1.1.1  | Umformatierung der Eingangsdaten . . . . .  | 28        |
| 3.1.1.2  | Hauptkomponenten zur statistischen Auswertung . . . . .                             | 29        |
| 3.1.2    | Grid_stat tool . . . . .  | 31        |
| 3.1.3    | MODE (Methode for Object-based Diagnostic Evaluation) . . . . .                     | 32        |
| 3.1.3.1  | Objektdefinierung . . . . .   | 33        |
| 3.1.3.2  | Klassifikation von Objektmerkmalen . . . . .  | 33        |
| 3.1.3.3  | Gesamtrelevanz als Funktion von Objektattributen . . . . .                          | 35        |
| 3.1.3.4  | Gegenüberstellung von MODE und SAL . . . . .  | 36        |
| <b>4</b> | <b>Qualitativer Verifikationszugang</b>   | <b>39</b> |
| 4.1      | Die Kontingenztafel . . . . .   | 39        |
| 4.2      | ROC (Relative Operating Characteristic) . . . . .                                   | 41        |
| 4.2.1    | Interpretation von ROC für ein Modell . . . . .                                     | 41        |
| 4.2.1.1  | Ausweitung auf größere zeitliche Intervalle . . . . .                               | 44        |
| 4.2.2    | ROC für mehrere Modelle . . . . .   | 46        |
| 4.2.2.1  | Aspekte zur Vergleichbarkeit unterschiedlicher Modelle . . . . .                    | 48        |
| 4.2.3    | Saisonale Vergleiche 18-stündiger Prognosen . . . . .                               | 50        |
| 4.2.4    | Saisonale Vergleiche 30-stündiger Prognosen . . . . .                               | 52        |
| 4.3      | Stratifikation nach Intensitätswerten des Niederschlags . . . . .                   | 53        |
| 4.3.1    | Auswertung der Schwellwertabhängigkeit der ausgewählten Verifikationsmaße . . . . . | 56        |
| 4.3.2    | Häufigkeitsverteilung der beobachteten Ereignisse . . . . .                         | 57        |
| 4.3.3    | Jahreszeitliche Vergleiche der Schwellwertabhängigkeit . . . . .                    | 58        |
| 4.4      | Fuzzy- Verifikation . . . . .   | 60        |
| 4.4.1    | Anwendungen von Fuzzy- Methoden . . . . .   | 61        |
| 4.4.2    | Fuzzy in räumlich limitierten Bereichen . . . . .                                   | 63        |
| <b>5</b> | <b>Quantitativer Verifikationszugang</b>  | <b>65</b> |
| 5.1      | BIAS (Mittlerer Fehler) . . . . .   | 65        |
| 5.1.1    | Direkter Modellvergleich anhand der täglichen Abweichungen . . . . .                | 66        |
| 5.1.2    | Auswertung und Vergleich monatsgemittelter Werte . . . . .                          | 67        |
| 5.1.3    | Verteilung von Fehlerhäufigkeiten . . . . .   | 68        |
| 5.1.4    | Stratifizierung des Bias innerhalb der COPS- Domäne . . . . .                       | 71        |

|          |  |           |
|----------|--|-----------|
| <b>6</b> | <b>Objektorientierter Verifikationszugang über MODE</b>                                      | <b>75</b> |
| 6.1      | Auswertung von extremen Niederschlagsereignissen . . . . .                                   | 75        |
| 6.1.1    | Versuchsaufbau . . . . .   | 76        |
| 6.1.2    | Objektidentifikation aus Niederschlagsfeldern . . . . .                                      | 78        |
| 6.1.3    | Objektanalyse anhand der Auswertung von Schwerpunktdifferenzen                               | 79        |
| 6.1.4    | Verteilung der Schwerpunktdifferenzen in Abhängigkeit eines Re-<br>levanzspektrums . . . . . | 81        |
| 6.1.5    | Untersuchung der Intensitätsverteilungen von Niederschlagsobjekten                           | 85        |
| 6.1.5.1  | Einfluss des Analysesystems VERA . . . . .   | 88        |
| <b>7</b> | <b>Zusammenfassung und Ausblick</b>  | <b>89</b> |



# 1 Einleitung

Diverse Fragestellungen der Verifikation zur Evaluierung der Modellgüte von Wettervorhersagemodellen haben zum Ziel, meteorologische Variablen in Form zweidimensionaler diskreter Parameterfelder gegenüber dem Eintreten von Ereignissen auszuwerten und quantitativ über Maßzahlen zu beschreiben. Eine der wesentlichen Problematiken ist dabei die Wahl einer geeigneten Referenz zur Beschreibung der tatsächlich eingetretenen Verhältnisse, somit gewissermaßen einer Repräsentation der Realität. Doch wie sieht diese “Realität” nun im Einzelnen eigentlich aus? Instrumente zur Messung der Parameter bilden im Grunde nur die unmittelbare Umgebung ihres Standortes ab, die Werte an sich werden aber als repräsentativ für ein wesentlich ausgedehnteres Umfeld interpretiert. Die an ein Analysesystem gestellte Herausforderung besteht nun in der räumlichen Auswertung und Interpretation der Messdaten, vor allem in komplexen Gebieten. Analysesysteme sind häufig an Prognosesysteme gekoppelt und verwenden die Feldstrukturen der Vorhersagen gewissermaßen als erste Abschätzung der realen Verhältnisse. Diese “First-Guess” - Felder sind aber direkt vom Modell abhängig und beeinflussen die Analyse maßgeblich. Die Folge ist ein zwangsläufiges Mitführen der systemimmanenten Fehler mit Auswirkungen auf die Ergebnisse bei der Verifikation (Hagedorn, 2010) [20].

Das am Institut für Meteorologie und Geophysik an der Universität Wien entwickelte Analysesystem VERA (Vienna Enhanced Resolution Analysis) versucht hingegen, das System des Analyseverfahrens gänzlich von den Abhängigkeiten der Prognosemodelle zu entkoppeln und ohne ein First- Guess- Feld auszukommen. Die dabei verwendete Methode geht auf die Technik zur Einbindung von Fingerprints (Steinacker et al., 2006) zurück. Diese Voraussetzung der Modellunabhängigkeit spielt vor allem bei einem Vergleich von Modellen unterschiedlicher Konzeptionierungen eine entscheidende Rolle. Eine der Zielsetzungen dieser Arbeit ist die Verifikation von Modellprognosen über die Auswertung der Parameterfelder unter Einbindung von VERA- Analysen als Referenz mit den sich damit ergebenden Vorteilen der Modellunabhängigkeit und der analogen Behandlung der Modelle. Der hierzu verwendete Referenz- Datensatz mit der Bezeichnung JDC (JointD-PHASE COPS), wurde über eine Re-Analyse der VERA- Felder zur Ergänzung der schnell verfügbaren GTS (Global Telecommunication System) Stationsdaten auf Basis von mehr als 11000 Stationen innerhalb Europas im Rahmen des Projekts VERITA erstellt. VE-

RITA (NWPmodel VERIfikation over complex Terrain with VERA) wird seit 2007 am Institut für Meteorologie und Geophysik in Wien mit dem Ziel der Verifikation von Bodenparametern hochauflösender numerischer Vorhersagemodelle, betrieben. Die vorliegende Arbeit als integraler Bestandteil des Projekts erfüllt dabei einen Teil der Vorgaben und Zielsetzungen.

Die Datenbasis auf Seiten der Modelle bilden Modellläufe des Forschungs- und Entwicklungsprojets von WWRP (World Weather Research Programme) mit der Bezeichnung MAP D-PHASE (Mesoscale Alpine Programme - Development of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine Region). Die Motivation von MAP D-PHASE (Rotach et al., 2009) [34] lag dabei in der Erprobung von hochauflösenden Modellapplikationen hinsichtlich der Erfassung von Niederschlagsereignissen innerhalb eines Zeitraums zwischen Juni und November 2007 über Mitteleuropa. Die zeitliche und räumliche Überschneidung mit COPS (Convective and Orographically-induced Precipitation Study), (Wulfmeyer et al., 2008)[44], ermöglicht zudem eine Ausweitung der Verifikationsstudie zur Untersuchung der Einflüsse von Mittelgebirgen in der räumlich stark limitierten Region.

### 1.1 Zielsetzung und Überblick über die Kapitelinhalte

Im Mittelpunkt der vorliegenden Untersuchung steht neben den Erfordernissen eines Vergleichs modellspezifischer Merkmale der Verifikation, die Anwendung und Integration des am NCAR (National Center for Atmospheric Research) DTC (Developmental Testbed Center) entwickelten Software- Paketes MET (Model Evaluation Tools). Zur Sicherstellung einer optimalen Vergleichbarkeit der Modelle werden diese auf ein 8 km VERA- Gitter interpoliert, sowie räumlich auf ein gemeinsames Gebiet beschränkt. Der Fokus liegt dabei in Anlehnung an die Projektziele von MAP D-PHASE und COPS auf der Auswertung von Niederschlag, speziell aber der Untersuchung der Prognosequalität während konvektiver Ereignisse. Ungeachtet der Bevorteilung von Niederschlag durchlaufen aber auch andere wichtige meteorologische Parameter den gesamten Verfahrensablauf der Interpolation, sowie die Implementierung und Auswertung mit MET, werden allerdings hier nicht explizit graphisch dargestellt oder interpretiert. Die optimale Vorbereitung erlaubt indes eine einfache Adaption in möglichen zukünftigen Projekten.

Kapitel 2 gibt einen Überblick über die verwendeten Modelldaten, sowie die Aufbereitung der Datensätze durch Interpolation auf VERA - Gitterpunkte und die Eingrenzung auf ein einheitliches Gebiet. Kapitel 3 beschäftigt sich mit der Beschreibung des Verifikationspakets MET. Statistische Auswertungen und Ergebnisse der Verifikationsstudie sind in Bezug auf eine qualitative Betrachtungsweise Kapitel 4 zu entnehmen. Untersuchungen

und Bewertungen einer quantitativen Betrachtungsweise sind im Kapitel 5 beschrieben. Abschnitt 6 beschäftigt sich mit einer Methode zur objektorientierten Verifikation von Niederschlag. Das letzte Kapitel bringt eine Zusammenfassung der gewonnenen Ergebnisse sowie einen Ausblick.



## 2 Aufbereitung der Datenbasis

Während MAP D-PHASE zwischen Juni und November 2007 wurde eine Vielzahl von zum Teil im Teststadium befindlichen oder auch bereits im operationellen Dienst eingesetzten numerischen Wettervorhersage-Modelle (NWP-Modelle) von verschiedenen europäischen und internationalen Wetterdiensten und Universitäten zum Einsatz gebracht. Die räumlichen Auflösungen der Modelle und damit auch die auflösbaren atmosphärisch meteorologischen Strukturen bewegen sich im Bereich der Meso- $\beta$ -Skala bis hin zur Meso- $\gamma$ -Skala, von etwas mehr als 10 km bis zu Gitterpunktabständen kleiner als 2 km. Unter ihnen befinden sich sowohl hydrostatische als auch nicht-hydrostatische Modelle, wobei ein eindeutiger Trend weg von den hydrostatisch approximierten festzustellen ist. Die Formulierung von Modellen über eine hydrostatische Approximation ist im Allgemeinen auf eine statisch stabile Atmosphäre beschränkt und generell nicht in der Lage, konvektive Umlagerungen in labiler Schichtung aufzulösen (z.B. Adrian und Frühwald, 2002) [3]. In der Folge müssen sämtliche konvektiven Prozesse in hydrostatischen Modellen parametrisiert werden um statische Instabilitäten abzubauen. Ein weiterer wichtiger Aspekt, vor allem für hochauflösende Modelle, ist die Kopplung zwischen der orographischen Skala und dem hydrostatischen Approximationsfehler, der umso größer ausfällt, je kleiner die Skala wird. Im Hinblick auf hochauflösende Modellformulierungen mit zunehmend feineren Höhenstrukturen in der Modelltopographie bedeutet das, dass eine hydrostatische Betrachtungsweise nicht mehr möglich oder zumindest nicht mehr sinnvoll ist. Durch die erhöhte Topographie werden interne Schwerewellen in Form hydrostatischer Leewellen induziert, die sich in weitere Folge verikal ausbreiten.

### 2.1 Beschreibung der verwendeten Modelle

Bei allen hier verwendeten Modellen handelt es sich um Limited- Area- Models (LAMs) mit der Konvention seitens der Organisatoren von D-PHASE, alle während des Projekts eingesetzten Modelle innerhalb eines bestimmten vorgegebenen geographischen Ausschnitts zu betreiben. An den Modellrändern eines LAMs ist es allerdings notwendig, die Randbedingungen zu kennen. Diese Randwerte können entweder auch aus LAMs über einem größeren geographischen Gebiet mit schwächerer Auflösung oder aber aus global

betriebenen Modellsystemen stammen. Man spricht von “nesting”, also dem Einbetten kleinerer LAMs in größere. Das Modell dessen Randbedingungen zu bestimmen sind, ist meist in ein größeres Modell gebettet, welches wiederum in ein weiteres, beispielsweise globales Modell genestet ist.

Auch die Anfangszustände werden vom jeweils antreibenden Modell durch Interpolation der Analysen generiert, sind aber wegen des Unterschiedes in der horizontalen und vertikalen Auflösung nicht sehr gut definiert. Dies führt zu einer Einschwingperiode (spin-up) von wenigen Stunden, bis sich die Strömung auf die hochaufgelöste Topographie eingestellt hat. Zusätzlich stützen sich die Analysen auf ein Hintergrundfeld, dem sogenannten “first-guess-Feld”. Die so eingebrachten zusätzlichen Informationen verbessern zwar die Analysequalität, entstammen aber dem numerischen Modell selbst, wodurch die Analyse indirekt auch modellabhängig ist.

Die Untersuchung der Zusammenhänge, in Bezug auf den Zusammenschluss der Modelle zu Modellketten und der sich ergebenden Abhängigkeiten durch die jeweilige Definition des Grundzustandes und der Randbedingungen der verschachtelten Modelle unterschiedlicher Auflösung ist Gegenstand dieser Arbeit und wird ausführlich diskutiert.

### 2.1.1 ALADIN - France

Das regionale Modell des französischen Wetterdienstes Aire Limitée Adaption dynamique Développement InterNational (ALADIN) ist ein spektrales, hydrostatisch approximiertes-LAM für Kurzfristprognosen (Bubnova et al., 1993) [11]. Neben Météo-France wird ALADIN im Rahmen eines internationalen Projekts in weiteren, überwiegend europäischen, Ländern von den lokalen Wetterdiensten weiterentwickelt. Die französische Version hat eine horizontale Auflösung von 9,5 km mit 46 vertikalen Schichten. Der dynamische Kern sowie die Randwerte an den seitlichen Begrenzungslinien des Modells entstammen dem von Météo-France betriebenen globalen Modell, ARPEGE. ALADIN wird operationell vier mal täglich gestartet. Für die im Rahmen von MAP D-PHASE betriebene Version über einem kleineren als den operationell verwendeten Ausschnitt stehen aber jeweils nur die 00 UTC- Läufe mit einer Vorhersagedauer von 30h und einem Zeitschritt von einer Stunde zur Verfügung (Bazile, 2009) [7]. Für die Anforderungen des mesoskaligen Modells AROME wird zusätzlich eine nicht-hydrostatische Applikation betrieben, die jedoch nicht operationell im Einsatz ist. Die französische Version von ALADIN während COPS und MAP D-PHASE ist somit das einzige hydrostatische NWP- Modell der getroffenen Auswahl. Weiterführende Informationen zum Aufbau des Modells finden sich in Fischer et al., 2005 [19], in YESSAD, 2010 [45] oder in Bouttier, 2010 [9].

### 2.1.2 AROME

AROME (Application of Research to Operations at Mesoscale) ist ein spektrales, nicht-hydrostatisches Prototypmodell für zukünftige operationell betriebene numerische Wettervorhersage -Modelle (NWP-Modelle) des französischen Wetterdienstes, Météo-France (Bouttier et al., 2006) [8]. Die Dynamik entspricht der des französischen, mesoskaligen Forschungsmodells Meso-NH. Mit einer horizontalen Auflösung von 2,5 km und 41 vertikalen Modellflächen werden zum Termin 00 UTC stündliche Prognosen für 30h gerechnet. Wie bei den meisten mesoskaligen Modellen wird auch bei AROME hochreichende Konvektion nicht parametrisiert, sondern soll vom Modell direkt aufgelöst werden. Das antreibende Modell von AROME ist eine nicht hydrostatische Version von ALADIN, die zumeist als ALADIN-NH bezeichnet wird. Während COPS und MAP D-PHASE im Sommer 2007 wurde AROME getestet, ein operationeller Betrieb findet seit Ende 2008 statt. Informationen zum Modellaufbau finden sich in YESSAD, 2010 [45], für Informationen zur Version von AROME während COPS und MAP D-PHASE wird Seity, 2009 [36] empfohlen.

### 2.1.3 COSMO-EU

COSMO-EU mit der früheren Bezeichnung Lokal-Modell Europa LME beruht auf den primitiven hydro-thermodynamischen Eulerschen Gleichungen; es ist ein nicht- hydrostatisches, kompressibles Modell. Der Grundzustand für die Formulierung der Modellgleichungen ist zeitlich konstant und hydrostatisch balanciert (Schulz und Schattler, 2009) [35]. COSMO-EU und das Lokal-Modell Kurzzeitfrist LMK (neuer COSMO-DE) mit einem kleineren räumlichen Ausschnitt und einer höheren Auflösung sind Anwendungen des Modellsystems LM, das innerhalb des internationalen Konsortiums COSMO (COnsortium for Small-Scale MOdelling) weiterentwickelt wird. COSMO-EU mit einer horizontalen Auflösung von 7 km und 40 Modellflächen wird alle 6 Stunden um 00, 06, 12 und 18 UTC gestartet und umfasst einen Vorhersagezeitraum von  $t+78$  Stunden für die Hauptläufe um 00 und 12 UTC, und 48h-Prognosen für die Termine 06 und 18 UTC. Die Speicherung der Ausgabefelder erfolgt stündlich. Für ein LAM wird ein antreibendes Modell zur seitlichen Steuerung benötigt, in diesem Fall entstammen die zeitlich variablen Randwerte dem hydrostatischen Modell GME.

### 2.1.4 COSMO-DE

Das Modell COSMO-DE ist physikalisch sehr ähnlich aufgebaut wie das COSMO-EU, also ein nicht- hydrostatisches, kompressibles Modell unter Verwendung der primitiven hydro-thermodynamischen Eulerschen Gleichungen (Baldauf et al. 2006) [6]. Der horizontale Gitterpunktsabstand reduziert sich von 7 km auf 2,8 km, vertikal wird auf 50 Modell-

schichten aufgestockt. Eine der wesentlichen Änderungen neben der höheren Auflösung liegt in der physikalischen Parametrisierung, wobei auf die Parametrisierung für hochreichende Konvektion gänzlich verzichtet wird und somit eine explizite Auflösung zumindest von hochreichender Konvektion in den groben Strukturen vom Modell gefordert wird. Flache Konvektion hingegen wird weiterhin parametrisiert. Graupel als zusätzliche Kategorie von skaligem Niederschlag steht neben fünf weiteren auch von COSMO-EU verwendeten Kategorien von Hydrometeoren für wolkenphysikalische Umwandlungsprozesse zur Verfügung. Als antreibendes Modell dient COSMO-EU. Das Modell wird alle drei Stunden zu den Terminen 00, 03, 06, 09, 12, 15, 18 und 21 UTC gestartet. Zu jedem Termin wird eine 18-stündige Prognose mit einer stündlichen Speicherung der Ausgabefelder erstellt.

### 2.1.5 COSMO 7

COSMO 7, ein weiteres nicht- hydrostatisches LAM aus der COSMO-Familie, wird von Meteo Swiss betrieben (MeteoSwiss, 2009) [30]. Der grundsätzliche physikalische Aufbau des Modells entspricht dem von COSMO-EU, allerdings werden die Modellgleichungen auf 60 vertikalen Schichten und auf einem sich von COSMO-EU unterscheidenden Gitter mit einer Maschenweite von 6,6 km gelöst. Neben der Verwendung eines anderen Projektionsmittelpunktes, wird auch ein wesentlich kleinerer Ausschnitt über Mitteleuropa verwendet (mehr zu Projektionen im Abschnitt 2.3.2). Als antreibendes Modell zur seitlichen Steuerung des LAMs dient nicht das GME, sondern das vom European Centre for Medium-Range Weather Forecast (ECMWF) entwickelte *Integrated Forecast System* (IFS), ein globales Modell, das wegen seiner Herkunft auch oft als “Europäisches Modell” oder als “ECMWF” bezeichnet wird. COSMO 7 läuft zweimal täglich zu den Hauptterminen 00 und 12 UTC, die stündlich abgespeicherten Vorhersagefelder werden für einen Zeitraum von 72h erstellt.

### 2.1.6 COSMO 2

COSMO 2 ist ein hochaufgelöstes nicht- hydrostatisches Modell des Schweizer Wetterdienstes (Meteo Swiss) und läuft seit Februar 2008 operationell (MeteoSwiss, 2009) [30]. Das während COPS und MAP D-PHASE im Teststadium befindliche Modell mit dem Ausschnitt des Alpenraums und einer horizontalen Auflösung von 2,2 km hat ähnlich wie COSMO-DE kein Parametrisierungsschema für hochreichende Konvektion. COSMO 2 ist in das gröbere LAM des Schweizer Wetterdienstes COSMO 7 genestet, welches somit die Randbedingungen für das 60 Schichten umfassende hochaufgelöste Modell liefert. Der Mittelpunkt der Projektion ist gegenüber dem antreibenden Modell verschoben (siehe Abschnitt 2.3.2). Jeweils zu den Terminen 00, 03, 06, 09, 12, 15, 18 und 21 UTC werden acht

mal täglich 24-stündige Prognosen gerechnet. Weiterführende Informationen zum allgemeinen Aufbau aller Modelle der COSMO- Modellfamilie finden sich zudem in Schättler et al., 2008 [2].

### 2.1.7 CMC-GEM-L

Das Canadian Meteorological Centre (CMC) - Global Environment Multiscale (GEM) Modell des kanadischen Wetterdienstes besteht aus einer dreifach genesteten Modellkette, beginnend mit einem globalen Modell und einem Gitterpunktabstand von  $0.3^\circ$  (McTaggart-Cowan et al., 2009) [28]. Im Rahmen von MAP D-PHASE wurden zwei nicht-hydrostatische LAMs über den im Projekt geforderten räumlichen Begrenzungen als Testmodelle betreiben (R. McTaggart-Cowan et al., 2007) [27]. CMCGEML - der Buchstabe L (low) soll andeuten, dass es sich um dasjenige LAM mit der schlechteren Auflösung handelt - hat eine räumliche Auflösung von  $0,135^\circ$  (15km), die Randbedingungen werden vom globalen GEM Modell geliefert. Als weitere Modellanwendung dieser Gruppe ist hier noch das CMCGEMH - der Buchstabe H (high) für die höhere Auflösung - mit einer räumlichen Auflösung von 2,5 km für eine Prognosedauer von 18h zu nennen. Diese Modellapplikation konnte für unsere Anwendungen aufgrund der zeitlichen Differenz zu den Startzeitpunkten anderer Modelle nicht in Betracht gezogen werden, da für bessere Vergleichbarkeit der Modelle untereinander jeweils nur 00 UTC Läufe Verwendung finden, CMCGEMH aber jeweils nur um 06 UTC startet. Auf diese Voraussetzung wird weiter unten näher eingegangen.

## 2.2 Modelldaten

Die Speicherung der Ausgabedaten eines Modells kann in verschiedener Weise und in unterschiedlichen Dateiformaten erfolgen. Ausgabedaten der Modelle aus der COSMO Familie etwa werden als GRIB (Gridded Binary) abgespeichert, die Météo France Produkte ALADIN und AROME im Dateiformat FM (FASTA) oder als NetCDF. Je nach Modell erfolgt auch eine Separation der gespeicherten Dateien in Modellschichten und Bodenfelder mit zum Teil unterschiedlichen Dateiformaten. Parameter, die nicht als prognostische Variablen vom Modell berechnet werden, werden im Post-processing durch Verknüpfung der direkten Ausgabeparameter nachgerechnet, wobei oft eine Konvertierung in ein anderes Dateiformat notwendig ist. Somit liegen modellspezifisch die abgespeicherten Daten in verschiedensten Formaten vor.

### 2.2.1 Datenformat der Modelle in MAP D-PHASE

Die Vorgabe von Seiten der Organisatoren von MAP D-PHASE war eine Vereinheitlichung der Ausgabeformate sowohl in Bezug auf die direkte Weiterverwendung in “real-time” Visualisierungen und Verifikationen, als auch auf die Sicherung der Daten und Aufbewahrung der gesamten erzeugten Felder für spätere Verifikationsstudien. Als Dateiformat wird das internationale, binäre Standardformat GRIB1 in der ersten Edition verwendet, ein von der WMO (World Meteorological Organisation) festgelegter Code zur platzsparenden Speicherung gegitterter, vorwiegend meteorologischer, Daten.

Die Beschreibung und Abspeicherung der meteorologischen Größen erfolgt getrennt für Parameter die sich auf die Erdoberfläche beziehen und für jene zur Beschreibung des Zustandes der atmosphärischen Schichten, repräsentiert durch Modellflächen. Zur Unterscheidung werden in den Dateinamen jedes Termins die Abkürzungen *surf* für Parameter am Erdboden und *plev* für vertikal geschichtete Größen verwendet. Stündlich akkumulierte Niederschlagswerte und die 2-Meter Temperatur werden als eigene Datei unter der Abkürzung *tpt2* abgespeichert. Bei den in dieser Arbeit verwendeten Daten zur Untersuchung der Modellgüte handelt es sich ausschließlich um Bodenfelder, meteorologische Parameter der verschiedenen Druck- bzw. Modellflächen werden nicht verwendet und seien hier nur erwähnt. Dateien jedes Termins enthalten alle vom Modell gerechneten Felder über den gesamten modellspezifischen Vorhersagezeitraum in stündlicher Form.

## 2.3 Modellgitternetz und Struktur der VERA-Gitters

Numerische Wettervorhersagemodelle, hier im speziellen LAMs mit einer für das Projekt MAP D-PHASE spezifischen geographische Begrenzung, verwenden unterschiedliche Gitterpunkte. Einerseits ergibt sich das Gitternetz aus der verwendeten Projektion, z.B. bei der Verwendung rotierter Koordinaten mit Pol im Pazifik, andererseits gibt es keine Vorgabe, wo genau ein Gitterpunkt zu liegen hat. Ein weiterer Aspekt resultiert aus der horizontalen Auflösung mit modellspezifisch unterschiedlichen Maschenweiten.

### 2.3.1 Gitterstruktur von VERA und dessen Projektion

Bei der von VERA verwendeten Projektion handelt es sich um eine stereographische Projektion auf eine Tangentialebene, wie sie in Mayer 2010 [26] beschrieben ist. Bei dieser Art der Projektion liegt gedanklich eine auf die Kugeloberfläche tangentielle Ebene mit nur einem Berührungspunkt. Jedem Punkt auf der Kugel kann ein Punkt auf der Ebene zugewiesen werden. Der Projektionspunkt liegt dabei exakt auf der gegenüberliegenden Seite der Erdkugel, die Verbindungslinie des Projektionspunktes und des Berührungs-

punktes der Tangentialebene mit der Kugel geht durch den Kugelmittelpunkt selbst und steht senkrecht auf die aufgespannte Ebene. Somit kommt es zu einer, zum Abstand vom Nullpunkt (Berührungspunktes mit der Tangentialebene) proportionalen, Verzerrung der Längen. Da es sich hier um einen Berührungspunkt im Gegensatz zu einer Berührungslinie handelt, gibt es keine längentreue Breite wie bei anderen Projektionen, sondern nur einen einzigen, unverzerrt darstellbaren Punkt, den Nullpunkt mit den Koordinaten  $(\lambda_0, \varphi_0)$ . Das auf der Ebene festgelegte Koordinatensystem wird in kartesischen Koordinaten geführt, die Abstände zwischen den Gitterpunkten können somit direkt in der Einheit Kilometer (km) vom Nullpunkt angegeben werden.

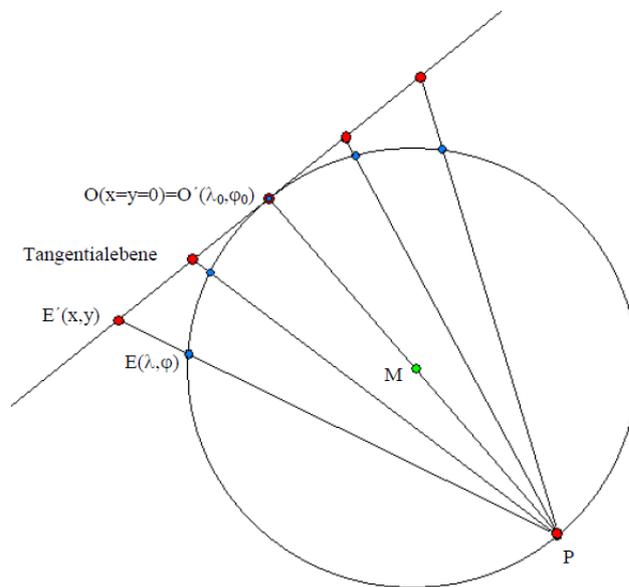


Abbildung 2.1: *Stereographische Projektion auf eine Tangentialebene. Die Tangentialebene berührt die Erdoberfläche im Punkt  $(\lambda_0, \varphi_0)$ . Jeder Punkt auf der Erdoberfläche  $E(\lambda, \varphi)$  lässt sich auf der Tangentialebene im Punkt  $E'(x, y)$  abbilden. Der Ausgangspunkt der Strahlen  $P$  bildet einen Gegenpol zum Punkt  $O=O'$ .*

VERA ist genau auf so einer Tangentialebene formuliert. Bei einer horizontalen Auflösung von 8 km entspricht das einem Gitterpunktsabstand von 8 km auf der Ebene, die Gitterlinien stehen jeweils senkrecht aufeinander. Bei der Transformation zwischen den Systemen der Tangentialebene und der geographischen Koordinaten  $(\lambda, \varphi)$  kommt es zu einer Verzerrung der Abstände, wodurch sich Gitternetzlinien nach der Projektion ins jeweils andere System krümmen und kein rechtwinkliger Zusammenhang, bis auf den Berührungspunkt, mehr besteht.

Um von geographischen Koordinaten in Koordinaten der Tangentialebene zu transfor-

mieren gelten die Beziehungen

$$k = \frac{2R}{1 + \sin \varphi_0 \sin \varphi - \cos \varphi_0 \cos \varphi \cos(\lambda - \lambda_0)} \quad (2.1)$$

$$x = k \cos \varphi \sin(\lambda - \lambda_0) \quad (2.2)$$

$$y = k [\cos \varphi_0 \sin \varphi - \sin \varphi_0 \cos \varphi \cos(\lambda - \lambda_0)] \quad (2.3)$$

wobei  $k$  einen Term definiert, der aus den Gleichungen für die horizontale und vertikale Koordinate herausgehoben werden kann.

Für die Transformation in die umgekehrte Richtung, also der Projektion von der Tangentialebene auf die Erdoberfläche lässt sich über die Hilfsvariablen

$$\rho = \sqrt{x^2 + y^2} \quad (2.4)$$

$$c = 2 \arctan\left(\frac{\rho}{2R}\right) \quad (2.5)$$

folgender mathematisch etwas aufwendigere Zusammenhang herstellen

$$\lambda = \lambda_0 - \arctan\left(\frac{x \sin c}{\rho \cos \varphi_0 \cos c - y \sin \varphi_0 \sin c}\right) \quad (2.6)$$

$$\varphi = \arcsin\left(\cos c \sin \varphi_0 + \frac{y \sin c \cos \varphi_0}{\rho}\right) \quad (2.7)$$

Hier werden beide Richtungen der Koordinatentransformation benötigt. Einerseits muss das vordefinierte 8km VERA-Gitter für eine spätere Interpolation der Modellparameter auf VERA-Gitterpunkte in geographische Koordinaten umgerechnet werden, andererseits muss aber auch eine Transformation der interpolierten Modelldaten auf VERA-Gitterpunkte der Tangentialebene erfolgen, um graphische Darstellungen besser mit VERA-Analysen vergleichen zu können.

Ausgehend vom Berührungspunkt der Tangentialebene mit der Kugeloberfläche bei  $\lambda_0 = 7^\circ$  und  $\varphi_0 = 48.8^\circ$ , stehen VERA-Analysen für einen Ausschnitt von -832 bis +832 km (209 Gitterpunkte) in x-Richtung und -768 bis +768 km (193 Gitterpunkte) in y-Richtung mit einem Gitterpunktabstand von  $\Delta x = \Delta y = 8$  km zur Verfügung. Dieser Bereich umfasst das geographische Gebiet von Mitteleuropa und beinhaltet den gesamten Alpenraum sowie den für D-PHASE relevanten Ausschnitt, in dem laut Konvention alle Modelle betrieben wurden. Ein Beispiel des von VERA analysierten Ausschnittes kann Abbildung 2.2 entnommen werden.

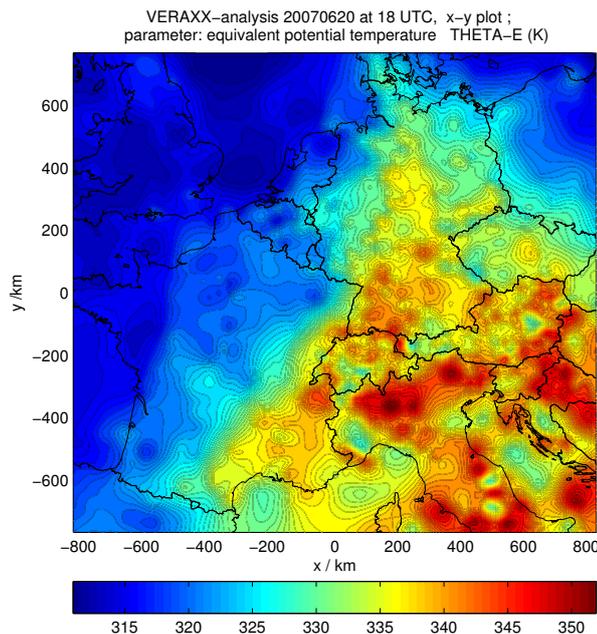


Abbildung 2.2: *Beispiel einer VERA-Analyse.  $\lambda_0 = 7^\circ$  und  $\varphi_0 = 48.8^\circ$ , Ausschnitt: von -832 bis +832 (208 Gitterpunkte) in x-Richtung und -768 bis +768 (192 Gitterpunkte) in y-Richtung mit einem Gitterpunktsabstand von  $\Delta x = \Delta y = 8$  km.*

### 2.3.2 Gitterstrukturen und Projektionen der Modelle

Ein Wettervorhersagemodell ist nicht in der Lage, Felder von Vorhersageparametern räumlich kontinuierlich auszugeben. Vielmehr erfolgt die Definition des Modells auf einer endlichen Anzahl diskreter Gitterpunkte zu festgelegten Zeitpunkten. Mit der Wahl des numerischen Verfahrens zur Diskretisierung der Bewegungsgleichungen, also der nicht-linearen partiellen Differentialgleichungen, wird oft eine Definition der Modellgleichungen auf äquidistanten Stützstellen in horizontaler Richtung vorausgesetzt, um das numerische Verfahren stabil zu halten. Im Gegensatz dazu ist in der Vertikalen die Auflösung ohnedies sehr variabel, da manche Parameter wie Druck oder Dichte exponentiell mit der Höhe abnehmen, andere wie die horizontalen Windkomponenten oder die Temperatur größenordnungsmäßig eher konstant bleiben (Haimberger, 2006) [21].

Die genaue geographische Lage der Gitterpunkte ist modellspezifisch. Abhängigkeiten ergeben sich aus der Einbindung von Randbedingungen des jeweils antreibenden Modells und einer möglichst recheneffizienten Einsetzbarkeit. Die Ausgabefelder von ALADIN, AROME und CMCGEML liegen auf einem regelmäßigen zylindrischen lat/lon Gitter vor, die Gitterpunkte verlaufen entsprechend entlang von Längen- und Breitenkreisen, wobei die Gitterpunktsabstände auf der Kugel Angaben der Winkel  $\lambda$ ,  $\varphi$  in Polarkoordinaten entsprechen. Zur graphischen Darstellung der Daten genügt es, die Koordinatenachsen in gleiche Winkelunterteilungen zu skalieren. Die Längen- und Breitenkreise stehen in dieser

Darstellung parallel zueinander, die Modelldomäne zeigt sich als Rechteck.

### 2.3.2.1 Rotiertes sphärisches Gitter

Alle Modelle der COSMO- Modellkette sind auf einem rotierten sphärischen Gitter definiert. Der Nordpol der Koordinaten unterscheidet sich dabei vom geographischen und liegt je nach Lage des Modellausschnitts im Pazifik. Die genaue Position kann Tabelle 2.1 entnommen werden. Die Verwendung eines rotierten sphärischen Gitters anstatt des üblichen geographischen Netzes hat für regionale Wettervorhersagemodelle den Vorteile, dass es sehr flexibel und rechenökonomisch eingesetzt werden kann. Durch die Verschiebung des Pols kommt es zu einer künstlichen Verlagerung des Äquators in das Modellgebiet. Die Meridiane verlaufen am Äquator selbst exakt, in geringen Breiten noch mit guter Näherung parallel, wodurch für viele Zwecke in kleinen Modellgebieten die Meridionalkonvergenz vernachlässigt werden kann. Für meteorologische Anwendungen und Darstellungen ist allerdings eine Umrechnung ins geographische System erforderlich.

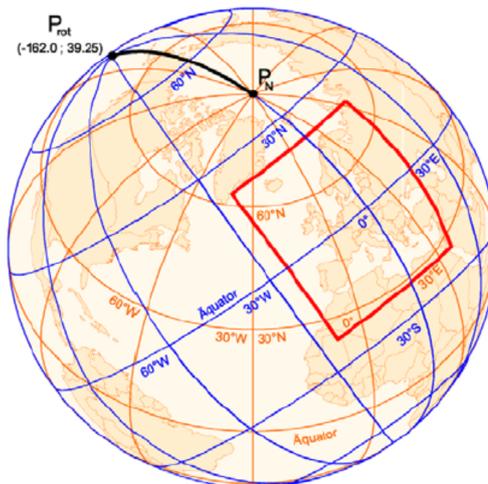


Abbildung 2.3: Demonstration der Verhältnisse eines rotierten sphärischen Koordinatensystems: Längen- und Breitenkreise des unrotierten geographischen Koordinatensystems (orange Linien) sowie rotierte Längen- und Breitenkreise (blaue Linien) im sphärischen Koordinatensystem (Quelle: Renken, 2009 [32])

| Modell   | $\lambda_N$ [°] | $\varphi_N$ [°] |
|----------|-----------------|-----------------|
| COSMO 7  | -170            | 32.5            |
| COSMO 2  | -170            | 43.0            |
| COSMO-EU | -170            | 40.0            |
| COSMO-DE | -170            | 40.0            |

Tabelle 2.1: Lage des Koordinaten-Nordpols für COSMO - Modelle

Um die Koordinaten  $(\lambda, \varphi)$  des rotierten Systems in geographische umzurechnen gilt folgende Beziehung aus der sphärischen Trigonometrie, beschrieben beispielsweise in Schulz und Schattler (2009) [35]

$$\lambda_g = \lambda_N - \arctan \left\{ \frac{\cos \varphi \sin \lambda}{\sin \varphi \cos \varphi_N + \cos \varphi_N \cos \varphi \cos \lambda} \right\}, \quad (2.8)$$

$$\varphi_g = \arcsin \{ \sin \varphi \sin \varphi_N + \cos \varphi \cos \lambda \cos \varphi_N \}, \quad (2.9)$$

wobei  $\lambda_N$  und  $\varphi_N$  dem Koordinaten-Nordpol und  $(\lambda_g, \varphi_g)$  den geographischen Koordinaten entsprechen. Bei der Programmierung in Matlab ist gegebenenfalls darauf zu achten, dass bei  $\lambda_g$  noch der Winkel von 180 Grad addiert werden muss. Der Grund dafür ist, dass in Matlab der arctan nicht in allen vier Quadranten ausgewertet wird und sonst die Lösung um  $180^\circ$  verschoben ist.

Zu beachten ist auch noch, dass die horizontalen Windkomponenten  $(u, v)$  auf das rotierte Gitter bezogen sind und sich somit von den Vektorkomponenten des zonalen Windes  $u_g$  und des meridionalen Windes  $v_g$  des geographischen Koordinatensystem unterscheiden. Die Komponenten des horizontalen Windes sind daher mit den Beziehungen

$$u_g = u \cos \delta + v \sin \delta, \quad (2.10)$$

$$v_g = -u \sin \delta + v \cos \delta, \quad (2.11)$$

ins geographische System zu transformierten.  $\delta$  bezeichnet dabei den Winkel zwischen den Längenkreisen im geographischen und im rotierten System und ist definiert als

$$\delta = \arctan \left\{ \frac{\cos \varphi_N \sin(\lambda_N - \lambda_g)}{\cos \varphi_g \sin \varphi_n - \sin \varphi_g \cos \varphi_N \cos(\lambda_N - \lambda_g)} \right\}. \quad (2.12)$$

### 2.3.3 Modellgebiete

Der Begriff Limited- Area- Model (LAM) bezeichnet wie oben erwähnt die Definition eines Wettervorhersagemodelles auf einem geographisch regionalen Bereich mit festgelegten Begrenzungslinien, somit einem abgegrenzten Gebiet auf der Erdoberfläche. Dabei wurde für MAP D-PHASE ein Bereich festgelegt, in dem alle beteiligten Modelle betrieben werden sollten. Dieser Ausschnitt umfasst ein Gebiet zwischen  $2^\circ$  und  $18^\circ$  in Ost-West Richtung, und  $43^\circ$  bis  $50^\circ$  in Nord-Süd Richtung und wird allgemein hin als D-PHASE Region bezeichnet. Abgebildet wird der gesamte Alpenbogen mit den jeweiligen Anteilen in Österreich, Slowenien, Italien, Frankreich und der Schweiz, sowie die großräumigen Einzugsgebiete der Alpen und Alpenvorländer Mitteleuropas. Die D-PHASE Region bietet

aber auch eine Überschneidung mit dem Gebiet von COPS, einem geographisch stark limitierten Gebiet an der Grenze zwischen Deutschland und Frankreich entlang des Rheins, mit den Mittelgebirgen der Vogesen in Ostfrankreich und dem Schwarzwald im Westen Deutschlands. Zu dieser geographischen Überschneidung kommt auch noch eine zeitliche für die Monate zwischen Juni und August 2007, wodurch ein weiteres Gebiet für Anwendungen und Vergleiche zur Verfügung steht.

Einige Modelle, vor allem Modelle mit sehr kleinen Gitterpunktabständen, decken nicht den gesamten von MAP D-PHASE geforderten Bereich ab. Gründe dafür sind zum Einen, dass einige Modelle operationell ohnehin in einem ähnlichen Gebiet betrieben werden und somit eine zumindest teilweise Deckung vorliegt, zum Anderen vor allem engmaschige Modelle mit Gitterpunktabständen um 2 km speziell den Anforderungen während COPS genügen und wegen der zeitlichen Übereinstimmung auch in MAP D-PHASE betrieben wurden.



Abbildung 2.4: Geographische Begrenzungen von COPS, D-PHASE und VERAXX: die hier verwendete COPS-Region entspricht nicht zur Gänze dem Gebiet von COPS während der Durchführung des Feldexperiments 2007. Quelle: Wikipedia, Alexrk2, 2010 [4], Bild verändert.

### 2.3.4 Verwendete Modellläufe

Die Vorhersagedauer eines Modells hängt von diversen Faktoren ab. Hochauflösende Modelle zielen eher auf die Vorhersage kleinskaliger Phänomene mit teilweise kurzer Lebensdauer und geringer Vorhersagbarkeit ab und sind meist nicht zur Prognose der großräumigen Zirkulationsmuster konzipiert. Die Vorhersagedauer bewegt sich bei manchen Modellen im Bereich einiger Stunden (Kürzestfristprognose), bei anderen bis hin zu wenigen Tagen, d.h. innerhalb der vorliegenden Auswahl 72h, mit der Tendenz zu größerer räumlicher Auflösung.

Grundsätzlich werden in dieser Arbeit Modelle mit einer Prognosedauer  $\leq 18$  Stunden, und jene mit einer Prognosedauer  $> 18$  Stunden jeweils zusammengefasst, wobei die zugrundeliegenden Modellläufe nur um 00 UTC betrachtet werden. somit ist zwischen zwei Experimenten zu unterscheiden: i) Verifikation 30-stündiger Prognosen und ii) Verifikation 18-stündiger Prognosen. Durch obige Beschränkung durch die Vorhersagedauer ergibt sich eine differente Anzahl an Modellen, die die Kriterien des jeweiligen Experiments erfüllen. Für 30-stündige Prognosen können vier Modelle verwendet werden, für 18-stündige die gesamte Anzahl der oben Ausgewählten, also insgesamt sieben Modelle.

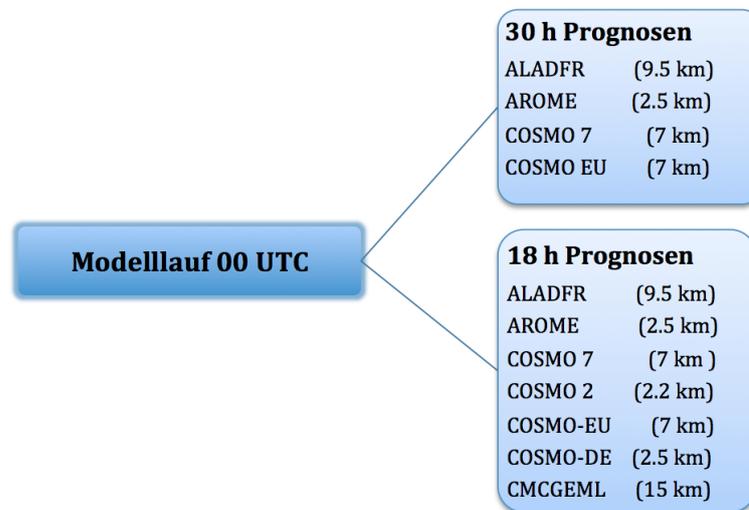


Abbildung 2.5: Verwendete Modelle für Tests mit 30-stündigen und 18-stündigen Prognosen

Manche Modelle starten mehrmals täglich, z.B. alle 3 oder 6 Stunden, andere nur einmal pro Tag. Grund für die ausschließliche Verwendung von Modellläufen mit der Initialisierungszeit 00 UTC ist die gemeinsame Verfügbarkeit aller verwendeten Modelle zu diesem Termin. Tabelle 2.2 zeigt zur Verdeutlichung die Startzeitpunkte der verwendeten Modelle.

| Modell        | Modelllauf (UTC) |    |    |    |    |    |    |    |
|---------------|------------------|----|----|----|----|----|----|----|
|               | 00               | 03 | 06 | 09 | 12 | 15 | 18 | 21 |
| ALADIN France | X                |    |    |    |    |    |    |    |
| AROME         | X                |    |    |    |    |    |    |    |
| COSMO 7       | X                |    |    |    | X  |    |    |    |
| COSMO 2       | X                | X  | X  | X  | X  | X  | X  | X  |
| COSMO-EU      | X                |    | X  |    | X  |    | X  |    |
| COSMO-DE      | X                | X  | X  | X  | X  | X  | X  | X  |
| CMC GEM L     | X                |    |    |    |    |    |    |    |

Tabelle 2.2: Modellläufe der ausgewählten Modelle

## 2.4 Anpassung und Ergänzung der Modellvariablen an VERA

Neben den horizontalen Windkomponenten  $u$  und  $v$  sowie den Niederschlagsfeldern in Form einstündig akkumulierter und qualitätskontrollierter Werte, stehen auch Felder der potentiellen Temperatur  $\Theta$ , der äquivalentpotentiellen Temperatur  $\Theta_e$ , des reduzierten Druckes  $p_{red}$  und des Mischungsverhältnisses  $m$  als operationelle Ausgabeparameter von VERA zur Verfügung. Die 2-Meter Temperatur wird einerseits zwar von allen Modellen bereitgestellt, jedoch nicht operationell von VERA ausgegeben und kann daher auch nicht für Verifikationsstudien herangezogen werden. Das Hauptaugenmerk in Bezug auf Temperaturmaße in VERA liegt in der Analyse von Feldern der potentiellen und äquivalentpotentiellen Temperatur.

### 2.4.1 Potentielle und äquivalentpotentielle Temperatur

Wird ein Luftpaket bei isentroper Zustandsänderung vertikal verlagert, das heißt von einem Druckniveau auf eine anderes gehoben oder abgesenkt, ändert sich dabei die potentielle Temperatur  $\Theta$  nicht. Das Temperaturmaß der potentiellen Temperatur ist also invariant gegenüber Druckschwankungen. Wird das Luftpaket vertikal bis zum Erdboden verlagert wo  $p = p_0 = 1000pPa$  gilt, entspricht der Wert von  $\Theta$  dem der absoluten Temperatur  $T$ .

$$\Theta = T \left( \frac{p_0}{p} \right)^\kappa \quad (2.13)$$

mit

$$\kappa = \frac{R_L}{c_p} \quad (2.14)$$

dem Isentropenexponenten als Quotienten der Gaskonstante trockener Luft  $R_L = 287,058 \frac{J}{kg \cdot K}$  und der spezifischen Wärme bei konstantem Druck  $c_p = 1004 \frac{J}{kg \cdot K}$ , wodurch  $\kappa \approx \frac{2}{7}$  ist.

Werden Kondensation und Verdunstung zugelassen, also die Luftfeuchte mitberücksichtigt, gelten bei isentroper Zustandsänderung diese Zusammenhänge nicht mehr. Das Temperaturmaß der potentielle Temperatur muss um einen exponentiellen Faktor erweitert werden. Man spricht dann von der äquivalentpotentiellen Temperatur

$$\Theta_e = \Theta \exp\left(\frac{L \cdot q}{c_p \cdot T}\right) \quad (2.15)$$

mit

$$q = 0,622 \frac{e}{p - 0,377e} \quad (2.16)$$

der spezifischen Feuchte und  $e$  dem Dampfdruck. Dabei bezeichnet  $L$  die Phasenumwandlungswärme bei Verdunstung oder Kondensation von Wasser. Die äquivalentpotentielle Temperatur dient auch als Maß des statischen Energieinhalts einer Luftmasse und eignet sich daher gut zur Identifizierung von Luftmassengrenzen. Vergleichbares trifft auf die potentielle Temperatur nicht zu, sie unterliegt innerhalb einer Luftmasse starken Schwankungen, stellt aber etwa ein gutes Maß zur Lokalisierung von Föhneffekten dar.

## 2.4.2 Reduzierter Druck

Die Ausgabe des Luftdrucks der einzelnen Modelle kann auf verschiedene Weise erfolgen. Dabei liegen die Werte der Druckfelder entweder auf der untersten Modellfläche oder der Erdoberfläche selbst vor. Bereitgestellt werden zudem auch die Felder des auf Meeresniveau reduzierten Druckes. Sehr modellspezifisch erweist sich jedoch die Art der Reduktion, die von den verschiedenen Wetterdiensten unterschiedlich gehandhabt wird. Da in den gespeicherten Daten sämtliche zur Druckreduktion benötigten Informationen enthalten sind, erschließt sich die Möglichkeit, diese für alle Modelle zu vereinheitlichen.

An dieser Stelle sei angemerkt, dass sich die nichtreduzierten Luftdruckswerte in den Modelldaten auf das Niveau der Erdoberfläche, genauer gesagt aber auf die Modelltopographie beziehen, da in dieser Arbeit ausschließlich die Bodfelder der Modelle mit der Abkürzung *surf* untersucht werden. Druckwerte der untersten Modellfläche sind demnach in den Dateien mit der Abkürzung *plev* (Druckflächen) zu finden.

Als Reduktionsmethode wird eine Standardreduktion verwendet, wie sie in Reuter et al. (1997), 162 ff [33], zu finden ist. Diese Standardmethode wird auch für die Reduktion des Druckes aus den Synop-Meldungen in VERA eingesetzt, wodurch die Vergleichbarkeit mit den auf die selbe Weise reduzierten Druckwerten der Modelle sinnvoll erscheint.

Die Methode beruht auf der Verknüpfung der idealen Gasgleichung

$$p = \rho R_L T_v \quad (2.17)$$

mit  $R_L$  der individuellen Gaskonstante für trockene Luft und

$$T_v = T(1 + 0,608 \cdot q) \quad (2.18)$$

der Virtuellen Temperatur, mit der hydrostatischen Beziehung

$$\frac{\partial p}{\partial z} = -g\rho \quad (2.19)$$

Durch Umformen der Gasgleichung nach  $\rho$  und Einsetzen ergibt sich

$$\frac{\delta p}{p} = -\frac{g \cdot \delta z}{R_L \cdot T_v}, \quad (2.20)$$

die Integration dieser Beziehung führt zur barometrischen Höhenformel einer nicht isothermen Atmosphäre

$$\ln \frac{p_0}{p} = -\frac{g \cdot (z_0 - z)}{R_L \cdot T_{v,m}}. \quad (2.21)$$

$T_{v,m}$  ist dabei die virtuelle Mitteltemperatur der Schicht. Mit der Annahme einer polytropen Atmosphäre, also einem konstanten Temperaturgefälle in vertikaler Richtung mit dem Gradienten

$$\gamma = -\frac{\partial T_v}{\partial z} = -0.0065 \frac{K}{m}, \quad (2.22)$$

kann die virtuelle Schichtmitteltemperatur geschrieben werden als  $T_{v,m} = \frac{T_{v_0} + T_v}{2}$  und man erhält schließlich die Formel für den reduzierten Druck

$$p_0 = p \cdot \exp\left(\frac{g \cdot z}{R_L \cdot T_v + R_L \gamma \frac{z}{2}}\right). \quad (2.23)$$

Wird ein geographischer Ausschnitt betrachtet, der nicht genügend klein ist um der Annahme einer konstanten Schwerebeschleunigung gerecht zu werden, wird ein weiterer Korrekturfaktor eingeführt, der sowohl die Breiten-, als auch die Höhenabhängigkeit berücksichtigt. Die Abhängigkeit der Schwerebeschleunigung von der geographischen Breite  $\varphi$  lässt sich über die Beziehung

$$g(\varphi) = 9.80616 \cdot [1 - 0.0026373 \cdot \cos(2\varphi) + 0.0000059 \cdot \cos^2(2\varphi)] \quad (2.24)$$

mit Bezug auf das Referenzniveau der Meeresoberfläche ausdrücken. Durch zusätzliche Abhängigkeit von der Höhe unter Einbindung der Breitenabhängigkeit ergibt sich schließ-

lich

$$g(h, \varphi) = g(\varphi) \cdot \left( (3.085462 \cdot 10^{-4} + (2.27 \cdot 10^{-7} \cdot \cos(2\varphi))) \cdot h \right) \cdot 0.01 \quad (2.25)$$

Die Breitenabhängigkeit ergibt sich aus der Zunahme der Zentrifugalkraft mit abnehmender Breite, sie erreicht am Äquator ihr Maximum und nimmt zum Pol hin ab. Die Höhenabhängigkeit rührt hingegen vom Abstand zum Massenmittelpunkte her, wodurch sich die Gravitationskraft in radialer Richtung vom Massenmittelpunkt weg verringert. Beide Effekte haben zahlenmäßig eher geringe Auswirkungen auf die Schwerebeschleunigung.

### 2.4.3 Mischungsverhältnis

Das Verhältnis der Dichtewerte von reinem Wasserdampf zur Dichte der trockenen Luft bezeichnet man in der Meteorologie als Mischungsverhältnis  $m$ .

$$m = \frac{\rho_w}{\rho_L} \quad (2.26)$$

Mit der spezifischen Feuchte  $q$  besteht folgender Zusammenhang

$$q = m/(1 + m), \quad (2.27)$$

wodurch sich  $m$  und  $q$  ineinander umrechnen lassen. Definiert man zusätzlich noch den Dampfdruck, also den Partialdruck des Wasserdampfes

$$e = R_w \rho_w T, \quad (2.28)$$

mit  $R_w$  der individuellen Gaskonstanten für Wasserdampf, lässt sich Gleichung 2.26 umschreiben zu

$$m = 0.622 \frac{e}{p - e}. \quad (2.29)$$

Die Konstante ergibt sich aus dem Verhältnis der beiden Gaskonstanten der trockenen Luft ( $287.04 \frac{J}{kg \cdot K}$ ) und der reinen Wasserdampfes ( $461.5 \frac{J}{kg \cdot K}$ ). Die Zahlenwerte des Mischungsverhältnisses und der spezifischen Feuchte sind größenmäßig praktisch gleich und werden leicht verwechselt. Die Dimension beider Größen ist Eins, die Angabe erfolgt häufig in  $g/kg$ . Für weiterführende Berechnungen müssen die Werte aber unbedingt in  $kg/kg$  umgerechnet werden.

#### 2.4.4 Höhenkorrektur für $\Theta$ und $\Theta_e$

VERA verwendet für die vertikale Interpretation der meteorologischen Parameter eine idealisierte Topographie, bei der in einem bestimmten Bereich (hier innerhalb einer Gitterbox von 8x8 km) jeweils die niedrigste Erhebung lokalisiert und als repräsentative Höhenstufe des Bereichs aufgefasst wird. Durch Verbinden der Talsolen und Becken wird die sogenannte Minimumtopographie definiert (Suklitsch, 2004).

Zwischen der VERA Minimumtopographie und den unterschiedlichen Modelltopographien besteht gewissermaßen eine räumlich-vertikale Diskrepanz. Auf der anderen Seite entsprechen aber auch die Modelltopographien der einzelnen Modelle selbst nicht der tatsächlichen Topographie und sind mehr oder weniger geglättet. Nach der im nächsten Abschnitt besprochenen Interpolation der Modelldaten auf VERA-Gitterpunkte liegen die Wertepaare eines Parameters zwar horizontal auf ein und demselben Gitterpunkt, vertikale Unterschiede müssen über Korrekturfaktoren aber erst angepasst werden.

Höhenkorregiert wird im speziellen die potentielle Temperatur über die Beziehung

$$\Theta_{korrr.} = \Theta - \Delta h \cdot 0.0033 \quad (2.30)$$

wobei  $\Delta h$  der Differenz zwischen der Modelltopographie und der Minimumtopographie von VERA entspricht. Der Zahlenwert von 0.0033 mit der Einheit  $[\frac{K}{m}]$  bezeichnet den vertikalen Temperaturgradienten der potentiellen Temperatur der Standardatmosphäre mit einer Abnahme von 0.33 K/100m. Bei der äquivalentpotentiellen Temperatur kann hingegen von Höhenkonstanz ausgegangen werden, wodurch keine Korrektur notwendig ist.

## 2.5 Interpolation der Modellfelder auf VERA-Gitterpunkte

Sowohl die Modellfelder, als auch die Parameter der Analyse liegen in gegitterter Form vor. Dabei verwendet jedes Modell ein eigenes, in Abhängigkeit von der jeweiligen Projektion und der verwendeten Auflösung, konstruiertes Gitternetz auf dessen Knotenpunkten die Werte der Parameter definiert sind. Analysen in Form von VERA-Feldern liegen hingegen auf einer Tangentialebene vor. Mit den in Abschnitt 2.3.1 beschriebenen Zusammenhängen ist es möglich, zwischen den Systemen der Kugeloberfläche und der Tangentialebene umzurechnen, also eine Projektion von Feldern und Gitterpunkten ins jeweils andere System. Dies führt wie schon besprochen zu einer Verzerrung der Darstellung.

Ein nicht unwesentlicher Vorteil ist, die Modelldaten sowie die Analysefelder auf ein und

demselben Koordinatensystem behandeln zu können. Es handelt sich aber immer noch um diskrete Felder, z.B. um einen Temperaturwert an einem ganz bestimmten Gitterpunkt. Um die prognostizierten Felder und die Felder der Analyse überhaupt vergleichbar zu machen, müssen Wertepaare am selben Punkt vorliegen, also an jedem Gitterpunkt jeweils ein Prognose- und ein Beobachtungswert. Auf welchem Punkt ein Wertepaar gebildet wird, ist grundsätzlich nicht von Bedeutung. Ziel ist aber die Verifikation von Modellwerten mit VERA-Analysen, wobei also die Wertepaarbildung auf VERA Gitterpunkten am adäquatesten erscheint.

Liegt der Gitterpunkt der Analyse zwischen den Punkten mit Vorhersagewerten, was im Allgemeinen der Fall ist, wird eine Interpolation zwischen den Punkten erforderlich. Den einfachsten Fall stellt die lineare Interpolation dar. Im eindimensionalen Fall werden die Funktionswerte zweier benachbarter Punkte so verbunden, dass über eine lineare Funktion die Funktionswerte an jedem Punkt im Intervall definiert sind. Wird dieses Konzept auf die nächsthöhere Dimension ausgeweitet, so müssen nicht wie bisher Geraden, sondern Flächen betrachtet werden. Definiert man die Ebene auf der die Gitterpunkte liegen als Gitterpunktsebene und die Ebene der jeweils zu den Gitterpunkten gehörenden Funktionswerte als Werteebene, so ist die Interpolation im Grunde nichts anderes als eine Projektion eines Punktes der Gitterpunktsebene auf die Werteebene, sodass jeder beliebige Punkt einem Wert in der Werteebene entspricht. Die Funktionswerte definieren also eine Fläche, oder sind zumindest einzelne Punkte dieser Fläche, wobei die genaue Festlegung abhängig von verschiedenen Ansätzen ist, in welcher Weise die Punkte zu verbinden sind. Triangulation ist einer dieser Ansätze und geht der Interpolation voraus. Dabei werden jeweils drei Punkte linear verbunden, sodass sie eine Ebene aufspannen. Das Verbinden von drei Punkten einer Punktwolke ist jedoch nicht eindeutig. Eine mögliche Einschränkung und damit Eindeutigkeit liefert das Kriterium von Delaunay, bei der der Umkreis der drei Punkte keine weiteren Punkte beinhalten darf, oder anders ausgedrückt, die Dreiecke so gewählt werden, dass es zu einer Maximierung der kleinsten Innenwinkel kommt. Bei der *Delaunay-Triangulation* wird ein Maschennetz aus Dreiecksflächen aufgespannt, die Interpolation geschieht im linearen Fall durch Auffinden des Dreieckes, das den Interpolationspunkt enthält. Zwischen den drei Eckpunkten wird linear interpoliert. In Matlab geschieht dies über den Befehl `griddata`, der eine triangulationsbasierte Interpolation, in unserem Fall linear, beinhaltet. Lineare Interpolation zählt zu den effizientesten Interpolationsmethoden, was vor allem für große Datensätze entscheidend ist und die Berechnungszeiten minimiert.

Interpoliert wird auf VERA-Gitterpunkte mit einem horizontalen Gitterpunktabstand von 8 km. Der Hauptanteil der verwendeten Modelle hat eine wesentlich feinere oder zumindest ähnliche Auflösung zwischen 2 und 10 km. Es gilt zu bedenken, dass Fehler bei

der Interpolation unvermeidbar sind. Steht mehr Information zwischen den Gitterpunkten zur Verfügung, das heißt, wird von einem feineren auf ein gröberes Gitter interpoliert, sind die Fehler eher gering. Umgekehrt sind bei der Interpolation auf ein feineres Gitter größere Defizite zu erwarten.

### 2.5.1 Gemeinsamer Bereich der jeweils verwendeten Modelle

Geographische Ausschnitte der Modelle füllen, wie in Abschnitt 2.3.3 erwähnt, nicht den gesamten geforderten Bereich aus oder sind dispositioniert. Die unterschiedliche Anzahl der Gitterpunkte nach der Interpolation auf VERA liegt zum Teil im Bereich von Faktor zwei, zwischen dem flächenmäßig größten und kleinsten Modell. Besonders im sehr komplexen Gebiet der Alpen mit starken regionalen Unterschieden und Einflüssen, kann der Vergleich versetzter oder ungleich großer Modellgebiete zu zusätzlichen Schwierigkeiten führen und ist mit Sicherheit nicht repräsentativ für die Beurteilung der Modellgüte.

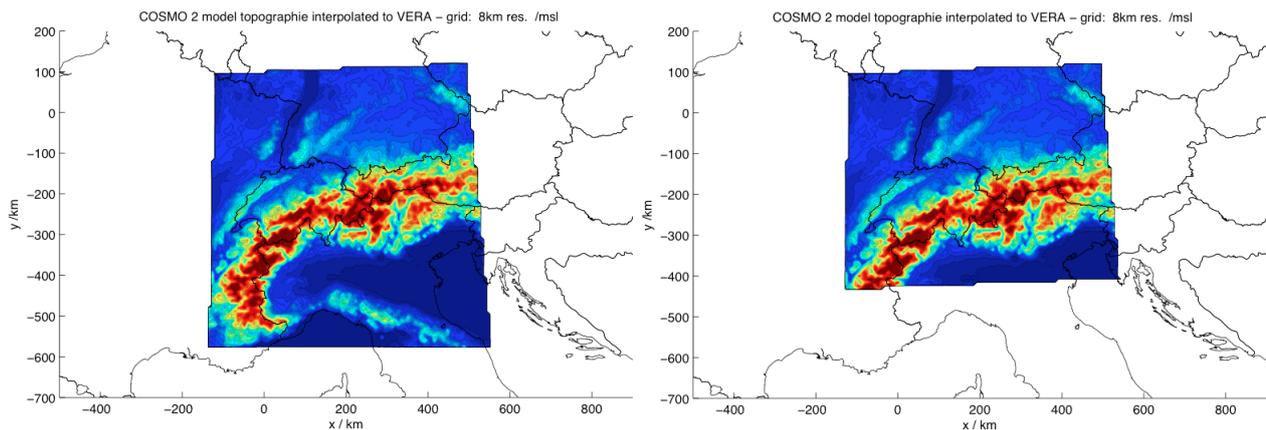


Abbildung 2.6: Gemeinsame Modellgebiete 30-stündiger Prognosen (links) und 18-stündiger Prognosen (rechts)

Für einen einheitlichen Modellvergleich werden die Felder eingegrenzt, und zwar auf das größte gemeinsame Gebiet der jeweils verwendeten Modelle. Beim Übereinanderlegen mehrerer Modellfelder ergibt sich eine charakteristische Fläche mit genau den Gitterpunkten, die von allen beteiligten Modellen geteilt werden. Für Versuche mit 30-stündigen Prognosen ergibt sich daraus ein Gebiet mit 7392 Gitterpunkten, für 18-stündige Prognosen verkleinert sich dieses Gebiet durch die Beteiligung aller Modelle der Auswahl nochmals, sodass 5432 Gitterpunkte zur Verfügung stehen. Die so eingegrenzten Modellgebiete entsprechen in weiterer Folge den Gebieten der Verifikation und sind in Abb. 2.6 dargestellt. Die Ausweitung der Gitterpunkte zur Festlegung des größeren Ausschnitts findet ausschließlich im Süden statt.

Technisch erfolgt die Definition des gemeinsamen Bereiches der Modelle über die Subtraktion von Parameterfeldern. Gitterpunkte außerhalb jedes Modellbereichs werden mit Füllwerten besetzt, d.h. NaN. Technisch erfolgt die Eingrenzung der Modelldomänen durch Multiplikation der Parameterfelder mit einer Matrix, die ausschließlich die Einträge 0 und 1 enthält. Ausschnitte kleinerer Gebiete wie COPS oder Subdomänen noch geringerer Ausdehnungen im Inneren werden nicht auf diese Weise erstellt, sondern über die Definition von Polylinien aus geographischen Koordinaten in MET, genauer beschrieben im Abschnitt 5.1.4.



## 3 Model Evaluation Tools (MET)

Alle in dieser Arbeit für den Zweck einer vergleichenden Verifikationsstudie verwendeten Verifikationsmaße werden mit MET berechnet. Die Abkürzung MET steht dabei für Model Evaluation Tools. Grob beschrieben handelt es sich um eine Sammlung von an sich unabhängigen und eigenständig arbeitenden Programmkomponenten zur Verifikation von numerischen Vorhersagemodellen (NWP-Modellen). Entwickelt wurde MET am NCAR (National Center for Atmospheric Research) -DTC (Developmental Testbed Center) in Boulder, USA.

Die erste offizielle Version von MET (METv1.0) ist im Januar 2008 erschienen, seit dem Frühjahr 2009 steht die überarbeitete Version METv2.0 zur Verfügung, welche somit die Grundlage zur Berechnung der einzelnen Verifikationsmaße bildet.

Das Ziel der Entwickler und die Philosophie hinter MET ist das Bereitstellen eines für jeden nicht-kommerziellen Zweck frei zugänglichen, jedoch “state-of-the-art”, also auf dem Stand der Entwicklung gehaltenem Software Pakets für Modellentwickler in Bezug auf die Anforderungen der Verifikation. Der Begriff “state-of-the-art” beinhaltet aber auch die Einbindung erst kürzlich entwickelter Verifikationsmethoden sowie von alternativen Ansätzen zur Ermittlung der Prognosegüte. Einer dieser neueren Ansätze ist beispielsweise eine Methode zur objektorientierten Verifikation von räumlichen Gebilden, wie sie etwa Niederschlagsgebiete darstellen. MET fasst diesen Ansatz und dem Begriff MODE (Methode for Object-based Diagnostic Evaluation) innerhalb einer eigenen Distribution zusammen (Davis et al. 2006a,b [14, 15] und Brown et al. 2007 [10]). Für eine ausführliche Beschreibung wird auf den Abschnitt 3.1.3 verwiesen. Ein weiterer neuer Ansatz zur Verifikation räumlicher Niederschlagsvorhersagen ist die Aufspaltung von Niederschlagsfeldern in ihre einzelnen Skalen über eine zweidimensionale diskrete Haar Wavelet-Zerlegung (Casati et al., 2004). Niederschlagsereignisse wie lokale Schauer und Ereignisse aus einem Frontalsystem mit unterschiedlichen physikalischen Antrieben (Konvektion oder großräumige Hebung) können so getrennt von einander betrachtet werden.

### 3.1 Überblick über die Komponenten von MET

MET gliedert sich grundsätzlich in 3 Module, jedes dieser Module enthält mehrere Komponenten, sodass insgesamt 10 Programme zur Verfügung stehen. Die drei Module lassen sich nach Abb. 3.1 unterteilen in: (i) Aufbereitung und Umformatierung der Eingangsdaten, (ii) Hauptkomponenten von MET zur Verifikation und in (iii) Analyse der Ausgabe. Die erste Gruppe fasst Programme zur Umformatierung und Veränderung der Eingangsdaten zusammen, um beispielsweise Datensätze in ein für MET lesbares Format zu bringen. Zudem sei hier angemerkt, dass sowohl die erste, als auch die momentan aktuelle Version von MET (METv2.0) ausschließlich auf Linux oder IBM basierenden Systemen betrieben werden kann.

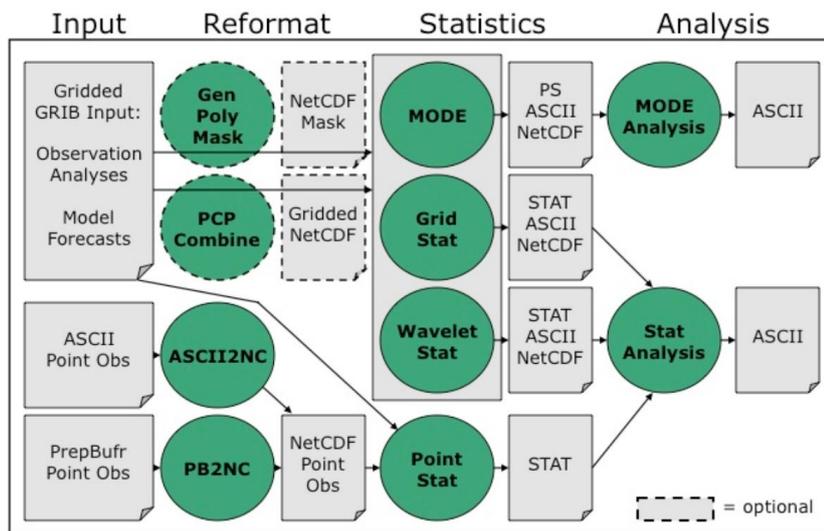


Abbildung 3.1: Struktureller Aufbau von MET mit den einzelnen Modulen. Grüne Felder repräsentieren die in MET integrierten Programme und Module, graue Felder die jeweiligen Ein- und Ausgangsformate. Quelle: MET User’s Guide, 2009 [29]

#### 3.1.1 Kurze Charakterisierung der einzelnen Programmkomponenten

##### 3.1.1.1 Umformatierung der Eingangsdaten

Alle hier angeführten Programme sind für die statistische Auswertung mit MET nicht zwingend notwendig und somit optional zu verwenden. Um jedoch Verifikation mit einer der Hauptkomponenten (siehe Abschnitt 3.1.1.2) von MET betreiben zu können, ist es wichtig, dass die Daten im richtigen Format vorliegen.

- PCP\_Combine akkumuliert stündliche Niederschlagswerte aus einer GRIB-Datei zu

Summenniederschlägen (z.B. 3-stündig, 6-stündig). Die Ausgabe der akkumulierten Werte erfolgt in eine separate netCDF Datei.

- `Gen_Poly_Mask` begrenzt den eingehenden Datenbereich auf einen vom Benutzer frei wählbaren geographischen Ausschnitt. Voraussetzung dabei ist das Vorhandensein einer Datei, die einzelne Punkte der geographischen Breite und Länge als Wertepaare beinhaltet. Somit können beliebig viele Wertepaare definiert werden, die vom Programm dann verbunden eine Fläche ergeben, solange der jeweils erste und letzte Punkt die selbe lat/lon Koordinate aufweist. In einem Unterordner von MET existieren einige dieser Dateien, jedoch nur für Nordamerika, im speziellen für die Vereinigten Staaten mit Unterteilungen in die Bundesstaaten. Für Anwendungen auf dem europäischen Kontinent wird zur Eingrenzung der Datengebiete das eigenständige Anlegen einer solchen Datei empfohlen.
- `ascii2nc` ist ein Programm zum Umwandeln von Punktbeobachtungen im Textformat, in ein für MET lesbares netCDF Format für Stationsmeldungen. Diese müssen im Textformat bezüglich der Reihenfolge der einzelnen Parameter und der Definition der Station selbst, wie etwa der Höhe über dem Meer oder der geographischen Position, einer genauen Vorgabe folgen. Jedem Parameter muss im Textformat eine eigene Zeile zugeordnet werden. Nach der Umwandlung in netCDF stehen diese Dateien für die weitere Verwendung mit `point_stat` als “Beobachtung” eines Zeitpunktes bereit. Siehe Abb. 3.1.
- `pb2nc` wandelt ähnlich wie `ascii2nc` Punktbeobachtungen in ein für MET lesbares netCDF Format um. Bei den Eingangsdaten handelt es sich jedoch diesmal nicht um Daten in einem Textformat, sondern um Beobachtungsdaten die in BUFR<sup>1</sup> gespeichert sind.

### 3.1.1.2 Hauptkomponenten zur statistischen Auswertung

Bei den in diesem Abschnitt beschriebenen Programmen handelt es sich um die vier Hauptkomponenten von MET zur statistischen Auswertung von meteorologischen Feldern. Diese setzen sich aus den unabhängigen Programmkomponenten `point_stat`, `grid_stat`, `mode` und `wavelet_stat` zusammen. Unabhängig bedeutet in diesem Zusammenhang die separate Ausführung der Komponenten ohne interne Abhängigkeiten oder Zugriffe auf andere Module oder Programmteile.

- `point_stat` kombiniert gegitterte Modellfelder mit punktuellen Beobachtungsdaten. Die Verwendung von Punktbeobachtungen erfordert allerdings ein Interpolati-

---

<sup>1</sup>BUFR (Binary Universal Form for the Representation of meteorological data)[42]

onsschema wobei die Richtung der Interpolation dabei so definiert ist, dass Werte der Vorhersage auf die Beobachtungspunkte interpoliert werden und für die Verifikation dann unregelmäßig verteilte Wertepaare von Vorhersage- und Beobachtungswerten zur Verfügung stehen. Es wird also nicht versucht, die unregelmäßig verteilten Beobachtungswerte auf ein regelmäßiges Gitter zu interpolieren. Damit unterscheidet sich dieses Verfahren grundlegend von VERA als Analyseschema. Zur Interpolation stehen dem Anwender von `point_stat` mehrere Interpolationsschemata verschiedener Ordnung zur Verfügung. Für die unregelmäßig verteilten Datenpunkte wird eine Vielzahl an konventionellen Verifikationsmaßen, sowohl für kontinuierliche, als auch für kategorische Parameter erzeugt, wobei zu den meisten auch Konfidenzintervalle berechnet werden, um die Unsicherheits- bzw. Schwankungsbreite der Maße abzuschätzen. `point_stat` benötigt die Angabe einiger stationsbezogener Parameter, z.B. ADPSFC für die Verwendung von Bodenmeldungen oder der Angabe einer Stationskennung. Die Auffassung der VERA-Analysen als Punktbeobachtungen wäre zwar grundsätzlich denkbar, die Implementierung in `point_stat` aber eher aufwendig.

- `grid_stat` ermöglicht es hingegen, regelmäßig auf einem Gitter verteilte Prognosewerte mit gegitterten Analysen zu vergleichen, solange beide auf demselben Gitter liegen. Ein Interpolationsschema wie es in `point_stat` verwendet wird ist daher nicht erforderlich. Die MET Komponente `grid_stat` berechnet im Grunde die selbe Vielzahl an Verifikationsmaßen wie `point_stat` und erfüllt dadurch die besten Voraussetzungen für eine vergleichende Verifikationsstudie hochaufgelöster numerischer Wettervorhersagemodelle mit VERA. Eine genauere Beschreibung und einige technische Details von `grid_stat` erfolgen im nächsten Abschnitt.
- MODE (Methode for Object-based Diagnostic Evaluation) ist ein Verfahren zur objektorientierten Verifikation räumlicher Niederschlagsgebiete (Davis et al. 2006a,b [14, 15] und Brown et al. 2007 [10]). Durch Anbringen eines Glättungsfilters auf die ursprünglichen Niederschlagsfelder und anschließender Festlegung eines Schwellwertes (thresholding), werden "Objekte" generiert, auf die die anfänglichen Niederschlagsdaten wieder aufgeprägt werden. Dieser Vorgang erfolgt sowohl für Vorhersage- als auch für Analysefelder in gleicher Weise, sodass ein Vergleich verschiedener Objektparameter in beiden Systemen angestellt werden kann. Objektparameter wie der Abstand der Massenschwerpunkte, die Orientierung der Objektachsen oder der Vergleich der Komplexität sind nur einige Beispiele um Niederschlagsgebilde zu klassifizieren und die Vorhersagen zu beurteilen. Mit MODE werden in dieser Arbeit einige Test durchgeführt, eine genauere Beschreibung folgt weiter unten.

- Ein weiterer neuer Ansatz zur Verifikation räumlicher Niederschlagsvorhersagen ist die Aufspaltung von Niederschlagsfeldern in ihre einzelnen Skalen über eine zweidimensionale diskrete Haar Wavelet-Zerlegung (Casati et al., 2004) [12]. Die vierte in MET implementierte Hauptkomponente `wavelet_stat` ermöglicht dem Benutzer eine getrennte Betrachtung und Verifikation von Ereignissen auf unterschiedlichen räumlichen Skalen. Meteorologische Phänomene innerhalb der einzelnen Skalen werden nicht selten von unterschiedlichen Antriebsmechanismen generiert. Betrachtet man etwa Niederschlagsereignisse aus Frontalsystemen oder konvektiv induzierte Gebilde, die durch die Skalenzerlegung vergleichbar werden und einen wesentlich tieferen Einblick bezüglich der Leistungsfähigkeit der Modelle gestattet. `wavelet_stat` wurde bis dato nicht zur Verifikation innerhalb des Projekts VERITA herangezogen, daher liegen auch keinerlei Erfahrungswerte im Umgang mit dieser MET- Distribution vor.

### 3.1.2 Grid\_stat tool

Wenn, wie im Fall der Anwendung von `grid_stat`, von einer großen Anzahl an Verifikationsmaßen als Resultat die Rede ist, beziehen sich die berechneten Maßzahlen zur Beurteilung der Modellgüte auf verschiedene Statistiken unterschiedlicher Variablengruppen. Eine Unterteilung erfolgt in Maßzahlen für kategorische Variablen, für Variablen mit kontinuierlichem Charakter und in Maßzahlen für Wahrscheinlichkeitsvorhersagen. Betrachtet man einige der wichtigsten meteorologischen Parameter wie Luftdruck, Temperatur oder Niederschlag, lässt sich allein durch die räumliche Verteilung der Parameter eine Ähnlichkeit zwischen dem Tempertur- und dem Druckfeld durch ihre Kontinuität erkennen, wohingegen die Niederschlagsverteilung als binäre Variable (JA / NEIN) gesehen werden kann und damit einen kategorischen Charakter aufweist. Jede kontinuierliche Variable kann allerdings durch Anbringen von Schwellwerten (engl.: thresholds) zu einer kategorischen Variablen umformuliert werden, z.B. Temperaturen unter dem Gefrierpunkt ( $T < 273.15\text{K}$ ). Die aktuelle Version METv2.0 beinhaltet ausschließlich die dichotome Betrachtungsweise kategorischer Variablen (d.h. 2 Kategorien), eine Ausweitung auf multi-kategorielle Vorhersagen in späteren Versionen ist geplant. Die Berechnung und Auswertung von Maßzahlen für probabilistische Vorhersagen mit Wahrscheinlichkeiten zwischen 0 und 100 Prozent für das Eintreten eines Ereignisses ist mit MET zwar möglich, wird aber in dieser Arbeit und im weiteren Kontext nicht behandelt.

Eine zentrale Größe für viele Verifikationsmaße kontinuierlicher Variablen ist der Vorhersagefehler, der als Differenz zwischen Prognose und Beobachtung, in unserem Fall Analyse, als  $(f - o)$  definiert ist und oft als Mittlerer Fehler (ME), oder auch als BIAS bezeichnet wird. In der Literatur trifft man aber häufig auf eine umgekehrte Definiti-

on des BIAS, wodurch es leicht zu Verwechslungen kommen kann. Obige Definition der Richtung des BIAS entspricht auch der Definitionsrichtung in MET und somit auch der in dieser Arbeit, wobei ein positiver Vorhersagefehler einer Überschätzung, ein negativer einer Unterschätzung des Parameters durch das Modell entspricht.

Niederschlagsereignisse werden aufgrund ihrer hohen zeitlichen und räumlichen Variabilität, vor allem aber bei Ereignissen auf kleineren Skalen, zum Teil fälschlicherweise als Fehlprognose interpretiert, obwohl das Zielgebiet womöglich nur knapp verfehlt wurde und beispielsweise ein Gitterpunkt im unmittelbaren Umfeld getroffen wurde. Je nach Anwendung und Festlegung spezifischer Kriterien an die Prognose selbst bzw. an die geforderte Genauigkeit einer Vorhersage, kann diese oftmals als durchaus akzeptabel angesehen werden, insbesondere bei hochauflösenden Modellen und häufig sehr feinskaligen Strukturen. Eine mögliche Herangehensweise an solche Probleme bieten verschiedene “neighborhood” Verifikationsmethoden, die innerhalb eines definierten Fensters eine Prognose noch als *hit* werten. MET beinhaltet die Berechnungen von Maßzahlen mittels dieser Fuzzy- Verifikationsmethoden, die genauer im Abschnitt 4.4 beschrieben werden.

Der technischen Durchführung und Steuerung der einzelnen Komponenten wird in dieser Arbeit nur wenig Aufmerksamkeit beigemessen. Im Vordergrund soll vielmehr die Darstellung der Methodik sowie die Interpretation der Ergebnisse der vergleichenden Verifikationsstudie stehen. Für eine wesentlich ausführlichere Beschreibung wird an dieser Stelle auf die technische Beschreibung [22] als Begleitliteratur zur vorliegenden Arbeit sowie auf den MET user’s Guide[29] verwiesen.

#### 3.1.3 MODE (Methode for Object-based Diagnostic Evaluation)

MET bietet neben zahlreichen konventionellen Methoden der Vorhersageverifikation auch die eingangs erwähnten neueren Ansätze einer objektorientierten Verifikation oder der Separation der Parameterfelder bezüglich deren Skalen (Casati et al., 2004) [12]. Der objektorientierte Zugang ist an sich nichts Neues. Seit längerem werden die über nur wenige Parameter beschreibbaren Ereignisse tropischer Wirbelstürme in relativ einfache Objekte übergeführt, um die Zugbahnvorhersagen beurteilen und verifizieren zu können. In der Literatur sind zahlreiche weitere Beispiele objektorientierter Verifikationsansätze für unterschiedliche Ereignisse zu finden, wie etwa Ansätze zur Verifikation von Lee- Zyklonen (Smith und Mullen, 1993) [38] oder des Seewindes (Case et al., 2004) [13], um nur einige zu nennen. Neuere Studien zielen eher auf die Verifikation hochaufgelöster Vorhersagemodelle ab, mit den dadurch zusätzlichen Problemen feinerer Strukturen. Die ersten, die Niederschlag sowohl in der Vorhersage als auch in der Beobachtung als Objekte klassifizierten und neben zahlreichen anderen Parametern beispielsweise die Entfernungen zwischen benachbarten Objekten untersuchten, waren Ebert und McBride (2000) [17]. Die in MO-

DE verwendete Methode bindet zumindest teilweise an den Überlegungen von Ebert und McBride an. Wesentliche Unterschiede bestehen vor allem in der unterschiedlichen Technik zur Objektdefinierung und der Ausweitung auf andere Objektparameter, Davis et al. (2006a) [14]. Innerhalb von MODE soll die Auffindung eines Objekts möglichst in der Weise erfolgen, in der ein menschlicher Beobachter ein Niederschlagsgebiet eingrenzen würde, also eher in schematischer und geglätteter Form. Um diesen Forderungen gerecht zu werden, sind einige Schritte notwendig, bevor mit diversen Fragestellungen der Verifikation in Bezug auf die generierten Objekte fortgefahren werden kann.

### 3.1.3.1 Objektdefinierung

Zur Auffindung einzelner Niederschlagsobjekte verwendet MODE neben der Begrenzung durch die Niederschlagsintensität einen Glättungsfilter, der auf die Originaldaten aufgebracht wird. Dieser Prozess wird auch als “convolution thresholding” bezeichnet (MET User’s Guide, 2009) [29]. Die Objektgenerierung erfolgt in drei Schritten:

1. Anbringen eines Glättungsoperators auf die originalen Niederschlagsfelder (convolution)
2. Eingrenzung der geglätteten Felder mittels eines Schwellwerts (thresholding), um Objekte zu definieren
3. Aufprägen der Rohdaten auf die geglätteten Felder (Objekte)

Durch den Glättungsoperator, in der Regel eine zirkulare Funktion, wird sichergestellt, dass die generierten Objekte an den Rändern keine scharfen Kanten aufweisen und somit die Eingrenzung weich und etwa in der Art erfolgt, wie sie ein Beobachter vornehmen würde. Zudem werden neben räumlich isolierten Niederschlagsgebieten und Bereichen geringer Intensitäten auch stark begrenzte intensive Niederschlagsgebiete gefiltert. Möglichkeiten zur Veränderung ergeben sich über die Parameter  $R$  und  $T$ .  $R$  bezeichnet hier den Einflussradius der Glättung (convolution radius) und  $T$  den Schwellwert zur Eingrenzung des geglätteten Feldes und Definition der Objekte. Die richtige Wahl der beiden Steuerungsparameter hängt in erster Linie von den anwendungsspezifischen Vorgaben ab. Ein Einflussradius von  $R = 4\Delta x$  wird in der Praxis indes als kleinster sinnvoller Wert betrachtet, da von numerischen Vorhersagemodellen nur in wenigen Fällen realistische Strukturen kleiner dem 7-fachen Gitterpunktsabstand zu erwarten sind (Skamarock, 2004) [37].

### 3.1.3.2 Klassifikation von Objektmerkmalen

Bei einer objektorientierten Verifikation ist im Gegensatz zu Gitterpunktmethoden die Beschreibung von Parametern zur Klassifikation der Objekte erforderlich. Zahlreiche dieser

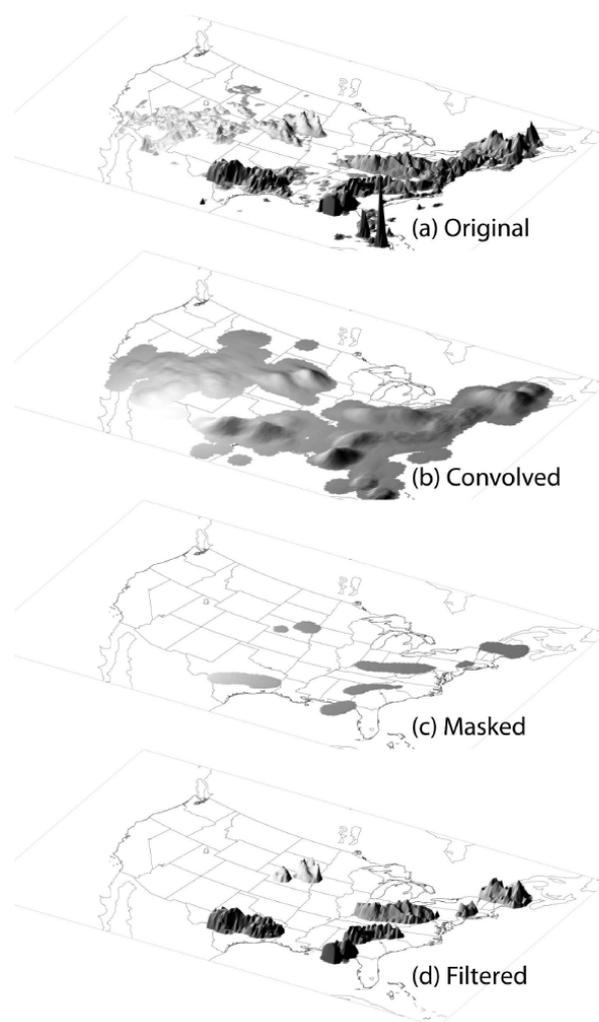


Abbildung 3.2: Schematische Darstellung der Definition von Niederschlagsobjekten mit MO-DE: a) zeigt die originalen, unveränderten Niederschlagsdaten, dreidimensionl dargestellt; b) Niederschlag nach Anbringung eines Glättungsoperators auf die originalen Daten; c) Eingrenzung des geglätteten Feldes über einen Schwellwert (threshold); d) zeigt die endgültig generierten Objekte nach Aufbringung der Niederschlagsintensitäten innerhalb der Objekte. Quelle: Davis et al. (2006a) [14], FIG. 2., p 1776

deskriptiven Objektmerkmale sind geometrischer Natur (Größe, geographische Position, diverse Formfaktoren oder die Komplexität). Die Maßzahl der Komplexität wird hier aus dem Vergleich der Objektfläche mit der Fläche der konvexen Einhüllenden gewonnen. Andere wiederum beschreiben die Niederschlagsverteilung innerhalb der Objekte. Die Definition von Objektparametern wird aus verschiedenen Gründen benötigt, einerseits um mehrere Niederschlagsobjekte innerhalb einzelner Prognose- oder Analysefelder gewissermaßen zu verschmelzen (merging), wenn angenommen werden kann, dass sie Teil eines übergeordneten Systems sind, zum Anderen um Objekte der Prognose mit Analyseobjek-

ten zu vergleichen und auf die Leistungsfähigkeit des Prognosesystems rückzuschließen (Brown et al., 2007) [10].

Neben der Beschreibung von lage- und intensitätsbezogenen Parametern einzelner Objekte werden diverse Attribute über den Vergleich der Niederschlagsgebilde von Analysen und Prognosen erzeugt. Diese paarweisen Vergleichsparameter werten im Allgemeinen Ähnlichkeiten durch Bildung von Differenzen oder Verhältnissen der Objektmerkmale auf beiden Seiten auf. Als Beispiele können hier neben zahlreichen anderen etwa die Fläche der Überdeckung oder der Abstand zwischen den Masseschwerpunkten genannt werden.

### 3.1.3.3 Gesamtrelevanz als Funktion von Objektattributen

Fuzzy- Verifikationsmethoden erlauben der Vorhersage eine gewisse Unschärfe aufzuweisen, wodurch eine Übereinstimmung mit der Beobachtung nicht unbedingt in exakter Weise bezüglich räumlicher, zeitlicher oder intensitätsbezogener Skalen erfolgen muss. Der Informationsgewinn aus diesen Methoden bzw. die Beurteilung einer Prognose liegt in der Angabe von Maßzahlen, die das Ausmaß der Übereinstimmung mit den Beobachtungen in Hinblick auf bestimmte Gesichtspunkte und die geforderten Voraussetzungen ausdrücken (vgl. 4.4). MODE verwendet sowohl zum Verschmelzen mehrere Objekte eines Feldes, als auch zum Vergleich der Objektpaare aus Analyse und Prognose ein Fuzzy-logic Verfahren. Der Unterscheid zu anderen Fuzzy-Methoden ist die Auffassung der benachbarten Ereignisse direkt als Wahrscheinlichkeit, wobei analyse- und prognoseseitig eine gewisse Wahrscheinlichkeit besteht, dass ein Ereignis entweder ein- oder auch nicht eintritt.

$$\langle I \rangle_s = \langle P \rangle_s \quad (3.1)$$

In der Beziehung 3.1 bezeichnet  $\langle I \rangle_s$  das Eintreten eines Ereignisses in der Umgebung einer Gitterbox, und  $\langle P \rangle_s$  die Abschätzung der Wahrscheinlichkeit eines Ereignisses in einer Gitterbox über den Anteil der Ereignisse in der Umgebung, wobei als Umgebung ein vorher fest definiertes Fenster gleicher Seitenlängen um jede Gitterbox bestimmt wird. Das Wahrscheinlichkeitsspektrum liegt dabei in kontinuierlicher Form zwischen 0 und 1 vor. Anders ausgedrückt bedeutet das, dass eine Prognose sowohl korrekte als auch inkorrekte Teile zum selben Zeitpunkt enthalten kann (Ebert, 2008) [16].

Brown et al. (2007) [10] zeigt eine in MODE implementierte Methode zur Berechnung der Gesamtrelevanz. Diese Maßzahl soll einen zusammenfassenden Überblick über die Relevanz eines prognostizierten Objekts hinsichtlich unterschiedlicher objektbeschreibender Parameter geben, die hier mit  $\alpha = (\alpha_1, \dots, \alpha_n)$  bezeichnet sind. Um zu beurteilen, welche Werte des i-ten Attributes wie aufzufassen sind, werden Interessenskarten  $I_{i(\alpha_i)}$  zur Zuordnung spezifischer Werte definiert, die üblicherweise den Wertebereich  $[0, 1]$  oder

[-1, 1] abdecken. Ein Wert von 0 deutet dabei auf keinerlei Interesse hin, 1 hingegen auf hohes Interesse. Innerhalb des Wertebereichs eines Attributes sind wachsende und fallende Funktionen zur Beschreibung der Interessenswerte definiert. Die Frage nach der Zuverlässigkeit eines Parameter wird über ein Konfidenzmaß  $C_i(\alpha)$  in Abhängigkeit des Attributvektors  $\alpha$  ausgedrückt und als Wert des Bereiches [0, 1] angegeben. Als Beispiel wird in diesem Zusammenhang ein Anemometer zur Messung der Richtung und Geschwindigkeit des Windes gegeben und auf die geringere Vertrauenswürdigkeit der Windrichtung bei geringen Geschwindigkeiten hingewiesen.

$$T(\alpha) = \frac{\sum_i w_i C_i(\alpha) I_i(\alpha_i)}{\sum_i w_i} \quad (3.2)$$

Bevor letztlich der Objektvergleich über den Wert der Gesamtrelevanz als Funktion des Vektors von Objektattributen über Gleichung 3.2 beschrieben wird, werden skalare Gewichtungsfaktoren  $w_i$  zur empirischen Bewertung des Einflusses der Attribute angegeben. Es besteht somit die Möglichkeit einer stärkeren Wichtung derjenigen Objektparameter, denen möglicherweise größerer Einfluss beim Vergleich der prognostizierten Objekte mit den Objekten der Analyse zur Abschätzung der Prognosequalität beigemessen werden soll als anderen. Beispielsweise wird die Schwerpunktsdifferenz gegenüber anderen Attributen häufig stärker gewichtet und über die Auffassung einer geringen räumlichen Distanz als Maß einer guten Prognose, vor allem in kräftigen Systemen, begründet.

Für eine detailliertere Ausführung und den Möglichkeiten zur Steuerung des Programms kann wieder auf die technische Beschreibung [22] als Begleitliteratur zur vorliegenden Arbeit sowie auf den MET user´s Guide[29] verwiesen werden.

#### 3.1.3.4 Gegenüberstellung von MODE und SAL

Eine erst kürzlich etablierte Methode zur objektorientierten Verifikation von quantitativen Niederschlagsprognosen ist der von Wernli et al., (2008) [43] entwickelte Ansatz SAL. Die Auswertungskriterien beschränken sich auf die drei in der Bezeichnung abgekürzten Parameter und beziehen sich im Allgemeinen auf die Beschreibung der Struktur (S), der Amplitude (A) und der Lage (L) der Objekte im Hinblick auf einen Objektvergleich. Die Herangehensweise zur Definition der Objekte sowie die Auswertung der Parameter unterscheidet sich dabei zum Teil aber grundlegend von der hier besprochenen Methode MODE oder dem früheren Ansatz von Ebert und McBride (2000).

Die SAL- Methode definiert Objekte, ebenso wie MODE, über einen Schwellwert. Die Beschränkung der Niederschlagsfelder erfolgt aber direkt, ohne die Felder in irgendeiner Weise zu glätten. Das Konzept von MODE setzt diesen Schritt allerdings zwingend

voraus, um die identifizierten Formen und Gebilde möglichst einfach zu halten. Ein weiterer wichtiger Unterschied ist die Anpassung des Schwellwerts an die Niederschlagsfelder selbst, durch Vormultiplizieren eines festen Faktors (z.B. 1/15) zum Maximalwert der Niederschlagsrate innerhalb des gesamten Datenbereichs. Infolgedessen kommt es für jeden Termin sowohl modell-, als auch analyseseitig zu einer individuellen Anpassung des Schwellwerts mit möglicherweise unterschiedlichen Werten. MODE wiederum verwendet einen einstellbaren aber festen Wert zur Untersuchung spezieller Fragestellungen bezüglich bestimmter Intensitäten. Vergleiche der generierten Objekte in beiden Feldern werden aber generell unterschiedlich gehandhabt. MODE vergleicht die Objekte zum Einen sequentiell im direkten Abgleich mit den Objekten im jeweils anderen Feld und zusätzlich über die Bildung von Clustern unter der Verwendung eines Fuzzy-logic Algorithmus, der SAL Ansatz hingegen gleicht jeweils die gesamten Domänen durch Mittelung der Eigenschaften ab.

Die Ermittlung des Amplitudenfaktors  $A$  sowie eines Teilfaktors des Lageparameters  $L_1$  ist gänzlich unabhängig von der Objektdefinition, zudem werden zur Bestimmung des Massenmittelpunkts der Niederschlagsfelder die ursprünglichen Daten ohne Beschränkung der Intensität verwendet. Der Amplitudenfaktor wird somit über die Differenz der Mittelwerte der Niederschlagsraten aus den gesamten Datensätzen der Prognose und der Analysen mit zusätzlicher Normierung über die Summe der Mittelwerte beider Felder festgelegt.  $L_1$  entspricht dem normierten Abstand der Schwerpunkte der gesamten Felder,  $L_2$  berücksichtigt zusätzlich die Schwerpunkte der einzelnen Objekte bezüglich deren Entfernung zum gemittelten Massenschwerpunkt eines Feldes. Die Strukturkomponente  $S$  vergleicht gewissermaßen die über die Maximalwerte des Niederschlags innerhalb der Objekte normierten Objektvolumina. Über die Normierungen sind die verschiedenen Parameter innerhalb bestimmter Wertebereiche beschränkt und vergleichbar. Die Motivation zur Verwendung von Relativmaßen liegt in der direkten Vergleichbarkeit quantitativer Niederschlagsprognosen unterschiedlicher Intensitäten.

MODE berechnet im Gegensatz dazu noch eine Reihe weiterer Attribute wie etwa Winkeldifferenzen, Überschneidungsflächen oder ein Maß der Komplexität der Gebilde, jedoch stets auf einzelne Objekte oder auf mehrere Objekte innerhalb eines Clusters bezogen. Es wird hier auch versucht, die einzelnen Attribute zusammenfassend über die Gesamtrelevanz aus dem Vergleich zweier Objekte oder Cluster darzustellen.



## 4 Qualitativer Verifikationszugang

Bei der Verifikation in den zur Verfügung stehenden Gebieten von D-PHASE oder COPS wird nicht allen Parametern die gleiche Bedeutung beigemessen. Während COPS stand vor allem die Untersuchung orographisch induzierter und konvektiver Niederschlagsereignisse im Vordergrund. Ähnlich gesteckt waren auch die Ziele von MAP D-PHASE zur hydrologischen und atmosphärischen Untersuchung und Simulation von Starkniederschlagsereignissen im Alpenraum. Der Fokus in Bezug auf die Modellierung in beiden Feldexperimenten galt also der richtigen qualitativen und quantitativen Erfassung von Niederschlag und betont die zentrale Rolle dieses Parameters. Somit ist es auch sinnvoll, der Bevorzugung von Niederschlag gegenüber anderen meteorologischen Parametern nachzugeben und in Analogie mit COPS und MAP D-PHASE auch die Verifikation dahingehend zu forcieren, ohne aber die verbleibenden Parameter gänzlich zu vernachlässigen.

Eine Stratifikation der Methoden erfolgt hier grundsätzlich über einen rein qualitativen, oder rein quantitativen Zugang zur Verifikation. Qualitative Auswertungen kategorischer Variablen (z.B. Niederschlag) ermitteln vor allem räumliche Charakteristika, ohne jedoch auf eine mengenmäßige oder richtungsbezogene Abweichung Rücksicht zu nehmen. Im Vordergrund steht die Evaluierung einer Prognose ausschließlich über das räumliche und zeitliche Eintreten oder Nichteintreten eines Ereignisses. Aussagen über eine mögliche mengenmäßige Über- oder Unterschätzung, die Magnitude der Fehler sowie deren Orientierung sind über eine quantitative Analyse der Prognosen sowohl für kategorische als auch kontinuierliche Variablen zu bestimmen. Ziel dieses Kapitels ist aber eine statistische Auswertung zur Beantwortung spezifischer Fragestellungen der qualitativen Verifikation, Auswertungen in Bezug auf die quantitative Erfassung der Prognosegüte erfolgt im nächsten Kapitel.

### 4.1 Die Kontingenztabelle

Im Mittelpunkt bei der qualitativen Verifikation steht eine Kontingenztabelle. Sie stellt gewissermaßen eine Methode dar, um prognostizierte und beobachtete Ereignisse für kategorische Variablen in Bezug zu setzen. Die Einträge der Kontingenztabelle bilden häufig die Basis diverser Verifikationsmaße. Variablen mit kategorischem Charakter sind wie

oben beschrieben solche, bei denen ein Ereignis, z.B. Niederschlag, entweder eintritt oder nicht eintritt, im Gegensatz zu kontinuerlichen Variablen wie Druck oder Temperatur als integrale Größen mit gleichmäßigen, stetigen Verteilungen. Jede kontinuierliche Variable kann aber durch Anbringen eines Schwellwerts als kategorische aufgefasst werden. Der einfachste Fall einer Kontingenztabelle unterscheidet vier Fälle: Die Anzahl der Gitterpunkte bei denen i) ein prognostiziertes Ereignisses auch eingetreten ist (*hits*), ii) ein nicht vorhergesagtes Ereignis auch nicht eingetreten ist (*correct rejections*), iii) sowie die beiden gemischten Terme einer Vorhersage, die nicht eingetreten ist (*false alarm*), iv) und umgekehrt der Beobachtung eines Ereignisses, das nicht prognostiziert wurde (*misses*). Eine

|          |     | Observed      |                     | total        |
|----------|-----|---------------|---------------------|--------------|
|          |     | yes           | no                  | total        |
| Forecast | yes | <b>hits</b>   | <b>false alarms</b> | fcst. yes    |
|          | no  | <b>misses</b> | <b>cr</b>           | fcst. no     |
| total    |     | obs. yes      | obs. no             | <b>total</b> |

Tabelle 4.1: *Kontingenztabelle: Beschreibung der orangen Felder im Text. obs. yes und obs. no bezeichnen die jeweiligen Spaltensummen, fcst. yes und fcst. no die Zeilensummen. Summe aller Gitterpunkte als Summe der Spalten- oder Zeilensumme wird mit total bezeichnet.*

solche Tabelle stellt an sich keine eigene Verifikationsmethode dar, vielmehr kann mit den Einträgen an den einzelnen Positionen eine Vielzahl an statistischen Maßen (*scores*) berechnet werden.

$$POD = \frac{hits}{hits + misses} \tag{4.1}$$

$$POFD = \frac{falsealarms}{falsealarms + cr} \tag{4.2}$$

POD (probability of detection) oder die hit rate setzt *hits* (JA/JA) ins Verhältnis zur Summe aus *hits* und nicht vorhergesagten, aber eingetretenen Ereignissen (*misses*). Der Nenner gibt die Anzahl der Gitterpunkte wieder, bei denen ein Ereignis tatsächlich eingetreten ist und könnte ebensogut direkt der Analyse entnommen werden. POD gibt den Anteil der richtig prognostizierten zur gesamten beobachteten Anzahl der Ereignisse wieder und erreicht den Bestwert von  $POD = 1$ , wenn in einem Vorhersagegebiet alle prognostizierten Ereignisse auch beobachtet werden können. POFD (probability of false detection) oder die *false alarm rate* setzt nicht eingetretene Prognosen ins Verhältnis zur Summe der nicht eingetretenen mit den nicht vorhergesagten und auch nicht beobachte-

ten Ereignissen. Der Nenner liefert hier den Anteil der Gitterpunkte, die nicht von einem Ereignis betroffen sind. Dieser Wert kann auch wieder direkt aus der Analyse ermittelt werden. Somit kann mit Hilfe des POFD der Anteil der nicht eingetretenen Ereignisse bestimmt werden, die inkorrekt prognostiziert wurden. Anders ausgedrückt: Anteile von Fehlalarmen an den Gitterpunkten ohne Niederschlag.

## 4.2 ROC (Relative Operating Characteristic)

Bei der Anwendung von MET zur Verifikation der Modelle mit VERA-Analysefeldern wird für jeden Zeitschritt, in unserem Fall für einen bestimmten Termin an jedem Tag, eine eigene Kontingenztabelle angelegt und POD und POFD berechnet. Die jeweils eingetragenen Werte stammen aus dem Vergleich des Modellfeldes mit dem Analysefeld für alle übereinstimmenden Gitterpunkte, die, wie im Abschnitt 2.5.1 beschrieben, nur von der Anzahl der verwendeten Modelle des jeweiligen Experiments abhängt. Die graphische Gegenüberstellung von POD und POFD führt zu der häufig verwendeten Darstellung einer ROC- Kurve (Relative Operating Characteristic) wie sie etwa in Mason und Graham (1999) [25] oder in Fawcett (2006) [18] zu finden ist.

### 4.2.1 Interpretation von ROC für ein Modell

Die von MET berechneten POD- und POFD- Werte können in einem ROC- Diagramm mit der *hit rate* auf der Ordinate und der *false alarm rate* auf der Abszisse eingetragen und jedem Tag eines Monats ein bestimmter Punkt zugeordnet werden. Die Lage des jeweiligen Punktes gibt Auskunft über die Qualität der Prognose eines Termins. Liegt der Punkt in der linken oberen Ecke, bedeutet das zum Einen, dass an keinem Gitterpunkt Niederschlag aufgetreten sein darf der nicht prognostiziert wurde (*misses=0*), zum Anderen, dass an keinem Punkt Niederschlag prognostiziert sein darf, der nicht eingetreten ist (*false alarms=0*). In so einem Fall ist  $POD=1$  und  $POFD=0$ , der dadurch entstehende Punkt kennzeichnet den Bestwert. Punkte mit einem Verhältnis von  $POD/POFD=1$  (*likelihood ratio*) liegen auf einer 45° Geraden und haben keinen Skill. Unterhalb liegende Punkte haben negativen Skill (Mason und Graham, 1999)[25].

Um Aussagen über die Prognosequalität eines Modells über einen längeren Zeitraum zu tätigen, können auch Punkte eines Mittelungszeitraums eingetragen werden. Dabei ist zu beachten, dass die POD- und POFD-Werte der einzelnen eingetragenen Punkte (Tage) nicht einfach arithmetisch gemittelt werden können. Durch Summieren der Einträge aus den Kontingenztabelle der einzelnen Tage entsteht eine neue Tabelle, die für die Berechnung einer neuen *hit rate* und einer *false alarm rate* herangezogen werden. Die Frage, warum die POD- oder POFD - Werte nicht einfach summiert und durch die Anzahl der

Tage ( $n$ ) dividiert werden können, lässt sich dadurch beantworten, dass die Spaltensummen in der Kontingenztabelle, also die Werte im Nenner, differieren. Am Beispiel der Berechnung von POD zeigt sich eine Variation der Werte im Nenner, da sich die Anzahl der Gitterpunkte mit Niederschlag für jeden Termin verändert. Damit ändert sich sozusagen die Grundgesamtheit für die Berechnung der relativen Werte. Gleich bleibt in jeder Tabelle immer nur die Gesamtsumme aller vier Felder und symbolisiert die Anzahl der übereinstimmenden Gitterpunktspaare zwischen Modell- und Analysefeld. Die über einen Zeitraum gemittelten Werte für POD und POFD berechnen sich nach Gleichung 4.3 bzw. Gleichung 4.4.

$$POD_{gesamt} = \frac{\sum_{i=1}^n hits}{\sum_{i=1}^n hits + \sum_{i=1}^n misses} \neq \frac{POD_1 + POD_2 + POD_3 \dots}{n} \quad (4.3)$$

$$POFD_{gesamt} = \frac{\sum_{i=1}^n false\ alarms}{\sum_{i=1}^n false\ alarms + \sum_{i=1}^n cr} \neq \frac{POFD_1 + POFD_2 + POFD_3 \dots}{n} \quad (4.4)$$

Werden alle Werte, wie in Abbildung 4.1 dargestellt, in ein ROC- Diagramm eingetragen, so lässt sich an den mit Zahlen nummerierten Punkten ablesen, wie ein Modell an einem bestimmten Tag (Termin) gearbeitet hat, beziehungsweise durch Eintragen des Punktes der einen Zeitraum repräsentiert, die Arbeitsweise innerhalb einer Zeitspanne. Das Balkendiagramm im unteren Bildabschnitt der Abbildung 4.1 zeigt die absolute Anzahl der Gitterpunkte mit Niederschlag (rot) und jener ohne Niederschlag (gelb), die Summe beider Werte ergibt die Gesamtzahl der beteiligten Gitterpunkte, d.h. 5435. Die Niederschlagsschwelle liegt bei 0.1 mm und entspricht der unteren Grenze der geforderten Messgenauigkeit bei der Niederschlagserfassung. Je näher ein Punkt der linken oberen Ecke kommt, umso weniger Fehlalarme (geringer POFD) bei einem gleichzeitig hohem Anteil an richtigen Prognosen (hoher POD) werden vom Modell ausgegeben.

Ein sehr nützlicher und häufig verwendeter relativer Index ist die Fläche unterhalb der ROC-Kurve (Mason und Graham, 1999) [25]. Da in unserem Fall immer nur ein Punkt in der Graphik verwendet wird, die ROC-Kurve aber eigentlich aus der Verbindung des Ursprungs (0/0) mit dem jeweiligen Punkt und weiter zum Punkt mit den Koordinaten (1/1) entsteht, lässt sich die Fläche aus einfachen geometrischen Überlegungen berechnen. Dabei besteht die Gesamtfläche aus zwei Teilflächen, der Fläche unterhalb der 45°-Geraden mit genau der halben Gesamtfläche und der eines allgemeinen Dreiecks. Wenn die Summe dieser beiden Flächen größer 0.5 ist, wird vom Modell eine brauchbare Prognose erstellt.

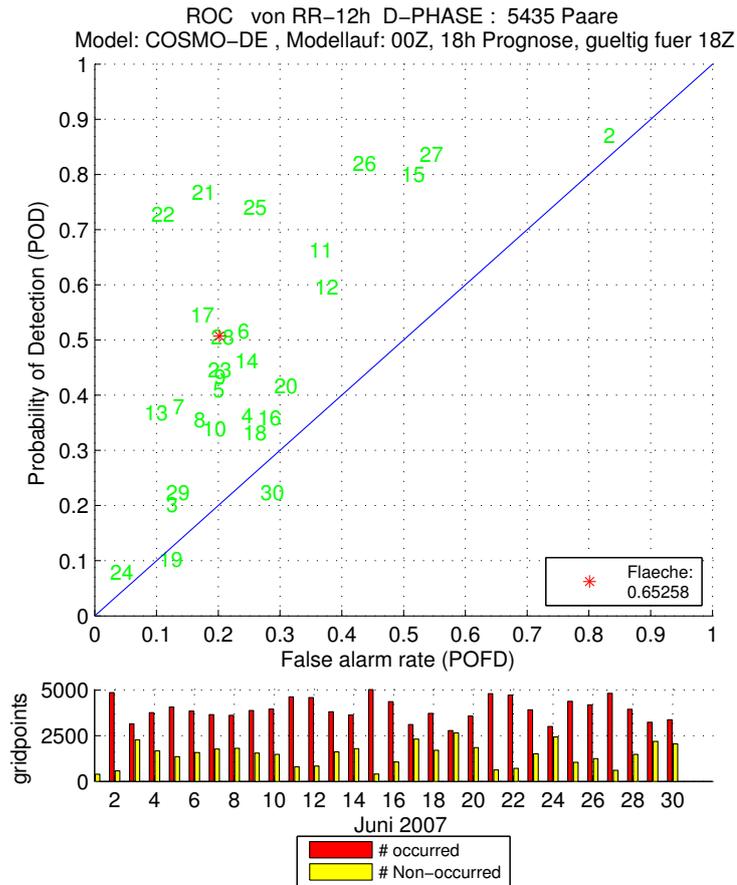


Abbildung 4.1: *Relative operating characteristic (ROC) des Modells COSMO-DE für 12-stündig akkumulierte Niederschlagswerte aus 18-stündigen Prognosen im Juni 2007 über den gemeinsamen Ausschnitt aller sieben Modelle. Fläche unter der Kurve der Verbindungslinie (0/0), \*, (1/1), Zahlen (grün) repräsentieren den Punkt am jeweiligen Tag. Der zu einem Zahlenwerten gehörende Punkt ist an der linken unteren Ecke der Ziffer zu verstehen. Niederschlagsschwelle: >0.1mm/12h*

Punkte in der linken oberen Ecke erreichen eine Flächenwert von 1 und kennzeichnet den Maximalwert des Parameters.

$$\alpha = \arcsin \left( \frac{POD}{\sqrt{POD^2 + POFD^2}} \right) - \frac{\pi}{4} \quad (4.5)$$

$$h = \sqrt{POD^2 + POFD^2} \cdot \sin \alpha \quad (4.6)$$

$$F = \frac{ah}{2} = \frac{\sqrt{2}h}{2} \quad (4.7)$$

$$F_{ges} = F + 0.5 \quad (4.8)$$

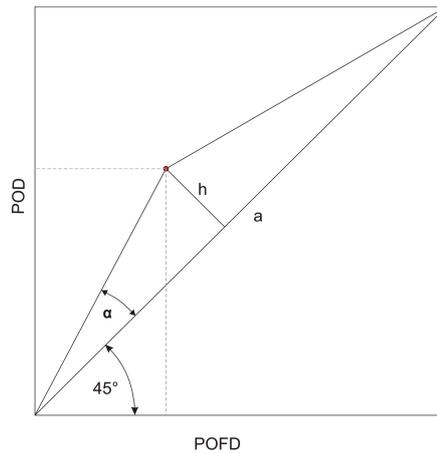


Abbildung 4.2: Schematische Darstellung der Verhältnisse zur Berechnung der Fläche unter der Kurve

Die in den Gleichungen 4.5 und 4.6 verwendeten POD- und POFD-Werte entsprechen den aktuellen Werten des jeweiligen Punktes.

Die Anwendung auf Abb. 4.1 ergibt einen Flächenindex von 0.653 für den mit \* gekennzeichneten Punkt des Monatsmittels von COSMO-DE für Juni 2007 innerhalb des D-PHASE Ausschnitts. Bei der Betrachtung der Verteilung zeigt sich eine hohe Dichte an Punkten im zweiten und dritten Quadranten. Einige Tage wie etwa der 21. oder 22. Juni erreichen sehr gute Werte, angezeigt über die relative Nähe zum Bestwert. Ein rein subjektiver Modellvergleich anhand der Verteilungen der Punktwolken zeigt ebenso für COSMO 2 (hier nicht dargestellt) eine sehr ähnlich dichte Anhäufung an Punkten und nur wenige Tage mit hohen POFD- Werten. Verteilungen der anderen Modelle sind hingegen wesentlich breiter entlang der 45° Geraden, wobei ALADIN France dabei die schwächsten Resultate mit einer deutlichen Neigung zu hohen POFD liefert.

#### 4.2.1.1 Ausweitung auf größere zeitliche Intervalle

Das eben vorgestellte Verfahren zur Untersuchung und Darstellung der Verhältnisse einzelner Tage, im Speziellen aber bestimmter Termine, in einem ROC- Diagramm lässt sich auch auf größere Zeitintervalle ausweiten. Ein wichtiges Kriterium ist dabei aber der Erhalt der Übersichtlichkeit aufgrund der hohen Informationsdichte in der Darstellung, sodass ein Vergleich mehrerer Monate einer Modellanwendung noch sinnvoll ist.

Das in der Abbildung 4.3 dargestellte ROC- Diagramm zeigt den zeitlichen Vergleich des DWD Modells COSMO-DE über eine Periode von drei Monaten zwischen Juni und August 2007 innerhalb des D-PHASE Ausschnitts. Der mit (\*) gekennzeichnete Punkt entspricht dem Mittelwert des gesamten Zeitraums. Zur optischen Separierbarkeit der Einträge werden unterschiedliche Farben für die drei Monate eingeführt. Das Balkendia-

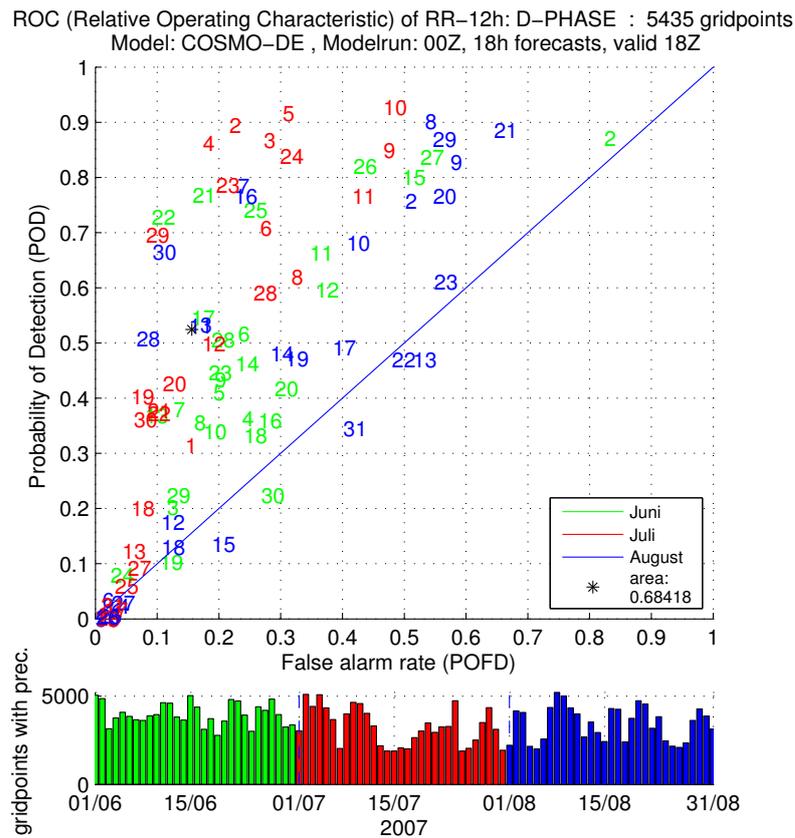


Abbildung 4.3: ROC des Modells COSMO-DE für 12-stündig akkumulierte Niederschlagswerte aus 18-stündigen Prognosen zwischen Juni und August 2007. D-PHASE Gebiet: gemeinsamer Ausschnitt aller sieben Modelle. Fläche unter der Kurve der Verbindungslinie (0/0), (\*), (1/1), Zahlen repräsentieren den jeweiligen Tag, unterschiedliche Farben die Monate. Der zu einem Zahlenwert gehörende Punkt ist an der linken unteren Ecke der Ziffer zu verstehen. Niederschlagsschwelle:  $>0.1\text{mm}/12\text{h}$

gramm im unteren Bildabschnitt zeigt wieder die Gitterpunkte mit Niederschlag in der Analyse und einer Akkumulationszeit von 12 Stunden.

Ein wieder eher subjektiv angestellter Vergleich zeigt eine recht gute Erfassung der Niederschlagsereignisse und eine brauchbare Modellkonzeptionierung für Juli mit hohen Trefferzahlen bei gleichzeitig eher geringen Werten von POFD (rot). Die Verteilung der Werte im August (blau) ist hingegen wesentlich breiter, die Anteile der Fehlalarme an den nicht eingetretenen Ereignissen deutlich größer, sodass insgesamt der Eindruck einer leichten Annäherung der Werte an die Diagonale entsteht. Die auch schon im letzten Abschnitt beschriebene Verteilung der Punkte im Juni (grün) liegt zwischen den beiden anderen Monaten.

Die Anwendung des ROC-Diagramms zur zeitlichen Gegenüberstellung beschränkt sich einerseits auf den Vergleich von nur wenigen Monaten. Auf der anderen Seite erscheint

aber auch ein Modellvergleich durch Eintragen mehrerer Modelle in ein ROC- Diagramm dieser Art nicht zweckmäßig, vor allem bei der Untersuchung längerer Perioden.

### 4.2.2 ROC für mehrere Modelle

Für ein Experiment, z.B. den Vergleich von 18-stündigen Prognosen eines Parameters über einen Monat, wird, wie schon mehrmals erwähnt, für alle Modelle die gleiche Anzahl an Gitterpunkten verwendet. Alle berechneten statistischen Maße zur Untersuchung der Qualität eines Prognosemodells sind also repräsentativ für exakt dasselbe Gebiet. Somit ist es möglich, die in Abschnitt 4.2.1 eingeführten mittleren POD- und POFD-Werte für ein Monat in ein Diagramm einzutragen und direkt zu vergleichen.

$$\text{likelihood ratio} = \frac{POD}{POFD} \quad (4.9)$$

Für den auch schon weiter oben angesprochene Fall, dass ein Punkt unterhalb der 45° Geraden liegt, also nach Gleichung 4.9 das *likelihood-Verhältnis*  $< 1$  ist und das Prognosesystem demnach negativen Skill hat (Mason und Graham, 1999)[25], ist nach Gleichung 4.5 der erste Term kleiner als  $\pi/4$  und somit  $\alpha < 0$ . Durch Mitführen des negativen Vorzeichens kommt es zu einer negativen Fläche des allgemeinen Dreiecks ( $F < 0$ ), wodurch nach Summieren der beiden Dreiecksflächen eine Gesamtfläche kleiner als 0.5 resultiert.

Bei manchen Modellen kommt es innerhalb eines Monats vor, dass an einzelnen Tagen entweder kein, bzw. ein fehlerhafter und nicht übermittelter Modelllauf stattgefunden hat und dadurch die Anzahl der Einträge zur Erstellung der neuen Kontingenztabelle leicht variieren kann. Für einen sauberen und exakten Modellvergleich dürften hinsichtlich der Berechnung der einzelnen Maße strenggenommen nur jene Tage einfließen, an denen alle beteiligten Modelle auch Werte geliefert haben. Die Anzahl der nicht vorhandenen Modellläufe beschränkt sich auf weniger als zwei innerhalb eines Monats und auf weniger als sechs innerhalb des gesamten betrachteten Zeitraums von sechs Monaten im Sommer und Herbst 2007. Die Abschätzung des relative Fehlers, der durch diese etwas unsaubere Betrachtung in Kauf genommen wird und in die Ergebnisse eingeht, gestaltet sich etwas schwierig. Beim Vergleich zwischen zwei Modellen wo jeweils zwei Läufe an unterschiedlichen Tagen eines Monats fehlen, kann als der Fall mit dem größten Fehlerpotential angesehen werden. Zwischen den beiden Modellen ergibt sich ein Unterschied von vier Läufen, der sich bei einer Monatsdauer von 30 Tagen als knapp 13 % Beteiligungsunterschied ausdrücken lässt. Hinzu kommt allerdings, dass sich ein fehlender Tag in den Modelldaten nicht zwangsläufig mit einem anderen vergleichen lässt, da es in jedem Fall einen Unterschied macht, ob Niederschlag aufgetreten ist oder nicht. In der Regel kann aber von einem wesentlich geringeren Fehlermaß als dem oben genannten Beteiligungsun-

terschied ausgegangen werden, der nur dann eintreten würde, wenn ein Modell am selben Tag bei gleichzeitiger Absenz eines anderen fehlerfrei arbeiten würde. Wichtig ist hier aber festzuhalten, dass diese Fehler in den Ergebnissen zwar nicht quantitativ berücksichtigt werden, die Fehlerquelle an sich aber als Unsicherheitsfaktor bei der Interpretation berücksichtigt werden sollte.

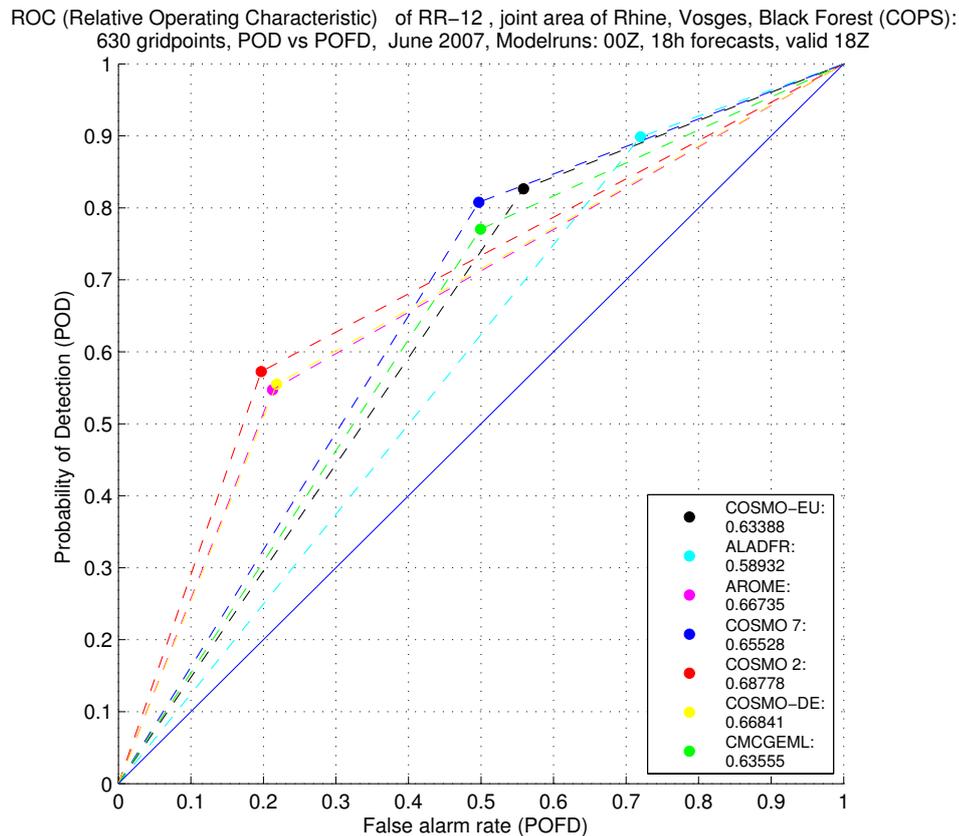


Abbildung 4.4: *Relative operating characteristic (ROC) für 12h akkumulierten Niederschlag aller ausgewählten Modelle mit einer Prognosedauer von 18 Stunden. Modellläufe um 00UTC für die Region COPS. Zahlenwerte in der Legende unterhalb der Modellbezeichnungen entsprechen der Fläche unter der Kurve. Niederschlagschwelle:  $>0.1\text{mm}/12\text{h}$*

Für eine qualitative Verifikation von Niederschlagsfeldern empfiehlt sich zur gemeinsamen Darstellung mehrerer Modelle in einer Graphik eine Darstellung der ROC-Kurve, in der jeweils die zeitlich gemittelten Punkte der einzelnen Modelle eingetragen sind. Abb. 4.4 zeigt die monatsgemittelten Werte von POD und POFD für Juni 2007 im Gebiet COPS für 12-stündig akkumulierten Niederschlag, jeweils um 18 UTC, aller beteiligten Modelle. Dabei zeigt sich ein signifikanter Unterschied. Offenbar gibt es eine Gruppe von Modellen, die wesentlich geringere Werte der *false alarm rate* sowie auch geringere POD-Werte aufweisen als andere. Bei dieser Gruppe handelt es sich um zwei Modelle der COSMO Familie, COSMO 2 und COSMO-DE, sowie um das französische Forschungsmodell AROME, also

um diejenigen Modelle, die innerhalb der Auswahl die kleinsten Gitterpunktabstände und somit die höchsten Auflösungen haben. Der Unterschied zu den anderen Modellen liegt neben der feineren räumlichen Auflösung vor allem bei der unterschiedlichen Behandlung von Niederschlag in Bezug auf die Parametrisierung der Prozesse, die vom Modell wegen zu kleinen räumlichen Strukturen nicht aufgelöst werden können. Die kleinsten auf einem Gitter auflösbaren Skalen entsprechen dem doppelten Gitterpunktabstand, alle kleineren Skalen müssen parametrisiert werden. Für ein numerisches Wettervorhersagemodell mit einem Gitterpunktabstand von 9 km bedeutet das in Bezug auf die Konvektion, dass ein Phänomen mindestens eine Größe von 18 km haben muss um explizit aufgelöst werden zu können. Bei feineren Modellen mit 2 km Auflösung liegt die kleinste aufgelöste Skala bei 4 km, also durchaus schon in der Größenordnung der Cumuluskonvektion. Wie schon im Kapitel 2.1 beschrieben wird für diese Modelle auf eine physikalische Parametrisierung für hochreichende Konvektion gänzlich verzichtet. Vom Modell wird also eine explizite Auflösung zumindest hochreichender Konvektion in den groben Strukturen gefordert, flache Konvektion hingegen wird weiterhin parametrisiert.

### 4.2.2.1 Aspekte zur Vergleichbarkeit unterschiedlicher Modelle

Bei der Betrachtung von Abb. 4.4 oder ähnlichen Darstellungen werden einige Fragen aufgeworfen, die es im nächsten Schritt zu beantworten gilt. Zum Einen stellt sich die Frage nach dem Zusammenhang zwischen POD, POFD und der Modellauflösung mit oder ohne der Verwendung eines Parameterisierungsschemas für hochreichende Konvektion, zum Anderen die Frage, ob die Flächen unterhalb der Kurven unterschiedlicher POD und POFD-Werte die selbe Signifikanz haben. Die gröbere Auflösung der Modelle führt vor allem bei konvektiven Ereignissen zu einer flächenmäßigen Überschätzung der Niederschlagsgebiete, da die feinen Strukturen auch durch die Parametrisierung nicht zur Gänze aufgelöst werden können. Gleichzeitig werden aber oft die Niederschlagsraten an den Gitterpunkten, an denen Niederschlag aufgetreten ist, unterschätzt. Anders formuliert kommt es zur Glättung der kleinräumigen Niederschlagsgebiete und zum Über- und Unterschwingen. Genau das zeigt sich auch an den POD bzw. POFD-Werten. Durch die räumliche Überschätzung der Größe von Niederschlagsgebieten ist es einfacher, einen Treffer zu erzielen, der Wert von POD ist hoch. Auf der anderen Seite deuten zu großflächig prognostizierte Niederschlagsfelder auf eine Unsicherheit seitens des Modells hin, wodurch der Anteil der "falschen Alarme" an den Gitterpunkten ohne positiver Beobachtung (POFD) auch groß ist. Der so entstehende Punkt im ROC-Diagramm liegt im rechten oberen Quadranten. Bei den Konvektionserlaubenden Modelle hingegen liegen die Verhältnisse etwas anders. Die kleineren und teilweise realistischeren Strukturen in der Prognose führen zu einer signifikanten Abnahme der POFD-Werte, die Vorhersagen sind schärfer und der Anteil an

Überprognosen nimmt ab. Demzufolge kommt es auch zu einer Abnahme des Anteils der Gitterpunkte, wo Niederschlag richtig vorhergesagt wurde. POD nimmt also auch ab. Der Grund dafür ist, dass zwar die generellen Strukturen wiedergegeben werden, es aber oft an der räumlichen Exaktheit fehlt. Genau an dieser Stelle greift eine weitere Technik der Verifikation ein, die unter dem Begriff Fuzzy-Verifikation zusammengefasst werden kann, wobei Ereignisse mit geringer räumlicher Versetzung noch als richtige Prognosen gewertet werden. Mehr zu Fuzzy-Verifikationsmethoden findet sich im Abschnitt 4.4

Zur Beantwortung der zweiten Frage, ob der Vergleich der Flächen unter der Kurve legitim ist wenn sich POD und POFD ändern, stößt das Konstrukt der pathologischen ROC-Kurve mit nur einem Punkt an seine Grenzen. Grundsätzlich gilt eine große Fläche als wünschenswert. Denkbar wäre aber der Fall, dass zwei ROC-Kurven dieselbe Fläche bei deutlich unterschiedlichen Werten von POD und POFD aufweisen. Entscheidend ist in diesem Fall der Betrag von POFD, wobei ein geringer POFD-Wert einem hohen vorzuziehen ist, da der hohe Anteil der falschen Alarme die Unschärfe der Prognose widerspiegelt (Fawcett, 2006) [18].

Abbildung 4.4 zeigt eine Erhöhung der Flächen unterhalb der ROC-Kurve für konvektionsauflösende Modelle bei gleichzeitiger Abnahme der POFD-Werte gegenüber von konvektionsparametrisierenden Modellen. Bei der paarweisen Untersuchung innerhalb einer Modellgruppe ergibt sich eine Erhöhung der Fläche, gleichzusetzen mit einer Verbesserung von 3% für COSMO 2 gegenüber COSMO 7, COSMO-DE schneidet bei der Flächenbetrachtung um 3.5% besser ab. Der größte Unterschied ist zwischen den französischen Modellen ALADIN France und AROME mit knapp 8% zu verzeichnen. Bei ALADIN tritt der eben behandelte Fall ein, dass über 70% der Vorhersagepunkte "falsche Alarme" sind, also die räumliche Ausdehnung der Niederschlagsobjekte erheblich zu groß eingeschätzt wurde, was wiederum einer unscharfen Prognose gleichkommt. Durch die räumlich großzügige Abdeckung des Vorhersagegebietes ist es im Gegenzug dazu einfach, diejenigen Gitterpunkte auch zu treffen, die tatsächlich Niederschlag erhalten haben, was sich in einem hohen POD-Wert zeigt. Ohne Frage sind in diesem Fall die Prognosen von AROME vorzuziehen, auch wenn sich kein so markanter Flächenunterschied zeigen würde wie hier.

Generell liegen die Anteile der korrekten Prognosen (*hits*) der konvektionserlaubenden Modelle dicht gedrängt zwischen 55 und 58%, dem ein vergleichsweise geringer Anteil an Überprognosen von knapp 20% gegenübersteht. Die Werte von POFD bei den konvektionsparametrisierenden Modellen bewegen sich im Bereich von 50%, was bedeutet, dass doppelt so viele Gitterpunkte vom Modell mit Niederschlag belegt werden, als Ereignisse zu beobachten sind. Auf der anderen Seite treffen knapp 80% der Prognosen auch tatsächlich ein. Ausnahme stellt dabei eben ALADIN mit einer Trefferquote von 90% dar,

im Gegensatz dazu stehen aber 70% falsche Alarme.

Wohlgermerkt handelt es sich beim oben betrachteten Fall um Niederschlag mit einer Akkumulationszeit von 12 Stunden. Bei Summenniederschlägen kürzerer Zeitintervalle kommt es zur Reduktion der Fläche unterhalb der Kurve und zu einer generellen Abnahme von POD sowie auch von POFD. Der immanente Unterschied in Bezug auf die Lage der Punkte und auf die Auflösung von Konvektion bleibt in der ROC Darstellung aber weiterhin bestehen. Je kürzer die Akkumulationszeit gewählt wird, umso präziser muss das zeitliche Eintreten eines Ereignisses prognostiziert werden, eine längere Summation entspricht gewissermaßen einem zeitlichen Toleranzfenster oder einer Glättung, da ein Ereignis auch etwas früher oder später eintreten kann, ohne das Ergebnis zu beeinflussen.

Auffallend ist aber der generell recht deutliche Unterschied von ALADIN France zu allen anderen verwendeten Modellen. Dieser liegt möglicherweise an der Formulierung des dynamischen Modellkerns zur Lösung des Gleichungssystems. Wie schon in Abschnitt 2.1.1 angedeutet, handelt es sich bei der Französischen Version von ALADIN um die einzige hydrostatisch approximierte Modellapplikation der getroffenen Auswahl, alle anderen Modelle sind hingegen nichthydrostatisch formuliert.

### 4.2.3 Saisonale Vergleiche 18-stündiger Prognosen

Bei der Ausweitung der eben festgestellten Zusammenhänge und der deutlichen Unterscheidung, ob ein numerisches Wettervorhersagemodell in der Lage ist, konvektive Niederschlagsereignisse explizit aufzulösen, oder ob diese als subskalige Prozesse parametrisiert werden müssen, könnte bei der Betrachtung weiterer Monate der konvektiven Jahreszeit auf ähnliche Verhältnisse geschlossen werden. Unter dem Begriff "konvektive Jahreszeit" werden hier vor allem die Sommermonate von Juni bis August mit vorherrschend konvektiver Niederschlagsaktivität zusammengefasst. Dabei wird aber keinesfalls ausgeschlossen, dass Konvektion auch im Herbst oder Winter auftreten kann, wo vorwiegend transiente barokline Systeme wetterbestimmend sind und eher stratiforme Niederschlagsereignisse dominieren. Zu erwarten wären also ähnliche Ergebnisse mit signifikanten Unterschieden zwischen den konvektionsauflösenden und den parametrisierten Modellen für die Sommermonate, sowie eine Angleichung der Verhältnisse im Herbst durch die Änderung der niederschlagsauslösenden Prozesse und dem zumindest teilweisen Fehlen hochreichender, konvektiver Gebilde.

Genau diese Vermutungen können auch bestätigt werden. Im Zeitraum zwischen Juni und August zeigt sich in Abb. 4.5, wie auch schon bei der Betrachtung von Juni (Abb. 4.4), der deutliche Unterschied zwischen der Parametrisierung und der modellseitigen Auflösung hochreichender Konvektion von COSMO 2, COSMO-DE und AROME. Im Herbst hingegen kommt es neben der Angleichung auch zu einer deutlichen Verringerung des An-

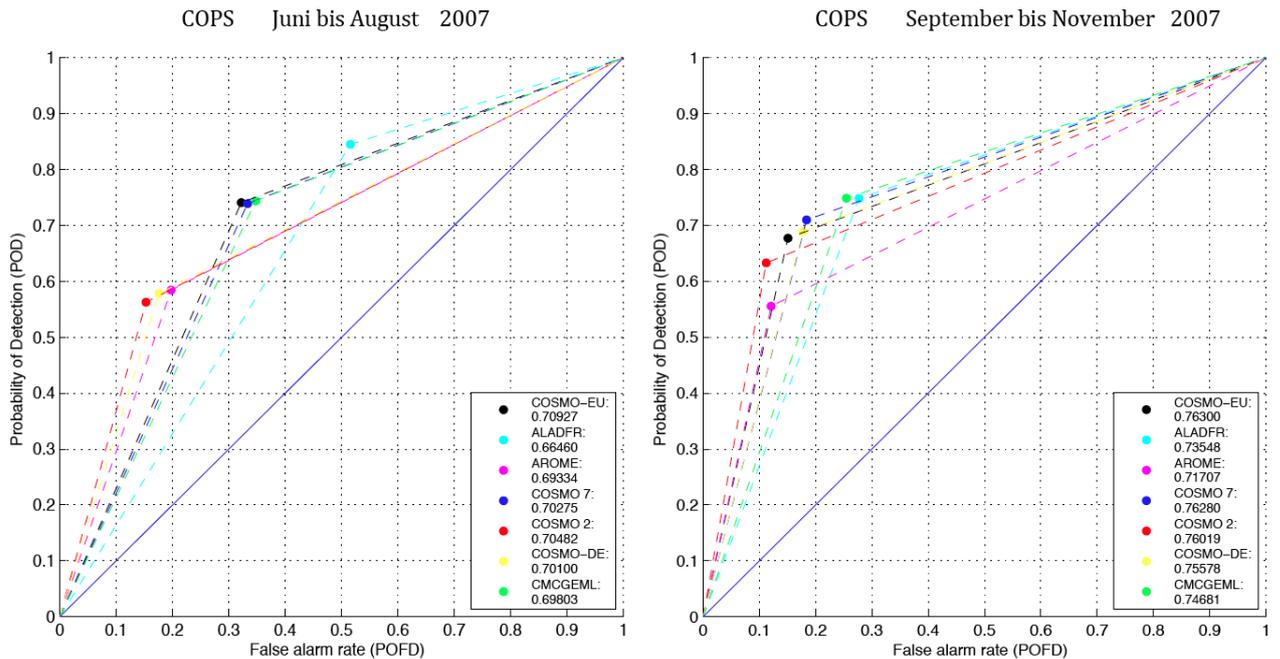


Abbildung 4.5: *Relative operating characteristic (ROC) für 12h akkumulierten Niederschlag aller ausgewählten Modelle mit einer Prognosedauer von 18 Stunden. Modellläufe um 00UTC für die Region COPS, vergl. Abb. 4.4 . Links: Sommer 2007 von Juni bis August. Rechts: Herbst 2007 von September bis November, Mittelungszeitraum für die eingetragenen Punkte jeweils drei Monate. 630 Gitterpunkte; Niederschlagsschwelle:  $>0.1\text{mm}/12\text{h}$*

teils von falschen Alarmen in den Prognosen aller Modelle. Die hochauflösenden Modelle COSMO 2 und COSMO-DE zeigen zudem auch eine Erhöhung der Trefferquote, bei allen anderen kommt es zu einer Reduktion von POD. Durch eine insgesamt Verschiebung der Punkte hin zum Bestwert (linke obere Ecke) erhöhen sich die Flächen unterhalb der Kurve im Schnitt um 5.3%. Die Flächenunterschiede bei der paarweisen Untersuchung innerhalb einer Modellgruppe lassen bei den Verhältnissen im Herbst auf keine Bevorzugung der hochauflösenden Modelle schließen, wie das in der konvektiven Jahreszeit der Fall ist. Die generelle Erhöhung der Flächen unterhalb der ROC-Kurve fällt für größere Modelle durch den markanteren Rückgang der POFD Werte sogar etwas größer aus.

Defizite zeigen sich also vor allem bei der Konvektionsparametrisierung der größeren Modelle durch die sehr “liberale” Verteilung von Niederschlag im Sommer. Durch Änderung der auslösenden Prozesse im Herbst verschwindet dieses Defizit, wodurch es zum Angleich an die Prognosegüte der hochauflösenden Modelle kommt. Wichtig ist hier also festzuhalten, dass hochauflösende Modelle über den gesamten Zeitraum ähnliche Ergebnisse liefern, die Insuffizienzen der größeren Modelle durch die Parametrisierung im Herbst reduziert werden, was zu einer Erhöhung der Prognosegüte führt und es nicht umgekehrt etwa zur Verschlechterung der Qualität der konvektionserlaubenden Modelle

kommt.

### 4.2.4 Saisonale Vergleiche 30-stündiger Prognosen

Die Durchmischung der Atmosphäre, im Speziellen der Troposphäre ab einer Höhe von etwa einem Kilometer, lässt sich vorwiegend durch Konvektion, also durch thermisch turbulente Umlagerungen erklären (Liljequist et al. 1984) [23]. Einer der wesentlichen antreibenden Mechanismen ist dabei die Erwärmung des Erdbodens und der unteren atmosphärischen Luftschichten infolge von Strahlungsabsorption. Ein weiterer Effekt zur Ausbildung von Konvektion wird über die relative Abkühlung höherer Luftschichten im Vergleich zu unteren Luftschichten hervorgerufen. Vor allem im Sommer bewirkt die Einstrahlung neben einem kräftigen Tagesgang der Temperatur der bodennahen Luftschichten, insbesondere über Landflächen, auch ein Konvektionsmaximum in den Nachmittagsstunden.

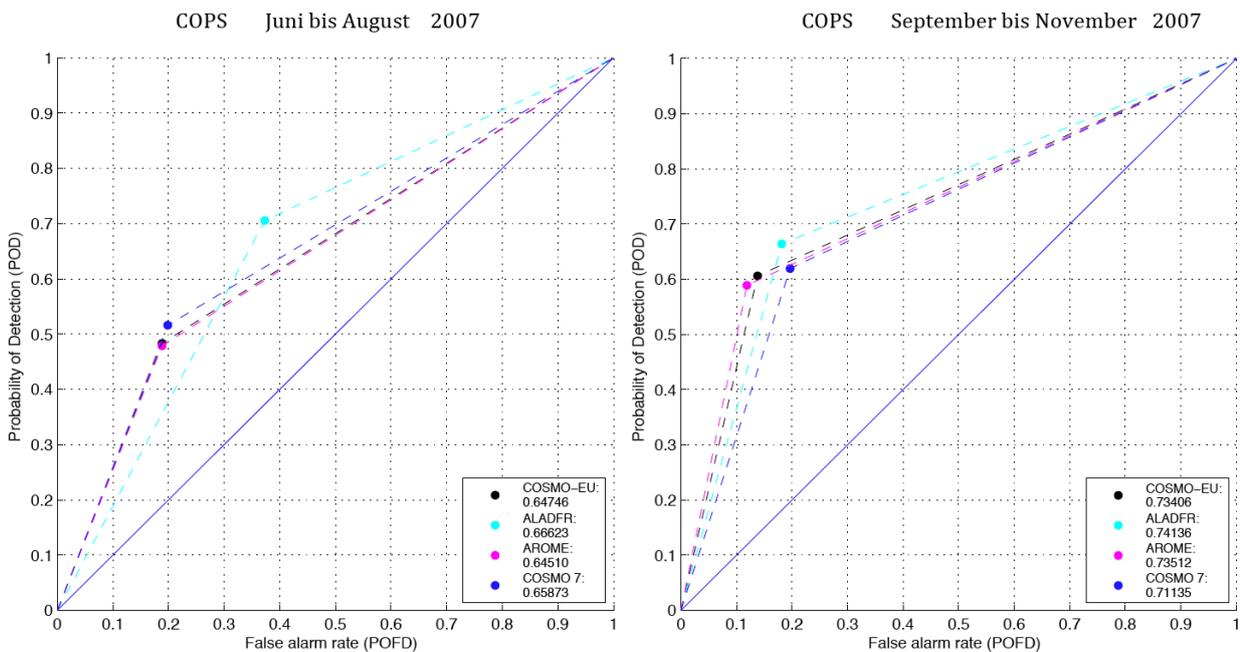


Abbildung 4.6: *Relative operating characteristic (ROC) für 12h akkumulierten Niederschlag und 30 stündigen Prognosen. Modellläufe um 00UTC für die Region COPS mit einer Gültigkeit um 06 UTC des Folgetags. Links: Sommer 2007 von Juni bis August. Rechts: Herbst 2007 von September bis November, Mittelungszeitraum für die eingetragenen Punkte jeweils drei Monate. 630 Gitterpunkte; Niederschlagsschwelle: >0.1mm/12h*

Untersucht man nun aber nicht mehr Niederschlagsereignisse in deren Akkumulationszeit auch das Maximum der konvektiven Aktivität fällt, sondern jene während der Nachtstunden zwischen 18 und 06 UTC, so haben die oben gefundenen Ergebnisse keine

Gültigkeit mehr und können daher nicht verallgemeinert werden. Die deutlich geringere konvektive Aktivität lässt keine Bevorzugung der konvektionserlaubenden Modelle gegenüber der Modelle mit Parametrisierung mehr zu. In den Nachtstunden tritt allerdings ein weiterer Effekt in Bezug auf die thermische Turbulenz auf. Über Meeresoberflächen ist der Tagesgang der Temperatur gegenüber Landflächen zwar eher gedämpft, die schwache nächtliche Abkühlung der höheren Luftschichten bei gleichzeitig nahezu unveränderten Verhältnissen in Bodennähe reicht aber zur Bildung labiler Verhältnisse und in weiterer Folge zur Konvektion, die über maritimen Gebieten während der Nachtstunden sogar ihr Maximum erreicht (Liljequist et al. 1984) [23]. Der geringe flächenmäßige Anteil der Meeresoberfläche im Verhältnis zur betrachteten Landfläche ist für diesen Effekt aber nicht ausreichend und kann daher auch bei der Untersuchung des D-PHASE Ausschnitts nicht explizit festgestellt werden.

Der räumlich sehr begrenzte Ausschnitt von COPS in Abbildung 4.6 zeigt eine Auswertung der über einen Zeitraum von jeweils drei Monaten gemittelten POD- und POFD-Werte in einem ROC- Diagramm des 12-stündig summierten Niederschlags für 06 UTC. Die jeweilige Prognosedauer beträt 30 Stunden. Die bei AROME modellseitig geforderte Auflösung von Konvektion zeigt keinen signifikanten Vorzug gegenüber den parametrisierenden Modellen. Auch aus der Untersuchung anderer Akkumulationsstufen des Niederschlags innerhalb der Abend- und Nachtstunden lassen sich sehr ähnliche Schlüsse ziehen. Markant ist wieder der deutliche Kontrast von ALADIN France über die hydrostatische Näherung. Die sehr liberale räumliche Abschätzung von Niederschlag führt einerseits zwar zu vergleichsweise hohen Trefferquoten, andererseits aber auch zu hohen Werten von POFD.

## 4.3 Stratifikation nach Intensitätswerten des Niederschlags

Die Untersuchungen der letzten Abschnitte in Bezug auf die Niederschlagsschwelle beschränken sich in erster Linie auf die geforderte Messbarkeitsgrenze der Intensität von 0.1 mm, bezogen auf eine bestimmte Zeitspanne der Akkumulation. Insbesondere bei Fragestellungen der qualitativen Verifikation kategorischer Variablen wie Niederschlag, treten aber nicht zuletzt wegen dieser sehr geringen Niederschlagsschwelle einige Schwierigkeiten auf, deren Ursache beispielsweise bei der Interpolation zu finden ist. Durch die Interpolation sehr inhomogener Felder kommt es vor allem bei großen Gradienten zu Vorgängen des Über- bzw. Unterschwingens der Werte. Dieser Effekt tritt bei allen Parametern auf, kann in den meisten Fällen aber nur sehr schwer korrigiert werden. Niederschlag als positiv definierter Parameter erlaubt zwar eine Korrektur, allerdings nur in Bezug auf Unterschwin-

gungsvorgänge durch Nullsetzen aller negativen Werte. Gerade durch die verstärkte Inhomogenisierung der Niederschlagsfelder unter Ausbildung scharfer Gradienten durch die Konvektion, speziell während der Sommermonate, kann dieser Effekt zum Teil sehr markante Auswirkungen auf die Verifikationsmaße selbst sowie auch auf den Modellvergleich mit sich bringen. Die Abschätzung der Fehlermagnitude bei Überschwingungsvorgängen gestaltet sich jedoch sehr schwierig. Fest steht aber in jedem Fall, dass mit zunehmender Niederschlagsrate das Ausmaß der Beeinflussung abnimmt. Bei einem Schwellwert von 0.1 mm ist das größte Fehlerpotential zu erwarten ist.

In den Einstellungsoptionen von MET zur statistischen Auswertung lassen sich sehr einfach mehrere Schwellwerte für einen Parameter anbringen. Die direkte Beschränkung der Vorhersage- und Analysefelder über Schwellwerte kann einerseits zur Untersuchung von Extremereignissen, etwa von Starkniederschlägen mit Raten oberhalb einer gewissen Schranke, oder aber zur Erhebung von Trendkurven bezüglich verschiedener Maßzahlen der Verifikation herangezogen werden.

Abb. 4.7 zeigt den Zusammenhang zwischen verschiedenen Verifikationsmaßen der kategorischen Variable des 12- stündig akkumulierten Niederschlags in Abhängigkeit von Niederschlagsschwellen, gemittelt über die Sommermonate Juni bis August 2007 im Gebiet COPS. Die grundsätzliche Form der Abbildung sowie die Wahl der Niederschlagsschwellen sind auch in Amodei und Stein (2009) [5] zu finden. Die berechneten Werte der Verifikationsmaße wurden nach dem im Abschnitt 4.2.1 eingeführten Schema über die Erstellung einer neuen Kontingenztafel aus der Summation der Tageinträge der ursprünglichen Kontingenztabellen ermittelt. Ein weiterer hier verwendeter Aspekt ist die Angleichung der Terminanzahl über ein Ausschlussverfahren, im Gegensatz zur bisher verwendeten Betrachtung und der Akzeptanz von zumindest geringen Fehlern über die leicht unterschiedliche Anzahl der beteiligten Tage eines Experiments (vergl. Abschnitt 4.2.2). Liefert ein Modell an einem Tag keine Prognose, so werden zur Angleichung auch die Modellläufe aller anderen beteiligten Modelle nicht in den Berechnungen berücksichtigt. Der Vergleich der Maßzahlen und somit der Modelle im weiteren Sinne kann also als weitgehend exakt, sowohl geographisch über die Verwendung derselben Gitterpunkte, als auch zeitlich durch die Angleichung der Terminanzahl, angenommen werden. Eine Einschränkung in Hinblick auf die Exaktheit des Modellvergleichs besteht aber weiterhin über die Schwächen der Interpolation.

Die in der Abbildung verwendeten Verifikationsmaße beinhalten neben den schon ausführlich beschriebenen Parametern von POD und POFD auch Maßzahlen wie den Critical Success Index (CSI) oder die Accuracy (ACC).

$$CSI = \frac{hits}{hits + misses + false\ alarms} \quad (4.10)$$

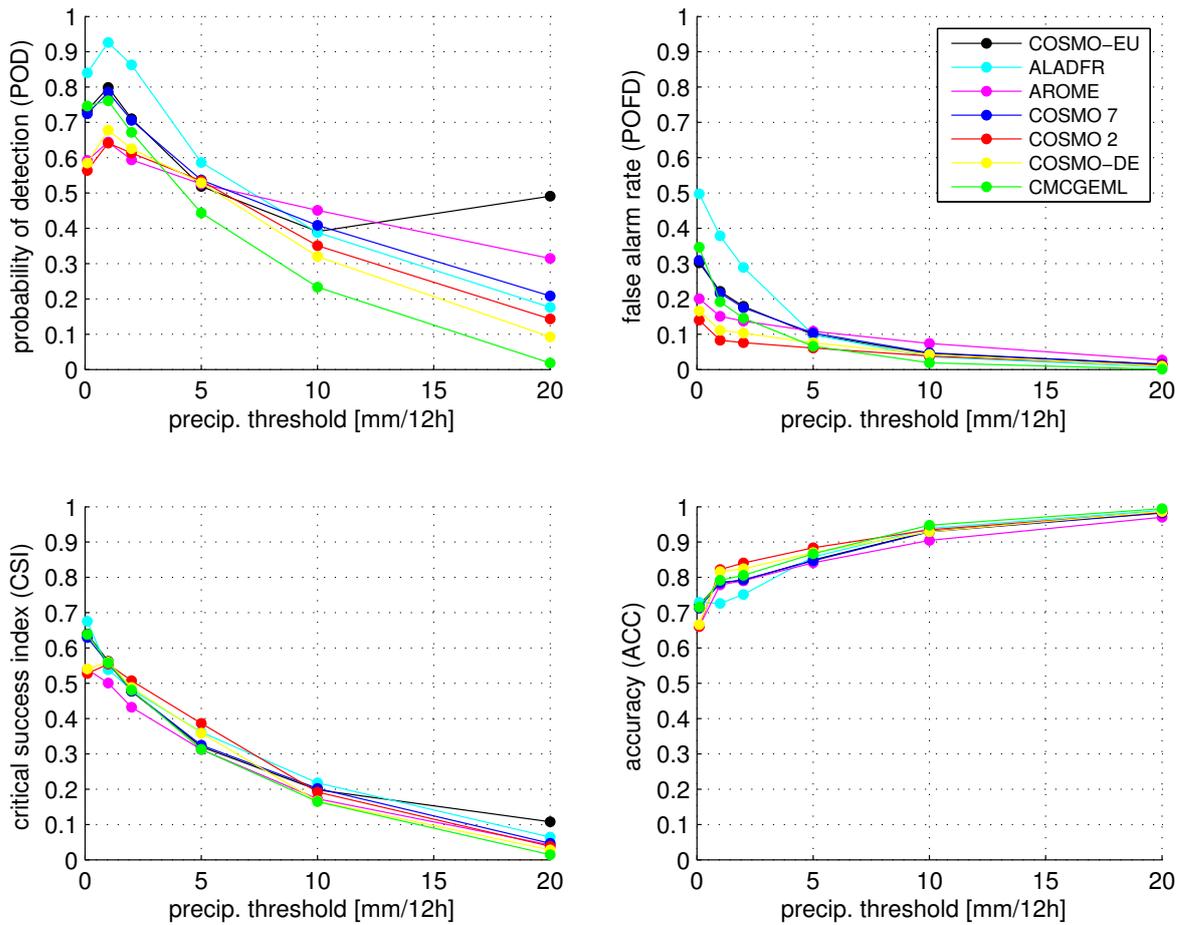


Abbildung 4.7: *POD*, *POFD*, *CSI* und *ACC* als Funktion der Niederschlagsschwellen  $\geq 0.1$ ,  $\geq 1$ ,  $\geq 2$ ,  $\geq 5$ ,  $\geq 15$  und  $\geq 20$  mm/12h., Modellläufe um 00 UTC, 18-stündige Prognosen für RR-12h zwischen **Juni und August 2007**. Gebiet: **COPS** (630 Gitterpunktpaare).

$$ACC = \frac{hits + cr}{total} \quad (4.11)$$

CSI gibt dabei den Anteil der richtigen Prognosen an der Summe der gesamten beobachteten Ereignisse mit der Anzahl der falschen Alarme wieder. Im Nenner wird also im Gegensatz zu POD zusätzlich noch die Anzahl der falschen Alarme der Prognose berücksichtigt. Der Bestwert von  $CSI = 1$  wird dann erreicht, wenn die Anteile der falschen Alarme und die nicht vorhergesagten aber eingetretenen Ereignisse in den Prognosen verschwinden. ACC bezeichnet die sogenannte *Accuracy* als Verhältnis der korrekten Prognosen zur gesamten Gitterpunktsanzahl, somit den Gesamtanteil der richtigen Vorhersagen. Bei der Prognose seltener Ereignisse sollte aber bedacht werden, dass die am häufigsten besetzte Kategorie der korrekten Nichtvorhersagen  $cr$  in hohem Maße an Einfluss gewinnt und den Wert der Accuracy stark beeinflusst. Der Bestwert wird über  $ACC = 1$  festgelegt,

wieder durch Ausbleiben der Kategorien *misses* und *false alarms* in der Kontingenztafel.

### 4.3.1 Auswertung der Schwellwertabhängigkeit der ausgewählten Verifikationsmaße

Die Entwicklungen der Verifikationsmaße in Abhängigkeit von Niederschlagsschranken aus Abb. 4.7 zeigen einige sehr interessante Aspekte. Bei Erhöhung der Schwelle von 0.1 auf 1 mm/12h steigt der Wert von POD, der Anteil von Fehlalarmen nimmt hingegen deutlich ab. Dies führt insgesamt zu einer erheblichen Verbesserung in Hinblick auf die Qualität der Prognosen und zwar für alle Modelle. Im Schnitt zeigt sich eine Zunahme des POD um 6.4% bei einem gleichzeitigen Rückgang der POFD- Werte um 8.7%. Durch Eintragen der Punkte in ein ROC- Diagramm und Auswertung der Fläche unterhalb der Kurve, steigt auch der Flächeninhalt im Mittel um 7.5%.

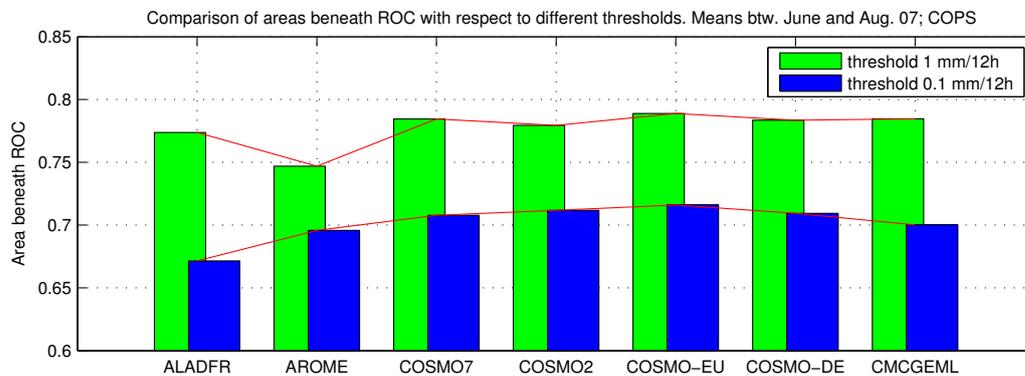


Abbildung 4.8: Vergleich der Flächenwerte unter der ROC- Kurve der Niederschlagsschwellen  $\geq 0.1$  mm/12h (blau) und  $\geq 1$  mm/12h (grün). Mittelungszeitraum: Juni bis August 2007 im Gebiet COPS. Modellläufe: 18-stündige Prognosen, initialisiert um 00 UTC. Rote Linie verbindet die Werte zur Veranschaulichung der Differenzen.

Eine Gegenüberstellung der Flächeninhalte unterhalb der ROC- Kurven, resultierend aus den zwischen Juni und August gemittelten Werten von POD und POFD in Abhängigkeit der Modelle, ist in Abb. 4.8 dargestellt. Die generell deutliche Vergrößerung der Fläche wird vom hydrostatischen Modell ALADIN France mit 10% am deutlichsten wiedergegeben, den kleinsten Wert zeigt hingegen AROME mit lediglich 5%. Die aufgezeigten Defizite bei der qualitativen Verifikation von Niederschlag sehr geringer Raten, z.B. 0.1 mm/12h, sind zwar teilweise über die oszillierende Wirkung infolge großer Gradienten durch die Interpolation zu erklären, spiegeln aber auf der anderen Seite die Unsicherheiten und Schwierigkeiten der numerischen Modellierung in der quantitativen Niederschlagsprognose bezüglich der Festlegung der räumlichen Ausdehnungen und Definition von Umrandungszonen von Niederschlagsgebieten wider.

### 4.3.2 Häufigkeitsverteilung der beobachteten Ereignisse

Starkniederschläge oder Ereignisse mit hohen Intensitäten, die beispielsweise eine Grenze von mehr als 20mm/12h überschreiten, treten in der Regel seltener auf als Niederschläge mit geringerer Rate. Hinzu kommt häufig die starke räumlich Begrenzung solcher Ereignisse vor allem durch die Konvektion in den Sommermonaten. Bei der Auswertung und Interpretation der Schwellwertabhängigkeiten der in Abb. 4.7 betrachteten Verifikationsmaße, liegt daher besonders für größere Schwellen die gleichzeitige Auswertung der Häufigkeitsverteilung nahe.

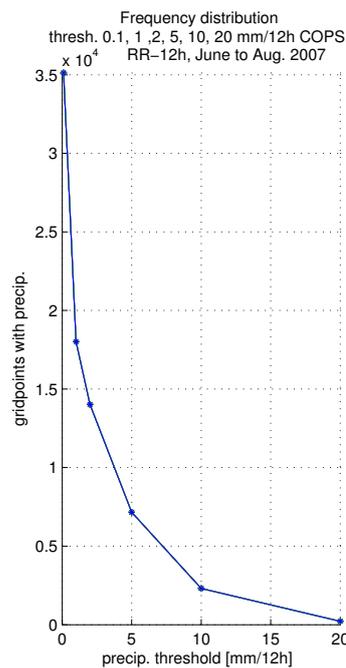


Abbildung 4.9: Häufigkeitsverteilung der Gitterpunkte mit Niederschlag aus Abb.: 4.7

Abbildung 4.9 zeigt die Verteilung der absoluten Häufigkeit als Funktion von Niederschlagsschwellen. Hierbei handelt es sich um die Zahl derjenigen Gitterpunkte, die im Minimum Niederschlag einer bestimmten Intensität erhalten haben. Der Wert der absoluten Häufigkeit lässt sich direkt aus der Kontingenztabelle ablesen, wenn die Anzahl der richtigen Prognosen (*hits*) mit der Zahl der eingetretenen, vom Modell aber nicht prognostizierten Ereignisse (*misses*), addiert wird. Es handelt sich aber dennoch um beobachtete Ereignisse, also um eine modellunabhängige Größe.

Bei der Auswertung der Häufigkeitsverteilungen wird die auch schon weiter oben diskutierte und eingeführte Gleichstellung der Modelle durch die Angleichung der beteiligten Termine verständlich. Notwendig wird dieser Schritt, um einer möglichen Diskriminierung der Modelle, die Niederschlag hoher Intensitäten falsch prognostiziert haben, gegenüber den Modellen, bei denen an solchen Tagen keine Prognose vorliegt, entgegenzuwirken. Die-

ses an sich grundsätzliche Problem zeigt sich aber ganz speziell bei hohen Schwellwerten, da solche Ereignisse wesentlich selten auftreten. Vor allem in räumlich stark begrenzten geographischen Gebieten, wie etwa dem Ausschnitt während COPS, sind selbst bei größeren Mittelungszeiträumen nur äußerst wenige Gitterpunkte beteiligt, deren Niederschlagsintensität eine Schwelle von 20 mm/12 überschreitet. Im vorliegenden Fall beschränkt sich deren Anzahl auf knapp 220 Gitterpunkte für einen Mittelungszeitraum von Juni bis August, und auf gerade mal 53 für die Herbstmonate.

### 4.3.3 Jahreszeitliche Vergleiche der Schwellwertabhängigkeit

Wie auch schon im Abschnitt 4.2.3 diskutiert wurde, verlieren die konvektionserlaubenden Modelle im Herbst bei Abnahme der Konvektion und Übergang zu zunehmend stratiformer Niederschlagsbildung ihre Vorrangstellung und damit ihren Gewinn gegenüber Modellen mit schwächerer Auflösung unter Verwendung von Parametrisierungsschemata. Jedoch galten diese Untersuchungen ausschließlich in Bezug auf eine qualitative Auswertung von Niederschlagsereignissen oberhalb der sehr niedrigen Schwelle von 0.1 mm/12h.

Durch die Abhängigkeit der Verifikationsmaße von der Niederschlagsschwelle zeigt sich im Sommer hingegen eine recht deutliche Verbesserung der Maßzahlenwerte bei geringer Erhöhung des Schwellwerts auf 1 mm/12h, bei weiterer Erhöhung der Schwellen sinkt neben der Häufigkeit der beobachteten Ereignisse auch das Vermögen der Modelle, diese in qualitativer Weise bezüglich Lage und Zeitpunkt des Eintretens auch richtig zu prognostizieren. Die Untersuchung der Verhältnisse im Herbst lässt sehr ähnliche Aussagen zu und zeigt ebenso die recht deutliche Verbesserung der Maße bei geringfügiger Erhöhung der Niederschlagsschwelle. Erstaunlich ist aber die Entwicklung für höhere Schwellenwerte. Dem generell wesentlich rascheren Abfallen der POFD- Werte im Herbst, steht in Abbildung 4.10 ein kontinuierliches Anheben der sehr gebündelten Kurvenverläufe von POD, besonders aber bis zu einem Wert von 5 mm/12h, gegenüber. Sehr ähnliche Verhältnisse zeigen sich auch bei CSI, dessen Maximalwerte in Richtung höherer Schwellenwerte verschoben sind. Die durchschnittliche Erhöhung von POD um 21% bei einer Schwelle von 5 mm/12h bei gleichzeitigem Rückgang des Anteils von Fehlalarmen um 6.9% führt insgesamt zu einer Flächenvergrößerung unterhalb der ROC- Kurve von knapp 14%.

Die Abb. 4.11 zeigt die Änderung der Flächenwerte unterhalb der entsprechenden ROC- Kurven bei einem Schwellenwert von 5 mm/12h in Abhängigkeit der Modelle. COSMO7 mit knapp 16% und ALADIN- France mit 12% stellen dabei die stärksten und schwächsten Erhöhungen der Flächen dar.

Neben der Änderung der auslösenden Prozesse nimmt auch die Häufigkeit der Niederschlagsereignisse im Herbst deutlich ab. Dies geht aus dem hier nicht näher skizzierten Verlauf der absoluten Häufigkeit des beobachteten Niederschlags in Abhängigkeit von

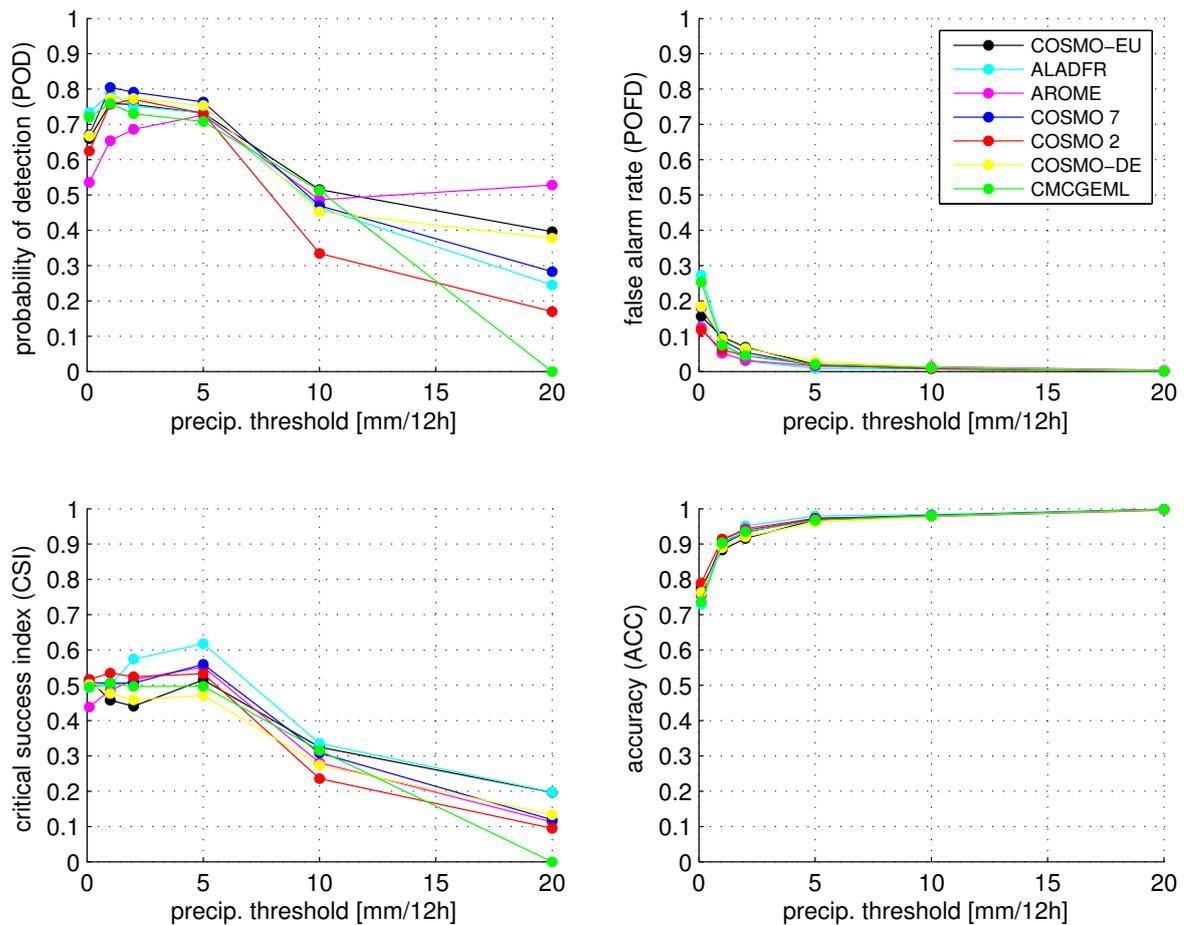


Abbildung 4.10: *POD*, *POFD*, *CSI* und *ACC* als Funktion der Niederschlagsschwellen  $\geq 0.1$ ,  $\geq 1$ ,  $\geq 2$ ,  $\geq 5$ ,  $\geq 15$  und  $\geq 20$  mm/12h. Modellläufe um 00 UTC, 18-stündige Prognosen für RR-12h zwischen **September und November 2007**. Gebiet: **COPS** (630 Gitterpunktpaare).

Schwellenwerten (vgl. Abb. 4.9) hervor, wobei sich die Zahl insgesamt etwa halbiert, bei höheren Schwellenwerten sogar auf ein Viertel abfällt. Stratiforme Niederschlagsfelder sind im Vergleich zu den sehr lokal begrenzten Strukturen konvektiver Zellen eher homogener mit geringeren Amplituden und schwächeren Gradienten. Die glättende Wirkung großer Akkumulationszeiträume führt zwar auch zu einem homogenisierenden Effekt, die generellen Strukturen der Konvektion bleiben aber dennoch erhalten. Scharfe Gradienten der Isohyeten sind typisch für kräftige Konvektion, wobei die sehr enge Drängung der Linien gleicher Niederschlagsintensitäten bei gleichzeitig starker lokaler Begrenzung der Ereignisse für die Verifikation bezüglich verschiedener Niederschlagsschwellen nur einen sehr kleinen Spielraum zulässt. In flacheren, homogeneren Feldern hingegen sind die Abstände zwischen den Isohyeten deutlich größer und daher Niederschlag oberhalb einer bestimmten Schranke auch leichter zu verifizieren. Als Konsequenz scheinen im Herbst

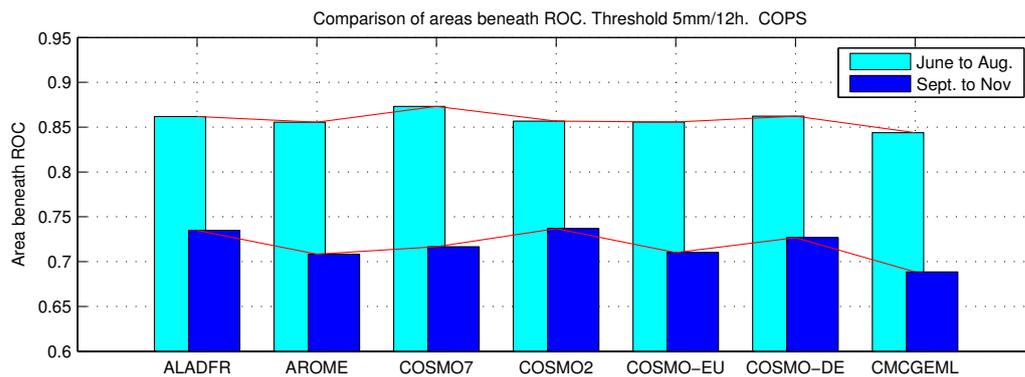


Abbildung 4.11: Vergleich der Flächenwerte unter der ROC- Kurve der Niederschlagsschwellen  $\geq 5 \text{ mm}/12\text{h}$  für die Mittelungszeiträume Juni bis August 2007 (cyan) und September bis November (blau) im Gebiet COPS. Modellläufe: 18-stündige Prognosen, initialisiert um 00 UTC. Rote Linie verbindet die Werte zur Veranschaulichung der Differenzen.

die Unsicherheiten bei räumlichen Niederschlagsprognosen etwas geringer auszufallen.

Sehr ähnliche Folgerungen lassen sich auch aus äquivalenten Analysen zur Untersuchung der Verhältnisse für das Gebiet von D-PHASE schließen. Auffällig ist dabei ein steiler Anstieg von POD bei der Erhöhung des Schwellwerts von 0.1 auf 1 mm/12h für alle Modelle um durchschnittlich knapp 20% innerhalb des Herbstzeitraums. Der größere Einzugsbereich und die daraus folgende Erhöhung der Werte in der absoluten Häufigkeitsverteilung der Gitterpunkte mit Niederschlag, nimmt auch Einfluss auf die Vertrauenswürdigkeit der Verifikationsmaße, insbesondere bei höheren Schwellwerten.

### 4.4 Fuzzy- Verifikation

Vorhersagefelder hochauflösender numerischer Modellapplikationen zeigen zwar mitunter wesentlich realistischere Strukturen eines meteorologischen Phänomens, schneiden bei der Verifikation mit Standardmethoden aber häufig schlechter ab als Modellprognosen mit größeren Auflösungen (z.B. Ebert, 2008 [16]). Kleinskalige Ereignisse sind demnach bezüglich der räumlich und zeitlichen Exaktheit des Auftretens wesentlich schwieriger zu prognostizieren. Hinzu kommt, dass sich die Dispositionierung eines Vorhersageobjekts mit geringer Ausdehnung gleich mehrfach in den Auswertungsparametern niederschlägt. Einerseits wird das Ereignis an einer Stelle prognostiziert, wo es aber nicht eintritt (false alarm), andererseits dort beobachtet, wo es nicht prognostiziert worden ist (misses). Dies führt also zu einer doppelten Bestrafung (double penalty) des Prognosesystems bei der Verifikation mit traditionellen Methoden (z.B. Amodei und Stein, 2009 [5] oder Ebert, 2008 [16]).

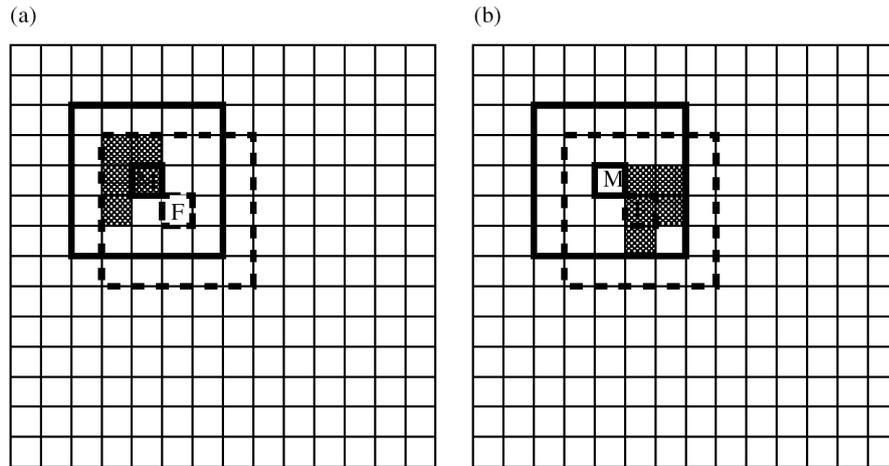


Abbildung 4.12: Darstellung eines beobachteten Niederschlagsereignisses (a) und eines prognostizierten (b). Gepunktete Gitterpunkte kennzeichnen ein Ereignis. Die großen Quadrate legen ein Fenster um die Punkte F und M fest. (Quelle: Amodei und Stein, 2009 [5])

In Abb. 4.12 werden genau diese Zusammenhänge expliziert. Der prognostizierte Wert am Punkt F kann einerseits nicht beobachtet werden, andererseits wird aber ein Punkt mit positiver Beobachtung nicht vorhergesagt (M). Die grundlegende Überlegung von Fuzzy-Verifikationstechniken ist eine etwas “unscharfe” Betrachtungsweise beim Abgleich der Gitterpunkte, um teilweise räumlich nur knapp verfehlte Prognosen nach verschiedenen Kriterien zumindest teilweise als Treffer aufzufassen. Der Hauptunterschied zwischen den verschiedenen Techniken besteht in der Entscheidung, wann eine Prognose noch nahe genug und daher noch brauchbar ist.

Innerhalb der vorliegenden Untersuchung wurden zwei Fenstergrößen mit Seitenlängen von 3x3 und 5x5 Gitterpunkten festgelegt. Als mindeste Voraussetzung gilt allgemein der Fall, dass ein Ereignis zumindest an einem Gitterpunkt innerhalb des festgelegten Fensters eintritt. Das hier verwendete Kriterium zur Auffassung einer Prognose als Erfolg, wird an eine 50% Bedingung geknüpft. Diese gibt den relativen Anteil an der Gesamtgröße des Fensters wieder, der mindestens mit Ereignissen abgedeckt sein muss, um die Prognose anzunehmen. Sind etwa von 9 Gitterpunkten des Fensters (3x3) nur 4 Gitterpunkte von einem Niederschlagsereignis größer dem eingestellten Schwellwert, z.B. 0.1mm/xh, betroffen, so ist der Grad der Bedeckung (coverage) mit  $4/9$  anzugeben und erreicht das vorgegebene Kriterium von 50% nicht.

#### 4.4.1 Anwendungen von Fuzzy- Methoden

Über die Festlegung einer Fenstergröße um einen Punkt und in weiterer Folge die Verschiebung dieser Umgebungsfelder an jeden weiteren Gitterpunkt eines geographischen

Ausschnitts, wird sowohl der Prognose, als auch den analysierten Feldern eine gewisse Unschärfe erlaubt, sodass die Erfordernisse eines exakten Abgleichs von Gitterpunkten auf beiden Seiten relativiert werden. Durch diesen Effekt einer etwas verschwommenen Wahrnehmung ist es einfacher, ein prognostiziertes Ereignis auch in den Feldern der Analyse aufzufinden. Zu erwarten wäre also eine Optimierung der Verifikation über eine Modifikation der Maßzahlen in Richtung des jeweiligen Bestwerts. Bei der Auswertung und Adaptierung der oben eingeführten Verifikationsmethoden an die Anforderungen der Fuzzy-Verifikation kann diese Annahme auch bestätigt werden.

Die Dimensionierung eines Fensters steht in unmittelbarem Zusammenhang mit der Leistung des Fuzzy- Systems über die Verbesserung eines Verifikationsmaßes mit zunehmender Fenstergröße. Die getesteten Fenstergrößen von 3x3 und 5x5 Gitterpunkten zeigen demnach eine Erhöhung von POD sowie eine Abnahme von POFD, die beim größeren Fenster deutlicher ausfallen. Dieser Zusammenhang ist allerdings nicht allgemein gültig und in erster Linie an die Wahl der Zusatzbedingung geknüpft. Die direkte Beziehung besteht nur im Fall des einfachsten Kriteriums, wobei ein Ereignis irgendwo innerhalb des Fensters eintreten kann. Das 50% Kriterium erlaubt diesen Zusammenhang nicht mehr, da bei zunehmender Fenstergröße dieses oft nicht mehr zu erfüllen und eine Verifikation dadurch auch nicht mehr möglich ist.

Eine Ausweitung der im Abschnitt 4.2.3 durchgeführten Untersuchungen und dem saisonalen Vergleich der ROC- Diagramme mit einer eingestellten Niederschlagsschwelle von 0.1mm/12h auf die vorgestellte Methode der Fuzzy- Verifikation mit einer Fenstergröße von 5x5 Gitterpunkten zeigt etwa im Gebiet von COPS eine durchschnittliche Zunahme der Flächenwerte unterhalb der ROC- Kurve um 3.7% zwischen Juni und August und eine Vergrößerung um 4.1% im Vergleichszeitraum der Herbstmonate (hier nicht dargestellt). Es erfolgt also eine Verschiebung der Punkte in Richtung der linken oberen Ecke, wobei im Vergleich die Flächenzunahme der konvektionsauflösenden Modelle knapp 1% geringer ausfällt als für parametrisierende. Verantwortlich für die größeren Flächenwerte zeigt sich im Sommer eher die deutliche Abnahme von POFD, im Herbst hingegen eine Verschiebung der Punkte in Richtung größerer POD- Werte.

Die Ausweitung auf eine Untersuchung der Verifikationsmaße in Abhängigkeit von Niederschlagsschwellwerten wird in Abb. 4.13 dargestellt. Für POD ergibt sich im Sommer nur eine leichte Verbesserung um etwa 2 % über den gesamten Bereich der eingestellten Schwellen. Die Balken der Abbildung zeigen hier die durchschnittlichen über alle Modelle gemittelten Änderungen eines Verifikationsmaßes. Eine zusätzliche Untersuchung der einzelnen Modelle deutet zudem auf eine etwas geringere Steigerung für konvektionsauflösende Modelle bei sehr niedrigen Schwellenwerten hin (nicht dargestellt). Die Werte von POFD fallen hingegen im Sommer über den gesamten Schwellenbereich, wobei der

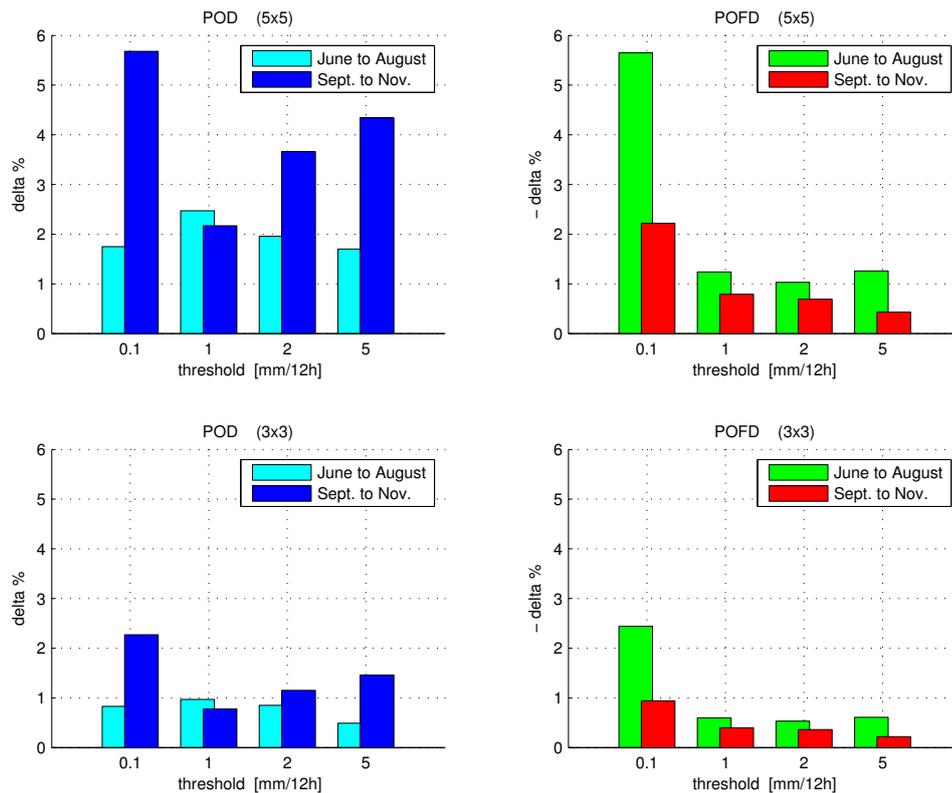


Abbildung 4.13: *Mittlere Änderung von POD und POFD für Sommer und Herbst 2007 als Funktion von Schwellwerten im Gebiet COPS. Die Balken entsprechen jeweils den über alle Modelle gemittelten Werten von POD oder POFD eines Schwellwerts. delta% ist als Verbesserung zur konventionellen Methode zu interpretieren, für POFD bedeutet das eine Abnahme. Fenstergrößen: 5x5 bzw. 3x3 Gitterpunkte.*

größte Unterschied im Bereich der kleinsten Schwelle mit knapp über 5% zu finden ist. Die Auswirkung des Konzeptes der Fuzzy- Verifikation auf die Verhältnisse im Herbst zeigt eine sehr deutliche Anhebung von POD bis zu einem Schwellwert von 5 mm/12h, die Abnahme der ohnehin schon geringen Anteile von Fehlalarmen an den Gitterpunkten ohne Niederschlag beträgt speziell für kleine Schwellen etwa 2% und nimmt für größere weiter ab. Sehr ähnliche, allerdings deutlich geringere Auswirkungen lassen sich durch die Einstellung der kleineren Fenstergröße, d.h. 3x3 Gitterpunkte, ermitteln. Auch bei der Ausweitung auf das Gebiet von D-PHASE liegen die Verhältnisse sehr ähnlich und zeigen insgesamt eine Optimierung der Verifikationsmaße.

#### 4.4.2 Fuzzy in räumlich limitierten Bereichen

Die Anwendung von Fuzzy- Methoden über die Verschiebung des Fensters kann ungeachtet der Vorteile jedoch auch einige Probleme mit sich bringen, insbesondere bei der Verifikation in räumlich sehr begrenzten Gebieten. Der Grund liegt an der Definition des Umge-

bungsfeldes eines Gitterpunktes in Randbereichen eines Ausschnitts. Um die Fenstergröße konstant zu halten, muss einerseits auf einige Gitterpunkte entlang der Begrenzungslinien verzichtet werden, sodass sich insgesamt das Verifikationsgebiet verringert. Andererseits können aber auch die Umgebungsfelder um die Randgitterpunkte gelegt werden, wodurch sich das Gebiet vergrößert. Das Ausmaß der Größenänderung hängt in erster Linie von den Abmessungen des ursprünglichen Gebietes ab. Insbesondere bei kleinen Ausschnitten und dem Einsatz großer Fenster sind die größten Änderungen zu erwarten. Weiters ist zu beachten, ob außerhalb des Verifikationsgebietes überhaupt noch Daten vorhanden sind, um dieses auszudehnen. Das Lösungskonzept von MET unterscheidet dazu zwei Fälle. Sind weitere Daten auch außerhalb des betrachteten Gebietes vorhanden, wird das Gebiet vergrößert, ansonsten verkleinert. Als Folge werden möglicherweise Fehler über die Betrachtung eines größeren Gebietes induziert, sodass die Fuzzy- Verifikation in Grenzfällen teils sogar schlechter abschneidet als traditionelle Methoden.

# 5 Quantitativer Verifikationszugang

Bisher gesteckte Auswertungsziele zur Evaluierung der Prognosequalität hochauflösender numerischer Wettervorhersagemodelle galten in erster Linie der räumlich und in weiterer Folge auch zeitlich richtigen Erfassung meteorologischer Ereignisse. Ein Ereignis kann etwa über kategorische Variablen festgelegt werden. Die Entscheidung des Eintretens oder nicht- Eintretens an einem Gitterpunkt erfolgt dabei in binärer Form mit Niederschlag als wohl wichtigsten Vertreter dieser Kategorie. Die Definition von Ereignissen ist ferner auch über die Einführung von Schwellwerten zur Begrenzung kontinuierlicher Variablen zu erreichen, wird hier aber nicht näher expliziert.

Die in diesem Kapitel eingeführten und beschriebenen quantitativen Auswertungen konzentrieren sich hingegen weniger auf die korrekte räumlich/zeitliche Erfassung, als auf die richtige Einschätzung der Intensität eines Ereignisses oder auf die adäquate Erfassung der Amplituden im Fall kontinuierlicher Variablenfelder.

## 5.1 BIAS (Mittlerer Fehler)

Die einfache Begriffsbildung und die zentrale Rolle bei der Definition zahlreicher anderer Verifikationsmaße, vor allem für kontinuierliche Variablen, machen den BIAS zu einer sehr anschaulichen Maßzahl in der Verifikation. Die verschiedenen Bezeichnungen in der Literatur wie mean error (mittlerer Fehler) oder Vorhersagefehler beziehen sich meist auf den BIAS als gemittelte Summe der Differenz zwischen Modell und Beobachtung. Oft findet man die Bezeichnung auch in Bezug auf andere Maße wie den BIAS score (frequency bias) für kategorische Variablen oder den multiplikativen BIAS. Diese sollten aber auf keinen Fall verwechselt werden. Die Festlegung der Richtung des Differenzterms ist nicht unbedingt zwingend vorgeschrieben, in der Literatur finden sich immer wieder auch gegenteilige Definitionen.

MET definiert die Richtung des BIAS, wie auch weiter oben erwähnt, als Differenz zwischen Modell und Beobachtung (F-O)

$$BIAS = \frac{1}{N} \sum_{i=1}^N (F_i - O_i), \quad (5.1)$$

wobei der Gitterpunktsindex  $i$  von 1 bis  $N$  läuft. Positive Werte des mittleren Fehlers lassen demnach auf eine Überschätzung des Parameters durch das Modell schließen, negative auf eine Unterschätzung. Durch die Mittelung kann es allerdings zu Kompensationseffekten der Fehler auf beiden Seiten der Nulllinie kommen, sodass ein scheinbar guter BIAS bei eigentlich schlechter Vorhersagequalität angezeigt werden kann. Als weitere Folge können auch keine Aussagen über die größten Amplituden der Fehler einzelner Gitterpunkte gemacht werden. Eine Methode, die diese kompensierenden Wirkungen beleuchten soll, um die Aussagekraft eines Mittelwerts zu untermauern, wird in Abschnitt 5.1.3 vorgestellt.

### 5.1.1 Direkter Modellvergleich anhand der täglichen Abweichungen

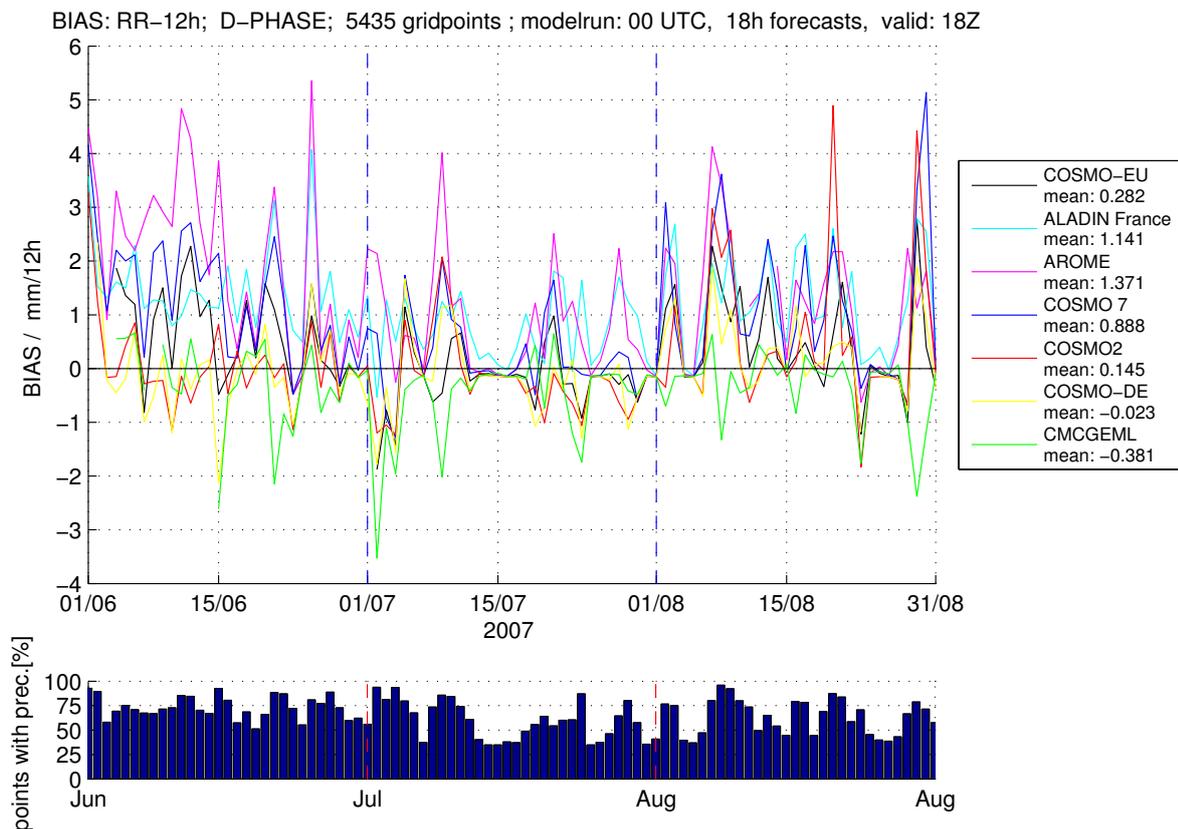


Abbildung 5.1: *BIAS der Tageswerte 18-stündiger Prognosen, Modellläufe jeweils um 00 UTC für den gemeinsamen Ausschnitt der verwendeten Modelle (D-PHASE). Die Mittelwerte des BIAS unterhalb des Modellnamens in der Legende beziehen sich auf den gesamten dargestellten Zeitraum von Juni bis August 2007. Zusätzlich ist der relative Anteil der Gitterpunkte des Gebietes geplottet, die Niederschlag in den 12 Stunden vor dem jeweiligen Termin erhalten haben. Schwelle: 0.1mm/12h.*

Möglichkeiten der Darstellung und Interpretation sind viele denkbar. Der erste und einfachste Fall ist die Darstellung der räumlich über alle Gitterpunkte gemittelten Tageswerte auf einer Zeitskala. Ein Vorteil dabei ist die direkte Erfassung der Arbeitsweise eines Modells zu einem bestimmten Termin und die Möglichkeit des Vergleichs zwischen den Modellen, wenn, wie in Abb. 5.1, mehrere in einem Plot dargestellt werden. Vor allem bei der Betrachtung längerer Zeiträume und bei Zunahme der Modellanzahl steigt auch die Komplexität der Abbildung, wobei es immer schwieriger wird, gezielte Informationen aus der Darstellung zu gewinnen. Die Abbildung zeigt Tageswerte des BIAS für alle Modelle der Auswahl über einen Zeitraum von drei Monaten zwischen Juni und August 2007 im Gebiet D-PHASE. Im unteren Bildabschnitt sind die relativen Anteil der Gitterpunkte mit Niederschlag größer einer Schwelle von  $0.1 \text{ mm}/12\text{h}$  geplottet. Ein Vergleich der Kurvenverläufe liefert eine deutliche Überschätzung der Niederschlagsintensität durch die französischen Modelle ALADIN und AROME über den gesamten Zeitraum, mit teils erheblichen Fehlermagnituden der höher aufgelösten Modellapplikation AROME. Das kanadische Modell CMCGEML zeigt hingegen den untersten Kurvenverlauf und insgesamt die markanteste Unterschätzung der Intensität für die Akkumulationszeit von 12 Stunden. Die hier nicht dargestellten kumulierten Niederschlagsereignisse kürzerer Zeitintervalle lassen auf sehr ähnliche Aussagen schließen und zeigen analoge zeitliche Entwicklungen des Bias. Die Amplituden der Fehler sowie die Anteile der relativen Häufigkeit der Ereignisse nehmen deutlich ab.

### 5.1.2 Auswertung und Vergleich monatsgemittelter Werte

Die eben gezeigte direkte Darstellung der Tageswerte des Bias führt vor allem bei der Ausweitung auf größere Betrachtungszeiträume zu sehr komplexen Abbildungen. Ein weiterer Ansatz ist hingegen die Auswertung von temporalen Mittelwerten eines bestimmten Zeitraumes, z.B. einem Monat. Die zeitliche Mittelung führt zu einer starken Glättung gegenüber den Tageswerten des BIAS, wobei die anfangs besprochene mögliche räumliche Kompensation der Fehler jetzt auch auf die zeitliche Ebene Auswirkungen hat. Die wesentlich glatteren Strukturen erlauben aber eine einfachere Interpretation saisonaler Tendenzen und einen zugänglicheren Vergleich der Modelle untereinander.

Abbildung 5.2 zeigt zeitlich gemittelte BIAS- Werte über die gesamte Dauer des Experiments von Juni bis November 2007 im Gebiet D-PHASE. Generell zeigt sich eine recht deutliche Korrelation der Kurvenverläufe zwischen den Modellen innerhalb der Modellketten. Höher aufgelöste Modellapplikationen mit der geforderten Fähigkeit der direkten Auflösung von Konvektion weisen zudem teils geringere Streuungen der monatsgemittelten Werte auf, vor allem während der Sommermonate. Ein Vergleich der Mittelwerte über den gesamten Zeitraum zeigt einheitlich positive Werte des Bias, also eine modellseiti-

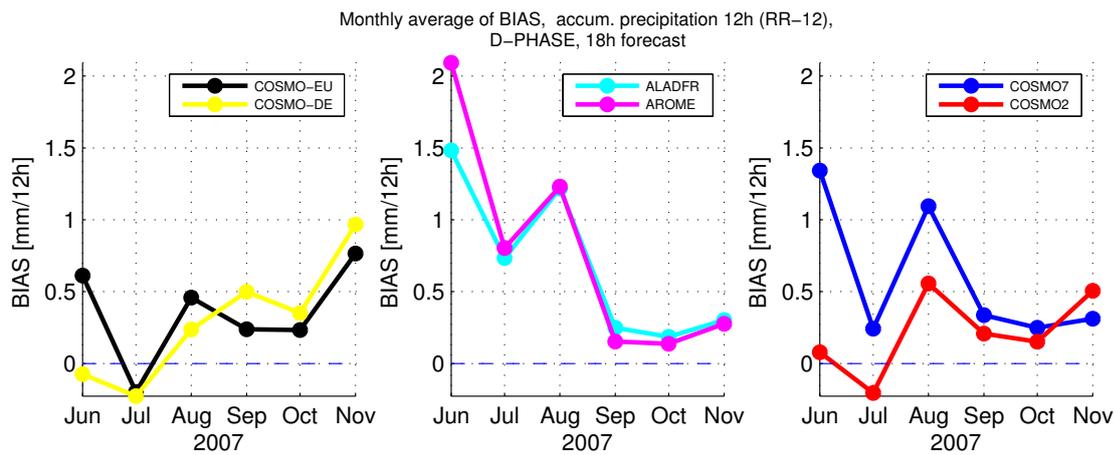


Abbildung 5.2: Monatsmittelwerte des BIAS: 12-stündig akkumulierter Niederschlag für den gemeinsamen Ausschnitt der verwendeten Modelle (D-PHASE). Die Modelle sind nach nesting-Stufen und Zugehörigkeit zu einer Modellfamilie paarweise geplotet. Niederschlagsschwelle: 0.1 mm/12h

ge Überschätzung der Niederschlagsintensität über Zentraleuropa. Die oben erwähnten markantesten Überschätzungen durch die französischen Modelle ALADIN- France und AROME im Zeitraum zwischen Juni und August nehmen im Herbst hingegen deutlich ab. Ähnliche Resultate liefern auch die Modelle der schweizer Modellkette über die Abnahme der Fehleramplituden beim Übergang zu vermehrt stratiformer Niederschlagstätigkeit. Reduziert man die Akkumulationszeit, so werden auch die Absolutwerte des Bias geringer. Auffallend ist zudem eine zunehmende Korrelierung der Abweichungskurven innerhalb der Modellketten bei Abnahme des Zeitintervalls der Akkumulation, mit Ausnahme von ALADIN und AROME. Das hochauflösende Modell AROME zeigt vor allem zu Beginn der Periode durchgehend eine starke systematische Überschätzung der Intensität, zurückzuführen auf anfängliche Schwierigkeiten bei der quantitativen Erfassung von Niederschlag in der Testphase des Forschungsmodells.

Die Favorisierung konvektionsauflösender Modelle hinsichtlich der besseren räumlichen und zeitlichen Erfassung von Niederschlagsereignissen während primär konvektiver Aktivität in den Sommermonaten, lässt sich, mit Ausnahme von AROME, durch die präzisere Abbildung der Niederschlagsintensität somit auch auf eine Bevorzugung dieser Modelle bei der quantitativen Erfassung der Ereignisse ausweiten.

### 5.1.3 Verteilung von Fehlerhäufigkeiten

Mittelwerte haben die Eigenschaft, den durchschnittlichen Wert einer Menge anzugeben, ohne dabei aber selbst getroffen werden zu müssen. Verdeutlicht wird dies bei der Bildung des Mittelwertes einer Menge aus natürlichen Zahlen. Dabei kann der Mittelwert durchaus

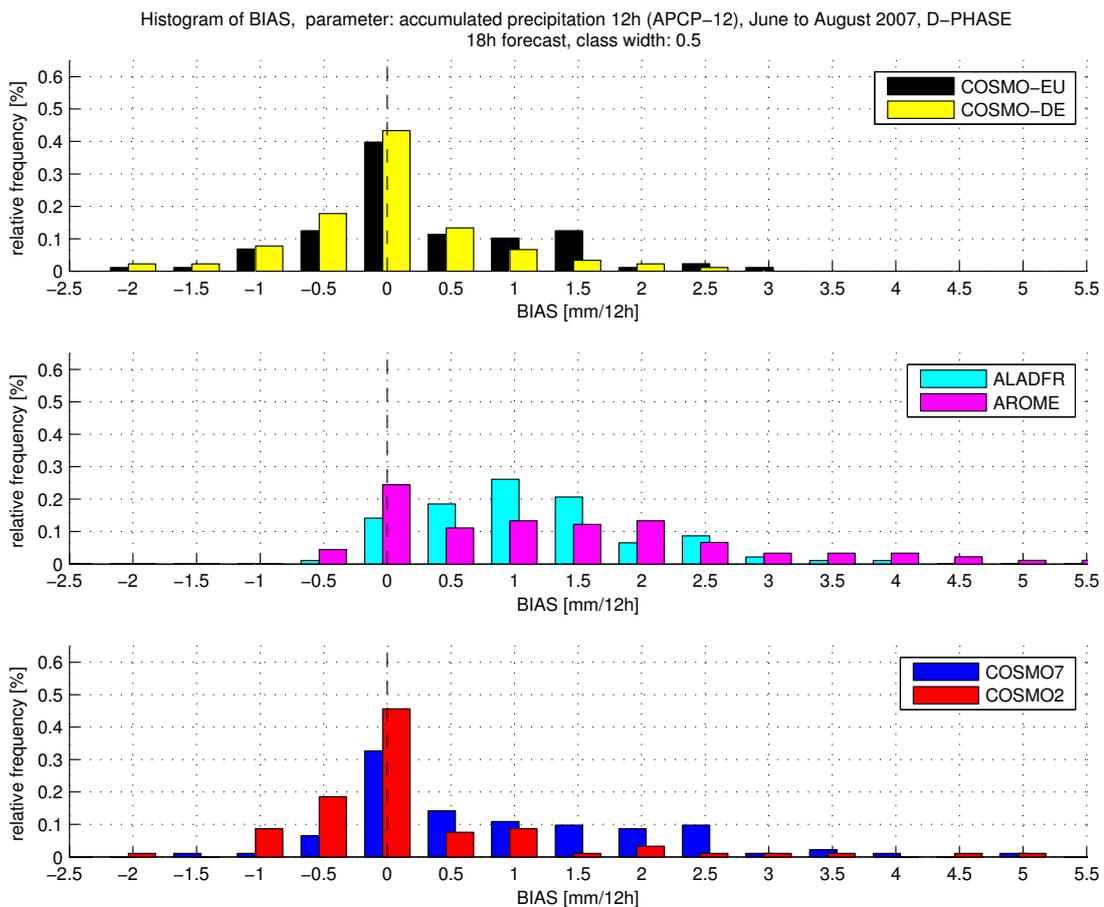


Abbildung 5.3: Verteilung von Fehlerhäufigkeiten für 12-stündig akkumulierten Niederschlag im Gebiet D-PHASE zwischen Juni und August 2007. Klassenbreite 0.5 mm/12h (z.B. Klasse 1 entspricht einer Abweichung zwischen 0.75 und 1.25 mm/12h).

aus der Menge der reellen Zahlen stammen, auch wenn dieser Wert in der Grundgesamtheit nicht existiert. z.B. die durchschnittliche Größe eines Haushalts in Österreich liegt bei 2.29 Personen (Statistik Austria, Stand 2010). Bei Einbeziehung sowohl positiver als auch negativer Abweichungen einer Meßgröße deutet ein Mittelwert von Null im Fall eines Fehlermaßes nicht unbedingt auf die perfekte Erfassung eines Problems hin. Wird in einem meteorologischen Modell ein Parameter innerhalb eines Zeitraums oder eines räumlichen Ausschnitts im selben Maße über- wie unterschätzt, scheint es, als ob der Parameter perfekt abgeschätzt wurde. Eine entscheidende Rolle spielt aber auch die Magnitude der Abweichungen vom Mittelwert, wobei aber wiederum kleine Abweichungen in beide Richtungen zum selben Ergebnis führen können wie große. Diese Zusammenhänge geben Anlass für die Einführung eines weiteren Ansatzes zur Untersuchung der Aussagekraft des Mittelwerts eines Parameters, der Häufigkeitsverteilung der Fehler.

Durch Festlegung einer Klassenbreite werden die Merkmale eines Parameters gruppiert,

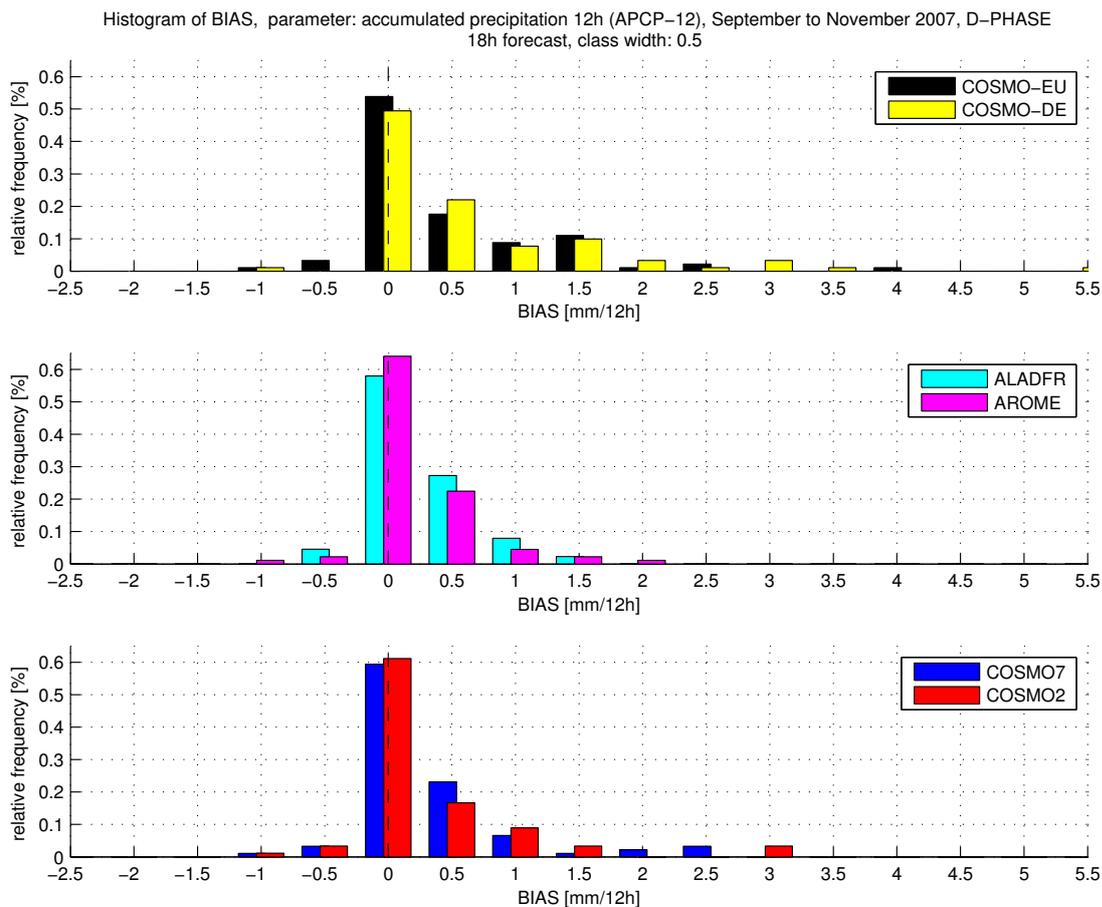


Abbildung 5.4: Verteilung von Fehlerhäufigkeiten für 12-stündig akkumulierten Niederschlag im Gebiet D-PHASE zwischen *September und November 2007*. Klassenbreite *0.5 mm/12h*

und die relative Häufigkeit der eingetretenen Ereignisse innerhalb der Klassen ausgewertet und als Balkendiagramm dargestellt. Die Summe der Häufigkeiten der einzelnen Balken ergibt dabei genau eins. Zur Beschreibung der Fehlerverteilung werden die zwei Eigenschaften der Schärfe (engl. sharpness) und andererseits der Schiefe einer Verteilung (engl. skewness) verwendet. Eine geringe Verteilungsschärfe ist gleichzusetzen mit hoher Varianz.

Abb. 5.3 zeigt die Verteilung von Fehlerhäufigkeiten der Sommermonate 2007 in der D-PHASE Region. Die paarweise Gegenüberstellung der Applikationen der Modellketten signalisiert im Fall von COSMO die stärksten Ausprägungen der relativen Häufigkeit in der Kategorie zwischen -0.25 und 0.25 mm/12, wobei die Besetzung dieser Klasse der geringsten Abweichung für hochaufgelöste Modelle deutlich größer ausfällt. Die sehr symmetrischen Verteilungen von COSMO-EU und COSMO-DE weisen zudem die geringste Varianz und dadurch die größte Schärfe der Auswahl auf. Modelle von Meteo Swiss hingegen zeigen trotz hoher Werte der besten Kategorie zusätzlich auch Anteile in Klassen höherer Bias-Werte und damit insgesamt auch größere Varianz. Interessant sind aber vor

allem die relativ großen Anteile bei leichter Unterschätzung der Intensität und auf der anderen Seite die teils großen Überschätzungen, jedoch mit sehr geringen Häufigkeiten. Eine subjektive Einschätzung der Symmetrieeigenschaften lassen in diesem Fall auf rechtsschiefe Verteilungen schließen. ALADIN und AROME verhalten sich hingegen gänzlich anders. Die ebenfalls rechtsschiefen aber erheblich flachen Verteilungen zeigen teils hohe Anteile der relativen Häufigkeit in größeren Kategorien und somit insgesamt eine markante Überschätzung der Niederschlagsintensität.

Die Darstellung der relativen Anteile der Amplitudenfehler im Vergleichszeitraum der Herbstmonate September bis November in Abb. 5.4 zeigt in Relation zur konvektiven Jahreszeit geringere Varianzen durch die Erhöhung der Verteilungsschärfe. Die Häufigkeiten der Kategorie der geringsten Abweichungen sind insgesamt zwar größer, im Fall von COSMO-DE fällt diese im Vergleich zu COSMO-EU aber leicht ab. Auffallend ist aber die starke Asymmetrie aller Verteilungen mit sehr geringen Anteilen bei negativen Klassenzahlen und in Summe einer generellen Überschätzung der Niederschlagsintensität. Besonders ausgeprägt sind die Auswirkungen bei Abnahme der Konvektion im Herbst aber in Bezug auf die französischen Modelle ALADIN und AROME.

#### 5.1.4 Stratifizierung des Bias innerhalb der COPS- Domäne

Die Wahl des Verifikationsgebiets bei der Auswertung und dem Vergleich meteorologischer Parameter hat teils maßgeblichen Einfluss bei der Überprüfung der Leistungsfähigkeit und Qualität einer Prognose. Besonders anschaulich wird diese Sensibilisierung bezüglich des gewählten geographischen Ausschnitts über die Einflüsse von Gebirgen. Die verschiedenen Modellapplikationen verwenden eine von der jeweiligen Auflösung abhängige Topographie, somit eine modellabhängig und individuelle Repräsentierung eines Gebirgszuges als idealisierte Abbildung der realen Form des Geländes. Mit größer werdender Auflösung nähert sich die Modelltopographie zunehmend der realen Topographie an. Umgekehrt werden also Gebirgszüge in schwächer auflösenden Modellen geglättet. Die Dämpfung der Amplituden und Abschwächung der Gradienten führt neben der Erniedrigung des Kammniveaus auch zur Streckung der räumlichen Ausdehnung eines Massivs und in weiterer Folge zu teils erheblichen Auswirkungen auf die Modellierung meteorologischer Parameter.

Zur gezielten Untersuchung dieser Einflussfaktoren auf die Prognosequalität werden die Modelle im Gebiet COPS mit den Mittelgebirgen der Vogesen in Ostfrankreich und dem Schwarzwald in Baden Württemberg im Süd-Westen Deutschlands analysiert. MET ermöglicht eine Verifikation innerhalb einer oder mehrerer über Polylinien aufgespannte Flächenstücke des Datenbereichs, deren Festlegung über die Angabe und Verbindung der Koordinatenpunkte erfolgt. Zur weiteren Unterteilung des COPS- Bereichs und der Isolation der Gebirgszüge wird eine Einteilung in drei etwa gleich große Subdomänen vor-

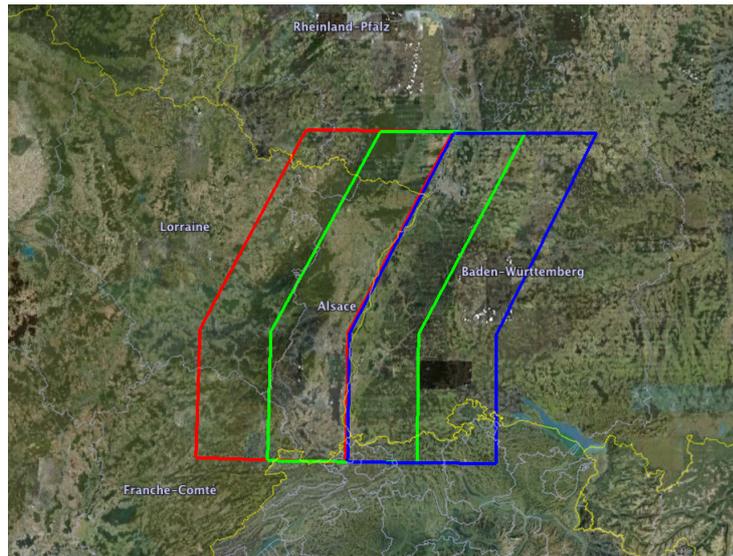


Abbildung 5.5: COPS- Gebiet mit den Subdomänen Vogesen (rot), Rheintal (grün) und Schwarzwald (blau). Die Einhüllende aller entspricht dem Gebiet von COPS. (Quelle: Google Earth, verändert)

genommen, deren Bezeichnungen das jeweilige Einzugsgebiet kennzeichnen. Das Rheintal entlang der Kammlinien der Vogesen und des Schwarzwalds umfasst 310 Gitterpunkte, die Gebiete der Vogesen mit 316 und die Schwarzwald- Domäne mit 314 Gitterpunkten liegen jeweils eine halbe Ausschnittsbreite nach Westen bzw. Osten verschoben.

Modellspezifische Auswertungen und Vergleiche des Bias für die Untergebiete werden in Abb. 5.6 dargestellt. Wie auch schon bei Auswertungen des wesentlich größeren D-PHASE- Bereichs hervorgeht, ist die dominierende Richtung des Bias, insbesondere aber für Modelle mit gröberer Auflösung, positiv. Die modellseitige Über- oder Unterschätzung der Ereignisintensität lässt sich hier allerdings auf die Einflussnahme der beiden Mittelgebirge bei der Untersuchung des räumlich sehr begrenzten Ausschnitts von COPS und dessen Subdomänen zurückführen. Wulfmeyer et al. (2008) [44] spricht in diesem Zusammenhang von einem Luv/Lee-Effekt über die systematische Überprognose der Intensität von Niederschlagsereignissen im Luv eines Gebirges und einer Unterschätzung im Lee.

Dies lässt sich auch über die Auswertung der Differenzfelder der Monatsmittelwerte am Beispiel des Augusts 2007 für die Modelle COSMO7 und COSMO2 in Abb. 5.7 bestätigen. Durch die zusätzliche Einbindung der Modelltopographie in die Graphik ist eine genaue Lokalisierung der Maxima des Bias in Gebirgsnähe vor der Kammlinie im Luv des Gebirgszuges möglich, mit Annahme einer vorherrschenden Windrichtung aus West. Der leeseitige Effekt der Unterschätzung der Intensität tritt hingegen im Rheintal aufgrund des steilen Abfalls der Vogesen deutlich stärker hervor als auf der flacheren Leeseite des

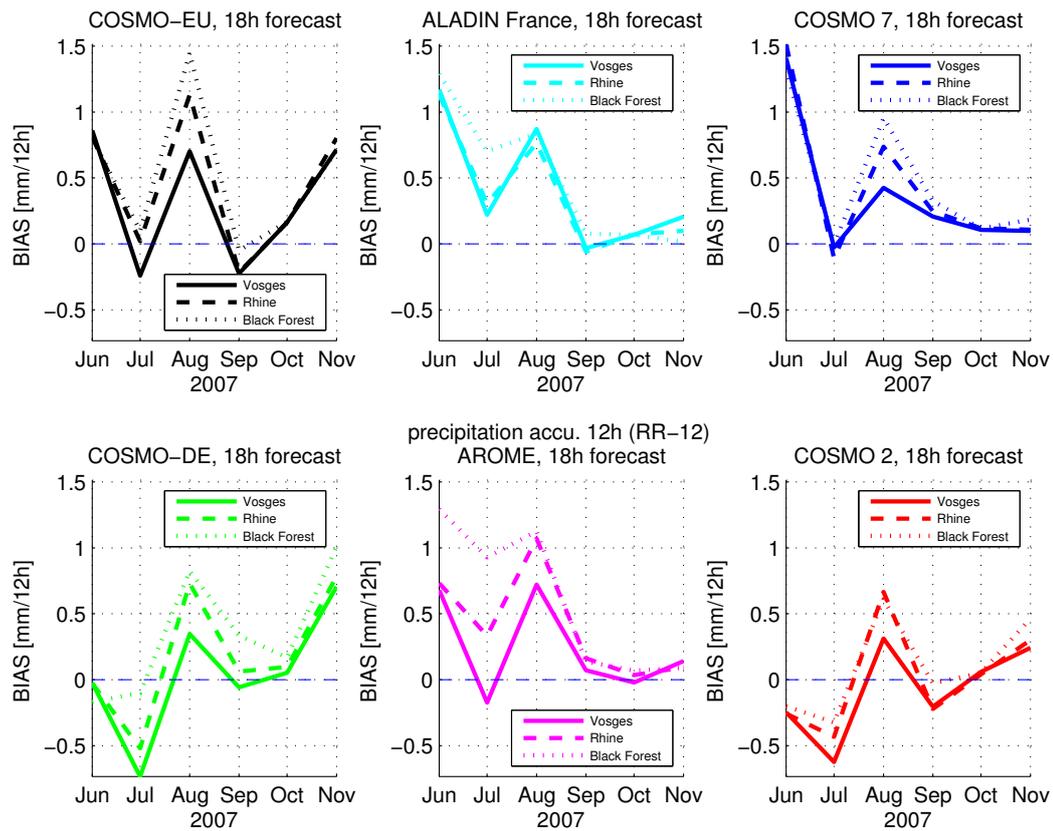


Abbildung 5.6: Monatsmittelwerte des Bias innerhalb der COPS- Subdomänen Vogesen, Rheintal und Schwarzwald. 18-stündige Modellprognosen des 12-stündig akkumulierten Niederschlags. Niederschlagsschwelle: 0.1 mm/12h.

Schwarzwaldes. COSMO2 zeigt insgesamt eine deutlich geringere Überprognose im Luv, jedoch auch einige weitere Abweichungsmaxima und Minima, die nicht an die Einflüsse der Gebirge geknüpft scheinen. Zudem ergeben auch die Auswertung und die Gegenüberstellung der Modelle des DWD, d.h. COSMO-EU und COSMO-DE, starke Ähnlichkeiten über die Kopplung der systematischen Fehler der Modelle an lokale Erhebungen des Geländes.

Vergleiche der Ergebnisse aus Abb. 5.6 zur Stratifizierung des Bias innerhalb der Subdomänen von COPS mit den Differenzfeldern der schweizer Modelle COSMO7 und COSMO2 im August 2007, deuten auf eine generelle Überschätzung der prognostizierten Intensität bezüglich aller drei Gebiete hin. Die kleinsten Werte treten im Gebiet der Vogesen und die größten im Schwarzwald auf. Die etwas unglückliche Positionierung der gewählten Untergebiete führt durch die Beteiligung von sowohl luv- als auch leeseitig beeinflussten Berghängen innerhalb eines Gebietes teilweise zu kompensierenden Effekten. Auffallend ist aber dennoch eine Staffelung der Kurven, wobei mit Ausnahme von ALADIN France die Werte im Gebiet der Vogesen häufig den untersten Kurvenverlauf aufweisen. Die Verläufe

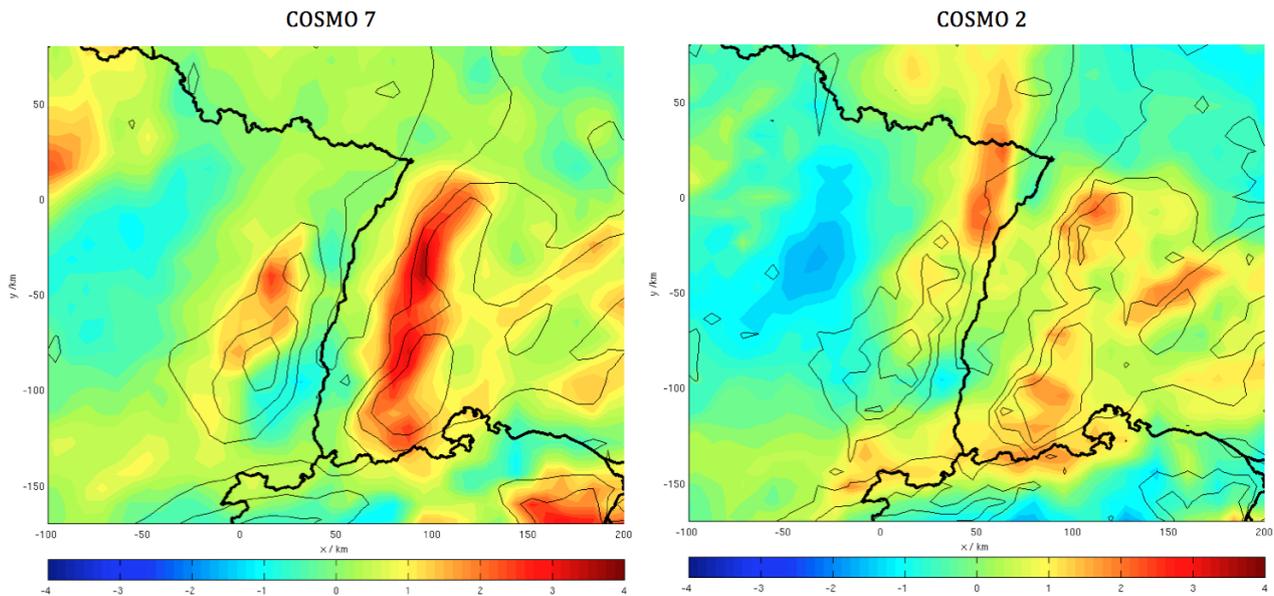


Abbildung 5.7: Differenzfelder der Monatsmittelmerte für August 2007 (Modell - VERA) des 12-stündig akkumulierten Niederschlags im Gebiet COPS. Modellläufe um 00 UTC mit einer Prognosedauer von 18h. Dünne schwarze Linien zeigen die Modelltopographie, Differenzfelder werden als Farbflächen der Einheit mm/12h dargestellt.

vom Schwarzwald hingegen zeigen oftmals die höchsten modellseitigen Überschätzungen der Intensität und bilden daher die obere Verlaufskurve.

Auswertungen des 24-stündig akkumulierten Niederschlags des 00 UTC Laufes und einer Gültigkeit um 06 UTC des Folgetages berücksichtigen den gesamten prognostizierten Tagesniederschlag 30-stündiger Vorhersagen innerhalb eines Monats. Die ähnlichen Verhältnisse mit der Überprognose im Luv- Bereich sowie die modellseitige Unterschätzung der Niederschlagsmenge im Lee treten sehr deutlich hervor und weisen zudem wesentlich größere Magnituden im Vergleich zu Auswertungen der 12-stündigen Akkumulierung auf. Eine Generalisierung dieser Zusammenhänge über eine Ausweitung auf kürzere Akkumulationszeiten lässt sich allerdings nicht finden. Differenzfelder des 3-stündigen Niederschlags zeigen zwar teilweise, jedoch in abgeschwächter Form, die direkte Kopplung der Ereignisse an die Orographie, wobei aber auch vermehrt Anteile eines inversen Effekts auftreten. Neben der Einflussnahme der Orographie auf Niederschlagsereignisse und den damit verbundenen Schwierigkeiten bei der Modellierung, besonders bei Konvektion, werden im weiteren natürlich auch andere meteorologische Parameter wie Luftdruck oder die Komponenten des Winds beeinflusst, deren Auswertung hier aber nicht vorgenommen wird.

# 6 Objektorientierter Verifikationszugang über MODE

MODE zur objektorientierten Analyse von Niederschlagsgebieten als eigene Distribution von MET wurde bereits im Abschnitt 3.1.3 ausführlich behandelt sowie entsprechend auf die verwendeten Techniken und Methodiken zur Objektdefinition und ferner auf die Parameter zur Beschreibung dieser Gebilde hingewiesen. In diesem Kapitel sollen praktische Beispiele zur Untermauerung der Bedeutung objektorientierter Verifikation oder dessen Vorzug gegenüber konventionellen Verifikationsmethoden im Vordergrund stehen. Im Speziellen erfolgt die Evaluierung der Modelle vorerst an einzelnen Modellen, im Weiteren aber auch über die Gegenüberstellung mehrerer Modelle anhand genau festgelegter Rahmenbedingungen, deren Beschreibung nachfolgend im Mittelpunkt steht.

## 6.1 Auswertung von extremen Niederschlagsereignissen

Die Festlegung der Rahmenbedingungen eines Experiments wird im Allgemeinen, speziell aber bei der Anwendung von MODE, als zentraler Aspekt betrachtet, um sich nicht in einer Vielzahl von Einstellungsmöglichkeiten zu verlieren. Ziel dieser Untersuchung ist die Evaluierung der Modellgüte von Starkniederschlägen, im Speziellen aber von konvektiven Niederschlägen hoher Intensität. Ein Großteil der im Sommer eintretenden Niederschlagsereignisse entstammt konvektiven Zellen. Treten hohe Niederschlagsintensitäten innerhalb kurzer Zeitintervalle auf, so können diese häufig zu kleinräumigen lokalen Überflutungen führen. Die Auflösung dieser Ereignisse in Bezug auf die zeitliche und räumliche Exaktheit des Eintretens von Extremereignissen wird von Vorhersagemodellen zur rechtzeitigen Warnung möglicher betroffener Gebiete gefordert. Gegenstand dieses Experiments ist die Untersuchung der räumlichen Schärfe von Niederschlagprognosen und der Vergleich verschiedenartiger Modelle und Auflösungen mit oder ohne der Verwendung von Parametrisierungsschemata.

### 6.1.1 Versuchsaufbau

Als Untersuchungszeitraum werden die Sommermonate zwischen Juni und August 2007 verwendet. Die räumliche Ausdehnung des betrachteten Gebietes erstreckt sich über Zentraleuropa und entspricht dem im Abschnitt 2.5.1 definierten D-PHASE Ausschnitt bei der Beteiligung aller verwendeten Modelle. Betrachtet werden 18-stündige Niederschlagsprognosen einer Akkumulationszeit von 12 Stunden (RR\_12h). Wie oben erwähnt, werden ausschließlich Modellläufe mit einer Initialisierungszeit um 00 UTC verwendet, bei der Betrachtung 18-stündiger Prognosen ergibt sich somit ein Zeitpunkt der Gültigkeit von 18 UTC. Die mitunter wichtigsten in MODE zu treffenden experimentspezifischen Anpassungen sind eine adäquate Einstellung des Einflussradius der Glättung  $R$  sowie die eines Schwellwerts  $T$  zur Definition der Niederschlagsobjekte innerhalb der geglätteten Felder.  $R$  wurde hierbei auf einen Wert von 4 Gitterpunkten eingestellt, also dem kleinstmöglichen Glättungsradius, bei dem noch realistische Strukturen zu erwarten sind, vgl. Abschnitt 3.1.3.1. Der Schwellwertparameter  $T$  erhält hier einen Wert von größer gleich 25 mm/12h. Wichtig ist hier nochmals festzuhalten, dass der Schwellwert  $T$  auf die geglätteten Felder angewendet wird und ausschließlich der Festlegung der Objektkontouren dient. Nach Aufprägung der originalen Niederschlagsfelder auf die Objekte können durchaus Werte auftreten, die erheblich kleiner sind als  $T$ , der überwiegend größere Teil der Niederschlagsraten innerhalb der Objekte wird allerdings den Schwellwert übertreffen. Eine Schwelle von 25 mm/12h bedeutet also nicht zwangsläufig die Untersuchung von Niederschlag einer Rate von 25 mm/12h oder mehr. Infolge der Vereinheitlichung durch die Interpolation der Modellwerte auf VERA- Gitterpunkte entspricht der Glättungsradius von 4 Gitterpunkten einer einheitlichen Distanz von 32 km für alle Modelle in gleichem Maße.

Zusammenfassend bestehen die wichtigsten Einstellungen bislang aus der Definition von  $R$ ,  $T$  und des geographischen Ausschnitts. Angewandt auf den zu betrachtenden Zeitraum zwischen Juni und August 2007 ergeben sich insgesamt 11 Fälle, auf die alle getroffenen Voraussetzungen zutreffen. Diese können Tabelle 6.1 entnommen werden.

| Monat      | Juni 2007 |    | Juli 2007 |    |    |    | August 2007 |    |     |       |    |
|------------|-----------|----|-----------|----|----|----|-------------|----|-----|-------|----|
| Tag        | 15        | 21 | 02        | 04 | 09 | 23 | 08          | 20 | 21  | 29    | 30 |
| Ereignisse | 3         | 1  | 4         | 1  | 1  | 1  | 2           | 1  | 1   | 1     | 2  |
| IOP        | 3b        | -  | 5b        | 6  | 7b | 10 | 14b         | -  | 17a | SOP-8 | -  |

Tabelle 6.1: Auswahl von 11 Fällen zwischen Juni und August 2007 im Gebiet D-PHASE mit Anzahl der jeweiligen Niederschlagsereignisse, sowie den während COPS definierten IOP's und SOP's

Neben der Aufzählung der Tage wird zusätzlich eine Verbindung zu den speziell während

COPS festgelegten IOP'S (Intensive Observation Period) und SOP's (Special Observing Period) hergestellt, die die Zeiträume konzentrierter und intensiver Beobachtungen für den wesentlich kleineren Ausschnitt von COPS während besonders auffälliger Wetterlagen widerspiegeln und mit einigen Fällen der getroffenen Auswahl zusammenfallen. Diese Auflistung sollte wegen der Konzentration auf das Gebiet von COPS allerdings nur als Vergleich betrachtet werden.

Eine hinreichende Voraussetzung zur Festlegung, ob ein Tag, im Speziellen aber der 18 UTC Termin, für das vorliegende Experiment in Frage kommt, ist das Vorhandensein eines oder mehrerer Niederschlagsobjekte im Feld der Analyse. Diese Vorgehensweise erlaubt eine von den spezifischen Modellen unabhängige Ausgangslage von Objektfeldern jedes Termins, im weiteren Sinne gewissermaßen eine Stratifizierung auf Basis von Beobachtungen. Der umgekehrte Fall einer Stratifizierung auf Vorhersagebasis wäre durchaus denkbar, aufgrund der mangelnden Vergleichbarkeit mit anderen Modellen hier aber nicht zielführend. Liegen mehrere Objekte zu einem Analysetermin vor, werden diese als eigenständige Ereignisse durch die zum Teil unterschiedlichen Antriebsmechanismen aufgefasst. Die Zahl der zur Verfügung stehenden Ereignisse erhöht sich somit von 11 auf insgesamt 18.

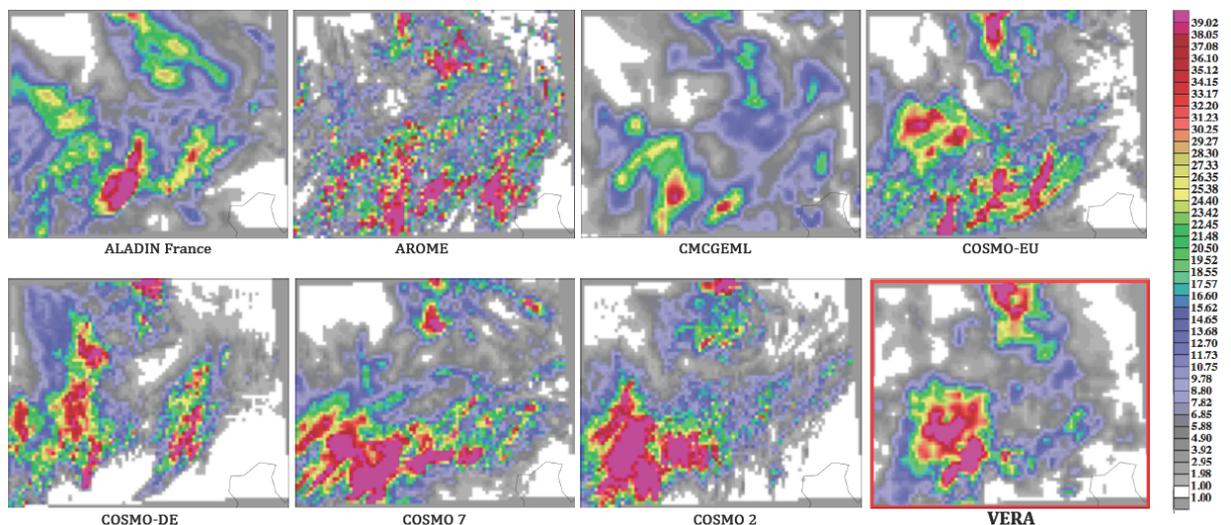


Abbildung 6.1: Vorhersagefelder des 12-stündig akkumulierten Niederschlags aller verwendeten Modelle am 8.8.2007, 18 UTC, Modellinitialisierungen 00 UTC (18-stündige Prognosen). Kleinste dargestellte Isohyete: 1 mm/12h; größte dargestellte Isohyete: 40 mm/12h. Gebiet entspricht dem D-PHASE Bereich, hier genauer dem größten gemeinsamen Gebiet aller beteiligten Modelle.

### 6.1.2 Objektidentifikation aus Niederschlagsfeldern

Zur Verdeutlichung des Ablaufs zur Identifikation von Objekten bezüglich der getroffenen Annahmen und Einstellungen wird der Vorgang anhand eines Beispiels für den 8. August 2007 gezeigt. Die Ausgangslage wird über die von den einzelnen Modellen prognostizierten Felder und über die Felder der VERA-Analyse des 12-stündig akkumulierten Niederschlags um 18 UTC mit einer Prognosedauer von 18 Stunden festgelegt, dargestellt in Abbildung 6.1. Die vorherrschende Wetterlage, geprägt von einem weit nach Süden ausgreifenden Höhentrog mit zunehmender Abschnürung über Frankreich und Entkoppelung vom Höhensystem, führte in Verbindung mit einer Frontalzone zu beträchtlichen Niederschlagsmengen und zahlreichen Überflutungen innerhalb der Zentral- und Westschweiz. Ein weiteres Niederschlagsgebiet hoher Intensität befand sich über Deutschland, im östlichen Baden-Württemberg. Die Abgebildete Skala zeigt Raten zwischen 1 mm/12h als untere Grenze und 40 mm/12h als größte dargestellte Isohyete.

Die direkte Gegenüberstellung der Modellfelder mit dem Feld der Analyse von VERA ermöglicht neben einer rein qualitativen Übersicht auch durchaus interessante Einblicke. Das Forschungsmodell des französischen Wetterdienstes AROME etwa zeigt sehr feine Strukturen über große Bereiche, die hingegen bei Modellen mit vergleichbaren Auflösungen wie COSMO2 und COSMO-DE nicht, bzw. nicht in diesem Ausmaß zu finden sind. Demgegenüber stehen VERA-Analysen mit einer Auflösung von 8 km. Auftretende Glättungseffekt in der Analyse sind unvermeidbar und neben der geforderten Auflösung der Felder auch vom Analyseschema selbst abhängig. Kleinskalige Prozesse etwa in Form lokaler kovektiver Zellen werden folglich auch von den 8 km VERA-Analysen nicht zur Gänze aufgelöst, obwohl die in den Modellfeldern dargestellten Strukturen durchaus eingetreten sein könnten. Ein weiterer interessanter Aspekt aus der direkten Gegenüberstellung ergibt sich aus dem Vergleich der Niederschlagsfelder der COSMO-Modellfamilie, die eine starke Kopplung der feineren Modellapplikationen mit den Antriebsmodellen aufweist. Die Modellkette von COSMO-EU und COSMO-DE des Deutschen Wetterdienstes DWD ist in das globale hydrostatische Modell GME genestet, währenddessen die Anwendungen der Modelle von Meteo Swiss, COSMO7 und COSMO2, als antreibendes Modell ECMWF verwenden. Dieser Unterschied ist im wesentlichen neben der Verwendung leicht unterschiedlicher Auflösungen der bedeutendste beim Vergleich der beiden Modelle und deutet offenbar sowohl auf die Relevanz als auch auf die Dominanz des antreibenden Modells bezüglich der nachfolgenden Modellkette hin. Mit anderen Worten, wenn hochauflösende Modelle zu Beginn schon mit falschen Informationen und Feldern beliefert werden, können sie ihre Fähigkeiten und Potentiale nicht oder zumindest nicht in vollem Umfang ausspielen, und der Gewinn gegenüber schwächer auflösenden Modellen schwindet.

Über die oben getroffenen Einstellungen für R und T lässt sich Abb. 6.1 der Nieder-

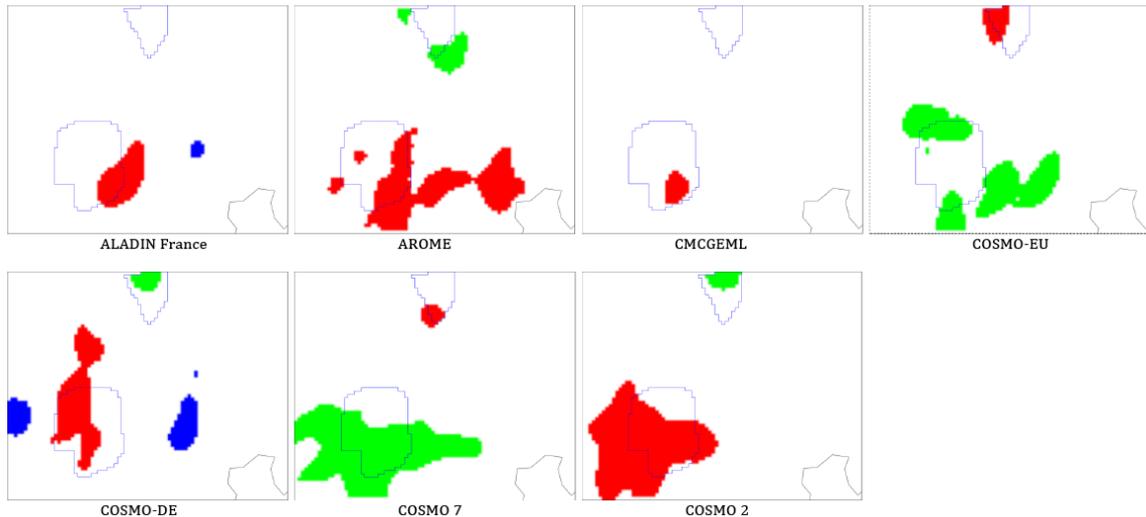


Abbildung 6.2: Darstellung der aus Abb. 6.1 generierten Niederschlagsobjekte mit MODE: Farbige Felder entsprechen den jeweiligen Niederschlagsobjekten der Modelle, die konturenhaften Darstellungen den Objekten der Analyse. Blau dargestellte Objekte sind als "falsche Alarmer" aufzufassen. Radius des Glättungsfilters  $R=4gp$  bei einem Schwellwert von  $T=25\text{ mm}/12h$ . Einstellungen auf beiden Seiten sind äquivalent.

schlagsfelder in eine objektbasierte Darstellung überführen. Abbildung 6.2 zeigt dabei eine Überblendung der Vorhersageobjekte (bunte Flächen) mit den Objekten der Analyse (Objektkonturen). Bei der Überlagerung der Objekte wird eine gewisse Unschärfe zugelassen, die nach den Kriterien eines Fuzzy-logic Systems definiert ist (vgl. Abschnitt 3.1.3.3). Werden die Annahmen des Fuzzy-Systems nicht mehr erfüllt, das heißt, liegt ein Prognoseobjekt beispielsweise zu weit von der Objektkontur der Analyse entfernt, so werden die Niederschlagsobjekte der Vorhersage blau abgebildet und in weiterer Folge als Fehlalarme aufgefasst. Die Farbgebung aller anderen Objekte (rot, grün) wird programmintern nach keinen besonderen Vorgaben vergeben und enthält daher auch keine zusätzliche Information.

### 6.1.3 Objektanalyse anhand der Auswertung von Schwerpunktsdifferenzen

Mehrere Niederschlagsobjekte innerhalb eines Feldes können unter bestimmten Voraussetzungen zu einem Cluster zusammengefasst werden. Die Kriterien dieses Prozesses, der hier unter dem Begriff "merging" zusammengefasst wird, beinhalten einen Abgleich mit den Objekten im jeweils anderen Feld, wobei wieder eine gewisse Unschärfe über ein Fuzzy-Engine System zulässig ist. Durch die Spezifikation der objektbeschreibenden Parameter an das neu entstandene Clusterobjekt lässt sich dieses beschreiben und mit einem Ob-

jekt auf der Gegenseite vergleichen. Eine konvexe objekteinhüllende Begrenzungsline legt Form und Orientierung des Clusterobjektes fest, veranschaulicht über einen gedanklichen "Lassowurf" um die beteiligten Objekte. Die konvexe Hülle erlaubt ausschließlich positive Krümmungen bis hin zu Geraden, sodass Cluster mit sehr einfachen Formen entstehen. Der Prozess des mergings wird aus Gründen der Vergleichbarkeit nur auf der Modellseite durchgeführt, wodurch sichergestellt wird, dass die Analyseobjekte für jedes Modell gleich bleiben. Diese Vorgehensweise ist eine sinnvolle Einschränkung, da sich die Objekte der Analysen in Hinblick auf Form, Orientierung oder der Lage der Schwerpunkte so nicht ändern.

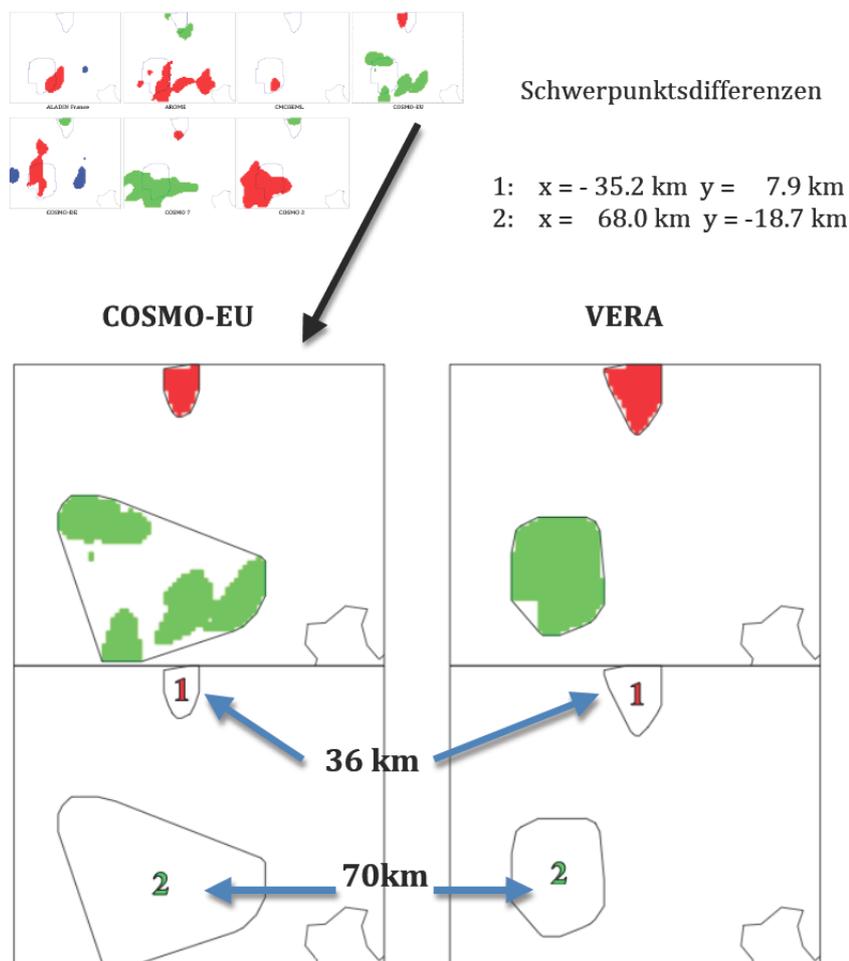


Abbildung 6.3: Clusterbildung der Niederschlagsobjekte auf der Vorhersageseite: Beispiel entspricht den Verhältnissen des Modells COSMO-EU am 8.8.2007 für eine 18-stündige Niederschlagsprognose der Akkumulationszeit von 12 Stunden aus Abb. 6.2. Schwerpunktsdifferenzen in  $x$ - und  $y$ -Richtung ermittelt aus der Koordinate des Schwerpunktes des Vorhersageobjekts abzüglich der Koordinate des Analyseobjektschwerpunktes ( $x_f - x_o$ ) und ( $y_f - y_o$ ). Die in der Abbildung eingetragenen Abstände zwischen den Objekten entsprechen den direkten Entfernungen zwischen den Schwerpunkten in Kilometer.

Eines der möglichen Anwendungsgebiete ist die Auswertung der Objektschwerpunkte über den Vergleich der Schwerpunktdifferenzen. Der Schwerpunkt eines Objektes wird über die objektinterne Niederschlagsverteilung ermittelt und entspricht hier dem massengewichteten Schwerpunkt. Vor allem bei der Clusterbildung wird somit der gemeinsame Schwerpunkt ausschließlich auf Basis der Niederschlagsverteilungen der beteiligten Objekte bestimmt und nicht etwa über die vom Cluster aufgespannte Fläche. MODE wertet die Objektschwerpunkte als x- und y Koordinate aus, ausgehend von der linken unteren Ecke des Bildbereiches. Durch die Differenzbildung der Koordinaten des Prognose- und Analyseobjektschwerpunktes in x- und y-Richtung getrennt von einander, werden zwei richtungsabhängige Größen erzeugt.

$$\Delta x = x_f - x_o \quad (6.1)$$

$$\Delta y = y_f - y_o \quad (6.2)$$

Eine Richtungsabhängigkeit der Schwerpunktdifferenzen in x- und y- Richtung bedeutet in diesem Zusammenhang, dass nach Festlegung der Subtraktionsrichtung nach Gleichung 6.1 und 6.2, ein negativer Wert einer Verschiebung des Prognoseobjektschwerpunkts nach Westen (x- Richtung) oder Süden (y- Richtung) entspricht, ein positiver einer nach Norden oder Osten. Ein Beispiel einer Clusterbildung von Vorhersageobjekten wird in Abbildung 6.3 anhand der Prognose des 12- stündigen Niederschlags von COSMO-EU dargestellt. Die berechneten Schwerpunktdifferenzen zeigen für Objekt 1 eine etwa 35 km westlich und 8 km nördliche Verschiebung des Vorhersageobjekts, für das Clusterobjekt mit der Kennzahl 2 eine knapp 70 km Verschiebung nach Osten sowie eine Dislokation um etwa 18 km nach Süden. Über die Beziehung

$$z = \sqrt{|x|^2 + |y|^2} \quad (6.3)$$

lässt sich der direkte Abstand zwischen den Schwerpunkten angeben, in der Abbildung mit den blauen Pfeilen angedeutet.

#### 6.1.4 Verteilung der Schwerpunktdifferenzen in Abhängigkeit eines Relevanzspektrums

Am Beispiel des 8.8.2007 wurde in den letzten beiden Abschnitten die grundsätzliche Vorgehensweise zur Objektdefinierung, zur Clusterbildung mehrerer Objekte einer Prognose und letztlich zur Berechnung von Schwerpunktdifferenzen demonstriert. Die Auswertung auf die in Tabelle 6.1 vorgestellten 11 Fälle, denen insgesamt 18 Ereignisse zuzuordnen

sind, wird im folgenden Kontext erläutert. Die oben getroffene Voraussetzung zur Festlegung eines Falls über das Vorhandensein zumindest eines Niederschlagsobjektes in der Analyse, somit gewissermaßen die Stratifizierung auf Basis von Beobachtungen, beinhaltet jedoch nicht das gleichzeitige Auftreten eines Objekts in den Prognosen. Werden die Kriterien zur Objektdefinierung im Feld der Vorhersage nicht oder nicht für alle Ereignisse des vorliegenden Falls erfüllt, so kann ein Objekt nicht erzeugt und ferner auch kein Vergleich objektbezogener Parameter angestellt werden. Dies ist auch der Grund für die modellspezifische Anzahl der vergleichbaren Objekte zwischen Analyse und Prognose der nachfolgenden Studie zur Untersuchung der räumlichen Exaktheit von Niederschlagsobjekten hoher Intensität. Durch die Zusammenfassung mehrerer Prognoseobjekte zu einem Cluster repräsentiert die modellspezifische Anzahl an Objekten gewissermaßen auch ein Gütemaß und somit einen Vergleichsparameter zwischen den Modellen.

Die Zuordnung eines Prognoseobjekts zu einem Objekt der Analyse ist nicht zwangsläufig eindeutig. Können mehrere Objekte in einem Feld bestimmt werden, ergeben sich auch mehrere Möglichkeiten beim Vergleich objektbeschreibender Parameter, im Speziellen also auch bei der Ermittlung von Schwerpunktsdifferenzen. Die Frage nach einem Kriterium, welches Prognoseobjekt überhaupt mit welchem Analyseobjekt assoziiert werden kann, wird hier über die Einführung eines Relevanzspektrums bezüglich des paarweisen Vergleiches von Objekten berücksichtigt. Dieses Kriterium stützt sich auf die schon im Abschnitt 3.1.3.3 eingeführte Maßzahl der Gesamtrelevanz, die eine Verknüpfung objektbeschreibender Attribute und Eigenschaften mit spezifischer Wichtung unter Einbindung eines Vertrauensbereichs darstellt und hier nochmals skizziert ist.

$$T(\alpha) = \frac{\sum_i w_i C_i(\alpha) I_i(\alpha_i)}{\sum_i w_i} \quad (6.4)$$

Die Anpassung der Gewichtungsfaktoren der Objektattribute  $w_i$  ist experimentspezifisch, wobei hier die in Tabelle 6.2 getroffenen Einstellungen verwendet wurden. Standardmäßig sieht MODE eine stärkere Gewichtung des geringsten Abstandes zwischen zwei Objekten mit einem Faktor von 4 vor, infolgedessen dafür aber keine Gewichtung des Intensitätsverhältnisses. Bei der Untersuchung von Extremniederschlägen spielen neben den räumlichen Faktoren des Auftretens eines Ereignisses auch die richtige Prognose der Niederschlagsintensität eine entscheidende Rolle, die in der vorliegenden Untersuchung über eine stärkere Gewichtung des Intensitätsverhältnisses auf Kosten einer schwächeren Interpretation der Relevanz des geringsten Objektabstandes berücksichtigt wird. Die Summe der Gewichtungsfaktoren muss sich auf keinen bestimmten Wert belaufen. Bei der Berechnung der Gesamtrelevanz nach Gleichung 6.4 erfolgt hierfür eine Normierung der Gewichtungsfaktoren über die Gesamtsumme der Gewichte.

Als durchaus kritischer Aspekt sei an dieser Stelle angemerkt, dass bei der Vergabe der Wichtungsfaktoren zwangsläufig ein gewisses Maß an Willkürlichkeit einfließt. Die große Freiheit an Einstellungsmöglichkeiten verleitet den Benutzer möglicherweise dazu, auf Kosten der Objektivität einer Auswertung Ergebnisse in bestimmte Richtungen zu lenken. Andererseits ergeben sich daraus aber auch Möglichkeiten zur sehr subtilen Untersuchung nach ganz speziellen Kriterien.

| Attributsbezeichnung   | Gewichtung |
|--|------------|
| Schwerpunksdifferenz   | 2          |
| geringster Abstand zwischen Objekten                               | 2          |
| geringster Abstand zwischen konvexen Objekthüllen                  | 0          |
| Winkeldifferenz  | 1          |
| Flächenverhältnis der Objekte                                      | 1          |
| Flächenverh. der Überschneidungsflächen / Flächenverh. der Objekte | 2          |
| Komplexitätsverhältnis   | 0          |
| Intensitätsverhältnis  | 2          |
|  | 10         |

Tabelle 6.2: Gewichtung von Objektattributen

Die Einbeziehung der Gesamtrelevanz beim Vergleich zweier Objekte ist also ein durchaus notwendiges Kriterium zur Einschätzung und Beurteilung, inwiefern die Vergleichsparameter, im vorliegenden Fall der Abstand zwischen den Objektschwerpunkten repräsentiert werden, bzw. welches Vertrauen einem gewissen Abstand entgegengebracht werden kann. Abbildung 6.4 zeigt eine Darstellung der räumlichen Verteilung der Schwerpunksdifferenzen aus dem Vergleich von Analyseobjekten mit Objekten der Modelle als Scatterplot aus insgesamt 11 Fällen (18 Ereignisse), deren Gesamtrelevanz bei der Gegenüberstellung der Objektpaare einen Wert von 80% übersteigt.

Die oben beschriebene modellspezifisch unterschiedliche Anzahl an Punkten durch Erreichen der geforderten Kriterien kann auch bei der Mittelwertbildung, etwa am Beispiel von CMCGEML mit nur vier Punkten, zu einem geringen Stichprobenumfang führen und die Vertrauenswürdigkeit bei der Interpretation des ermittelten Wertes herabsetzen. Die Farbgebung eines Punkts spiegelt den Wert der Gesamtrelevanz in direkter Weise wider. Zusätzlich zur Streubreite der Punkte ist zur Mittelwertbildung auch der Wert der Gesamtrelevanz von Bedeutung und sinnvoll, diesen als Gewichtungsfaktor in die Berechnung mit einzubeziehen. Weit entfernte Punkte und Punkte mit vergleichsweise geringer Relevanz werden somit schwächer gewichtet.

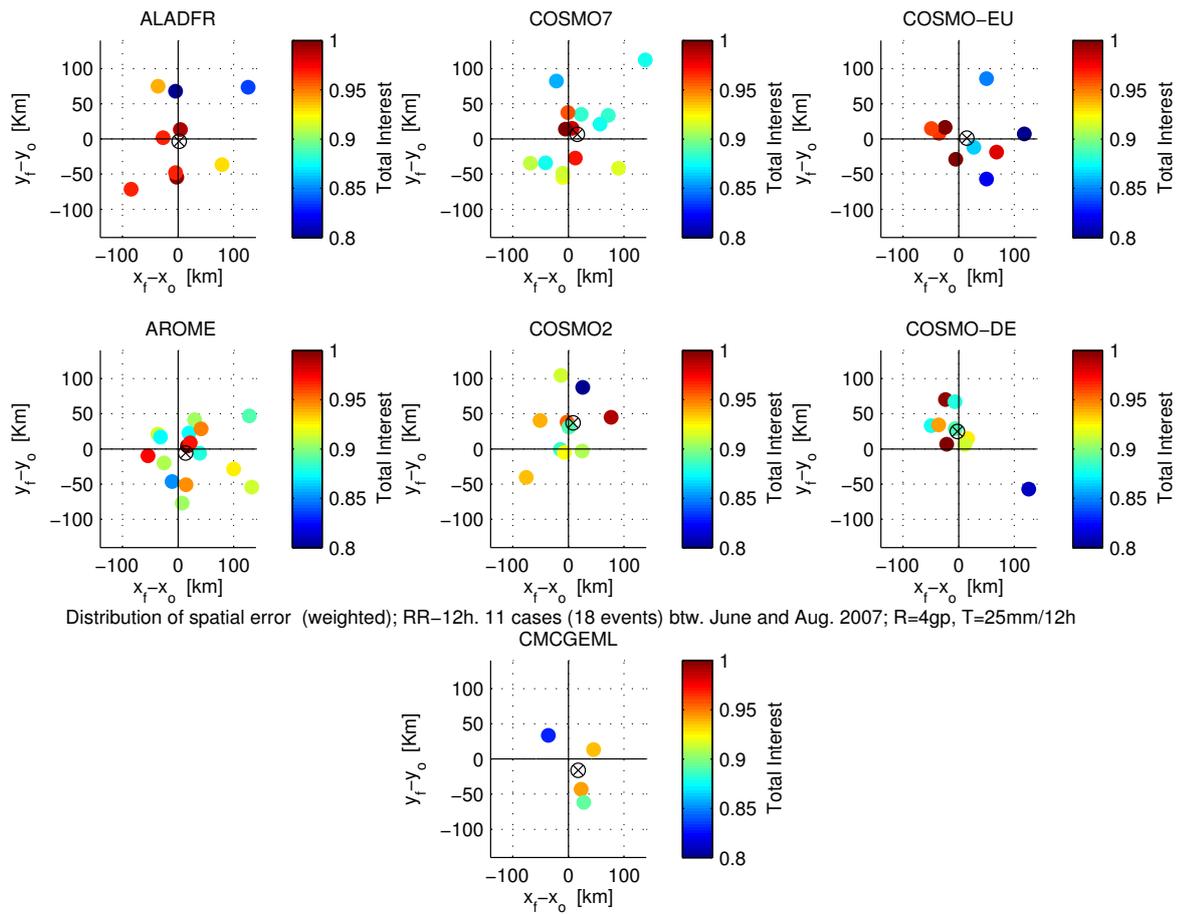


Abbildung 6.4: Räumliche Verteilung der Schwerpunktsdifferenzen von Niederschlagsobjekten. Differenzen in  $x$ - und  $y$ - Richtung in km. Die unterschiedliche Anzahl an Punkten ergibt sich aus der Festlegung der Rahmenbedingungen, im Speziellen der Parameter  $R=4$  gp und  $T=25$  mm/12h. Grenze der Gesamtrelevanz (total interest  $T(\alpha)$ ) ist 0.8.  $\otimes$  zeigt den über die Gesamtrelevanz gewichteten Mittelwert der beteiligten Punkte.

Die Auffassung der modellspezifischen Anzahl an Punkten in der Darstellung als eigenes Gütemaß ermöglicht eine Auswertung und Rangfolge der Modelle. Kann jedem Objekt der Analyse (insgesamt 18) ein entsprechendes Prognoseobjekt oder in weiterer Folge ein Cluster aus mehreren Objekten zugeteilt werden, so wird nach den festgelegten Kriterien auch eine gute Prognose geliefert. Wenn allerdings kein geeignetes Objekt gefunden werden kann, so liegen die Gründe entweder bei einer deutlichen Unterschätzung der Intensität des Extremereignisses oder aber bei einer zu starken Dispositionierungen der Objekte und Ablehnung durch die Kriterien. AROME mit insgesamt 17 Objekten liegt dabei deutlich voran, gefolgt von den Modellen der schweizer Modellkette. Die geringe Anzahl bei CMCGEML lässt sich über eine generelle Unterschätzung der Niederschlagsintensität des Modells erklären.

| AROME | COSMO 7 | COSMO 2 | ALADIN | COSMO-DE | COSMO-EU | CMCGEML |
|-------|---------|---------|--------|----------|----------|---------|
| 17    | 14      | 11      | 9      | 9        | 9        | 4       |

Tabelle 6.3: Anzahl der Prognoseobjekte oder Cluster aus Abb. 6.4

Im Vergleich weist ALADIN- France in Abb. 6.4 bei gleichzeitig großer Streuung der Punkte einen Mittelwert sehr nahe am Ursprung auf. Dieser sollte allerdings aufgrund der kompensierenden Eigenschaften bei der Mittelwertbildung nicht überbewertet werden. Generell zeigt sich eine leicht verminderte Varianz bei hochauflösenden Modellen im Vergleich zu den Antriebsmodellen der selben Modellkette. Eine Betrachtung der Mittelwerte der COSMO- Modellkette zeigt eine relative Verschiebung der prognostizierten Niederschlagsobjekte zu den Objekten der Analyse nach Osten für die konvektionsparametrisierenden Modelle COSMO7 und COSMO-DE in der Größenordnung zwischen 10 und 15 km und eine wesentlich deutlichere für die hochauflösenden Modelle COSMO2 und COSMO-DE in nördliche Richtung mit Werten im Bereich von 30 bis 40 km.

| ALADFR  | COSMO-EU | AROME    | COSMO7   | CMCGEML  | COSMO-DE | COSMO2   |
|---------|----------|----------|----------|----------|----------|----------|
| 3.97 km | 14.58 km | 14.88 km | 17.45 km | 24.02 km | 24.79 km | 37.99 km |

Tabelle 6.4: Länge des mittleren Verbindungsvektors

Tabelle 6.4 zeigt einen Überblick über die aus Gleichung 6.3 ermittelte Distanz der über die Gesamtrelevanz gewichteten Mittelwerte der Schwerpunktdifferenzen der einzelnen Modelle.

### 6.1.5 Untersuchung der Intensitätsverteilungen von Niederschlagsobjekten

Die bisherigen Auswertungen und Analysen auf Basis einer objektbasierenden Verifikationsmethode, im speziellen dem Vergleich von Objektschwerpunkten, erlauben eine quantitative Einschätzung der räumlichen Exaktheit von Niederschlagsprognosen. Massengemittelte Objektschwerpunkte nehmen zwar indirekt Bezug auf die Niederschlagsverteilung, spiegeln aber keineswegs die realen Verteilungen und Verhältnisse wider. Eine räumlich exakte Positionierung des Niederschlagsmaximums infolge der richtigen Darstellung der Verteilung im Inneren eines Objekts, lässt allerdings keine qualifizierte Aussage über die adäquate Abbildung der Niederschlagsintensität zu, der bei der Prognose von Extremereignissen besonderes Interesse gilt.

Die Beschreibung und der Vergleich von Niederschlagsverteilungen innerhalb der Objekte wird über Perzentile zum Ausdruck gebracht. Perzentile teilen eine Verteilung gleichmäßig in 100 Abschnitte gleicher Größe. Der einem Perzentil entsprechende Wert gleicht

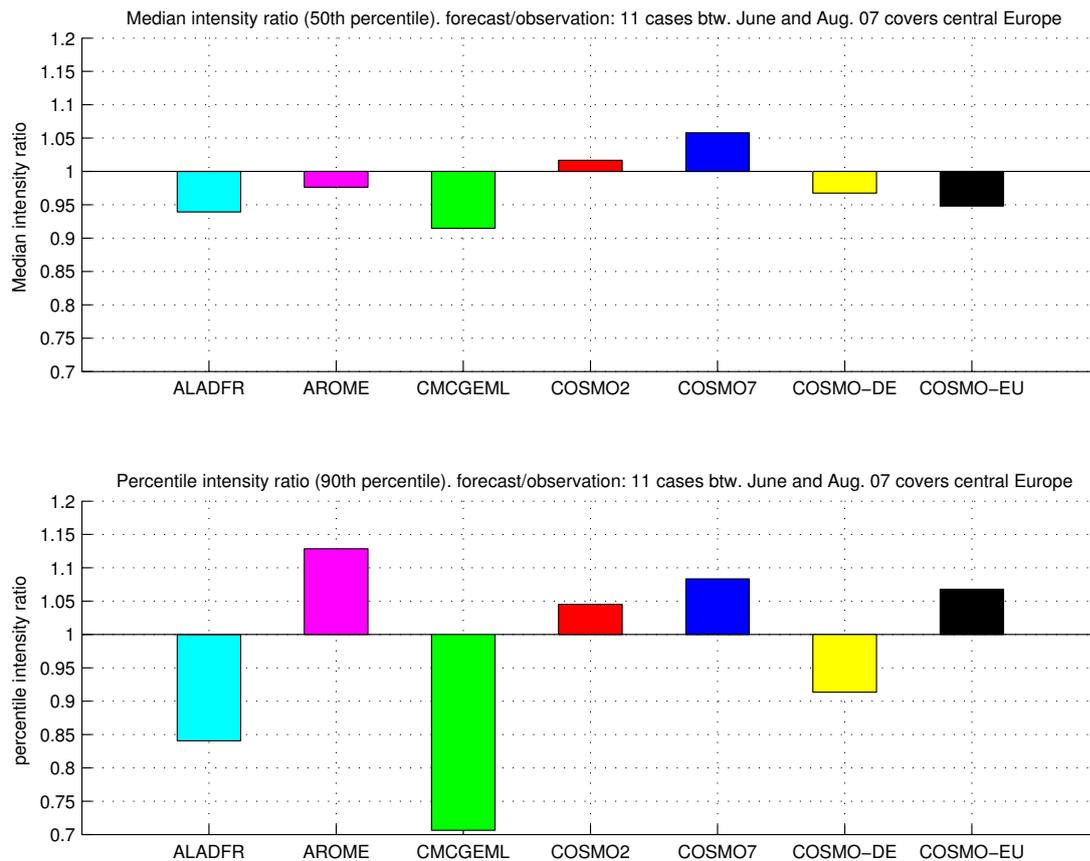


Abbildung 6.5: Vergleich der Niederschlagsverteilung innerhalb der Objekte. Oben: Median Intensitätsverhältnis (50. Perzentil). Unten: Intensitätsverhältnis des 90. Perzentils. Festlegung des Verhältnisses über  $F/O \dots > 1$  wenn  $F > O$ , entspricht dem Fall einer Überprognose. Dargestellt sind jeweils die Mittelwerte aus insgesamt 11 Fällen (18 Ereignissen) zwischen Juni und August 2007 über Mitteleuropa.

einem Punkt der Verteilung in der Weise, dass der jeweilige prozentuelle Anteil der Fälle unterhalb dieses Wertes liegt. Das 50. Perzentil teilt also die Verteilung und gibt den Wert an, an dem 50% der Fälle ober- und unterhalb liegen. Dieser Punkt entspricht dem Median. Ist eine Verteilung nicht Normalverteilt, so unterscheidet sich der Median vom Mittelwert und darf nicht mit diesem verwechselt werden, da zur Bestimmung des Median im Gegensatz zum Mittelwert die Magnitude der Werte eine untergeordnete Rolle spielt. Allgemein, und somit auch im konkreten Fall der Untersuchung von Niederschlagsverteilungen, bedeutet das eine hohe Stabilität des Medians gegenüber Ausreißern, während der Mittelwert wesentlich stärkeren Schwankungen unterliegt. Im Fall der Extremintensität, ausgedrückt über das 90. Perzentil, liegen genau 90% der Werte unterhalb und lediglich 10% über diesem Wert.

Als Vergleichsparameter wird das Verhältnis zwischen den Perzentilwerten der Vorhersage und den Analyseperzentilen als  $F/O$  eingeführt. An dieser Stelle sei angemerkt, dass

die direkte Ausgabe von MODE sehr wohl Intensitätsverhältnisse der verschiedenen Perzentile erzeugt, zur Verhältnisbildung aber der jeweils kleinere durch den größeren Wert geteilt wird. Durch die Einführung einer Richtungsfestlegung über F/O können Abweichungen aber in beide Richtungen zur Differenzierung zwischen Über- und Unterprognosen der Niederschlagsintensität unterschieden und in weiterer Folge verglichen werden.

Abbildung 6.5 zeigt die über 11 Fälle gemittelten Intensitätsverhältnisse der einzelnen Modelle. Beide Teilbilder sind im selben Skalenbereich geplottet und direkt vergleichbar. Bei der Betrachtung des Intensitätsverhältnisses des 50. Perzentils zeigen sich durchgängig deutlich geringere Abweichungen vom Wert des idealen Verhältnisses. Konvektionsauflösende Vorhersagemodelle sind demnach in der Lage, den Median der Intensitätsverteilung von extremen Niederschlagsereignissen auf eine Genauigkeit von weniger als 5% zu schätzen. Die jeweiligen Antriebsmodelle weisen zwar deutlich größere Varianz im Bereich von etwa 12% auf, die Richtung der Abweichung ist aber identisch und deutet auf eine generell ähnliche Arbeitsweise der Modelle innerhalb der Modellgruppen hin. Mit Ausnahme von COSMO7 und COSMO2 wird der Verteilungsmedian eher leicht unterschätzt, die größte Abweichung ist dabei beim Modell des kanadischen Wetterdienstes CMCGEML mit 8% zu vermerken.

Zur Einschätzung der Modellgüte bezüglich der Prognose extremer Intensitäten dient die zweite Darstellung in Abbildung 6.5 des 90. Perzentils. Die allgemein größeren Amplituden der Abweichungen deuten auf eine erhöhte Unsicherheit bei der Vorhersage der Extremintensität hin, zudem ist keine bevorzugte Richtung mehr festzustellen. CMCGEML mit der schwächsten Modellauflösung innerhalb der Gruppe zeigt die signifikanteste Unterschätzung der Extremintensität von knapp 30%, gestützt von den Aussagen der Abbildung 6.4, wodurch infolge der notorischen Unterschätzung insgesamt auch nur vier Objekte für Vergleichszwecke zur Verfügung stehen. Ein weiterer interessanter Aspekt ist die zum Teil sprunghafte Änderung der Abweichungen in die entgegengesetzte Richtung, wie sie etwa bei den Modellen AROME und COSMO-EU auftritt. In beiden Fällen wird der Median der Intensitätsverteilung unterschätzt, der Wert des 90. Perzentils aber deutlich überschätzt. Die Magnitude der Abweichungen wird zur Bestimmung eines Perzentils nicht direkt berücksichtigt. Die Schwankungsbreite der Niederschlagsintensität der oberen 50% des Medians ist durch den Median selbst als untere Grenze beschränkt, nach oben kann aber, bis auf einen physikalisch realistischen Maximalwert, keine feste Schwelle angegeben werden. Die Umkehrung des Vorzeichens in die Richtung einer Überprognose der Extremintensität über das 90. Perzentil kann demnach über zwei Prozesse erklärt werden. Zum Einen über das Auftreten sehr feinskaliger Strukturen und gleichzeitig stark fluktuierender Intensitäten mit zum Teil sehr hohen Raten, zum Anderen aber auch über wesentlich glattere Strukturen aber generell hohen Magnituden. Der erste Fall

einer sehr sprunghaften, feinskaligen Struktur wird von AROME gezeigt. COSMO-EU hingegen weist deutlich weniger Struktur auf und entspricht daher eher dem anderen Fall einer generellen Überschätzung der Extremintensität.

### 6.1.5.1 Einfluss des Analysesystems VERA

Vor allem bei der Beurteilung und Einschätzung der Intensitätsverteilung von Extremereignissen stellt sich die Frage nach dem Auflösungsvermögen des Analysesystems. Die hohe räumliche Variabilität von Niederschlagsereignissen sowie die großen Gradienten der Intensität erfordern ein sehr enges Messnetz um die Verhältnisse richtig abzubilden. Ereignisse auf sehr kleinen Skalen können oft nicht oder zumindest nicht vollständig erfasst werden. Als Konsequenz werden die Felder der Analyse beeinflusst und in weiterer Folge geglättet. VERA mit einem Gitterpunktsabstand von 8 km ist demnach nicht in der Lage, eine Intensitätsverteilung exakt abzubilden. Die größten Auswirkungen des Effekts der Glättung sind also bei der Maximalintensität zu erwarten. Eine gedämpfte Verteilung der Intensität in der Analyse hat allerdings maßgeblichen Einfluss auf die Auswertung von Extremereignissen zur Einschätzung der Prognosegüte über den eben angestellten Vergleich der Niederschlagsverteilungen in Abb. 6.5. Modellseitige Überschätzungen werden daher teilweise zu stark dargestellt, Unterschätzungen zu schwach.

## 7 Zusammenfassung und Ausblick

Die Auswahl und teilweise auch Entwicklung statistischer Verfahrenstechniken zur Realisierung einer vergleichenden Verifikationsstudie bezüglich deterministischer Vorhersagemodelle stellt einen zentralen Aspekt in der vorliegenden Arbeit dar. Der Fokus liegt dabei in Anlehnung an die Zielsetzung der Projekte MAP D-PHASE und COPS in der Auswertung und Beurteilung von Niederschlagsprognosen, speziell in Hinblick auf konvektive Ereignisse. Niederschlag als kategorische Variable erlaubt zudem sowohl eine Stratifikation der Methoden in einen rein qualitativen Zugang zur Auswertung räumlicher Charakteristika, als auch einen rein quantitativen Verifikationszugang zur Evaluierung der Ereignisintensität. Neben konventionellen Techniken wurden auch erst kürzlich entwickelte Methoden wie der objektorientierte Ansatz von MODE oder ein Konzept der Fuzzy-Verifikation implementiert. Der Schwerpunkt der Auswertungen wurde auf mögliche Unterschiede der Modellapplikationen bei der Behandlung von Niederschlagsereignissen in Abhängigkeit von Modellauflösung, der Verwendung von Parametrisierungsschemata, Einfluss der Jahreszeit oder aber der Näherung über die hydrostatische Approximation gelegt.

Die Auswertungen ergeben ein deutlich besseres Abschneiden der konvektionsauflösenden Modelle bezüglich der richtigen qualitativen Erfassung von Niederschlagsereignissen vor allem im Sommer bei vorwiegend konvektiven Auslösungsprozessen. Deutlich wird dieser Zusammenhang bei der Betrachtung der ROC-Kurven innerhalb eines Mittelungszeitraums zwischen Juni und August, wobei Gruppierungen der Modelle mit und ohne Parametrisierungsschema aufgezeigt werden. ALADIN France als einziges hydrostatisches Modell hebt sich hier nochmals ab, zwar mit signifikant höheren Anteilen an eingetroffenen Prognosen, im Gegenzug aber starker Neigung zur flächenmäßigen Überschätzung. Die hochauflösenden Modelle COSMO-DE, COSMO 2 und AROME weisen demgegenüber die geringsten POFD Anteile auf, durch die feineren Strukturen unter Berücksichtigung kleinerer Skalen und den damit verbundenen Schwierigkeiten einer exakten Positionierung der Ereignisse, nehmen aber auch die Werte von POD deutlich ab. Dennoch zeigen diese Modelle eine leichte Erhöhung der Flächenwerte unterhalb der ROC-Kurven. Festzuhalten ist hier, dass sich die Verhältnisse im Herbst beim Übergang zu vermehrt stratiformer Niederschlagsbildung angleichen und in weiterer Folge nicht mehr von einer Bevorzugung der konvektionserlaubenden Modelle ausgegangen werden kann. Erhöht man nun aber

den untersten Schwellwert der Niederschlagsintensität von 0.1 auf 1mm/xh, kommt es zu einer markanten Vergrößerung der Flächenwerte aller Modelle. Dies zeigt die zusätzlichen Unsicherheiten der genauen Einschätzung der Ausdehnung der Ereignisse hinsichtlich sehr geringer Schwellwerte. Im Herbst wird diese Optimierung zusätzlich auch auf höhere Schwellwerte ausgedehnt. Die Einbeziehung von Fuzzy- Methoden gewährt den Prognosen eine gewisse Unschärfe bei der Ereignispositionierung und bringt eine weitere Verbesserung der Maßzahlen in der Größenordnung von einigen Prozent.

Die Favorisierung der hochauflösenden Modelle in Bezug auf eine adäquate räumlich und zeitliche Erfassung von Niederschlagsereignissen, lässt sich über Auswertungen des Modellfehlers (BIAS), vor allem in den Sommermonaten, auch auf eine Bevorzugung dieser Modelle bei der quantitativen Erfassung der Intensitäten ausweiten. Die durchschnittlichen Abweichungen über den gesamten Zeitraum zeigen dabei eine generelle Überschätzung der Intensität mit geringeren Werten bei engmaschigen Modellanwendungen. Eine Analyse der Verteilungen über die Häufigkeitsanteile des Bias führt zudem zu einer Erhöhung der Verteilungsschärfe dieser Modellgruppe.

Der in MET implementierte objektorientierte Verifikationsansatz MODE wurde anhand einer Untersuchung von extremen Niederschlagsereignissen zwischen Juni und August 2007 über Zentraleuropa realisiert und sowohl auf räumliche Differenzen der Objektschwerpunkte, als auch bezüglich der Verteilungsfunktionen der Niederschlagsintensität analysiert. Vor allem bei der Einschätzung des Medians einer Intensitätsverteilung zeigen sich signifikant geringere Abweichungen der Modellanwendungen mit dem direkten Vermögen der Konvektionsauflösung. Die Vorhersage der Extremintensität über das 90. Perzentil der Verteilung hingegen weist auf höhere Unsicherheit bei der Modellierung hin, wobei vor allem das hydrostatische Modell ALADIN sowie die Anwendung mit der schwächsten Auflösung, CMCGEML, die Maximalintensität am deutlichsten unterschätzen. Die räumliche Auswertung der Schwerpunksdifferenzen ergibt in diesem Zusammenhang jedoch keine unmittelbare Verbesserung für feinskalige Modelle.

Die Integration von MET zur Beantwortung zahlreicher Problemstellungen der Modellverifikation in der vorliegenden Arbeit eröffnet neben der Möglichkeit der direkten Auswertung unzähliger Maßzahlen auch einfache Zugänge zu komplexeren Auswertungen und neueren Verfahrenstechniken wie MODE oder Ansätzen über die Aufspaltung von Niederschlagsfeldern in ihre einzelnen Skalen über eine *wavelet*- Zerlegung. Diese könnten etwa in zukünftigen und weiterführenden Arbeiten eingesetzt werden. Ebenso ermöglicht die vorangegangene Aufbereitung der Datensätze und Anpassung an die Erfordernisse von MET eine einfache Ausdehnung der Verfahren auf weitere meteorologische Parameter, die zwar zum Teil schon erarbeitet, durch die Fokussierung auf Niederschlagsereignisse hier jedoch nicht näher diskutiert wurden.

# Literaturverzeichnis

- [1] WMO: A guide to the code form FM 92-IX Ext. GRIB Edition 1. 1994 (17). – Forschungsbericht. – WMO TD-No.611
- [2] DEUTSCHER WETTERDIENST, OFFENBACH: A Description of the Nonhydrostatic Regional COSMO-Model. 2008. – User’s Guide. – Part IIV
- [3] ADRIAN, G. ; FRÜHWALD, D.: Design der neuen Modellkette GME/LM. In: *promet* (2002), Juni, Nr. 3/4, S. 106 – 110. – Jahrg. 27
- [4] ALEXRK2: Europe relief laea location map. In: *Wikipedia* (2010), März. [http://de.wikipedia.org/wiki/Datei:Europe\\_relief\\_laea\\_location\\_map.jpg](http://de.wikipedia.org/wiki/Datei:Europe_relief_laea_location_map.jpg). – 24.3.2011
- [5] AMODEI, M. ; STEIN, J.: Deterministic and fuzzy verification methods for a hierarchy of numerical models. In: *Meteor. Appl.* 16 (2009), S. 191 – 203
- [6] BALDAUF, M. u. a.: Kurze Beschreibung des Lokal-Modells Kürzestfrist LMK und seiner Datenbanken des DWD, Version 1.0 / Deutscher Wetterdienst, Offenbach. 2006 (1.0). – Forschungsbericht. – Stand 16.8.2006
- [7] BAZILE, E.: dphase -aladfr: ALADIN-France operational model forecasts run by Meteo-France for the MAP D-PHASE project / World Data Center for Climate. Version: 2009. [http://dx.doi.org/10.1594/WDC/dphase\\_aladfr](http://dx.doi.org/10.1594/WDC/dphase_aladfr). 2009. – Forschungsbericht. – 13.9.2010
- [8] BOUTTIER, F. u. a.: Progress of the AROME mesoscale NWP project. In: *Research activities in atmospheric and oceanic modeling* 36 (2006), S. 11–12. – Edited by J. Cote
- [9] BOUTTIER, F.: The Meteo-France NWP system: description, recent changes and plans. In: *GMAP: Numerical Weather Prediction Research Group* (update April 2010). <http://www.cnrm.meteo.fr/gmap/>. – 15.1.2011
- [10] BROWN, B.G. u. a.: Application of the MODE object-based verification tool for the evaluation of model precipitation fields. In: *AMS 22nd Conference on Weather*

*Analysis and Forecasting und 18th Conference on Numerical Weather Prediction.*  
Park City, Utah, 25-29 June 2007

- [11] BUBNOVA, R. u. a.: International project ARPEGE/ALADIN. In: *EWGLAM Newsletter* 22 (1993), S. 117 – 130
- [12] CASATI, B. u. a.: A new intensity-scale approach for the verification of spatial precipitation forecasts. In: *Meteor. Appl.* 11 (2004), S. 141–154
- [13] CASE, J. L. u. a.: An objective technique for verifying sea breezes in high-resolution numerical weather prediction models. In: *Weather and Forecasting* 19 (2004), S. 690–705
- [14] DAVIS, C.A. u. a.: Object-based verification of precipitation forecasts, Part I: Methodology and application to mesoscale rain areas. In: *Monthly Weather Review* 134 (2006a), S. 1772–1784
- [15] DAVIS, C.A. u. a.: Object-based verification of precipitation forecasts, Part II: Application to convective rain systems. In: *Monthly Weather Review* 134 (2006b), S. 1785–1795
- [16] EBERT, E.: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. In: *Meteorol. Appl.* 15 (2008), S. 51–64
- [17] EBERT, E. ; MCBRIDE, J. L.: Verification of precipitation in weather systems: Determination of systematic errors. In: *J. Hydrol.* 239 (2000), S. 179–202
- [18] FAWCETT, T.: An introduction to ROC analysis. In: *Pattern Recognition Letters* 27 8 (2006), S. 861–874
- [19] FISCHER, C. u. a.: An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. In: *Q. J. R. Meteorol. Soc* (2006), Nr. 131, S. 3477–3492
- [20] HAGEDORN, R.: On the relative benefits of TIGGE multi-model forecasts and reforecast-calibrated EPS forecasts. In: *ECMWF Newsletter* 124 (2010), S. 17 – 23
- [21] HAIMBERGER, L.: Vorlesung Numerische Methoden der Wettervorhersage - Teil I Grundlagen. (2006), S. 79. – Institut für Meteorologie und Geophysik, Universität Wien - unveröffentlicht

- 
- [22] KIESENHOFER, S.: *Verifikation hochaufgelöster numerischer Wettervorhersagemodelle mit MET und VERA im Rahmen von MAP D-PHASE; Technische Beschreibung*. – Institut für Meteorologie und Geophysik, Universität Wien - unveröffentlicht
- [23] LILJEQUIST, G. H. ; CEHAK, K.: *Allgemeine Meteorologie*. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 1984. – 3. Auflage
- [24] MALARDEL, S. u. a.: Representation of clouds in AROME. (2006), 13-15 November. – ECMWF Workshop on parameterization of clouds in large-scale models; 13-15 November 2006
- [25] MASON, S. J. ; GRAHAM, N. E.: Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels. In: *Weather and Forecasting* 14 (1999), S. 713–725
- [26] MAYER, D.: *Transformationsformeln für die Vera-Tangentialebene, VERA-Notiz Nr.75, Institut für Meteorologie und Geophysik*. 2010
- [27] MCTAGGART-COWAN, R. u. a.: MAP D-PHASE - High Resolution Guidance in Steep Terrain. (2007). <http://www.atmos.albany.edu/facstaff/rmctc/>. – 25.9.2010
- [28] MCTAGGART-COWAN, R. u. a.: Development and tropical transition of an Alpine lee cyclone. Part I: Case analysis and evaluation of numerical guidance. In: *Monthly Weather Review* 138 (2009), S. 2281–2307
- [29] MET USER'S GUIDE: *Model Evaluation Tools Version 2.0 (METv2.0) User's Guide*, March 2009. [http://www.dtcenter.org/met/users/docs/users\\_guide/MET\\_Users\\_Guide\\_v2.0\\_rev2.pdf](http://www.dtcenter.org/met/users/docs/users_guide/MET_Users_Guide_v2.0_rev2.pdf). – 11.10.2010
- [30] METEOSWISS (Hrsg.): *Operational implementation of COSMO-7 and COSMO-2 at MeteoSwiss*. MeteoSwiss, <http://cosmo-model.cscs.ch/>. – 23.11.2010
- [31] MITTERLMAIER, M. ; NIGEL, R.: Intercomparison of Spatial Verification Methods: Identifying Skillful Spatial Scales Using the Fractional Skill Score. In: *Weather and Forecasting* 25 (2010), S. 343–354
- [32] RENKEN, A. L.: *Analyse der REMO-Simulationsdaten zum Klimawandel in Schleswig-Holstein und Anwendung der Klimaszenarien auf das Kielstau- Einzugsgebiet mit dem öko-hydrologischen Modell SWAT*, Christian-Albrechts-Universität zu Kiel, Diplomarbeit, 2009
- [33] REUTER, H. u. a. ; RAIT, Wilhelm (Hrsg.): *Lehrbuch der Experimentalphysik, Band 7, Erde und Planeten*. Berlin : L. Bergmann and C. Schaefer, 1997. – 132–310 S.

- [34] ROTACH, M. W. u. a.: MAP D-PHASE: Real-Time Demonstration of Weather Forecast Quality in the Alpine Region. In: *BAMS* 90 (2009), S. 1321 – 1336
- [35] SCHULZ, J. P. ; SCHATTLER, U.: Kurze Beschreibung des Lokal-Modells Europa COSMO-EU (LME) und seiner Datenbanken auf dem Datenserver des DWD / Deutscher Wetterdienst, Offenbach. 2009. – Forschungsbericht. – Stand 15.1.2009
- [36] SEITY, Y.: dphase -arome: AROME 30hours forecasts run by Meteo-France CNRM/GAME for the MAP D-PHASE project / World Data Center for Climate. Version: 2009. [http://dx.doi.org/10.1594/WDC/dphase\\_arome](http://dx.doi.org/10.1594/WDC/dphase_arome). 2009. – Forschungsbericht. – 13.9.2010
- [37] SKAMAROCK, W. C.: Evaluating mesoscale NWP models using kinetic energy spectra. In: *Monthly Weather Review* 132 (2004), S. 3019–3032
- [38] SMITH, B. B. ; MULLEN, S. L.: An evaluation of sea level cyclone forecasts produced by NMC's nested-grid model and global spectral model. In: *Weather and Forecasting* 8 (1993), S. 37–56
- [39] STANSKI, H. R. u. a.: Survey of common verification methods in Meteorology. In: *WMO World Weather Watch Technical Report No. 8, WMO/TD No. 358 2* (1989), July
- [40] STEINACKER, R. u. a.: A Mesoscale Data Analysis and Downscaling Method over Complex Terrain. In: *Monthly Weather Review* 134 (2006), S. 2758 – 2771
- [41] SUKLITSCH, M.: *Objektive Analyse der Temperatur der Niederungen*, Institut für Meteorologie und Geophysik, Universität Wien, Diplomarbeit, 2004
- [42] THORPE, W.: A guide to the WMO code form FM 94 BUFR / Fleet Numerical Meteorology and Oceanography Center, Monterey, California 93943 U.S.A. Monterey, California 93943 U.S.A, . – Forschungsbericht
- [43] WERNLI, H. u. a.: SAL - A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. In: *Monthly Weather Review* 136 (2007), S. 4470 – 4487
- [44] WULFMAYER, V. u. a.: The convective and orographically induced precipitation study. A Research and Development Project of the World Weather Research Program for Improving Quantitative Precipitation Forecasting in Low-Mountain Regions. In: *BAMS* 89 (2008), S. 1477 – 1486
- [45] YESSAD, K.: BASICS ABOUT ARPEGE/IFS, ALADIN AND AROME IN THE CYCLE 37 OF ARPEGE/IFS. In: *METEO-FRANCE/CNRM/GMAP/ALGO* (24.

November 2010). <http://www.cnrm.meteo.fr/gmapdoc/IMG/pdf/ykarpbasics37.pdf>. – 13.4.2011

# Danksagung

Ich bedanke mich bei Ass. Prof. Mag. Dr. Manfred Dorninger für die Mitarbeit im Projekt VERITA und die unzähligen Besprechungen innerhalb unserer Arbeitsgruppe mit meinen Kollegen Mag. Theresa Gorgas und Sarah Umdasch. Danken möchte ich auch Mag. Niko Filipovic und Mag. Stefan Sperka für die Unterstützung in technischen Fragen vor allem bei der Programmierung mit Matlab sowie für die aufgewendete Zeit für Diskussionen und Erklärungen.

Ganz besonder möchte ich aber meiner Frau Michaela für die mentale Unterstützung während dieser Zeit danken.

# Lebenslauf

## Persönliche Daten:

Name: Stefan Kiesenhofer  
Geburtsdatum: 2.8.1982  
Geburtsort: Linz  
Staatsbürgerschaft: Österreich  
Familienstand: verheiratet  
Ehefrau: Michaela Kiesenhofer, Flugbegleiterin  
Eltern: Elisabeth Kiesenhofer, ehem. Angestellte  
Rudolf Kiesenhofer, ehem. Angestellter  
Bruder: Markus, Student

## Ausbildung:

1993 – 1997 Hauptschule Freistadt „Marianum“  
1997 – 2002 HTBLA Eisenstadt, Maschinenbau/Ausbildungszweig  
Flugtechnik  
Juni 2002 Abschluss mit Matura  
2003 – 2006 Studium der Meteorologie und Geophysik, Universität  
Wien, erster Studienabschnitt  
2006 – 2011 Studium der Meteorologie und Geophysik, Universität  
Wien, zweiter Studienabschnitt

## Beruflicher Werdegang:

August 1998 Ferialpraxis Fa. Mandl & Berger, Abt. Werkzeugbau  
August 1999 Ferialanstellung Fa. Mandl & Berger, Abt. Werkzeugbau  
Juli 2001 Ferialpraxis Fa. Air Linz, Luftfahrzeugwartung  
Juni – Sept. 2003 Ferialanstellung Flughafen Linz/Hörsching,  
Flugbetriebsinstandhaltung  
August 2008 Ferialpraxis ORF Fernseh-Wetterredaktion  
Februar 2010 Ferialpraxis Austro Control, Abteilung MET  
Juni 2009 – Dez. 2010 Projektmitarbeit am Institut für Meteorologie und  
Geophysik, Projekt „VERITA“