



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Eignung räumlicher Verifikationsmethoden für die Verifikation
verschiedener meteorologischer Parameter einer WRF
Ensemblevorherage mit den Analysedaten des VERA Systems in
den Wintermonaten 2016/17

verfasst von / submitted by

Judith Svacina, BSc

angestrebter akademischer Grad/in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 614

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Meteorologie

Betreut von / Supervisor:

Ass.-Prof. Mag. Dr. Manfred Doringner

Abstrakt

Die Nachfrage nach spezifischen Wettervorhersagen steigt immer stärker an. Kunden fordern detaillierte Vorhersagen beispielsweise für Veranstaltungen, Firmenstandorte, Windparks oder Ernteeinfuhr. Um ein Wettervorhersagemodell und somit dessen Prognosen zu verbessern, muss das Modell überprüft und gegebenenfalls angepasst werden. Ein wichtiger Schritt ist die Verifikation des Modells. Durch die feine Auflösung moderner Vorhersagemodelle erhöht sich die räumliche Variabilität der Prognose. Durch kleine räumliche Veränderungen wird das Modell bei traditionellen Verifikationsmethoden auf einem Gitterpunkt häufig bestraft (double penalty). Um die tatsächliche Qualität der Modelle zu testen, werden räumliche Verifikationsmethoden angewendet, wodurch die Umgebung bzw. die gesamte Domäne mitberücksichtigt wird.

Der zentrale Aspekt dieser Arbeit ist die Überprüfung der Eignung räumlicher Verifikationsmethoden (Upscaling, Fractions Skill Score, Structure-Amplitude-Location (SAL) und Displacement and Amplitude Score (DAS)) bei der Betrachtung kontinuierlicher meteorologischer Variablen. Das wird mit statistischen Kenngrößen, wie der Häufigkeitsverteilung oder dem Bias bestimmt. Es hat sich gezeigt, dass die Windgeschwindigkeit mit den Neighborhood Methoden (Upscaling, Fractions Skill Score) verifiziert werden kann. Das liegt vor allem daran, dass die Windgeschwindigkeit ähnliche Eigenschaften wie der Niederschlag besitzt. Das ist vorteilhaft, da viele räumliche Verifikationsmethoden speziell für Niederschlagsereignisse entwickelt wurden. Einer der entscheidenden Punkte bei der räumlichen Verifikation (speziell für SAL, DAS) ist die Definition von Objekten. Die Objekte lassen sich so wählen, dass sie Tiefdruckgebiete, Fronten oder Gebiete höherer Windgeschwindigkeit, darstellen. Vor allem in kleinen Domänen ist es besonders wichtig, dass die Vorhersage- und Analyseobjekte nicht zu groß werden. Je kleiner der Ausschnitt der Domäne und je größer das Objekt ist, desto schwieriger ist es, die tatsächliche Qualität des Vorhersagemodells zu beschreiben. Im Extremfall, wenn das Objekt die gesamte Domäne bedeckt, ergibt sich bei der Anwendung von SAL eine Mittelung oder ein sehr kleiner Amplitudenfehler (DAS).

Abstract

Customers demand detailed forecasts for events, company locations, wind farms or harvest import. In order to improve a weather forecast model and thus its forecast, the model must be checked and, if necessary, adapted. An important step is the verification of the model. The fine resolution of modern prediction models increases the spatial variability of the forecast. However, the model is often penalized (double penalty) in traditional verification methods. In order to test the actual quality of the models, spatial verification methods are applied.

The central aspect of this work is to check how suitable the spatial verification methods like Upscaling, Fractions Skill Score, Structure Amplitude-Location (SAL) and Displacement and Amplitude Score (DAS) are when continuous meteorological variables are considered. This is determined via statistical parameters, e.g. the frequency distribution or bias. It is shown that the wind speed can be verified well with the Neighborhood Methods (Upscaling, Fractions Skill Score). This is mainly due to the fact that wind speed has similar properties to precipitation. This is important because many spatial verification methods have been developed specifically for precipitation events. One of the key points in spatial verification (especially for SAL, DAS) is the definition of objects. Objects can be selected to represent low pressure areas, fronts or areas of higher wind speed. Especially in small domains it is very important that the objects of the forecast and analysis do not become too large. The smaller the domain and the larger the object, the more difficult it is to describe the actual quality of the weather prediction model. If the object covers the entire domain, this results in an averaging for the SAL or a very small amplitude error (DAS).

Danksagung

Ich möchte mich herzlich bei meinem Betreuer Ass.-Prof. Mag. Dr. Manfred Dorninger bedanken. Vielen Dank für die anregenden Diskussionen und hilfreichen Tipps, die mir bei der Erstellung der Masterarbeit halfen.

Mein besonderer Dank gilt meiner Familie, insbesondere meinen Eltern, die mir mein Studium ermöglicht und mich in all meinen Entscheidungen stets unterstützt haben. Ein großer Dank geht an meinen Freund, der immer für mich da war und mich motiviert hat.

An dieser Stelle möchte ich mich auch bei meinen Studienkollegen bedanken. Der Zusammenhalt und die gegenseitige Unterstützung und Motivation während des Studiums, haben die Zeit sehr besonders gemacht.

Inhaltsverzeichnis

1	Einleitung	4
1.1	Motivation und Fragestellung	4
1.2	Räumliche Verifikation	5
2	Daten	12
2.1	Modelldaten	12
2.2	Vienna Enhanced Resolution Analysis (VERA)	14
2.3	Datenaufbereitung	15
3	Ausgewählte Verifikationsmethoden	18
3.1	Upscaling	18
3.2	Fractions Skill Score (FSS)	20
3.3	Structure-Amplitude-Location (SAL)	22
3.4	Displacement and Amplitude Score	26
3.5	Software	27
4	Ergebnisse und Diskussion	29
4.1	Datenanalyse	29
4.2	Upscaling	39
4.3	Fractions Skill Score	43
4.4	Structure-Amplitude-Location	45
4.5	Displacement and Amplitude Score	59
5	Zusammenfassung	66
	Literaturverzeichnis	68

Tabellenverzeichnis

4.1	DAS des Vorhersagelaufs vom 10.01.2017 00 UTC für Domäne d01 und d02	63
-----	--	----

Abbildungsverzeichnis

1.1	Vorhergesagte und beobachtete Position von Wetterereignissen	7
1.2	RMSE bei zunehmender räumlichen Skala	8
1.3	Entwicklung des Prognosefehlers	9
1.4	Kategorisierung der räumlichen Verifikationsmethoden	9
2.1	Domänen des WRF Vorhersagemodells	13
2.2	Analysedaten des VERA Systems der potentiellen Temperatur	14
2.3	Gewichtsfunktion Cressman-Interpolation	16
2.4	Interpolation mit diversen Cressmanradien	17
3.1	Prinzip des Upscalings	19
3.2	Kontingenztabelle	19
3.3	Prinzip des Fractions Skill Scores	21
3.4	Fractions Skill Score und Skillful Scale	22
3.5	SAL-Diagramm	25
4.1	Bias der Windgeschwindigkeit vom 11/2016 bis 03/2017	30
4.2	Bias der Windgeschwindigkeit vom 11/2016 bis 03/2017	31
4.3	Equitable Threat Score der gemittelten Windgeschwindigkeit vom 11/2016 bis 03/2017	32
4.4	Mittelwert der Windgeschwindigkeit verschiedener Fallbeispiele	35
4.5	Großwetterlage: GFS Analyse für den 11.01.2017 06 UTC	36

4.6	Mittelwert Domäne d02 Niederschlag, Luftdruck, Äquivalentpotentielle Temperatur, Mischungsverhältnis vom 10.01.2017 00 UTC	37
4.7	Relative Häufigkeitsverteilung der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC	38
4.8	Equitable Threat Score für die Windgeschwindigkeit vom 10.01.2017 00 + 30-33 h (11.01. 06-09 UTC)	40
4.9	Differenz des ETS beider Domänen für die Windgeschwindigkeit vom 10.01.2017 00 +30-33 h (11.01. 06-09 UTC)	41
4.10	Frequency Bias der Windgeschwindigkeit vom 10.01.2017 00 +30-33 h (11.01. 06-09 UTC)	42
4.11	Fractions Skill Score der Windgeschwindigkeit vom 10.01.2017 00 +30-33 h (11.01. 06-09 UTC	44
4.12	SAL: Niederschlag des Vorhersagelaufs vom 10.01.2017 00 UTC	46
4.13	Objekte der Windgeschwindigkeit vom 10.01.2017 +38 h	48
4.14	SAL: Struktur der Windgeschwindigkeit	49
4.15	SAL: Amplitude der Windgeschwindigkeit	50
4.16	SAL: Räumliche Verschiebung der Windgeschwindigkeit	51
4.17	SAL-Diagramm Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 (28-35 h)	53
4.18	Objekte der äquivalentpotentiellen Temperatur vom 10.01.2017 + 38 h	54
4.19	Objekte des reduzierten Luftdrucks vom 10.01.2017 + 38 h	55
4.20	SAL: Äquivalentpotentielle Temperatur des Vorhersagelaufs vom 10.01.2017 00 UTC	56
4.21	SAL: Reduzierter Luftdruck des Vorhersagelaufs vom 10.01.2017 00 UTC	57
4.22	DAS: Phasen der Berechnung des Vorhersagelaufs vom 10.01.2017 00 +38 h	60
4.23	DAS Windgeschwindigkeit > 3 m/s des Vorhersagelaufs vom 10.01.2017 00 UTC	61
4.24	DAS Windgeschwindigkeit > 5 m/s des Vorhersagelaufs vom 10.01.2017 00 UTC	62
4.25	DAS Diagramm der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC	64

Kapitel 1

Einleitung

1.1 Motivation und Fragestellung

Hochauflösende Wettervorhersagemodelle produzieren Prognosen, welche Wetterereignisse detailliert repräsentieren (Ebert, 2009). Mit der feineren Auflösung von Vorhersagemodellen tritt auch eine höhere räumliche Variabilität innerhalb der Prognose auf Gilleland et al. (2009). Das Modell kann das Wetterereignis zwar richtig entwickeln, allerdings wird es in Raum und Zeit oft nicht richtig positioniert (Gofa et al., 2018). Folglich wird die tatsächliche Qualität dieser Vorhersagen durch punktspezifische Verifikationsverfahren oft nicht richtig dargestellt. Der Verschiebungsfehler führt dazu, dass das Modell häufig doppelt bestraft wird (double penalty) (Gilleland et al., 2009). Herkömmliche Verifikationsverfahren berücksichtigen nicht, dass das Wetterereignis von benachbarten Gitterpunkten abhängig ist und als zusammenhängendes Objekt behandelt werden muss. Räumliche Verifikationsmethoden, welche ereignisbasiert oder objektorientiert sind, helfen die tatsächliche Qualität von hochaufgelösten numerischen Wettervorhersagemodellen aufzuzeigen (Ebert (2009), Casati (2009)).

Diese Masterarbeit überprüft die Eignung von räumlichen Verifikationsmethoden auf verschiedene Parameter in einer kleinen Domäne. In den Fokus rückt vor allem die Windgeschwindigkeit des 10m Windes. Weiters werden auch die äquivalentpotentielle Temperatur, der reduzierte Luftdruck und das Mischungsverhältnis betrachtet. Die räumlichen Verifikationsmethoden werden zusätzlich an Ensemblevorhersagen getestet. Außerdem werden Möglichkeiten zur Darstellung der Ergebnisse gesucht.

Eine Problematik ist, dass viele Methoden (vgl. Kapitel 3) verlangen, dass der Parameter, den man verifizieren möchte, als ein klar definiertes Objekt vorliegen muss (Wernli et al. (2008), Keil and Craig (2009)). Der Niederschlag ist eine positiv semidefinite Größe und kann daher Nullwerte annehmen. Dadurch bilden alle Werte größer als Null zusammen ein oder mehrere natürliche Objekte. Die Windgeschwindigkeit besitzt die gleiche Eigenschaft der positiven Semidefinitheit, allerdings treten Kalmen eher selten auf. Die Windgeschwindigkeit

kann nicht als ein abgeschlossenes Objekt verstanden werden, denn das Modell liefert Werte größer als Null auf (fast) jedem Gitterpunkt innerhalb der Domäne. Andere kontinuierliche Variablen wie die Temperatur oder der Luftdruck sind keine abgeschlossenen Objekte, da die Werte in ihrer SI-Basiseinheit immer größer als Null sind. Ein weiterer wichtiger Punkt ist das Verhältnis zwischen der Größe des Objekts und der Domäne. Zusätzlich hängt die Vorhersagequalität davon ab, wie genau die Modelle die räumlichen Strukturen von Wetterereignissen wiedergeben können (Keil and Craig, 2007).

Räumliche Verifikationsmethoden lassen sich gut auf deterministische Vorhersagen anwenden. Die Vorhersagedaten für diese Masterarbeit basieren auf einer Ensemblevorhersage (vgl. Kapitel 2). Betrachtet man die Ensemblevorhersage als eine Sammlung deterministischer Vorhersagen, kann jeder der Member und der Mittelwert mittels räumlichen Verifikationsmethoden verifiziert werden (Jolliffe and Stephenson, 2012). Bei einer hohen Anzahl an Ensemblemember ist es für die Interpretation umso wichtiger, wie die Ergebnisse aufbereitet werden. Für die Verifikation von Ensemblevorhersagen als Gesamtheit gibt es noch nicht sehr viele Ansätze. Eine Möglichkeit liefert die Ensembleversion des SAL (vgl. Kapitel 3.3) nach Radanovics et al. (2018).

Die verwendete Ensemblevorhersage wurde für die Vorhersage von Vereisung an Windkraftanlagen generiert (vgl. Kapitel 2). Dabei liegt der Fokus auf für Vereisung relevanter Parameter wie Feuchte, Temperatur oder Wassergehalt auf Nabenhöhe der Windkraftanlagen. Die Ensemblemember werden zusätzlich aus einem Satz verschiedener Parametrisierungen angetrieben.

1.2 Räumliche Verifikation

Numerische Wettervorhersagemodelle sind in der heutigen Meteorologie nicht mehr wegzudenken. Durch die stetig steigende Rechenleistung der Computersysteme verbesserte sich auch die Qualität der sogenannten NWP (numerical weather prediction) Modelle und somit auch die Modelle kontinuierlich (Shuman, 1989). Der Fokus der Vorhersage lag zunächst auf dem Druckfeld, heutzutage werden bereits kleinräumige Ereignisse wie Gewitterzellen vorhergesagt. Der Fortschritt liegt vor allem in der Datenassimilation, die durch den Ausbau des Beobachtungsnetzwerk sowie der Entwicklung neuer Methoden (z.B. 4D Var Datenassimilation) verbessert wurde (Tsuyuki and Miyoshi, 2007). Weitere Gründe sind die Verbesserung der Parametrisierungsansätze und die Reduktion der horizontalen Gitterpunktabstände auf dem Vorhersagegitter (vgl. Shuman (1989)). Gegenwärtig besitzen viele numerische Wettervorhersagemodelle eine sehr hohe räumliche Auflösung im km-Bereich und prognostizieren Wetterereignisse auf feineren Skalen mit geringer Vorhersagbarkeit (Jolliffe and Stephenson, 2012). Die meisten NWP Modelle produzieren räumliche Felder, deren Merkmale als Objekte interpretiert werden können. Häufig korrelieren alle Werte auf den Gitterpunkte innerhalb der Domäne (Marzban and Sandgathe, 2006). Wetterereignisse, welche nicht durch das Modellgitter aufgelöst werden können, da sie zu feinskalig sind, werden durch sogenannte Parametrisierungen modelliert. Modelle mit hoher räumlichen

Auflösung erscheinen sehr nützlich, da diese besondere Merkmale besser darstellen (Keil and Craig, 2009). Allerdings führt ein feineres Modellgitter nicht zwangsläufig zu einer signifikant besseren Prognosen auf einem einzelnen Gitterpunkt (Done et al., 2004).

Um die Qualität der Prognosen zu bestimmen, wird die Vorhersage mit Beobachtungs- oder Analysedaten verglichen und verifiziert. Ein sogenanntes traditionelles Verifikationsmaß vergleicht kontinuierliche Vorhersagedaten wie Temperatur, Wind oder Luftdruck (Jolliffe and Stephenson, 2012) mit Beobachtungen auf einem Gitterpunkt, die in Raum und Zeit übereinstimmen (Ebert, 2008). Beobachtungen und Analysedaten fungieren gleichermaßen als mögliche Vergleichsdaten. In dieser Masterarbeit werden Analysedaten des Vienna Enhanced Analysis (VERA) Systems verwendet (vgl. Kapitel 2.2).

Ein Beispiel für ein traditionelles Verifikationsmaß für die Vorhersagegüte ist der Root-Mean-Squared-Error (RMSE). Es handelt sich um ein quadratisches Fehlermaß, bei dem große Fehler stärker gewichtet werden (Stanski et al., 1989). Auf Grund seiner Einfachheit wird der RMSE häufig in diversen Wettervorhersagezentren angewendet (Wernli et al., 2008). Das Maß ist wie folgt definiert (Jolliffe and Stephenson, 2012):

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (X_{i,fcst} - X_{i,obs})^2} \quad (1.1)$$

wobei $X_{i,fcst}$ der vorhergesagte und $X_{i,obs}$ der beobachtete (oder analysierte) Wert ist. M ist die Anzahl der Gitterpunkte. Je kleiner das Maß wird, desto besser ist das Ergebnis. Eine Vorhersage ist perfekt, wenn der RMSE gleich Null ist.

Mit zunehmender räumlicher Auflösung numerischer Wettervorhersagemodelle werden die aufgelösten Skalen kleiner und es ist schwieriger eine punktgenaue Prognose zu erhalten (Ebert (2009), Keil and Craig (2009)). Traditionelle Verifikationsmaße beurteilen daher eine Vorhersage auf feinem Gitter schlechter als eine Vorhersage auf einem gröberen Gitter, auch wenn das hochaufgelöste Modell subjektiv besser abschneidet (Dorninger et al., 2018). Es kann beispielsweise vorkommen, dass Niederschlagsobjekte mit der richtigen Intensität und Struktur für einen bestimmten Gitterpunkt vorhergesagt, die Position des Objekts allerdings versetzt beobachtet wird (Wernli et al., 2009), siehe Abbildung 1.1. Die traditionellen Verifikationsmethoden geben keine Information darüber, ob die Vorhersage und Beobachtung lediglich räumlich verschoben aufgetreten sind (Jolliffe and Stephenson, 2012). Bei der Verifikation eines Gitterpunkts wird ein Verschiebungsfehler gleich behandelt wie jeder andere Fehler in der Vorhersage. Allerdings wird das Modell auf Grund des Verschiebungsfehlers doppelt bestraft, was als *double penalty* Problem bezeichnet wird (Gilleland et al., 2009). Erstens wird das Modell durch die Verifikation schlechter bewertet, da es an dem betrachteten Gitterpunkt Niederschlag vorhergesagt hat, dieser aber nicht eingetreten ist. In der Verifikation wird dieser Fall als *false alarm* bezeichnet. Zusätzlich wird das Modell schlechter evaluiert, weil Niederschlag an einem benachbarten Gitterpunkt beobachtet wird (*missed event*) (Wilks, 2011). Das double penalty Problem tritt hauptsächlich bei der Verifikation von hochaufgelösten Wettervorhersagemodellen auf (Jolliffe and Stephenson, 2012), was die

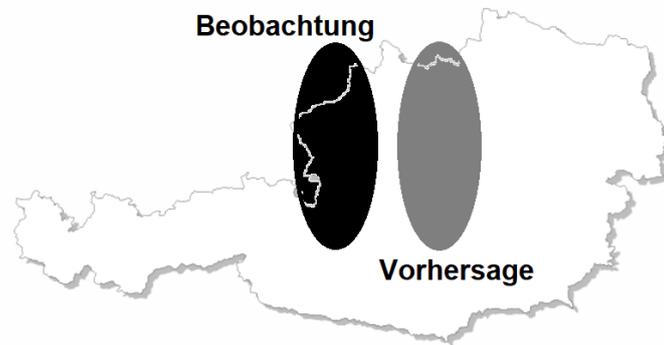


Abbildung 1.1: Schematische Darstellung von Beobachtung und Vorhersage eines Wetterereignisses mit der gleichen Form, aber räumlich versetzt. Der Umriss für die Österreichkarte stammt von www.stepmap.de

Interpretation der Verifikationsergebnisse durch herkömmliche Verfahren erschwert, welche die räumliche und zeitliche Abhängigkeit der beobachteten Wetterereignissen zwischen den Gitterpunkten nicht berücksichtigen (Marzban and Sandgathe (2006), Keil and Craig (2007)). Die schlechteren Ergebnisse der Verifikation stellen demnach nicht die tatsächliche Qualität der hochauflösten Modelle dar (Ebert, 2009) und weisen auf eine Limitierung der Verifikation auf einem Gitterpunkt hin (Keil and Craig, 2009).

Um dieses Problem zu vermeiden oder zumindest zu minimieren, werden räumliche Verifikationsmaße oder objektorientierte Methoden verwendet. In den vergangenen Jahren wurden einige Verifikationsmethoden entwickelt, um hochauflösten numerische Wettervorhersagemodelle besser verifizieren und evaluieren zu können. Einen Überblick über die Möglichkeiten findet man in Gilleland et al. (2009) oder Ebert (2009)). Der Vorteil besteht darin, dass viele dieser neuen Methoden keine genaue Übereinstimmung zwischen Vorhersage- und Analysedaten benötigen. Sie betrachten die Umgebung der Gitterpunkte mit oder berücksichtigen die gesamte Domäne und verifizieren das Wetterereignis anhand von Merkmalen (Keil and Craig, 2009), wie der Form des Objekts, der Intensität oder der räumlichen Position.

Abbildung 1.2 zeigt, die Modellvorhersage verifiziert besser, je mehr Gitterpunkte in die Berechnung miteinfließen. Zunächst werden die Werte auf den betrachteten Gitterpunkten gemittelt und anschließend der RMSE berechnet. Neben der Mittelung über mehrere Gitterpunkte gibt es auch andere Vorgehensweisen (vgl. Kapitel 3). In der Abbildung entspricht die Verifikation auf einem einzelnen Gitterpunkt einer traditionellen Punkt-Verifikation und man erhält einen (relativ) hohen RMSE Wert. Bei der Zunahme der räumlichen Skala nimmt der RMSE ab, was für eine bessere Vorhersage spricht. Werden alle Gitterpunkte in einer Domäne berücksichtigt, wird der RMSE des Mittelwerts berechnet. Die Bewertung durch den RMSE alleine ist nicht zufriedenstellend, da er das double-penalty-Problem nicht berücksichtigt. Mit zunehmender Vorhersagezeit nehmen Fehler in der Vorhersage auf Grund von Unsicherheiten in den Eingangsdaten und durch Fehler in

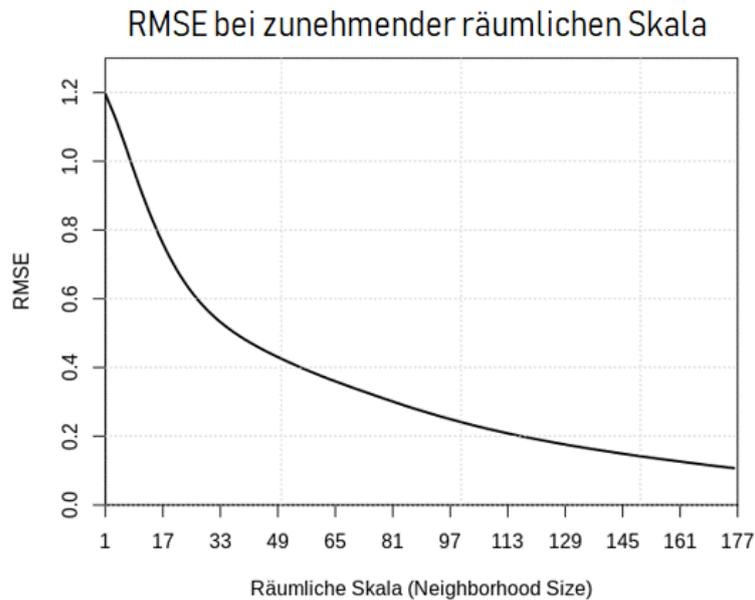


Abbildung 1.2: Root Mean Squared Error bei zunehmender räumlichen Skala. Auf der Abszissenachse ist die Anzahl der Gitterpunkt aufgetreten, welche bei der Berechnung berücksichtigt werden. Als Verifikationsmaß dient der RMSE, welcher auf der Ordinatenachse abgebildet ist. Je größer die räumliche Skala wird, desto bessere Ergebnisse werden erhalten.

der Modellformulierung, zu (vgl. Bjerknes (1911), Orrell et al. (2001)). Ensemblevorhersagen schätzen diese Unsicherheiten ab (Leutbecher and Palmer, 2007), dazu sei auf Kapitel 2.1 verwiesen. Abbildung 1.3 zeigt den schematischen Verlauf eines Vorhersagefehlers (Santoalla, 2018). Zunächst nimmt der Fehler rasch zu, mit zunehmender Vorhersagedauer nähert sich der Fehler einem Sättigungsgrad. Die modifizierte Vorhersage konvergiert gegen das klimatologische Mittel. Folglich wächst der Prognosefehler nicht ins Unendliche, sondern kann maximal die klimatologischen Extrema erreichen.

Um die Vielzahl an verschiedenen räumlichen Verifikationsmethoden besser vergleichen und verbessern zu können, wurde im Jahr 2007 das *Spatial Verification Method Intercomparison Project* (ICP, ral.ucar.edu/projects/icp/) ins Leben gerufen. Die Mehrheit der dort vorgestellten Methoden wird über zwei allgemeine Typen klassifiziert: Filtermethoden und Verschiebungsmethoden (Jolliffe and Stephenson, 2012). Bei ersterem wird ein skalenabhängiger Filter auf das Vorhersagefeld und/oder Analysefeld angewendet und anschließend anhand verschiedener Fehlermaße verifiziert (Jolliffe and Stephenson, 2012). Die Verschiebungsmethoden messen die räumliche Verschiebung oder Verzerrung, welche auf das Vorhersagefeld angewendet werden müsste, damit es mit dem analysierten Feld übereinstimmt (Jolliffe and Stephenson, 2012). Während des Vorgangs werden bestimmte Eigenschaften, z.B. Amplitude oder räumliche Verschiebung, des Modell- und Analysefelds betrachtet und verglichen. Diese zwei Kategorien wurden im Rahmen des Projekts des ICP in 4 weitere Kategorien (Field-Deformation, Neighborhood, Scale-Separation, Feature-Based) eingeteilt. In Phase 2 wurden diese um eine Kategorie (Distance Metrics) erweitert (Dorninger et al.,

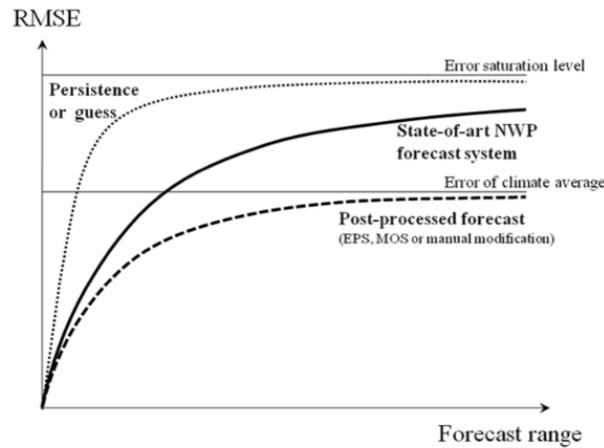


Abbildung 1.3: Schematische Darstellung der prognostizierten Fehlerentwicklung eines numerischen Wettervorhersagemodells (State-of-art NWP). Zusätzlich ist Persistenz (Persistence or guess) eingezeichnet, deren Fehler zu einem höheren Fehlersättigungsgrad konvergieren als modifizierte Prognosen, die zu einem niedrigeren RMSE-Niveau konvergieren (Post-processed forecast) (vgl. ECMWF Forecast User Guide (Santoalla, 2018)).

2018) (vgl. Abbildung 1.4). Allerdings muss eine Methode nicht zwingend nur einer Kategorie angehören (Gilleland et al., 2009). Phase 2 startete 2014 unter dem Namen *Mesoscale*

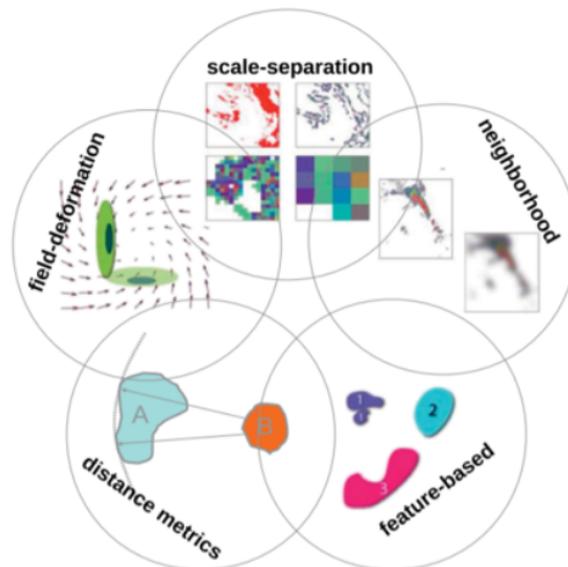


Abbildung 1.4: Kategorisierung der Räumlichen Verifikationsmethoden: Methoden, welche sich ähnlicher sind, liegen näher beisammen, wobei Überschneidungen (Kreise) nicht auszuschließen sind. Bild: Dorninger et al. (2018).

Verification Intercomparison over Complex Terrain (MesoVICT, Dorninger et al. (2018)). Während in Phase 1 das bessere Verständnis von räumlichen Verifikationsmethoden im Vordergrund stand, liegt der Schwerpunkt in Phase 2 auf der Anwendung der Methoden, vor allem bei komplexeren meteorologischen Fragestellungen (Dorninger et al., 2018). Dazu zählen etwa der Umgang mit Ensemblevorhersagen oder anderen meteorologischen Parametern neben dem Niederschlag oder auch der Umgang mit Modellen im komplexen Gelände.

1.2.1 Kategorisierung räumlicher Verifikationsmethoden

Neighborhood Methoden auch Fuzzy Methoden genannt, gehören zu den Filtermethoden und verwenden Tiefpassfilter, die höhere Frequenzen dämpfen. Ein Beispiel ist der Glättungsfilter (Dorninger et al., 2018), wodurch das betrachtete Objekt unscharf wird. Die Vorhersagen müssen nur näherungsweise in Raum, Zeit oder Intensität mit dem Analysefeld übereinstimmen, um mit einem guten Score bewertet zu werden (Ebert, 2008). Die verschiedenen Neighborhood-Methoden werden durch Entscheidungsmodelle unterschieden, mit denen ausgesagt wird, ob eine Prognose nahe genug und somit als nützlich für den Benutzer angesehen werden kann (Ebert, 2008). Die Methode Upscaling (vgl. Zepeda-Arce et al. (2000), Yates et al. (2006)) besagt, eine nützliche Vorhersage ähnelt der Analyse, wenn sie auf gröberen Skalen gemittelt wird. Der Fractions Skill Score (vgl. Roberts and Lean (2008)) sagt aus, dass eine Vorhersage nützlich ist, wenn die Häufigkeit von vorhergesagten und analysierten Ereignissen ähnlich sind. Das soll vor allem Benutzer helfen, für ihre Zwecke sinnvolle Methoden zu verwenden.

Neighborhood Methoden vergleichen Vorhersage und Analyse eines Gitterpunkts innerhalb eines definierten Fensters (Dorninger et al., 2018), das bedeutet benachbarten Gitterpunkte fließen in die Verifikation mit ein. Es gibt keine eindeutige Antwort, wie viele Gitterpunkte mitberücksichtigt werden sollen (Ebert, 2008). Umfasst die Gitterbox nur einen einzelnen Gitterpunkt, so wird eine traditionelle Punkt-Verifikation durchgeführt (Jolliffe and Stephenson, 2012). Durch Variation der Größe der räumlichen Skala kann die Prognosequalität als Funktion der Skala bestimmt werden (Ebert, 2008). Des Weiteren wird die Qualität für mehrere Intensitätsschwellen berechnet (Ebert, 2009). Eine Anwendung aus der Kategorie der Neighborhood Methoden wurde auch für Abbildung 1.2 verwendet. Neighborhood Methoden können allerdings keine Fehlerquellen diagnostizieren Ebert (2009), daher empfiehlt es sich die Ergebnisse mit anderen räumlichen Verifikationsmethoden zu vergleichen.

Feature-Based Methoden oder objektorientierte Methoden gehören zu den Verschiebungsmethoden. Über einen Schwellenwert werden Wetterereignisse in Vorhersage- und Analysefeld als Objekte definiert und diese anschließend miteinander verglichen (Dorninger et al., 2018). Das Objekt wird anhand verschiedener voneinander unabhängiger Attribute wie Intensität, Struktur oder Ausdehnung kategorisiert und bewertet (Jolliffe and Stephenson, 2012). Wichtig ist, dass die Prognose- und Analysefelder auf einem gemeinsamen Gitter liegen und die darauf betrachteten Objekte eindeutig definiert sind (Jolliffe and

Stephenson, 2012). Ein Beispiel für ein natürliches Objekt ist ein Niederschlagsereignis. Der Niederschlag hat als untere Grenze den Wert Null, die direkt zur Definition der Objekte verwendet werden können. Die Windgeschwindigkeit ist wie der Niederschlag eine positiv semidefinite Größe und kann zusätzlich nicht unendlich groß werden. Jedoch sind Kalmen viel seltener, wodurch die untere Grenze für die Definition der Objekte ungeeignet ist. Die Windgeschwindigkeit hat keine klar definierten Grenzen und muss vor der Verifikation mit Feature Based Methoden in ein Objekt umgewandelt werden.

Field-Deformation Methoden gehören zu den Displacement Methoden, die die Vorhersage als eine Schätzung des Analysefeldes, mit Fehlern in der Intensität, räumlichen Ausdehnung, betrachtet (Jolliffe and Stephenson, 2012). Ähnlich zu den Feature-Based-Methoden wird das Wetterereignis als Objekt betrachtet. Der Unterschied liegt darin, dass der Fehler nicht für einzelne Skalen oder Attribute betrachtet werden, sondern für das Objekt an sich. Dafür wird das Objekt des Vorhersagefeldes in Raum und Form verändert, um es an das analysierte Objekt anzupassen (Gilleland et al., 2009). Das Objekt darf verschoben, gedreht oder auch neu skaliert werden. Eine geringere Verformung bedeutet eine bessere Prognose (Jolliffe and Stephenson, 2012). Die Verschiebung des Vorhersagefeldes erfolgt über ein Vektorfeld, welches zuvor berechnet werden muss. Keil and Craig (2009) verwenden für die Berechnung einen sogenannten Optical Flow Algorithmus (siehe Kapitel 3.4). Anschließend erhält man Informationen über Verschiebung, Rotation etc. des gesamten Feldes und verifiziert das Differenzfeld (Keil and Craig, 2009). Es muss nicht zwingend das Vorhersagefeld an das Analysefeld angepasst werden, es ist auch umgekehrt möglich.

Distance-Metrics Methoden gelten als eine Mischung von Field Deformation und Feature Based Methoden. Vorhersage- und Analysefeld werden in binäre Abbildungen umgewandelt (Jolliffe and Stephenson, 2012), anschließend wird die Distanz zwischen Vorhersage- und Analysefeld für jeden einzelnen Gitterpunkt verifiziert (Dorninger et al., 2018). Beispielsweise wird die Hausdorff-Distanz verwendet, um binäre Bilder zu vergleichen (Jolliffe and Stephenson, 2012).

Scale-Separation Methoden gehören zu den Filtermethoden. Mittels eines räumlichen Bandpassfilters (z.B. Fourier- oder Wavelet-Zerlegung (Jolliffe and Stephenson, 2012)) werden in Vorhersage- und Analysefeld einzelne Wellenlängen isoliert. Diese Wellenlängen werden mit Wetterphänomenen verschiedener Wetterlagen in Verbindung gebracht (Casati, 2009). Welche Methode aus den obigen fünf Klassifikationen für die räumliche Verifikation verwendet werden soll, ist benutzerabhängig. Eine genauere Erklärung der in dieser Masterarbeit verwendeten Maße findet sich unter Kapitel 3.

Kapitel 2

Daten

Dieses Kapitel stellt die Daten des Modells und dessen Vergleichsdaten vor. Die Vorhersagedaten wurden durch eine Ensemblevorhersage generiert, welche für Vereisungsprognosen an Windkraftanlagen entwickelt wurde. Als Vergleichsdaten werden die Analysedaten des Vienna Enhanced Resolution Analysis (VERA) Systems, bereitgestellt von Dr. Dieter Mayer, verwendet.

2.1 Modelldaten

In Zusammenarbeit des Instituts für Meteorologie und Geophysik an der Universität Wien (IMGW) mit der Zentralanstalt für Meteorologie und Geodynamik (ZAMG), dem Stromunternehmen VERBUND Hydro Power GmbH und dem Schweizer Unternehmen Meteotest wurde im Rahmen des Projekts *ICE Control* ein Modell zur Vorhersage von Vereisung an Windkraftanlagen entwickelt. Ziel des Projekts ist es, eine quantifizierbare Verbesserung in der Vorhersage von Vereisungen zu erreichen und damit einhergehend einen kosteneffektiven Betrieb der Windkraftanlagen zu gewährleisten. Die Vereisungsvorhersage basiert auf einer Modellkette, die aus dem Meteorologischen Vorhersagemodell des *Weather Research and Forecasting Model* (WRF) und daran angehängt das Makkonen Modell, ein Vereisungsmodell (vgl. Makkonen (1981)), besteht. Zum Abschätzen der Unsicherheiten dient eine Ensemblevorhersage. Informationen zur Modellkette und zu den Auswirkungen der Vereisung an Windkraftanlagen in diesem Kapitel finden sich - sofern nicht anders angegeben - auf der Homepage des Instituts für Meteorologie und Geophysik (img.univie.ac.at/forschung/meteorologie/projekte/ice-control, zuletzt aufgerufen 17.11.2019) und der Zentralanstalt für Meteorologie und Geodynamik (www.zamg.ac.at/cms/de/wetter/news/ice-control-projekt-zur-prognose-von-vereisung-an-windkraftanlagen, zuletzt aufgerufen am 17.11.2019).

Die Motivation der Entwicklung lag in der Problematik der Vereisung der Rotorblätter. Kommt es zum Eiswurf, besteht ein hohes Sicherheitsrisiko (Dierer et al., 2011) und die

Anlage muss abgeschaltet werden (Hau, 2013). Das führt zu ungeplanten Stillstandszeiten und damit verbundenen erhöhten Ertragsverlusten (Dierer et al., 2011). Abhilfe verschafft die Beheizung der Rotorblätter, um von Beginn an Vereisung zu vermeiden (Diekmann and Rosenthal, 2014). Eine exakte Prognose für das Eintreten der Vereisung soll die Kosten, verursacht durch Vereisung, minimal halten. Ist die Vorhersage ungenau, fallen weitere Kosten an: Heizkosten für die Beheizung der Rotorblätter, obwohl keine Vereisung auftritt oder zusätzliche Personalkosten, obwohl kein Personal zum Beheizen der Anlage benötigt wird.

Für die Modellierung des meteorologischen Modells WRF werden zwei genestete Domänen mit 12.5 km und 2.5 km verwendet. Abbildung 2.1 zeigt die Ausschnitte Europa und Deutschland für die beiden Domänen. Im Zentrum der Domänen befindet sich der Windpark Ellern (siehe Verbund (2017)) mit 18 Windkraftanlagen in der Region Hunsrück in Rheinland-Pfalz in Deutschland, welcher von VERBUND Hydro Power GmbH betrieben wird. Das Modell generiert Vorhersagen aus 10 Ensemblemitgliedern und dem Kontrolllauf. Die Mitglieder bestehen aus einem Set von verschiedenen physikalischen Parametrisierungen. Im Userguide von WRF (2017) können die möglichen Parametrisierungen wie die Wolkenphysik, Konvektion oder Wechselwirkungen in der Landnutzung nachgelesen werden.

Das Modell liefert Vorhersagen für die Wintersaisons November 2016 bis März 2017 und November 2017 bis März 2018. Im Folgenden werden die Modelldaten aus der ersten Saison verwendet.

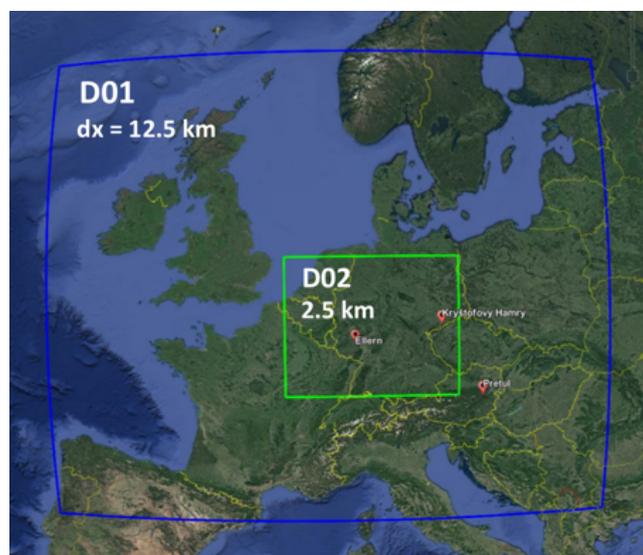


Abbildung 2.1: Zwei Domänen des meteorologischen Modells WRF: Domäne 01 mit räumlicher Auflösung 12.5 km, Domäne 02 mit räumlicher Auflösung 2.5 km. Die Domänen wurden so gewählt, dass der Windpark Ellern in Deutschland im Zentrum liegt.

2.2 Vienna Enhanced Resolution Analysis (VERA)

Die Verifikation wird anhand der *Vienna Enhanced Resolution Analysis* Daten durchgeführt. Eine gute Beschreibung des Analyseverfahrens findet sich auf der offiziellen Website (VERA (1995), <https://www.univie.ac.at/amk/vera>), welche für dieses Kapitel zusammengefasst wurde. Das Grundkonzept des VERA Systems wurde 1995 am Institut für Meteorologie

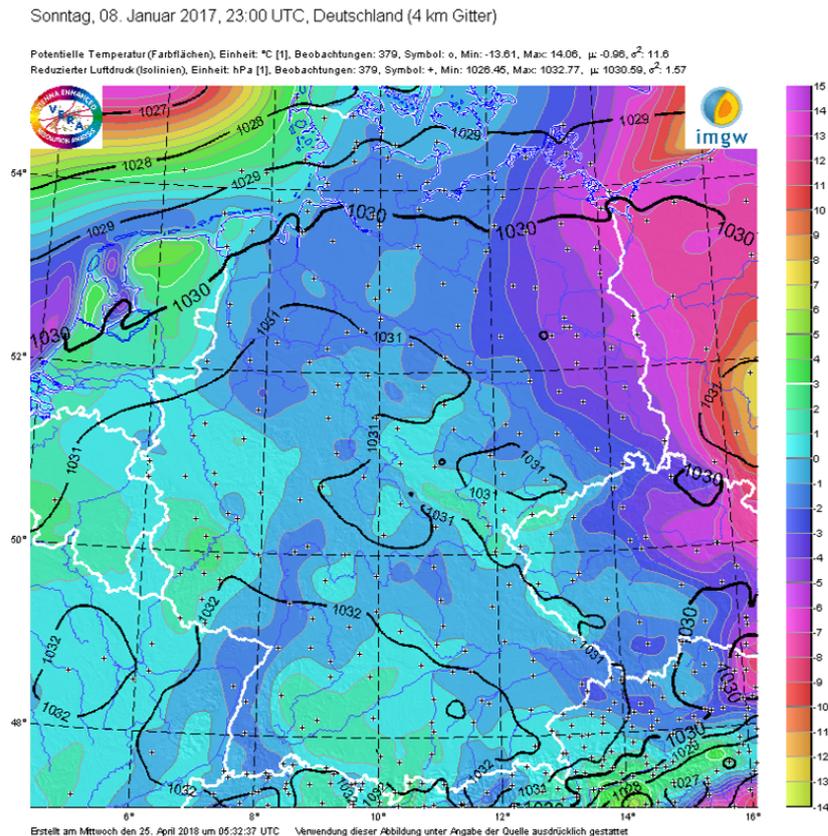


Abbildung 2.2: Analysesystem des VERA Systems der Potentiellen Temperatur in °C (Farbflächen) auf einem 4 km Gitter für den Ausschnitt Deutschland. Zusätzlich sind Isolinien des reduzierten Luftdrucks abgebildet (schwarze Linien). Die Grafik wurde von Dr. Dieter Mayer zur Verfügung gestellt.

und Geophysik der Universität Wien unter der Leitung von Prof. Dr. Reinhold Steinacker entwickelt. Die verwendete Methode wird laufend verbessert und Neuerungen hinzugefügt. Die Technik zählt zur objektiven Datenanalyse und analysiert meteorologische Daten über komplexer Orographie, z.B. den Alpen. Skalare und vektorielle Beobachtungsfelder, die als unregelmäßig verteilte Punkte vorliegen, werden auf ein regelmäßiges Gitter gebracht. Der Vorteil des Verfahrens ist, dass für die Berechnung kein statistisches Vorwissen notwendig ist, sondern nur Informationen der Stationsdaten benötigt werden.

Mathematisch basiert die Datenanalyse auf zwei Schritten: Erstens wird eine Fehlerkorrektur berechnet, welche die einzelnen Messwerte auf ihre räumliche physikalische Konsistenz testet

und die Messwerte von Fehlern befreit. Als zweiter Schritt wird eine thin-plate-spline Interpolation (vgl. Keller and Borkowski (2019)) durchgeführt. Die Quadrate der 1. Ableitung (Steigung) und 2. Ableitung (Krümmung) des Feldes werden minimiert, um ein möglichst gradientschwaches und glattes Analysefeld zu erhalten, wie es auch in der Natur angestrebt wird. Die Topographie des zu analysierenden Feldes wird zusätzlich durch sogenannte *Fingerprints* berücksichtigt (Steinacker et al., 2006). Dadurch werden subskalige Informationen, welche vom Messnetz nicht aufgelöst wurden, in die Analyse miteinberechnet.

Die Analyse wurde auf einem 4 km Gitter für Deutschland und auf einem 16 km Gitter für den Ausschnitt von Europa von Dr. Dieter Mayer zur Verfügung gestellt. Für die Verifikation wird das Vorhersagefeld für beide Auflösungen auf das 4 km Gitter interpoliert. Der Ausschnitt wird von der hochaufgelösten Domäne des Vorhersagemodells übernommen. Als Beispiel für die VERA zeigt Abbildung 2.2 die potentielle Temperatur für den 8. Jänner 2017 um 23 UTC.

2.3 Datenaufbereitung

2.3.1 Cressman-Interpolation

Die Interpolation der Vorhersagedaten auf das Analysegrid mit 4 km Auflösung wird mit einer inverse-distance Methode nach George P. Cressman (Cressman, 1959) durchgeführt. Es gilt, je weiter ein Modellpunkt von einem Messort (Analysepunkt) entfernt ist, mit einem umso kleineren Gewicht geht er in die Analyse ein. Das Prinzip dieser Abstandsgewichtung besagt, dass die dafür verwendete Gewichtung nur abhängig von der Distanz ist (Barth et al., 2008).

In die Berechnung fließen nur Gitterpunkte innerhalb eines Einflussradius mit ein, Punkte außerhalb des Radius haben keinen Einfluss und werden nicht berücksichtigt. Der Radius ist definiert als Abstand zwischen Modell- und Analysepunkt und wird als Cressmanradius d bezeichnet (Cressman, 1959).

$$d = |r_{fcst} - r_{ana}| \quad (2.1)$$

Daraus folgt, je kleiner der Radius, desto weniger Gitterpunkte fließen in die Interpolation mit ein. Ein kleinerer Radius entspricht einer geringeren Glättung. Ist der Einflussradius groß, so wird das Modellfeld stärker geglättet (McNulty, 2011).

$$w_j = \frac{N^2 - d^2}{N^2 + d^2} \quad d < N \quad (2.2)$$

$$w_j = 0 \quad d \geq N \quad (2.3)$$

wobei N der Radius des betrachteten Einflussbereichs ist.

Die Korrelation zwischen Gewichtung und Abstands des Analyse- und Modellpunkts ist in Abbildung 2.3 dargestellt und kann näherungsweise als Glockenkurve angesehen werden

(Cressman, 1959).

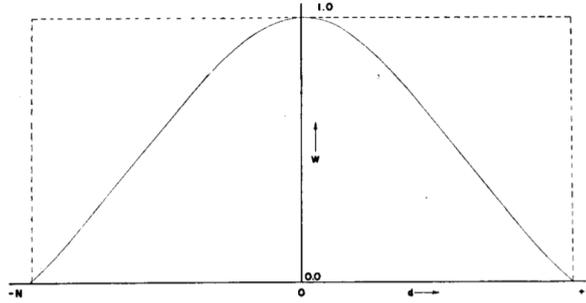


Abbildung 2.3: Die Darstellung der Gewichtsfunction W aufgetragen gegen den Abstand d ergibt eine Glockenkurve. Bild: Cressman (1959)

2.3.2 Radius der Cressman-Interpolation

Die feinskalige Domäne $d02$ wurde durch die Interpolation von einem Vorhersagegitter mit der Auflösung von 2.5 km auf das gröbere Analysegrid mit 4 km hinaufskaliert. Das Vorhersagefeld wird dadurch geglättet und verliert an Information. Die gröbere Domäne $d01$ wurde auf höhere Skalen interpoliert und zusätzlich auf den Ausschnitt von $d02$ reduziert. Die Cressman-Methode funktioniert am besten, wenn das Vorhersagefeld auf einem größeren Gitter als das Analysefeld vorliegt (Barth et al., 2008). Wegen des geringen Unterschieds zwischen der Auflösung des Vorhersagefeldes in $d02$ und des Analysefeldes kann die Interpolationsmethode angewendet werden. Es hat sich herausgestellt, dass mindestens 12 Gitterpunkte innerhalb des Einflussradius liegen sollen. Die hochaufgelöste Domäne $d02$ erhält einen Cressmanradius von 33 km. Die gröbere Domäne $d01$ erhält einen kleineren Radius von 27 km. Es sind weniger Werte notwendig, um Informationen eines groben Modells auf ein feinmaschigeres Gitter zu interpolieren. Wie sinnvoll der errechnete Cressmanradius von $cr = 33$ km ist, lässt sich rechnerisch überprüfen.

Das Vorhersage- und Analysefeld liegt nach der Interpolation auf einer Gitterfläche mit $dx = [-440, 440]$ und $dy = [-440, 440]$. Die Fläche ergibt

$$A = 884 \text{ km} \cdot 884 \text{ km} = 781456 \text{ km}^2 \quad (2.4)$$

Die Beobachtungen, welche für die Analyse zur Verfügung stehen, schwanken für den Bereich für Deutschland zwischen 190 und 420 Beobachtungen. Nimmt man ein Mittel von 305 Beobachtungen im Analysefeld an, so ergibt sich eine Beobachtungsdichte von

$$\rho = \frac{305 \text{ Obs}}{A} = 3.9 \cdot 10^{-4} \text{ Obs/km}^2 \quad (2.5)$$

$$= 3.9 \text{ Obs}/(100 \text{ km} \times 100 \text{ km}) \quad (2.6)$$

Im Analysefeld liegen rund 0.0004 Beobachtungen pro km^2 bzw. 4 Beobachtungen auf einem

100 km x 100 km Gitter.

Wie viel Beobachtungen theoretisch innerhalb der vom Cressmanradius aufgespannten Fläche ($\pi \cdot r^2$) liegen, lässt sich über die Dichte ρ bestimmen.

$$\rho_{cr}(cr) = \rho_{cr}(33) = \rho \cdot (33^2 \pi) \text{ km}^2 = 1.3 \text{ Obs/Fläche} \quad (2.7)$$

$$\rho_{cr}(cr) = \rho_{cr}(7) = \rho \cdot (7^2 \pi) \text{ km}^2 = 0.001 \text{ Obs/Fläche} \quad (2.8)$$

Bei einem Cressmanradius von 33 km liegen zumindest ein bis zwei Beobachtung innerhalb der durch den Radius aufgespannten Fläche. Bei $cr = 7$ sind es nur 0.001 Beobachtungen. Folglich ist die Verwendung des größeren Cressmanradius für die Interpolation sinnvoll. Das Ergebnis wird durch Abbildung 2.4 unterstützt. Die

Vergleich Interpolation der äquivalentpotentiellen Temperatur d02 für unterschiedliche Cressmanradien

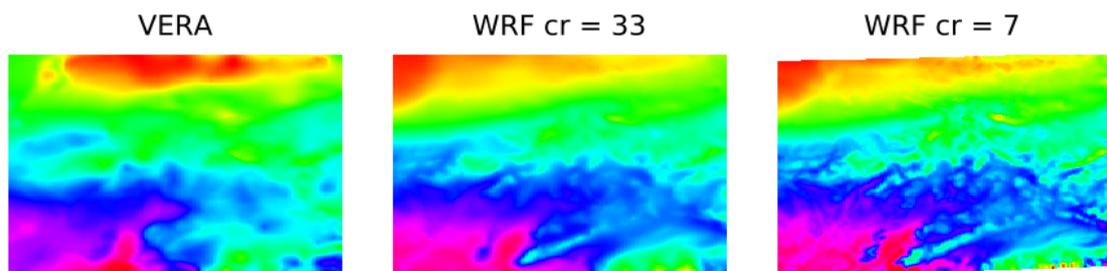


Abbildung 2.4: Äquivalentpotentielle Temperatur interpoliert mit verschieden großen Einflussradien. Links wird das VERA-Analysefeld gezeigt, in der Mitte das Modellfeld mit dem vorgeschlagenen Radius $cr = 33$, rechts das Modellfeld mit $cr = 7$.

Abbildung zeigt die äquivalentpotentielle Temperatur für verschiedene Cressmanradien für die hochaufgelöste Domäne d02. Die Abbildung zeigt das VERA-Analysefeld, das Modellfeld interpoliert mit einem Cressmanradius von $cr = 33$ und $cr = 7$. Die Struktur der Temperaturverteilung im Modellfeld ist bei kleinerem Cressmanradius feiner und detaillierter. Jedoch zeigt sich auch, dass die stärker geglätteten Felder ähnlicher zum Analysefeld sind und somit die gleichen Skalen repräsentieren.

Die grafische Darstellung unterstützt zusätzlich die mathematische Berechnung, dass ein Cressmanradius von 33 km für die Interpolation sinnvoll ist.

Kapitel 3

Ausgewählte Verifikationsmethoden

3.1 Upscaling

Upscaling gehört zu den ältesten Neighborhood-Methoden und wurde bereits zu Beginn der 2000er Jahre entwickelt (vgl. Zepeda-Arce et al. (2000), Yates et al. (2006)). Auf Grund ihrer Einfachheit wird die Methode häufig als Referenz verwendet (Ebert, 2009).

Das Entscheidungsmodell für Upscaling besagt, gute Vorhersagen ähneln den Beobachtungen, wenn sie auf gröbere Skalen gemittelt werden (Zepeda-Arce et al., 2000). Das bedeutet, die Methode mittelt Analyse- oder Vorhersagewerte, welche auf dem gleichen Gitter vorliegen müssen, innerhalb einer bestimmten räumlichen Skala und verifiziert den Mittelwert. Dabei wird über jeden Gitterpunkt der Domäne ein Fenster gelegt (vgl. Abb. 3.3). Alle Werte innerhalb des Bereiches fließen in die Berechnung des Mittelwerts mit ein. Das betrachtete Fenster wird fortlaufend um eine Gitterdistanz verschoben, um den Mittelwert für jeden Gitterpunkt zu bestimmen. Anschließend wird das Fenster kontinuierlich vergrößert, wodurch Prognose und Analyse auf immer größer werdenden Skalen gemittelt werden. Folglich wird die Auflösung des Feldes immer gröber (Zepeda-Arce et al., 2000). Upscaling nimmt an, dass eine leicht von der Analyse verschobene Vorhersage sinnvoll ist, der Grad der Verschiebung ist durch die räumliche Skala angegeben Ebert (2008). Ein Nachteil ist der Verlust von kleinskaliger Variabilität durch die Mittelung auf immer gröber werdenden Skalen (Gilleland et al., 2009).

Die Anzahl der Gitterpunkte der räumlichen Skala ist abhängig von dem betrachteten Parameter und kann vom Benutzer frei gewählt werden. Um sicherzustellen, dass der Gitterpunkt immer genau in der Mitte des Fensters liegt, muss die Anzahl der Gitterpunkte ungerade sein. Anschließend wird der Mittelwert mit verschiedenen Verifikationsmaßen aus der Kontingenztafel (siehe Abb. 3.2) bestimmt. Die Tabelle gibt die Häufigkeit an, wie oft ein Wetterereignis (nicht) eingetreten ist bzw.(nicht) vorhergesagt wird. Es gibt 4 Beziehungsmöglichkeiten zwischen Analyse und Vorhersage: Ein Wetterereignis wird korrekt vorhergesagt (*hit*), ein Wetterereignis wird vorhergesagt, tritt aber nicht ein (*false alarm*),

Wetterereignis wird nicht vorhergesagt, tritt aber ein (*missed event*) und Wetterereignis wird nicht vorhergesagt und tritt nicht ein (*correct rejection*) (Wilks, 2011). Ein Vorteil bei

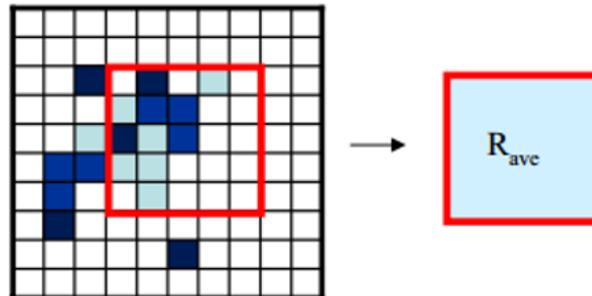


Abbildung 3.1: Prognose und Analysefeld werden innerhalb eines definierten Fensters gemittelt und dieser Wert verifiziert. Bild: Weusthoff et al. (2008)

		Analyse	
		JA	NEIN
Prognose	JA	hit	false alarm
	NEIN	miss	correct rejection

Abbildung 3.2: 2x2 Kontingenztabelle

Upscaling im Vergleich zu Punkt-Vorhersagen ist, dass die Qualität der Vorhersage gleichzeitig für mehrere Schwellenwerte bestimmt werden kann (Ebert, 2009). Die Schwellenwerte können z.B. dynamisch über Quantile bestimmt oder auch mit Hilfe des Definitionsbereiches des Parameters als festen Wert gesetzt werden. Allerdings variieren bei der dynamischen Variante zur Bestimmung der Intensitätsschwellen die Werte je nach Wetterlage.

Typische traditionelle Verifikationsmaße, welche bei Upscaling zum Einsatz kommen, sind u.a. Root Mean Squared Error (RMSE), Threat Score, Equitable Threat Score oder Frequency Bias.

(Equitable) Threat Score (TS und ETS)

Zepeda-Arce et al. (2000) verwendete den Threat Score (TS), um Niederschlagsvorhersagen auf verschiedenen räumlichen Skalen zu verifizieren. Das Verifikationsmaß ist auch als Critical Success Index (CSI) bekannt (Davis et al., 2006) und wird wie folgt berechnet:

$$TS = \frac{\text{hit}}{\text{hit} + \text{miss} + \text{false alarm}} \quad \epsilon [0, 1] \quad (3.1)$$

Der Nenner beschreibt die vorhergesagten und eingetretenen Ereignisse. Der Threat Score (TS) misst das Verhältnis der korrekt prognostizierten Ereignisse zu den vorhergesagten und

oder eingetretenen Ereignissen. Das Maß ist empfindlich gegenüber richtigen Vorhersagen und berücksichtigt keine *correct rejection* Fälle (Ebert et al., 2015). Es liegt im Intervall von 0 bis 1, wobei 1 eine perfekte Prognose bedeutet (Wilks, 2011).

Eine Variante des TS ist der Equitable Threat Score (ETS) oder Gilbert Skill Score, bei dem zusätzlich ein Korrekturterm $\text{hit}_{\text{random}}$ eingeführt wird, der die Stichprobe von Zufallstreffern beseitigen soll (Ebert, 2009). Das Maß bleibt empfindlich gegenüber korrekten Vorhersagen.

$$\text{ETS} = \frac{\text{hit} - \text{hit}_{\text{random}}}{\text{hit} + \text{miss} + \text{false alarm} - \text{hit}_{\text{random}}} \in \left[-\frac{1}{3}, 1\right] \quad (3.2)$$

mit

$$\text{hit}_{\text{random}} = \frac{1}{n}(\text{hit} + \text{miss})(\text{hit} + \text{false alarm}) \quad (3.3)$$

n ist der Stichprobenumfang aller Ereignisse. Im Gegensatz zum Threat Score hängt der Equitable Threat Score daher auch von *correct rejection* Fällen ab (Wilks, 2011).

3.1.1 Frequency Bias

Der Frequency Bias berechnet das Verhältnis der Häufigkeit von vorhergesagten Ereignissen zur Häufigkeit der eingetretenen Ereignisse (Wilks, 2011).

$$\text{Bias} = \frac{\text{hit} + \text{false alarm}}{\text{hit} + \text{miss}} \in [0, \infty) \quad (3.4)$$

Ist der Frequency Bias 1, so treten Wetterereignisse bei Vorhersage und Analyse gleich oft auf. Ist der Wert kleiner 1, wird das Ereignis zu selten vorhergesagt (*underforecast*), ist der Wert größer 1, wird das Ereignis zu oft vorhergesagt (*overforecast*) (Wilks, 2011). Der Frequency Bias gibt nicht an, wie gut das Wetterereignis bezüglich Attribute (z.B. Intensität) vorhergesagt wird; sonst misst er lediglich die Häufigkeit.

3.2 Fractions Skill Score (FSS)

Der Fractions Skill Score (FSS) zählt zu den Neighborhood-Methoden. Das Fehlermaß kann für die Genauigkeit der Vorhersage eines einzelnen Fallbeispiels berechnet werden oder auch für längere Zeiträume betrachtet werden (Mittermaier and Roberts, 2010). Vorhersage- und Analysefeld werden unter Angabe eines Schwellenwertes in ein binäres Feld eingeteilt, um den Einfluss eines Bias zu verringern (Roberts and Lean, 2008). Zunächst wird die Häufigkeit, wie oft ein Ereignis eingetreten oder vorhergesagt wurde, innerhalb eines definierten Fensterbereichs für jeden Gitterpunkt bestimmt. Anschließend werden die Felder wie auch bei der Upscaling Methode auf immer größeren räumlichen Skalen miteinander verglichen. Aus dieser Frequenz wird der Fractions Brier Score berechnet (Jolliffe and Stephenson, 2012). Das Maß ist eine Abwandlung des Brier Scores und wird ähnlich zum

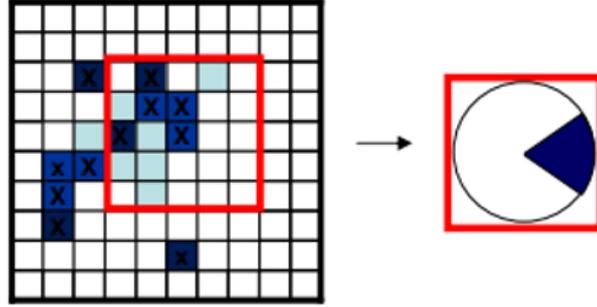


Abbildung 3.3: Innerhalb eines definierten Fensters wird die Häufigkeit des Ereignisses ermittelt und dieser Wert verifiziert. Bild: Weusthoff et al. (2008)

mittleren quadratischen Fehler (mean squared error MSE) definiert und misst die Genauigkeit von Wahrscheinlichkeitsprognosen (Wilks, 2011):

$$\text{FBS} = \frac{1}{N} \sum_N (P_{fcst} - P_{ana})^2 \quad \epsilon [0, 1] \quad (3.5)$$

mit N als der Anzahl der Gitterpunkte innerhalb der betrachteten Domäne. P_{fcst} und P_{ana} bezeichnen den Anteil der prognostizierten und eingetretenen Ereignisse, welche innerhalb der betrachteten räumlichen Skala auftreten. Der Fractions Brier Score ist 0 bei einer perfekten Vorhersage und liefert für eine schlechte Prognose den Wert 1. Jedoch ist der Fractions Brier Score abhängig von der Häufigkeit der Ereignisse (Roberts and Lean, 2008). Daher wird nicht direkt der Brier Score verglichen, sondern dessen Skill Score berechnet, der Fractions Skill Score (Roberts and Lean, 2008):

$$\text{FSS} = 1 - \frac{\text{FBS}}{\frac{1}{N} (\sum_N P_{fcst}^2 + \sum_N P_{ana}^2)} \quad \epsilon [0, 1] \quad (3.6)$$

Der Fractions Skill Score liegt zwischen 0 und 1, wobei 1 eine perfekte Prognose bedeutet.

Eine Besonderheit des Fractions Skill Scores ist, dass dieser mit zunehmender räumlichen Skala eine Prognose besser beurteilt. In Abbildung 3.4 ist der typische Verlauf des Fractions Skill Scores mit zunehmender räumlichen Skala dargestellt. Roberts and Lean (2008) zeigen auf, dass der Wert am geringsten ist, wenn nur ein einzelner Gitterpunkt betrachtet wird. Der Wert steigt mit zunehmender räumlichen Skala bis das Verifikationsmaß sich asymptotisch einem Maximalwert annähert, dort nimmt der FSS ohne Bias den Wert 1 an. Der Wert des Maximums wird berechnet mit $n = 2N - 1$, wobei N die Anzahl der Gitterpunkte an der längeren Seite der Domäne ist.

Die Vorhersage wird als sinnvoll oder nützlich bewertet, wenn gilt:

$$\text{FSS}_{\text{useful}} = 0.5 + \frac{f_0}{2} \quad (3.7)$$

f_0 entspricht der Häufigkeit des Parameters in der gesamten Domäne. Gleichung 3.7 berechnet

die kleinstmögliche räumliche Skala, welche eine sinnvolle Prognose liefert, den sogenannten *skillful scale* (Mittermaier and Roberts, 2010).

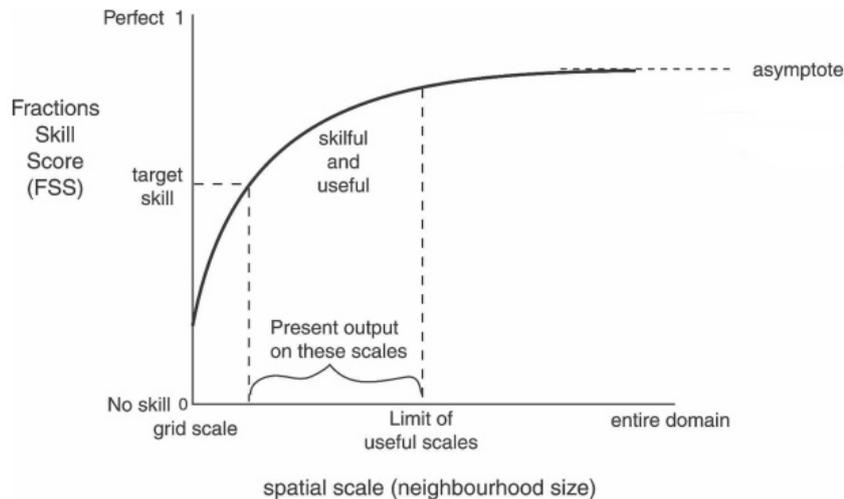


Abbildung 3.4: Verlauf des Fractions Skill Scores (FSS) mit zunehmender Skalengröße (neighbourhood size). Der FSS nähert sich asymptotisch einem Maximum. Die Skillful Scale (target skill) zeigt die kleinstmögliche Skala an, auf der die Vorhersage bereits sinnvoll ist. Bild: Roberts and Lean (2008).

3.3 Structure-Amplitude-Location (SAL)

Ein Vertreter der Feature Based Methoden ist die Structure-Amplitude-Location (SAL) Technik. Die Methode vergleicht Objekte im Vorhersage- und Analysefeld anhand voneinander unabhängiger Eigenschaften wie Struktur, Amplitude und räumliche Verschiebung, welche die Qualität der Vorhersage messen (Wernli et al., 2008). Der Vorteil gegenüber anderen objektbasierenden Methoden ist, dass es keine räumliche Zuordnung zwischen den Objekten des Analyse- und des Vorhersagefeldes notwendig ist (Dorninger and Gorgas, 2013). Die Technik wird gerne in vielen europäischen meteorologischen Zentren zur Evaluation und Verifikation von Niederschlagsvorhersagen verwendet (Jolliffe and Stephenson, 2012).

Zur Definition der Objekte schlägt Wernli et al. (2008) einen dynamischen Schwellenwert vor:

$$R^* = \frac{1}{15} R^{\max} \quad (3.8)$$

wobei R^{\max} der maximale Wert über der gesamten Domäne ist. Allerdings kann auch ein fixer benutzerabhängiger Schwellenwert gewählt werden.

3.3.1 Komponenten

Die detaillierte Methodik kann in Wernli et al. (2008) und Wernli et al. (2009) nachgelesen werden. Die folgenden Informationen über die Komponenten der SAL-Technik stammen - sofern nicht anders angegeben - aus den genannten Veröffentlichungen.

Struktur liefert Informationen über Größe und Form der Objekte. Sie wird über die normierte Differenz eines Volumens berechnet:

$$S = \frac{V(R_{\text{mod}}) - V(R_{\text{ana}})}{0.5[V(R_{\text{mod}}) + V(R_{\text{ana}})]} \quad \epsilon [-2, 2] \quad (3.9)$$

R_{ana} bezeichnet das Objekt des Analysefeldes, R_{mod} das Objekt des Vorhersagemodells. V ist das mittlere gewichtete Volumen in Abhängigkeit des Ereignisfeldes R , für das gilt:

$$V(R) = \frac{\sum_{n=1}^M R_n V_n}{\sum_{n=1}^M R_n} \quad (3.10)$$

wobei V_n ein skaliertes Volumen ist.

$$V_n = \sum_{(i,j) \in \mathcal{R}_N} \frac{R_{ij}}{R_n^{\text{max}}} = \frac{R_n}{R_n^{\text{max}}} \quad (3.11)$$

R_{max} ist der maximale Wert innerhalb des Objekts und R_{ij} sind die Werte auf dem Gitter mit R_n als die Summe über alle R_{ij} .

Die Struktur nimmt Werte zwischen -2 und $+2$ an, wobei 0 eine perfekte Vorhersage darstellt. S ist positiv, wenn das Objekt zu großflächig und/oder zu flach vorhergesagt wird und kann negative Werte erreichen, wenn es zu klein prognostiziert wird.

Amplitude liefert Informationen über die Genauigkeit der Menge bzw. Intensität des Objekts. Mathematisch ist die Amplitude die normierte Differenz des Mittelwertes über die gesamte Domäne:

$$A = \frac{D(R_{\text{mod}}) - D(R_{\text{ana}})}{0.5[D(R_{\text{mod}}) + D(R_{\text{ana}})]} \quad \epsilon [-2, 2] \quad (3.12)$$

R ist das Ereignisfeld und $D(R)$ der Mittelwert der Objekte innerhalb der Domäne:

$$D(R) = \frac{1}{n} \sum_{(i,j) \in \mathcal{D}} R_{ij} \quad (3.13)$$

wobei R_{ij} die Werte auf dem Gitter sind und n die Anzahl der Objekte in der Domäne.

Der Definitionsbereich der Amplitude liegt zwischen -2 und $+2$, wobei 0 eine perfekte Vorhersage bedeutet. Weiters, ist $A = +1$, so überschätzt die Vorhersage die Analyse um

den Faktor 3, ist $A = -1$, so unterschätzt die Vorhersage die Analyse um den Faktor 3.

Ort (Location) liefert Informationen über die räumliche Verschiebung zwischen Vorhersage- und Analyseobjekt. Die Komponente besteht aus zwei Teilen: L_1 und L_2 . L_1 misst den normierten Abstand zwischen den Massenschwerpunkten von Vorhersage und Analyse (Jolliffe and Stephenson, 2012). L_1 liefert keine Aussage über die Position der Objekte - diese können auf komplett unterschiedlichen Seiten der Domäne auftreten. Daher wird als zweiter Teil der Abstand zwischen den Objekten L_2 betrachtet (Jolliffe and Stephenson, 2012). Bei der Interpretation müssen beide Komponenten gemeinsam betrachtet werden.

$$L_1 = \frac{|x(R_{\text{mod}}) - x(R_{\text{ana}})|}{d_{\text{max}}} \in [0, 1] \quad (3.14)$$

$$L_2 = 2 \left[\frac{|r(R_{\text{mod}}) - r(R_{\text{ana}})|}{d_{\text{max}}} \right] \in [0, 1] \quad (3.15)$$

mit d_{max} als längsten Abstand zwischen zwei Eckpunkten der Domäne. x bezeichnet den Massenschwerpunkt des gesamten Feldes, r ist der massengewichtete mittlere (normierte) Abstand zwischen dem Massenschwerpunkt der gesamten Domäne s_n und dem Massenschwerpunkt eines einzelnen Objekts s .

$$r = \frac{\sum_{n=1}^M R_n (s - s_n)}{\sum_{n=1}^M R_n} \quad (3.16)$$

$$R_n = \sum_{(i,j) \in \mathcal{R}_n} R_{ij} \quad (3.17)$$

Beide Teile liegen zwischen 0 (perfekte Vorhersage) und 1 (schlechte Vorhersage).

Nach Addition der beiden Komponenten $L = L_1 + L_2$ erhält man Werte zwischen 0 und +2. Eine perfekte Vorhersage wird nur dann erreicht, wenn beide Teilkomponenten Null sind (Wernli et al., 2008). Außerdem kann die räumliche Verschiebung nie negativ werden. Generell gilt, eine Vorhersage ist dann perfekt, wenn alle drei Komponenten 0 sind ($S = A = L = 0$).

Abbildung 3.5 zeigt das SAL-Diagramm, welches Informationen über Struktur und Amplitude zusammenfasst. Die Idee wurde von Dorninger and Gorgas (2013) übernommen. Auf der x-Achse ist die Struktur, auf der y-Achse die Amplitude abzulesen. Mit Hilfe des SAL-Diagramms kann schnell eine Aussage über das Objekt getroffen werden. Liegt ein Messpunkt im 1. Quadranten (rechts oben), so sind beide Komponenten positiv. Das vorhergesagte Objekt ist zu groß oder flach ($S > 0$) und die Vorhersagemenge wird überschätzt ($A > 0$). Ähnliches gilt für die anderen drei Quadranten. Zusätzlich kann die räumliche Verschiebung als dritte Dimension miteinbezogen.

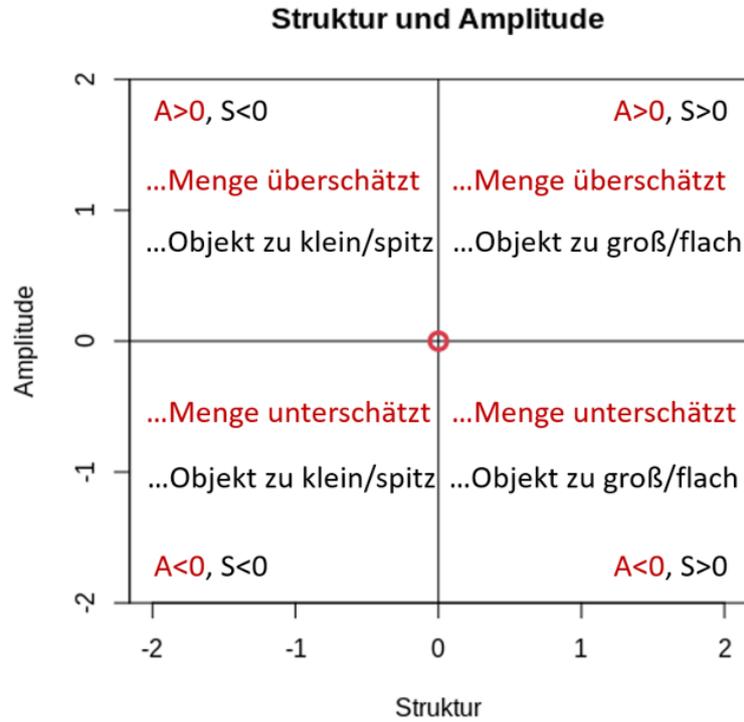


Abbildung 3.5: Schematische Darstellung eines SAL-Diagramms nach Abbildung von Dorninger and Gorgas (2013). Bei $A > 0$ wird die Menge oder Intensität des vorhergesagten Objekts überschätzt, bei $A < 0$ unterschätzt. Bei $S > 0$ ist das Objekt zu groß oder flach, bei $S < 0$ ist das Objekt zu klein/spitz.

3.3.2 Ensemble SAL (eSAL)

Die Standardversion nach Wernli et al. (2008) wird auf ein deterministisches numerisches Wettervorhersagemodell angewendet. Radanovics et al. (2018) erweiterte die Version auf die Anwendung von Ensemblevorhersagen, wobei die Ensemblemember nicht einzeln, sondern als eine Gesamtheit betrachtet werden. Die Ensembleversion kann weiterhin auf deterministische Modelle angewendet werden und funktioniert auch für ein Ensemble aus Analysedaten (Radanovics et al., 2018).

Zunächst wird wie in Formel 3.9 das gewichtete Volumen für jeden einzelnen Ensemblemember berechnet und anschließend über alle Member gemittelt. Die Struktur liegt zwischen -2 und +2 und hat die gleichen Eigenschaften wie die Standardmethode (vgl. 3.5).

$$eS = \frac{\langle V(R_{\text{mod}}) \rangle - \langle V(R_{\text{ana}}) \rangle}{0.5[\langle V(R_{\text{mod}}) \rangle + \langle V(R_{\text{obs}}) \rangle]} \quad \epsilon [-2, 2] \quad (3.18)$$

Der Mittelwert der Ensemblemember ist über den Mittelungsoperator $\langle \rangle$ gegeben. Die Amplitude wird im ersten Schritt analog zu Gleichung 3.12 über die normierte Differenz des Mittelwerts berechnet. Die Amplitude des eSAL ergibt sich durch die Mittelung über alle Ensemblemember. Auch diese Komponenten liegt zwischen -2 und +2, wobei 0 die perfekte

Vorhersage bedeutet.

$$eA = \frac{\langle D(R_{\text{mod}}) \rangle - \langle D(R_{\text{ana}}) \rangle}{0.5[\langle D(R_{\text{mod}}) \rangle + \langle D(R_{\text{ana}}) \rangle]} \quad \epsilon [-2, 2] \quad (3.19)$$

Der normierte Abstand L_1 zwischen Massenschwerpunkt von Vorhersage und Analyse ergibt sich aus dem Ensemblemittelwert des Massenschwerpunkts. Der Abstand L_2 zwischen den Objekten wird über den Continuous Ranked Probability Score (CRPS) definiert. Der CRPS ist der größte Unterschied des eSAL zur Standardmethode.

$$eL_1 = \frac{|\langle x(R_{\text{mod}}) \rangle - \langle x(R_{\text{ana}}) \rangle|}{d_{\text{max}}} \quad \epsilon [0, 1] \quad (3.20)$$

$$eL_2 = 2 \times CRPS \left[P \left(\frac{r(R_{\text{mod}})}{d} \right), P \left(\frac{r(R_{\text{ana}})}{d} \right) \right] \quad \epsilon [0, 1] \quad (3.21)$$

wobei $P(R/d)$ die empirischen Verteilungsfunktionen der Ensemblemember und der Analyse sind (Hersbach, 2000). Der CRPS ist das probabilistische Äquivalent des mittleren absoluten Fehlers (MAE) (Radanovics et al., 2018) und misst die Differenz zwischen den beiden Verteilungsfunktionen.

$$CRPS = CRPS(P_{\text{mpd}}, P_{\text{ana}}) = \int_{-\infty}^{\infty} \left[P_{\text{mod}}(x) - P_{\text{ana}}(x) \right]^2 dx \quad (3.22)$$

Der CRPS ist so definiert, dass er nur minimal sein kann, wenn die beiden Verteilungsfunktionen gleich sind. Das ist nur möglich, wenn eine perfekte deterministische Vorhersage vorliegt (Hersbach, 2000).

3.4 Displacement and Amplitude Score

Der Displacement and Amplitude Score zählt zu den Field Deformation Methoden. Die Methode untersucht, wie gut das Vorhersage- und Analysefeld in Position und Form übereinstimmen Keil and Craig (2007). Dabei wird das eine Feld (z.B. Vorhersagefeld) gedreht und verzerrt, um es an das zweite Feld (z.B. Analysefeld) bestmöglichst anzupassen (Jolliffe and Stephenson, 2012). Die Deformation wird sowohl für die Vorhersage, als auch für die Analyse durchgeführt. Damit können false alarms und missed events berücksichtigt werden (Keil and Craig, 2009).

Keil and Craig (2007) berechnen die Verschiebungsmerkmale (Amplitude und Position) mittels sogenannter optical-flow-Technik basierend auf einem pyramidical matching Algorithmus. Zuerst wird das Vorhersagefeld verpixelt und dann mit Hilfe eines Vektorfeldes verschoben (Jolliffe and Stephenson, 2012), um es an das Bild des Analysefeldes anzupassen. Das Vektorfeld wird so gewählt, dass es den RMSE zwischen Vorhersage- und Analysefeld minimiert (Keil and Craig, 2009). Die Größe der berechneten Vektoren ist ein Maß für den Verschiebungsfehler DIS, die Intensität zwischen den Feldern vor und nach der Deformation

ist ein Maß für den Amplitudenfehler AMP (Keil and Craig, 2007). Dieser Vorgang kann beliebig oft wiederholt werden, wobei die Auflösung des Gitters immer feiner wird. Die beiden Fehlergrößen erhält man nach jedem Durchgang. Anschließend berechnet man ein gewichtetes Mittel dieser Ergebnisse. Die Formeln sind Keil and Craig (2009) entnommen:

$$\text{AMP} = \frac{1}{n_{\text{ana}} + n_{\text{fcst}}} \left(n_{\text{ana}} \overline{\text{AMP}_{\text{ana}}} + n_{\text{fcst}} \overline{\text{AMP}_{\text{fcst}}} \right) \quad (3.23)$$

$$\text{DIS} = \frac{1}{n_{\text{ana}} + n_{\text{fcst}}} \left(n_{\text{ana}} \overline{\text{DIS}_{\text{ana}}} + n_{\text{fcst}} \overline{\text{DIS}_{\text{fcst}}} \right) \quad (3.24)$$

mit den Mittelwerten der Fehlergrößen $\text{AMP}(x, y)$ und $\text{DIS}(x, y)$ jedes Durchgangs und des Stichprobenumfangs n :

$$\overline{\text{AMP}_{\text{ana}}} = \frac{1}{n_{\text{ana}}} \left[\sum_A \text{AMP}_{\text{ana}}(x, y)^2 \right]^{1/2} \quad (3.25)$$

$$\overline{\text{DIS}_{\text{obs}}} = \frac{1}{n_{\text{ana}}} \sum_A \text{DIS}_{\text{ana}}(x, y) \quad (3.26)$$

Gleichungen 3.25 und 3.26 gelten für den Analyserraum *ana*, das bedeutet, das Vorhersagefeld wird in Richtung der Analyse verändert. Analoges gilt für den Vorhersageraum *fcst*.

Der Displacement and Amplitude Score (DAS) ergibt sich aus der Kombination von AMP und DIS (Gleichungen 3.23, 3.24).

$$\text{DAS} = \frac{\text{DIS}}{D_{\text{max}}} + \frac{\text{AMP}}{I_0} \quad (3.27)$$

Der Amplitudenfehler wird mit der charakteristischen Intensität I_0 normiert. Diese kann benutzerabhängig gewählt werden, häufig wird der Mittelwert des RMS des Analysefelds dafür verwendet (Keil and Craig, 2009). Der Verschiebungsfehler wird mit dem maximalen Abstand D_{max} normiert, damit die Amplitude des analysierten und vorhergesagten Feldes übereinstimmt. DAS liegt im Definitionsbereich von 0 bis unendlich, meist hat das Fehlermaß eine Größenordnung von 1.

3.5 Software

Dieses Unterkapitel beschreibt die Software für die Berechnung der Methoden und inklusive der Programmiersprache und der verwendeten Routinen.

3.5.1 Interpolation

Wie in Kapitel 2.3.1 beschrieben, wird die Interpolation der Modelldaten auf die Analysedaten die Methode nach Cressman verwendet. Für die Berechnung wurde eine Routine in Fortran 90 von Ass.-Prof. Mag. Dr. Manfred Dorninger angepasst.

3.5.2 Räumliche Verifikation

Zur Berechnung vieler räumlicher Verifikationsmethoden steht in der Programmiersprache R das Package SpatialVx zur Verfügung. Mit Hilfe dieses Packages wurden die Ergebnisse der Methoden Upscaling (3.1), Fractions Skill Score (3.2) und Structure-Amplitude-Location (3.3) berechnet. Zusätzlich liefert das Package SpatialVx die Möglichkeit die Ergebnisse grafisch darzustellen. Die Visualisierungen der Ergebnisse dieser Masterarbeit wurden manuell erstellt, auch wenn die Ideen aus dem Package übernommen wurden.

Die Ensembleversion eSAL ist eine aktualisierte Version der SAL Routine des Packages SpatialVx von Dr. Sabine Radanovics (researchgate.net) frei verfügbar.

Der Displacement and Amplitude Score (3.4) nach Keil and Craig (2007) ist nicht in SpatialVx implementiert. Die Methode basierend auf einem Optical Flow Algorithm wurde als Python 2 Programm von Dr. Christian Keil zur Verfügung gestellt, welche für die Berechnung der Ergebnisse in Python 3 umgeschrieben und angepasst wurde.

Kapitel 4

Ergebnisse und Diskussion

4.1 Datenanalyse

Das betrachtete Vorhersagemodell liefert von November 2016 bis März 2017 täglich eine 60-stündige Wettervorhersage. Folgende meteorologische Parameter werden von einer Auswahl an räumlichen Verifikationsmethoden (vgl. 2) verifiziert: Windgeschwindigkeit des 10m-Windes, 6-stündiger akkumulierter Niederschlag, äquivalentpotentielle Temperatur und reduzierter Luftdruck. Der Fokus der Masterarbeit liegt auf der Windgeschwindigkeit. Als erstes wird der Mittelwert der Windgeschwindigkeit über den gesamten Zeitraum betrachtet. Abbildung 4.1(Links) zeigt die gemittelte Windgeschwindigkeit der hochaufgelösten Domäne d02 (siehe Abb. 2.1 Ausschnitt Deutschland) für jeden Vorhersagezeitpunkt des Kontrolllaufs und des Mittelwerts der Ensemblemember. Zusätzlich wird der Mittelwert der Analysedaten des VERA Systems für den jeweiligen zum Vorhersagezeitpunkt passenden Analysetermin gegenübergestellt. Während der Wintermonate weht auf Grund der stabilen atmosphärischen Nachtverhältnisse der Wind schwächer. Untertags steigt die Windgeschwindigkeit wieder an. Auffällig ist, dass der Tagesmittelwert für den zweiten Tag sowohl für die Vorhersage als auch für die Analyse niedriger ist. Dieses Muster ist sehr unüblich, da es sich hier um einen Mittelwert über mehrere Monate handelt. Vor allem für die Analyse müsste sich die mittlere Windgeschwindigkeit nach 24 Stunden wiederholen. Abbildung 4.1 zeigt den Bias der Windgeschwindigkeit zwischen Ensemblemittelwert und Analyse für das gesamte Winterhalbjahr von November 2016 bis März 2017. Die Differenz ist für jede Vorhersagestunde gegen den entsprechenden Analysetermin aufgetragen. Der Mittelwert der Ensemblemember zeigt eine Überschätzung der Windgeschwindigkeit für fast jeden Zeitpunkt, der Kontrolllauf unterschätzt den 10m-Wind bei höheren Windgeschwindigkeiten. Abbildung 4.2 zeigt den Bias zwischen den Ensemblemembern und der Analyse. Der Bias einiger Member nimmt mit höheren Windgeschwindigkeit stark zu und zeigt eine Überschätzung der Windvorhersage (vgl. 08_0802_031.3_LUC03.04, 10_0802_031.3_LUM02_02, 28_0802_031.3_LUC02_01). Andere Member sind nahezu ident mit dem Kontrolllauf (vgl. (08_0101_031.3_LUM03_04, 08_0605_031.3_LUC03_04,

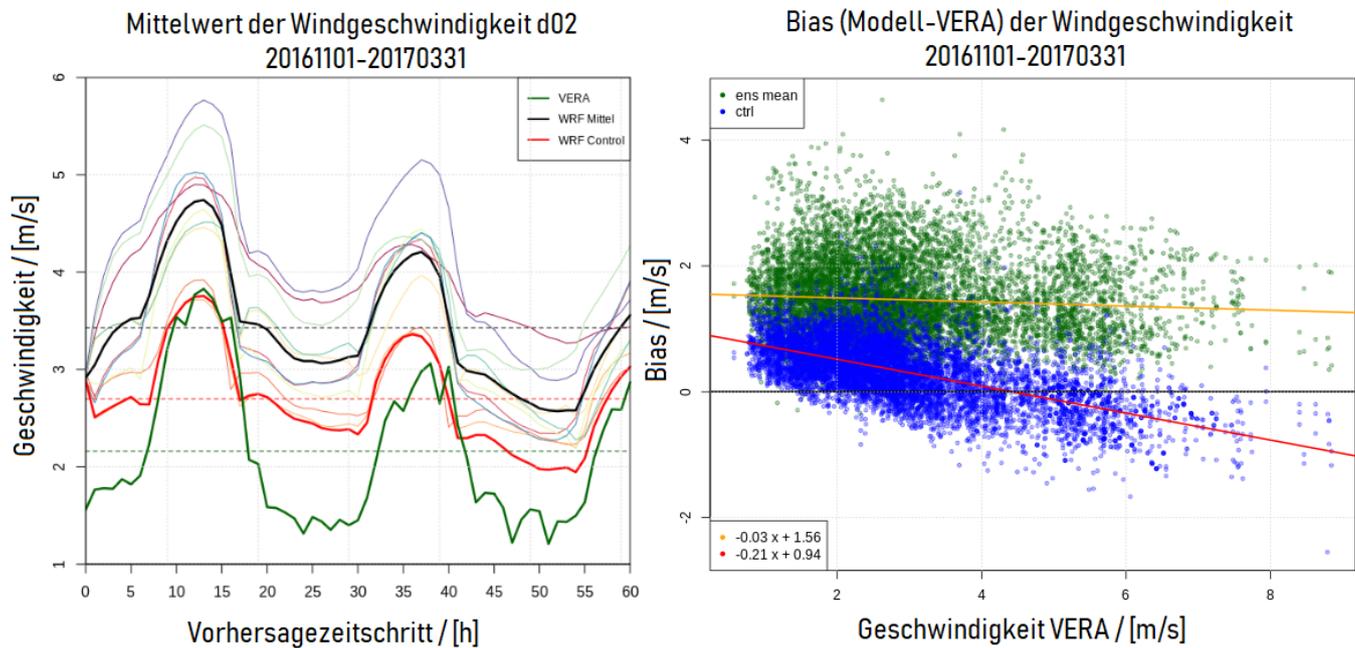


Abbildung 4.1: Links: Mittlere Windgeschwindigkeit der Domäne d02 der Analysedaten des VERA Systems (grün), WRF Kontrolllauf (rot) und der Ensemblemittelwert (Mittelwert schwarz). Rechts: Bias der Windgeschwindigkeit für jeden Vorhersagezeitschritt von November 2016 bis März 2017 zwischen dem Ensemblemittelwert des WRF Vorhersagemodells und der Datenanalyse des VERA Systems (grün) sowie zwischen dem Kontrolllauf des WRF Vorhersagemodells und der Datenanalyse des VERA Systems (blau). Die Gleichung der linearen Regression für den Ensemblemittelwert (orange) lautet $-0.03x + 1.56$ und $-0.21x + 0.94$ für den Kontrolllauf (rot).

10_0605_031_3.LUC02_01). In dieser Gruppe wird Bias mit zunehmender Windgeschwindigkeit negativ. Das bedeutet, bei niedrigen Windgeschwindigkeiten erfolgt eine Überschätzung in der Prognose, bei stärkerem Wind erfolgt eine Unterschätzung. Die restlichen Ensemblemittel sind ähnlich zum Mittelwert der Ensemblevorhersage. Die Ensemblemittel können anhand des Bias in 3 Gruppen eingeteilt werden.

Aus der Datenanalyse gehen zwei Besonderheiten hervor: Die mittlere Windgeschwindigkeit der Analysedaten des VERA Systems liegen im unteren Bereich der Ensemblevorhersage. Der Bias zeigt, dass der Großteil der Ensemblemittel höhere Windgeschwindigkeiten als die Analyse aufweisen. Der Kontrolllauf weist einen geringeren Bias auf. Folglich liegt die mittlere Windgeschwindigkeit der Analysedaten des VERA Systems im unteren Bereich der Ensemblevorhersage. Die meisten Ensemblemittel weisen höhere Windgeschwindigkeiten auf, vor allem in der Nacht ist der Unterschied zwischen den Analyse- und Modelldaten sehr groß. Das lässt sich auf eine schlechte Erfassung der sehr stabilen unteren Luftschichten im Modell schließen. Zweitens können die Ensemblemittel in drei Gruppen mit ähnlicher Vorhersage des 10 m-Windes zusammengefasst werden. In den ersten 10 Vorhersagezeitschritten ist die Bündelung besonders stark ausgeprägt und der Trend somit

Bias (Modell-VERA) der Windgeschwindigkeit 20161101-20170331

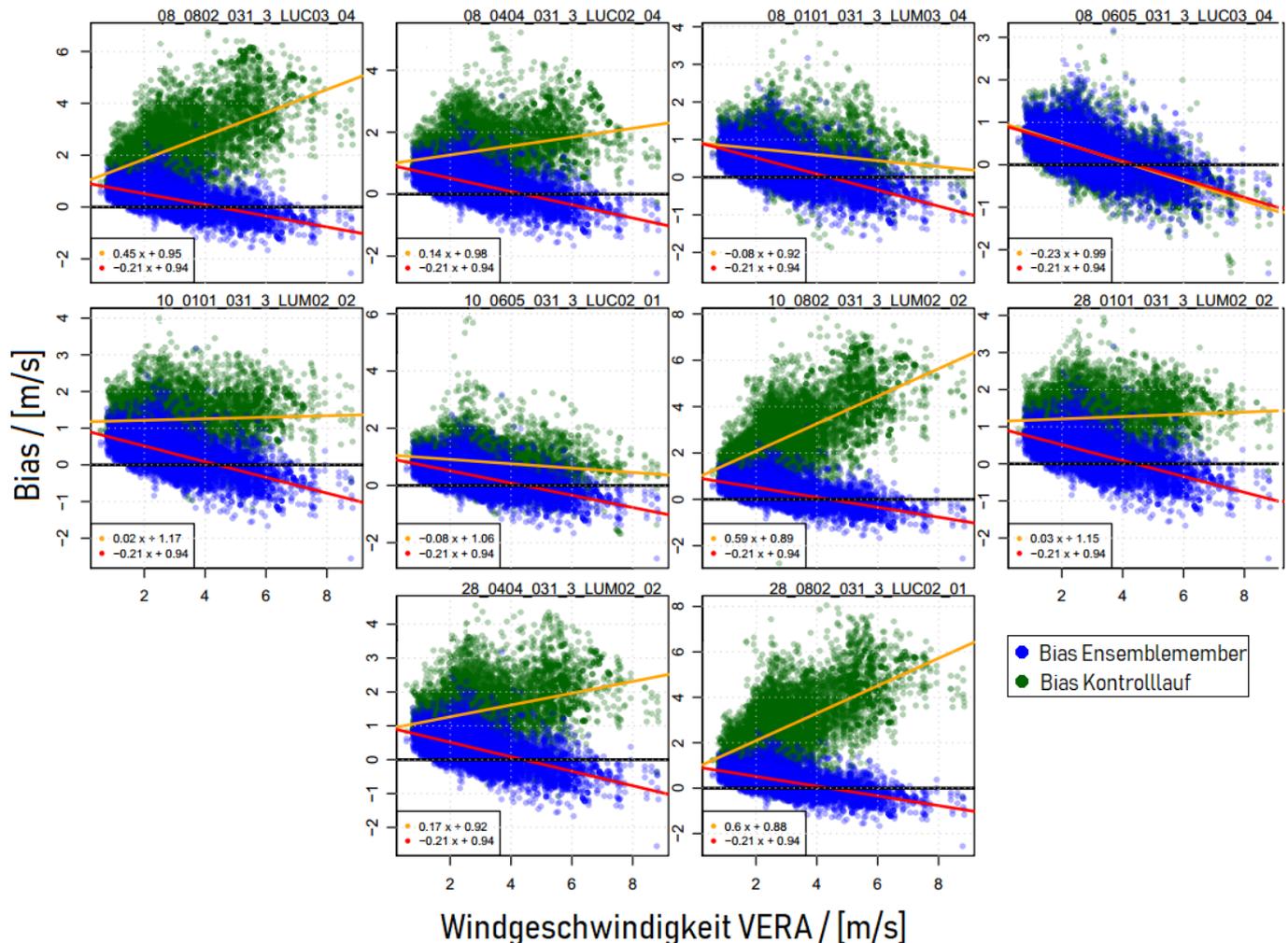


Abbildung 4.2: Bias der Windgeschwindigkeit zwischen jedem Ensemblemember des WRF Vorhersagemodells (grün) und der Datenanalyse des VERA Systems und der Bias des Kontrolllaufs des WRF Vorhersagemodells (blau) und der Datenanalyse des VERA Systems. Zusätzlich sind die Regressionsgeraden (orange = Ensemblemember, rot = Kontrolllauf) eingezeichnet und die Parameter der Regressionsgeraden dazu angegeben.

gut zu erkennen. Das lässt sich auch auf die Wahl der Ensemblemember zurückführen.

4.1.1 Equitable Threat Score: Räumliche Verifikation über das Winterhalbjahr

Die Fehlermaße von Upscaling werden über die Maße aus der Kontingenztafel definiert. Der Fractions Skill Score betrachtet die Häufigkeit eines Wetterereignisses. Ob ein Ereignis eingetreten ist oder nicht, wird in beiden Fällen durch einen Schwellenwert bestimmt.

Equitable Threat Score: d02 Mittelwert der Windgeschwindigkeit 20161101-20170331

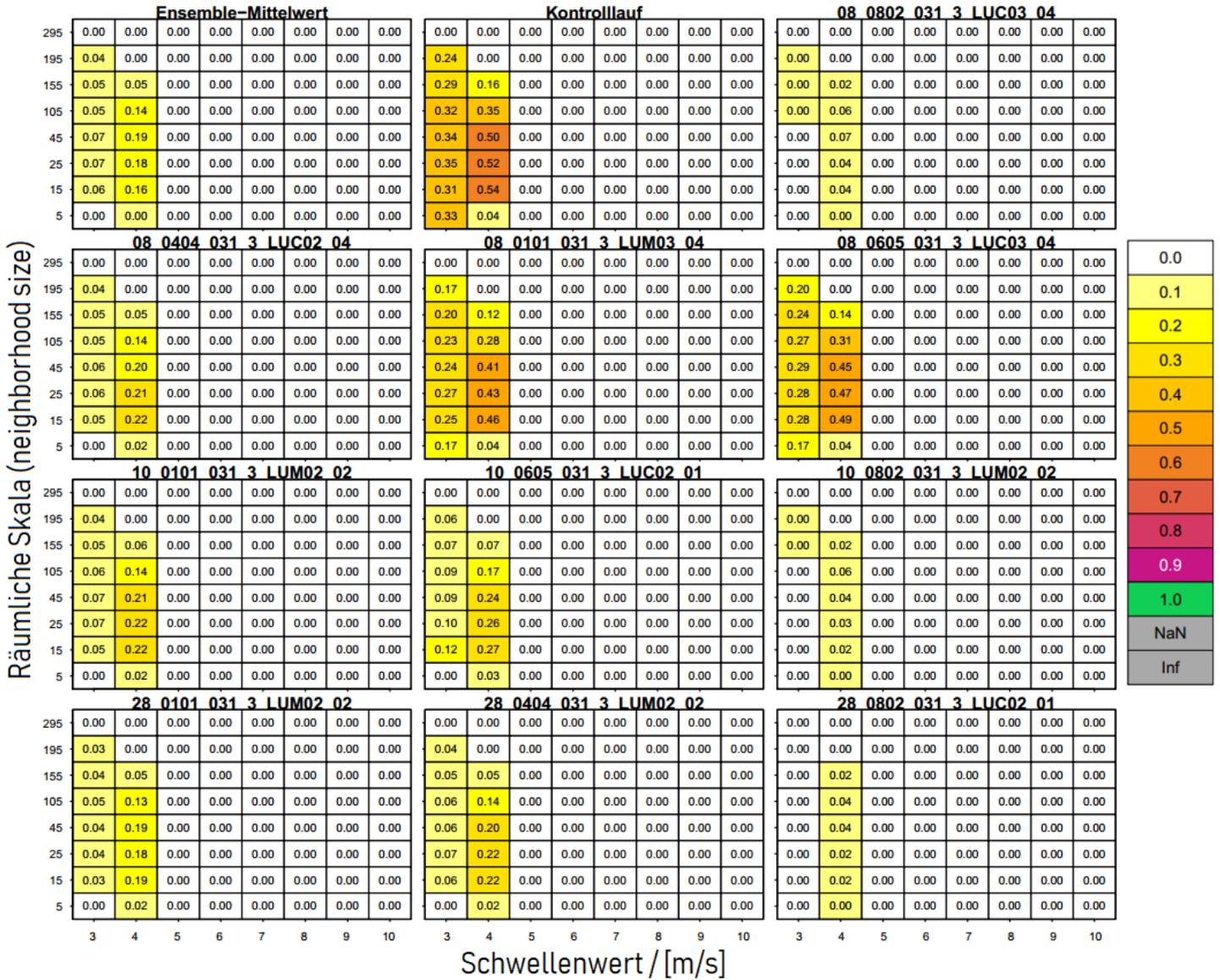


Abbildung 4.3: Equitable Threat Score (ETS) der mittleren Windgeschwindigkeit des 10 m Windes von 11/2016 bis 03/2017. Links oben ist der Mittelwert über die Ensemblevorhersage dargestellt, mitte oben der Kontrolllauf, die weiteren Matrizen sind die Ensemblemember. Die x-Achse beschreibt die Schwellenwerte in m/s, die y-Achse gibt die Anzahl der Gitterpunkte an (räumliche Skala). Der Index weist jedem Wert des ETS in der Matrix eine Farbe zu: weiß bedeutet kein Skill ($ETS \leq 0$), grün ist eine (fast) perfekte Vorhersage ($0.9 < ETS \leq 1$), die Werte zwischen den Extremfällen sind mit gelb schwach und mit violett gut.

Der Schwellenwert dient außerdem dazu, um das Vorhersage- und Analysefeld von einem kontinuierlichem Vektorfeld in das gewünschte binäre Feld umzuwandeln, um die Vorhersagen und Analysen in die Kontingenztabelle einordnen zu können. Zusätzlich geben

die betrachteten Neighborhood Methoden die Zuverlässigkeit der Vorhersage für verschiedenen räumliche Skalen an. Die untere Grenze des Schwellenwertes wurde mit 3 m/s festgelegt, da niedrigere Windgeschwindigkeiten nicht relevant sind. Die obere Grenze wurde mit 10 m/s festgelegt, da höhere Windgeschwindigkeiten seltener auftreten. Die räumlichen Skalen reichen von 5 bis 295 Gitterpunkten. Die Auflösung des gemeinsamen Gitters von Analyse und Vorhersagemodell beträgt 4 km. Die obere Grenze wurde experimentell festgelegt. Größere Skalen wurden anhand der erhaltenen Ergebnisse als nicht sinnvoll erachtet.

Die Verifikationsergebnisse werden anhand eines Intensity-Scale-Diagramms dargestellt, ähnlich zu Casati et al. (2004). Die Idee stammt aus Ebert (2008), wo das Diagramm für die Ergebnisse verschiedener Neighborhood-Methoden verwendet wurde. Die Darstellung ist für alle Neighborhood-Methoden in dieser Masterarbeit gleich gewählt. Eine andere Möglichkeit der Darstellung ist über eine Kurvenschar, wobei jede Kurve einen Schwellenwert repräsentiert. Für diese Masterarbeit wird nur die erste Variante verwendet.

Ein Vorteil von Upscaling und Fractions Skill Score ist, dass die Daten für mehrere Stunden bis Monate gemittelt und verifiziert werden können. Für die Verifikationsergebnisse in Abbildung 4.3 wurde die Windgeschwindigkeit über das gesamte Winterhalbjahr von November 2016 bis März 2017 gemittelt. Der Mittelwert wurde vom Equitable Threat Score (ETS) bewertet (vgl. Upscaling in Kapitel 3.1).

Upscaling mittelt die Analyse- und Vorhersagedaten, was den Einfluss höherer Werte abschwächt. Daher erhält man für hohe Schwellenwerte und für große Skalen stets niedrigere Verifikationsergebnisse. Zusätzlich werden die Eingangsdaten über das Winterhalbjahr gemittelt. Das bedeutet, dass vor der Verifikation höhere Werte herausgefiltert werden. Die doppelte Glättung der Vorhersage- und Analysedaten macht sich in den Ergebnissen bemerkbar. Die ETS-Werte für die mittlere Windgeschwindigkeit ergeben ab einem Schwellenwert von 5 m/s den Wert Null. Die Vorhersage besitzt für höhere Schwellenwerte keinen Skill. Bei einem ETS-Wert von 1 ist die Vorhersage perfekt. Der höchste ETS-Wert innerhalb des betrachteten Zeitraums hat der Kontrolllauf mit einem Maximalwert von 0.54 bei einem Schwellenwert von 4 m/s und einer feinen räumlichen Skala (15 Gitterpunkte bei einer Auflösung von 4 km). Auch die Ensemblemember 08_0101_031_3_LUM03_04 und 08_0605_031_3_LUC03_04 besitzen höhere ETS-Werte. Die Ensemblemember 08_0802_031_3_LUC03_04, 10_0802_031_3_LUM02_02 und 28_0802_031_3_LUC02_01 weisen die niedrigsten ETS-Werte auf.

Diskussion

Mittels Upscaling sollen jene Skalen herausgefunden werden, welche die besten Verifikationsergebnisse liefern. Die gesuchten Skalen sind anhand des farblichen Musters der Matrizen gut erkennbar. Im obigen Fall liefern die räumlichen Skalen mit 15 bis 105 Gitterpunkte bei einer ursprünglichen Auflösung von 4 km die besten Verifikationsergebnisse für einen Schwellenwert von 4 m/s. Das bedeutet, dass eine Vorhersage von 4 m/s (und höher) über das Winterhalbjahr gesehen in vielen Skalen mit den Analysedaten des VERA Systems

übereinstimmt und bezogen auf die Analyse eine gute Vorhersage liefert.

Die Ensemblemember lassen sich in die gleichen Gruppen wie bei jenen bezogen auf den Bias einteilen. Diese Übereinstimmung des Verhaltens der Ensemblemember lässt darauf schließen, dass die mittlere Windgeschwindigkeit mit der Methode verifiziert werden kann.

Es ist nicht sinnvoll, die SAL-Methode oder den DAS zu verwenden, wenn der Mittelwert über das gesamte Winterhalbjahr betrachtet wird. Dabei werden die Wetterereignisse als Objekt betrachtet. Mittelt man über mehrere Stunden oder auch länger, verliert man Informationen über Form, Intensität oder Lage der Objekte. Gerade diese Eigenschaften will man mit diesen Methoden überprüfen.

4.1.2 Fallbeispiel: Frontdurchgang am 11.1.2017 von 06-09 UTC

Wie gut sich die ausgewählten räumlichen Verifikationsmethoden eignen, soll nun anhand verschiedener Fallbeispiele getestet werden. Manche Fallbeispiele beschreiben einen Frontdurchgang mit auffrischendem Wind und höheren Niederschlagsmengen, andere eine gradientschwache Wetterlage mit geringer Windgeschwindigkeit und geringem Niederschlag in der gesamten Domäne. Abbildung 4.4 zeigt die mittleren Windgeschwindigkeiten des 60-stündigen Vorhersagelauf der hochaufgelösten Domäne d02 für vier Fallbeispiele. Die Wetterlage für den Vorhersagelauf vom 12.12. 00 UTC bis 14.12.2016 12 UTC entspricht einer Hochdrucklage, die Wetterlage für den 03.01.2017 00 UTC bis 05.01.2017 12 UTC beschreibt einen Frontdurchgang mit anschließender Hochdruckausbildung. Zwischen dem 10.01. 00 UTC und 12.01.2017 12 UTC zieht eine Front über die Domäne d02. Der Zeitraum des Vorhersagelaufs am 21.02. 00 UTC bis 23.02.2017 12 UTC beschreibt eine dynamische Wetterlage mit mehreren Frontendurchgängen. Die bereits angesprochene Gruppierung der Ensemblemember ist bei allen Fallbeispielen stark ausgeprägt. Außerdem nimmt der Spread der Ensemblevorhersage, wie zu erwarten war, mit höheren Windgeschwindigkeiten zu. Die Zeitreihe der Analysedaten des VERA Systems zeigt im Vergleich zu den Ensemblemember des WRF Vorhersagemodells keine glatte Kurve, da die Analyse von der Anzahl an Beobachtungsdaten abhängig und daher stärkeren Schwankungen ausgesetzt ist.

Im Folgenden werden die Ergebnisse für die Wetterlage vom 10.01. 00 UTC - 12.01.2017 12 UTC näher beschrieben. Das Fallbeispiel wurde ausgewählt, da die Zunahme der mittleren Windgeschwindigkeit mit Frontdurchgang im Gebiet der Domäne d02 am 11.01.2017 von etwa 06-09 UTC sehr eindeutig ist (vgl. Abb. 4.4). Außerdem weist die Ensemblevorhersage während der Zunahme des Windes einen geringen Ensemblespread im Vergleich zu den anderen Fallbeispielen (z.B. 21.02.-23.02.2017) auf. Erst nach Durchzug der Front ist der Spread höher.

Großwetterlage Westeuropa liegt in einer zonalen Höhenströmung. Im Osten greift ein Höhentrog bis nach Afrika aus. Dadurch dringen kalte arktische Luftmassen weit in den Süden vor. In 500 hPa Höhe befindet sich östlich von Island ein Tiefdruckgebiet,

Mittelwert der Windgeschwindigkeit d02 für 0-60 h

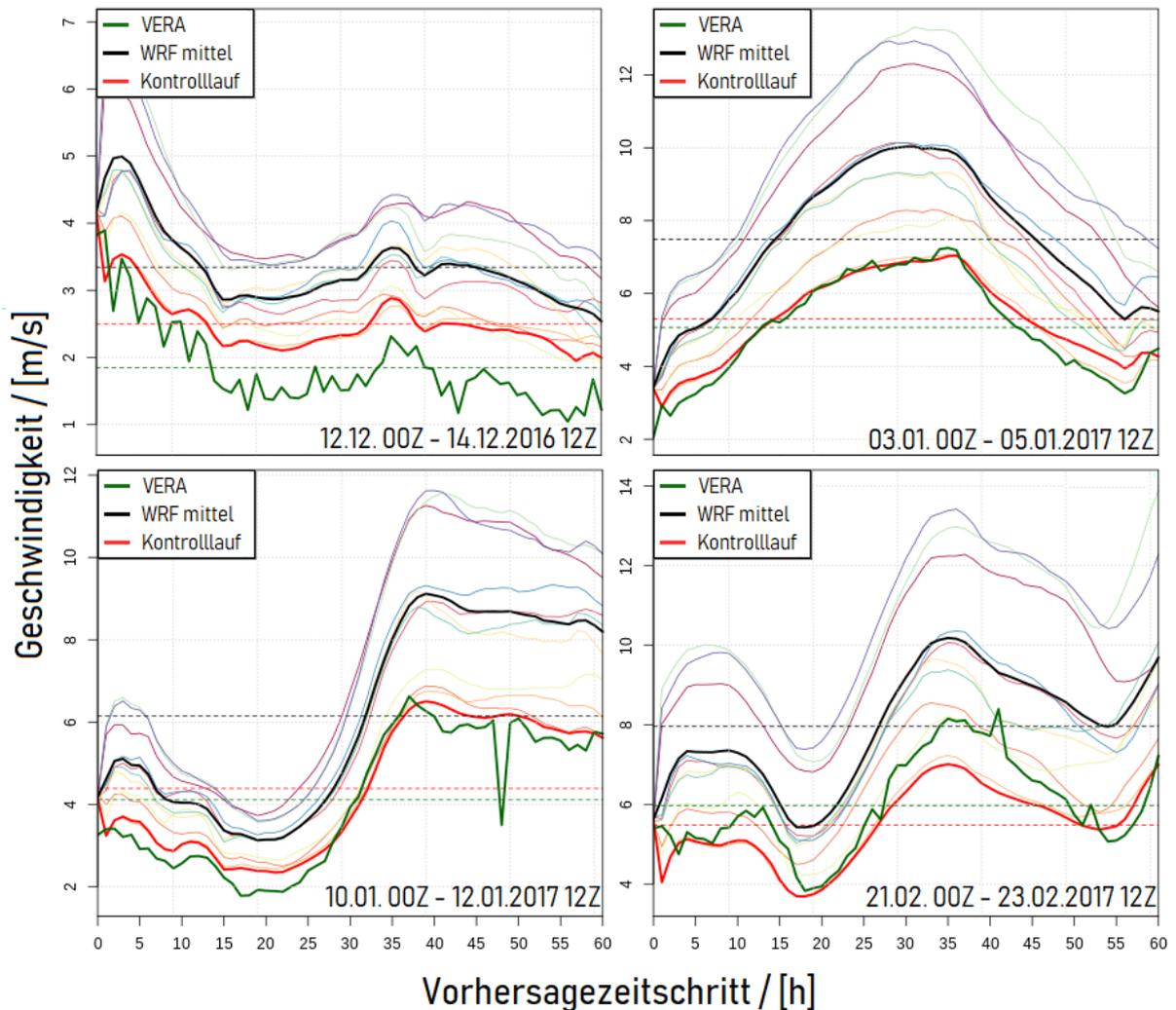


Abbildung 4.4: Windgeschwindigkeit der Analysedaten des VERA Systems (grün), des Kontrolllaufs des WRF Vorhersagemodells (rot) und der Ensemblemember des WRF Vorhersagemodells (Member bunt, Mittelwert schwarz) gemittelt über die hochaufgelöste Domäne d02 für die Vorhersageläufe vom 12.12.2016 00 UTC (12.12.-14.12.), 03.01.2017 00 UTC (03.01.-05.01.), 10.1.2017 00 UTC (10.01.-12.01.) und 21.2.2017 00 UTC (21.2.-23.2.). Die Schwankungen in den Analysedaten sind bedingt durch die unterschiedliche Anzahl an Beobachtungsdaten, welche in die Analyse miteinfließen.

welches in einen Höhengtrog eingebettet ist. An der Rückseite des zugehörigen, gut ausgeprägten Tiefdruckgebiets am Boden fließen erneut kalte polare Luftmassen von Nordwesten nach Mitteleuropa. Aus dem Tief entspringt ein Frontensystem, welches am 11. Jänner 2017 (06-09 UTC) über Deutschland zieht. Die Analyse der 850 hPa pseudopotentiellen Temperatur des Global Forecast System GFS zeigt die unterschiedlichen Luftmassen, welche sich am 11.01.2017 um 06 UTC über Europa befinden (Abb. 4.5). Man erkennt eine

Zunge energiereicher Luft nordwestlich von Deutschland. Auf Grund der sehr schmalen Warmluftzone (Zone energiereicher Luft) lässt sich auf ein bereits okkludiertes System schließen. Abbildung 4.6 zeigt frontrelevante Parameter des Vorhersagelaufs vom 10.01.2017

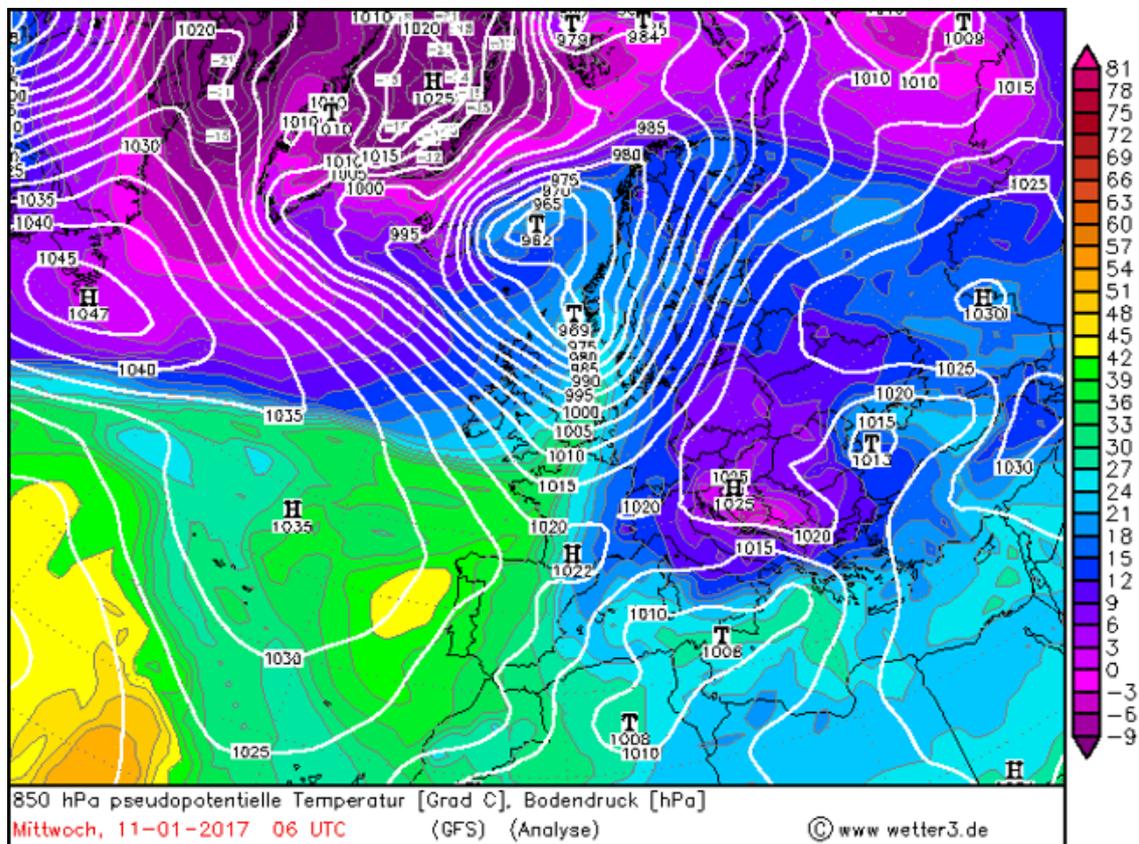


Abbildung 4.5: Analyse der Äquivalentpotentielle Temperatur in 850 hPa am 11.01.2017 06 UTC. Die Analyse wurde mittels des Wettervorhersagemodells Global Forecast System (GFS) generiert, welches vom National Centers for Environmental Prediction (NCEP) produziert wird. Die Analysekarte ist öffentlich zugänglich (www1.wetter3.de/archiv_gfs_dt.html, zuletzt aufgerufen am 23.2.2020).

00 UTC. Der reduzierte Luftdruck nimmt während des Vorhersagelaufs präfrontal ab und nach dem Durchzug der Front wieder zu. Die äquivalentpotentielle Temperatur nimmt stark zu, was auf einen Luftmassenwechsel hindeutet. Der Anstieg des Mischungsverhältnisses deutet auf eine Zunahme der Feuchte auf Grund der herrannahenden Front hin. Nach Durchzug der Front geht der Parameter wieder zurück, wenn auch nur geringfügig. Der Niederschlag nimmt nach der 30. Vorhersagestunde (entspricht 11.01.2017 06 UTC) zu. Der 10 m-Wind frischt auf und bleibt für den Rest des Vorhersagelaufs erhöht. Abbildung 4.7 zeigt die Häufigkeitsverteilung der Windgeschwindigkeit der hochaufgelösten Domäne d02. Die Analysedaten des VERA Systems und der Kontrolllauf des WRF Vorhersagemodells sind in allen Teilfenster eingefügt; somit können die Ensemblemember des Vorhersagemodells besser miteinander verglichen werden. Die häufigsten Windgeschwindigkeiten treten bei 3 bis 5 m/s auf. Die Analysedaten des VERA Systems besitzen ein zweites, geringeres Maximum bei

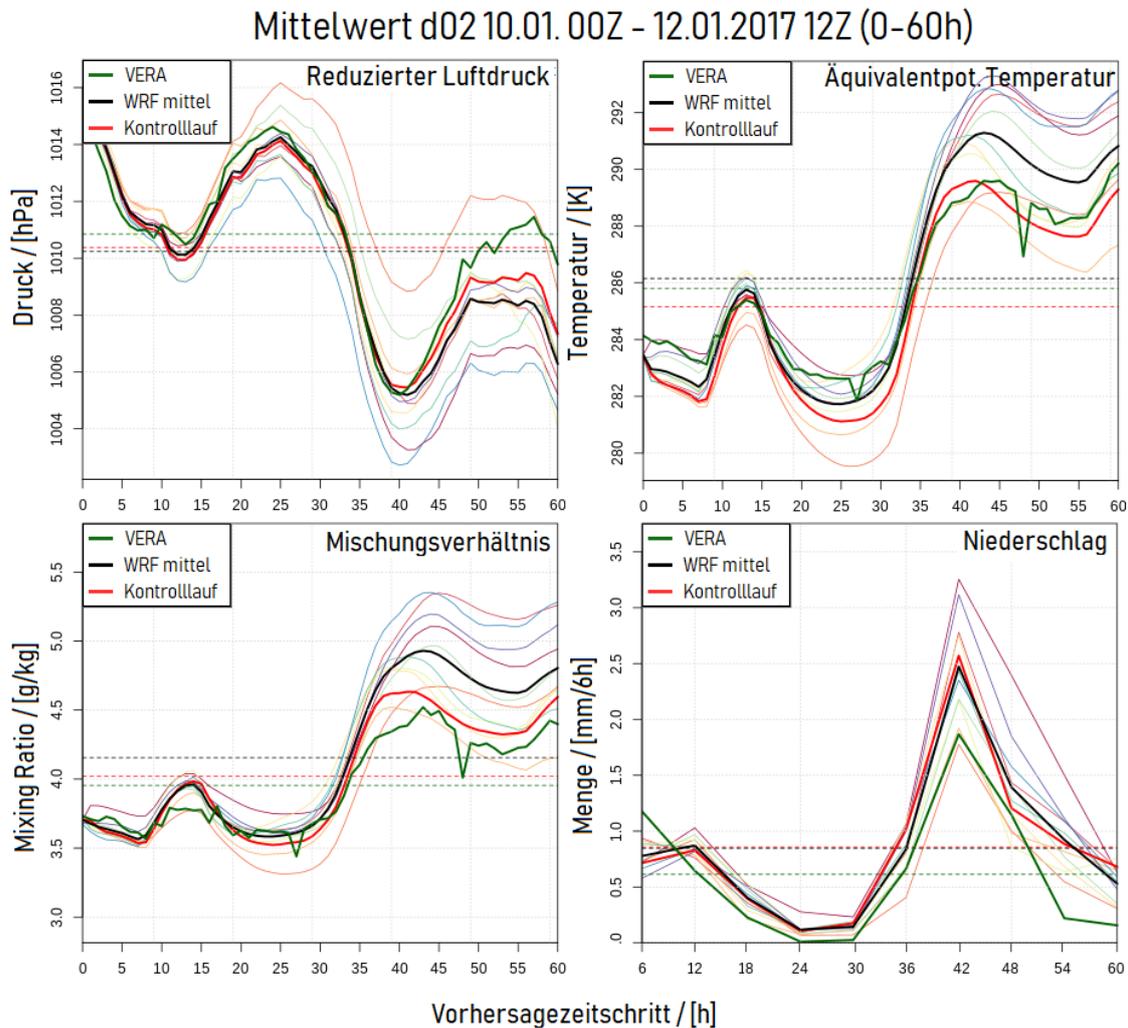


Abbildung 4.6: Mittelwert des 60-stündigen Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02 für den 6-stündigen akkumulierten Niederschlag, den reduzierten Luftdruck, die äquivalentpotentielle Temperatur und das Mischungsverhältnis.

niedrigeren Windgeschwindigkeiten. Der Nordwesten der hochaufgelösten Domäne d02 liegt über der Nordsee. Auf Grund der niedrigen Dichte an Beobachtungsstationen über dem Meer fließen weniger Werte in die Analysedaten ein, was zu niedrigeren Windgeschwindigkeiten führt. Die Anzahl an Gitterpunkten mit schwächerem Wind nimmt zu und zeigt sich als zweites Maximum in den Analysedaten. Die Werte liegen zwischen 0 und 10 m/s. Die Verteilungen der Analysedaten und des Kontrolllaufs des Modells sind ähnlich (vgl. Abb. 4.7). Die Verteilung liefert keine Informationen, ob die Werte räumlich und zeitlich übereinstimmen. Einige Ensemblemember haben einen spitze Verteilung und sind ähnlich zur Verteilung der Analyse. Andere Ensemblemember besitzen eine flachere Verteilung. Die mittlere Windgeschwindigkeit nimmt im Vergleich zur Analyse deutlich höhere Werte an. Die unterschiedlichen Ergebnisse sind erwünscht, da eine Ensemblevorhersage die Unsicherheit in der Prognose abschätzen soll. Die Häufigkeitsverteilung dieses Fallbeispiels zeigt erneut eine

Häufigkeitsverteilung: d02 Mittel der Windgeschwindigkeit vom 10.01.2017 (0-60h) (#1803648)

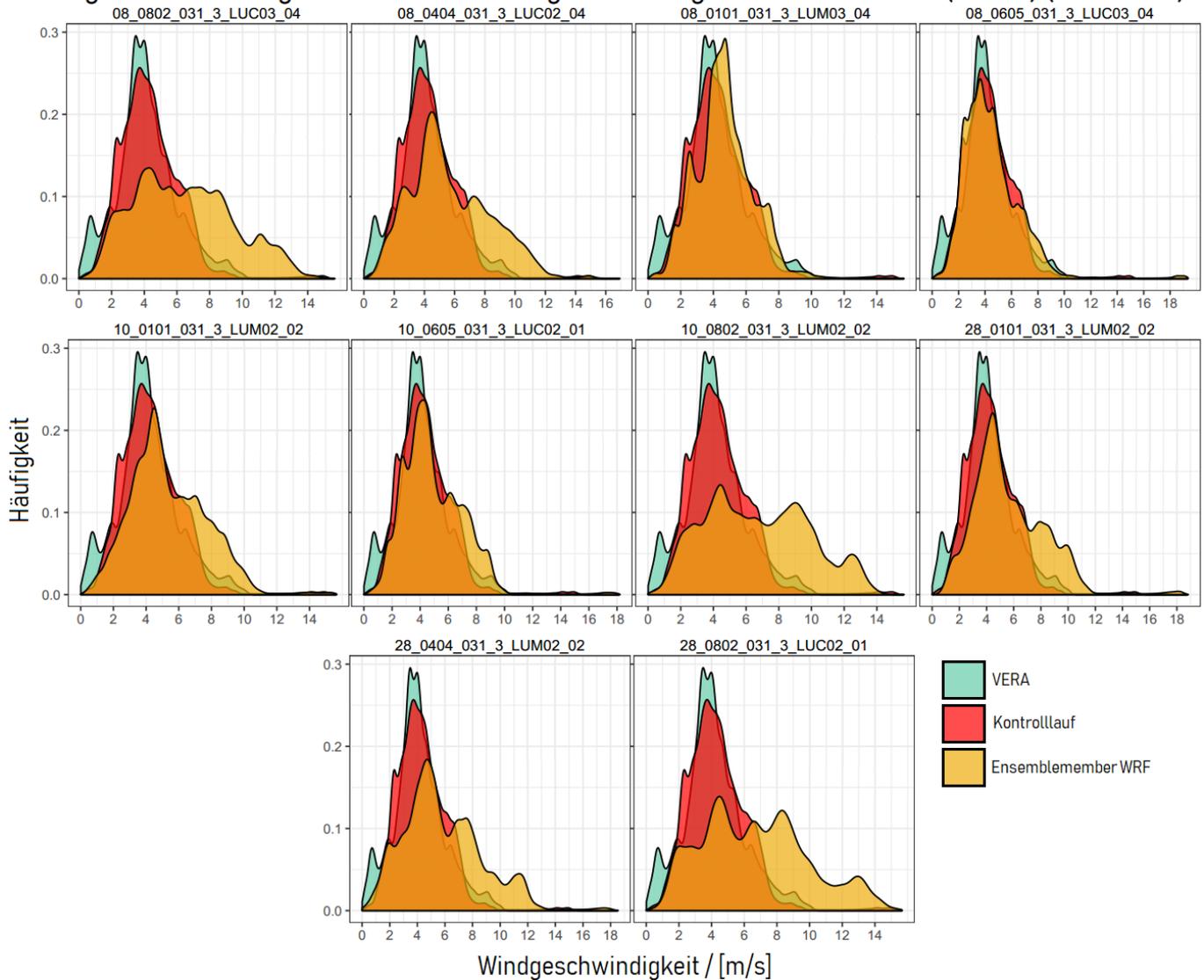


Abbildung 4.7: Relative Häufigkeitsverteilung des 10 m-Windes des Vorhersagelaufs vom 10.01. 00 UTC - 12.01.2017 12 UTC des Kontrolllaufs (rot) und der Ensemblemember (orange). Zusätzlich ist für den gleichen Zeitraum die relative Häufigkeitsverteilung der Analysedaten des VERA Systems (grün) dargestellt. Die Verteilungen des Kontrolllaufs und der Analysedaten sind in allen Fenstern gleich und dienen zur besseren Vergleichbarkeit.

Einteilung der Ensemblemember in drei Gruppen. Die Ensemblemember lassen sich in die gleichen Gruppen wie bereits beim Biaszusammenfassen.

Im Folgenden wird immer der Vorhersagelauf vom 10.01.2017 00 UTC betrachtet.

4.2 Upscaling

Abbildung 4.8 zeigt den Equitable Threat Score für den Vorhersagelauf vom 10.01.2017 00 UTC für die Vorhersagestunden 30-33 (11.01.2017 06-09 UTC). Zu diesem Zeitpunkt zieht die Front über die hochaufgelöste Domäne d02. Die Darstellung der Verifikationsergebnisse erfolgt wie in Abbildung 4.3 als Intensity-Scale-Diagramm. Die Schwellenwerte (3 bis 10 m/s) und die räumliche Skalen (5 bis 295 Gitterpunkte bei 4 km Auflösung) wurden wie zuvor gewählt. Vor der Berechnung wurde ein Mittel über die gewünschten Vorhersagestunden berechnet.

Abbildung 4.9 zeigt die Differenz der ETS-Werte der Windgeschwindigkeit zwischen der hochaufgelösten Domäne d02 mit 2.5 km Auflösung und der gröberen Domäne d01 mit 12.5 km. Beide Domänen wurden auf das Gitter des Analysefeldes mit einer Auflösung von 4 km interpoliert. Sie besitzen nach der Interpolation die gleiche räumliche Auflösung. Domäne d01 wird auf den Ausschnitt der kleineren Domäne reduziert. Zur Differenzenbildung wird Domäne d01 von d02 abgezogen. Ist die Differenz positiv, so besitzt d02 höhere ETS-Werte. Ist die Differenz negativ, so besitzt d02 niedrigere ETS-Werte im Vergleich zu d01. Der Großteil der Differenzen liegt im positiven Bereich. Das bedeutet, dass die Vorhersage der hochaufgelösten Domäne d02 im Vergleich zur Vorhersage der Domäne d01 während des Frontdurchgangs durch den ETS besser bewertet wird.

Abbildung 4.10 zeigt den Frequency Bias der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC für den Mittelwert der Vorhersagestunden 30-33 (11.11.2017 06-09 UTC). Der Frequency Bias liefert Informationen, ob die Ensemblemember ein Wetterereignis zu häufig oder zu selten vorhersagen. Bei einem Wert größer als 1 wird die Anzahl überschätzt, kleiner als 1 unterschätzt. Für das betrachtete Fallbeispiel wird das Ereignis vom WRF Vorhersagemodell generell zu häufig prognostiziert. Die Gruppe der Ensemblemember, die vom ETS am besten bewertet werden, zeigen für manche räumliche Skalen bei höheren Schwellenwerten eine Unterschätzung. Tendenziell werden hohe Werte des ETS und somit exaktere Prognosen eher bei niedrigen Windgeschwindigkeiten berechnet. Im Gegensatz zum ETS des Mittelwerts über das Winterhalbjahr, erhält man für höhere Schwellenwerte Werte größer als Null. Der Kontrolllauf und die Ensemblemember 08_0101_031_3_LUM03_04, 08_0605_031_3_LUC03_04 und 10_0605_031_3_LUC02_01 liefern die höchste Werte mit einem Maximum von 0.9, was bereits einer sehr guten Prognose entspricht (1 = perfekt).

4.2.1 Diskussion

In der Häufigkeitsverteilung und im Bias (vgl. Abb. 4.10) ist zu erkennen, dass die Ensemblemember des WRF Vorhersagemodells das gleiche Verhalten zeigen. In den Verifikationsergebnissen, vor allem bezüglich des ETS, ergibt sich das gleiche Muster. Die Ensemblemember lassen sich wie bei der Häufigkeitsverteilung (vgl. Abb. 4.7) in die gleichen Gruppen zusammenfassen. Die Ensemblemember mit einer spitzen Häufigkeitsverteilung liefern die höchsten ETS-Werte, was einer guten Vorhersagequalität entspricht. Ensemblemember mit einer flachen Verteilung ergeben die niedrigsten ETS-Werte.

Equitable Threat Score: d02 Windgeschwindigkeit vom 10.01.2017 + 30-33h

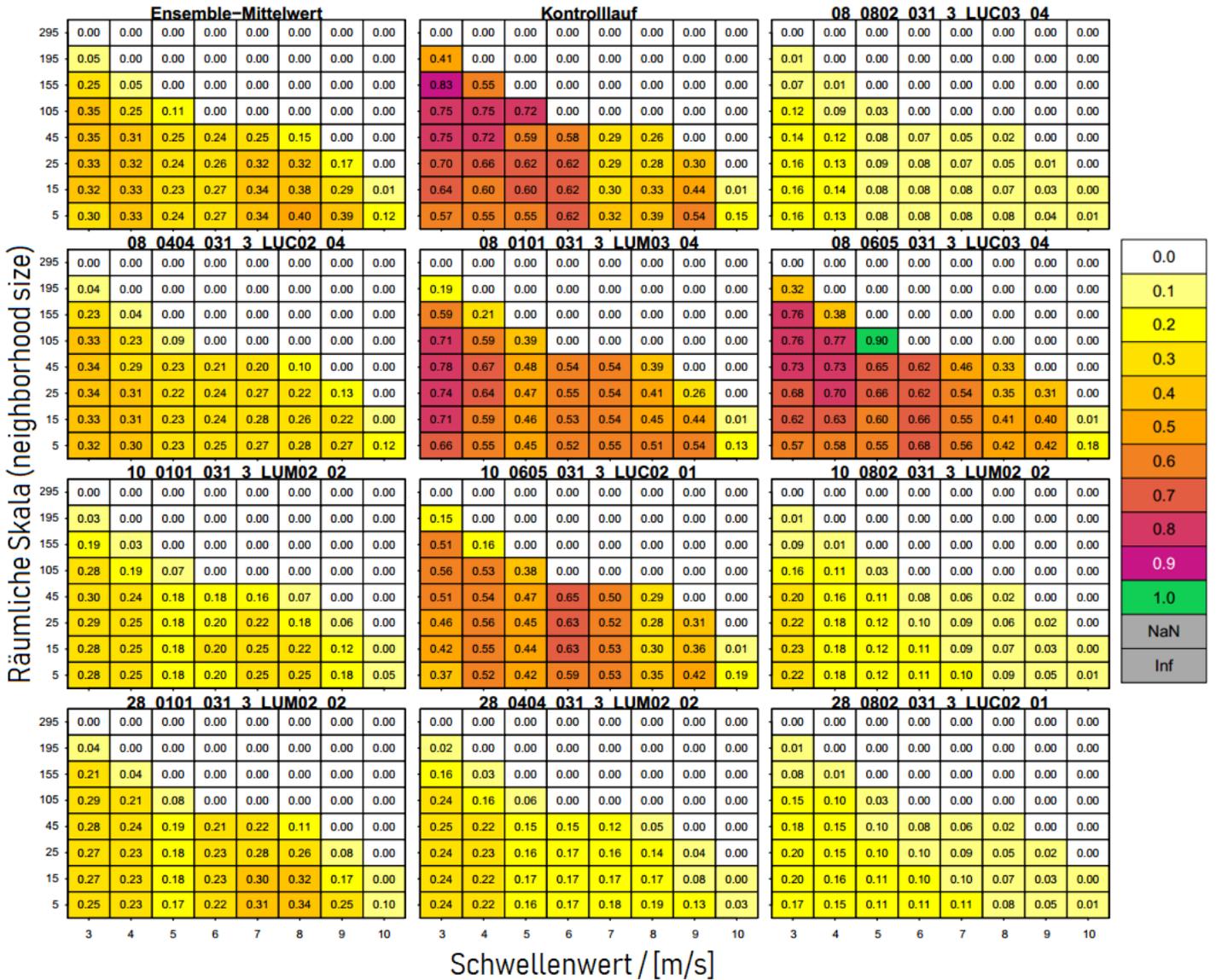


Abbildung 4.8: Equitable Threat Score (ETS) für die Windgeschwindigkeit für den Vorhersagezeitpunkt: 10.01.2017 00 +30-33 h (11.01.2017 06-09 UTC). Links oben: Mittelwert der Ensemblevorhersage, mitte oben: Kontrolllauf, die weiteren Matrizen sind die Ensemblemember der Vorhersage. Die x-Achse beschreibt den Schwellenwert in m/s, die y-Achse gibt die räumliche Skala an. Der Index weist jedem Wert des ETS in der Matrix eine Farbe zu: weiß bedeutet kein Skill ($ETS \leq 0$), grün ist eine (fast) perfekte Vorhersage ($0.9 < ETS \leq 1$), die Werte zwischen den Extremfällen sind mit gelb schwach und mit violett gut.

Anhand der Häufigkeitsverteilung sieht man, dass die häufigsten Windgeschwindigkeiten des Fallbeispiels zwischen 3 und 5 m/s liegen. Höhere Windgeschwindigkeiten sind daher seltener oder treten in der gesamten Domäne nicht auf. Möglicherweise liegt auch nur ein

ETS Differenz d02-d01: Windgeschwindigkeit vom 10.01.2017 + 30-33h

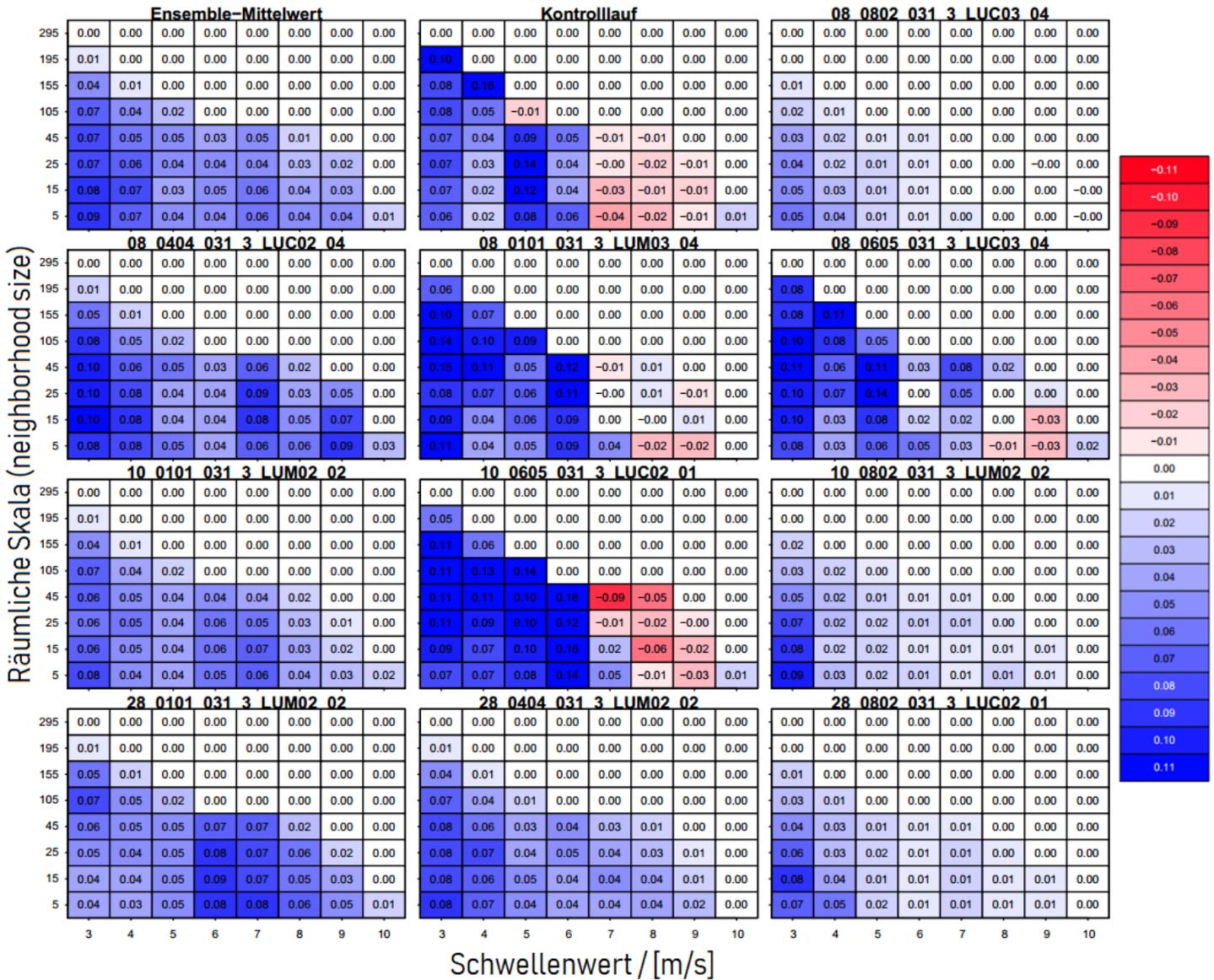


Abbildung 4.9: Differenz des Equitable Threat Scores (ETS) der Windgeschwindigkeit für den Vorhersagezeitpunkt 10.01.2017 00 +30-33 h (11.01.2017 06-09 UTC) zwischen der hochaufgelösten Domäne d02 und der gröberen Domäne d01. Die ETS-Werte der Domäne d01 werden von d02 abgezogen. Sind die ETS-Werte von d02 größer als d01, ist die Differenz positiv. Sind die ETS-Werte von d02 kleiner als d01, ist die Differenz negativ. Links oben: Mittelwert der Ensemblevorhersage, mitte oben: Kontrolllauf, die weiteren Matrizen sind die Ensemblemember der Vorhersage. Die x-Achse beschreibt den Schwellenwert in m/s, die y-Achse gibt die räumliche Skala an. Ist die Differenz positiv (blau), so schneidet d01 bei der Beurteilung durch den ETS besser ab. Bei negativen Differenzen (rot) besitzt die Domäne d02 niedrigere FSS-Werte.

Frequency Bias: d02 Windgeschwindigkeit vom 10.01.20217 + 30-33h

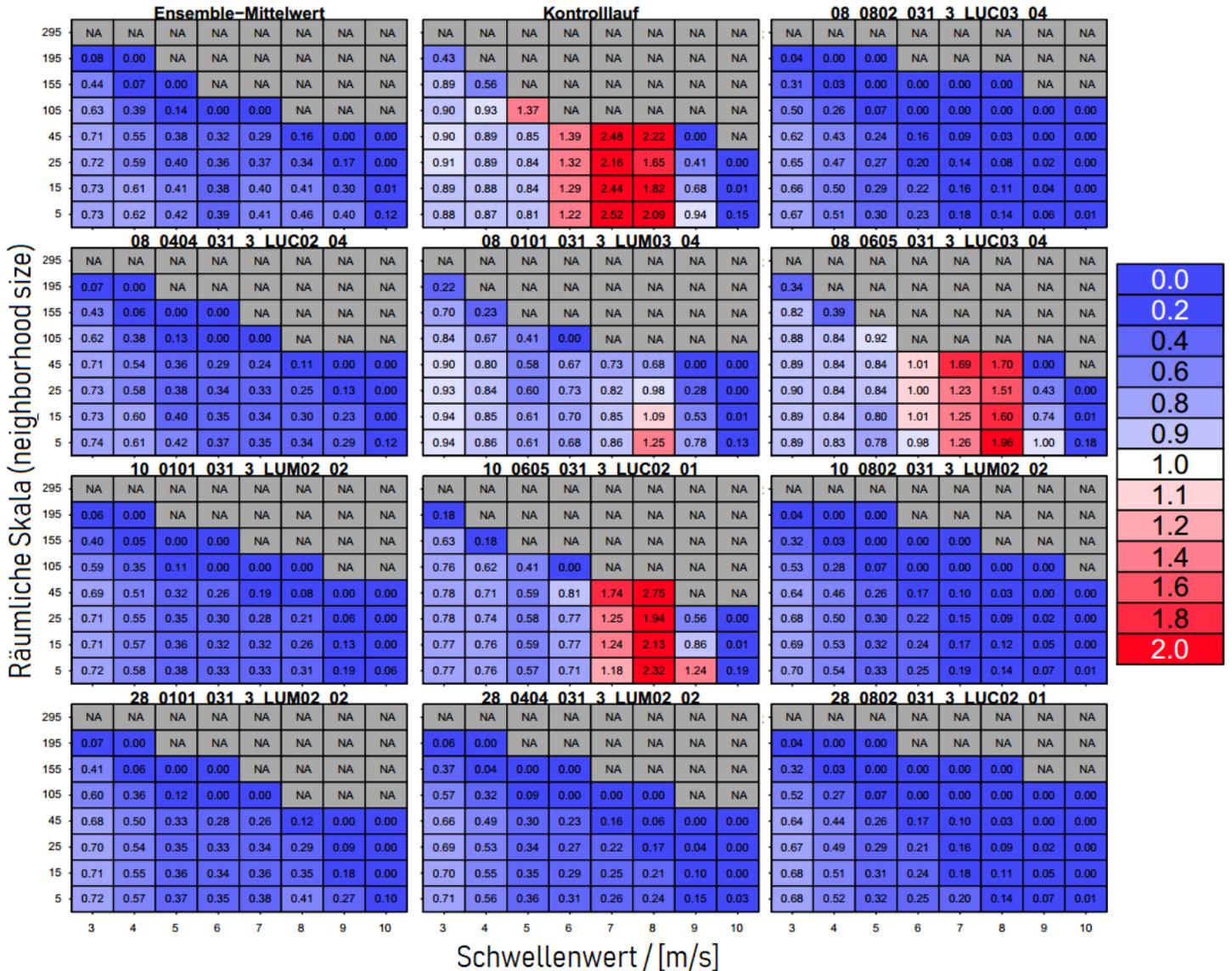


Abbildung 4.10: Frequency Bias der Windgeschwindigkeit für den Vorhersagezeitpunkt: 10.01.2017 00 +30-33 h (11.01.2017 06-09 UTC). Links oben: Mittelwert der Ensemblevorhersage, mitte oben: Kontrolllauf, die weiteren Matrizen sind die Ensemblemember der Vorhersage. Die x-Achse beschreibt den Schwellenwert in m/s, die y-Achse gibt die räumliche Skala an. Der Wert 1 entspricht einer perfekten Vorhersage (weiß), ≤ 1 entspricht einer zu seltenen Vorhersage (blau), ≥ 1 bedeutet zu oft vorhergesagt (rot).

Vorhersagewert innerhalb einer räumlichen Skala vor. Wird dieses Signal richtig vorhergesagt, handelt es sich meist um einen Zufallstreffer, der den Frequency Bias in diesem Fall stärker beeinflusst. Die Bewertung durch den Frequency Bias ist für höhere Schwellenwerte daher nicht aussagekräftig.

Die Differenzen der ETS-Werte weisen ein ähnliches Muster wie der Frequency Bias auf. Ein Unterschied zwischen ETS und Frequency Bias ist die Berücksichtigung von *false alarm* Ereignissen. Erhöht sich die Anzahl der nichteingetretenen Prognosen, so nimmt der Zähler des Frequency Bias und der Nenner des ETS zu. Der Frequency Bias wird höher und das Maß neigt dazu größer als 1 zu werden. Das Ereignis wird zu oft vorhergesagt. Der ETS wird kleiner und die Wahrscheinlichkeit steigt, dass der ETS einer anderen Domäne größer ist. Die Differenz neigt dazu negativ zu werden.

4.3 Fractions Skill Score

Der Fractions Skill Score (FSS) ist ein probabilistische Ansatz und liefert Informationen, wie sinnvoll die Vorhersage für eine bestimmte Skala ist. Abbildung 4.11 zeigt den FSS für die Windgeschwindigkeit der Domäne d02 des Vorhersagelaufs vom 10.01.2017 00 UTC für den Mittelwert der Vorhersagestunden 30 bis 33 (11.01.2017 06-09 UTC). Das Fehlermaß wird wie zuvor als ein Intensity-Scale Diagramm dargestellt. Zusätzlich wird die Vorhersage durch die Skillful Scale beurteilt (vgl. Abb. 3.4 und Gleichung 3.7). Die für die Vorhersage sinnvolle Skalen liegen oberhalb der Skillful Scale. Liegt der Wert bei niedrigen räumlichen Skalen, so wird die Vorhersage für feinen Skalen als sinnvoll bewertet. Für jeden Schwellenwert liegen die höchsten FSS-Werte bei niedrigen Skalen. Hohe Werte bewerten die Vorhersage gut (1 = perfekte Vorhersage). Folglich wird die Vorhersage mit zunehmender räumlichen Skala besser bewertet. Erneut ergeben sich drei unterschiedliche Fälle. Bei einigen Ensemblemember nimmt der Skill kontinuierlich mit zunehmendem Schwellenwert ab, wodurch die besten FSS-Werte in der linken oberen Ecke der Matrix liegen und nach rechts unten abnehmen. Diese Gruppe ist geprägt von niedrigeren FSS-Werten und einer eher schlechteren Vorhersage. Zusätzlich zeigt die Kurve der Skillful Scale bei einem Schwellenwert von 5 m/s steil nach oben. Das bedeutet, die Vorhersage ist nur bei niedrigen Schwellenwerten sinnvoll. Zu dieser Gruppe zählen jene Ensemblemember, die bereits durch andere Verifikationsmethoden schlechter bewertet wurden. Zusätzlich kann der Ensemblemember 10_0101_031_3_LUM02_02 auf Grund der Skillful Scale zu dieser Kategorie gezählt werden, allerdings handelt es sich um einen Grenzfall. Zur zweiten Variante zählen der Mittelwert der Ensemblemember und die Member 08_0404_031_3_LUC02_04 und 28_0101_031_3_LUM02_02. Auf den ersten Blick unterscheiden sie sich zum ersten Muster wenig. Allerdings nehmen die FSS-Werte mit steigendem Schwellenwert weniger rasch ab. Diese Eigenschaft spiegelt sich in der Skillful Scale wieder, da die Vorhersage auch für feinere Skalen bei höheren Schwellenwerten als sinnvoll gilt. Das dritte Muster wird durch den Kontrolllauf und den restlichen Ensemblemember beschrieben und ist geprägt von höheren FSS-Werte. Die Ensemblemember dieser Gruppe erhalten eine bessere Bewertung.

Für Variante 2 und 3 nimmt das Fehlermaß bei höheren Schwellenwerten wieder etwas zu. Der FSS liefert beispielsweise für den Ensemblemember 10_0605_031_3_LUC02_01 höhere FSS-Werte bei einem Schwellenwert von 9 m/s, als bei 8 m/s. Die Vorhersage dieser Ensemblemember besitzt bei mittleren Windgeschwindigkeiten (zwischen 5 und 7 m/s) einen

Fractions Skill Score: d02 Windgeschwindigkeit vom 10.01.2017 + 30-33h

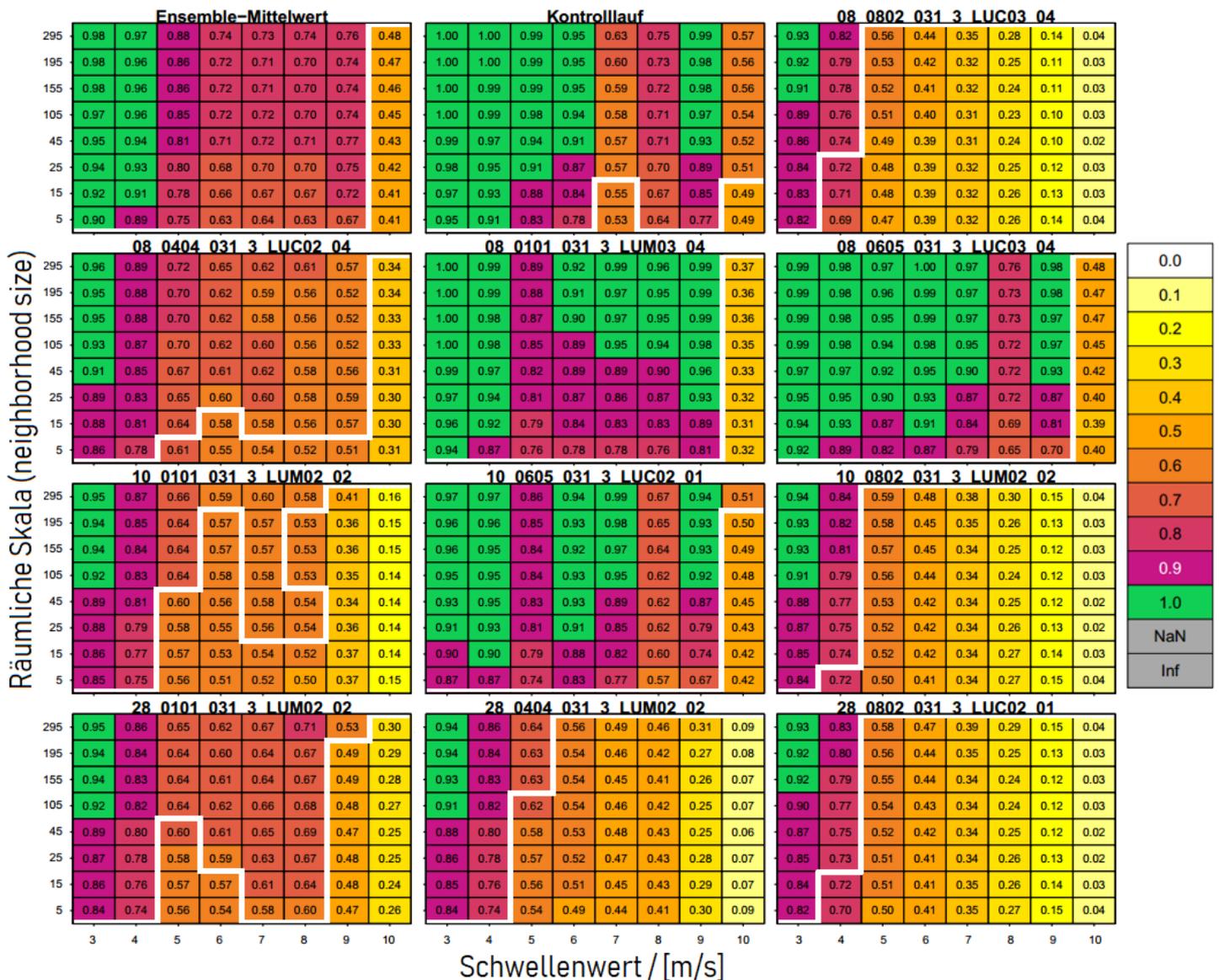


Abbildung 4.11: Fractions Skill Score der Windgeschwindigkeit für den Vorhersagezeitpunkt: 10.01.2017 00 +30-33 h (11.01.2017 06-09 UTC). Links oben: Ensemblemittelwert, mitte oben: Konrolllauf, die weiteren Matrizen stellen die Ensemblemember dar. Die x-Achse beschreibt den Schwellenwert in m/s, die y-Achse gibt die räumliche Skala an. Der Wert 0 (weiß) bedeutet kein Skill, 0.91-1 (grün) entspricht einer (fast) perfekten Vorhersage. Der Skillful Scale (weiß) ist als Linie eingetragen, Skalen unterhalb der Kurve gelten als nicht sinnvoll.

niedrigeren Skill. Wird die Skillful Scale für diese Ensemblemember betrachtet, sind die Schwankungen der FSS-Werte nebensächlich. Von Interesse ist nur, welche Skalen oberhalb der kritischen Kurve liegen. Für die beschriebene zweite und dritte Variante befinden sich nahezu alle Skalen über der Linie. Bei einem Schwellenwert von 10 m/s werden nur beim

Kontrolllauf sinnvolle Skalen angegeben.

4.3.1 Diskussion

Anhand der Muster des Fractions Skill Score lassen sich, wie beim ETS, die Ensemblemember in die drei gleichen Gruppen zusammenfassen. Allerdings fällt für das Fallbeispiel vom 10.01.-12.01.2017 ein Ensemblemember aus dem Schema heraus (vgl. 10_0101_031_3.LUM02_02). Im Großteil der Fälle bleibt die Gruppierung jedoch bestehen. Der Vergleich mit der Häufigkeitsverteilung und dem Bias zeigt in den Ensemblemember ein ähnliches Verhalten. Daraus lässt sich schließen, dass der Fractions Skill Score für die Verifikation von Windgeschwindigkeiten geeignet ist.

Höhere Windgeschwindigkeiten sind seltener. Mit zunehmendem Schwellenwert nimmt die Häufigkeit der Ereignisse ab. Dadurch wird die Vorhersage bei hohen Schwellenwerten häufig schlechter bewertet. Ähnlich zum ETS nimmt auch der Skill der Vorhersage mit zunehmendem Schwellenwert ab. Der FSS bewertet generell höhere Skalen besser als feinere. Allerdings nimmt der FSS des Kontrolllaufs und andere Ensemblemember bei mittleren Schwellenwerten ab, um bei höheren Werten wieder zuzunehmen. Die Verschlechterung wird tragend, wenn auch die Skillful Scale in diesem Bereich zunimmt. Abbildung 4.11 zeigt, dass die Kurve der Skillful Scale für den Kontrolllauf sehr flach ist. Es gilt, je flacher die Kurve der Skillful Scale, desto besser wird die Vorhersage bewertet. Folglich wird der Kontrolllauf auch für hohe Schwellenwerte gut bewertet.

Neben dem Kontrolllauf zeigen auch anderer Ensemblemember einen niedrigeren Skill für die mittleren Windgeschwindigkeiten. Möglicherweise wird die Lage der Front falsch vorhergesagt, was sich auf die mittleren Windgeschwindigkeiten auswirkt. Der Equitable Threat Score liefert eine ähnliches Bild, vor allem der Schwellenwert von 5 m/s zeigt niedrigere ETS-Werte.

4.4 Structure-Amplitude-Location

4.4.1 Niederschlag

Die SAL-Technik wurde für die Verifikation von Niederschlagsobjekten entwickelt. Niederschlagsereignisse sind natürliche Objekte mit einer klar definierten unteren Grenze von 0 mm. Für die Verifikation werden Spuren von Niederschlag aus den Objekten entfernt, indem ein Schwellenwert von 0.1 mm gewählt wird. Abbildung 4.12 zeigt die SAL-Komponenten des 6-stündigen akkumulierten Niederschlags des Vorhersagelaufs vom 10.01.2017 00 UTC der hochaufgelösten Domäne d02. Die Verifikationsergebnisse von Struktur, Amplitude und räumlicher Verschiebung werden als Dichteplot dargestellt. Die Streuung der Ensemblemember bildet die Varianz. Das heißt, ein geringer Spread zwischen den Ensemblemember führt zu einem kleineren Interquartilsabstand des Boxplots. Im Gegensatz zu den Neighborhood Methoden können die einzelnen Ensemblemember auf

Grund der Darstellung nicht mehr voneinander unterschieden werden. Der Kontrolllauf und die Komponenten der Ensembleversion eSAL werden getrennt vom Dichteplot als Linien eingetragen. Die Darstellung ist an Radanovics et al. (2018) angelehnt. Struktur und

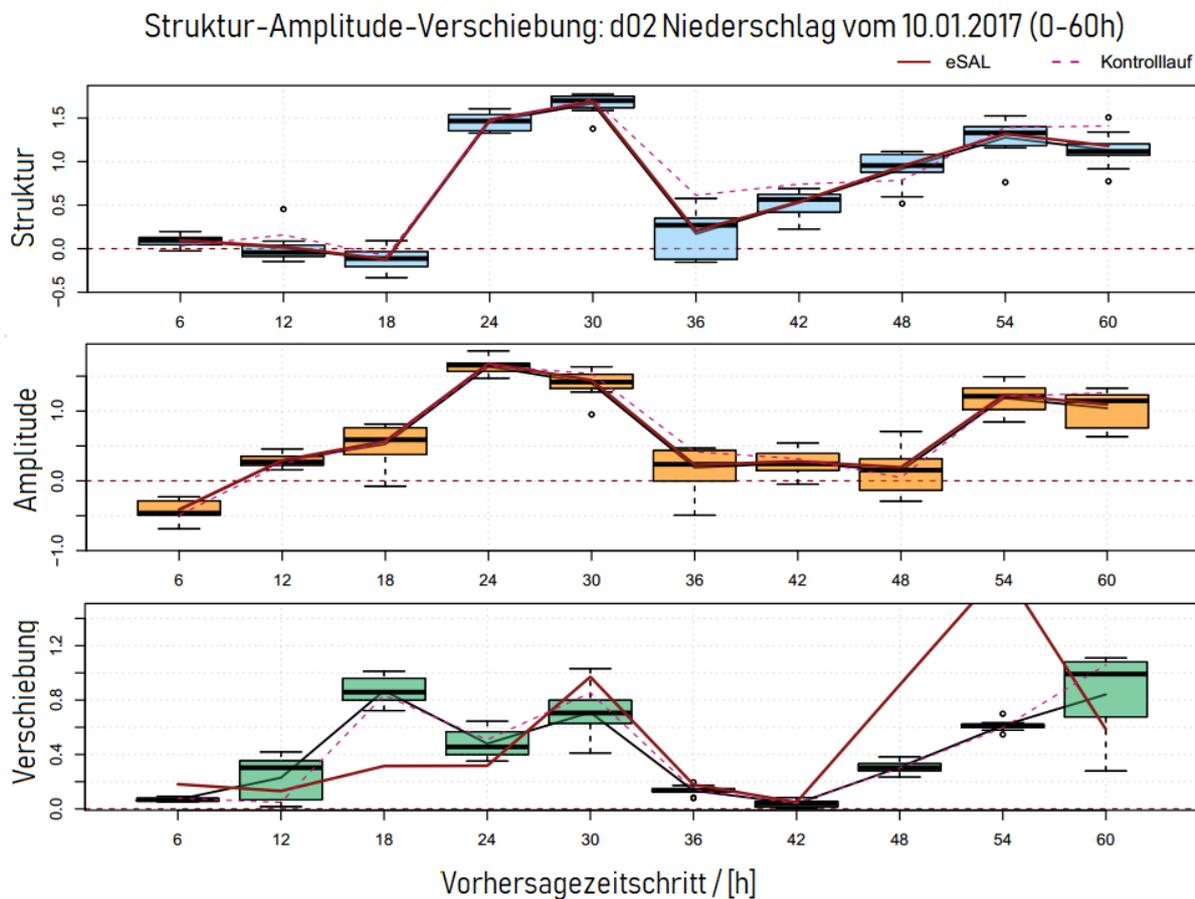


Abbildung 4.12: SAL-Komponenten für den 6-stündigen akkumulierten Niederschlag des Vorhersagelaufs vom 10.01.2017 00 UTC (10.01. 00 UTC - 12.01.2017 12 UTC). Die Ensemblemember werden durch den Whiskersplot repräsentiert. Der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt behandelt. Oben ist die Struktur, in der Mitte die Amplitude, unten die räumliche Verschiebung abgebildet. Der Schwellenwert liegt bei 0.1 mm.

Amplitude sind fast immer positiv. Das heißt, die Niederschlagsobjekte werden in den meisten Fällen zu großflächig vorhergesagt und die Intensität überschätzt. Beide Komponenten werden über einen Mittelwert bestimmt (vgl. Gleichung 3.9 und 3.12). Daher können die Ergebnisse des SAL mit der mittleren Niederschlagsmenge (vgl. Abb. 4.6) für den Zeitraum des Vorhersagelaufs vom 10.01.2017 00 UTC verglichen werden.

In den Vorhersagestunden 18-30 vor dem Frontdurchgang liegen die gemittelten Niederschlagswerte der Vorhersage unterhalb des 60-stündigen Mittelwerts der Analysedaten des VERA Systems (Zeitraum 10.01. 00 UTC - 12.01. 12 UTC). Für die Amplitude und die

Struktur ergeben sich höhere Werte und die Vorhersage wird bezüglich Form und Intensität schlechter bewertet. Im Vergleich zu den anderen Zeitpunkten ist der Ensemblespread und der Bias zur Analyse geringer. Daraus lässt sich schließen, dass falsche Vorhersagen bei niedrigen Niederschlagsmengen stärker bestraft werden als bei hohen Niederschlagsmengen. Ein ähnliches Ergebnis ergibt sich, wenn die mittlere Niederschlagsmenge gegen Ende des Vorhersagelaufs für die Vorhersagestunden 54 bis 60 (12.01. 06-12 UTC) wieder abnimmt. Neben der Form und Intensität wird innerhalb dieser beiden Zeiträume die Position der Niederschlagsobjekte ebenfalls schlechter bewertet.

Die Vorhersage ist perfekt, wenn alle drei Komponenten Null sind. Das wird am ehesten in den Vorhersagestunden 12 (10.01. 12 UTC) und 36 (11.01. 12 UTC) erreicht. Der zweite Fall entspricht der Zeit nach dem Frontdurchgang, welcher mit den Neighborhood Methoden verifiziert wurde.

Struktur und Amplitude der Ensembleversion und des Medians der Standardmethode besitzen ähnliche Werte. Die räumliche Verschiebung des eSAL ist größer als der Median des SAL, wenn die mittlere Niederschlagsmenge gering ist.

4.4.2 Windgeschwindigkeit

Zunächst muss die Windgeschwindigkeit des 10 m-Windes als Objekt definiert werden, ansonsten fließen alle Werte in der gesamten Domäne in die Verifikation mit ein. Die räumliche Verschiebung kann sich nur mehr am Massenschwerpunkt des Objekts orientieren. In der Berechnung der räumlichen Verschiebung wird nur die Komponente L1 berücksichtigt, L2 wird sehr klein oder Null. Die Struktur und die Amplitude berechnen das Mittel über die gesamte Domäne. Das detailliertere Windgeschwindigkeitsfeld von Vorhersage und Analyse wird geglättet. Folglich kann eine Vorhersage zu gut bewertet werden. Das SAL verlangt, dass das zu verifizierende Wetterereignis als Objekt vorliegt. Die Windgeschwindigkeit hat keine klar definierten Grenzen. Zur Definition des Objekts dient ein Schwellenwert. Abbildung 4.13 zeigt verschiedene Objekte der Windgeschwindigkeit der Domäne d02 auf Grund unterschiedlicher Schwellenwerte (0 m/s, 3 m/s, 5 m/s, 7 m/s). In der Grafik wird die Windgeschwindigkeit als ein zusammenhängendes Objekt dargestellt. Die höchsten Windgeschwindigkeiten befinden sich im Nordwesten der Domäne, die niedrigsten im Südosten. Mit höherem Schwellenwert wird das Objekt kleiner und zerfällt in mehrere Objekte. Das Objekt des Ensemblemember verändert kaum seine Größe und Form. Dieser Ensemblemember weist generell höhere Windgeschwindigkeiten auf. In der Häufigkeitsverteilung (Abb. 4.7) treten Windgeschwindigkeiten bis 14 m/s auf. Abbildung 4.14 zeigt die Struktur des Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02 für drei Schwellenwerte 3 m/s, 5 m/s und 7 m/s. Die Struktur liefert Informationen über die Größe und Form des Objekts. Abbildung 4.15 zeigt die Amplitude für den gleichen Vorhersagelauf und der gleichen Domäne mit den gleichen Schwellenwerten. Die Amplitude gibt Information über die Intensität der Objekte. Die Werte sind für alle drei Fälle fast immer positiv. Laut Definition der Amplitude bedeuten positive Werte, dass die Vorhersage die

Objekte der Windgeschwindigkeit d02 vom 10.01.2017 +38h

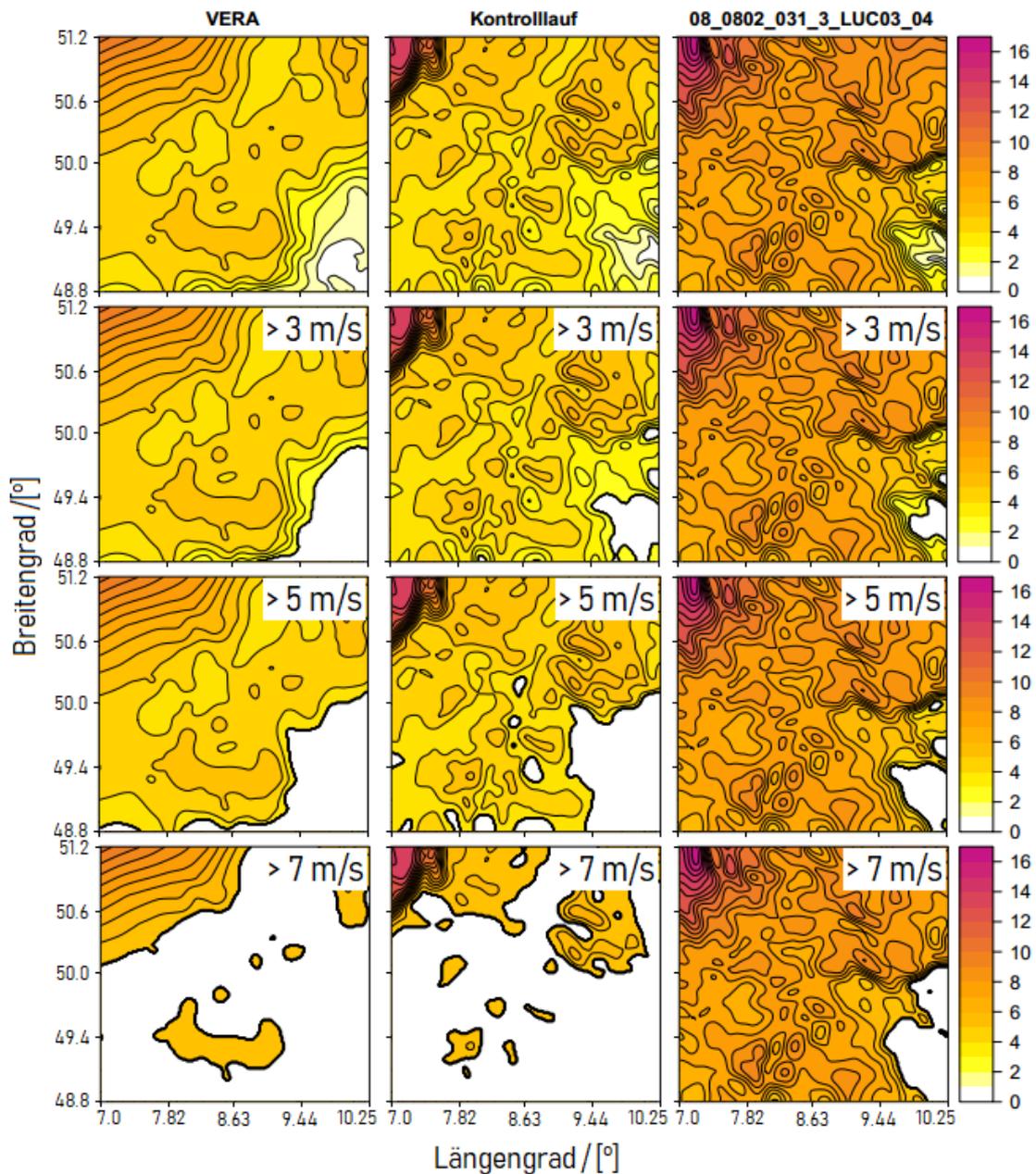


Abbildung 4.13: Objekte der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC für die 38. Vorhersagestunde (entspricht 11.01. 14 UTC). Links sind die Analysedaten des VERA Systems dargestellt, in der Mitte der Kontrolllauf der WRF Ensemblevorhers, recht ein beliebiger Ensemblemember. Die obere Reihe sind die interpolierten Windgeschwindigkeiten, die anderen Objekte werden mit einem Schwellenwert von 3 m/s, 5 m/s und 7 m/s definiert.

Struktur: d02 Windgeschwindigkeit vom 10.01.2017 0-60 h

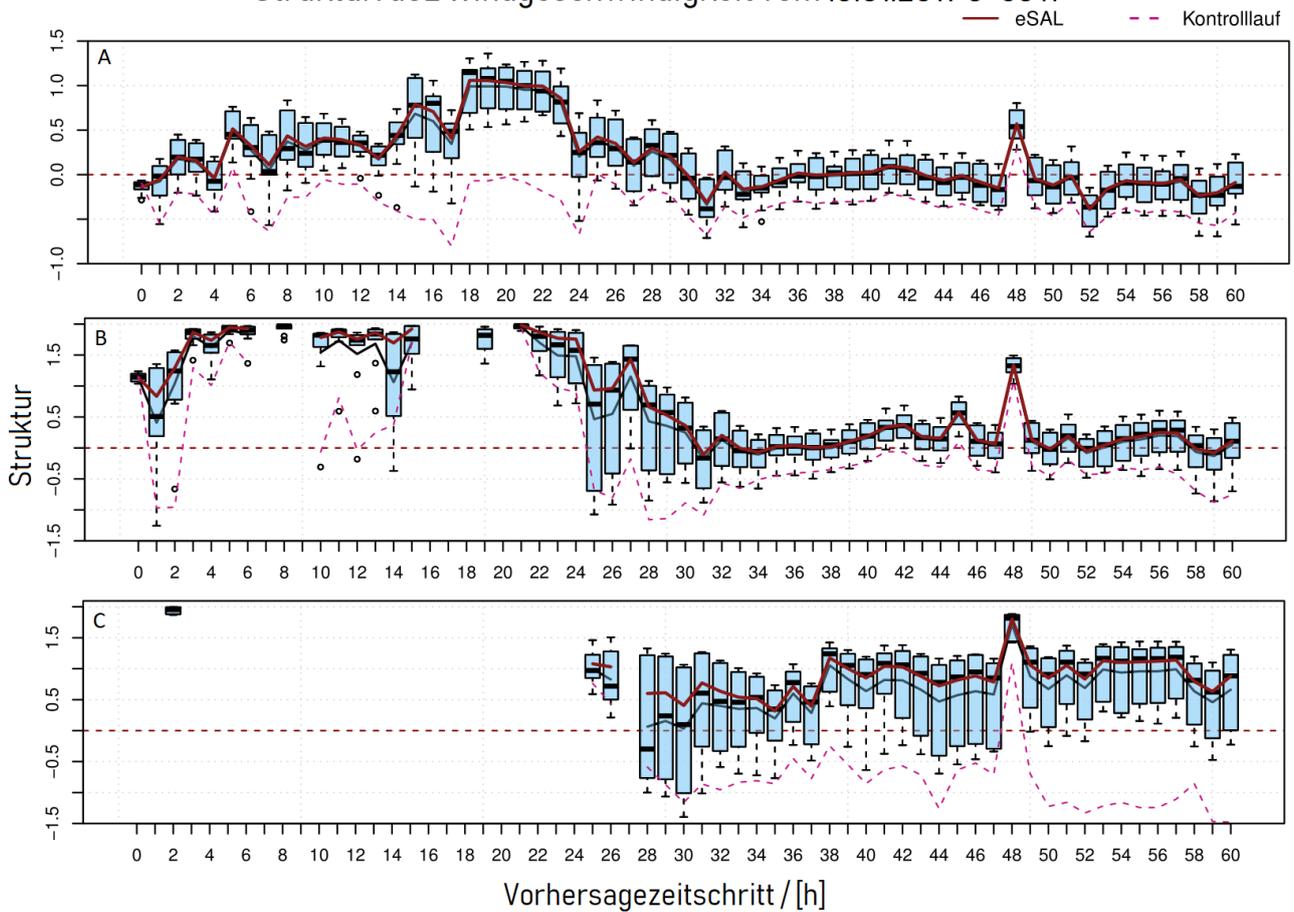


Abbildung 4.14: Struktur des SAL der Windgeschwindigkeit für verschiedene Schwellenwerte von 3 m/s, 5 m/s und 7 m/s als Zeitreihe des Vorhersagelaufs vom 10.01.2017 00 UTC. Die Ensemblemember werden durch den Whiskersplot repräsentiert, der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt dargestellt.

Windgeschwindigkeit überschätzt. SAL von +1 bedeutet, dass die Vorhersage um den Faktor 3 überschätzt wird. Die Struktur und Amplitude werden als Mittelwerte definiert. Daher können die Komponenten mit dem Mittelwert der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC aus Abbildung 4.4 verglichen werden. Der mittlere Wind weht zunächst schwach (etwa 4 m/s). Die Ergebnisse von Amplitude und Struktur sind meist positiv, somit werden Größe und Intensität des Windobjekts überschätzt. Besitzt ein Objekt viele Werte nahe des Schwellenwertes, dann führen bereits kleine Unterschiede zu einer Änderung der Form und Intensität des Objekts. Dann sind die Ergebnisse variabel. Mit dem Frontdurchgang (11.11. 06-09 UTC) nimmt die Windgeschwindigkeit zu und bleibt erhöht (11.01. ab 09 UTC). Das Objekt wird durch die durchziehende Front definierter und die Form und Intensität der Objekte stimmen besser überein. Die Werte der Amplitude bleiben dennoch positiv. Das heißt, die Vorhersage überschätzt die Windgeschwindigkeit, die Größe des Objekts wird besser

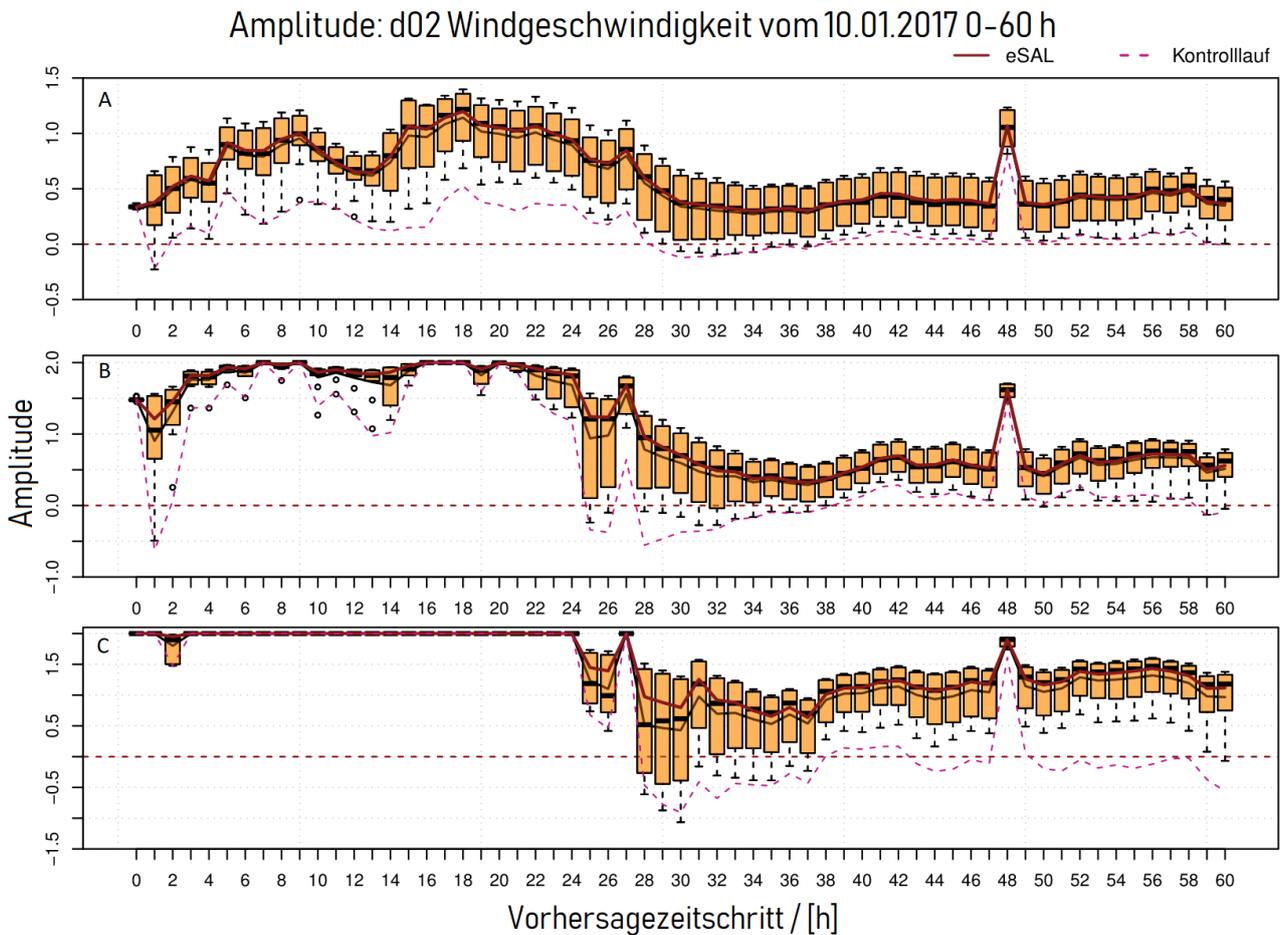


Abbildung 4.15: Amplitude des SAL der Windgeschwindigkeit für verschiedene Schwellenwerte von 3 m/s, 5 m/s und 7 m/s als Zeitreihe des Vorhersagelaufs vom 10.01.2017 00 UTC. Die Ensemblemember werden durch den Whiskersplot repräsentiert, der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt dargestellt.

bewertet.

Mit zunehmendem Schwellenwert verschieben sich die Werte der Struktur und Amplitude weiter in den positiven Bereich. Werden auf Grund des Schwellenwerts die Objekte kleiner, vergrößert sich der Interquartilsabstand der Boxen der Struktur. Die Objekte der Ensemblemember unterscheiden sich stärker voneinander und damit auch die Bewertung bezüglich der Größe des Objekts. Mit zunehmenden Schwellenwert steigt die Wahrscheinlichkeit, dass sich kein Objekt im Analyse- und oder Vorhersagefeld befindet. Die Struktur liefert keine Werte. Bei fehlendem Vergleichsobjekt hat die Amplitude ihren maximalen Wert. Der Wert ist positiv, wenn kein Objekt im Analysefeld vorliegt (Überschätzung), und negativ, wenn im Vorhersagefeld kein Objekt vorhanden ist (Unterschätzung).

Die Anzahl an Beobachtungstationen, die in die Analyse einfließen, haben einen Einfluss auf

die analysierte Windgeschwindigkeit. Bei einer geringeren Stationsdichte werden niedrigere Werte berechnet. Der negative Ausreißer in der 48. Vorhersagestunde des Vorhersagelaufs vom 10.01.2017 (entspricht der Analyse vom 12.01.2017 00 UTC) liefert einen höheren Wert in allen Komponenten des SAL (vgl. Abb. 4.14–4.16). Das zeigt die Wichtigkeit einer hohen Beobachtungsdichte für die Datenanalyse. Die Struktur und Amplitude der

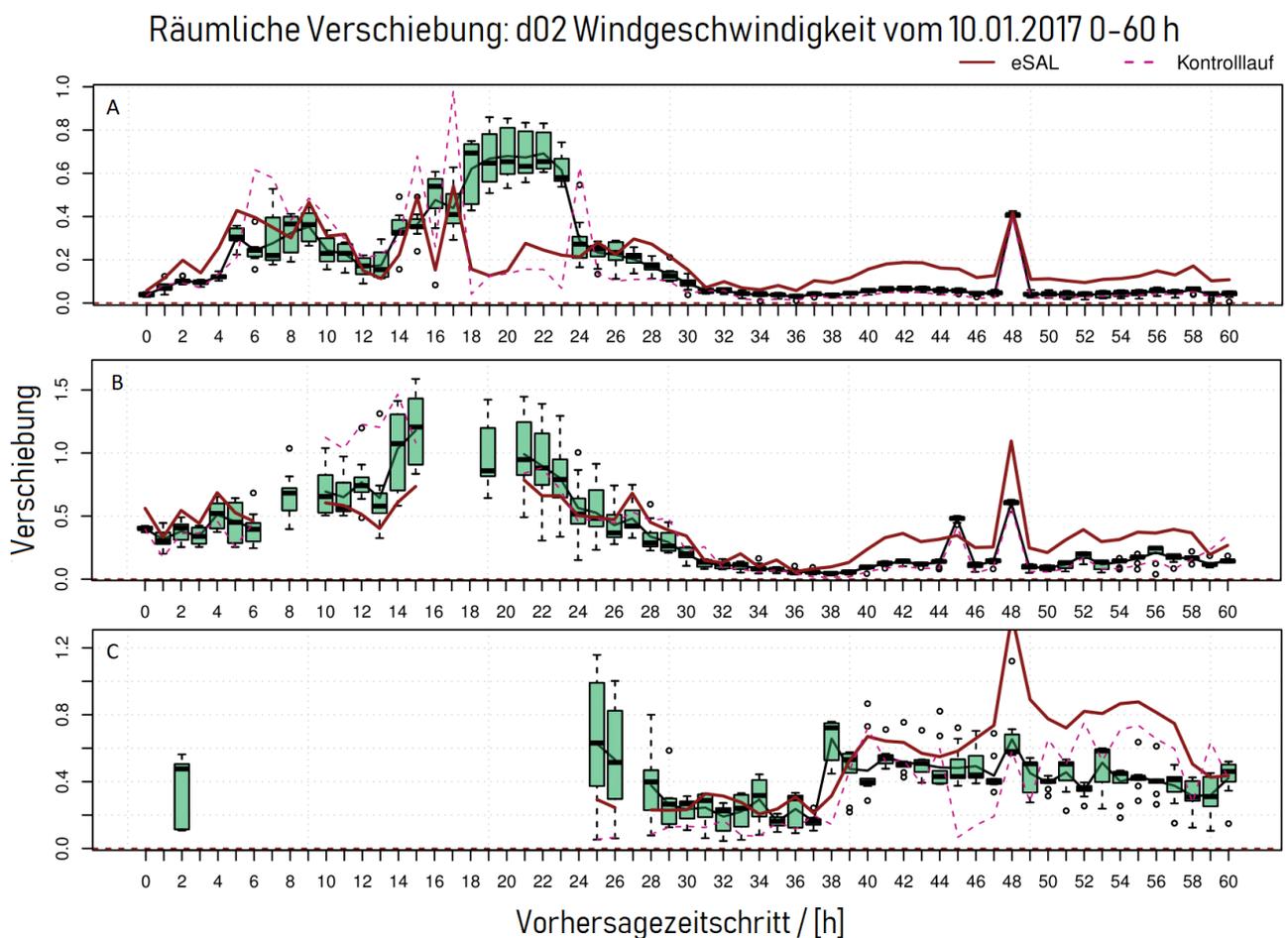


Abbildung 4.16: Räumliche Verschiebung des SAL der Windgeschwindigkeit für verschiedene Schwellenwerte vom 3 m/s, 5 m/s und 7 m/s als Zeitreihe des Vorhersagelaufs von 10.01.2017 00 UTC. Die Ensemblemember werden durch den Whiskersplot repräsentiert, der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt dargestellt.

Ensembleversion nach Radanovics et al. (2018) liegen nahe des Medians der Standardmethode. Bei höheren Schwellenwerten liegen die Komponenten des eSAL knapp oberhalb des Medians. Der Kontrolllauf liegt für beide Komponenten am unteren Rand des Dichteplots. Die Größe des Objekts wird meist zu klein vorhergesagt (Werte kleiner 0). Bezüglich der Amplitude wird Kontrolllauf für alle Schwellenwert am besten bewertet (Werte nahe Null). Auffällig ist, dass die Kurve des Kontrolllaufs sich mit zunehmendem Schwellenwert kaum verändert. Allerdings kehrt sich sein Vorzeichen um und die Intensität der Windgeschwindigkeit wird für einen Schwellenwert von 7 m/s leicht unterschätzt.

Abbildung 4.16 zeigt die räumliche Verschiebung des Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02 für drei Schwellenwerte 3 m/s, 5 m/s und 7 m/s. Die räumliche Verschiebung liefert Informationen über den Ort des Objekts. Auf Grund der Definition (vgl. Gleichung 3.14-3.15) kann der Verschiebungsfehler nur positive Werte aufweisen. Wie bereits bei der Struktur und Amplitude sind die Ergebnisse in der ersten Hälfte der Vorhersagezeit variabler. Auf Grund der Änderung der Form der Objekte ändert sich auch die räumliche Lage.

In der zweiten Hälfte des Vorhersagelaufs ergibt sich für niedrige Schwellenwerte nahezu perfekte Ergebnisse mit einer geringen Streuung der Ensemblemember. Die räumliche Verschiebung zwischen Analyse- und Vorhersagefeld ist während und nach dem Frontdurchgang gering. Mit zunehmendem Schwellenwert nimmt auch der Fehler der räumlichen Verschiebung zu.

Die räumliche Verschiebung des eSAL zeigt eine stärkere Abweichung zum Median und der Verteilung der Ensemblemember. Bereits bei der Verifikation von Niederschlagsobjekten mittels der SAL-Technik wurde beobachtet, dass sich die räumliche Verschiebung der Ensembleversion anders verhält, als bei der Struktur und Amplitude.

SAL-Diagramm

Eine weitere Darstellungsmöglichkeit ist das SAL-Diagramm. Es beinhaltet die gleichen Informationen wie die Abbildungen 4.14 bis 4.16, jedoch ist der betrachtete Zeitraum geringer. Das Diagramm unterstreicht Unterschiede zwischen der Standardmethode und der Ensembleversion des SALs.

Abbildung 4.17 stellt das SAL-Diagramm des Vorhersagelaufs vom 10.01.2017 00 UTC für die Vorhersagestunden 28-35 (entspricht der Analyse vom 11.01. 04-11 UTC) dar. Gezeigt werden die mittleren Ergebnisse des Ensemblemember, die Ergebnisse des Kontrolllaufs und des eSAL. Das SAL-Diagramm wurde bereits von Wernli et al. (2008) und später auch von Radanovics et al. (2018) zur Darstellung ihrer Ergebnisse verwendet. Die Amplitude des Kontrolllaufs in Abbildung 4.17 liegt nahe Null, die Intensität der vorhergesagten Windgeschwindigkeit stimmt gut mit den Analysedaten des VERA Systems überein. Das Objekt wird für niedrige Schwellenwerte zu klein dargestellt ($S > 0$) und für hohe Schwellenwerte zu groß ($S < 0$). Die Ergebnisse der Standardmethode und Ensembleversion liegen allesamt im 1. Quadranten. Auf Grund von Abbildung 3.5 folgt, dass das Objekt für alle Schwellenwerte bezüglich Größe und Intensität überschätzt wird. Der größte Unterschied liegt in der räumlichen Verschiebung, wobei die Ensembleversion höhere Werte liefert.

4.4.3 Äquivalentpotentielle Temperatur und reduzierter Luftdruck

Die Objekte der äquivalentpotentiellen Temperatur und des reduzierten Luftdrucks werden über den Mittelwert des Analysefelds definiert. Für jeden Vorhersagezeitschritt wird der Mittelwert neu berechnet und als Schwellenwert verwendet. Das daraus resultierende Objekt des Temperaturmaßes (vgl. Abb. 4.18) beschreibt energiereichere Luftmassen. Für

Struktur-Amplitude-Verschiebung: d02 Windgeschwindigkeit vom 10.01.2017 28-35 h

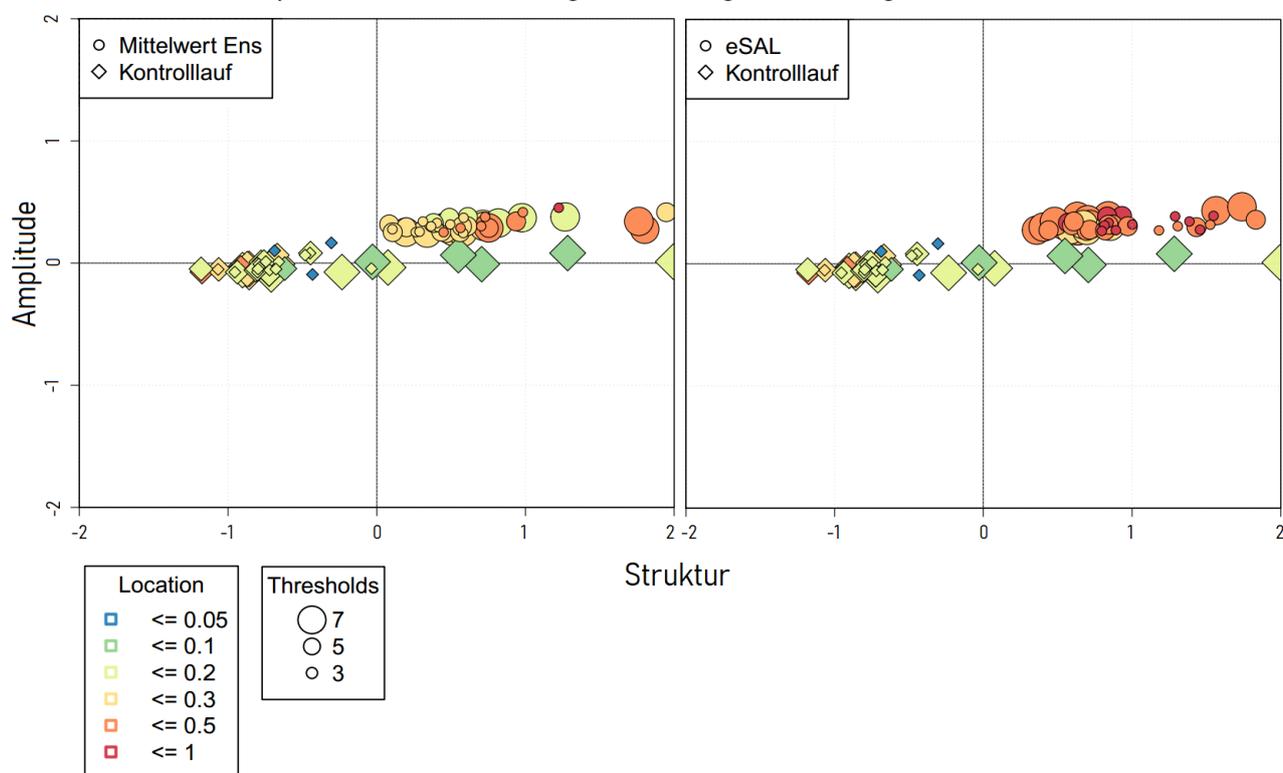


Abbildung 4.17: SAL-Diagramm der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC für die Vorhersagestunden 28 bis 35 (11.01. 04-11 UTC). Die Marker repräsentieren den Mittelwert der Ensemblemembers (links), das eSAL (rechts) und den Kontrolllauf. Die Amplitude wird auf der y-Achse aufgetragen, die Struktur auf der x-Achse, die räumliche Verschiebung durch Farben gekennzeichnet. Die Größe der Punkte zeigt an, welcher Schwellenwert für die Objektbildung verwendet wird. Liegen nun alle Werte im 1. Quadranten, so werden die Objekte in Größe und Intensität überschätzt ($S > 0, A > 0$).

den Vorhersagelauf vom 10.01.2017 00 UTC der Domäne d02 kann das Objekt mit der durchziehenden Front assoziiert werden. Das Objekt des Drucks (vgl. Abb. 4.19) wird durch Werte kleiner als der Mittelwert definiert. Es zeigt das Tiefdruckgebiet, mit dem die Front über Europa zieht. Zusätzlich wird für die Berechnung des reduzierten Luftdrucks der Schwellenwert vom Objekt abgezogen. Eine Front und ein Tiefdruckgebiet zählen auf Grund ihrer Ausdehnung von 200 bis 2000 km nach Orlanski's Skalenanalyse (Orlanski, 1975) zur Mesoskala. Dadurch liegt in einer kleinen Domäne meist nur ein Objekt vor. Will man den tieferen Druck darstellen, ergibt sich bei dieser Vorgehensweise ein Objekt mit negativen Werten. Zur weiteren Berechnung werden die Absolutwerte herangezogen, wodurch sich der niedrigste zum höchsten Wert umkehrt. Wird ein Tiefdruckgebiet verifiziert, ist neben der Lage auch seine Intensität von Interesse. Wegen der Umkehrung des Vorzeichens liefert die Amplitude Informationen über die Stärke des Tiefdruckkerns. In Abbildung 4.20 werden SAL-Komponenten für die äquivalentpotentielle Temperatur

Äquivalentpotentielle Temperatur >288 K vom 10.01.2017 +38 h

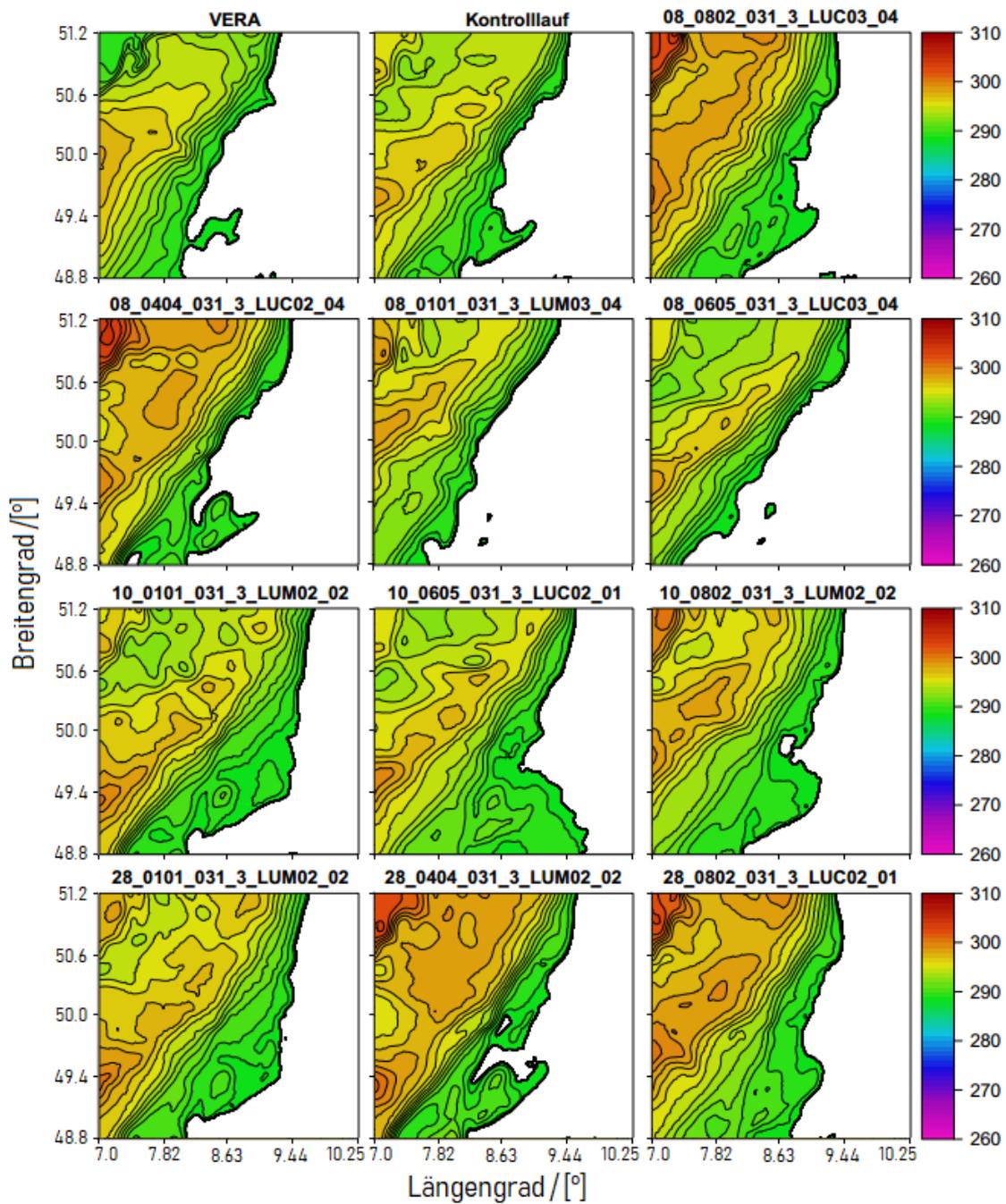


Abbildung 4.18: Objekte der äquivalentpotentiellen Temperatur des Vorhersagelaufs vom 10.01.2017 für die 38. Vorhersagestunde (entspricht der Analyse am 11.01. 14 UTC). Es werden die Analysedaten des VERA Systems, der Kontrolllauf der WRF Ensemblevorhersage und die Ensemblenmember der Domäne d02 mit dem Schwellenwert von 288 K dargestellt.

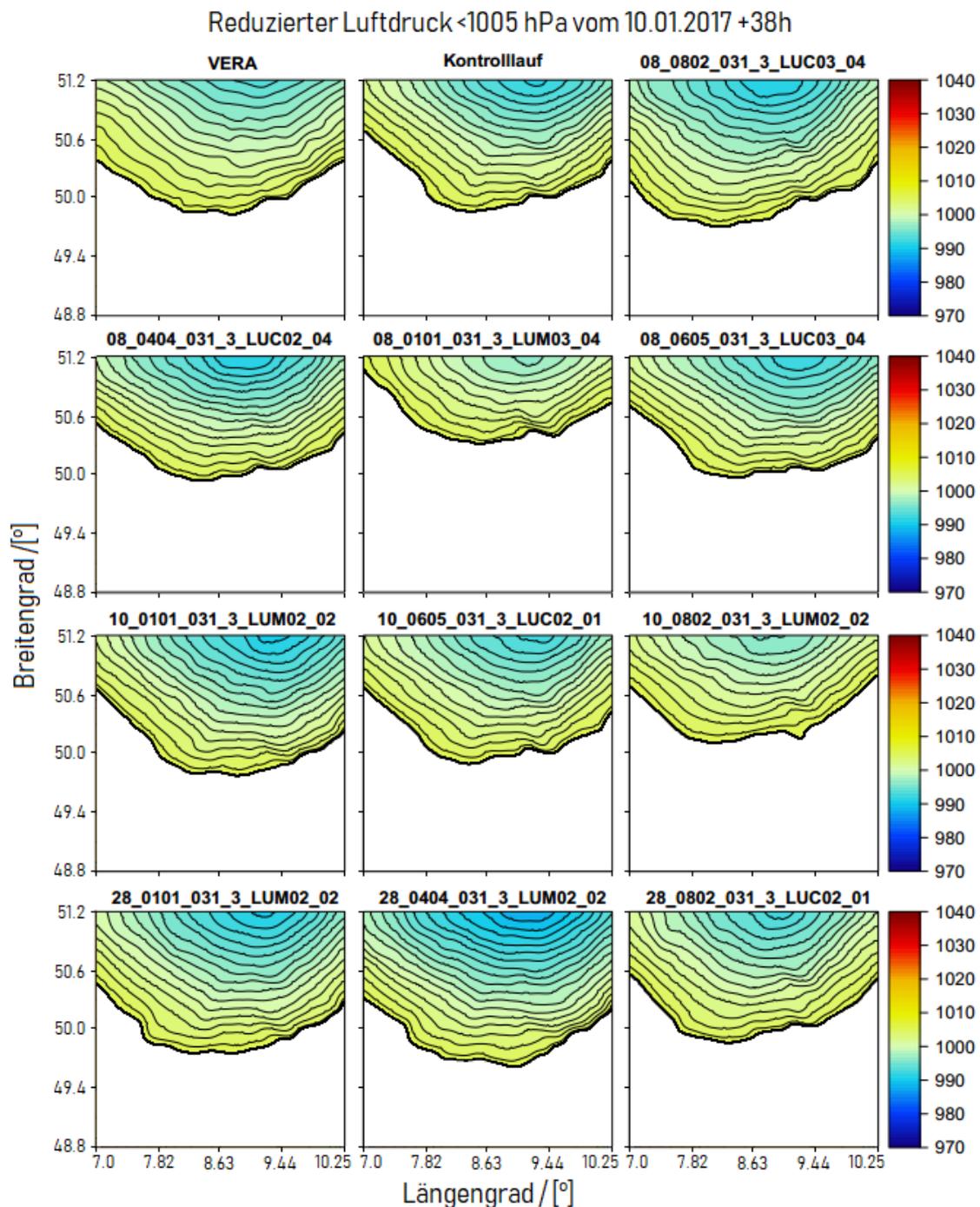


Abbildung 4.19: Objekte des reduzierten Luftdrucks des Vorhersagelaufs vom 10.01.2017 für die 38. Vorhersagegestunde (entspricht der Analyse am 11.01. 14 UTC). Es werden die Analysedaten des VERA Systems, der Kontrolllauf der WRF Ensemblevorhersage und die Ensemblenmember der Domäne d02 mit der oberen Grenze von 1005 hPa dargestellt.

dargestellt. Die Struktur, Amplitude und räumliche Verschiebung zeigen großteils eine

Struktur-Amplitude-Verschiebung: d02 Äquivalentpotentielle Temperatur vom 10.01.2017 0-60 h

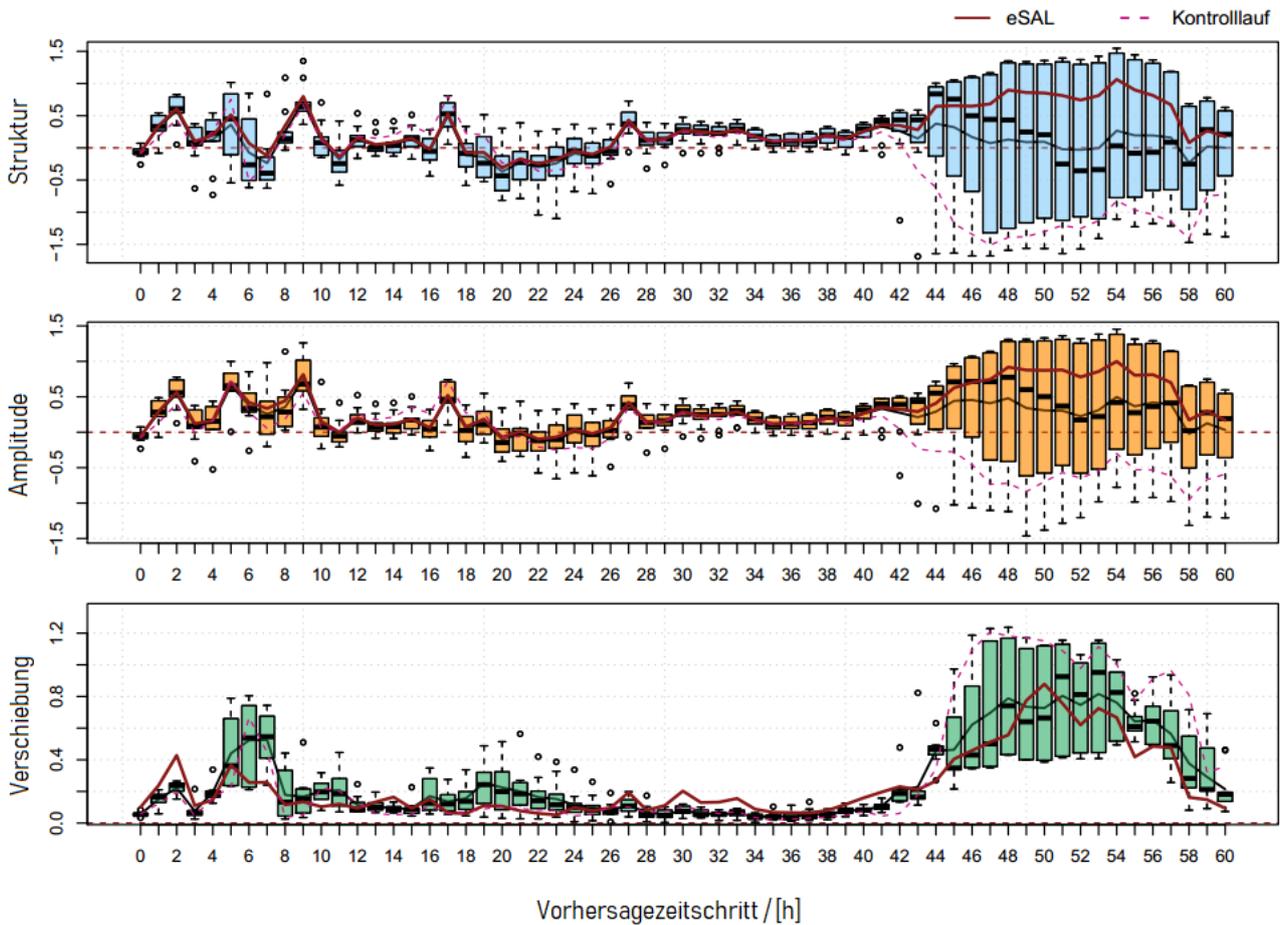


Abbildung 4.20: SAL-Komponenten der äquivalentpotentielle Temperatur des Vorhersagelaufs vom 10.01.2017 00 UTC (10.01. 00 UTC - 12.01.2017 12 UTC). Die Ensemblemember werden durch den Whiskersplot repräsentiert, der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt behandelt. Oben ist die Struktur, in der Mitte die Amplitude, unten die räumliche Verschiebung abgebildet. Der Schwellenwert wird für jeden Vorhersageschritt über den Mittelwert des Analysefeldes bestimmt.

gute Übereinstimmung der Objekte zwischen Analyse und Vorhersage. Der Mittelwert der äquivalentpotentiellen Temperatur in Abbildung 4.6 nimmt für die Vorhersagezeitschritte 10-15 und 30-36 zu (entspricht der Analyse vom 10.01. 10-15 UTC und 11.01. 06-12 UTC) zu. Dann liefern alle drei SAL-Komponenten Werte um Null und eine sehr geringe Unsicherheit der Ensemblemember. Folglich wird der Frontendurchgang, repräsentiert durch das Objekt der äquivalentpotentiellen Temperatur, als eine fast perfekte Vorhersage bewertet. Gegen Ende des Vorhersagelaufs (ab Vorhersagestunde 44) nimmt die Unsicherheit der Ensemblemember stark zu, erkennbar an dem großen Interquartilsabstand des Boxplots. Nach Durchzug der Front weist die Domäne relativ gleiche Werte der äquivalentpotentielle Temperatur auf. Durch den Mittelwert als Schwellenwert entstehen sehr unterschiedliche

Objekte jedes Ensemblemember, die zu größeren Unsicherheiten führen. Wie bereits bei

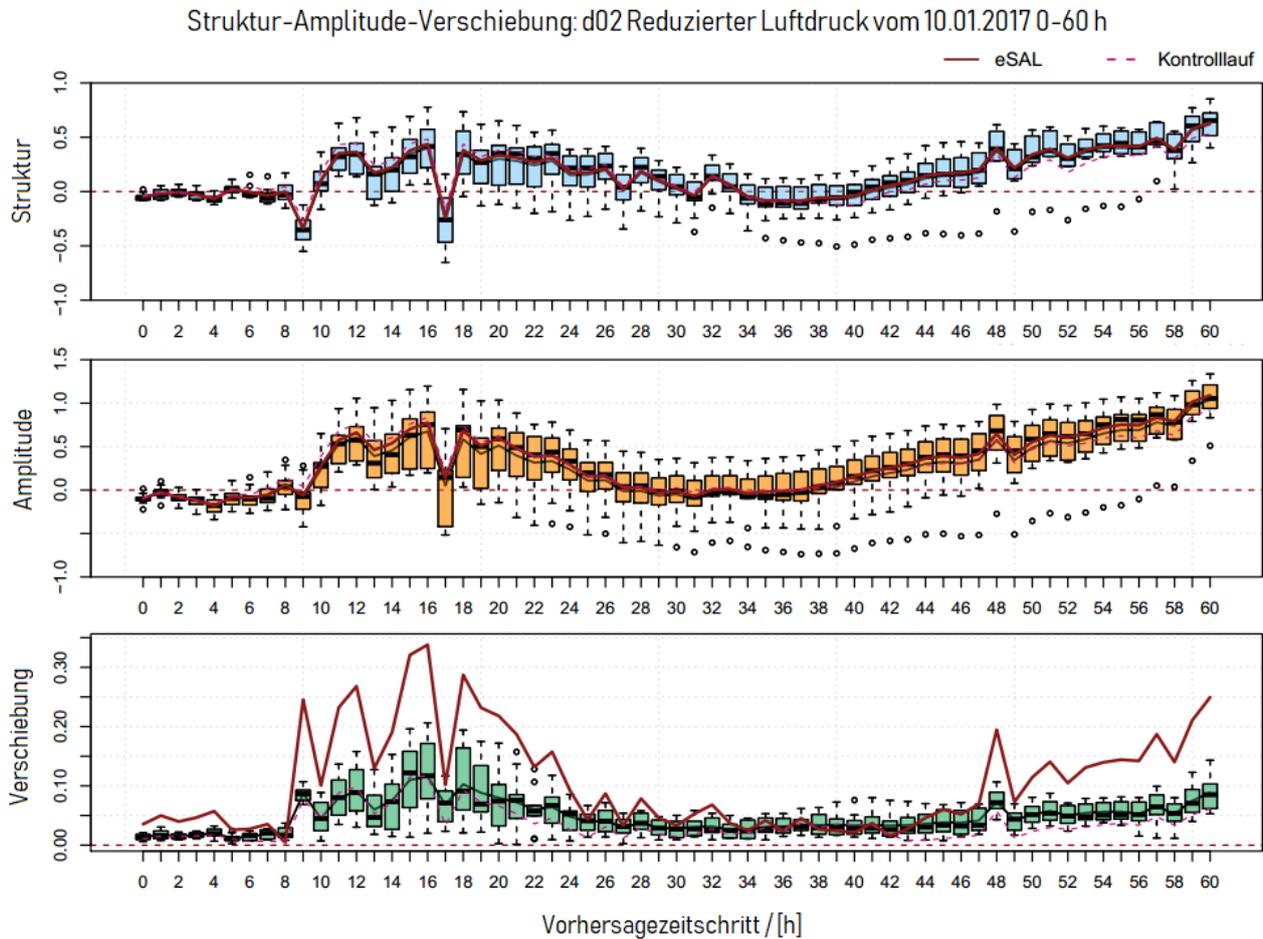


Abbildung 4.21: SAL-Komponenten des reduzierten Luftdrucks des Vorhersagelaufs vom 10.01.2017 00 UTC (10.01. 00 UTC - 12.01.2017 12 UTC). Die Ensemblemember werden durch den Whiskersplot repräsentiert, der Kontrolllauf (rosa strichliert) und das eSAL (rote Linie) werden getrennt behandelt. Oben ist die Struktur, in der Mitte die Amplitude, unten die räumliche Verschiebung abgebildet. Der Schwellenwert wird für jeden Vorhersageschritt über den Mittelwert des Analysefeldes bestimmt.

der Windgeschwindigkeit sind der Median der Ensemblemember und die Ensembleversion aller Komponenten ähnlich. Allerdings hat das eSAL für Struktur und Amplitude ab der Vorhersagestunde 44 höhere Werte als der Median. Die Ensembleversion überschätzt die Größe und Intensität des Objekts. Das eSAL der räumlichen Verschiebung ist ähnlich zum Median des SAL. Das unterscheidet sich von den Ergebnissen der Windgeschwindigkeit des 10 m-Windes und von jenen des Niederschlags.

Abbildung 4.21 stellt die SAL-Komponenten des reduzierten Luftdrucks dar. Der Vergleich mit dem Mittelwert des Drucks zeigt, dass bei Abnahme des Luftdrucks Struktur, Amplitude und räumliche Verschiebung Werte nahe Null liefern. Die Vorhersage wird sehr gut bewertet.

Nimmt der Druck in der Domäne wieder zu, so neigt die Vorhersage die Analyse zu überschätzen. Spricht man von einer Überschätzung bezüglich der Intensität, werden die Druckwerte niedriger vorhergesagt. Das liegt an der inversen Definition des Objekts. Auf Grund der Skala des reduzierten Luftdrucks verläuft die Kurve der Parameter viel glatter, denn die Werte sind weniger variabel als bei kleinskaligen Parametern. Der Verlauf des Medians aller drei Komponenten des SAL und der Ensembleversion sind erneut ähnlich. Bei Zunahme des Luftdrucks wird die Vorhersage durch die Ensembleversion eSAL schlechter bewertet. Der Wert von eL ist um den Faktor 2 höher. Allerdings hat der Druck die niedrigsten Werte der räumlichen Verschiebung.

4.4.4 Diskussion

Die Bewertung der SAL-Komponenten lassen sich gut mit dem Mittelwert aus den Abbildungen 4.4 und 4.6 beschreiben. Falsche Vorhersagen bei niedrigen Werten von Niederschlag oder Windgeschwindigkeit werden stärker bestraft. Das gleiche gilt für die äquivalentpotentielle Temperatur und den reduzierten Luftdruck. Nimmt die Temperatur zu, so ist die Vorhersagequalität des Temperaturmaßes besser. Die Vorhersage des Luftdrucks wird besser bewertet, wenn der Druck abnimmt. Das Objekt des Luftdrucks ist invers definiert. Zur Definition des Objekts wurde der Schwellenwert von allen Werten abgezogen und nur der Betrag der Werte kleiner Null berücksichtigt. Dann wird der tiefste Druckwert nominal zum höchsten. Somit gilt auch für den Luftdruck, dass die Vorhersage von niedrigeren Werten des Objekts (höhere Druckwerte) stärker bestraft wird.

Die Windgeschwindigkeit hat eine feste untere Grenze. Für die Definition der Objekte wurde ein fixer Schwellenwert verwendet. Die Objekte von Temperatur und Druck werden mittels Mittelwert aus Vorhersage- und Analysefeld definiert. Es empfiehlt sich, einen Schwellenwert höher als das Minimum von Vorhersage bzw. Analyse zu wählen. Ein zu hoher Schwellenwert kann jedoch alle Werte in der Domäne eliminieren, sodass kein Objekt mehr vorliegt.

Die Verifikationsergebnisse der Parameter hängen vom Größenverhältnis des Objekts zur Domäne ab. Betrachtet man die hochaufgelöste Domäne d02. Ist das Objekt so groß wie die Domäne selbst, so fließen alle Werte des Parameters in die Verifikation mit ein. Die Struktur wird über das Volumenmittel berechnet. Dann liegt das gewichtete Volumen nahe 1, für die Struktur ergeben sich Werte um Null. Die Amplitude berechnet sich aus der Differenz der Mittelwerte. Werden alle Gitterpunkte berücksichtigt, ergibt sich ein stark geglätteter mittlerer Fehler nahe Null. Die räumliche Verschiebung hängt von der Gitterpunktsanzahl entlang einer Domänenseite ab. Bei räumlich ähnlichen Objekten ist der Abstand zueinander und der Massenschwerpunkt zwischen Analyse- und Vorhersagefeld gering. Daraus folgt eine kleine räumliche Distanz. Mehrere kleine Objekte anstelle eines zusammenhängenden Objekts sind aussagekräftiger. Es ist seltener, dass in einer sehr kleinen Domäne mehrere Objekte auftreten. Beispielsweise zeigt der Luftdruck unabhängig vom Schwellenwert nur ein Objekt an. Das liegt auch an der Größenordnung des Drucks. Liegen in der Domäne keine signifikanten Wetterereignisse wie Front oder Tiefdruckgebiet vor, so sind die Werte relativ gleich. Dann

entstehen sehr viele unterschiedliche Objekte für jeden Ensemblemember, die zu größeren Unsicherheiten führen.

Radanovics et al. (2018) gibt an, dass die räumliche Komponente des eSAL tendenziell niedriger oder gleich dem Median ist. Der Grund dafür liegt in der Berechnung der räumlichen Komponente des eSAL. Berechnet werden die absoluten Abstände des Mittelwerts von Vorhersage und Analyse und der Continuous Ranked Probability Score zweier Verteilungsfunktionen. Das gilt solange, bis Ausreißer den Mittelwert stärker beeinflussen als den Median. Nur die äquivalentpotentielle Temperatur hat einen eL kleiner oder gleich des Medians. Das eL der anderen meteorologischen Parameter, einschließlich des Niederschlags, ist größer als der Median des SAL. Das ist vor allem dann der Fall, wenn die Streuung der Ensemblemember geringer ist. In einer kleinen Domäne können bereits wenige Ausreißer den Mittelwert verändern.

4.5 Displacement and Amplitude Score

Der Displacement and Amplitude Score bewertet mittels Bildvergleich wie gut Objekte in Position und Form übereinstimmen. Das kann durch die Verschiebung von Objekten realisiert werden. Für die Berechnung des Maßes werden zwei normierte Fehlermaße für Amplitude und Verschiebung addiert. Ähnlich zum SAL spielt auch die Größe der Domäne und der Skala des meteorologischen Parameters eine Rolle. Abbildung 4.22 zeigt das Objekt der Windgeschwindigkeit für die Analysedaten des VERA Systems am 12.01.2017 00 UTC und für den Kontrolllauf der WRF Ensemblevorhersage des Vorhersagelaufs vom 10.01.2017 00 UTC für die 38. Vorhersagestunde. Beide Objekte werden mit einem Schwellenwert von 3 m/s definiert. Diese Darstellung ist bereits aus Abbildung 4.13 bekannt. Im Nordwesten der Domäne hat die Vorhersage höhere Werte, im Südosten gibt es wenige bis keine Windgeschwindigkeiten über 3 m/s. Der Kontrolllauf zeigt eine höhere Windgeschwindigkeit im Nordwesten der Domäne. Das Objekt im Analysefeld ist etwas kleiner als im Vorhersagefeld. Die Deformation wird sowohl im Vorhersage- als auch im Analysefeld durchgeführt, um false alarms und missed events zu berücksichtigen. Informationen, wo das Vorhersage- bzw. Analysefeld am stärksten verändert werden müssen, zeigen die deformierten Felder. Daraus ergeben sich der Verschiebungs- (DIS) und Amplitudenfehler (AMP). Diese zählen gemeinsam mit dem deformierten Vorhersage- und Analysefeld zu den signifikanten Phasen des verwendeten Algorithmus.

Die maximale Distanz D_{max} (vgl. Gleichung 3.27) liegt bei 64 km. Für den pyramidal matching Algorithmus wurden 4 Level verwendet. Das entspricht 16 Gitterpunkte bei einer Auflösung von 4 km. Die größte Skala berücksichtigt 64 Pixel. Für die charakteristische Intensität wurde der Mittelwert des RMS aller Analysewerte im Objekt gewählt.

Das Analysefeld zeigt einen höheren Verschiebungsfehler im Nordosten und im Westen der Domäne d02. Der Amplitudenfehler ist im Nordwesten am größten. Dort sind die höchsten Windgeschwindigkeiten. Der DIS der Vorhersage hat das stärkste Signal im Nordwesten. Der

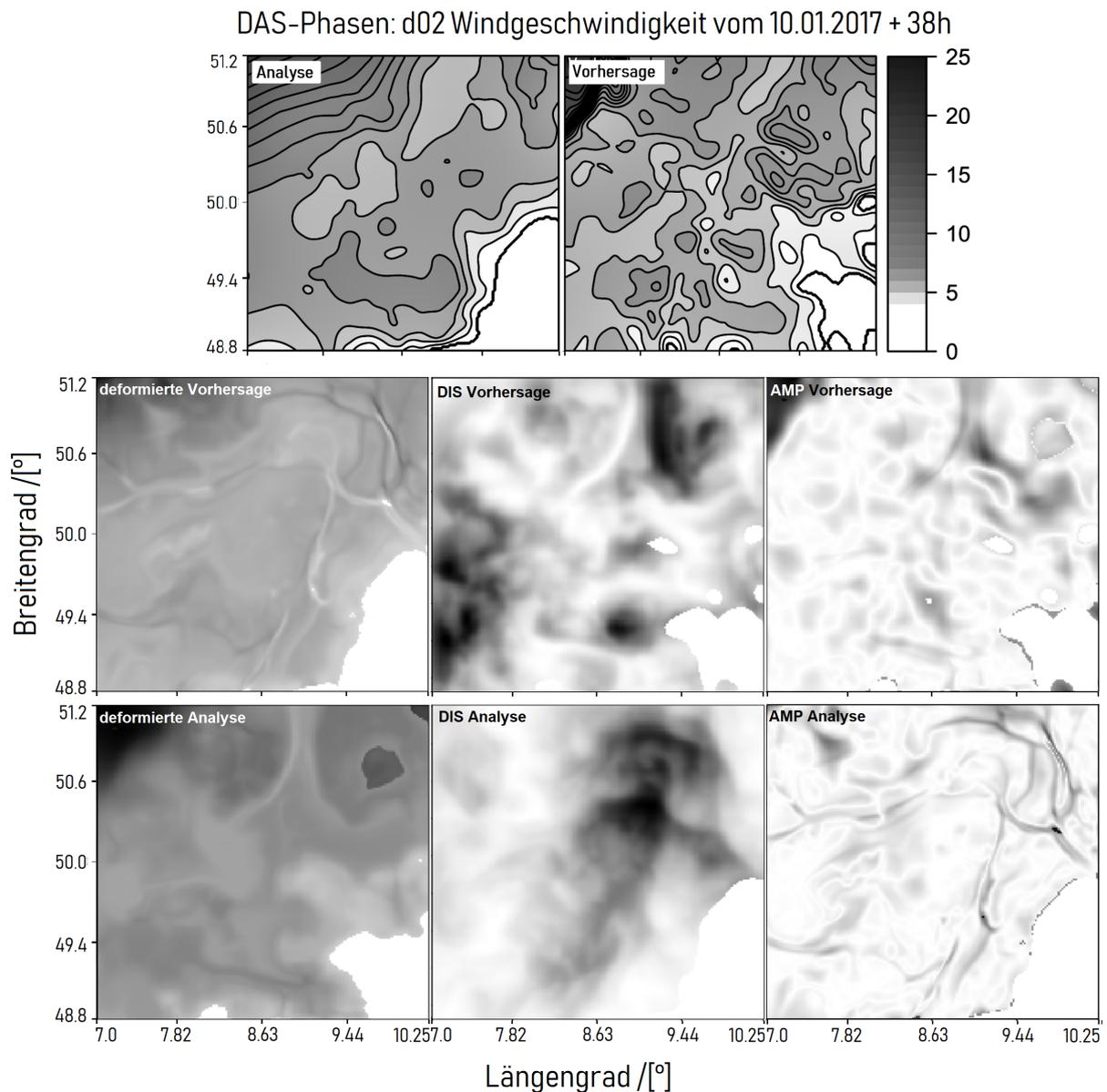


Abbildung 4.22: Phasen des Optical Flow Algorithmus zur Berechnung des DAS für die 10m-Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC +38 h der Domäne d02. In der oberen Reihe zeigt das Objekt die Windgeschwindigkeit mit einem Schwellenwert von 3 m/s des Analysefeldes und des Vorhersagefeldes des Kontrolllaufs. Die mittlere Reihe zeigt die deformierte Vorhersage im Analyse- und des daraus resultierenden Verschiebungs- und Amplitudenfehlers. Untere Reihe zeigt die deformierte Analyse im Vorhersageraum und den Ergebnissen DIS und AMP. Je dunkler das Feld, desto stärker wird es deformiert.

AMP zeigt eine durchgezogene Struktur.

Abbildung 4.23 zeigt die Ergebnisse des DAS, DIS und AMP für die gesamte Vorhersagedauer.

Wie bereits bei der SAL-Methode werden die Ensemblemember des WRF Vorhersagemodells als Dichteplot dargestellt. Die rote Kurve entspricht dem Kontrolllauf. DAS ergibt sich

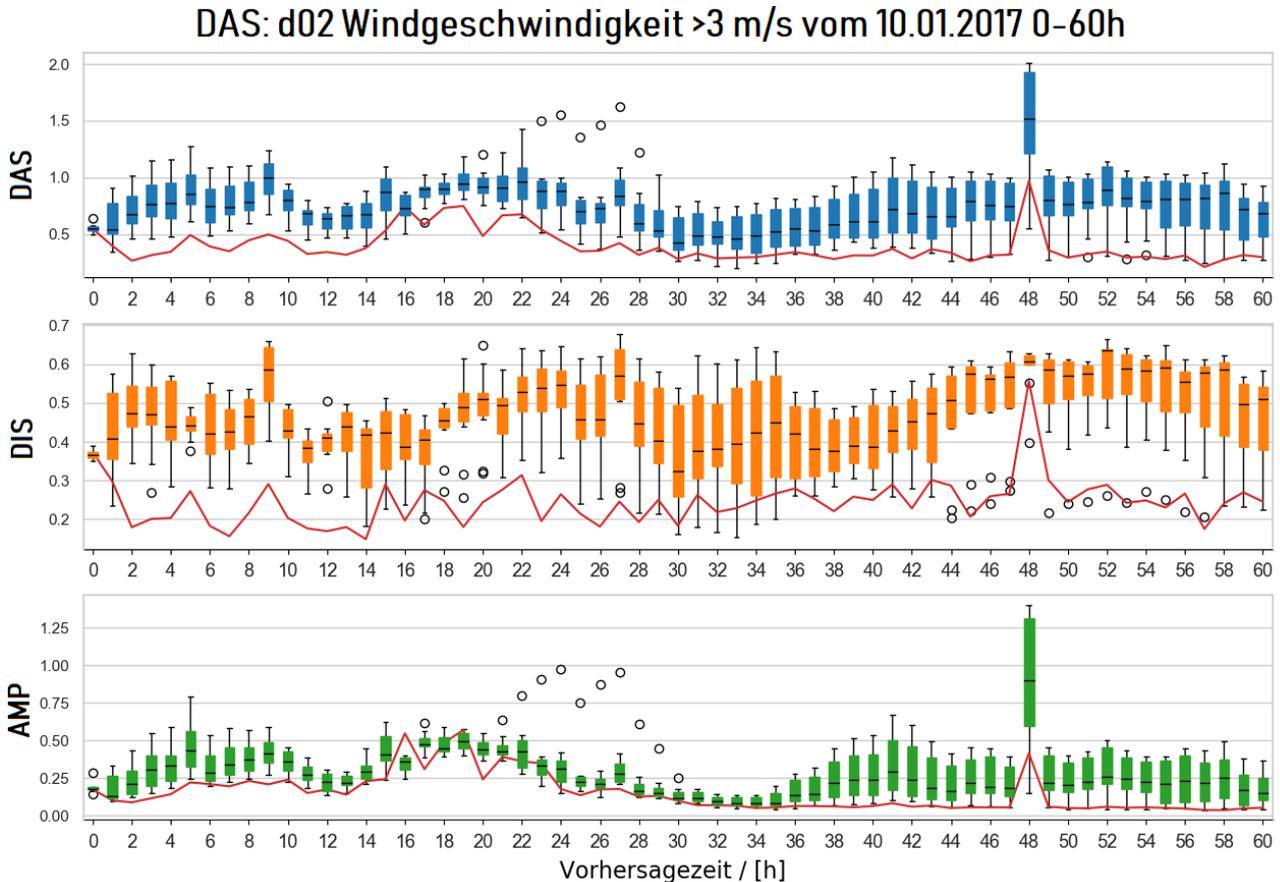


Abbildung 4.23: Displacement and Amplitude Score DAS der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02, Verschiebungsfehler DIS und Amplitudenfehler AMP. Der Schwellenwert beträgt 3 m/s.

aus der gewichteten Summe des Verschiebungsfehlers DIS und des Amplitudenfehlers AMP. Der Verschiebungsfehler hat gemittelt über die gesamte Vorhersagezeit einen Wert von 0.45 (vgl. Tabelle 4.1) und ist größer als der AMP mit 0.28. Der Score wird vom DIS stärker gewichtet. Der mittlere Wert des DAS beträgt 0.74. Die Vorhersage ist bei einem Wert von 0 perfekt, der Fehler kann theoretisch unendlich werden. Meist hat das Maß einen Wert der Größenordnung 1. Die DAS-Werte liegen zwischen 0.25 und 1.25 (Außreisser wurden ignoriert). Der Kontrolllauf besitzt einen Wert von 0.39, was im Vergleich zu den Ensemblemember einer guten Vorhersagequalität entspricht. DAS liefert zu Beginn des Vorhersagelaufs vom 10.01.2017 00 UTC einen Wert um 0.5. Das Fehlermaß nimmt, anders als erwartet, mit fortschreitender Vorhersagezeit nicht zu. Auf Grund des sofortigen Fehleranstiegs lässt sich schließen, dass bereits zum Initialisierungszeitpunkt eine Fehlersättigung erreicht ist. Die Zunahme des Fehlers in der 48. Vorhersagestunde ist auf Grund der geringeren Beobachtungsdichte am 12.01. 00 UTC, welche in die Analyse einfließen.

Abbildung 4.24 zeigt den Displacement and Amplitude Score DAS, den Verschiebungsfehler

DAS: d02 Windgeschwindigkeit >5m/s vom 10.01.2017 0-60h

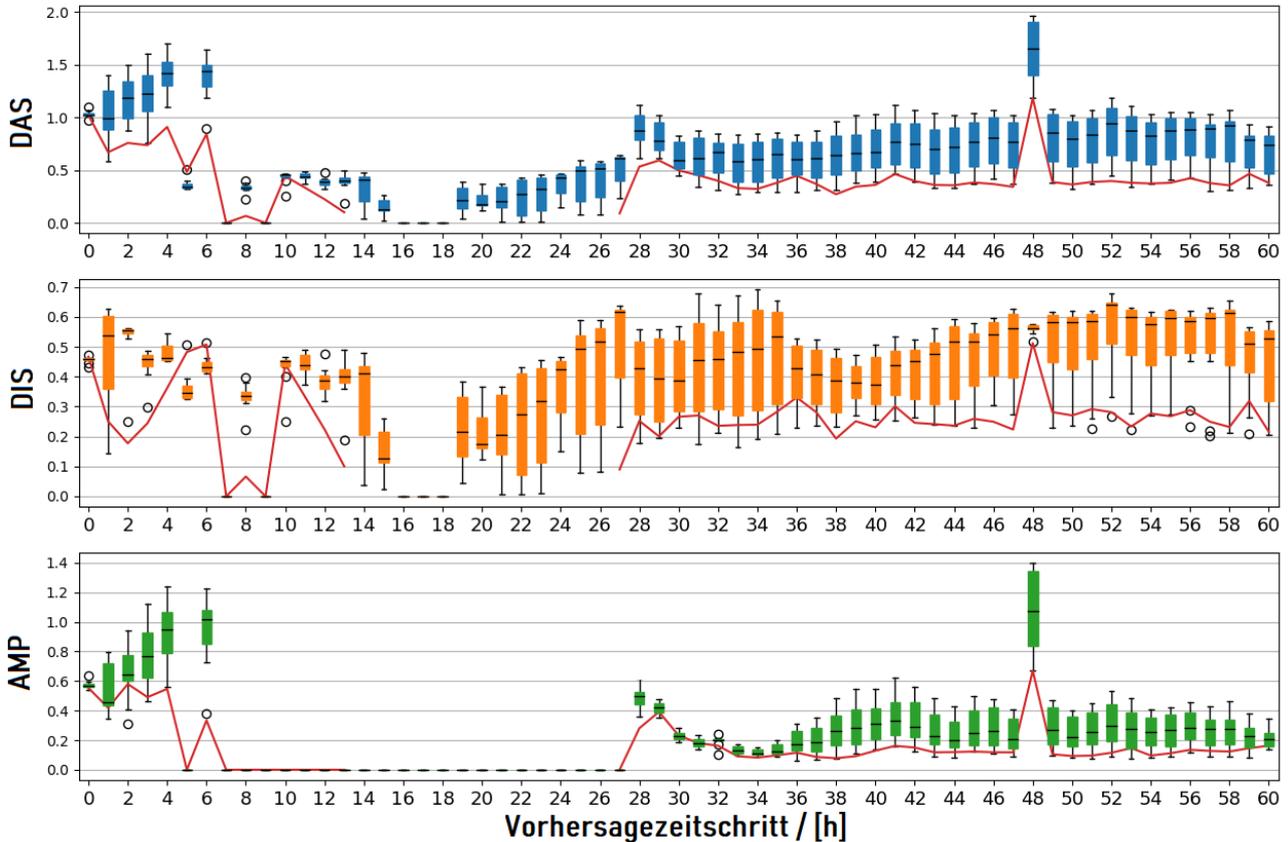


Abbildung 4.24: Displacement and Amplitude Score DAS der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02, Verschiebungsfehler DIS und Amplitudenfehler AMP. Der Schwellenwert beträgt 5 m/s.

DIS und den Amplitudenfehler DAS für Windgeschwindigkeit für die Vorhersagezeit vom 10.01.2017 00 UTC der Domäne d02 mit einem Schwellenwert von 5 m/s. Je höher der Schwellenwert wird, desto kleiner werden die Objekte. Außerdem steigt die Wahrscheinlichkeit, dass sich kein Objekt in der Domäne befindet (Vorhersagestunden 16-18). Dann wird auch kein Fehler zwischen den Objekten erkannt und man erhält einen DAS von Null. Das entspricht einer perfekten Vorhersage (hier: correct rejection). Durch das Entfernen der niedrigeren Windgeschwindigkeiten nimmt der Amplitudenfehler zu. Der Verschiebungsfehler bleibt in etwa gleich, da die höchsten Windgeschwindigkeiten im Allgemeinen im Nordwesten liegen. Die Position des Objekts bleibt in etwa gleich (vgl. Abb. 4.13). Mit größerem AMP und gleichbleibendem DIS wird auch der DAS größer. Ab der Vorhersagestunde 25 ist die mittlere Windgeschwindigkeit höher. Folglich ist der Unterschied zwischen den beiden Objekten unterschiedlichen Schwellenwerts geringer.

Die gemittelten Ergebnisse des kombinierten Fehlermaßes und seines Verschiebungs- und Amplitudenfehlers sind in Tabelle 4.1 für Domäne d01 und d02 für die Schwellenwerte von

3 m/s und 5 m/s zusammengefasst. DAS berechnet für die Objekte mit dem Schwellenwert

		AMP	DIS	DAS	AMP Kontr.	DIS Kontr.	DAS Kontr.
Obj. 3	d01	0.305567	0.460312	0.76588	0.174180	0.253903	0.428083
	d02	0.287761	0.455739	0.74350	0.151714	0.244483	0.396198
Obj. 5	d01	0.248015	0.42395	0.671965	0.169872	0.262559	0.432431
	d02	0.257128	0.44474	0.701867	0.166253	0.260469	0.426722
Obj. 7	d01	0.346635	0.365024	0.711659	0.359989	0.278629	0.638618
	d02	0.359239	0.380756	0.739996	0.349861	0.262894	0.612754

Tabelle 4.1: Mittel des Amplitudenfehlers AMP, Verschiebungsfehlers DIS und Fehlermaßes DAS des gesamten Vorhersagelaufes vom 10.01.2017 00 UTC 0-60h. Berechnet werden der Mittelwert der Ensemblemember und der Mittelwert des Kontrolllaufs. Die Ergebnisse werden für einen Schwellenwert von 3 m/s, 5 m/s und 7 m/s berechnet und für beide Domänen d01 und d02.

von 3 m/s höhere Fehler als für die Objekte mit einem Schwellenwert von 5 m/s. Ein höherer Schwellenwert erhöht die Wahrscheinlichkeit, dass sich kein Objekt in der Domäne befindet. In den Ergebnissen hat sich gezeigt, dass die Objekte mit dem Schwellenwert von 5 m/s einige Nullwerte erhalten. Die mittlere Fehlerdifferenz zwischen den beiden Objekten ist relativ gering. Der Kontrolllauf hat kleinere Fehler. Die Häufigkeitsverteilung (vgl. Abbildung 4.7) zeigt bereits, dass der Kontrolllauf eine ähnliche Verteilung zur VERA Analyse hat und daher eine bessere Prognose darstellt. Das Objekt mit dem Schwellenwert von 7 m/s zeichnet sich durch eine Zunahme des Amplitudenfehlers und einer Abnahme des Verschiebungsfehlers aus. Während das größere Objekt mit 3 m/s für d02 besser abschneidet, besitzt Objekt 5 für d02 einen höheren Fehlerwert.

DAS-Diagramm

Abbildung 4.25 stellt den Amplituden- und Verschiebungsfehler des Vorhersagelaufs vom 10.01.2017 00 UTC gegenüber. Die Darstellung liefert einen Überblick über das Verhalten verschieden großer Objekte bezüglich des Verschiebungs- und Amplitudenfehlers des DAS. Niedrigere DAS Werte liegen näher am Ursprung. Mit dieser Darstellung kann eine Aussage getroffen werden, ob die Vorhersage des Objekts wegen der Lage oder der Amplitude schlechter bewertet wird. Für die Objekte mit einem Schwellenwert von 3 m/s und 5 m/s berechnet der DIS höhere Werte als der AMP. Erst mit höherem Schwellenwert, wenn niedrigere Windgeschwindigkeiten herausgefiltert werden, verschieben sich die Wertepunkte und der AMP berechnet einen größeren Fehler. Das bedeutet, dass der DAS bei großen Objekten (Schwellenwert von 3 m/s) stärker vom DIS beeinflusst wird.

Am linken Rand liegt eine Säule mit Wertepunkten. Die entstehen, wenn bei sehr kleinen Objekten, die Windgeschwindigkeit klein ist, aber die räumliche Lage nicht stimmt. Das tritt vor allem bei Objekten mit einem höheren Schwellenwert auf.

DAS: Verschiebungs- und Amplitudenfehler der Windgeschwindigkeit vom 10.01.2017 00-60h

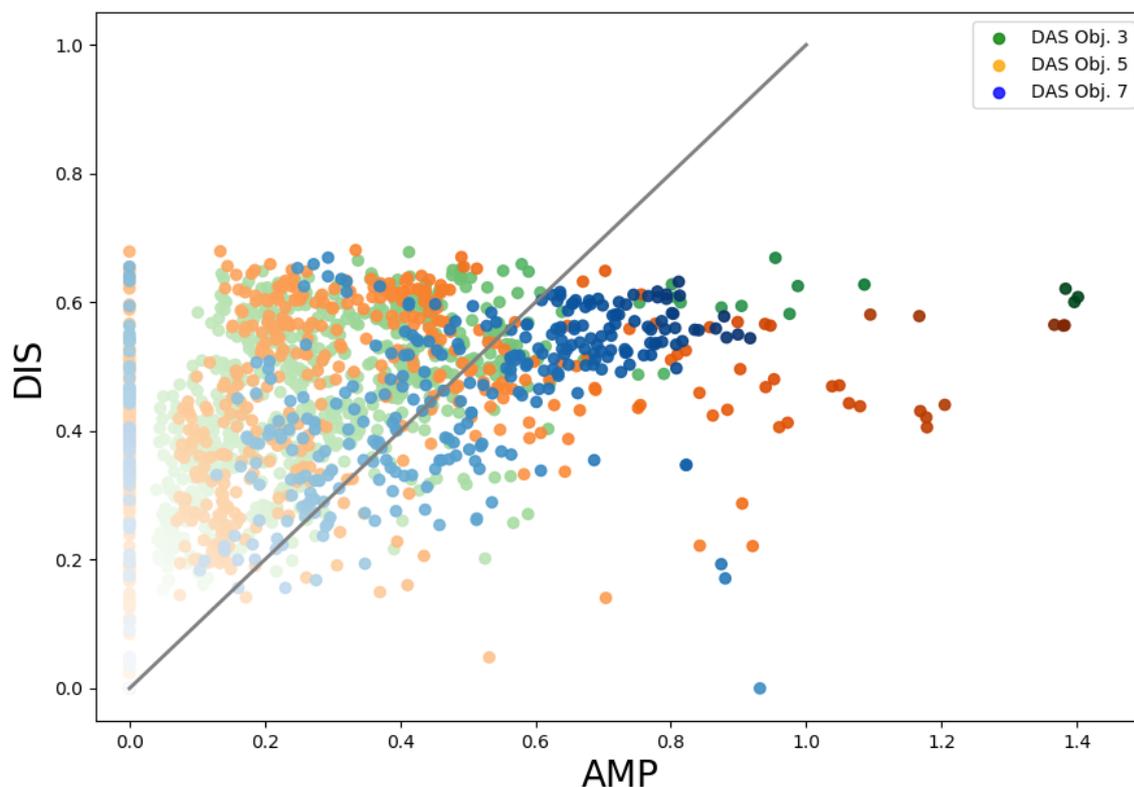


Abbildung 4.25: Verschiebungs- (DIS) und Amplitudenfehler (AMP) der Windgeschwindigkeit des Vorhersagelaufs vom 10.01.2017 00 UTC der Domäne d02 für die Objekte mit den Schwellenwerten 3 m/s, 5 ms und 7 m/s. Liegt ein Wertepunkt oberhalb der Linie (grau), so ist der DIS größer als der AMP, unterhalb der Linie ist der AMP größer. Ist der DAS klein, so liegen die Punkte nahe des Ursprungs. Sind die Punkte transparenter ist der DAS klein, je größer der Fehlerwert ist, desto intensiver die Färbung.

4.5.1 Diskussion

Das Fehlermaß nimmt, anders als erwartet, mit fortschreitender Vorhersagezeit nicht zu. Es wird angenommen, dass bereits zum Initialisierungszeitpunkt eine Fehlersättigung erreicht ist. Das liegt einerseits am heterogenen Windfeld, andererseits liegt es daran, dass das Objekt der Windgeschwindigkeit im Vergleich zur Domäne groß ist und auch weniger Raum für Verschiebungsfehler lässt. Die Windvorhersage für den mittleren Wind wird selten für die gesamte Domäne perfekt sein. Allerdings gibt es keinen maximalen Fehler des Windes. Das Objekt wird mit zunehmendem Schwellenwert kleiner und berücksichtigt die niedrigeren Windgeschwindigkeiten, die am häufigsten auftreten, nicht. Daher wird der Amplitudenfehler größer, der Verschiebungsfehler allerdings geringer. Die Methode eignet sich besser für die

Verifikation von Objekten mit höheren Windgeschwindigkeiten bzw. für Windmaxima als für den mittleren Wind mit niedrigem Schwellenwert. Bei niedrigen Windgeschwindigkeiten ist das Objekt zu groß im Vergleich zur Domäne.

Das DAS-Diagramm aus Abbildung 4.25 liefert die gleichen Informationen wie die Zeitreihe in Abbildung 4.23 dar. Der Nachteil ist, dass die Werte des DAS nicht direkt abzulesen sind.

Die anderen meteorologischen Parameter, welche nicht dargestellt sind, zeigen eine Zunahme des Fehlers mit der Zeit. Das DAS ist für alle Größen kleiner als der Fehler für die Windgeschwindigkeit. Außerdem ist der DIS größer als der AMP. Die äquivalentpotentielle Temperatur und der reduzierte Luftdruck repräsentieren Wetterereignisse auf großer Skala und besitzen in der kleinen Domäne d02 sehr konstante Werte. Das heißt, der größte Unterschied liegt in der Position der Objekte.

Kapitel 5

Zusammenfassung

Der zentrale Aspekt dieser Arbeit ist die Überprüfung, wie gut sich räumliche Verifikationsmethoden eignen, wenn kontinuierliche Variablen, wie die Windgeschwindigkeit, betrachtet werden.

Die verwendeten räumlichen Verifikationsmethoden behandeln unterschiedliche Fragestellungen. Die Methode Upscaling bewertet die Zuverlässigkeit der gemittelten Vorhersagewerte für verschiedene räumliche Skalen und Schwellenwerte. Das Entscheidungsmodell besagt, nützliche Vorhersage ähneln der Analyse, wenn sie auf gröbere Skalen gemittelt werden (Zepeda-Arce et al. (2000), Yates et al. (2006)). Der FSS vergleicht die Häufigkeiten von Vorhersage und Analyse auf einem bestimmten Teilabschnitt (fraction). Das Entscheidungsmodell besagt, nützliche Vorhersagen haben eine ähnliche Häufigkeit von prognostizierten und analysierten Ereignissen (Roberts and Lean (2008)). Zusätzlich gibt die Skillful Scale Information an, ab welchen räumlichen Skalen die Vorhersage sinnvoll ist (Mittermaier and Roberts (2010)). Die SAL-Methode vergleicht Vorhersage- und Analyseobjekte bezüglich ihrer Struktur, Amplitude und Ort (Wernli et al. (2008)). Der DAS vergleicht das Vorhersage- und Analysefeld in Position und Form anhand von Verschiebungsmerkmalen (Amplitude und Position) (Keil and Craig (2007)).

Viele räumliche Verifikationsmethoden wurden für Niederschlagsereignisse entwickelt. Die Windgeschwindigkeit hat viel Ähnlichkeit mit dem Niederschlag. Beide sind positiv semidefinit und besitzen eine natürliche untere Grenze. Des Weiteren haben sie die gleiche Größenordnung. Die Windgeschwindigkeit ist eine kontinuierliche heterogene Größe und viel kleinskaliger als der Niederschlag. Daher ist es sinnvoll, den Mittelwert über mehrere Stunden (Neighborhood Methoden) oder einen gesamten Vorhersagelauf (SAL, DAS) zu betrachten. Die Datenanalyse der Windgeschwindigkeit (Bias, Mittelwert, Häufigkeitsverteilung) von Vorhersage und Analyse zeigt: Die meisten Ensemblemembers überschätzen die mittlere analysierte Windgeschwindigkeit. Der Kontrolllauf wird von allen räumlichen Verifikationsmethoden am besten bewertet. Der Ensemblemittelwert liefert, wie zu erwarten, eine mittlere Qualität. Zweitens lassen sich die Ensemblemembers in drei Gruppen mit ähnlichen Vorhersagen einteilen. Das ist nur bei der Windgeschwindigkeit zu beobachten.

Die Ensemblemembers bestehen aus einem Set von physikalischen Parametrisierungen. Die Hypothese ist, dass die Parametrisierungen dazu führen, dass manche Ensemblemembers ähnlich zueinander sind. Für zukünftige Arbeiten könnte getestet werden, wie sich die Windgeschwindigkeit in höheren atmosphärischen Schichten verhält, um so Informationen über die Korrelation zwischen Modell und Parametrisierung bezüglich des Windes zu erhalten.

Upscaling berechnet Scores mit Hilfe der Kontingenztabelle. Dafür muss die Windgeschwindigkeit in ein binäres Feld umgewandelt werden, was durch die Verwendung eines Schwellenwerts passiert. Die Windgeschwindigkeit kann dann relativ einfach verifiziert werden. Das gilt auch für den Fractions Skill Score. Die Verifikationsergebnisse entsprechen den Erwartungen, welche auf Grund der Häufigkeitsverteilung und des Bias erstellt wurden (vgl. 4.2 und 4.3). Das unterstützt die Aussage, dass sich die Neighborhood Methoden für die Verifikation der Windgeschwindigkeit eignen. Die Neighborhood Methoden für die Verifikation von großskaligen meteorologischen Parameter wie der äquivalentpotentiellen Temperatur und dem reduzierten Luftdruck eignen sich für kleine Domänen weniger. Auf Grund der geringen Variabilität der Vorhersage- und Analyswerte werden sie meist für jede räumliche Skala gleich bewertet. Der Vorteil der Neighborhood Methoden ist, dass Analyse und Vorhersage nur näherungsweise übereinstimmen müssen. Wird eine größere Domäne betrachtet, würde man wahrscheinlich ein ähnliches Ergebnis erhalten. Diese Hypothese könnte man in Zukunft austesten. Mittels Upscaling sollen jene Skalen herausgefunden werden, die die besten Verifikationsergebnisse liefern. Die Ergebnisse zeigen, dass die in der Analyse aufgelösten Skalen in der Vorhersage vertreten sind. Diese wurden in einem Intensity-Scale-Diagramm dargestellt (vgl. Casati et al. (2004), Ebert (2008)). Der Vorteil liegt darin, dass die Ensemblemembers getrennt voneinander betrachtet werden. So ist z.B. die Gruppierung der Members sichtbar. Besteht das Vorhersagemodell aus einer sehr hohen Anzahl an Ensemblemembers, ist die Darstellung nachteilig. Der Fractions Skill Score gibt die Skalen an, für die die Vorhersage sinnvoll ist. Auch die FSS-Werte werden in einem Intensity-Scale Diagramm dargestellt. Zusätzlich kann die Skillful Scale eingetragen werden. Abbildung 3.4 zeigt, dass bei guten Vorhersagen das Maß mit größeren Skalen zunimmt. Räumliche Skalen, welche oberhalb dieser Skala liegen, gelten als sinnvoll. Je flacher die Kurve ist, desto zuverlässiger wird die Vorhersage.

Für objektorientierte Methoden muss das Windfeld zusätzlich in ein Objekt umgewandelt werden. Auch wenn die Windgeschwindigkeit eine natürlich untere Grenze von 0 m/s besitzt, so sind Kalmen eher selten, und die Windgeschwindigkeit tritt nicht als ein natürliches abgeschlossenes Objekt auf. Für die Definition der Objekte wird ein fixer Schwellenwert verwendet. Will man die Windgeschwindigkeit mit der SAL-Methode verifizieren, muss darauf geachtet werden, wie groß das Objekt gewählt wird. Es empfiehlt sich einen Schwellenwert größer als das Minimum der Analyse- und Vorhersagewerte zu wählen. Die SAL-Verifikationsergebnisse der Parameter hängen vom Größenverhältnis der Objekte zur Domäne ab. Mehrere kleine Objekte eines Systems anstelle eines zusammenhängenden Objekts sind aussagekräftiger, allerdings sind mehrere Merkmale in einer sehr kleinen Domäne seltener. Wernli et al. (2009) beschreibt in seiner Arbeit, dass die Wahl der Domäne einen tiefgreifenden Einfluss auf die Verifikationsergebnisse hat. Ist das Objekt so groß, sodass

es die Domäne vollständig ausfüllt, ergibt sich eine Mittelung. Der Parameter wird zu gut bewertet. Allerdings sollte die Domäne im Vergleich zum Objekt nicht zu groß sein. Dann steigt die Wahrscheinlichkeit, dass zwei unterschiedliche meteorologische Systeme erfasst werden. Ein System könnte das andere ausgleichen. Für zukünftige Arbeiten könnte man den Vorhersagelauf der größeren Domäne d01 auf beiden Domänen (d01 und d02) auswerten und gegenüberstellen. Die Objekte der äquivalentpotentiellen Temperatur können so gewählt werden, dass eine Front repräsentiert wird. Wird das Objekt des Drucks invers definiert, so kann ein Tiefdruckgebiet verifiziert werden. Die geringe Variabilität der Werte bleibt vorhanden, wodurch die Kurve der Ergebnisse glatter verläuft, als z.B. beim Niederschlag. Die Verifikationsergebnisse lassen sich gut mit dem Mittelwert aus der Datenanalyse beschreiben. Der Schwellenwert zur Definition der Objekte wird für jeden Vorhersagezeitschritt neu berechnet. Das Tiefdruckgebiet und die Front können gut verifiziert werden. Befindet sich in der Domäne keine Front, dann sind alle Temperaturwerte sehr ähnlich. Der Mittelwert als Schwellenwert definiert viele kleine Objekte, die zu größeren Unsicherheiten führen. Das Problem kann reduziert werden, indem der Schwellenwert über ein Perzentil bestimmt wird. Auch für die Berechnung des DAS ist es sinnvoll, die Windgeschwindigkeit in Objekte umzuwandeln. Der DAS gibt den kombinierten Verschiebungs- und Amplitudenfehler an. Die Fehlerwerte des mittleren Windes verändern sich nicht, da bereits zu Beginn des Vorhersagelaufs eine Fehlersättigung eingetreten ist. Verändert sich der Fehler doch, dann kann das ein Hinweis für z.B. fehlerhafte Werte sein. Die Methode ist für Verifikation von Windmaxima sinnvoller. Für das SAL und den DAS werden die Ensemblemembers in einem Dichteplot zusammengefasst, sodass die Ergebnisse als Zeitreihe dargestellt werden (vgl. Radanovics et al. (2018)). Der Vorteil ist, dass so die Bewertung der Objekte mit fortschreitender Vorhersagedauer verglichen werden. Im Gegensatz zu den Neighborhood Methoden eignet sich das SAL nicht, wenn die Vorhersage und Analyse über einen längeren Zeitraum gemittelt werden. Der Nachteil der Darstellung ist, dass die Information eines individuellen Ensemblemembers verloren geht. Für eine genauere Information kann das SAL-Diagramm herangezogen werden (vgl. Wernli et al. (2008), Radanovics et al. (2018)). Die Größen des DAS können im zweidimensionalen Diagramm dargestellt werden. Es liefert einen Überblick über das Verhalten verschieden großer Objekte bezüglich des Verschiebungs- und Amplitudenfehlers des DAS. Der Wert des DAS kann nicht direkt abgelesen werden.

Räumliche Verifikationsmethoden lassen sich vorrangig auf deterministische Vorhersagen anwenden. Die Ensemblemembers der verwendeten Vorhersagemodells werden als Sammlung deterministischer Vorhersagen angesehen und getrennt voneinander verifiziert. Zusätzlich wird der Mittelwert der Ensemblemembers betrachtet. Mit der Ensembleversion eSAL wird die Ensemblevorhersage in ihrer Gesamtheit betrachtet (Radanovics et al. (2018)). Während die Ergebnisse für Struktur und Amplitude meist nahe des Medians liegen, weicht die räumliche Verschiebung (eL) stärker vom Median und Mittelwert der Standardmethode ab. Die kleine Domäne könnte ein Grund für das Verhalten der eL-Komponente sein, so dass viel mehr Ausreißer die Ergebnisse verändern und den eL ansteigen lassen.

Literaturverzeichnis

- Barth, A., Troupin, C., Azcarate, A. A., Joassin, P., and Beckers, J.-M. (2008). Introduction to Optimal Interpolation and Variational Analysis, chapter 1.5: Cressman method. *GeoHydrodynamics and Environment Research*, pages 8–10.
- Bjerknes, V. (1911). Dynamic Meteorology and Hydrography, Part II. Kinematics. *Gibson Bros., Carnegie Institute, New York*.
- Casati, B. (2009). New Developments of the Intensity-Scale Technique within the Spatial Verification Methods Intercomparison Project. *Weather and Forecasting*, 25(1):113–143.
- Casati, B., Ross, G., and Stephenson, D. B. (2004). A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11(2):141–154.
- Cressman, G. P. (1959). An operational objective analysis system. *Monthly Weather Review*, 87:367–374.
- Davis, C., Brown, B., and Bullock, R. (2006). Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Monthly Weather Review*, 134(7):1772–1784.
- Diekmann, B. and Rosenthal, E. (2014). Energie: Physikalische Grundlagen ihre Erzeugung, Umwandlung und Nutzung. *Springer Spektrum*, 3.
- Dierer, S., Oechslin, R., and Cattin, R. (2011). Wind turbines in icing conditions: performance and prediction. *Advances in Science and Research*, pages 245–250.
- Done, J., Davis, C. A., and Weisman, M. (2004). The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, 5:110–117.
- Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M., Ebert, E., Brown, B., and Wilson, L. (2018). The Setup of the MesoVICT Project. *Bulletin of American Meteorological Society*, pages 1887–1906.
- Dorninger, M. and Gorgas, T. (2013). Comparison of NWP-model chains by using novel verification methods. *Meteorologische Zeitschrift*, 22:373–393.

- Ebert, E., Brooks, H., Brown, B., Ferro, C., Jenker, J., Jolliffe, I., Koh, T., Roebber, P., and Stephenson, D. (2015). Forecast Verification, www.cawcr.gov.au/projects/verification.
- Ebert, E. E. (2008). Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15:51–64.
- Ebert, E. E. (2009). Neighborhood Verification: A Strategy for Rewarding Close Forecasts. *Weather and Forecasting*, 24(6):1498–1510.
- Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., and Ebert, E. E. (2009). Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, 24(5):1416–1430.
- Gofa, F., Boucouvala, D., Louka, P., and Flocas, H. A. (2018). Spatial verification approaches as a tool to evaluate the performance of high resolution precipitation forecasts. *Atmospheric Research*, 208(September):78–87.
- Hau, E. (2013). Windkraftanlagen: Grundlagen, Technik, Einsatz, Wirtschaftlichkeit - 2. Auflage. *Springer Verlag*.
- Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction System. *American Meteorological Society*, 15:559–570.
- Jolliffe, I. T. and Stephenson, D. B. (2012). Forecast Verification: A Practitioner’s Guide in Atmospheric Science 2nd Edition. *Wiley-Blackwell*, page 288.
- Keil, C. and Craig, G. C. (2007). A Displacement-Based Error Measure Applied in a Regional Ensemble Forecasting System. *Monthly Weather Review*, 135(9):3248–3259.
- Keil, C. and Craig, G. C. (2009). A Displacement and Amplitude Score Employing an Optical Flow Technique. *Weather and Forecasting*, 24(5):1297–1308.
- Keller, W. and Borkowski, A. (2019). Thin plate spline interpolation. *Journal of Geodesy, Springe Verlag*.
- Leutbecher, M. and Palmer, T. N. (2007). Ensemble Forecasting. *ECMWF Research Department*, (514).
- Makkonen, L. (1981). Estimating Intensity of Atmospheric Ice Accretion on Stationary Structures. *Journal of Applied Meteorology*, 20:595–600.
- Marzban, C. and Sandgathe, S. (2006). Cluster Analysis for Verification of Precipitation Fields. *Weather and Forecasting*, 21(5):824–838.
- McNulty, R. P. (2011). Operational Weather Analysis - chapter 4: Introduction to objective analysis. wxonline.info/ebook/wxa.html.

- Mittermaier, M. and Roberts, N. (2010). Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score. *Weather and Forecasting*, 25(1):343–354.
- Orlanski, I. (1975). A Rational Subdivision of Scales for Atmospheric Processes. *Bulletin American Meteorological Society*, 56(5):527–530.
- Orrell, D., Smith, L., Barkmeijer, J., and Palmer, T. B. (2001). Model error in weather forecast. *Nonlinear Processes in Geophysics*, 8:357–371.
- Radanovics, S., Vidal, J., and Sauquet, E. (2018). Spatial Verification of Ensemble Precipitation: An Ensemble Version of SAL. *Weather and Forecasting*, 33:1001–1020.
- Roberts, N. M. and Lean, H. W. (2008). Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, 136(1):78–97.
- Santoalla, D. (2018). Forecast User Guide: 4.1 Forecast Error Growth, <https://confluence.ecmwf.int/x/KffUBw>, zuletzt aufgerufen am 1.3.2020.
- Shuman, F. G. (1989). History of Numerical Weather Prediction at the National Meteorological Center. *Weather and Forecasting*, 4:286–296.
- Stanski, H. R., Wilson, L. J., and Burrows, W. R. (1989). Survey of common verification methods in meteorology 2nd Edition. *World Weather Watch Tech. Rept. No.8, WMO/TD No.358, WMO, Geneva*.
- Steinacker, R., Ratheiser, M., Bica, B., Chimani, B., Dorninger, M., Gepp, W., Lotteraner, C., Schneider, S., and Tschannett, S. (2006). A Mesoscale Data Analysis and Downscaling Method over Complex Terrain. *Monthly Weather Review*, 134:2758–2271.
- Tsuyuki, T. and Miyoshi, T. (2007). Recent Progress of Data Assimilation Methods in Meteorology. *Journal of the Meteorological Society of Japan*, 85B:331–361.
- VERA (1995). Vienna Enhanced Resolution Analysis, www.univie.ac.at/amk/vera/, zuletzt aufgerufen am 1.3.2020.
- Verbund (2017). Windkraftanlage - Windpark Ellern, www.verbund.com/de-at/ueber-verbund/kraftwerke/unsere-kraftwerke/windpark-rheinland, zuletzt aufgerufen am 1.3.2020.
- Wernli, H., Hofmann, C., and Zimmer, M. (2009). Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique. *Weather and Forecasting*, 24(6):1472–1484.
- Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). SAL-A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, 136(11):4470–4487.

- Weusthoff, T., Ament, F., Arpagaus, M., Rotach, M., and Schütze, M. (2008). Fuzzy verification of precipitation forecasts during the DOP - on the benefit of high resolution models with explicit convection. *7th COPS Workshop*.
- Wilks, D. S. (2011). Statistical methods in the atmospheric sciences - 3rd Edition. *Elsevier Ltd*.
- WRF (2017). WRF Userguide, www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/contents.html.
- Yates, E., Anquetin, S., Ducrocq, V., Creutin, J., Ricard, D., and Chancibaul, K. (2006). Point and areal validation of forecast precipitation fields. *Meteorol. Appl.*, 13:1–20.
- Zepeda-Arce, J., Foufoula-Georgiou, E., and Droegemeier, K. K. (2000). Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *Journal of Geophysical Research: Atmospheres*, 105(D8):10129–10146.