



Sicherstellung der Entwässerung küstennaher, urbaner Räume unter Berücksichtigung des Klimawandels – das BMBF-Forschungsprojekt Stuck.

2. Zwischenbericht aus dem Arbeitspaket 1:
Niederschlagsvorhersagen und Starkregenwarnungen aus Radarmessdaten, Radarnowcast und numerischer Wettervorhersage



Zwischenbericht von hydro & meteo GmbH & Co. KG zum Arbeitspaket 1, Teil 2: AP 1.2 und 1.3 - Kombinierte Niederschlagsvorhersage, Bewertung der Vorhersagequalität für die Hochwasserwarnung.

Förderkennzeichen 033W031, gefördert vom Bundesministerium für Bildung und Forschung (BMBF) innerhalb des Förderschwerpunkts Nachhaltiges Wassermanagement (NaWaM).



Autoren:

Alrun Jasper-Tönnies
Thomas Einfalt
Alexander Strehz



hydro & meteo GmbH & Co. KG
Breite Straße 6-8, 23552 Lübeck

Weitere Projektpartner:



Landesbetrieb Straßen, Brücken und Gewässer (LSBG)
Sachsenfeld 3-5, 20097 Hamburg
Prof. Dr. Gabriele Gönnert



Technische Universität Hamburg (TUHH)
Institut für Wasserbau, Denickestraße 22, 21073 Hamburg
Prof. Dr.-Ing. Peter Fröhle



Universität Hamburg (UHH), bestehend aus:
Biozentrum Klein Flottbek, Abt. Angewandte Pflanzenökologie
Ohnhorststraße 18, 22609 Hamburg
Prof. Dr. Kai Jensen

CEN, Institut für Bodenkunde
Allendeplatz 2, 20146 Hamburg
Prof. Dr. Annette Eschenbach



Hamburgisches WeltWirtschaftsinstitut gGmbH (HWWI)
Heimhuder Straße 71, 20148 Hamburg
Malte Jahn

Inhalt

1	Einleitung.....	5
2	Datengrundlage	5
2.1	Radardaten und Radardatenprodukte	5
2.2	COSMO-DE und COSMO-DE-EPS.....	6
2.2.1	Stationsmessdaten.....	7
3	Niederschlagsvorhersagen.....	8
3.1.1	Ensemblevorhersagen	8
3.2	Nowcasts	9
3.2.1	Nowcast-Ensembles	10
3.2.2	Interpolierte Vorhersage	10
4	Untersuchung und Validierung.....	11
4.1	Indizes für die Bewertung der Vorhersagequalität.....	11
4.1.1	Qualitätskriterien	11
4.1.2	Grenzwerte Niederschlag für die Pegelwarnung.....	12
4.2	Allgemeine Auswertungen zur Vorhersagequalität	13
4.2.1	Auswertungen der Nowcast-Ensembles	13
4.2.2	Auswertungen COSMO-DE-EPS	17
4.3	Vorhersagequalität für die Pegelwarnung	20
4.3.1	Vorhersagequalität der Nowcasts für die Pegelwarnung.....	20
4.3.2	Vorhersagequalität der COSMO-DE-EPS für die Pegelwarnung	22
4.3.3	Auswertungen zum Blending.....	23
4.3.4	Schlussfolgerungen	24
5	Kombinierte Niederschlagsvorhersagen – Umsetzung	24
5.1.1	Umsetzung des Blendings	24
5.2	Umsetzung der kombinierten Ensemblevorhersagen	26
5.3	Validierung der kombinierten Niederschlagsvorhersagen im Vergleich zu COSMO-DE.....	29
5.4	Ergebnisse für die Pegelwarnung 2017	31
6	Zusammenfassung und Ausblick.....	32

7	Danksagung.....	33
8	Literaturverzeichnis.....	34

1 Einleitung

Der vorliegende Zwischenbericht stellt Ergebnisse aus dem Arbeitspaket 1 Meteorologie und Vorhersage, bearbeitet durch hydro & meteo GmbH & Co. KG im Verbundforschungsprojekt „Sicherstellung der Entwässerung küstennaher, urbaner Räume unter Berücksichtigung des Klimawandels“ (Stuck) dar.

Das Projekt Stuck ist Teil der BMBF-Fördermaßnahme „Regionales Wasserressourcen-Management für den nachhaltigen Gewässerschutz in Deutschland“ (ReWaM). Eine Übersicht über das Projekt Stuck mit Berichten von allen Projektpartnern findet sich im projektübergreifenden 2. Zwischenbericht (LSBG 2016).

Als ein wichtiger Teil eines nachhaltigen Hochwassermanagements soll im Arbeitspaket 1 untersucht werden, wie sich kurzfristige Niederschlagsvorhersagen und die daraus resultierenden Hochwasserwarnungen verbessern lassen. Besonders in städtisch geprägten Räumen besteht bei der Hochwasserwarnung in Zusammenhang mit konvektiven Starkregen ein großer Bedarf für verbesserte Verfahren. Durch die Kombination von radarbasierten Ensemblevorhersagen und Ensemblevorhersagen aus der numerischen Wettervorhersage wird versucht, Verbesserungen im Vergleich zur aktuellen Hochwasserwarnung zu erzielen. Die resultierenden Vorhersagen werden am Beispiel des Warnsystems Binnenhochwasser Hamburg (WaBiHa, www.wabiha.de) untersucht, das der Landesbetrieb Straßen, Brücken und Gewässer (LSBG) betreibt. Die kombinierten Vorhersagen werden zudem den Projektpartnern als Eingangsdaten für hydrologische Modelle zur Verfügung gestellt (Arbeitspaket 2: Hydrologie und Flächenmanagement).

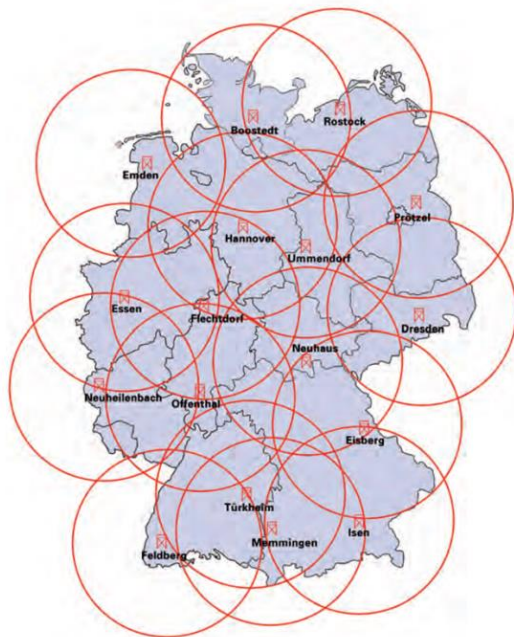
2 Datengrundlage

Für die Nutzung im Projekt Stuck werden vom Deutschen Wetterdienst meteorologische Daten bereitgestellt:

1. Radardaten und Radardatenprodukte
2. Numerische Wettervorhersagedaten
3. Stationsmessdaten

2.1 Radardaten und Radardatenprodukte

Im Rahmen des Projekts werden die Radarmessungen der Standorte Boostedt, Emden, Hannover und Rostock/Warnemünde (Abb. 1), sowie das angeeichte Radar-Deutschlandkomposit des DWD und der Radar-Nowcast des DWD mit einem Vorhersagezeitraum von bis zu zwei Stunden bezogen. Die Daten aller vier Radarstationen werden in Echtzeit abgerufen, archiviert und für die Weiterverarbeitung vorbereitet. Die Messungen der einzelnen Radarstandorte werden automatisiert geprüft und korrigiert. Aus den korrigierten Radardaten wird ein Komposit erstellt, welches den norddeutschen Raum abdeckt (Abb. 2). Anschließend werden die Radarmessungen unter Verwendung von Regenschreiberdaten an die am Boden gemessenen Werte angeeicht. Die kompositierten und angeeichten Radardaten werden im internen Bereich der Projektwebsite dargestellt. Eine Beschreibung der Aufbereitung der Radardaten inklusive Korrekturverfahren und der Aneichung mit Regenschreiberdaten ist im 2. Zwischenbericht des Projekts Stuck enthalten (Kapitel 4, LSBG 2016).



▲ Die 17 Radarstandorte des Deutschen Wetterdienstes - mit einer Reichweite von jeweils 150 km

Abb. 1: Das Radarnetzwerk des Deutschen Wetterdienstes bestehend aus 17 Standorten. Die Reichweite der einzelnen Radare von jeweils 150 km ist mit roten Kreisen gekennzeichnet (Quelle: DWD).

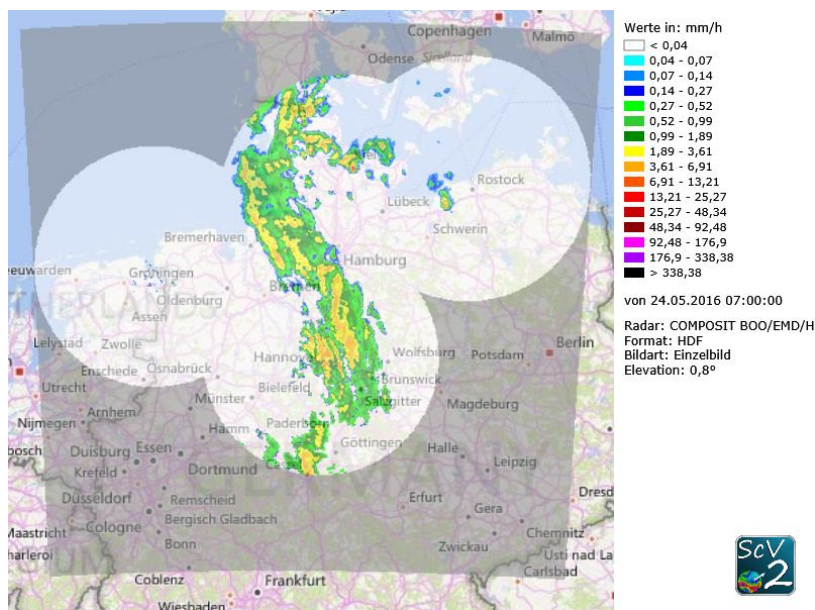


Abb. 2: Radarkomposit aus den Radardaten der Standorte Boostedt, Emden, Hannover und Rostock nach der Anreichung mit Stationsdaten des deutschen Wetterdienstes in mm/h.

2.2 COSMO-DE und COSMO-DE-EPS

Vom Deutschen Wetterdienst werden numerische Wettervorhersagen der Lokalm Modelle COSMO-DE und COSMO-DE-EPS bezogen. Die bezogenen Daten werden seit dem 7.9.2015 archiviert und direkt nach dem Abruf für die Weiterverarbeitung vorbereitet. Der räumliche Ausschnitt der bezogenen Modelldaten ist in Abb. 3 dargestellt.



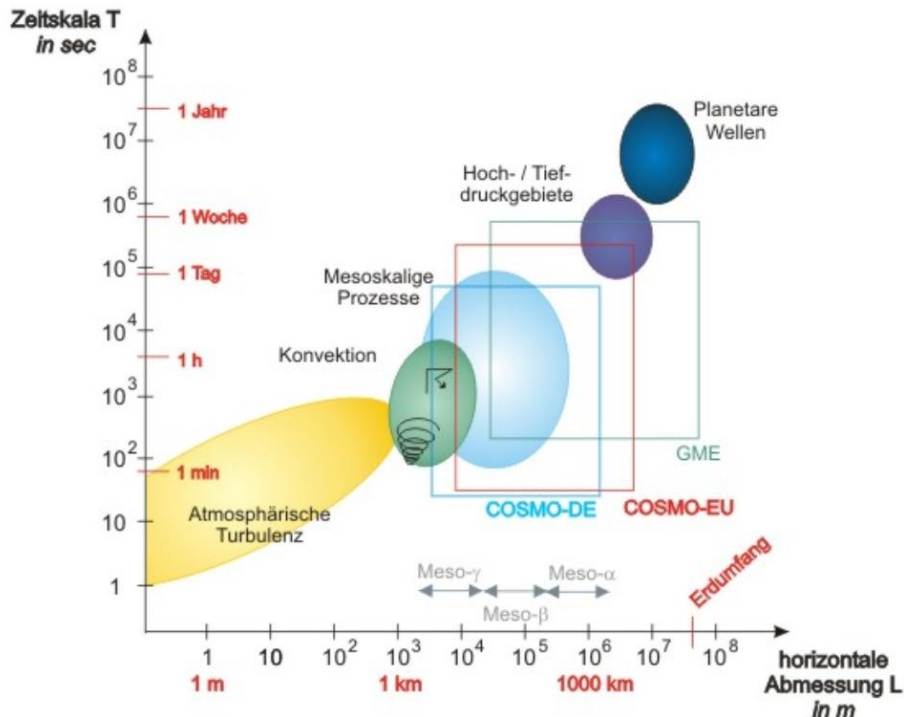
Abb. 3: Modellgebiet von COSMO-DE und Ausschnitt der im Projekt bezogenen Daten von COSMO-DE und COSMO-DE-EPS (rot umrahmter Bereich).

Das Modell COSMO-DE wurde vom Deutschen Wetterdienst als Modell zur kurzfristigen Wettervorhersage entwickelt (Baldauf et al., 2011). COSMO-DE ist ein nichthydrostatisches Modell, das mit einer horizontalen Auflösung von 0.025° (ca. 2.8 km) gerechnet wird, die es ermöglicht, hochreichende Konvektion explizit darzustellen. Das bedeutet, größere Gewitterzellen können zumindest grob simuliert werden und es kann bis zu einem gewissen Grad auf die Parametrisierung von Konvektion verzichtet werden. COSMO-DE wird vom Modell COSMO-EU angetrieben (Skalenbereiche siehe Abb. 4). Das Modell wird alle 3 Stunden gerechnet, der Vorhersagezeitraum betrug anfangs 18 Stunden und wurde auf 27 Stunden verlängert. Das Modellgebiet von COSMO-DE ist in Abb. 3 dargestellt.

Seit 2012 wird vom Deutschen Wetterdienst zusätzlich das Modell-Ensemble COSMO-DE-EPS operationell betrieben. COSMO-DE-EPS wird auf dem gleichen Gitter wie COSMO-DE (horizontale Gitterweite 0.025° , konvektionsauflösend) gerechnet und besteht aus 20 Ensembleläufen, die Variationen der Anfangsbedingungen, seitlichen Randbedingungen und Modellphysik beinhalten. Unter anderem werden auch die Parametrisierungen zur Konvektion und Wolkenphysik variiert. Vergleichsstudien zwischen COSMO-DE-EPS und dem deterministischen COSMO-DE ergaben eine erkennbare Verbesserung der Vorhersagequalität, gerade auch bei höheren Niederschlagsintensitäten (Kühnlein et al. 2014).

2.2.1 Stationsmessdaten

Messungen der Niederschlagshöhe stehen ab dem 01.01.2010 mit einer zeitlichen Auflösung von 5 Min von den Stationen Neuwiedenthal und Fuhlsbüttel für das Stadtgebiet Hamburg zur Verfügung. Seit dem 03.02.2016 werden die Daten dieser beiden Stationen nahezu in Echtzeit über das XML-Austauschformat des DWD in 1-minütiger Auflösung bezogen und archiviert. Eine Umwandlung in das Format UVF und die Weitergabe an das Arbeitspaket 2 läuft operationell seit Ende 2016.



Durch die verschiedenen Modelle aufgelöste Skalenbereiche und damit verbundene physikalische Prozesse.

Abb. 4: Durch die verschiedenen Modelle aufgelöste Skalenbereiche und damit verbundene physikalische Prozesse. Quelle: DWD

3 Niederschlagsvorhersagen

Die atmosphärischen Prozesse, die sich auf die Entstehung von Niederschlag auswirken, sind kleinräumig und komplex. Unter realen Bedingungen können diese Prozesse von Vorhersagemodellen nicht vollständig abgebildet werden. Die Vorhersage von Niederschlägen ist daher prinzipiell mit Unsicherheit verbunden. Bei konvektiven Niederschlagsereignissen wirkt sich diese Unsicherheit stärker aus als bei großflächigen, langanhaltenden Niederschlägen, wie sie z.B. in Zusammenhang mit einer Warmfront auftreten. Für die Niederschlagsvorhersage gibt es zwei grundlegend verschiedene Ansätze: Vorhersagen von numerischen Wettermodellen und Kurzfristvorhersagen (Nowcasts), die im Wesentlichen eine Extrapolation von Messdaten darstellen. Welcher Ansatz zu besseren Vorhersagen führt, ist vor allem vom Vorhersagehorizont abhängig. Im Projekt Stuck werden numerische Wettervorhersagen und Nowcasts genutzt und kombiniert, um für verschiedene Vorhersagehorizonte die jeweils besten Vorhersagen bereitzustellen.

In diesem Kapitel werden die radarbasierten Kurzfristvorhersagen vorgestellt. Die kombinierten Niederschlagsvorhersagen werden in Kapitel 5 beschrieben.

3.1.1 Ensemblevorhersagen

Ziel von Ensemblevorhersagen ist es, die Unsicherheiten der Vorhersage möglichst realistisch abzubilden und dadurch quantifizierbar zu machen. Bei einer Ensemblevorhersage lässt die Bandbreite und die Verteilung der Ensembleläufe Rückschlüsse darauf zu, wie stabil oder instabil die Wetterlage ist. Wenn Ensemblevorhersagen gemeinsam ausgewertet werden, können Größen wie der Median, Quantile oder Überschreitungswahrscheinlichkeiten abgeleitet werden. Mithilfe von Ensemblevorhersagen ist es beispielsweise möglich, in einem Warnsystem Warnungen in Abhängigkeit von der Wahrscheinlichkeit zu erstellen, mit der ein Ereignis eintritt oder ein Grenzwert überschritten wird. In

den folgenden Auswertungen liegt ein Schwerpunkt auf der Analyse von Ensemblevorhersagen, sowohl bei den numerischen Wettervorhersagen (mit dem Produkt COSMO-DE-EPS) als auch bei den Nowcasts, die als Ensemble mit 10 Ensembleläufen gerechnet werden.

3.2 Nowcasts

Nowcasts bzw. Kurzfristvorhersagen basieren auf der laufenden Analyse von aktuellen Messungen ‚in Echtzeit‘ – hier auf Niederschlagsmessungen von Radar und Regenschreibern. Damit diese ständig und in guter Qualität zur Verfügung stehen, müssen eine Reihe von Vorarbeiten durchgeführt werden. Diese wurden im 2. Zwischenbericht StucK (LSBG 2016) in Kapitel 4 dargestellt. Als Ergebnis liegen die Niederschlagsmessungen auf einem Gitter von 1 km x 1 km vor und werden alle 5 Minuten aktualisiert. Eine radarbasierte Kurzfristvorhersage hat einen Vorhersagehorizont von ca. 2-3 Stunden.

Für die Erstellung der Nowcasts wird ein Verfahren verwendet, das auf Zellerkennung (engl. cell tracking) basiert, und in der Software SCOUT (hydro&meteo 2009) implementiert ist. Darin werden die jeweils aktuellsten Niederschlagsmessungen mit zurückliegenden Messungen verglichen und die Bewegung der Niederschlagsfelder wird nachvollzogen. Einzelne Niederschlagszellen werden anhand einer Reihe von charakteristischen Eigenschaften wie Größe, Form und Reflektivität identifiziert. Ausgehend von den Masseschwerpunkten wird für jede wiedererkannte Niederschlagszelle ein Bewegungsvektor berechnet. Aus den Bewegungsvektoren wird ein Bewegungsvektorfeld erstellt. Darauf basierend wird mit einer Semi-Lagrange Methode berechnet, wie sich das Niederschlagsfeld in der Zukunft entwickeln könnte.

Im Projekt StucK wurde das Verfahren zur Interpolation des Bewegungsvektorfelds überarbeitet. Mit dem neuen Verfahren werden großräumige Divergenz und Rotation des Niederschlagsfeldes besser berücksichtigt. Abb. 5 zeigt ein Beispiel mit den Bewegungsvektorfeldern beider Methoden im Vergleich. Während in dem Beispiel in der linken Abbildung kaum Rotation erkennbar ist, ist in der rechten Abbildung eine zyklonale Rotation im Bewegungsvektorfeld sichtbar. Ein weiterer Vorteil der neuen Methode liegt in der größeren Robustheit gegenüber Ausreißern, die zum Beispiel durch eine ungenaue Zellerkennung entstehen können.

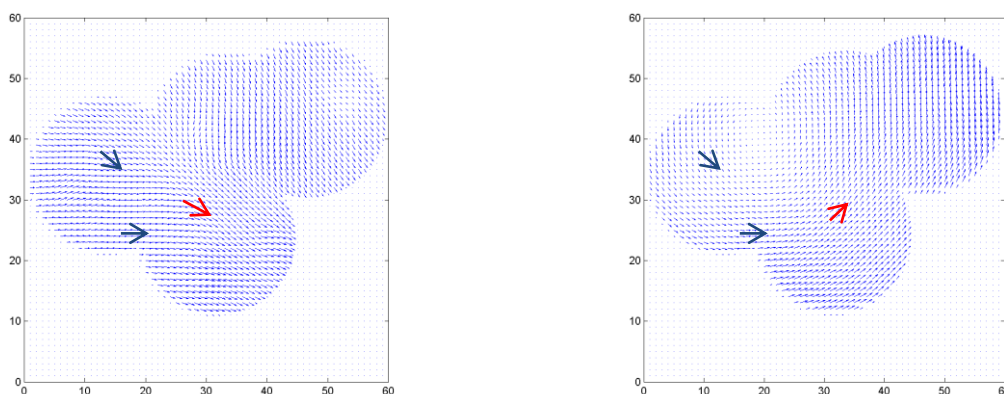


Abb. 5: Beispiel für unterschiedliche Bewegungsvektorfelder für das Radarkomposit. Das Feld links wurde mit dem ursprünglich in SCOUT implementierten Verfahren berechnet, das rechte Feld mit dem neuen Verfahren. Zur besseren Erkennung sind die blauen und roten Pfeile eingezeichnet, als Vergrößerung der darunterliegenden Bewegungsvektoren.

Eine weitere Überarbeitung betraf das Zellwachstum in der Nowcast-Routine. Darin wird für die vergangenen Radarmessungen nachvollzogen, wie sich die Größe und Intensität von Niederschlagszellen ändert. Diese Informationen werden zur Extrapolation des Niederschlagsfeldes genutzt. Allerdings verändern sich Niederschlagszellen sehr schnell in Form und Größe, insbesondere bei kleinräumigen konvektiven Ereignissen. Durch die reine Extrapolation der vorangegangenen Änderungen lassen sich daher die tatsächlichen Veränderungen des Niederschlagsfeldes über 1-3 Stunden in der Regel nicht gut vorhersagen. Die Wachstumsroutinen werden aber bei der Erstellung von Ensemble-Vorhersagen genutzt, um verschiedene mögliche Entwicklungen des Niederschlagsfeldes abzubilden. Im Zuge der Bearbeitung wurden die Einstellungen mehrerer Modellparameter, die sich auf die Stärke des Zellwachstums und die Variation bei der Ensembleberechnung auswirken, getestet und angepasst.

3.2.1 Nowcast-Ensembles

Ziel der Nowcast-Ensembles ist es, die Unsicherheiten in der Radarvorhersage realistisch abzubilden, so dass Warnungen auch in Abhängigkeit von der Wahrscheinlichkeit für eine Grenzwertüberschreitung erstellt werden können. Um ein Nowcast-Ensemble zu erzeugen, werden zunächst der Mittelwert und die Standardabweichung der Bewegungsvektoren und der Wachstumsraten der Zellen (Größe und Intensität) über die letzten Zeitschritte berechnet. Diese sind ein Maß für die Streuung der Ausgangswerte und werden dazu benutzt, um für jeden Ensemblelauf neue Bewegungsvektoren und Wachstumsraten zu generieren. Mit diesen unterschiedlichen Anfangsbedingungen werden die jeweiligen Ensembleläufe berechnet. Auf diese Weise enthält das Vorhersageensemble sowohl Unsicherheiten in der Zellerkennung mit SCOUT als auch reale Variationen in der Entwicklung der Zellen (Tessendorf & Einfalt 2012). Weiterhin werden Unsicherheiten durch schnell veränderliche Wetterlagen und Messunsicherheiten abgeschätzt und gehen in die Berechnung der Ensembles mit ein.

3.2.2 Interpolierte Vorhersage

Bei der Berechnung von Niederschlagsmengen aus Radardaten wird häufig angenommen, dass die zeitlichen und räumlichen Veränderungen zwischen zwei aufeinanderfolgenden Messungen (zeitlicher Abstand 5 Minuten) zu vernachlässigen sind. Unter dieser Annahme kann die Niederschlagssumme über 5 min durch eine einfache Umrechnung aus der Niederschlagsintensität abgeleitet werden. Insbesondere bei schnell ziehenden Niederschlagsgebieten kann diese Annahme allerdings zu Fehlern führen. In diesem Fall ist es sinnvoll, zwischen zwei aufeinanderfolgenden Messungen zu interpolieren („zeitliche Interpolation“). Das verwendete Verfahren wird in Jasper-Tönnies & Jessen (2014) beschrieben und auch bei der Erstellung der Nowcasts angewandt. Ein Beispiel vom 05.05.2015 ist in Abb. 6: Niederschlagssummen über 1 h von einer radarbasierten Kurzfristvorhersage für den 05.05.2015. Die linke Abbildung dargestellt. In der Stundensumme von einer Vorhersage ohne zeitliche Interpolation (linke Abbildung) scheint die Niederschlagszelle zu springen. Die Vorhersage mit interpolierten Niederschlägen (rechte Abbildung) zeigt eine kontinuierlichere Zugbahn.

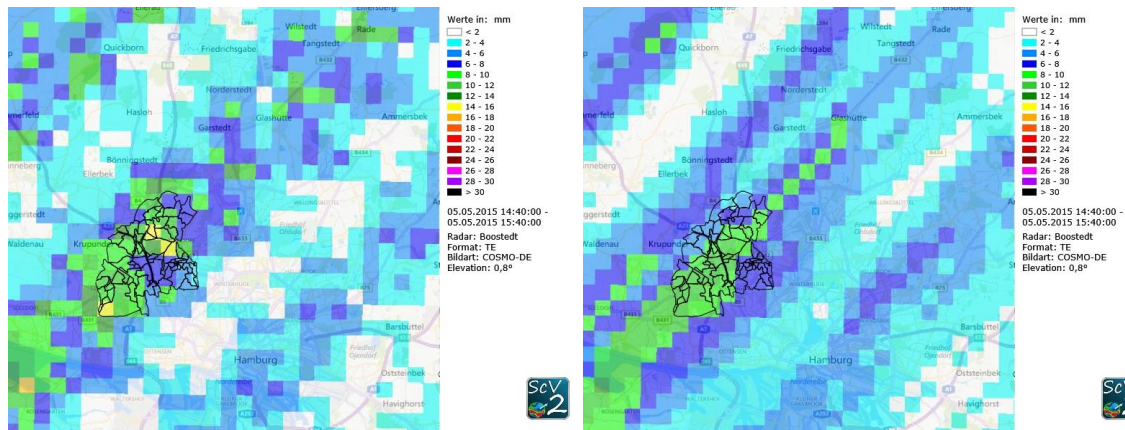


Abb. 6: Niederschlagssummen über 1 h von einer radarbasierten Kurzfristvorhersage für den 05.05.2015. Die linke Abbildung stellt die Stundensumme der Vorhersage ohne zeitliche Interpolation, die rechte Abbildung die Stundensumme mit zeitlicher Interpolation dar.

4 Untersuchung und Validierung

Die numerischen Niederschlagsvorhersagen und die radarbasierten Kurzfristvorhersagen wurden gemeinsam ausgewertet. Untersucht wurde der Zeitraum: 01.06. - 31.08.2016. Im Sommer 2016 traten in Norddeutschland und Hamburg viele konvektive Niederschlagsereignisse auf, teilweise mit Starkregen und Pegelüberschreitungen an den Hamburger Pegeln. Fragestellungen der Auswertungen waren insbesondere:

- Wie gut ist die Vorhersagequalität der Nowcasts im Vergleich zu COSMO-DE und COSMO-DE-EPS, in Abhängigkeit vom Vorhersagezeitraum?
- Wie können die Ensemble-Informationen von Nowcasts und COSMO-DE-EPS am besten für die Starkregenwarnung genutzt werden, z. B. über Ensemble-Quantile, Überschreitungswahrscheinlichkeiten,...?
- Welche Grenzwerte des Niederschlags sind relevant für die Erzeugung von Warnungen im Warnsystem WaBiHa? Ab welchen Niederschlagsmengen treten Überschreitungen der Warn Grenzen an den Hamburger Pegeln auf?

4.1 Indizes für die Bewertung der Vorhersagequalität

Ein wesentliches Kriterium für die Bewertung der Vorhersagequalität der Niederschlagsvorhersagen ist, wie gut damit die Hochwasserwarnung für die Hamburger Pegel ausfällt. Um dies auszuwerten werden zwei verschiedene Datensätze zugrunde gelegt. Einerseits wird untersucht, wie gut die Vorhersagequalität für die Überschreitung von kritischen Niederschlagssummen (siehe 4.1.2) ist (Vergleich Vorhersage Niederschlag – Messung Niederschlag). Andererseits wird geprüft, wie häufig auf Basis der Grenzwerte die Überschreitung von Warnschwellen an den Pegeln richtig vorhergesagt wird (Vergleich Vorhersage Niederschlag – Messung Pegel).

4.1.1 Qualitätskriterien

Die Auswertung erfolgt auf Basis der Kontingenztafel (Tab. 1). Diese teilt jede Vorhersage und jede Messung in die Kategorien A-D ein, als *Treffer* (A), *Falschen Alarm* (B), *Verpasst* (C) oder *Richtigerweise kein Alarm* (D). Aus der Anzahl aller Ereignisse, die in die jeweilige Kategorie fallen, können Wahrscheinlichkeiten dafür abgeleitet werden, wie häufig Ereignisse richtig vorhergesagt werden

bzw. falsche Alarme auftreten. Dafür gibt es die folgenden 2 Qualitätsindizes (Donaldson *et al.*, 1975):

1. *Trefferrate*: Die Wahrscheinlichkeit, dass aufgetretene Ereignisse rechtzeitig vorhergesagt wurden (engl. *Hit Rate* oder *Probability of Detection*): $POD = A / (A + C)$
2. *Anteil falsche Alarme*: die Wahrscheinlichkeit, dass sich die Vorhersage ‚der Schwellwert wird überschritten‘ als falscher Alarm erweist, (engl. *False Alarm Ratio*): $FAR = B / (A + B)$

Diese Indizes werden speziell zur Auswertung von seltenen Ereignissen verwendet. Der Anteil der falschen Alarme wird hier als Anteil aller Alarme ($A + B$) berechnet, ist also unabhängig davon, wie häufig richtigerweise kein Alarm ausgegeben wurde (D).

	Messung		
Vorhersage		Schwellwert überschritten	Schwellwert nicht überschritten
	Schwellwert überschritten	A	B
	Schwellwert nicht überschritten	C	D

$$POD = A / (A + C)$$

$$FAR = B / (A + B)$$

Tab. 1: Kontingenztabelle für die Bewertung von Vorhersagequalität

Die Indizes stellen ein gutes Maß für die Qualität von deterministischen Ereignisvorhersagen dar. Um zusätzlich die Qualität der Ensemblevorhersagen zu erfassen, wurden sie in Abhängigkeit von der Überschreitungswahrscheinlichkeit berechnet, die sich aus den Ensemblevorhersagen ergibt. Daraus resultiert für jede Überschreitungswahrscheinlichkeit von 0.05 – 1.0 (mit anderen Worten: in 1/20 - 20/20 Vorhersageläufen wird der Grenzwert überschritten) ein eigener Wert. Diese wurden in den folgenden Auswertungen gegeneinander aufgetragen, so dass die Ergebnisse für das ganze Ensemble in einer Abbildung zusammengefasst sind.

Zusätzlich gibt es Indizes, um die Qualität von Ensemble-Vorhersagen mit deterministischen Vorhersagen zu vergleichen, wie den Brier-Skill Score (Wilks, 1995), und Auswertungen, die die Verteilung der Ensembleläufe untersuchen (Reliability-Diagramme und Talagrand-Diagramme). Diese werden ergänzend gezeigt.

4.1.2 Niederschlagsgrenzwerte für die Pegelwarnung

Im Rahmen des Projekts wurde untersucht, welche Kriterien für die Pegelwarnung an den Hamburger Binnengewässern relevant sind und welche Grenzwerte des Niederschlags verwendet werden können, um einen kritischen Niederschlag, der zu einer Pegelwarnung führen kann, zu erkennen. Hierzu wurden zunächst die Niederschläge über dem Testgebiet (eine Fläche von 25 km x 25 km über dem Stadtgebiet Hamburg) mit den beobachteten Überschreitungen von Warnschwellen aller im Warn-dienst WaBiHa aufgeführten Pegel für den Sommer 2016 (01.06.-31.08.2016) verglichen. Als Grundlage an Niederschlagsdaten dienten die auf einem 5 km x 5 km-Gitter aggregierten Radarniederschläge. Die Daten wurden auf die Frage hin ausgewertet, welche Ereignisse durch Starkregen verursacht wurden, und ab welcher Niederschlagshöhe mit einer Überschreitung der mittleren Warngrenze an mindestens einem Pegel zu rechnen war.

In dem Zeitraum gab es 34 Ereignisse, in denen an mindestens einem der Hamburger Pegel eine Warngrenze überschritten wurde. Von der Auswertung ausgeschlossen wurden 3 Ereignisse, bei denen die Radardaten nicht vollständig vorlagen (Ausfall des Radars Boostedt), sowie 4 Pegelüberschreitungen, bei denen der Pegelstand durch ein vorhergehendes Niederschlagsereignis stark erhöht war und es durch einen sehr geringen Niederschlag erneut zu einer Überschreitung kam. Die übrigen Überschreitungen wurden in 19 Fällen durch kurze konvektive Starkregen verursacht und in 8 Fällen durch weniger intensive Niederschläge über einen Zeitraum von mehreren Stunden. Hierbei traten sowohl großräumige Niederschlagsgebiete auf als auch Wettersituationen mit vielen kleinen Niederschlagszellen, von denen mehrere über das Hamburger Stadtgebiet zogen. Die 19 Fälle mit kurzen Starkregen wurden genauer betrachtet. Daraus ergaben sich folgende Grenzwerte für den Niederschlag:

1. Niederschlagssumme über eine Stunde > 7 mm in mindestens einer Gitterzelle von $5 \text{ km} \times 5 \text{ km}$ (mittlerer Niederschlag).
2. Niederschlagssumme über eine Stunde > 10 mm an mindestens einem Einzelpixel ($1 \text{ km} \times 1 \text{ km}$)

In fast allen betrachteten Fällen waren beide Kriterien äquivalent. Da das 1. Kriterium (> 7 mm) über eine größere Fläche berechnet wird, ist es etwas robuster als das 2. Kriterium und wurde in den anschließend durchgeführten Auswertungen verwendet (siehe 4.3.1 und 0). Das 1. Kriterium wurde in 14 der 19 Ereignisse mit Pegelüberschreitungen erfüllt, in 3 Fällen wurde der Grenzwert nur knapp unterschritten und in 2 Fällen war die gemessene Niederschlagssumme deutlich < 7 mm. D.h. es ist nicht ausgeschlossen, dass ein Pegelstand über die (mittlere) Warngrenze steigt, ohne dass der Niederschlagsgrenzwert überschritten wurde. Größere Überschreitungen an mehreren Pegeln traten aber in diesen Fällen nicht auf. Zusätzlich war das Kriterium im Auswertezeitraum drei weitere Male erfüllt, ohne dass eine Pegelüberschreitung beobachtet wurde. D.h. in den meisten Fällen (14 von 17) kommt es bei einem Niederschlag über dem Grenzwert auch zu einer Pegelüberschreitung.

4.2 Allgemeine Auswertungen zur Vorhersagequalität

4.2.1 Auswertungen der Nowcast-Ensembles

Ein Beispiel für die vorhergesagte Niederschlagsintensität im Vergleich zur Radarmessung ist in Abbildung 7 gezeigt. Dargestellt ist die Radarmessung vom 26.04.16 um 11:40 Uhr und die Nowcast für den gleichen Zeitpunkt, die auf Basis der Radardaten bis 10:40 Uhr berechnet wurde (1 h-Vorhersage). Die Lage der Niederschlagszellen in Messung und Vorhersage stimmt relativ gut überein. Man erkennt allerdings, dass einige Niederschlagszellen unterschiedlich aussehen. Die Form und Größe der Niederschlagszellen in der Vorhersage entspricht noch weitgehend dem Stand der eine Stunde vorher gemessenen Niederschläge, während sich die realen Zellen verändert haben.

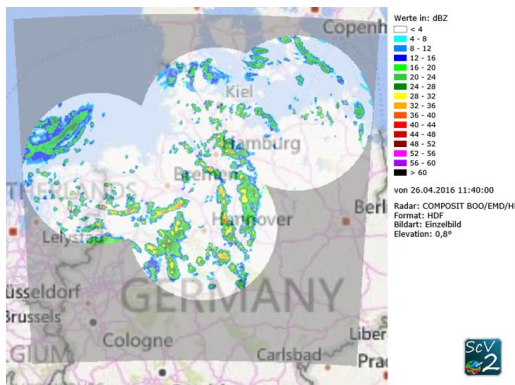
Dies veranschaulicht ein Grundproblem der kurzfristigen Niederschlagsvorhersage: während in vielen Fällen gut vorhergesagt wird, wie sich die Zellen weiterbewegen, können die Veränderungen von Form und Größe nur sehr eingeschränkt vorhergesagt werden. Eine weitere Eigenschaft speziell der Nowcasts erkennt man in den leeren Bereichen am westlichen und südlichen Rand des Komposits.

Die Niederschläge, die dort in der Radarmessung zu sehen sind, lagen zum Zeitpunkt der Vorhersageberechnung noch außerhalb der Sichtweite der Radars. Da die Nowcasts ausschließlich auf den Radarmessungen beruhen, kann für die entsprechenden Bereiche keine Vorhersage erstellt werden.

Bei der Berechnung von Ensemblevorhersagen wird die Intensität und Zugrichtung der Niederschlagszellen variiert. Abbildung 8 zeigt ein Beispiel von einem konvektiven Starkregenereignis vom

28.08.2016. Dargestellt sind die Niederschlagssummen über einen Vorhersagezeitraum von 2 h von 10 Ensemblevorhersagen von 14:00 UTC und von der Radarmessung (14:00-16:00 UTC). Das Beispiel zeigt das Einzugsgebiet der Kollau. Die Ensemblevorhersagen unterscheiden sich in der Gesamtniederschlagsmenge und der Position des maximalen Niederschlags. Die gemessene Niederschlagsmenge im Einzugsgebiet der Kollau liegt innerhalb der Bandbreite des Ensembles, etwas oberhalb des Medians der Ensemblevorhersagen.

Radarmessung 26.4.16
11:40 Uhr



Vorhersage
11:40 Uhr (+ 60 min)

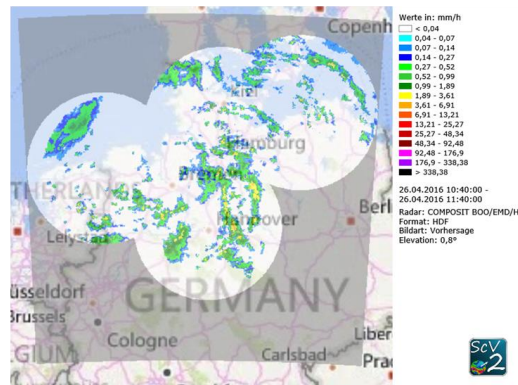


Abbildung 7: Beispiel 26.04.2016: Links: Radarmessung der Niederschlagsintensität um 11:40 Uhr in dBZ, rechts: vorhergesagte Niederschlagsintensität: Nowcast von 10:40 Uhr (+ 60 min) in mm/h. Die Farbskala der rechten Grafik ist über die Beziehung $Z[\text{dBZ}] = 200 \times R[\text{mm/h}]^{1.6}$ in die Farbskala der linken Grafik umgerechnet, so dass die Bilder optisch vergleichbar sind.

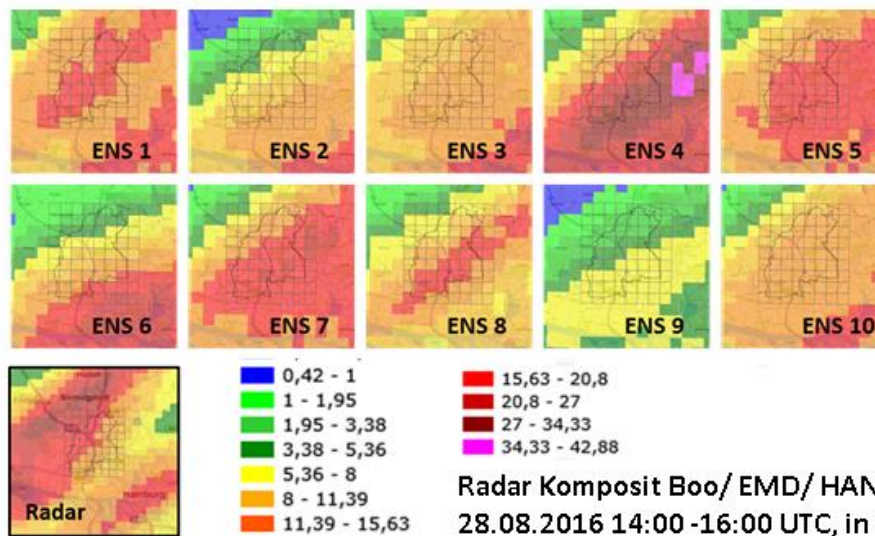


Abbildung 8: Vorhergesagte Niederschlagssumme von 10 Ensemblesnowcasts vom 28.08.2016 14:00 UTC über 2 h und die vom Radar gemessene Niederschlagssumme (u. l.) über den gleichen Zeitraum im Einzugsgebiet der Kollau.

In einer Auswertung über 3 Monate vom 01.06. - 31.08.2016 wurden die Nowcast-Ensembles mit gemessenen Niederschlägen verglichen. In diesem Abschnitt sind die Ergebnisse der statistischen Verteilung der Niederschläge und der Verteilung der Ensemblevorhersagen im Vergleich zu den Messungen dargestellt. Eine Auswertung der Vorhersagequalität in Hinblick auf Starkregen folgt in Abschnitt 4.3.

Abb. 9 zeigt die kumulierte Häufigkeitsverteilung von hohen Niederschlagssummen über 1 h von den Messdaten und Vorhersagedaten, sowohl von einer einzelnen Vorhersage (magenta) als auch vom Vorhersageensemble (rot). Der Bildausschnitt zeigt das höchste 1 % aller Niederschlagsmessungen. Der Unterschied zwischen der linken und der rechten Abbildung besteht darin, dass auf der linken Seite die Ergebnisse der Vorhersagen mit der zuvor verwendeten Wachstumsroutine dargestellt sind; die Vorhersagen rechts wurden mit der neuen Wachstumsroutine berechnet (siehe 3.2). In der linken Abbildung fallen die vorhergesagten Niederschläge höher aus als die gemessenen Niederschläge. Das deutet darauf hin, dass durch die Wachstumsroutine systematisch zu hohe Niederschläge produziert wurden. Bei den überarbeiteten Vorhersagen (rechte Seite) fällt die Niederschlagsverteilung ähnlich aus wie bei den Messungen, d. h. die Niederschlagsverteilung wurde verbessert und ein systematischer Fehler ist nicht mehr erkennbar. Die folgenden Auswertungen basieren auf der Vorhersage mit der neuen Wachstumsroutine.

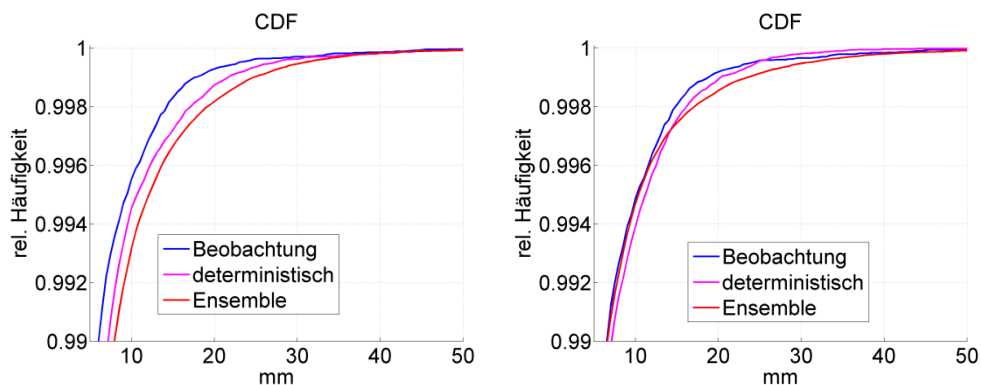


Abb. 9: Relative Häufigkeitsverteilung von gemessenen und vorhergesagten Niederschlägen (Niederschlagssummen über 1 h, kumulierte Darstellung). Blau: Radarmessung, magenta: Nowcast über 1 h (einzelner Vorhersagelauf), rot: Nowcasts über 1 h – Ensemble. Dargestellt sind Werte oberhalb des 0.99 Quantils, d.h. das oberste 1 % aller gemessenen bzw. vorhergesagten Niederschläge. Die linke Abbildung zeigt die Verteilung der Nowcasts bei Verwendung der alten Wachstumsroutine, die rechte die Ergebnisse nach der Überarbeitung der Wachstumsroutine.

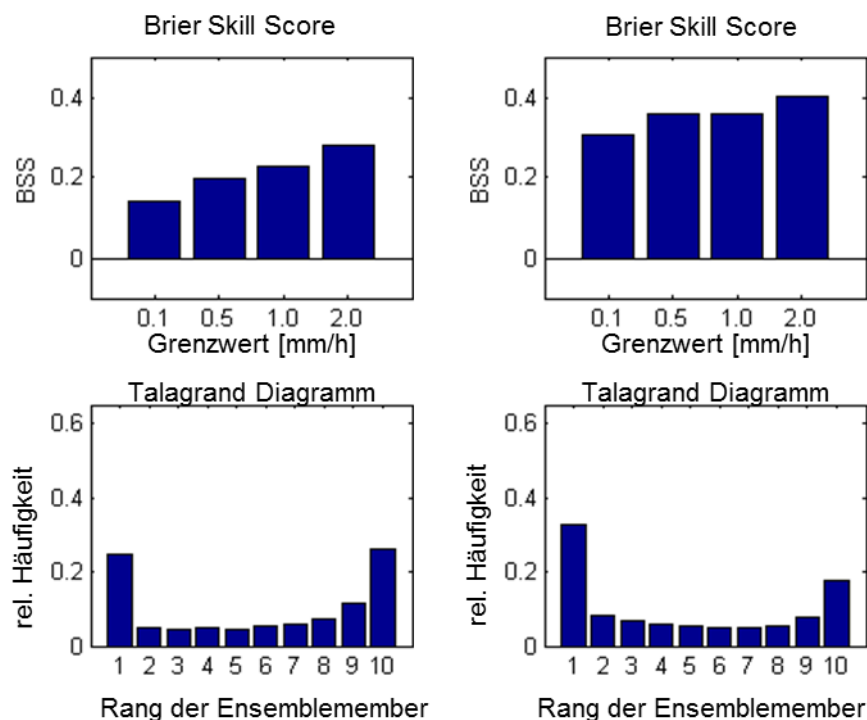


Abbildung 10: Brier Skill Score (BSS) und Talagrand Diagramm der Nowcasts für die 1. Stunde (links) und die 2. Stunde der Vorhersage (rechts). Die Höhe der Balken im Talagrand Diagramm gibt die normierte Häufigkeit der jeweiligen Ensemblevorhersagen an. Der Rang 1, ..., 10 der Vorhersagen steht für die vorhergesagte Niederschlagsmenge, sortiert vom niedrigsten bis zum höchsten Niederschlag.

In Abbildung 10 sind Ergebnisse von Auswertungen der Ensemblevorhersagen (Vergleich mit einer Einzelvorhersage und Bandbreite der Vorhersagen) dargestellt. Die obere Abbildung zeigt den Brier Skill Score (BSS) in Abhängigkeit von der Niederschlagsintensität. Dieser ist ein Maß dafür, wieviel zusätzliche Information die Ensemblevorhersagen im Vergleich zu einer einzelnen Vorhersage liefern. Je höher der Wert ist, desto mehr Nutzen ergibt sich aus der Ensemblevorhersage (0 = kein Zugesinn). Aus der Abbildung geht hervor, dass der BSS mit zunehmender Niederschlagsintensität ansteigt. Im Vergleich zwischen der 1. Stunde (links) und der 2. Stunde (rechts) der Vorhersage fällt außerdem der BSS in der 2. Stunde höher aus. Mit dem längeren Vorhersagezeitraum und mit höheren Niederschlägen steigt die Unsicherheit der Vorhersagen an. Das Ergebnis deutet darauf hin, dass die Ensemblevorhersagen gerade bei unsicheren Vorhersagen einen erhöhten Nutzen gegenüber einer einzelnen Vorhersage bringen. Das Talagrand-Diagramm (untere Abbildung) illustriert, wie gut die Bandbreite der Ensemblevorhersagen die Unsicherheit der Vorhersage abbildet. Die Höhe der Balken gibt an, wie häufig welche Ensemblevorhersage (sortiert nach der Größe, vom niedrigsten bis zum höchsten vorhergesagten Niederschlag) am besten zum gemessenen Niederschlag passt. Bei einem optimal verteilten Ensemble hätten alle Balken ungefähr die gleiche Höhe. Ausgewertet wurden ausschließlich Zeiten, zu denen Niederschlag gemessen oder vorhergesagt wurde. In 20-30 % dieser Fälle trifft die Ensemble-Vorhersage mit der niedrigsten Niederschlagsmenge am besten die Radarmessung. Die Verteilung zeigt, dass die niedrigste und die höchste Niederschlagsvorhersage aus dem Ensemble häufiger vorkommen als die mittleren Vorhersagen und somit die Bandbreite, die vom Ensemble vorhergesagt wird, nicht immer groß genug ist. Ein Grund hierfür könnte in der zeitlichen und räumlichen Ungenauigkeit der Nowcasts liegen: eine Niederschlagszelle, die an der falschen Stelle vorhergesagt wurde, kann dazu führen, dass an einem Ort die niedrigste Ensemblevorhersage am besten zutrifft (dort, wo sie fälschlicherweise vorhergesagt wurde) und an einem anderen Ort die höchste (dort, wo sie auftrat, aber nicht vorhergesagt wurde). Im Vergleich zwischen der 1. Stunde und der 2. Stunde der Vorhersage wird die Ensembleverteilung in der 2. Stunde schlechter. Gleichzeitig nimmt der Brier Skill Score zu.

Abbildung 11 zeigt eine weitere Auswertung des Nowcast-Ensembles über eine Stunde („Reliability“). Aus dieser Auswertung geht hervor, wie die vorhergesagte Wahrscheinlichkeit im Untersuchungszeitraum mit der tatsächlichen Häufigkeit von Ereignissen übereinstimmt. Ergebnisse, die genau auf der eingezeichneten Diagonalen liegen, bedeuten, dass die vorhergesagte Wahrscheinlichkeit mit der tatsächlichen Eintrittswahrscheinlichkeit übereinstimmt.

Datenpunkte im grau hinterlegten Bereich deuten an, dass die vorhergesagte Wahrscheinlichkeit nicht gut mit der tatsächlichen Eintrittshäufigkeit übereinstimmt. Die Auswertung wird für verschiedene Grenzwerte des Niederschlags (0.5-3 mm/h) gezeigt. Bei den niedrigen Grenzwerten von 0.5 und 1 mm/h liegen die Ergebnisse des Nowcast-Ensembles nah an der Diagonalen, bei den höheren Grenzwerten sind sie weiter entfernt. Das bedeutet z. B. für den Grenzwert 0.5 mm/h: wenn alle Ensemblevorhersagen Niederschlag oberhalb des Grenzwerts vorhersagen (vorhergesagte Wahrscheinlichkeit: 1.0) treten in ca. 90 % der Fälle tatsächlich Niederschläge oberhalb des Grenzwerts auf. Bei den höheren Grenzwerten sinkt dieser Wert auf ca. 70-80 %. Das heißt, für niedrige Grenzwerte stimmt die vorhergesagte Wahrscheinlichkeit gut mit der tatsächlichen Häufigkeit der Ereignisse überein, bei höheren Grenzwerten ist die vorhergesagte Wahrscheinlichkeit weniger zutreffend.

Doch auch bei den höheren Grenzwerten nimmt die beobachtete Häufigkeit kontinuierlich mit der vorhergesagten Wahrscheinlichkeit zu.

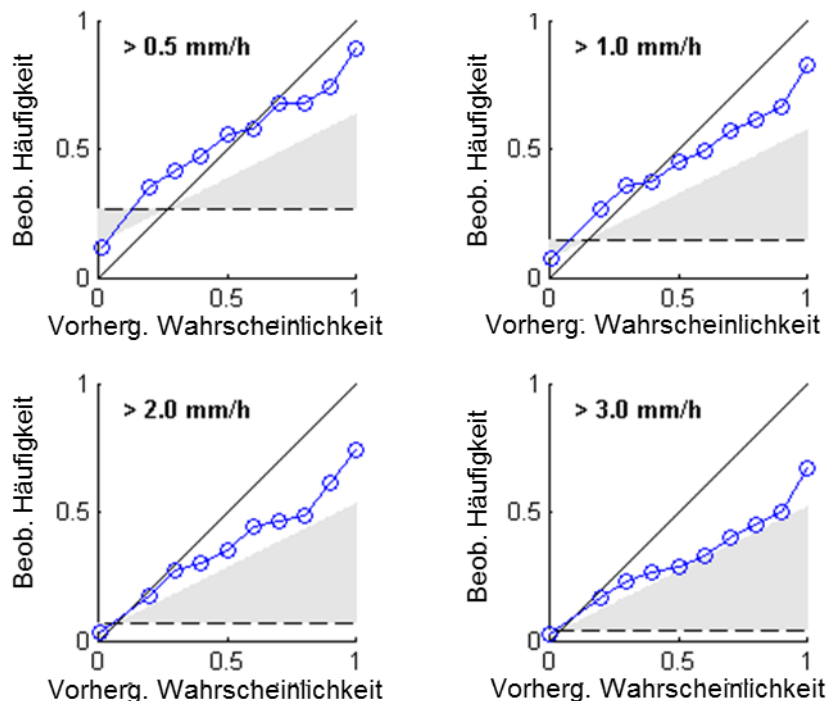


Abbildung 11: Reliability Diagramm der Nowcasts (1h-Vorhersage), Sommer 2016

4.2.2 Auswertungen COSMO-DE-EPS

Abbildung 12 zeigt ein Fallbeispiel mit einzelnen Ensemblevorhersagen von COSMO-DE-EPS im Vergleich zur Radarmessung für 3 aufeinanderfolgende Stunden. In diesem Zeitraum zieht ein Niederschlagsgebiet, dessen Zentrum sich zunächst nördlich von Fehmarn befindet, nach Osten. Es wird schwächer und schmaler, während im Süden des Gebiets eine neue Niederschlagszelle mit höherer Niederschlagsintensität entsteht. Keine der COSMO-DE-EPS Vorhersagen sagt die zeitliche Veränderung und die Zuggeschwindigkeit des Niederschlagsgebiets treffend voraus. Dennoch gibt es zu zwei Zeitpunkten eine Vorhersage, die relativ gut zu den Messungen passt. Von 18-19 Uhr passt der 3. Ensemblelauf recht gut, während das Niederschlagsgebiet im 1. Ensemblelauf zu weit östlich liegt, von 20-21 Uhr passt hingegen der 1. Ensemblelauf am besten zur Messung.

Die Ensemblevorhersagen von COSMO-DE-EPS wurden analog zu den Nowcast-Ensembles über den Zeitraum 01.06.16-31.08.16 ausgewertet. Untersucht wurden die Vorhersagezeiträume 4-6 h und 7-9 h mit jeweils 20 Ensembleläufen. Die Ergebnisse sind in Abbildung 13 und Abbildung 14 dargestellt. Auch bei den COSMO-DE-EPS Vorhersagen nimmt der BSS mit zunehmender Niederschlagsintensität und längerem Vorhersagezeitraum zu. Die beiden Vorhersagezeiträume unterscheiden sich aber nicht so deutlich wie bei den Nowcast-Ensembles. Das Talagrand Diagramm zeigt eine relativ gleichmäßige Verteilung der Vorhersagen von der zweitniedrigsten bis zur zweithöchsten Vorhersage. Hingegen kommt die niedrigste Vorhersage in dem Diagramm deutlich häufiger vor, passt also häufiger am besten zu den Messungen. Auch das Vorkommen der höchsten Vorhersage ist leicht erhöht. Ähnlich wie bei den Nowcast-Ensembles wird keine optimale Verteilung erzielt. Eine Ursache dürfte wiederum in der räumlichen und zeitlichen Ungenauigkeit der Vorhersagen liegen.

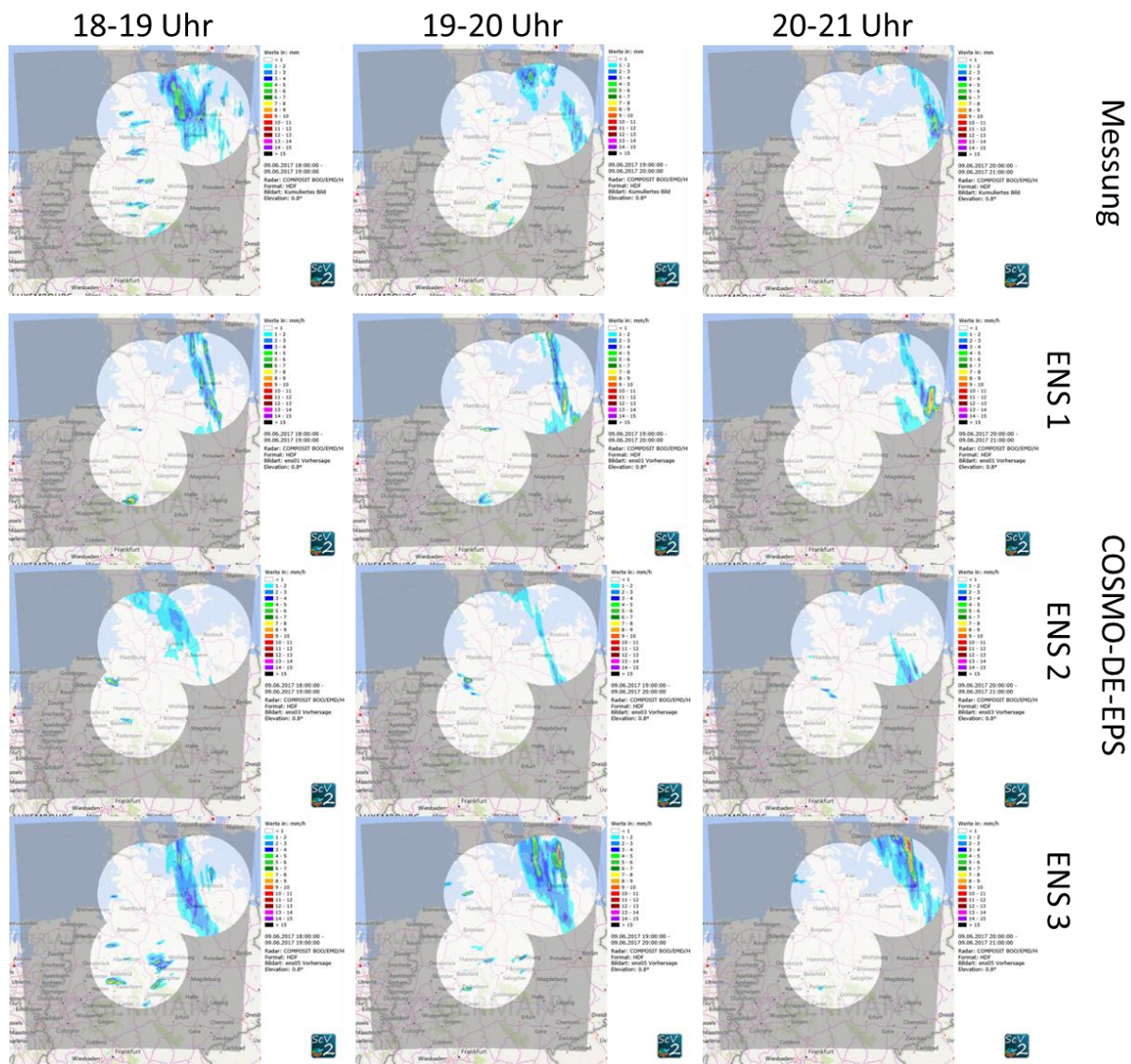


Abbildung 12 Fallbeispiel 09.06.2017: 3 Ensembleläufe von 18-21 Uhr von COSMO-DE-EPS und Radarmessungen (oben).

Abbildung 14 zeigt die vorhergesagten Wahrscheinlichkeit von COSMO-DE-EPS im Vergleich zur beobachteten Häufigkeit („Reliability Diagramm“, vgl. Auswertung der Nowcasts). Die Ergebnisse weichen teilweise deutlich stärker von der Diagonalen ab, als es bei den Nowcast-Ensembles der Fall ist. Das bedeutet, dass die vorhergesagte Wahrscheinlichkeit weniger über die tatsächliche Eintrittswahrscheinlichkeit aussagt. So wurde nur in ca 50 % der Fälle, in denen alle Ensembleläufe Niederschlag über 0.1 mm/h vorhergesagt haben, auch Niederschlag über diesem Grenzwert gemessen. Bei den höheren Grenzwerten von 1.0 und 2.0 mm/h macht sich bemerkbar, dass weniger Datenpunkte in die Auswertung eingehen als bei den Nowcasts, da die Vorhersagen von COSMO-DE-EPS nur alle 3 h produziert werden. Es kommen nur wenige Fälle vor, in denen mehr als 50 % der Ensemblevorhersagen Niederschlag oberhalb von 2.0 mm/h vorhersagen. In diesem Bereich liefert die Auswertung keine verwendbaren Ergebnisse zur tatsächlichen Eintrittswahrscheinlichkeit. Allerdings ist die Tatsache an sich, dass nur in sehr wenigen Fällen mehr als 50 % der Ensemblevorhersagen an dem gleichen Gitterpunkt Niederschlag oberhalb von 2.0 mm/h vorhersagen, ein wichtiges Ergebnis für die Interpretation der Ensemblevorhersagen. Auf diesen Aspekt wird in Kapitel 4.3 in Zusammenhang mit der Untersuchung der COSMO-DE-EPS und Nowcasts bei stärkeren Niederschlägen noch weiter eingegangen.

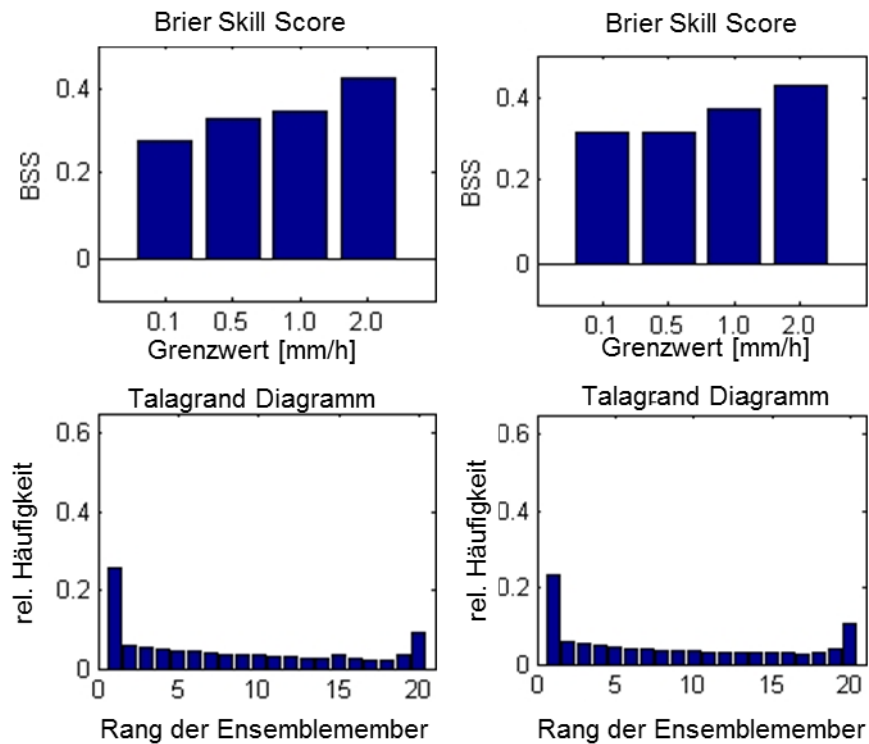


Abbildung 13: Brier Skill Score (BSS) und Talagrand Diagramm von COSMO-DE-EPS Vorhersagen, links von Vorhersagen für den Vorhersagezeitraum 4-6 h, rechts für den Vorhersagezeitraum 7-9 h

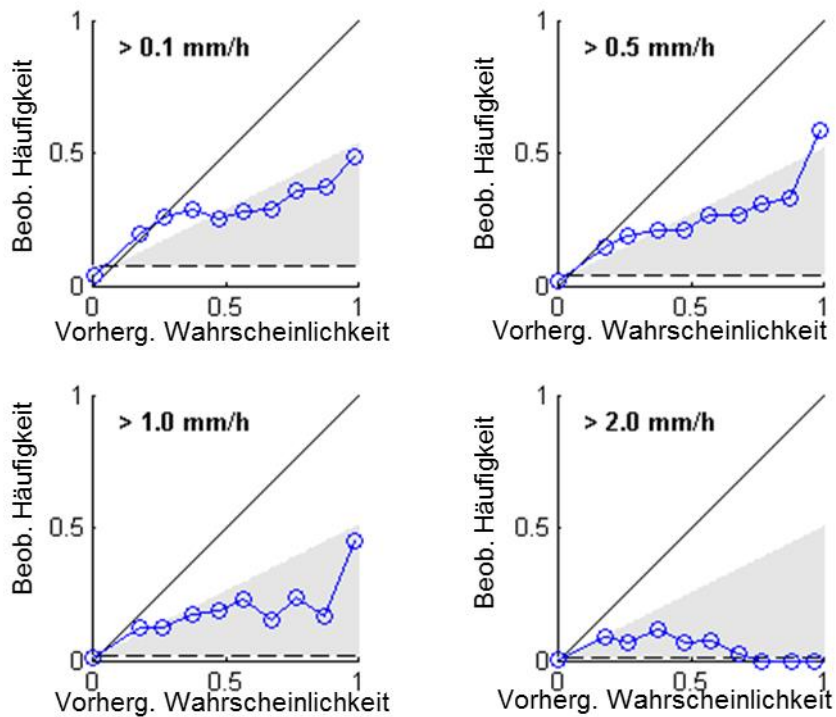


Abbildung 14: Reliability Diagramm COSMO-DE-EPS für den Vorhersagezeitraum 4-6 h

4.3 Vorhersagequalität für die Pegelwarnung

4.3.1 Vorhersagequalität der Nowcasts für die Pegelwarnung

Die Vorhersagequalität der Nowcasts wird speziell für die Anwendung der Pegelwarnung im Hamburger Warnsystem WaBiHa untersucht. Dafür werden die Abfragen im Warnsystem zugrunde gelegt:

- 1) Die Grenzwerte der Niederschlagsintensität für die Pegelwarnung (relevante Intensitäten für die Pegelwarnung, siehe Abschnitt 4.1.2): Akkumulierter Niederschlag > 7 mm in einer Stunde über einer Fläche von 5 km x 5 km.
- 2) Die Vorhersagen werden über dem Stadtgebiet Hamburg zusammengefasst. Für das gesamte Stadtgebiet wird nur ein Warnstatus aus der Niederschlagsvorhersage abgeleitet – wenn an mindestens einem Ort der Grenzwert überschritten ist.
- 3) Wenn eine Warnung an sich korrekt ist, aber der Zeitpunkt der Überschreitung nicht exakt vorhergesagt wurde, kann sie dennoch als *Treffer* gewertet werden: ein Zeitunterschied von bis zu einer Stunde zwischen vorhergesagter und gemessener Überschreitung wird toleriert. (Davon ausgenommen ist die erste Stunde der Vorhersage).

Sobald an einem Ort innerhalb von Hamburg eine Grenzwertüberschreitung vorhergesagt wird, wird der Status auf Warnung gesetzt. Wenn es tatsächlich zu einer Grenzwertüberschreitung im Hamburger Stadtgebiet kommt – also mindestens an einem Ort eine kritische Menge Niederschlag gemessen wurde – wird die Vorhersage als *Treffer* gewertet, sonst als *Falscher Alarm*. Analog dazu werden die Einteilungen *Verpasst* oder *Richtigerweise kein Alarm* vorgenommen.

Im Untersuchungszeitraum 01.06.2016 - 31.08.2016 wurde der Grenzwert von 7 mm 39 Mal überschritten, das entspricht 1.8 % der Messungen in diesem Zeitraum. Zusätzlich wurde der Grenzwert: 3 mm Niederschlag, akkumuliert über eine Stunde untersucht. Dieser Wert wurde in 108 Fällen überschritten, was einem Anteil von 4.9 % der Messungen entspricht – durch das häufigere Vorkommen ist dieses Kriterium robuster; die Ergebnisse werden daher ergänzend ausgewertet und mit den Ergebnissen auf Basis des höheren Grenzwertes verglichen.

In Abb. 15 werden die Ergebnisse der Vorhersagequalität aus diesem Zeitraum dargestellt, wobei die Trefferrate (POD) gegen den Anteil der Falschen Alarme (FAR) aufgetragen ist. Eine perfekte Vorhersage wäre in diesem Diagramm in der linken oberen Ecke (Trefferrate = 1, Anteil Falsche Alarme = 0), eine zufällige Vorhersage in der unteren rechten Ecke zu finden. Die Ensemblequantile bzw. Überschreitungswahrscheinlichkeiten aus der Ensemblevorhersage sind mit Kreisen markiert, die Quantile 0.1, 0.5 und 0.9 zusätzlich gekennzeichnet. Die Vorhersagequalität ist für die 1. Stunde des Vorhersagezeitraums (schwarz), die 2. Stunde (rot) und die 3. Stunde (blau) abgebildet. In der linken Abbildung ist die Auswertung für den Grenzwert 3 mm, in der rechten Abbildung für den Grenzwert 7 mm dargestellt. Die beste Vorhersagequalität hat die Vorhersage für die 1. Stunde. Bei der Auswertung mit Grenzwert 3 mm beträgt die Trefferrate über 0.8 und der Anteil der Falschen Alarme unter 0.4 – für niedrige Überschreitungswahrscheinlichkeiten aus dem Ensemble, bei höheren Überschreitungswahrscheinlichkeiten sinkt die Trefferrate und der Anteil der Falschen Alarme reduziert sich auf unter 0.2. In der 2. und 3. Stunde nimmt die Vorhersagequalität deutlich ab. Bei der Auswertung mit Grenzwert 7 mm ist die Vorhersagequalität insgesamt schlechter, für niedrige Überschreitungswahrscheinlichkeiten ergeben sich aber dennoch hohe Trefferraten und die Abnahme der Vorhersagequalität von der 1. bis zur 3. Stunde ist vergleichbar mit der linken Abbildung. Die Abnahme der Vorhersagequalität mit dem höheren Grenzwert entspricht den Erwartungen und hängt mit den Eigenschaften von konvektiven Niederschlägen zusammen: je höher die Grenzwerte sind, desto sel-

tener und meistens auch kürzer und kleinräumiger sind die Ereignisse, die in die Auswertung eingehen – diese sind weniger gut vorherzusagen als größere Ereignisse mit länger anhaltendem Niederschlag.

Ungefähr eine Stunde vor Eintritt eines konvektiven Starkregenens über 7 mm, der mit hoher Wahrscheinlichkeit zu einer Überschreitung der Warngrenze an einem Hamburger Pegel führt, beträgt die Wahrscheinlichkeit, dass er von mindestens einer der Ensemblevorhersagen vorhergesagt wird (Quantil: 0.1) im Untersuchungszeitraum über 80 %. Der Anteil der Falschen Alarme liegt in diesem Fall bei ca. 55 %. Wenn vor dem gleichen Ereignis 2 Stunden vor Ereigniseintritt gewarnt wird, ist die Trefferrate immer noch über 80 %, allerdings steigt der Anteil der Falschen Alarme deutlich an (auf ca. 70 %). Drei Stunden vor dem Ereignis kann man, wenn man einen Anteil von Falschen Alarmen um 80 % in Kauf nimmt, immerhin ca. 70 % der Ereignisse vorhersagen. Aus diesen Ergebnissen wird deutlich, wie stark die Vorhersagequalität der Nowcasts von dem Vorhersagezeitraum abhängt und wie wichtig es ist, die jeweils aktuellste Vorhersage für eine Warnentscheidung zugrunde zu legen.

Zusätzlich zur Auswertung der Vorhersagen auf Basis der Niederschlagsmessungen wurde geprüft, wie gut die in 4.1.2 aufgeführten Überschreitungen der Warngrenzen an den Hamburger Pegeln vorhergesagt wurden (Vergleich Vorhersage Niederschlag – Messung Pegel). Dazu wurden die 19 Pegelüberschreitungen im Zeitraum 01.06.-31.08.2016 betrachtet, die durch kurzen Starkregen verursacht wurden und mit den Vorhersagen einer einzelnen Kurzfristvorhersage verglichen. Mit der 1h-Vorhersage wurde vor 15 der 19 Pegelüberschreitungen richtig gewarnt. Bei einer Vorwarnzeit von einer halben Stunde erhöhte sich die Anzahl der korrekten Warnungen auf 17. Dieses Ergebnis ist konsistent mit den Ergebnissen des Vorhersageensembles und die Vorhersage des Ereignisses ‚Pegelüberschreitung‘ schneidet im Fall von konvektiven Starkregen ähnlich ab wie die Vorhersage des Ereignisses ‚gemessener Niederschlag > 7 mm‘.

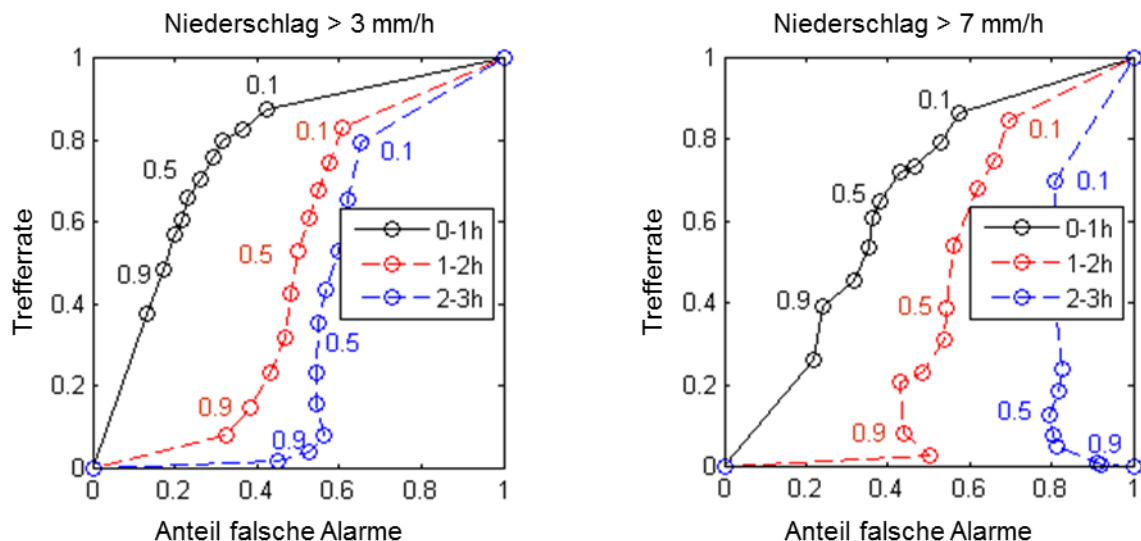


Abb. 15: Vorhersagequalität der Nowcasts als Trefferrate gegen den Anteil der falschen Alarme, ausgewertet auf Basis der Grenzwerte 3 mm bzw. 7 mm Niederschlag, aggregiert über eine Stunde. Die Vorhersagezeiträume betragen 0-1 h (schwarz), 1-2 h (rot) und 2-3 h (blau), Kreise entsprechen Quantilen der Ensemblevorhersagen.

4.3.2 Vorhersagequalität der COSMO-DE-EPS für die Pegelwarnung

Analog zu 4.3.1 wird die Vorhersagequalität der COSMO-DE-EPS Vorhersagen für die Pegelwarnung ausgewertet. Es werden die gleichen Kriterien wie bei der Auswertung der Nowcasts zugrunde gelegt, sodass ein direkter Vergleich zwischen den Vorhersagen möglich ist.

Ein Unterschied zu den Nowcasts ist, dass die COSMO-DE-EPS Vorhersagen wegen der langen Rechenzeit erst 3 Stunden, nachdem die Vorhersagen gestartet werden, auf dem Datenserver zur Verfügung stehen. Da sie in einem Intervall von 3 Stunden gerechnet werden, ist die aktuellste COSMO-DE-EPS Vorhersage im Normalfall zwischen 3 und 6 Stunden alt, so dass die ersten Stunden der Vorhersage bereits in der Vergangenheit liegen. In der Anwendung entspricht daher die Nowcast in der 1. Stunde einer COSMO-DE-EPS Vorhersage in der 4. bis 6. Stunde des Vorhersagezeitraums.

In Abb. 16 ist die Vorhersagequalität der COSMO-DE-EPS Vorhersagen für den Vorhersagezeitraum 4-6 h und 7-9 h dargestellt. Wie bei den Nowcasts nimmt auch hier die Vorhersagequalität mit der Dauer der Vorhersage ab, aber die Abnahme ist schwächer im Vergleich zu den Nowcasts. Anders als bei den Nowcasts befinden sich die Ergebnisse für die verschiedenen Quantile des Vorhersageensembles ungefähr auf einer senkrechten Linie, haben also einen ähnlichen Anteil falscher Alarme. Das bedeutet, wenn ein höheres Quantil des Ensembles betrachtet wird (In anderen Worten: mehrere Ensemblevorhersagen sagen einen Starkregen voraus), ist die Wahrscheinlichkeit, dass sich eine Warnung um einen Treffer und nicht um einen Fehlalarm handelt nicht erkennbar höher, als wenn nur ein Lauf Starkregen voraussagt. Andererseits nimmt die Trefferrate für höhere Quantile deutlich ab. Beim Grenzwert 7 mm/h erzielt das niedrigste Quantil (= 1/20 Läufen liegt über dem Grenzwert) für den Vorhersagezeitraum 4-6 h eine Trefferrate über 0.8 bei einem Anteil von Falschen Alarmen um 0.7. Für den Median liegt die Trefferrate unter 0.3, der Anteil Falscher Alarme aber immer noch bei ca. 0.6. Die deterministischen Vorhersagen von COSMO-DE schneiden etwas besser ab als der Median des Ensembles, beim Grenzwert 7 mm/h liegt die Trefferrate aber für beide Vorhersagezeiträume unter 0.3. Beim Grenzwert 3 mm/h fällt die Vorhersagequalität insgesamt besser aus als bei 7 mm/h, ansonsten sind die Ergebnisse ähnlich wie bei dem höheren Grenzwert.

Zusammengefasst lässt sich mit den COSMO-DE-EPS eine wesentlich bessere Trefferrate erzielen als mit COSMO-DE, d.h. es kann vor mehr Ereignissen richtig gewarnt werden, während der Anteil der falschen Alarme nicht wesentlich erhöht wird. Die besten Ergebnisse erhält man, wenn man ein niedriges Ensemblequantil (0.05 oder 0.1) betrachtet.

Ein Vergleich zwischen den Ergebnissen von COSMO-DE-EPS (Abb. 16) und Nowcasts (Abb. 15) zeigt, dass die Nowcasts in der ersten Stunde deutlich besser abschneiden als die COSMO-DE-EPS Vorhersagen. In der 2. Stunde ist der Unterschied in der Vorhersagequalität nicht mehr groß und in der 3. Stunde fallen die COSMO-DE-EPS Vorhersagen im Durchschnitt besser aus als die Nowcasts, unabhängig davon, welcher Grenzwert betrachtet wird.

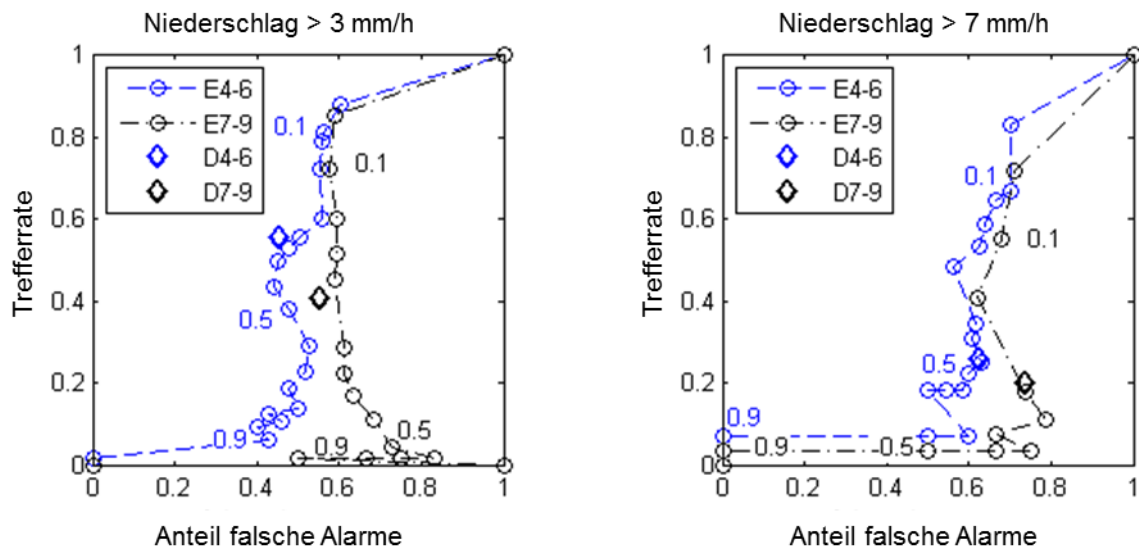


Abb. 16: Vorhersagequalität der COSMO-DE-EPS (Kreise entsprechen Quantilen der Ensemblevorhersagen) und COSMO-DE Vorhersagen (Rauten), ausgewertet auf Basis der Grenzwerte 3 mm bzw. 7 mm Niederschlag, aggregiert über eine Stunde. Die Vorhersagezeiträume betragen 4-6 h (blau) und 7-9 h (schwarz).

4.3.3 Auswertungen zum Blending

Blending („Vermischen“) von Radar-Nowcasts und numerischer Wettervorhersage wird angewandt, um einen kontinuierlichen Übergang zwischen den Vorhersagen zu erreichen und für einen Übergangszeitraum, in dem sowohl Nowcasts als auch numerische Wettervorhersagen vorliegen, die Vorhersagen zu verbessern. Die Methoden dafür reichen von einer einfachen Überlagerung der Vorhersagedaten bis hin zu aufwändigen und komplexen Methoden, die den Ansatz verfolgen, Fehler in den numerischen Wettervorhersagen zu verringern. Dafür wird der Umstand verwendet, dass die Radarmessungen und Nowcasts aktuellere Informationen enthalten als die numerischen Wettervorhersagen, da die Eingangsdaten für deren Berechnung aufgrund des Rechenaufwands und Datenvolumens bereits veraltet sind, wenn die Vorhersagen verfügbar werden. Das kann genutzt werden, um die Vorhersagen mit Messungen oder Nowcasts abzugleichen und die Unterschiede gegebenenfalls für eine Korrektur der numerischen Wettervorhersagen zu nutzen (siehe z.B. Atencia et al. 2010).

Es wurden Fallbeispiele der COSMO-DE-EPS Vorhersagen betrachtet (siehe Beispiel in 4.2.2) um einen Überblick über Abweichungen zwischen den COSMO-DE-EPS Vorhersagen und Messdaten zu gewinnen und zu prüfen, ob diese mithilfe der Nowcasts verringert werden können (z.B. durch räumliche Verschiebungen auf Basis der Schwerpunkte der Niederschlagsfelder, Korrektur der Intensitätsverteilung, Korrektur der Zuggeschwindigkeit,...). Einige Fallbeispiele zeigen ein Potential, die COSMO-DE-EPS mithilfe der aktuelleren Daten zu verbessern. Allerdings fallen die Probleme bei verschiedenen Niederschlagsereignissen und Vorhersageläufen sehr unterschiedlich aus - eine Korrektur, die in einigen Beispielen zu einer Verbesserung führen würde, würde in anderen Fällen die Vorhersage verschlechtern. Auf eine umfassende Analyse der einzelnen Fehlerquellen und Korrekturmöglichkeiten wurde an dieser Stelle verzichtet.

In einem Langzeittest wurde untersucht, ob die Vorhersagen verbessert werden können, indem mithilfe der aktuellen Informationen die COSMO-DE-EPS Läufe ausgewählt werden, die näher an den Radarmessungen bzw. Nowcasts liegen, während die schlechter passenden Läufe ausgeschlossen werden. Der Nachteil bei diesem Vorgehen liegt darin, dass die Gesamtanzahl der Vorhersageläufe in dem Ensemble reduziert wird. Die Ergebnisse der Langzeitauswertung zeigten, dass die genauere

Auswahl von im Durchschnitt besser passenden Läufen den Nachteil, die Anzahl der Läufe reduziert zu haben, nicht wett macht. Insbesondere zeigten einzelne Vorhersageläufe, die im Überschneidungszeitraum stark von den Messdaten abwichen, im späteren Vorhersageverlauf eine gute Übereinstimmung und umgekehrt.

Als Fazit der Auswertungen wird die beste Vorhersagequalität erreicht, wenn im Übergangszeitraum alle 20 Ensembleläufe berücksichtigt werden, unabhängig von der Übereinstimmung mit Radarmessungen/ Nowcasts. Um Starkregenereignisse (Niederschlag > 3 mm bzw. > 7 mm in einer Stunde) mit einer hohen Wahrscheinlichkeit voraussagen zu können, muss ein niedriges Quantil des Vorhersageensembles betrachtet werden. D.h. bereits wenn 1-2 Ensembleläufe Niederschlag oberhalb des Grenzwertes vorhersagen und alle anderen darunter liegen, kann eine Warnung sinnvoll sein.

4.3.4 Schlussfolgerungen

Aus den Ergebnissen in 4.3.1 und 4.3.2 lassen sich Schlussfolgerungen für die Konstruktion der kombinierten Niederschlagsvorhersagen ableiten, so dass sich eine möglichst gute Vorhersagequalität im Warnsystem ergibt:

- 1) In den ersten beiden Stunden stellen die Nowcasts eine deutliche Verbesserung gegenüber COSMO-DE/ COSMO-DE-EPS dar. Diese werden daher für die ersten beiden Stunden der kombinierten Niederschlagsvorhersagen verwendet.
- 2) Die besten Ergebnisse für den etwas längeren Vorhersagezeitraum erhält man, wenn man das niedrigste Ensemblequantil von COSMO-DE-EPS betrachtet. Damit ist die Trefferrate deutlich höher als bei der Warnung mit COSMO-DE. Für das Warnsystem bedeutet das, dass sich die besten Ergebnisse erzielen lassen, wenn alle Ensemblevorhersagen gerechnet und analysiert werden und der höchste Lauf die Entscheidung über die Warnung liefert. Wenn mehrere Kriterien betrachtet werden (im aktuellen Warnsystem WaBiHa gibt es mehrere Grenzwerte über verschiedene Vorhersagezeiträume), können unterschiedliche Ensembleläufe zu dem jeweils höchsten Wert führen.

Dieses Ergebnis gilt speziell für die Anwendung der Starkregenwarnung bei kurzen konvektiven Ereignissen, auf der der Fokus der Auswertung lag. Bei der Warnung vor großräumigen und länger anhaltenden Niederschlägen kann es sinnvoll sein, zusätzlich die höheren Ensemblequantile zu betrachten.

- 3) Blendingzeitraum: Im Übergangszeitraum wurde die Ergebnisse von COSMO-DE-EPS und Nowcasts, sowie die Kombination untersucht. In diesem Zeitraum war der Effekt von Ensemblevorhersagen gegenüber einer einzelnen Vorhersage deutlich erkennbar. Weiterführende Methoden führten bei einzelnen Ereignissen zu einer Verbesserung, bei anderen aber auch zu einer Verschlechterung und im Langzeittest ergab sich kein eindeutiger Effekt. Daher wird eine relativ einfache Umsetzung zur Kombination der beiden Vorhersageensembles gewählt, die im folgenden Kapitel beschrieben wird.

5 Kombinierte Niederschlagsvorhersagen – Umsetzung

5.1.1 Umsetzung des Blendings

Die COSMO-DE-EPS Vorhersagen werden zunächst auf das gleiche Gitter wie die Nowcasts interpoliert (1 km x 1 km), um diese mit den Nowcasts kombinieren zu können. Im Blending Zeitraum (3. und

4. Stunde der Vorhersage) werden die Nowcasts zu Stundensummen aggregiert, so dass sie auf die gleiche Weise vorliegen wie die COSMO-DE-EPS. Im nächsten Schritt erfolgt eine Überlagerung der Niederschlagsfelder von COSMO-DE-EPS und Nowcasts durch eine zeitabhängige gewichtete Mittelung, wie in Golding (1998) beschrieben. Die Gewichtungsfaktoren betragen

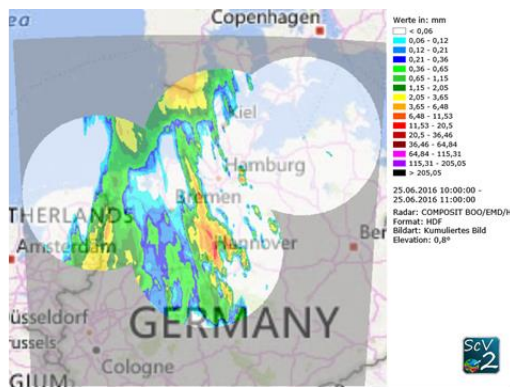
Für die 3. Stunde: COSMO-DE-EPS $G_c = 2/3$; Nowcasts $G_n = 1/3$

Für die 4. Stunde: COSMO-DE-EPS $G_c = 3/4$; Nowcasts $G_n = 1/4$

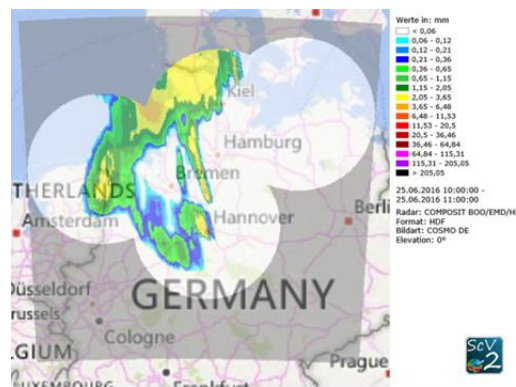
Wenn an einem Ort undefinierte Werte in den Nowcasts enthalten sind, ändern sich die Gewichtungsfaktoren auf $G_c = 1$; $G_n = 0$. Undefinierte Werte können auftreten, wenn ein Radar ausfällt oder wenn Niederschlagswerte eingehen, die mit Beginn der Nowcast-Berechnung noch außerhalb der Sichtweite des Radars lagen.

Da die COSMO-DE-EPS Vorhersagen für volle Stunden vorliegen, die Radar-Nowcasts hingegen alle 5 min aktualisiert werden, muss im Zeitraum des Blendings dieser zeitliche Unterschied überbrückt werden. Dafür werden die Niederschlagssummen der Nowcasts zunächst über volle Stunden gerechnet und mit den COSMO-DE-EPS überlagert. Die 3. Stunde wird dann entsprechend der tatsächlichen Länge der Nowcasts (2 h ab Vorhersagebeginn) verkürzt und die Niederschlagsmenge, die für diesen Zeitraum berechnet wurde, wird anteilmäßig reduziert. Die Vorhersagen im Blending Zeitraum werden alle 5 min mit dem Vorliegen der neuen Nowcasts aktualisiert.

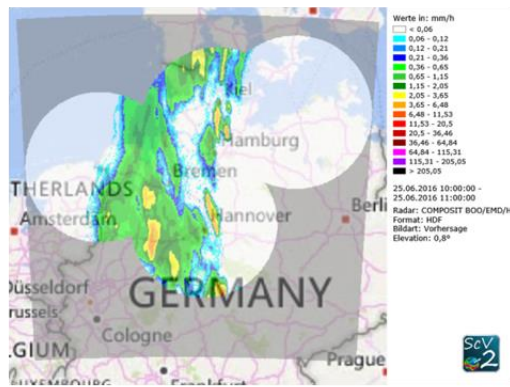
Ein Beispiel zur Überlagerung von Nowcast und COSMO-DE-EPS findet sich in Abb. 17. Dargestellt ist die Vorhersagesumme von einer Nowcast (Stundensumme der 2. Stunde der Vorhersage), einem COSMO-DE-EPS Lauf und die geblendete Vorhersagesumme, sowie die Radarmessung. In diesem Beispiel passt die geblendete Vorhersage besser zur Messung als die Einzelvorhersagen. Eine Langzeitauswertung folgt in Kapitel 5.3.



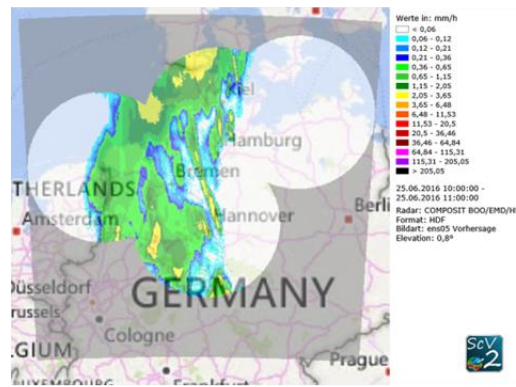
Niederschlagssumme in mm – Messung vom 25.06.2016 10:00-11:00 Uhr (korrigierte und angeeichte Radarmessungen)



Niederschlagssumme in mm – Vorhersage von 09:00 Uhr für den Zeitraum 10:00-11:00 Uhr Radarnowcast



Niederschlagssumme in mm – Vorhersage für 10:00-11:00 Uhr COSMO-DE-EPS (6:00 Uhr Lauf, verfügbar ab ca. 9:00 Uhr)



Niederschlagssumme in mm – Vorhersage von 09:00 Uhr für den Zeitraum 10:00-11:00 Uhr Blending Radarnowcast & COSMO-DE-EPS

Abb. 17: Beispiel zum Blending 25.06.2016: Vorhersagen (Nowcast, COSMO-DE-EPS & Blending) von 9:00 Uhr für den Zeitraum 10:00-11:00 Uhr im Vergleich zur Radarmessung (o. l.).

5.2 Umsetzung der kombinierten Ensemblevorhersagen

Aus den Auswertungen in Kapitel 4 ergeben sich die Vorhersagezeiträume, für die die verschiedenen Vorhersagen verwendet werden:

1. 1-2 h: Radar-Nowcasts
2. 3-4 h: Blending
3. > 4 h: COSMO-DE-EPS

Im 1. Zeitraum werden die Nowcasts mit einem Zeitschritt von 5 min gerechnet und alle 5 min aktualisiert. Dieser deckt die ersten beiden Stunden ab.

Im 2. Zeitraum werden die Nowcasts mit den COSMO-DE-EPS überlagert (Blending), siehe Kapitel 5.1.1.. Wenn die Vorhersage nicht zu einer vollen Stunde berechnet wird, wird die erste Stunde im Blending-Zeitraum verkürzt, so dass ab der 4. Stunde Vorhersagen über volle Stunden eingehen, so wie sie von COSMO-DE-EPS vorliegen.

Im 3. Zeitraum > 4 h bis zum Ende des Vorhersagezeitraums werden die COSMO-DE-EPS Vorhersagen mit Zeitschritt 1 h angehängt.

Es wird ein Vorhersageensemble erstellt, das aus 10 Nowcast-Ensembles und 20 Läufen von COSMO-DE-EPS zusammengesetzt ist. Die Ensembleläufe der Nowcasts werden doppelt verwendet, so dass jeweils 20 Vorhersagen produziert werden.

Von den kombinierten Vorhersagen werden folgende Produkte erstellt:

- 1) Kontinuierliche Zeitreihen über einen Vorhersagezeitraum von 0-20 h mit Zeitschritt 5 min in den ersten beiden Stunden, anschließend 1 h (20 Ensemblevorhersagen)
- 2) Fortlaufende Vorhersagebilder der Niederschlagsintensität (mm/h) über einen Vorhersagezeitraum von 0-20 h
- 3) Vorhergesagte Niederschlagsmengen, aggregiert über verschiedene Vorhersagezeiträume (0-2 h, 0-6 h, 0-12 h, 0-24 h), zusammengesetzt aus Radar-Nowcasts, Blending und COSMO-DE-EPS – Ensembledarstellung

Aus den Zeitreihen können alle weiteren Kriterien abgeleitet werden (Grenzwertüberschreitungen für das Warnsystem). Diese werden zudem an die Projektpartner für hydrologische Modellierungen weitergegeben. Die Vorhersagebilder und Vorhersagesummen dienen zur Visualisierung und geben einen Überblick über die aktuelle Vorhersagesituation. Beispiele der Vorhersagesummen von jeweils 3 Ensembleläufen im Vergleich zur Radarmessung sind in Abbildung 18 - Abbildung 20 dargestellt.

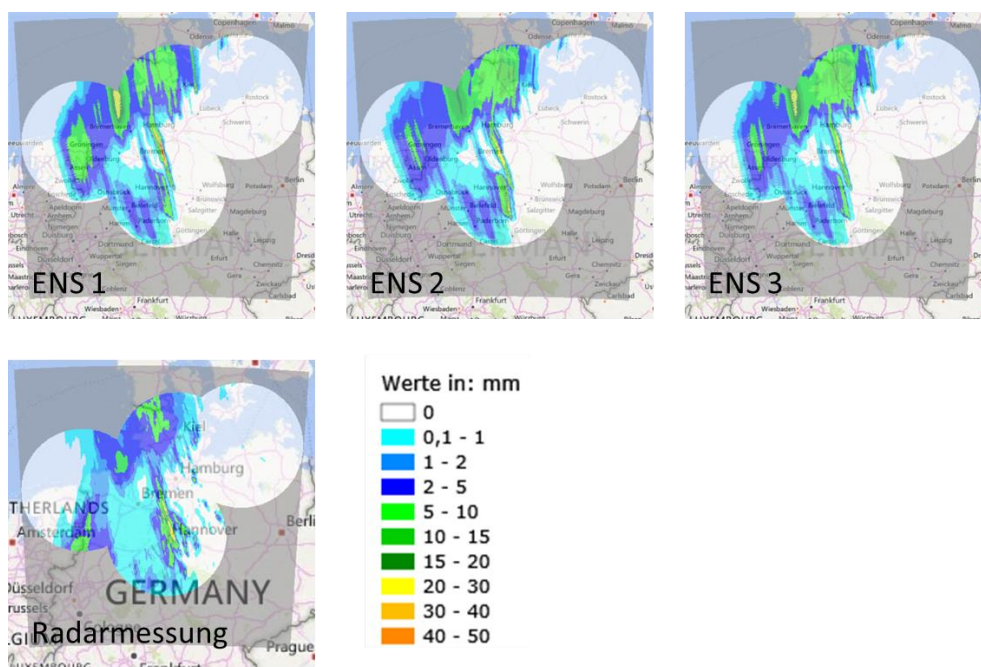


Abbildung 18: Beispiel Vorhersagesummen: Ensembledarstellung Niederschlag von 3 Ensembleläufen 25.06.2017 9:00 Uhr für 9:00-11:00 Uhr (0-2 h) und die Radarmessung von 9:00-11:00 Uhr (unten).

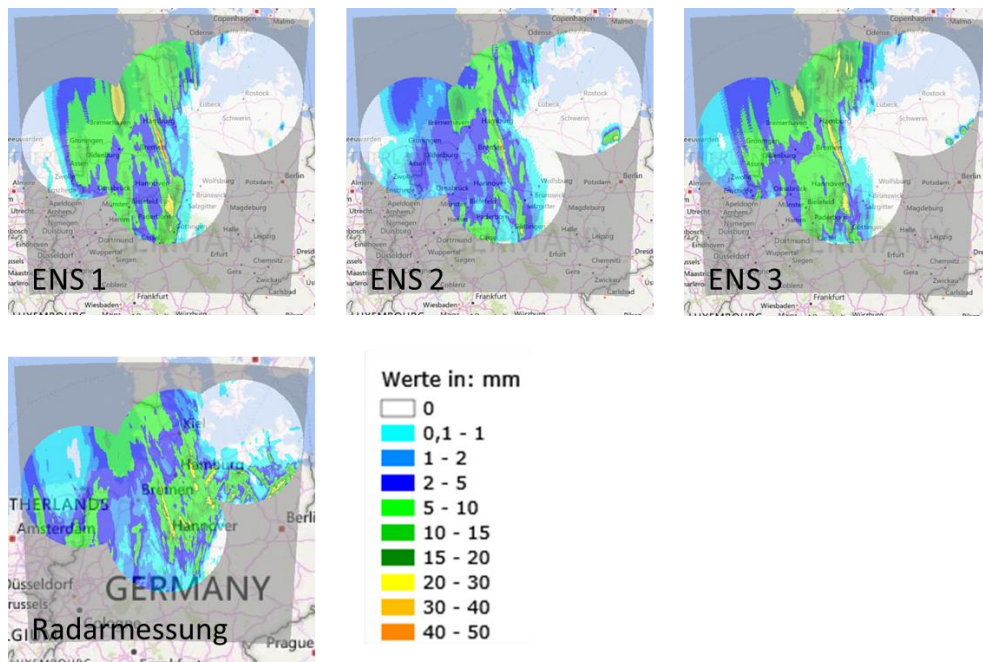


Abbildung 19: Beispiel Vorhersagesummen: Ensemblevorhersagen Niederschlag von 3 Ensembleläufen 25.06.2017 9:00 Uhr für 9:00-15:00 Uhr (0-6 h) und die Radarmessung von 9:00-15:00 Uhr (unten).

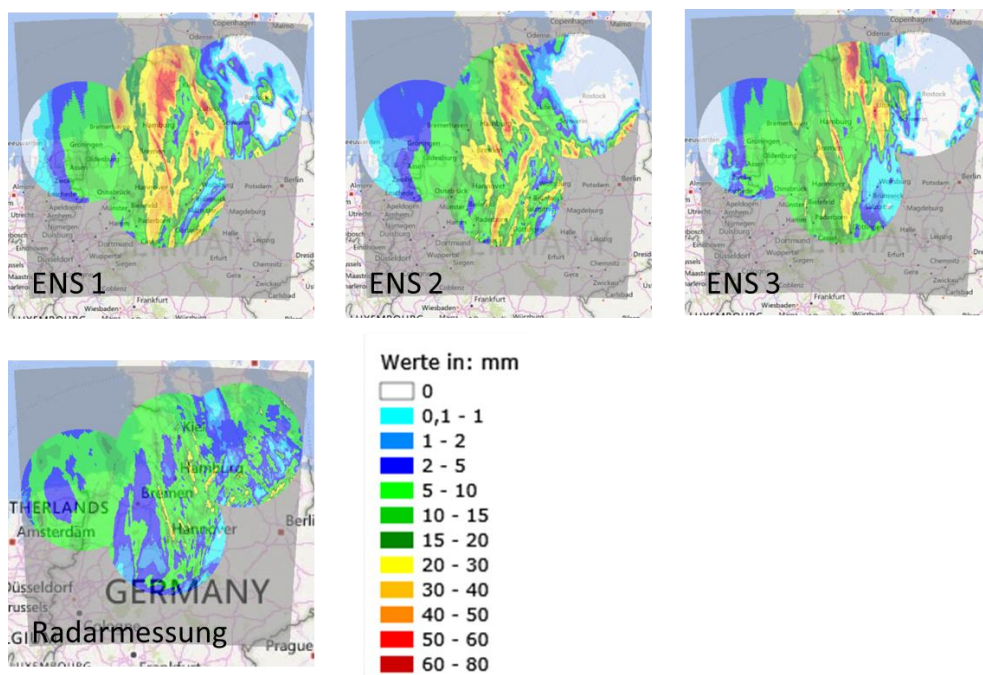


Abbildung 20: Beispiel Vorhersagesummen: Ensemblevorhersagen Niederschlag von 3 Ensembleläufen 25.06.2017 9:00 Uhr für 9:00-21:00 Uhr (0-12 h) und die Radarmessung von 9:00-21:00 Uhr (unten).

5.3 Validierung der kombinierten Niederschlagsvorhersagen im Vergleich zu COSMO-DE

Eine unabhängige Validierung der kombinierten Niederschlagsvorhersagen wurde für den Zeitraum 22.03.2017 - 22.06.2017 durchgeführt. Dafür wurden sowohl COSMO-DE-EPS Vorhersagen als auch das Produkt der kombinierten Niederschlagsvorhersagen, bestehend aus Ensemblesowcasts, Blending und COSMO-DE-EPS mit den deterministischen COSMO-DE Vorhersagen verglichen. Die Auswertung erfolgt analog zu den Auswertungen in 4.3.

In die Auswertung gehen etwas veränderte COSMO-DE-EPS Vorhersagen ein, da die Berechnung der Ensemblevorhersagen im März 2017 durch den Deutschen Wetterdienst umgestellt wurde. Unter der Annahme, dass sich dadurch keine wesentlichen, für das Verfahren relevanten Unterschiede ergeben, wurden die Schlussfolgerungen aus 4.3.4 ohne Änderungen umgesetzt. Die Ergebnisse wurden separat für die kombinierten Vorhersagen und für COSMO-DE-EPS ausgewertet und dargestellt. Um speziell die Veränderungen in COSMO-DE-EPS in Bezug auf die Starkregenvorhersage zu bewerten, ist der untersuchte Zeitraum zu kurz.

Im Zeitraum 22.03.2017 - 22.06.2017 wurden im Hamburger Stadtgebiet in 32 Fällen starke Niederschläge gemessen (akkumulierter Niederschlag über eine Stunde in einem Gebiet von 5 km x 5 km > 3 mm). Im gleichen Zeitraum fanden 11 Starkregenereignisse statt, die das in 4.1.2 beschriebene Warnkriterium erfüllen (akkumulierter Niederschlag über eine Stunde in einem Gebiet von 5 km x 5 km > 7 mm). Da die Anzahl der Starkregenereignisse zu niedrig ist, um damit die Qualitätsparameter Trefferrate und Anteil falsche Alarmer in Abhängigkeit von dem Ensemblequantil zu bestimmen, werden diese Ergebnisse nur für das Kriterium Niederschlag > 3 mm gezeigt. Auf die Ereignisse mit Niederschlag > 7 mm wird im darauffolgenden Abschnitt näher eingegangen.

In

Ab-

bildung 21 ist die Vorhersagequalität als Trefferrate (*hit rate*) gegen den Anteil der falschen Alarme (*false alarm ratio*) der kombinierten Vorhersagen dargestellt, wobei die Vorhersagezeiträume 1. bis 2. Stunde (Nowcasts), 3. bis 4. Stunde (Blending) und ab der 5. Stunde (COSMO-DE-EPS) separat dargestellt sind. Dazu ist die Vorhersagequalität der COSMO-DE Vorhersagen für die Vorhersagezeiträume 1-3 h, 4-6 h, 7-9 h und 10-12 h gezeigt. Wie erwartet nimmt die Vorhersagequalität der kombinierten Vorhersagen in den ersten Stunden ab, wobei die Trefferrate sinkt und der Anteil der falschen Alarme zunimmt. Die überlagerten Vorhersagen (3. und 4. Stunde, Blending) fallen ähnlich aus wie die COSMO-DE-EPS Vorhersagen für den gleichen Zeitraum. Im Vergleich zu der Auswertung aus 2016 geht eine geringere Anzahl von Starkregenereignissen ein und einzelne Ereignisse und Vorhersagen wirken sich stärker aus. So ist zu erklären, dass z. B. die COSMO-DE Vorhersage für den Vorhersagezeitraum 10-12 h besser ausfällt als für den Zeitraum 7-9 h. Die Trefferrate der COSMO-DE-EPS Vorhersagen fällt auch für die längeren Vorhersagezeiträume 7-9 h und 10-12 h gut aus. Allerdings steigt der Anteil der falschen Alarme im Vergleich zu den kürzeren Vorhersagen. Wenn man das niedrigste Quantil der kombinierten Ensemblevorhersagen betrachtet, ergibt sich eine deutlich bessere Vorhersagequalität als bei COSMO-DE: Die Trefferrate des niedrigsten Quantils liegt für alle Vorhersagezeiträume über 0.7, während die Trefferrate bei COSMO-DE zwischen 0.1 und 0.4 beträgt. Beim Anteil der falschen Alarme gibt es in einzelnen Vorhersagezeiträumen Abweichungen. Im Mittel fällt hingegen der Anteil der falschen Alarme bei den kombinierten Vorhersagen mit ca. 50-70 % ähnlich aus wie bei COSMO-DE. Die Ergebnisse sind konsistent mit den in Kapitel 4 beschriebenen Ergebnissen aus 2016. Die verbesserte Vorhersagequalität der kombinierten Vorhersagen gegenüber COSMO-DE ist eindeutig zu erkennen.

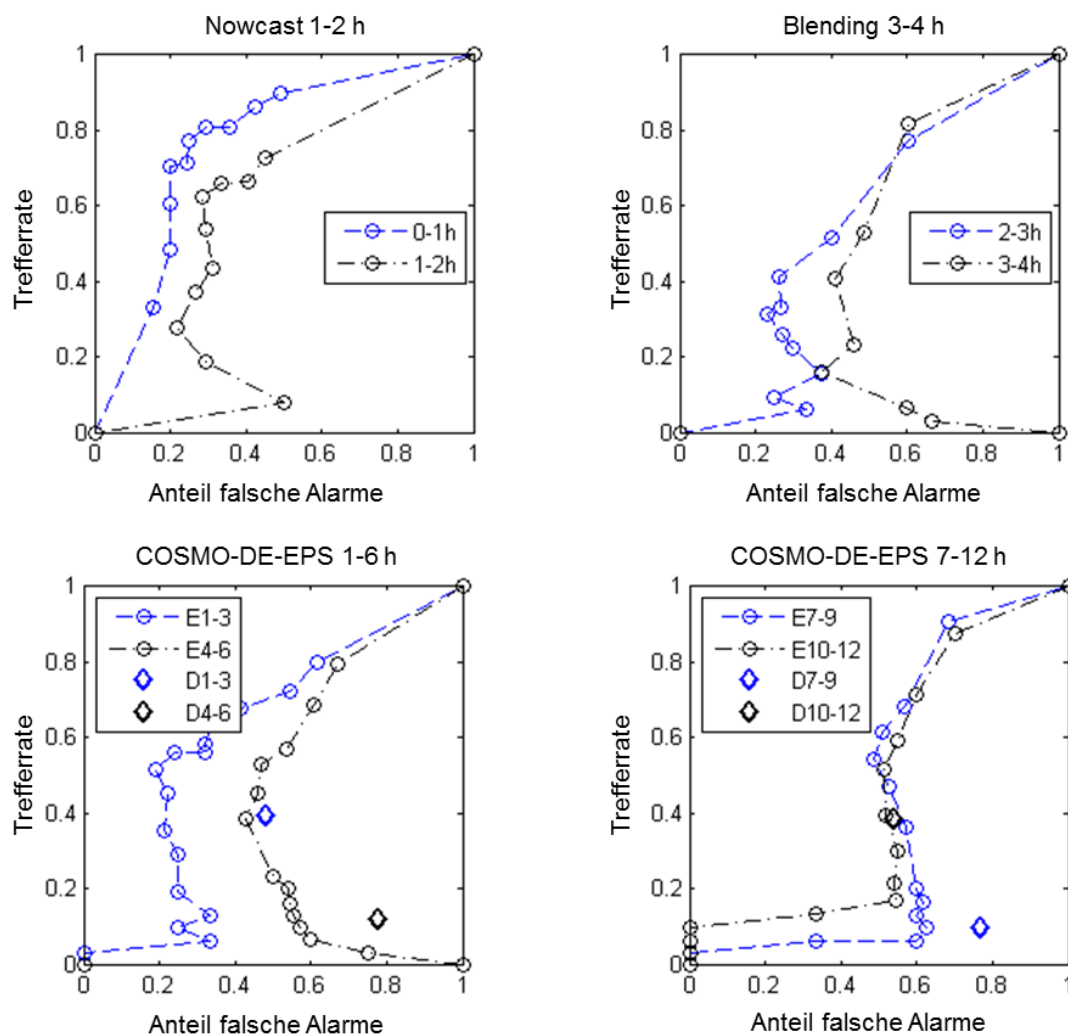


Abbildung 21 Vorhersagequalität (Trefferquote gegen den Anteil der falschen Alarme) von Niederschlagsvorhersagen von Nowcasts, Blending und COSMO-DE-EPS sowie COSMO-DE für verschiedene Vorhersagezeiträume. Ausgewertet wurde der Zeitraum 22.03.-22.06.2017. Die Kreise kennzeichnen Ensemblequantile. Das niedrigste Ensemblequantil weist jeweils die höchste *hit rate* auf. Oben links: die 1. und 2. Stunde des Vorhersagezeitraums von den Nowcasts. Oben rechts: die überlagerten Vorhersagen aus Nowcasts und COSMO-DE-EPS für die 3. und 4. Stunde. Unten: die COSMO-DE-EPS Vorhersagen (Kreise/ gestrichelt) und die COSMO-DE Vorhersagen (Rauten) für die Vorhersagezeiträume 1-3 h, 4-6 h, 7-9 h und 10-12 h.

5.4 Ergebnisse für die Pegelwarnungen 2017

Eine Auswertung der 11 Starkregenereignisse im Zeitraum 22.03.2017 - 22.06.2017, die das Warnkriterium für die Pegelwarnung erfüllten (akkumulierter Niederschlag über eine Stunde in einem Gebiet von 5 km x 5 km > 7 mm) wurde ergänzend zu den Auswertungen in 5.3 durchgeführt. Bei diesen Ereignissen kam es jeweils auch zu erhöhten Pegelständen oberhalb der mittleren Warnstufe an einem oder mehreren Pegeln im System WaBiHa. Im Untersuchungszeitraum traten 4 weitere Pegelüberschreitungen auf, die durch länger anhaltende Niederschläge verursacht wurden. Diese werden hier nicht näher betrachtet. In Abbildung 22 sind die Ergebnisse der Starkregenwarnung auf Basis der kombinierten Vorhersagen, bestehend aus Nowcast Ensembles, Blending und COSMO-DE-EPS, gezeigt. Eine Vorhersage wurde als Warnung - und somit als Treffer oder Falscher Alarm - gewertet, wenn von mindestens einem Ensemblelauf der kombinierten Vorhersagen Niederschlag oberhalb des Grenzwertes vorhergesagt wurde (d.h. das niedrigste Ensemble-Quantil wurde ausgewertet). Abbil-

Abbildung 22 zeigt, bei wie vielen der 11 Starkregenereignisse eine richtige Warnung ausgegeben wurde, in Abhängigkeit von der Vorhersagezeit. Die kombinierten Vorhersagen sind in blau (Ensemble-Nowcasts und Blending) und rot (COSMO-DE-EPS) dargestellt. Dazu sind die Ergebnisse von COSMO-DE (schwarz) eingezeichnet. Darunter ist die Anzahl der falschen Alarme im gleichen Zeitraum gezeigt, in Abhängigkeit von der Vorwarnzeit. Man sieht, dass die Anzahl der richtig vorhergesagten Ereignisse auch bei den längeren Vorwarnzeiten von 7-12 h durch COSMO-DE-EPS hoch ist. Allerdings steigt die Anzahl der falschen Alarme deutlich an. Je nach Anwendung kann es daher sinnvoll sein, die Vorwarnzeit zu begrenzen.

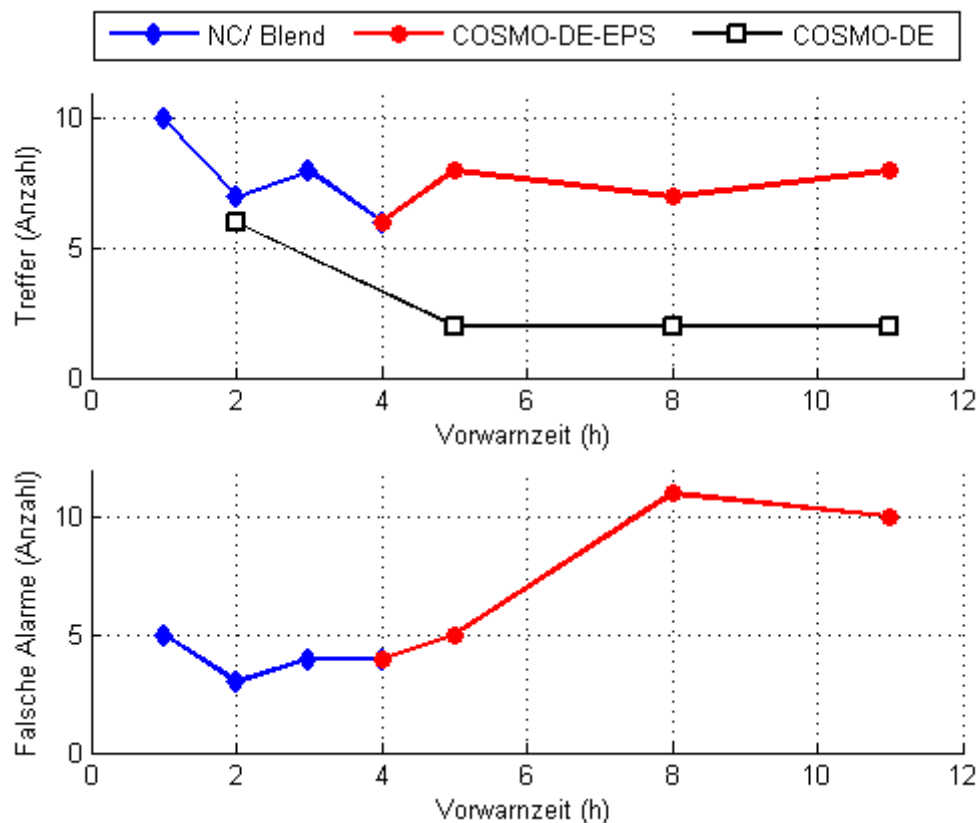


Abbildung 22 Ergebnisse der kombinierten Vorhersagen im Testzeitraum 22.03.-22.06.17. In diesem Zeitraum trat eine Gesamtanzahl von 11 Niederschlagsereignissen auf, die das Warnkriterium im Hamburger Stadtgebiet überschritten. Bei diesen Ereignissen kam es auch jeweils zu Pegelüberschreitungen mit Pegelständen oberhalb der mittleren Warnstufe im System WaBiHa. In der oberen Abbildung ist die Anzahl der richtigen Warnungen der kombinierten Vorhersagen im Vergleich zu COSMO-DE dargestellt. In der unteren Abbildung ist die Anzahl der Falschen Alarme der kombinierten Vorhersagen in Abhängigkeit vom Vorhersagehorizont dargestellt.

6 Zusammenfassung

Im Rahmen des Forschungsprojekts Stuck wurden im Arbeitspaket 1 kombinierte Niederschlagsvorhersagen erstellt, die auf Nowcast-Ensembles und COSMO-DE-EPS basieren. Die Vorhersagen werden im operationellen Betrieb alle 5 min aktualisiert und es werden Vorhersagezeitreihen und Vorhersagesummen produziert. Eine Reihe von Auswertungen zur Vorhersagequalität der kombinierten Vorhersagen wurde durchgeführt, wobei der Schwerpunkt auf der Analyse von kleinräumigen Starkregenereignissen mit hohen Niederschlagsintensitäten lag, bei einem Vorhersagezeitraum von 1-9 h. Im Vergleich mit der deterministischen numerischen Wettervorhersage COSMO-DE haben die kombinierten Vorhersagen eine deutlich verbesserte Vorhersagequalität. In den ersten beiden Stunden

werden die Vorhersagen durch die Verwendung der Nowcasts erheblich verbessert. Für längere Vorhersagen ergibt sich eine erkennbare Verbesserung durch die Verwendung der Ensembles (COSMO-DE-EPS) gegenüber der Einzelvorhersage (COSMO-DE).

Bei der Interpretation der Ensembles ist zu beachten, dass die vorhergesagten Überschreitungswahrscheinlichkeiten auf Basis der Ensembles nicht zuverlässig mit den tatsächlich beobachteten Wahrscheinlichkeiten übereinstimmen. Die besten Ergebnisse in der Vorhersage von kurzen für die Pegelwarnung relevanten Starkregenereignissen ergaben sich, wenn ein niedriges Ensemblequantil betrachtet wurde, das einer vorhergesagten Überschreitungswahrscheinlichkeit von 5 % -10 % entspricht. (D.h. in 1 bzw. 2 von 20 Läufen treten Niederschlagsintensitäten oberhalb des Grenzwerts auf.)

Die Ergebnisse lassen erwarten, dass die kombinierten Niederschlagsvorhersagen für eine Reihe von Anwendungen (u. a. die Warnungen im Warnsystem WaBiHa und hydrologische Simulationen in den Hamburger Einzugsgebieten) eine erkennbare Verbesserung gegenüber den bisher vorhandenen und genutzten Vorhersagen darstellen. Die Verwendung der Ensemblevorhersagen im operationellen Betrieb stellt einen größeren Aufwand dar im Vergleich zu einer einzelnen Vorhersage, da größere Datenmengen gespeichert und verarbeitet werden müssen. Da aber die Vorhersagequalität wesentlich verbessert wird, dürfte der zusätzliche Aufwand für viele Anwendungen – auch über die im Projekt Stuck untersuchten Anwendungen hinaus – gerechtfertigt sein.

7 Danksagung

Die Autoren und Projektpartner des Projekts „Stuck“ danken dem Bundesministerium für Bildung und Forschung (BMBF) für die Förderung des Verbundforschungsprojekts im Rahmen der Fördermaßnahme „Regionales Wasserressourcen-Management für den nachhaltigen Gewässerschutz in Deutschland (ReWaM). Projektpartner sind:

- Freie und Hansestadt Hamburg – Landesbetrieb Straßen, Brücken und Gewässer (Verbundprojektkoordination)
Prof. Dr. Gabriele Gönnert, Dieter Ackermann, Dr. Heiko Westphal, Fred Hesser, Sarah Gennis und Matthias Strauer
- Technische Universität Hamburg, Institut für Wasserbau
Prof. Dr.-Ing. Peter Fröhle, Sandra Hellmers, Edgar Nehlsen
- hydro & meteo GmbH & Co. KG, Lübeck
Dr. Thomas Einfalt, Alexander Strehz, Alrun Jasper-Tönnies
- Universität Hamburg
 - o Centrum für Erdsystemforschung und Nachhaltigkeit, Institut für Bodenkunde
Prof. Dr. Annette Eschenbach, Dr. Alexander Gröngroft, Kira Kalinski
 - o Biozentrum Klein Flottbek, Abteilung Angewandte Pflanzenökologie
Prof. Dr. Kai Jensen, Nikola Lenzewski
- Hamburgisches WeltWirtschaftsinstitut gGmbH
Malte Jahn, Andreas Röhlig, Marie-Christin Rische

8 Literaturverzeichnis

Atencia, A., Rigo, T., Sairouni, A., Moré, J., Bech, J., Vilaclara, E., ... & Garrote, L. (2010). Improving QPF by blending techniques at the Meteorological Service of Catalonia. *Natural Hazards and Earth System Sciences*, 2010, Vol. 10, p. 1443-1455.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., Reinhardt, T. (2011): Operational convective-scale numerical weather prediction with the COSMO model. *Monthly Weather Review* 139, 3887–3905. doi:10.1175/MWR-D-10-05013.1.

Donaldson R. J., Dyer R. M. & Krauss M. J. 1975 An objective evaluator of techniques for predicting severe weather events. Preprints: 9th Conf. Severe Local Storms, Norman, Oklahoma, Amer. Met. Soc., 321-326.

Golding, B. W. (1998): Nimrod: A system for generating automated very short range forecasts. *Meteorological Applications*, 5(01), 1-16.

hydro & meteo (2009): SCOUT. Version 3.3. Lübeck: hydro & meteo GmbH & Co. KG.

Jasper-Tönnies, A. & Jessen, M. (2014) Improved radar QPE with temporal interpolation using an advection scheme, 8th ERAD, 1 - 5 September 2014, Garmisch-Partenkirchen
http://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/377_JasperToennies.pdf

Kühnlein, C., Keil, C., Craig, G. C., Gebhardt, C. (2014): The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Q.J.R. Meteorol. Soc.* doi: 10.1002/qj.2238

LSBG (2016) Sicherstellung der Entwässerung küstennaher, urbaner Räume unter Berücksichtigung des Klimawandels - BMBF-Projekt Stuck, Projektbericht Gewässer und Hochwasserschutz Nr. 7/2016
https://www.stuck-hh.de/fileadmin/user_upload/Stuck-Zwischenbericht_12-12-2016.pdf

Tessendorf, Alrun; Einfalt, Thomas (2012): Ensemble radar nowcasts - a multi-method approach. In: Robert J. Moore, Steven J. Cole und Anthony J. Illingworth (Hg.): *Weather radar and hydrology. Proceedings of a symposium held in Exeter, UK, April 2011*. Wallingford: IAHS Press (IAHS publication, 351), S. 310–316.

Wilks, D. S. (1995): *Statistical Methods in the Atmospheric Sciences*. San Diego, London: Academic Press.