



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Verifikationsmethoden für
Wetterradaextrapolationen

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat)

Verfasser:	Stefan Mayer
Matrikelnummer:	0502165
Studienrichtung:	A415 Meteorologie
Betreuer:	Ao.Univ.Prof.Mag.Dr. Leopold Haimberger
Wien, am:	03. Mai 2011

Zusammenfassung

Wetterradarinformation ist im Bereich des Nowcastings auf Grund ihrer Aktualität, sowie ihrer hohen zeitlichen und räumlichen Auflösung nicht mehr wegzudenken. Mit dem Ziel, dem Meteorologen eine noch bessere Grundlage zur Kurzfrist-Niederschlagsprognose zu bieten, wurden in den letzten Jahrzehnten eine große Zahl an Wetterradarextrapolationen realisiert. Diese Arbeit zeigt, dass Verifikation für jedes Vorhersagesystem in vielerlei Hinsicht unverzichtbar ist, sei es um die aktuelle Güte zu überwachen, verschiedene Systeme miteinander zu vergleichen oder die Vorhersagemethodik zu optimieren. Wie kann man nun diese „Güte“ speziell für Wetterradarextrapolationen definieren? Diesem Problem wurde sich aus praxisbezogener, sowie theoretischer Sicht genähert, um somit umfassende Anforderungen an ein Extrapolationssystem und vor allem an das speziell dafür zu entwickelnde Verifikationschema zu definieren. In dieser Arbeit geht es darum, die Güte einer Extrapolation an sich zu erfassen. Das geht am besten, wenn die vorhergesagte und die beobachtete Variable dieselbe physikalische Größe sind. Daher wurden extrapolierte Radarbilder mit gemessenen Bildern und nicht mit beobachteten Niederschlägen verglichen. Beim Niederschlag ist das besonders gerechtfertigt, da der Zusammenhang zwischen Radarreflektivität und Bodenniederschlag sehr unsicher ist. Es zeigt sich, dass sich theoretisch definierte Qualitätsaspekte einerseits und die praxisorientierten Forderungen nach exakter Orts- und Intensitätsvorhersage andererseits, gut ergänzen. Deshalb greift das hier entwickelte Verifikationsschema für Wetterradarextrapolationen auf bewährte Maße aus der Verifikation von dichotomen Systemen, sowie klassische statistische Momente zurück. Ergänzend werden für probabilistische Vorhersagen übliche Methoden, wie zum Beispiel das Reliability-Diagramm verwendet. Dieses Schema wurde mittels eines MatLab-Prototyps auf jene Wetterradarextrapolationen angewendet, die bei Austro-Control GmbH zurzeit in Verwendung sind. Es zeigte sich, dass die verwendeten Methoden großteils uneingeschränkt, vereinzelt mit adaptierter Interpretation geeignet sind, um alle Qualitätsaspekte einer Wetterradarextrapolation zu beschreiben.

Abstract

Weather radar information play an important role in the field of nowcasting due to its actuality and its high temporal and spatial resolution. A lot of weather radar extrapolation systems implemented in the last decades offer an even better substructure for short term precipitation forecasting. This paper demonstrates that verification for every prognosis system is indispensable for monitoring actual quality, comparing different systems or optimizing prediction methods. But how can this certain "quality" be defined for weather radar extrapolations? This complex of problems is approached in realistic, as well as theoretical perception to define a comprehensive verification scheme for extrapolation systems. This paper tend to measure the quality of extrapolation itself. That runs most effective when the predicted and the observed variable are the same physical unit. That is why extrapolated weather radar images are compared with measured radar images and not with observed precipitation. This is especially justified to precipitation as the correlation between weather radar reflectivity and ground-based precipitation measurements is very uncertain. When verifying weather radar images it arised, that theoretically defined quality aspects according to Murphy (1993) on the one hand, and practice-oriented demands for accurate regional and intensity forecasts on the other hand are complementing one another, and a verification scheme was found to describe the quality of weather radar extrapolation comprehensively. This scheme applies established methods of the verification of dichotomical systems, as well as classical statistic approaches. There is a complementary use of common methods for probability predictions as reliability graphs. This scheme has been implemented as a MatLab-Prototyp and has been applied to weather radar extrapolations currently used at AustroControl GmbH. The methods turned out to be suitable mostly without restrictions. Only in special cases an adapted interpretation is needed to describe all aspects of quality.

Inhaltsverzeichnis

1	Zweck von Wetterradarextrapolationen	1
1.1	Einleitung	1
1.1.1	Erkennung von Gefahren im Nowcasting	1
1.1.2	Niederschlags-Kurzfristprognosen	2
1.2	Anforderungen an Wetterradarextrapolationen	3
1.2.1	Verfolgen einer Starkniederschlagszelle (Ortsvorhersage)	4
1.2.2	Mengenerfassung (Intensitätsvorhersage)	5
2	Extrapolationsalgorithmen	7
2.1	Übersicht	7
2.1.1	Area Tracking	7
2.1.2	Cell Tracking	10
2.2	Extrapolationen beim österreichischen Flugwetterdienst	15
2.2.1	Verwendete Wetterradar-Daten	15
2.2.2	Verwendete Extrapolationen	20
3	Anforderungen an die Verifikation	25
3.1	Zweck einer Verifikation	25
3.2	Verifikation: Probleme und Anforderungen	26
3.2.1	Eigenschaften der WXR- Messung (Theorie und Probleme)	26
3.2.2	Fertige Anforderungen- Schlussfolgerungen	46
4	Verifikationsmethodik	48
4.1	Bias	54
4.2	Accuracy	55
4.3	Reliability	57
4.4	Resolution & Sharpness	60
4.5	Discrimination	61
4.6	Skill	64
4.6.1	Gilbert Skill Score (Equitable Threat Score)	65
4.6.2	Heidke Skill Score	66
4.6.3	True Skill Statistics	66
4.6.4	Vergleich der Skill Scores	67
4.7	Festlegen des Verifikationsschemas	67

5	Verifikationssoftware	70
5.1	Anforderungen	70
5.2	Umsetzung	71
5.2.1	Calculations	72
5.2.2	Plots	76
6	Ergebnisse	84
6.1	Auswertung Winter 2008-2009	85
6.2	Auswertung Sommer 2009	101
7	Ergebnis und Ausblick	113
	Literatur	116

1 Zweck von Wetterradarextrapolationen

1.1 Einleitung

Besonders bei Prognosen im Kurzfristbereich stoßen numerische Wettervorhersagemodelle (NWP) schnell an ihre Grenzen, die in erster Linie durch limitierte Auflösungen sowie nicht exakt bekannte Anfangszustände bedingt sind. Daher muss im Bereich des Nowcastings (Das ist ein Kunstwort zusammengesetzt aus “Now” und “Forecasting”. Es bezeichnet die Prognosemethode bestehend aus Erfassung des Ist-Wetterzustandes und einer Vorhersage für die nächsten Stunden. In der Radarmeteorologie hat sich ein Vorhersagezeitraum von 0-2 Stunden als ideal erwiesen (Hailzl, 2006)) auf alternative Informationsquellen zurückgegriffen werden, die zwar nur den Istzustand der Atmosphäre zeigen, dafür den Anforderungen an Aktualität und hoher räumlicher Auflösung gerecht werden. Diese Bedingungen werden in erster Linie von Information, gewonnen durch Messstationsnetzwerke (TAWES, etc.) und Fernerkundungsdaten, erfüllt. Messstationsnetzwerke haben nach wie vor den großen Nachteil einer zu geringen Stationsdichte, vor allem in gebirgigem Terrain, wo noch dazu auf kleinen Distanzen starke Unterschiede bei den wichtigsten meteorologischen Größen (Temperatur, Wind, Niederschlag, etc.) beobachtet werden können.

Im Bereich der Fernerkundung sind satellitengestützte Messungen und das Wetterradar (WXR) die populärsten Methoden. Ergänzend können zum Beispiel auch Webcams verwendet werden, deren Dichte, vor allem in den letzten Jahren stark gestiegen ist.

1.1.1 Erkennung von Gefahren im Nowcasting

Hauptaufgabe einer Kurzfristprognose muss es sein, meist vom Anwendungsbereich abhängige Gefahren zu erkennen, um ihre Entwicklungen in den nächsten Stunden prognostizieren zu können. Im Allgemeinen wird vor allem in der konvektiven Jahreszeit, das Erkennen von Gewitterzellen, die enormes Gefahrenpotential in Form von Blitzentladungen, Sturmböen, Starkregen oder im Extremfall Hagelschlag bergen, ei-

ne zentrale Rolle spielen. Im Winterhalbjahr steht dagegen eine zeitlich und räumlich gut aufgelöste Prognose von stratiformen Niederschlag, insbesondere von Schneefall, beziehungsweise von gefrierendem Regen im Vordergrund. Es sei angemerkt, dass es aber natürlich auch im Winter zu konvektiv bedingten Gefahren kommen kann und diese dann aufgrund ihrer Seltenheit besonderes Gefahrenpotenzial (insbesondere Starkschneefall) bergen.

Um eine professionelle Kurzfristprognose produzieren zu können, ist es unerlässlich, den genauen Istzustand der entscheidenden Parameter in der Atmosphäre zu kennen. Dafür werden im besten Fall alle zur Verfügung stehenden Messsysteme genutzt und deren Informationen kombiniert, um sich ein umfassendes Bild zu kreieren. Jedes System liefert unterschiedliche Informationen, denen verschiedene Messmethoden, sowie verschiedene Betrachtungsweisen - von der Erdoberfläche aus betrachtet gewinnt man andere Informationen über die Atmosphäre als vom Weltall aus - zu Grunde liegen. Außerdem ist jedes Messsystem fehlerbehaftet. Diese Fehler sind durch Abgleichung mehrerer Systeme zumindest gedanklich reduzierbar. Nur durch Kombination aller Informationen ist eine ausreichende Erfassung des Istzustandes möglich.

Aber gerade bei der Erfassung der aktuellen Niederschlagssituation spielt Wetterradarinformation die bestimmende Rolle. Obwohl ein Wetterradarbild nicht die tatsächliche Niederschlagsverteilung, schon gar nicht am Boden, zeigt, ist es mit Hilfe von Wetterradarinformation und etwas Erfahrung doch möglich, wichtige Eigenschaften wie die räumliche Ausdehnung, die nähere Vergangenheit (Zugbahn, Intensitätsentwicklungen, usw.), sowie die Intensitätsverteilung (in drei Dimensionen) von Niederschlagsgebieten zu erkennen. Diese Informationen lassen eine viel umfangreichere Analyse zu, als das durch reines Betrachten von Niederschlagsmessungen am Boden möglich wäre.

1.1.2 Niederschlags-Kurzfristprognosen

Somit ist Wetterradarinformation im Bereich der Kurzfristvorhersage inzwischen unerlässlich. Sie bietet eine bis dato unerreichte zeitliche (in Österreich zur Zeit fünf Minuten) und räumliche Auflösung (in Österreich 1x1 km), sowie ausreichende Aktualität, um präzise Niederschlagsanalysen beinahe in Echtzeit unterstützen zu können. Um nun vom Istzustand zur Prognose zu gelangen muss der Meteorologe gedanklich extrapolieren.

Mit dem Ziel, dem Meteorologen eine noch bessere Grundlage zur Kurzfrist- Nie-

derschlagsprognose zu bieten, wurden in den letzten Jahrzehnten eine große Zahl an Wetterradarextrapolationen entwickelt und implementiert. Man kann diese als automatisierte Umsetzungen der gedanklichen Extrapolation des Meteorologen sehen, der damit “first guess” Entwicklungen erhält, in die er dann sein meteorologisches und geographisches (orographisches) Know-How einarbeiten kann, um so mit insgesamt weniger Arbeitsaufwand, zur bestmöglichen Niederschlags-Kurzfristprognose zu kommen. Extrapolationen sind somit bestens geeignet um den Arbeitsaufwand für eine Kurzfristprognose zu verringern, beziehungsweise Arbeitsabläufe zu optimieren.

1.2 Anforderungen an Wetterradarextrapolationen

Was sind also die Gütemerkmale einer Wetterradarextrapolation? Murphy (1993) beschreibt für meteorologische Vorhersagen ganz allgemein drei unterschiedliche Güte-typen (Tabelle 1.1).

Eine Vorhersage kann, im Sinne seines Gütetyps 1, als “gut” gewertet werden, wenn sie vom Meteorologen als sinnvoll betrachtet wird. Das Gütemaß ist hier somit die *Konsistenz* zur Referenzprognose des Meteorologen. Eine solche Beurteilung wird daher bei unterschiedlichen Meteorologen auch unterschiedliche Ergebnisse liefern. Der Gütetyp 2 zielt auf die *Qualität* der Vorhersage ab. Die Güte ist umso größer, je näher der vorhergesagte Zustand am gemessenen liegt. Nach Gütetyp 3 ist eine Vorhersage nur dann “gut”, wenn sie zur Entscheidungsfindung des Anwenders beiträgt. Dieser Typ beschreibt den *Wert* der Prognose für den Kunden (Murphy (1993)).

Nun muss man überlegen, welche Kriterien eine WXR-Extrapolation aufweisen muss, um im besten Fall, alle diese Gütefaktoren zu erfüllen.

Entscheidend dafür, dass ein Meteorologe eine WXR-Extrapolation als sinnvoll ansieht, sie also als konsistent bewertet, wird ein “natürliches” Aussehen der Echos, sowie ihre mittlere Bewegung sein, die sich in Grundzügen (vor allem bei stratiformen Echos) mit seiner Prognose decken sollte. Damit eine WXR-Extrapolation hohe Qualität erreichen kann, müssen im Wesentlichen die Position, die räumliche Ausdehnung, sowie die Intensitätsverteilung möglichst nahe an der Vergleichsmessung liegen. Die Bedeutungsgewichtung dieser einzelnen Qualitätsfaktoren ist leicht vom Charakter der Wettererscheinung (stratiform oder konvektiv) abhängig. Der Wert einer Vorhersage hängt stark vom Anwender und Einsatzzweck ab. Je nachdem ob eine Extrapolation, wie bisher angenommen, vom Meteorologen oder direkt vom Endkunden (zum Beispiel als Zusatzinformation zu einer meteorologisch aufbereiteten Textprognose)

Typ	Name	Definition
1	Consistency	Correspondence between forecasts and judgments
2	Quality	Correspondence between forecasts and observations
3	Value	incremental benefits of forecasts to users

Tabelle 1.1:

Kennzeichen einer “guten” Vorhersage laut Murphy (1993).

verwendet wird, müssen die Wert-Kriterien an die Prognose eigens definiert werden.

Murphy (1993) zeigt, dass alle drei Gütemerkmale über mehr oder weniger komplizierte Wechselwirkungen miteinander verknüpft sind. Zum Beispiel ist die Qualität einer Vorhersage einer der entscheidenden Faktoren um auf ihren Wert zu schließen. Daher sollen nun lediglich die Kriterien, die in erster Linie die Qualität bestimmen, näher beschrieben werden.

1.2.1 Verfolgen einer Starkniederschlagszelle (Ortsvorhersage)

Besonders wenn es um die Vorhersage der Entwicklung einer Starkniederschlagszelle geht, spielt die genaue Ortsvorhersage eine entscheidende Rolle. Die Gefahren (Blitzschlag, Starkregen, Hagelschlag, etc.), die von Gewittern ausgehen, beschränken sich im Allgemeinen auf Bereiche von einigen Kilometern. Je zuverlässiger man also die Zugbahn einer solchen Zelle bestimmen kann, umso gezielter können Warnungen gegeben, beziehungsweise etwaige Schutzmaßnahmen ergriffen werden. In den meisten Anwendungsbereichen sind in erster Linie solche Schutzmaßnahmen mit großem finanziellen Aufwand verbunden, zum Beispiel in der kommerziellen Luftfahrt. Genaue und vor allem zuverlässige Kurzfristvorhersagen von Gewitterzugbahnen entscheiden dort aber nicht nur über Wirtschaftlichkeit, sondern im Extremfall sogar über Menschenleben.

Eine gute Ortsvorhersage besteht, wie bereits erwähnt, aus der richtigen Position, die man bei konvektiven Zellen zum Beispiel aufgrund der Position des Intensitätsmaximums definieren kann, sowie einer korrekten horizontalen, räumlichen Ausdehnung (Form und Größe) der Zelle. Natürlich wäre es auch wünschenswert die vertikale Ausdehnung einer Starkechozelle vorherzusagen. Besonders wenn es um die Prognose von Hagel geht, kann auf 3D- Information kaum verzichtet werden. Außerdem wären genauere Schlüsse über das Lebenszyklus-Stadium möglich, in dem sich die Zelle befindet, woraus man wiederum Information über deren weitere Entwicklung erhalten

würde.

1.2.2 Mengenerfassung (Intensitätsvorhersage)

Die entscheidenden Merkmale einer konvektiven Zelle sind aber nicht nur ihre Position und Ausdehnung. Genauso maßgebend im Nowcasting sind ihre Intensitätswerte. Damit ist einerseits die horizontale und vertikale Intensitätsverteilung gemeint, sowie die Maximalintensität.

Prinzipiell muss die Gewitterzelle in ihrem Reifestadium eine mindestens 5 km mächtige Pixelsäule mit der Reflektivitätsstufe größergleich 46 dBZ bzw. 27 mm/h aufweisen. Eine Hagelzelle ist im 3D-Bild durch eine geneigte Pixelsäule zu erkennen, wobei das oberste Reflektivitätspixel um 3-5 km zum Untersten versetzt sein soll. Manchmal sind auch der typische Überhang oberhalb der Schwachechoregion und die Hakenform im Grundriss zu beobachten. Die Radartops müssen kälter als -60°C sein. Weiters muss die Reflektivitätsstufe größergleich 50 dBZ bzw. 50 mm/h eine mindestens 5 km mächtige Pixelsäule oberhalb der Nullgradgrenze aufweisen (Ladungstrennung findet erst oberhalb der 0°C -Grenze statt). Die Stufe größergleich 46 dBZ bzw. größergleich 27 mm/h soll eine vertikale Ausdehnung von mindestens 8 km besitzen (Hailzl, 2006). Um eine Gewitterzelle also genau analysieren zu können, ist 3D Information unerlässlich. Aus praktischen Gründen werden aber nach wie vor einige WXR-Extrapolationen aus 2D-Daten (MAXCAPPI-Darstellung) generiert (2.2.1), aus denen viele dieser Merkmale nicht abgeleitet werden können. Somit muss sich der Meteorologe in diesen Fällen mit einer Vorhersage der Intensitätsmaxima pro vertikaler Pixelsäule zufrieden geben, um daraus auf die ungefähre Dynamik (und die damit verbundenen Gefahren, Niederschlagsart und Intensität) zu schließen.

Eine gute Intensitätsvorhersage spielt aber nicht nur bei konvektiven Ereignissen eine wichtige Rolle. Im Falle von stratiformen Niederschlag, besonders wenn er als Schnee fällt, ist es essentiell, die Niederschlagssintensität für die nächsten Stunden zuverlässig vorherzusagen. Als gutes Beispiel kann man Prognosen für Straßenräumdienste im Winter heranziehen, für die ein paar Zentimeter Schnee mehr (im Vergleich zur mittelfristig eingeplanten Schneemenge) großen Mehraufwand im organisatorischen und finanziellen Sinn bedeuten, beziehungsweise bei unzureichender Kurzfristinformation auch zu wenig Kapazitäten zur Verfügung stehen, um die Räumarbeiten ausreichend schnell zu bewältigen, damit Verkehrsbehinderungen in Maßen gehalten werden.

Abgesehen vom Kundennutzen, spielt die zu erwartende Niederschlagsintensität auch für den Meteorologen eine entscheidende Rolle, wenn es um die Prognose der Schneefallgrenze geht. Bei länger anhaltendem, intensivem Niederschlag kommt es zur Abkühlung der Luftschicht, in die der (feste) Niederschlag fällt. Dieser Luftschicht wird bekanntlich, die zur Schmelzung der festen Hydrometeore notwendige, Energie entzogen. Somit sinkt die Schneefallgrenze lokal stark ab (Steinacker, 1983). Eine Kurzfristprognose der Niederschlagsintensität kann somit notwendig sein, um eine Änderung der Schneefallgrenze in den nächsten Stunden richtig prognostizieren zu können.

2 Extrapolationsalgorithmen

An dieser Stelle soll nun ein Überblick über die verschiedenen Algorithmen zur WXR-Extrapolation gegeben werden. Um in weiterer Folge ein passendes Verifikationsschema entwickeln zu können, ist es notwendig, über die unterschiedlichen Ansätze und daraus resultierenden Eigenschaften verschiedener Extrapolationstechniken informiert zu sein.

2.1 Übersicht

Grundsätzlich existieren zwei verschiedene Arten von Wetterradarextrapolationen.

2.1.1 Area Tracking

Die erste und gleichzeitig auch historisch ältere der beiden Möglichkeiten basiert ausschließlich auf klassischen Bildverarbeitungsmethoden, mit denen versucht wird, Pixel-Muster in mindestens zwei Bildern einander zuzuordnen und somit auf den Verlagerungsvektor zu schließen. Weit verbreitet sind Extrapolationsalgorithmen dieser Art, die den Mustervergleich mittels Kreuzkorrelation bewerkstelligen. Dieses Prinzip wurde bereits von Rinehart und Garvey (1978) oder Kessler und Russo (1963) vorgestellt und nennt sich TREC (tracking radar echo by correlation). Die Arbeitsweise dieses Extrapolationstyps ist in Abbildung 2.1 näher erläutert. Die folgende Beschreibung basiert auf Mecklenburg (2000).

Das erste, von zwei aufeinander folgenden WXR-Bildern wird durch ein reguläres Gitter geteilt und um jeden der entstehenden Gitterpunkte herum, wird die sogenannte „*tracking area*“ oder „*box size*“ definiert. Weiters wird ein Suchfenster (searching area) festgelegt, dessen Größe durch die maximal zulässige Verlagerungsgeschwindigkeit bestimmt wird. Dieses Suchfenster ist dann jener Bereich, der mit dem Muster in der dazugehörigen *tracking area* verglichen wird. Zu diesem Zweck wird das Muster der *tracking area* auf jede mögliche Position im Suchfenster verschoben und für alle Verschiebungen ein Testkriterium, meistens der Korrelationskoeffizient laut Gleichung

2.1 berechnet. Wobei die x_i die Pixel im ersten Bild, y_i die Pixel im zweiten Bild, $\bar{x}$, sowie $\bar{y}$ deren Mittelwerte und N die Gesamtpixelzahl repräsentieren.

$$R_{x|y} = \frac{S_{x|y}}{\sqrt{S_x} \sqrt{S_y}} \quad (2.1)$$

$$S_{x|y} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x}) (y_i - \bar{y}) \quad (2.1a)$$

$$S_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad \text{und} \quad S_y = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \quad (2.1b)$$

Der gültige Verlagerungsvektor für eine „tracking area“ ergibt sich dann durch jene Verschiebung, die zum maximalen Korrelationskoeffizienten führt.

Mecklenburg (2000) schlägt alternativ zum Korrelationskoeffizienten die mittlere, absolute Differenz (Gleichung 2.2) als Testkriterium vor.

$$DIFF_{x|y} = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (2.2)$$

Der Unterschied zwischen den beiden Testkriterien liegt in der Gewichtung der Abweichungen. Bei Verwendung des Korrelationskoeffizienten werden größere Intensitätsabweichungen durch die quadratische Abhängigkeit stärker gewichtet als kleinere. Daher bietet diese Methode Vorteile bei der Verlagerung konvektiver Systeme, wogegen die mittlere absolute Differenz eventuell bei stratiformen Echos zu bevorzugen ist (Mecklenburg, 2000).

Ein häufiges Problem der TREC-Algorithmen liegt in verrauschten Bewegungsvektorfeldern. Dadurch entstehen konvergente und divergente Bereiche im Bild, wodurch es zu Verzerrungseffekten (siehe Abbildung 2.7) kommen kann (Reyniers, 2008). Zur Korrektur dieser Fehler wird im Allgemeinen die zweidimensionale Kontinuitätsgleichung auf das Feld der Bewegungsvektoren angewendet. Diese Methodik wurde von Li et al. (1995) vorgeschlagen und nennt sich COTREC (continuity of TREC vectors). Dabei wird grundsätzlich die Kostenfunktion aus Gleichung 2.3, unter der Bedingung der zweidimensionalen Kontinuitätsgleichung (Gleichung 2.4), minimiert. u_0 und v_0 sind die Komponenten der Bewegungsvektoren, die durch TREC ermittelt wurden,

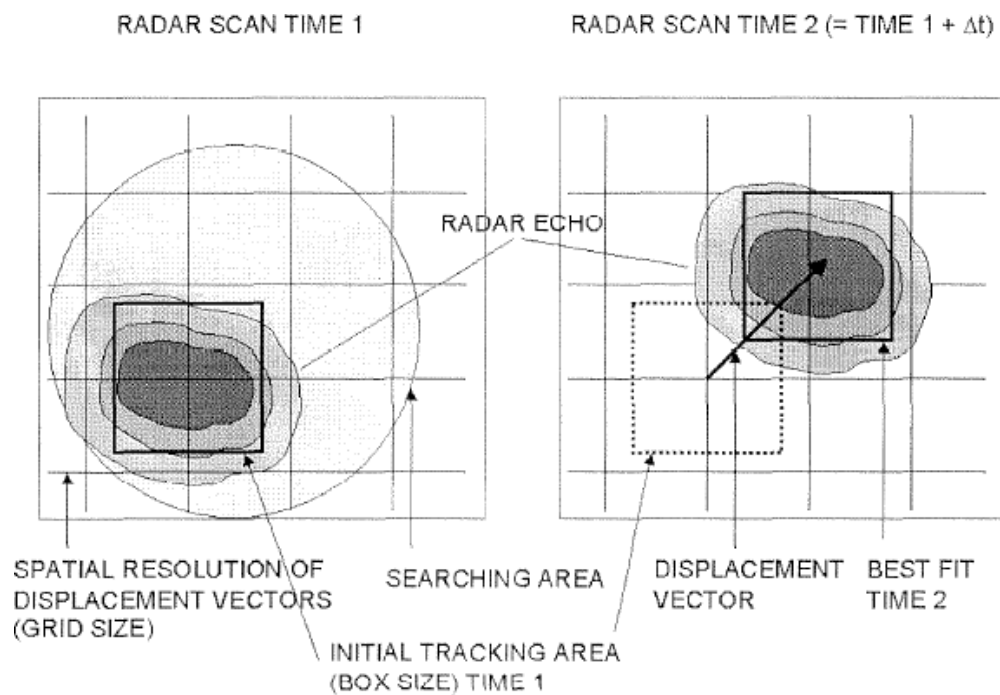


Abbildung 2.1:

Arbeitsweise des TREC-Algorithmus nach Rinehart und Garvey (1978). Die Abbildung wurde aus Mecklenburg (2000) übernommen. Man erkennt die Unterteilung des Bildes durch ein reguläres Gitter, sowie die „tracking area“ um einen der entstehenden Gitterpunkte. Diese „tracking area“ (auch „box size“ genannt) wird dann innerhalb des Suchradius („searching area“) verschoben und mit dem Pixelmuster im zweiten Bild verglichen. Jene Verschiebung, die den maximalen Korrelationskoeffizienten liefert, bestimmt den gültigen Bewegungsvektor.

die Komponenten u und v hingegen bilden das neue, divergenzfreie Vektorfeld.

$$J(u, v) = [(u - u_0)^2 - (v - v_0)^2] \, dx dy \quad (2.3)$$

$$\frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} = 0 \quad (2.4)$$

Generell kann man aber sagen, dass es sich bei Extrapolationsmethoden, die auf Mustererkennung basieren um universal einsetzbare Methoden handelt, die für die meisten Niederschlagsregime, adäquate Kalibrierung vorausgesetzt, brauchbare Ergebnisse liefern können.

Ergänzend zu dieser Gruppe von Extrapolationen sei noch die Verlagerung mittels „*optical flow*“ erwähnt, die zum Beispiel von Peura und Hohti (2004) genau beschrieben wird.

2.1.2 Cell Tracking

Die Algorithmen der zweiten Art verfahren gänzlich anders. Beim „*cell tracking*“ (CT) wird versucht die Bewegung einzelner Objekte oder Zellen, das sind im Allgemeinen zusammenhängende Bildbereiche hoher Intensität, zu beschreiben. Daher nennt man diese Gruppe von Extrapolationen auch *objektbasierte* Verfahren. Der beschriebene Ansatz bedingt, dass CT-Algorithmen nur auf konvektive Ereignisse angewendet werden können und somit im Gegensatz zu *area tracking*-Verfahren (Abschnitt 2.1.1) nicht universal verwendbar sind. Jeder CT-Algorithmus besteht aus zwei essentiellen Teilen (Reyniers, 2008). Den ersten Teil bildet ein *Detektionsalgorithmus*. Er definiert die einzelnen, zu verlagernden Zellen in den beiden WXR-Bildern. Zu diesem Zweck werden in der Regel ein oder mehrere Intensitätsschwellwerte verwendet (siehe Abbildung 2.2). Die Wahl dieser/dieses Schwellwerte(s) beeinflusst die weitere Verlagerung maßgeblich. Immerhin wird in weiterer Folge für jede detektierte Zelle ein eigener Verlagerungsvektor ermittelt.

Abbildung 2.2 zeigt nur einige wenige Lösungsansätze zu dieser zentralen Problematik beim *cell tracking*. Es soll an dieser Stelle nur ein Eindruck der Komplexität dieser Themas vermittelt werden. Für jede Zelle werden nach der Detektion charakteristische Merkmale, die von Verfahren zu Verfahren unterschiedlich definiert sind gespeichert. Übliche Merkmale sind zum Beispiel die Koordinaten des Massenschwerpunktes, die Fläche der Zelle oder die Höhe der Echo-Tops.

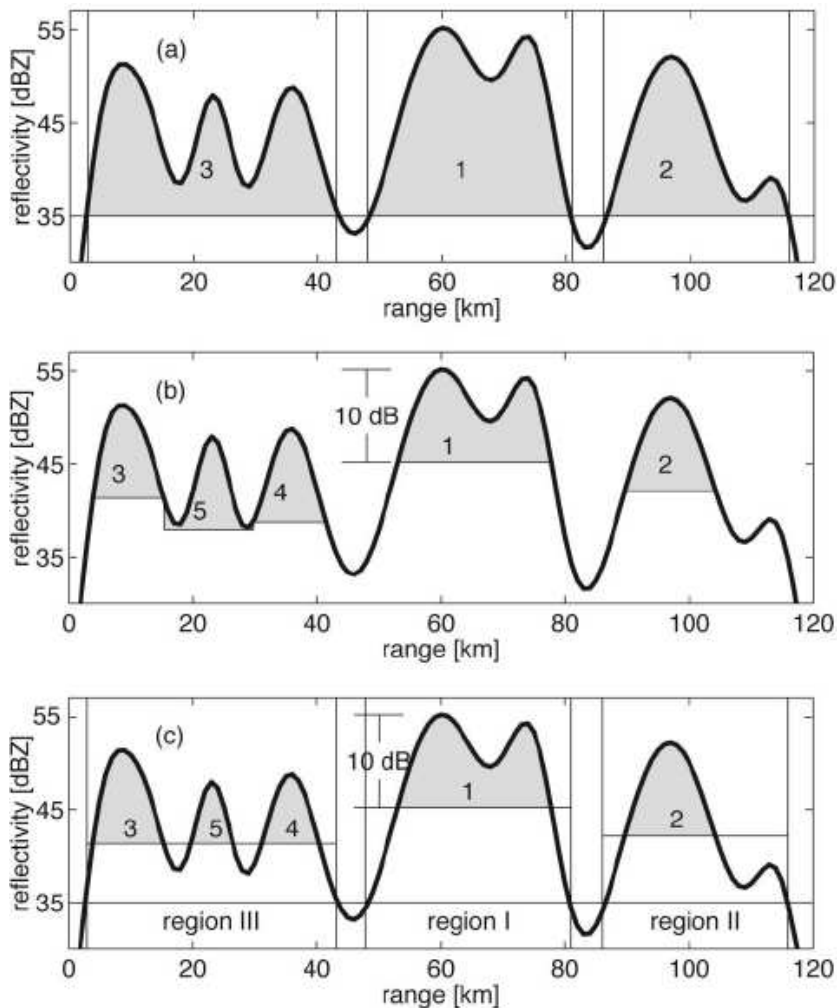


Abbildung 2.2:

Schematischer Vergleich von Methoden zur Zelldetektion mittels Schwellwerten nach Handwerker (2002): (a) Die einfachste dieser drei Methoden verwendet ausschließlich einen Schwellwert (35 dBZ) um einzelne Starkechozonen zu definieren. (b) Hier werden die oberen 10 dBZ jedes lokalen Maximums als Zelle detektiert. Dies kann laut Handwerker dazu führen, dass fälschlicherweise schwächere Reflektivitäten einer Zelle einer benachbarten, stärkeren Zelle zugeordnet werden. Dieser Effekt tritt bei Zelle Nummer 5 auf. (c) Die dritte und gleichzeitig aufwendigste Methode ist jene, die von Handwerker selbst vorgeschlagen wurde und im Rahmen des TRACE3D-Algorithmus verwendet wird. Als erstes wird hier das gesamte Bild durch den unteren Schwellwert (35 dBZ) in einzelne Regionen unterteilt. Danach teilt ein zweiter Schwellwert, der sich aus der Maximalreflektivität einer Region minus 10 dBZ berechnet, die Regionen in einzelne Zellen auf.

Nach der Detektion der einzelnen Zellen kommt der zweite wichtige Bestandteil eines „cell tracking“-Verfahrens zu Einsatz. Nun gilt es die Zellen aus dem ersten Bild im zweiten Bild wieder zu finden. Dieser Vorgang wird „*matching*“ genannt. Dazu werden nun die zuvor gespeicherten Zellenmerkmale verwendet. Innerhalb eines definierten Suchradius um die Ausgangszelle im ersten Bild werden alle Zellen, die sich im zweiten Bild innerhalb dieses Radius befinden, aufgrund ihrer Merkmale mit der Ursprungszelle verglichen und nach verfahrensabhängigen Kriterien wird ihr schlussendlich die „ähnlichste“ Zelle zugeordnet. Daraus wird in weiterer Folge der Verlagerungsvektor für jede Zelle gewonnen.

An dieser Stelle sollen nun einige bekannte CT-Algorithmen aufgelistet werden: (Die Auswahl wurde von Reyniers (2008) übernommen.)

- TITAN

„Thunderstorm Identification Tracking Analysis and Nowcasting“ ist wohl der am Weitesten verbreitete CT-Algorithmus. Deswegen soll er etwas genauer erläutert werden. Er wurde von Dixon und Wiener (1993) entwickelt und verwendet grundsätzlich dreidimensionale, für einzelne Elemente der Detektion aber auch zweidimensionale Radardaten in einem kartesischen Koordinatensystem. Zur Zellidentifikation wird/werden ein/zwei Schwellwert(e) verwendet. Der zweite Schwellwert wird notwendig, wenn zwei eindeutig eigenständige Zellen sehr nahe aneinander geraten, sich aber aus physikalischer Sicht nicht vereinen. In diesen Fällen kann es vorkommen, dass sich die Zellen bei Verwendung von nur einem Schwellwert überlappen. Weiters muss jede Zelle eine gewisse Größe aufweisen um detektiert zu werden. Für die einzelnen Zellen wird eine lange Liste an Erkennungsmerkmalen gespeichert. Die wichtigsten sind die Koordinaten des Reflektivitätsschwerpunkts, das Volumen der Zelle, ihre Größe und Form, wobei diese einerseits durch eine Ellipse beschrieben wird, die an die horizontale Projektion der Zelle angepasst wurde, andererseits inzwischen für einige Anwendungen auch durch ein entsprechendes Polygon (siehe Abbildung 2.3).

Um die gefundenen Zellen in zwei aufeinanderfolgenden Bildern einander zuzuordnen, werden von TITAN zwei Methoden verwendet. Als erstes wird nach Überlappungsflächen gesucht. Sollten sich Flächen zu den unterschiedlichen Zeitpunkten signifikant überlappen, so wird angenommen, dass es sich um dieselbe Zelle handelt. Bei vollständiger Überlappung, also perfekter Übereinstimmung einer Zelle in zwei aufeinanderfolgenden Bildern ergibt sich das Übereinstim-

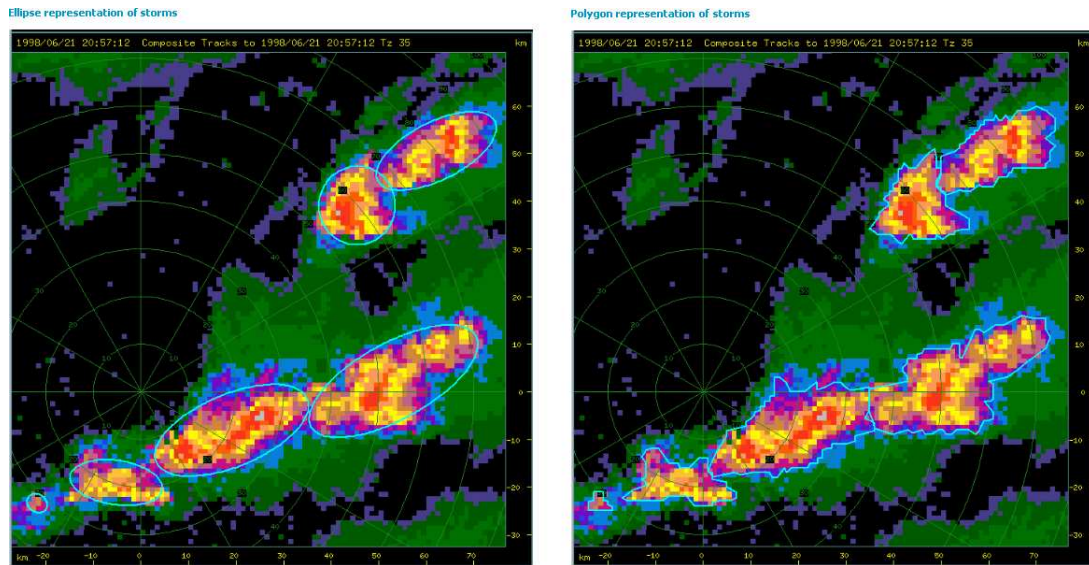


Abbildung 2.3:

TITAN speichert eine große Zahl an Parametern um eine Zelle zu beschreiben. Unter anderem wird die Form der Zelle als Ellipse, sowie als Polygon gespeichert. (Quelle: TITAN-Homepage)

mungsmaß zu 2.0. Im Gegensatz dazu lautet das Ergebnis 0.0 wenn keine Überlappung gefunden wurde. Üblicherweise wird ab einer Übereinstimmung von 0.6 angenommen, dass es sich um dieselbe Zelle handelt. Bleibt dieser Ansatz ohne Erfolg, so kommt ein optimiertes, kombinatorisches Zuordnungsschema zum Einsatz, das mögliche Zugbahnen der Zellenschwerpunkte vergleicht. Die Zuordnung kommt dabei im Wesentlichen durch die Minimierung einer Kostenfunktion zustande.

TITAN wurde ursprünglich als Analysetool entwickelt um Gewitterzellen objektiv detektieren und verfolgen zu können. Aufgrund der umfangreichen Informationen, die TITAN von den einzelnen Zellen speichert, bietet es dazu alle Möglichkeiten. Inzwischen ist das Software-Paket, das im Übrigen zu wissenschaftlichen und nicht gewerblichen Zwecken frei erhältlich ist, stark angewachsen und beinhaltet seit längerem eben auch eine Anwendung zur Erstellung von Kurzfristprognosen.

Für nähere Informationen sei auf die TITAN-Homepage verwiesen, auf der diese Kurzzusammenfassung aufbaut. Die entsprechende URL ist im Literaturverzeichnis vermerkt.

- SCIT

Ein weiterer, oft verwendeter CT-Algorithmus, der mit dreidimensionalen Radardaten arbeitet, nennt sich SCIT (Storm Cell Identification and Tracking) und wurde von Johnson et al. (1998) entwickelt. Im Unterschied zu TITAN werden zur Zelldetektion hier sieben Intensitätsschwellwerte (30, 35, 40, 45, 50, 55, 60 dBZ) verwendet. Somit ist SCIT in der Lage, nicht nur isolierte Gewitter, sondern auch die einzelnen Zellen in cluster- und linienförmig organisierten Systemen gut zu erfassen. Die Zellzuordnung wird durch ein Distanzkriterium errechnet, wobei sich das Suchfenster aus der Geschwindigkeit, sowie Bewegungsrichtung der Zelle während der letzten Zeitschritte ergibt. Dabei werden, analog zu TITAN, mögliche Zellteilungen („splitting“) und Vereinigungen („merging“) berücksichtigt. Der Verlagerungsvektor wird schlussendlich mit Hilfe einer „least square“-Methode linear aus den letzten zehn bekannten Positionen der Zelle bestimmt. SCIT ist im Gegensatz zu TITAN nicht kostenlos erhältlich.

- TRACE3D

Als nächstes soll der von Handwerker (2002) entwickelte TRACE3D-Algorithmus Erwähnung finden. Wie der Name bereits vermuten lässt, verwendet auch dieser Algorithmus dreidimensionale Radardaten in Polarkoordinaten. Zur Zelldetektion wird hier ein Schema mit zwei Schwellwerten verwendet, dass in Abbildung 2.2 bereits näher erläutert wurde. Das besondere daran ist, dass im Gegensatz zu TITAN oder SCIT, der zweite Schwellwert relativ zur Maximalintensität im entsprechenden Bildbereich gewählt wird, wodurch laut Handwerker die Zelldetektion deutlich verbessert werden kann. Die Zellzuordnung basiert, ähnlich zu SCIT, auf der Minimierung einer Kostenfunktion unter Zuhilfenahme der Zellbewegungen während der vergangenen Zeitschritte, die man zu Beginn aus VVP-Feldern (Volume Velocity Processing) gewinnen kann, um das Suchfenster zu definieren. Mögliche Zellteilungen und Vereinigungen werden auch hier berücksichtigt. Der Source-Code zu TRACE3D ist auf Anfrage frei verfügbar.

- CELLTRACK und TRT

Bei CELLTRACK und TRT handelt es sich um zwei unterschiedliche CT-Verfahren, die jedoch beide zweidimensionale Radardaten in der MAXCAP-PI (siehe Abschnitt 2.2.1) Darstellung verwenden. CELLTRACK wurde vom tschechischen Wetterdienst (CHMI) als Ergänzung für konvektive Wetterlagen

Radarstation	Latitude	Longitude
Wien/Raichenwarth	N48.074	E16.536
Salzburg/Feldkirchen	N48.065	E13.062
Patscherkofel	N47.467	E11.217
Zirbitzkogel	N47.072	E14.560

Tabelle 2.1:

Koordinaten der österreichischen Wetterradarstationen aus Kaltenböck et al. (2010a)

zum operationell verwendeten COTREC-Algorithmus entwickelt (Kyznarová und Novák, 2005; Novák, 2007). Zur Zellidentifikation wird dabei nur ein Schwellwert (44 dBZ) verwendet. Die Zellzuordnung verwendet das Verlagerungsvektorfeld aus dem COTREC-Algorithmus, um das Suchfenster zu definieren.

TRT („Thunderstorm Radar Tracking“) andererseits wurde von Hering et al. (2004) entwickelt und wird vom Schweizer Wetterdienst (MeteoSwiss) verwendet. Ähnlich zu TRACE3D wird hier ein adaptives Schwellwertschema verwendet, das es möglich machen soll, Zellen in allen Stadien ihres Lebenszyklus gut zu erfassen. Die Zellzuordnung basiert auf aus der Historie geschätzten Verlagerungsvektoren. Als Erkennungsmaß wird dann die Überlappung der verlagerten, „alten“ Zelle und der, im aktuelleren Bild gemessenen, verwendet. Im Falle einer Erstdetektion, das heißt wenn keine Zellhistorie vorhanden ist, greift das System auf Verlagerungsfelder zurück, die zuvor durch kreuzkorrelationbasierte Extrapolation ermittelt wurden.

2.2 Extrapolationen beim österreichischen Flugwetterdienst

2.2.1 Verwendete Wetterradar-Daten

Der österreichische Wetterradarverbund, der vom österreichischen Flugwetterdienst AustroControl GmbH (ACG) betrieben wird, besteht zur Zeit aus vier identischen Doppler-Radargeräten der Firma EEC (Enterprise Electronic Corporation, USA) (Kaltenböck, 2007). Die Reichweite eines einzelnen Radars beträgt theoretisch etwa 230 km. Sie arbeiten mit einer Frequenz von 5,620-5,625GHz (G-Band), die es ermöglicht, das gesamte Spektrum flüssiger sowie fester Niederschläge zu erfassen.

Die Koordinaten der einzelnen Stationen (Kaltenböck et al., 2010a) sind der Tabelle 2.1 zu entnehmen. Eine 360° Drehung der Radarantenne, bei der eine Elevation vollständig erfasst wird, dauert zirka 15 Sekunden. Das schrittweise Abtasten aller



Abbildung 2.4:
Österreichisches Wetterradarnetz aus Kaltenböck (2007)

16 Elevationen benötigt dementsprechend etwa vier Minuten. Der Zeitstempel des Bildes entspricht dem Startzeitpunkt des Abtastvorganges. Das Radarbild steht nach einer Bearbeitungszeit von etwa 40 Sekunden, also effektiv erst fünf Minuten nach dem zugewiesenen Aufnahme-Zeitpunkt zur Verfügung (Kaltenböck et al., 2010a). Diese Bearbeitungsphase ist notwendig, um von polaren in kartesische Koordinaten umzurechnen, sowie die Endprodukte zu generieren.

Es gibt eine Reihe von WXR-Produkten, die automatisch erzeugt werden. Die Auswahl des geeigneten Produkts richtet sich nach dem Zweck, den es zu erfüllen gilt. Für das genaue Analysieren und Vermessen einer Gewitterzelle werden gänzlich andere Darstellungen verwendet, als zur routinemäßigen Überwachung des gesamten österreichischen Luftraums. Für einen guten Überblick zu diesem Thema sei auf Hailzl (2006) verwiesen. An dieser Stelle soll nur die WXR-Austria-Composite-MAXCAPPI-Darstellung näher erläutert werden, weil alle im weiteren Verlauf verwendeten Extrapolationen auf sie aufsetzen.

Das WXR-Austria-Composite bezeichnet die Kombination aller vier zuvor erwähnten WXR-Stationen in einem Bild. Damit ist es möglich, das gesamte Bundesgebiet Österreichs (abgesehen von etwaigen Abschattungsgebieten, siehe Abschnitt 3.2.1.2) hinsichtlich Niederschlag zu überwachen. Dieses Bild besitzt eine Auflösung von 1x1 km und ist somit 824x648 Bildpunkte groß. Die Reflektivität, beziehungsweise die Nie-

derschlagsintensität wird dabei in 14 Klassen aufgelöst. Die Abstufung, sowie der Zusammenhang der verschiedenen Einheiten ist in Abbildung 2.5 dargestellt. Aufgrund des sehr breiten Wertebereichs, den das zurückgestreute Radarsignal annehmen kann, wird oft die logarithmische Einheit dBZ verwendet. Mit deziBel (dB) bezeichnet man ein logarithmisches Verhältnis zwischen zwei Messgrößen, das mit 10 skaliert ist. In der Radarmeteorologie wird der Radarreflektivitätsfaktor Z (siehe Abschnitt 3.2.1.2) meist auf den Wert $1 \text{ mm}^6 \text{ m}^{-3}$ bezogen (Hailzl, 2006). Nach Gleichung 2.5 lässt sich der Radarreflektivitätsfaktor daher in der Einheit dBZ schreiben. Die genaue Umrechnung in entsprechende Niederschlagsraten (in mm/h) soll in Abschnitt 3.2.1.1 ausführlich erläutert werden.

$$Z = 10 \log \left(\frac{Z}{1 \text{ mm}^6 \text{ m}^{-3}} \right) [\text{dBZ}] \quad (2.5)$$

Im weiteren Verlauf dieser Arbeit wird es notwendig sein, ein in 14 Intensitätsklassen aufgelöstes WXR-Bild in die Einheit dBZ zu transformieren. Zu diesem Zweck wurde, wie in Abbildung 2.5 dargestellt, der Zusammenhang der Intensitätsklassenmittelwerte in dBZ durch eine lineare Funktion angenähert, die durch Minimierung der Abweichungsquadrate gefunden wurde. Der vereinfachte Zusammenhang, der im Rahmen der Umsetzung verwendet werden soll, lautet:

$$Z [\text{dBZ}] \approx 4,0426 \cdot \text{Int} [\text{Klasse}] + 7,9014 \quad (2.6)$$

In der MAXCAPPI-Visualisierung (Maximum Constant Altitude Plan Position Indicator) wird die in den dreidimensionalen Radardaten enthaltene Information reduziert, um sie zweidimensional darstellen zu können. MAXCAPPI bedeutet, dass im Grundriss für jeden Pixel die Maximalintensität der vertikalen Niederschlagssäule dargestellt wird, also der größte Reflektivitätswert aller Höenschichten über dem jeweiligen Bildpunkt. Im Aufriss (Zonalschnitt), wenn vorhanden, wird die Maximalintensität jeder Spalte und im Kreuzriss (Meridionalschnitt) die jeder Zeile dargestellt. Für die Wetterradarextrapolation wird jedoch nur der Grundriss verwendet. In der Abbildung 2.6 stehen zum Abschluss alle relevanten Informationen zum WXR-Austria-Composite in der MAXCAPPI-Darstellung visualisiert und zusammengefasst zur Verfügung.

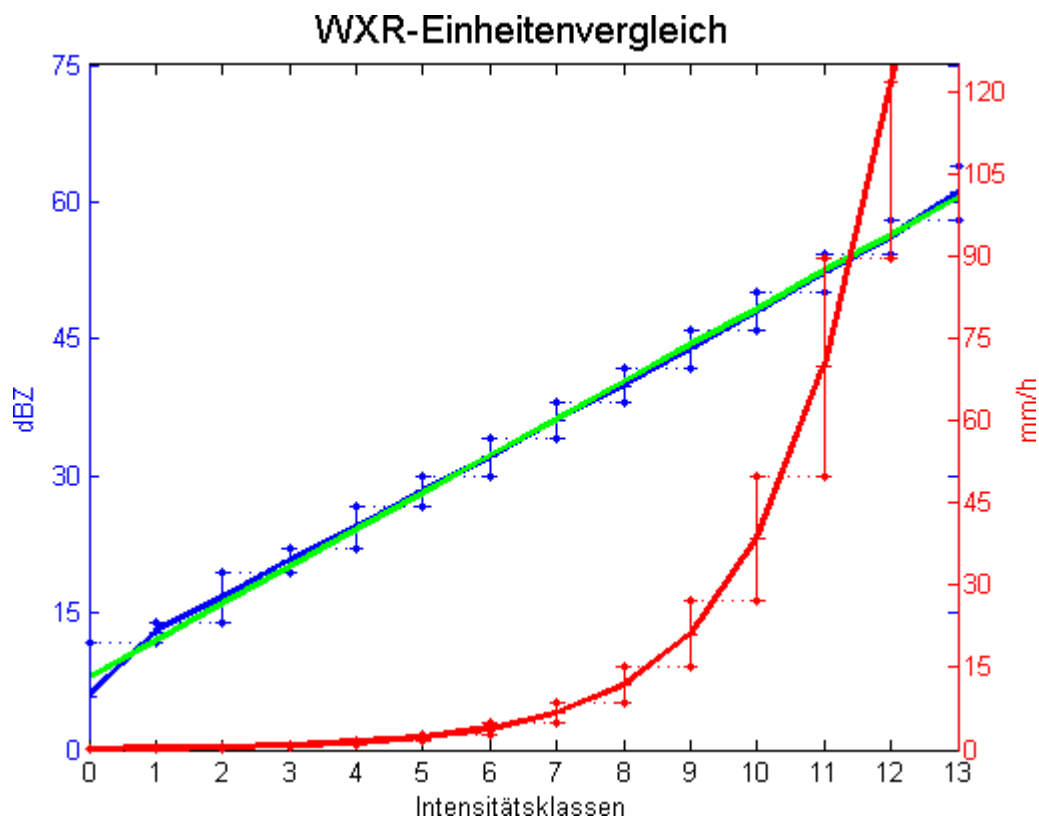


Abbildung 2.5:

Vergleich der gängigen Einheiten für WXR-Reflektivitäten: Im WXR-Austria-Composite MAXCAPPI wird die Reflektivität in 14 Intensitätsklassen aufgelöst und dargestellt. In diesen Klassen steigt die Niederschlagsrate (in [mm/h], im Diagramm Rot dargestellt) von unten nach oben exponentiell an, die Reflektivität (Blau dargestellt), im logarithmischen Maß dBZ gemessen, naturgemäß annähernd linear. Die grüne Kurve zeigt zusätzlich jene lineare Funktion, die den Zusammenhang der Intensitätsklassenmittelwerte in dBZ am Besten widerspiegelt. Sie wurde mittels Minimierung der Abweichungsquadrate gewonnen.

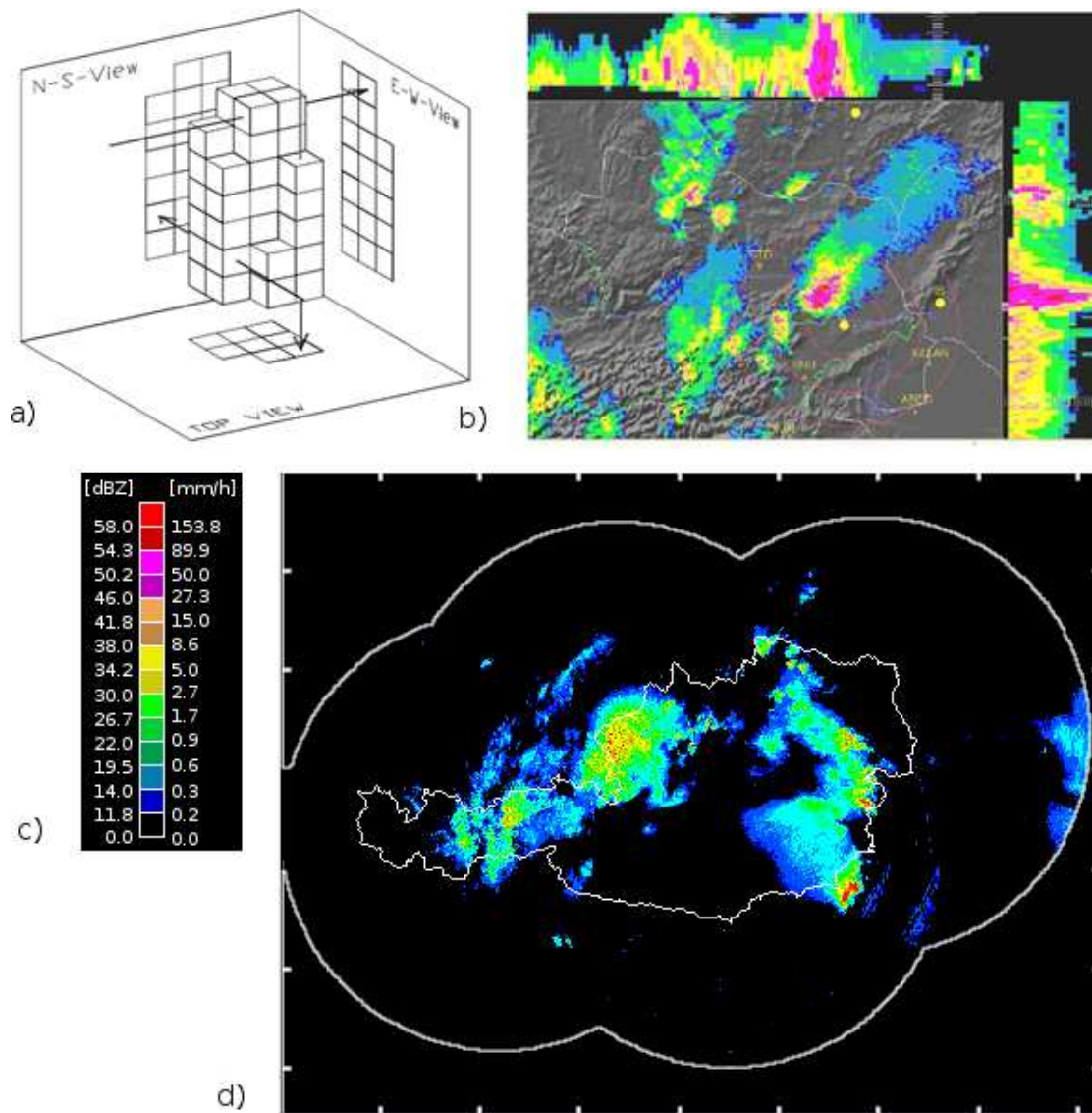


Abbildung 2.6:

Überblick: WXR-Austria-Composite MAXCAPPI: a) MAXCAPPI-Generierung aus Hailzl (2006). Es wird immer der maximale Wert der entsprechenden Niederschlags-säule geplottet; b) Ausschnitt aus einem vollständigen Austria-Composite MAXCAPPI mit Grund-, Auf- und Seitenriss aus Hailzl (2006). Zur Extrapolation wird jedoch nur der Grundriss verwendet; c) Logarithmisch abgestufte 14-teilige Reflektivitäts-/Niederschlagsratenskala; d) Erfassungsbereich des WXR-Austria-Composite, in das vier Radarstationen eingebunden sind.

2.2.2 Verwendete Extrapolationen

Im weiteren Verlauf dieser Arbeit werden ausschließlich TREC-Extrapolationen verwendet werden, die bei AustroControl GmbH berechnet wurden. Das Grundprinzip des TREC-Verfahrens wurde bereits in Abschnitt 2.1.1 erläutert. An diesem Punkt sollen daher nur noch die Feinheiten dieser spezifischen Extrapolationsart erwähnt werden. Ergänzend werden auch objektorientierte Verfahren (siehe Abschnitt 2.1.2) bei ACG angewandt. Diese sollen hier jedoch nicht näher beschrieben werden.

Die Qualität einer TREC-Extrapolation ist, wie sich in der Praxis gezeigt hat, stark von einzelnen Parametereinstellungen abhängig.

Im verwendeten Datensatz variieren fünf Parametergruppen, die in Tabelle 2.2 näher beschrieben werden. Dabei handelt es sich erstens um den zeitlichen Abstand zwischen den Ausgangsbildern, dem *Ausgangsintervall*. Es liegt die Annahme nahe, dass ein großes Ausgangsintervall (zum Beispiel 15 Minuten) mit zunehmender Vorhersagedauer bessere Ergebnisse liefert, weil zeitlich kleinskalige Abweichungen von der mittleren Bewegung des Niederschlagsregimes besser eliminiert werden können. Im extremen Kurzfristbereich (aus der Praxis geschätzt bis 30 Minuten) kann dies aber auch von Nachteil sein, hier ist das genaue Erfassen dieser zeitlichen Fluktuation durchaus erwünscht, sodass kleinere Ausgangsintervalle zu bevorzugen sind.

Der zweite Parameter beschreibt die Methode, mit der die Ausgangsbilder gefiltert werden. Er setzt sich aus einer Filterart (Gauß- beziehungsweise Binomialfilter), sowie der Stärke des Filters zusammen. Diese Stärke definiert sich durch die Größe, des den Pixel umgebenden Bildbereiches, der zur Berechnung herangezogen wird. Durch stärkere Filter gehen kleinskalige (räumliche) Phänomene weniger in die Extrapolation ein. Auch dies sollte theoretisch bei längerer Vorhersagezeit Vorteile bringen.

Auch das *Testkriterium* variiert im TREC-Datensatz. Es wird einerseits die klassische Kreuzkorrelation berechnet, andererseits auch die Summe der Abweichungsquadrate (sum of squares). Die spezifischen Vor- und Nachteile dieser Vergleichsmöglichkeiten sind nicht offensichtlich. Bei beiden werden größere Abweichungen durch die quadratische Abhängigkeit stärker gewichtet (siehe auch Abschnitt 2.1.1).

Der nächste Parameter hat sich in der Praxis als der einflussreichste erwiesen. Die *Nestgröße* beschreibt die Größe der „tracking area“. Auch hier gilt, dass durch größere Nester kleinskalige, räumliche Phänomene weniger Einfluss auf das Extrapolationsergebnis haben. Dieser Detailverlust wird mit wachsender Extrapolationszeit zum Vorteil, weil die mittlere Bewegung des Gesamtsystems besser erfasst werden

Beispiel	Parameter	Beschreibung	Variationen im Testda- tensatz
5min	Ausgangs- intervall	Zeit zwischen den, für die Berechnung herangezogenen Bildern	5min, 10min, 15min
b7	Filter	Die Ausgangsbilder werden vor der Berechnung mit Filtern geglättet: „g“ bezeichnet einen Gauß-, „b“ einen Binomialfilter; Die Zahl im Namen repräsentiert die Stärke des Filters.	g5, b7
corr	Testkriterium	Beschreibung siehe Abschnitt 2.1.1. Neben der klassischen Kreuzkorrelation (siehe Gleichung 2.1), die mit „corr“ gekennzeichnet ist, besteht auch die Möglichkeit die Summe der Abweichungsquadrate „sosq“ als Testkriterium zu verwenden.	corr, sosq
56	Nestgröße (tracking area)	Beschreibung siehe Abschnitt 2.1.1. Die tracking area ist immer quadratisch, ihre Größe wird in Pixel angegeben. „16“ bezeichnet also eine Fläche von 16x16 Bildpunkten. Die Nestgröße ist aus eigener Erfahrung der wichtigste Parameter.	16 32, 56
opt2	Such- algorithmus	Aus Zeitgründen wird zur Berechnung des Testkriteriums nicht jedes Pixel verglichen. Der Suchalgorithmus bestimmt welche Pixel zum Vergleich herangezogen werden.	opt2, grid2

Tabelle 2.2:

Verschiedene Parametergruppen der TREC-Extrapolation bei AustroControl GmbH

kann.

Der letzte Parameter dient ausschließlich der Optimierung des Berechnungsvorganges. Aus Zeitgründen wird zur Berechnung des Testkriteriums nicht jeder Pixel herangezogen. Der *Suchalgorithmus* beschreibt die Auswahl der zu betrachtenden Bildpunkte.

Um den Ressourcenaufwand in Grenzen zu halten wurden nicht alle möglichen Parameterkombinationen berechnet. Es wurden vier möglichst unterschiedliche Sets zusammengestellt und routinemäßig berechnet. Diese sind:

- WIIS_ARAD_IMG_5min_b7_corr56_opt2_*.png
- WIIS_ARAD_IMG_5min_g5_corr16_opt2_*.png
- WIIS_ARAD_IMG_5min_g5_corr32_opt2_*.png
- WIIS_ARAD_IMG_15min_g5_sosq56_grid2_*.png

Die Bedeutung dieser Dateinamen ergibt sich aus Tabelle 2.2 und wird im ergänzend Anhang (7) detailliert behandelt.

In Abbildung 2.7 sind beispielhaft drei TREC-Extrapolationen geplottet. Sie beruhen auf den gleichen Parametereinstellungen. Der Vorhersageschritt variiert jedoch zwischen 10 und 60 Minuten.

Aus den bis hier vorgestellten vier TREC-Extrapolationen (mit unterschiedlichen Parametereinstellungen) wird desweiteren in einem von Kaltenböck et al. (2010b) entwickelten Verfahren ein „optimiertes Bild“ errechnet. Das Optimierungsschema soll an dieser Stelle nur kurz angeschnitten werden, für weiterführende Informationen sei auf die erwähnte Literatur verwiesen.

Die Idee ist, die Güte der einzelnen Extrapolationsvarianten in Echtzeit zu ermitteln, sobald gemessene Vergleichdaten zur Verfügung stehen. Es wird weiter angenommen, dass die Güteunterschiede zwischen den Extrapolationen (erstellt mit verschiedenen Parametereinstellungen), die hauptsächlich von der synoptischen Lage oder genauer gesagt von der Beschaffenheit der Niederschlagsregime abhängig sind, sich in „kurzer“ Zeit nur wenig verändern. Das heißt, jene Parameterkombination, die zum letzten Vorhersagezeitpunkt die besten Ergebnisse geliefert haben, sollte auch für die aktuell anstehende Extrapolation am Besten geeignet sein. Außerdem wird die Optimierung für drei geographische Bereiche (Nord-, Süd- und Ostösterreich), sowie für

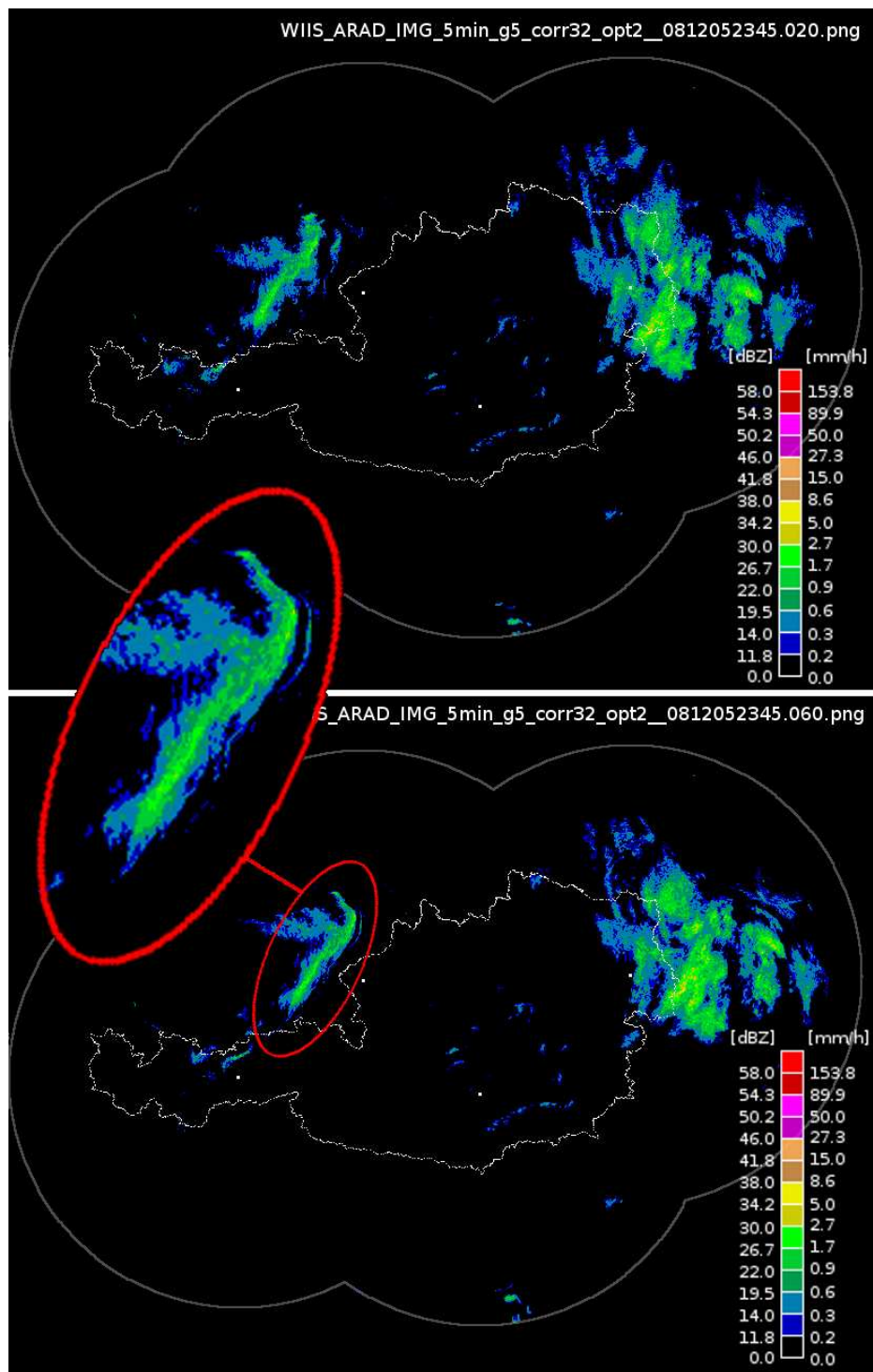


Abbildung 2.7:

Beispiele für TREC-Extrapolationen berechnet bei AustroControl GmbH: Die zwei gezeigten Extrapolationen vom 5. Dezember 2008, erstellt um 23:45h beruhen auf den selben Parametereinstellungen (5min_g5_corr32_opt2). Der Vorhersageschritt beträgt jedoch 20 Minuten im oberen und 60 Minuten im unteren Bild. In letzterem kann man bereits einige Konvergenz-, beziehungsweise Divergenzzonen entdecken (siehe Detail). Diese wären durch eine Erweiterung auf COTREC (siehe Abschnitt 2.1.1) in den Griff zu bekommen.

jeden Vorhersagezeitschritt einzeln errechnet. Auch diese optimierten Extrapolationen sollen im Rahmen dieser Arbeit mit Hilfe des zu erstellenden Prototyps genau verifiziert werden.

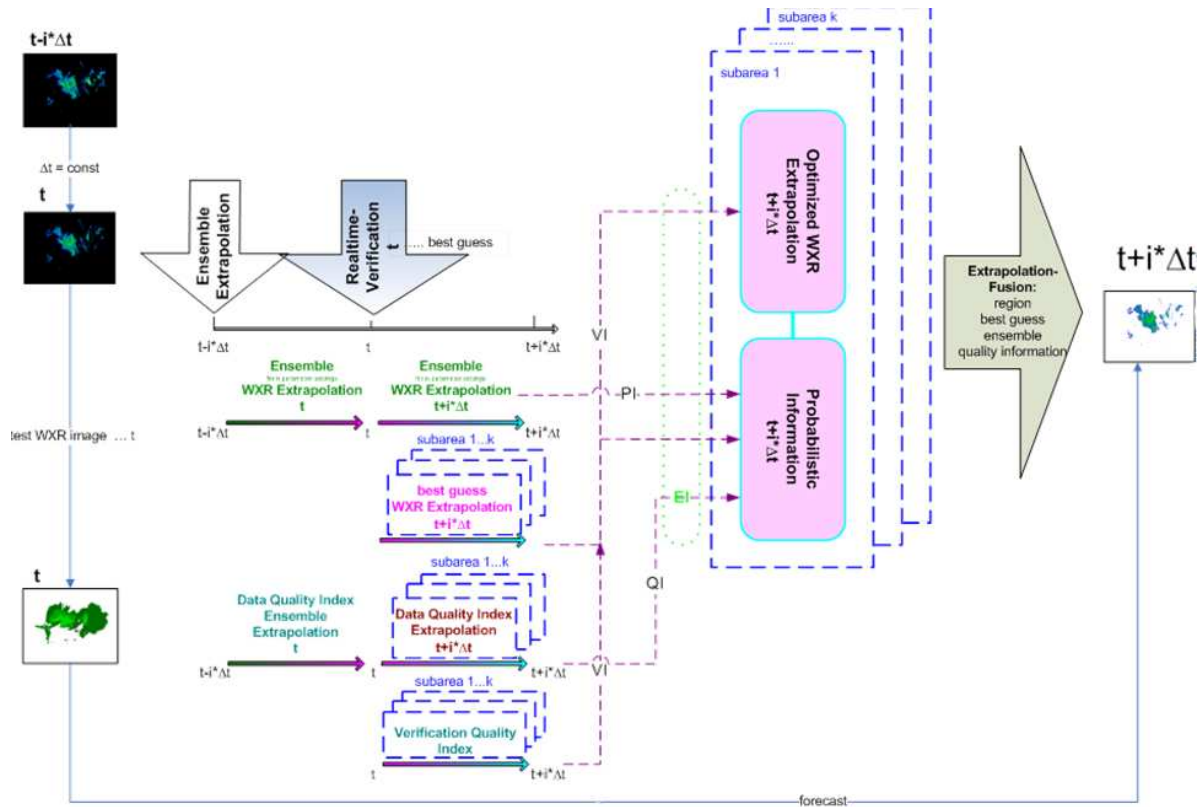


Abbildung 2.8:

Schematische Darstellung des Optimierungsschemas für WXR-Extrapolationen aus Kaltenböck et al. (2010b): In dieser Konfiguration werden TREC-Extrapolationen beim österreichischem Flugwetterdienst in Echtzeit optimiert. Die Methode beruht auf einer routinemäßigen Berechnung mehrerer Extrapolationen mit verschiedenen Parametereinstellungen (Ensemble-Extrapolation), sowie einer in Echtzeit ablaufenden Abschätzung der Güte jener Extrapolationen, die Vorhersagen für den aktuellen Zeitpunkt geliefert haben. Diese Abschätzung ist dann für die Optimierung der neuen Vorhersage maßgebend.

3 Anforderungen an die Verifikation

3.1 Zweck einer Verifikation

Verifikation ist der Prozess, der die Berechnung der *Qualität* (siehe Tabelle 1.1) einer Vorhersage zum Ziel hat (Wilks, 2006). Dabei wird die Vorhersage immer mit entsprechenden Messungen oder Beobachtungen verglichen. Die Verifikationsmethodik kann dabei qualitativen oder quantitativen Charakter besitzen (Atger et al., 2010). Im Folgenden soll nun erläutert werden, warum jedes Vorhersagesystem nur gemeinsam mit einer passenden Verifikation seine Aufgabe erfüllen kann.

Welchen Nutzen hat eine Prognose, wenn man nicht abschätzen kann, wie gut sie ist? Ein Meteorologe wird zur Erstellung seiner Vorhersage immer jene Informationsquellen heranziehen, die sich als zuverlässig und hochwertig erwiesen haben. Somit dient die Verifikation dem Vergleich mehrerer Vorhersagesysteme. Sie soll zeigen, wann (d.h. bei welcher synoptischen Lage) und wo (geographisch) welches Vorhersagesystem den anderen im Mittel überlegen ist. Desweiteren ist es wichtig, im Rahmen einer in Echtzeit ablaufenden Verifikation, die Vorhersagegüte zu überwachen, um einerseits genauere Aussagen über die aktuelle Güte machen zu können und andererseits rechtzeitig Situationen zu erkennen, in denen die Güte nachlässt. Außerdem ist ohne hinreichende Verifikation keine Verbesserung eines Vorhersagesystems möglich. Der erste Schritt, um etwas zu verbessern, muss es sein, zu erkennen wann und wo Fehler entstehen und wie sie beschaffen sind. (Atger et al., 2010)

Diese Ziele spiegeln sich auch in der von Brier und Allen (1951) erstmals vorgeschlagenen Unterteilung. Er unterscheidet drei Arten von Verifikation aufgrund ihrer Motivationen.

- administrativ
- wissenschaftlich (scientific)
- ökonomisch

Administrative Verifikation umfasst den Bereich der Echtzeitverifikation. Sie soll, wie schon beschrieben, den Vergleich der aktuellen Qualität mehrerer Vorhersagesysteme und das rechtzeitige Erkennen von Qualitätsschwankungen ermöglichen. Die systematische Verbesserung von Vorhersagemethoden, durch das Erkennen von individuellen Fehlern und ihrer Ursachen, liegt im Bereich der wissenschaftlichen Verifikation (Brier und Allen, 1951). Nur die ökonomisch motivierte Verifikation passt nicht ganz in das Konzept von Atger et al. (2010). Sie zielt darauf ab, den ökonomischen Wert einer Vorhersage als Entscheidungshilfe für den Kunden/Anwender zu erfassen. Somit nimmt sie eine besondere Stellung im “Randbereich” der Verifikation ein, weil durch statistische Verifikationsmethoden alleine, der Wert einer Vorhersage im Allgemeinen nicht bestimmt werden kann. Hierzu muss jeder Anwendungsfall speziell betrachtet werden (Murphy und Katz, 1997a).

3.2 Verifikation: Probleme und Anforderungen

Ausgehend von den in Abschnitt 1.2 definierten Anforderungen an eine WXR-Extrapolation (siehe 1.2) gilt es nun, die Anforderungen an das zu entwickelnde Verifikationsschema festzulegen. Ziel dieser Arbeit soll es sein, ein Verifikationsschema zu finden, das die *Qualität* einer WXR-Extrapolation erfassen kann. Wir beschränken uns somit auf die Bereiche der administrativen und wissenschaftlichen Verifikation.

Weiters ist es wichtig, einen Kompromiss zwischen Informationsgrad und Komplexität zu schließen. Es soll einerseits die Qualität der WXR-Extrapolation so gut als möglich beschrieben werden, andererseits soll das fertige Verifikationssystem, vor allem da es auch in Echtzeit (Überwachung der aktuellen Qualität) verwendbar sein soll, möglichst einfach aufgebaut sein. Besonderes Augenmerk liegt dabei auf einer übersichtlichen Darstellung der Ergebnisse, damit diese von fachkundigen Personen mit möglichst wenig Zeitaufwand interpretiert werden können.

3.2.1 Eigenschaften der WXR- Messung (Theorie und Probleme)

Ein Wetterradarbild spiegelt nur bedingt die Niederschlagsverteilung am Boden wider. Dies ist erstens durch die radiometrische Messung an sich bedingt und zweitens durch die Umrechnungsproblematik, die entsteht, wenn man versucht, aus der gemessenen Radarreflektivität eine meteorologisch verwendbare und vergleichbare Größe, wie die Niederschlagsrate zu berechnen (Schneider, 2008). Diese Umstände spielen

bei der Auswahl der Verifikationsmethodik eine entscheidende Rolle, in erster Linie, wenn es um die Auswahl der Vergleichsdaten geht. Daher werden sie im Rahmen der folgenden Punkte (3.2.1.1 und 3.2.1.2) etwas näher beschrieben, um danach eine Entscheidungsgrundlage für die Auswahl der zur Verifikation zulässigen Datenquellen zu haben.

3.2.1.1 Umrechnungsproblematik: Radarreflektivität in Niederschlagsrate

Um dieses Dilemma besser erläutern zu können, ist Theorie der Radarmessung von Nöten.

Die zentrale Beziehung in der Radarmeteorologie ist die Radargleichung. Sie beschreibt, wieviel der ausgesendeten Leistung zurückgestreut wird und somit empfangen werden kann. Ihre Herleitung wurde zum Beispiel von Hailzl (2006) ausführlich beschrieben, daher soll an dieser Stelle darauf verzichtet werden. Wir starten somit mit der Radargleichung für Volumensziele:

$$\bar{P}_r = \frac{P_t G^2 \lambda^2 \Theta^2 h_p}{512 \pi^2} \frac{1}{d^2} \sum_{vol} \sigma_i \quad (3.1)$$

(Das Formelwerk, in dieser Schreibweise wurde zum Großteil von Heuel (2004) übernommen.) Hierbei steht $\bar{P}_r$ für die Empfangsleistung, P_t für die Sendeleistung, G für den Antennengewinn (für nähere Informationen siehe Hailzl (2006)), λ für die Wellenlänge des Radarstrahls, Θ für den horizontalen Öffnungswinkel, h_p für die Pulslänge und d für die Entfernung des Volumenziels. Der letzte Faktor der Gleichung 3.1 beschreibt den Inhalt des Pulsvolumens.

$$\sum_{vol} \sigma_i = \eta = \frac{1}{V} \sum_i \sigma_i \left[\frac{m^2}{m^3} \right] \quad (3.2)$$

Die Rückstreuintensität (Radarreflektivität η , siehe Formel 3.2) ist die Summe der Rückstreuquerschnitte σ_i pro Einheitsvolumen V . Die Ausbreitung des Radarstrahls soll an dieser Stelle vereinfacht als Kegel angenommen werden, das Pulsvolumen hat also die Form eines Kegelstumpfes (siehe Abbildung 3.1). Der Puls beschreitet einen Hin- und einen Rückweg, er durchläuft somit jeden Punkt zweimal. Hat der Puls die Länge h_p , dann beträgt die Pulsvolumenslänge $\frac{h_p}{2}$ (Heuel, 2004). Somit ergibt sich das

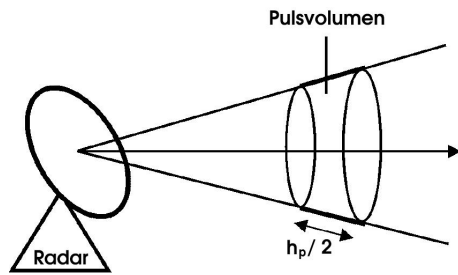


Abbildung 3.1:
Schematische Darstellung des vereinfachten Pulsvolumens (Heuel, 2004)

Pulsvolumen laut Formel 3.3:

$$V = \pi \left(d \frac{\Theta}{2}\right)^2 \frac{h_p}{2} \quad (3.3)$$

Damit Gleichung 3.1 gilt, wurde ein homogen ausgefülltes Pulsvolumen vorausgesetzt. Dies ist jedoch laut Heuel im Allgemeinen nicht realistisch. Bei größeren Pulsvolumina geht man daher von einem Intensitätsmaximum in der Mitte des Volumens und einer nach außen hin abnehmenden Intensität aus (Stichwort Gaußsche Glockenkurve). Dies soll nun auch in der Radargleichung für Volumensziele berücksichtigt werden.

$$\bar{P}_r = \frac{P_t G^2 \lambda^2 \Theta^2 h_p}{1024 \ln 2 \pi^2 d^2} \sum_{vol} \sigma_i \quad (3.4)$$

Die große Herausforderung liegt aber in der näheren Beschreibung der Radarreflektivität η . Sie hängt neben der Wellenlänge λ des Radarstrahls (und seiner Polarisation) maßgeblich von der Größenverteilung der Hydrometeore (und deren Form), sowie ihrem Aggregatzustand (der den komplexen Brechungsindex m bestimmt) ab (Hailzl, 2006; Blahak, 2004). Gleichung 3.5 zeigt die Abhängigkeiten der Radarreflektivität für kugelrunde Niederschlagspartikel:

$$\eta = \frac{\pi^5}{\lambda^4} |K|^2 \sum_{vol} D_i^6 [m^2] \quad mit \quad |K|^2 = \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \quad (3.5)$$

Diese Gleichung ist eine gute Näherung, solange die kugeläquivalenten Durchmesser der Hydrometeore kleiner sind als die Radarwellenlänge. Im C-Band Bereich (siehe 2.2.1) gilt dies bis zu einem Partikeldurchmesser von 3-5 mm (Hailzl, 2006). Wenn diese Voraussetzung erfüllt ist, kann man die Streuprozesse an den Hydrometeoren anhand der „Rayleigh-Approximation“ (reine Dipolstreuung) beschreiben. Sie liegt Gleichung 3.5 zu Grunde. Bei größeren Tropfen müsste die „Mie-Streuung“ ver-

wendet werden und bei Objekten, deren Durchmesser viel größer als die verwendete Radarwellenlänge ist, (im C-Band bei Durchmessern über 50 cm) gelten die optischen Reflexionsgesetze. Die Rückstrahlquerschnitte sind in diesem Bereich nicht mehr von der Wellenlänge abhängig (Hailzl, 2006). Da die Durchmesser von Niederschlagstropfen in der Regel kleiner als 5 mm sind, wollen wir uns in diesem Rahmen ausschließlich auf diesen Bereich konzentrieren, indem die „Rayleigh-Streuung“ angewendet werden darf.

In Gleichung 3.5 sieht man, dass die Radarreflektivität von der sechsten Potenz des Tropfendurchmessers D abhängt. Daher liefert zum Beispiel ein einziger Tropfen mit einem Durchmesser von 5 mm (und einer Masse von zirka 0,065 mg) genauso viel Rückstreuung wie 15625 Tropfen mit 1 mm (und einer insgesamten Masse von 8,18 mg) oder 1 Mio. Tropfen mit 0.5 mm Durchmesser (diese hätte eine Masse von 65,4 mg). Zum Beispiel zeigten Randeu et al. (2002) mit Distrometeruntersuchungen, dass sehr große Tropfen (Durchmesser bis zu 10 mm) zu einer signifikanten Niederschlagsüberschätzung in der Radarmessung führen. (Im Extremfall um den 4-fachen Wert).

$|K^2|$ wird als der dielektrische Faktor der Niederschlagssubstanz bezeichnet. Diese Materialkonstante spiegelt die Abhängigkeit des Rückstreuquerschnitts vom Aggregatzustand wider. Bei Wasser beträgt $|K^2| = 0,928$, für Eis hingegen gilt $|K^2| = 0,197$ (Hailzl, 2006). Somit sinkt der Rückstreuquerschnitt bei festem Niederschlag (Hagel, Schnee, Graupel, etc.) auf zirka $\frac{1}{5}$. Niederschlagszonen mit festem Niederschlag werden daher mit viel geringerer Intensität angezeigt, als jene mit flüssigem. Andererseits kommt es in der Schmelzzone unterhalb der Nullgradgrenze zum gegenteiligen Effekt (siehe Abbildung 3.2). Dort kann man eine Signalverstärkung beobachten, weil die zu Boden fallenden, festen Hydrometeore beim Anschmelzen rasch von einem Wasserfilm überzogen werden und somit den Rückstreuquerschnitt eines flüssigen Tropfens gleicher Größe erhalten (Hailzl, 2006).

Setzt man nun Gleichung 3.5 in unsere Radargleichung für Volumensziele (Gleichung 3.4) ein, so erhalten wir unsere endgültige Radargleichung für Volumensziele:

$$\bar{P}_r = \frac{\pi^3 P_t G^2 \Theta^2 h_p}{1024 \ln 2 \lambda^2} |K|^2 \frac{1}{d^2} \sum_{vol} D_i^6 \quad (3.6)$$

Der wichtigste Faktor in dieser Gleichung ist gleichzeitig der letzte. Wie bereits erwähnt, spielt hier die sechste Potenz der Partikeldurchmesser die entscheidende Rol-

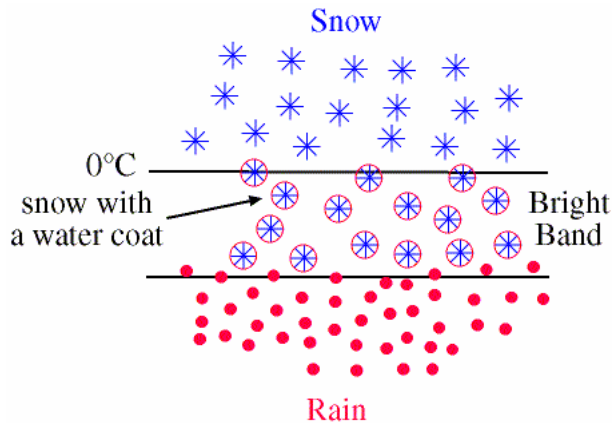


Abbildung 3.2:

Brightband-Effekt (Hailzl, 2006): Der Wasserfilm, der die festen Hydrometeore in der Schmelzschicht überzieht, führt zu einer eklatanten Signalverstärkung.

le, sie beeinflusst die empfangbare Leistung maßgebend. Aufgrund dieser Bedeutung wurde eine neue Größe definiert, der Radarreflektivitätsfaktor Z :

$$Z = \frac{1}{V} \sum_i D_i^6 = \int_0^\infty N(D) D^6 dD \left[\frac{mm^6}{m^3} \right] \quad (3.7)$$

Z (Formel 3.7) beschreibt also die Rückstreuung der Hydrometeore mit der Regentropfenverteilung $N(D)$ im Einheitsvolumen V (Wuest, 2005). Die wahre Reflektivität η kann nun folgendermaßen, mit Hilfe des Reflektivitätsfaktors ausgedrückt werden (siehe Formel 3.8). Es sei an dieser Stelle darauf hingewiesen, dass Z in der Literatur oft einfachheitshalber als Reflektivität bezeichnet wird.

$$\eta = \frac{\pi^5}{\lambda^4} |K|^2 Z \quad (3.8)$$

Um aber quantitative Aussagen über Niederschlagsereignisse machen zu können, ist es nun notwendig, diese Messgröße in eine meteorologische Größe, nämlich die Regenrate umzurechnen. Die Regenrate ist ein Maß für die Niederschlagsmenge pro Zeiteinheit. Sie wird meist in mm/h angegeben. Das ist nun, das zu Beginn dieses Unterkapitels angesprochene Umrechnungsproblem und auch eine der zentralen Herausforderungen in der Radarmeteorologie. Allgemein berechnet sich die Regenrate laut Formel 3.9, wobei $N(D)$ die Anzahl der Tropfen mit Durchmesser D pro m^3 pro mm (die Einheit von $N(D)$ ist also $m^{-3}mm^{-1}$) und $v(D)$ die dazugehörigen End-

fallgeschwindigkeiten (die natürlich wiederum vom jeweiligen Durchmesser abhängig sind) der Tropfen bezeichnet.

$$R = \frac{\pi}{6} \int_0^{\infty} v(D) D^3 N(D) dD \left[\frac{mm}{h} \right] \quad (3.9)$$

Die Beziehung zwischen dem Radarreflektivitätsfaktor (Gleichung 3.7) und der Niederschlagsrate (Gleichung 3.9), die sogenannte *Z/R-Korrelation*, ist also stark vom Tropfenspektrum abhängig. Der Reflektivitätsfaktor Z beinhaltet eine Abhängigkeit von der sechsten Potenz der Tropfendurchmesser, für die Regenrate ist andererseits das Volumen der Tropfen maßgebend, womit die Durchmesser der Tropfen hier mit der dritten Potenz in die Berechnung eingehen. Allein aufgrund dieser Tatsache kann es bei unterschiedlichen Niederschlagsereignissen, mit gleicher Radarreflektivität zu sehr unterschiedlichen Niederschlagsmengen kommen. (Messtechnische Probleme, die dabei ebenfalls eine große Rolle spielen, sollen anschließend diskutiert werden.) Daher wäre für eine exakte Umrechnung einer Radarreflektivität in die entsprechende Regenrate die Kenntnis der genauen Tropfenverteilung $N(D)$ von Nöten. Dieses Tropfenspektrum ist aber stark von den synoptischen Bedingungen, der Jahreszeit, dem Niederschlagstyp, sowie den lokalen Bedingungen abhängig. Auch während eines Niederschlagsereignisses bleibt es nicht konstant (Heuel, 2004). Es ist daher in der Praxis unbekannt und muss geschätzt werden. Welche Größen sind also nötig um eine Z/R -Beziehung herzuleiten? Laut Gleichung 3.7 für den Radarreflektivitätsfaktor, sowie Gleichung 3.9 für die Niederschlagsrate sind dafür das Tropfenspektrum $N(D)$, sowie die Endfallgeschwindigkeiten der Hydrometeore $v(D)$ notwendig.

Laut Heuel (2004) kann man mit Hilfe der empirischen Annahme, dass eine ausreichend lange Tropfen-Falldauer zu einem Gleichgewichtsspektrum führt, dieses mittels einer Exponentialfunktion darstellen. (Dies wurde unter anderem von Ohtake (1970) gezeigt.)

$$N(D) = N(0) \exp(-\Lambda D) \quad (3.10)$$

Der bekannteste Ansatz für Regentropfen ($D \geq 0,5 \text{ mm}$) stammt von Marshall und Palmer (1948). Dabei wird die spektrale Ansatzdichte $N(0) = 8000 \text{ m}^{-3} \text{ mm}^{-1}$ und $\Lambda = 4,1 R^{-0,21} (\frac{h}{mm})^{-0,21} \text{ mm}^{-1}$ gesetzt. R steht dabei für die Regenrate (einzusetzen in mm/h). Die Tropfenverteilung hängt also bei diesem Ansatz von der Regenrate ab (siehe Abbildung 3.3). Diese, schon relativ alte Parametrisierung baut auf sommer-

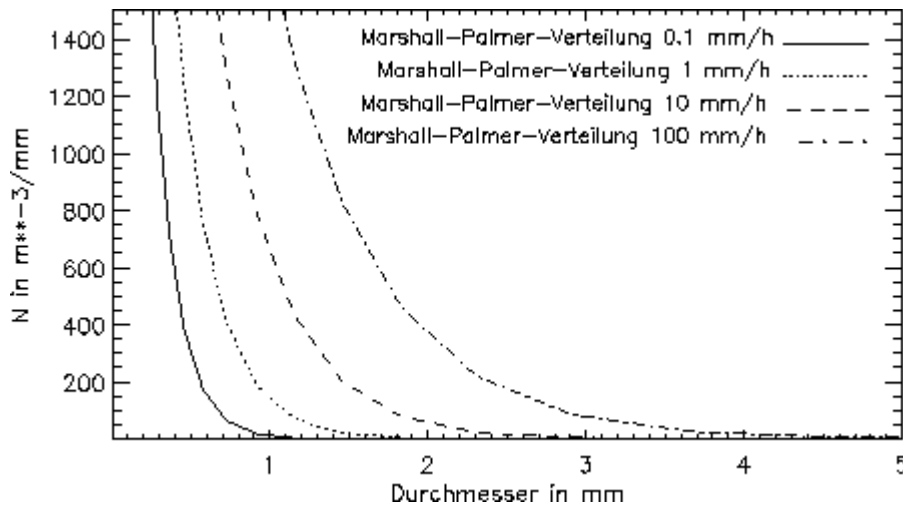


Abbildung 3.3:

Exponentialverteilung nach Marshall und Palmer (1948): Die verschiedenen Kurven zeigen die Abhängigkeit des Tropfenspektrums von der Regenrate.

liche Regenereignisse in Ottawa 1948 auf. Marshall und Palmer maßen dort relativ viele kleine Tropfen und wenige große. Daher wurden später verschiedene, alternative Parametrierungen entwickelt, die meist auf die dreiparametrische Gamma-Verteilung aufbauen und versuchen die oftmalige Überschätzung der Anzahl sehr kleiner Tropfen (durch die exponentielle Beschreibung nach Marshall und Palmer) zu berücksichtigen. Für nähere Informationen zu diesen Ansätzen sei auf einschlägige Literatur, zum Beispiel Heuel (2004) oder Grimbacher (2002), sowie Blahak (2004) verwiesen.

Zur Berechnung der Niederschlagsrate (Gleichung 3.9) benötigt man zusätzlich noch $v(D)$, die Endfallgeschwindigkeiten der Hydrometeore. Diese kann laut Hailzl (2006) als semi-empirisches Potenzgesetz folgender Form angenähert werden:

$$v(D) = a \left(\frac{D}{D_r} \right)^b \quad (3.11)$$

D_r bezeichnet den Referenzdurchmesser ($D_r = 1 \text{ cm}$). Für a und b findet man in der Literatur verschiedenste Werte. Einige gängige Werte, übernommen aus Hailzl (2006) finden sich in Tabelle 3.1 wieder.

Kennt man nun das Tropfenspektrum, sowie die Endfallgeschwindigkeiten der Hydrometeore kann man nun eine ZR-Beziehung berechnen. Aufgrund der großen Variabilität des Tropfenspektrums, durch die oben genannten Einflussfaktoren, kann es jedoch keine allgemein gültige ZR-Beziehung geben. In der Literatur finden sich

nach	a [$\frac{cm}{s}$]	b
Liu und Orville (1969)	2115	0,8
Kessler (1969)	1300	0,5
Atlas und Ulbrich (1977)	1767	0,67

Tabelle 3.1:

Potenzgesetzparameter zur Bestimmung der Tropfen-Endgeschwindigkeit (Hailzl, 2006)

Beziehung	Autor(en)
$Z = 206 R^{1,55}$	Kessler
$Z = 282 R^{1,45}$	Liu und Orville
$Z = 238 R^{1,5}$	Atlas et al.
$Z = 296 R^{1,47}$	Marshall und Palmer
$Z = 200 R^{1,6}$	Marshall und Palmer (Standardbeziehung)

Tabelle 3.2:

Gängige ZR-Beziehungen aus Hailzl (2006)

unzählige verschiedene Ausführungen, zum Beispiel listet Battan (1973) ganze 69 Beziehungen für verschiedene Situationen von unterschiedlichen Autoren auf. Sie alle sind aber von folgender Gestalt:

$$Z = a R^b \quad (3.12)$$

Ein paar der gängigsten ZR-Beziehungen, ausgewählt laut Hailzl (2006) sind in Tabelle 3.2 aufgelistet. Beim österreichischen Flugwetterdienst wird die sogenannte Standardbeziehung von Marshall und Palmer verwendet, die laut Hailzl genau genommen nur für stratiformen Niederschlag in mittleren Breiten gilt.

Dieses Kapitel soll zeigen, dass die Umrechnung von Radarreflektivitäten in Niederschlagsraten immer nur eine Näherung sein kann. Dieser Aspekt muss auf jeden Fall beachtet werden, wenn man versucht, Wetterradarextrapolationen gegen Bodenniederschlagsmessungen zu verifizieren.

Ergänzend sei an dieser Stelle angemerkt, dass diese Bodenniederschlagsmessungen natürlich nicht flächig vorliegen und außerdem der Vergleich meist auf 6- oder sogar 12-stündige Niederschlagssummen aufgesetzt werden müsste. Ihre Zeitskala ist somit für die Verifikation eines Nowcasting-Systems denkbar ungeeignet.

3.2.1.2 Messfehler Wetterradar

Des weiteren muss man in diesem Zusammenhang beachten, dass die Radarmessung, wie alle physikalischen Messverfahren, Fehlern unterworfen ist. Diese Fehler beeinflussen jeden Radarscan und daher auch jede Extrapolation, die auf diesen WXR-Daten aufbaut. Es wäre daher wünschenswert, eine Verifikationsmethodik zu finden, mit der es (zumindest qualitativ) möglich ist, Fehler in extrapolierten WXR-Bildern, die aufgrund mangelhafter Eingangsdaten entstanden sind, von Fehlern, die ihre Quelle in der Extrapolationsmethodik haben, unterscheiden zu können. Zumindest aber muss einem bewusst sein, dass eine Vorhersage nur so gut sein kann, wie die Daten auf denen sie beruht. Dass heißt, je höher die Qualität der Eingangsdaten, desto höher wird auch jene der Extrapolation sein. Aufgrund ihrer Bedeutung, vor allem wenn es um die Interpretation von Verifikationsergebnissen geht, folgt nun eine kleine Auswahl der gravierendsten WXR-Messfehler. Diese Liste erhebt keinen Anspruch auf Vollständigkeit. Die Auswahl beruht zum Teil auf einem unveröffentlichten, internen Projektbericht zum selben Thema, den der Autor im August 2009 zusammen mit K. Hennermann, R. Kaltenböck und M. Kerschbaum verfasst hat. Alle WXR-Bilder stammen aus dem Archiv der AustroControl GmbH (ACG).

Man kann grundsätzlich zwei Arten von Fehlern hinsichtlich ihrer Erscheinung, unterscheiden:

Störsignale und nichtmeteorologische Echos lassen Niederschlag in Gebieten vermuten, in denen keiner vorhanden ist, beziehungsweise führen in diesen Bereichen zu einer überschätzten Darstellung eines Niederschlagsfeldes. In der zweiten Fehlergruppe kommt es zum gegenteiligen Effekt, sodass in Folge von Abschattungen (durch Hindernisse wie Bodenerhebungen, u.ä.) und dämpfenden Effekten Niederschlagsfelder unterschätzt, oder erst gar nicht erfasst werden. Es entstehen *Lücken* im Erfassungsbereich der Radarstationen.

Als Erstes soll die Gruppe der Störsignale und nichtmeteorologischen Echos betrachtet werden:

3.2.1.2.1 Festechos

Als Festechos oder auch Clutter Echos bezeichnet man Echos, die nicht von Niederschlägen, sondern von anderen, den Radarstrahl reflektierenden Objekten stammen. Man kann sie weiter unterscheiden in:

- *Bodenechos* (Ground Clutter): Sie entstehen durch Festziele am Boden (Berge,

Gebäude, etc.) und stellen daher vor allem im Gebirge ein Problem dar. Sie treten in erster Linie bei Messungen in niedrigen Elevationen (für Einzelheiten zur Radarmessung an sich siehe 2.2.1) des Radarstrahls auf. Um sie von Niederschlag zu unterscheiden, kann man den zeitlichen Verlauf betrachten, denn clutterbehaftete Pixel fluktuieren zeitlich sehr stark. Um Bodenechos zu entfernen ist bei den von AustroControl in Österreich betriebenen Wetterradarstationen (siehe wieder 2.2.1) bereits ein Algorithmus (Clutter-Filter) implementiert (Hailzl, 2006).

- *Clear Air Echos*: Auch aus klarer Luft, das heißt aus Luftmassen ohne Bewölkung und Niederschlag, wird zuweilen Radarstrahlung zurückgestreut und kann somit auch detektiert werden. Dies kann durch Insekten- oder Vogelschwärme geschehen, aber auch durch kleine Staubpartikel oder durch Luftturbulenzen. Solche Echos werden als „Clear Air Echos“ bezeichnet. Sie treten meist an schönen, warmen Sommertagen in den untersten 1 bis 2 km der Atmosphäre auf (Gysi, 2010). Ihre Behandlung an dieser Stelle dient in erster Linie der Vollständigkeit. Ihr Störpotential fällt gegenüber anderen Störungsarten eher gering aus.
- *Brightband*: Dieser Effekt wurde bereits im Abschnitt 3.2.1.1 erläutert, daher wird an dieser Stelle nur auf seine Auswirkungen eingegangen. In einer MAXCAPPI-Visualisierung (siehe Abschnitt 2.2.1) kommt es durch diesen Effekt zu einer (ungefähr kreisrunden) Zone sehr hoher Reflektivität im Bereich der Radarstationen, die man nicht als Starkniederschlag interpretieren darf (Siehe Abbildung 3.5). Das Brightband ist bei stratiformen Niederschlägen, durch ihre größere horizontale Ausdehnung und geringere Reflektivität, offensichtlicher als bei konvektiven.
- *Flugzeugechos*: Bewegliche Festziele werden von herkömmlichen Clutter Filtern nicht erfasst. Flugzeuge erscheinen im Radarbild als signifikante Echos der Intensitätsstufen 1-4 (siehe 2.2.1). Aufgrund der Strahlbreiteverzerrung erscheinen sie im Seitenriss oft einige Kilometer hoch und wenn ein Flugzeug mehrmals vom Radarstrahl erfasst wird, entstehen in den Bildern mehr oder weniger miteinander verbundene Pixelketten. Ähnlich den „clear air echos“ kann diese Störungsart als eher nebensächlich betrachtet werden.

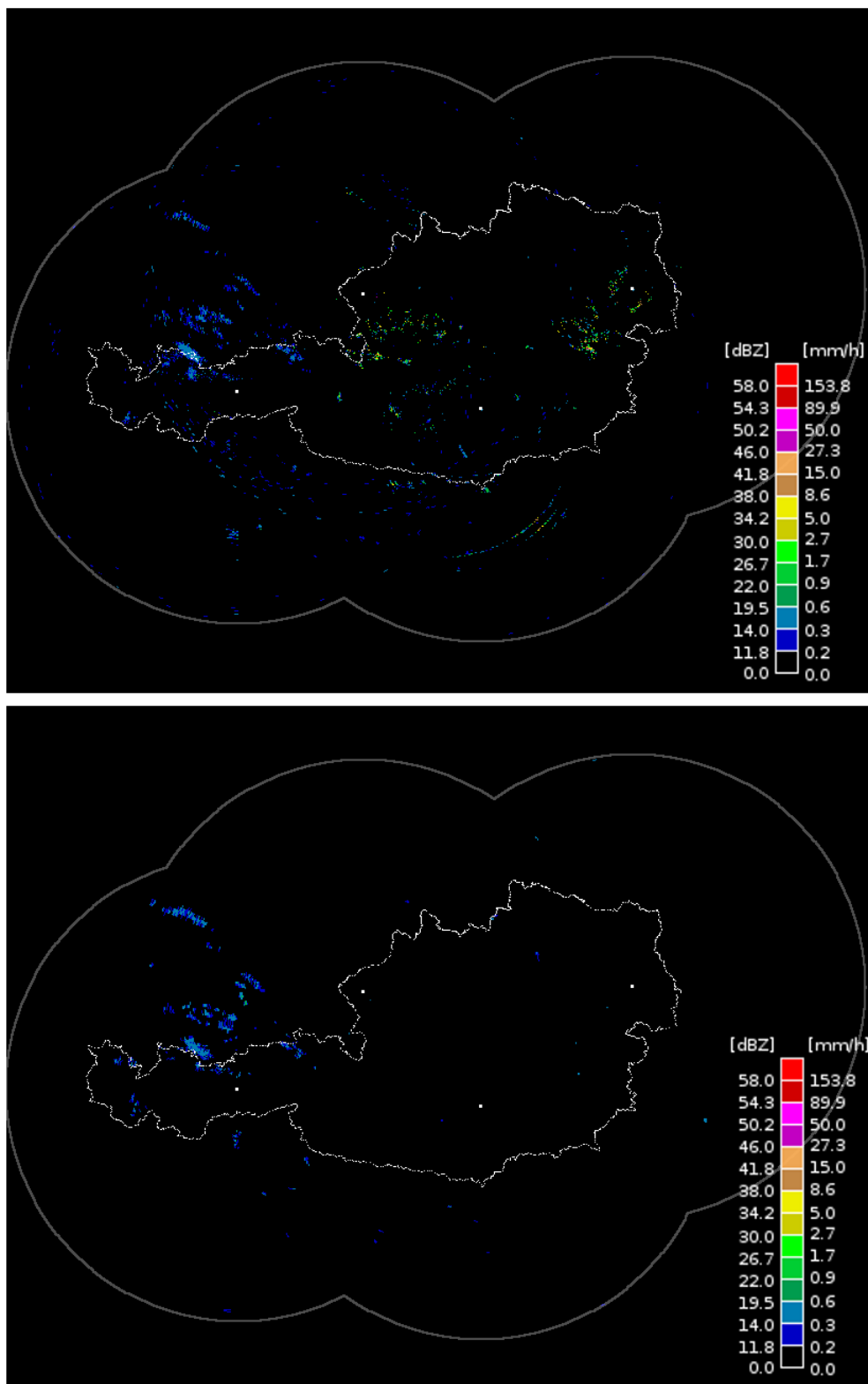


Abbildung 3.4:

Wirkung des Ground-Clutter-Filters: Um 00:00 UTC wird der Clutter-Filter zur Neu-Initialisierung bei allen vier WXR-Stationen (des Austria-Composite) ausgeschaltet (oberes WXR-Bild). Um 00:15 (unteres WXR-Bild) ist er wieder in Betrieb.

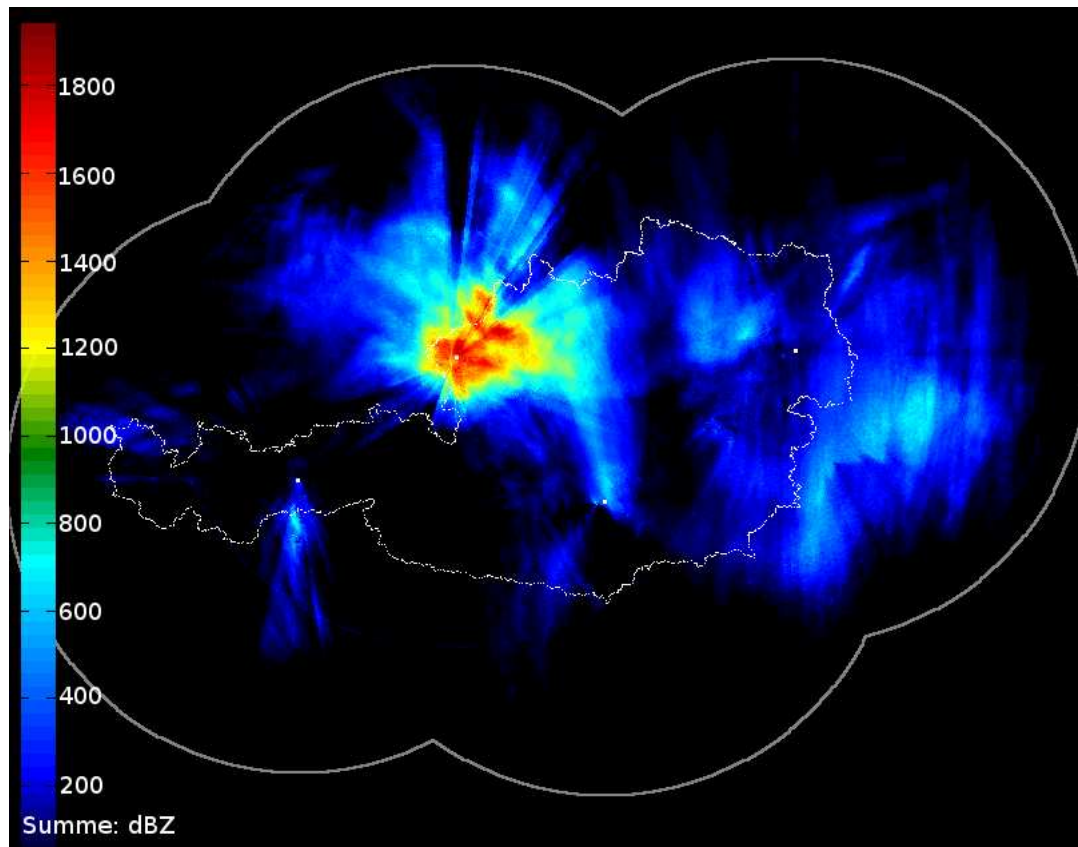


Abbildung 3.5:

Der Brightband-Effekt im Bereich der WXR-Station Salzburg/Feldkirchen wird durch das Aufsummieren mehrerer WXR-Austria-Composite-Bilder deutlicher sichtbar. Hier wurde nach Kaltenböck et al. (2010a) die Tagessumme vom 17. Dezember 2008 berechnet.

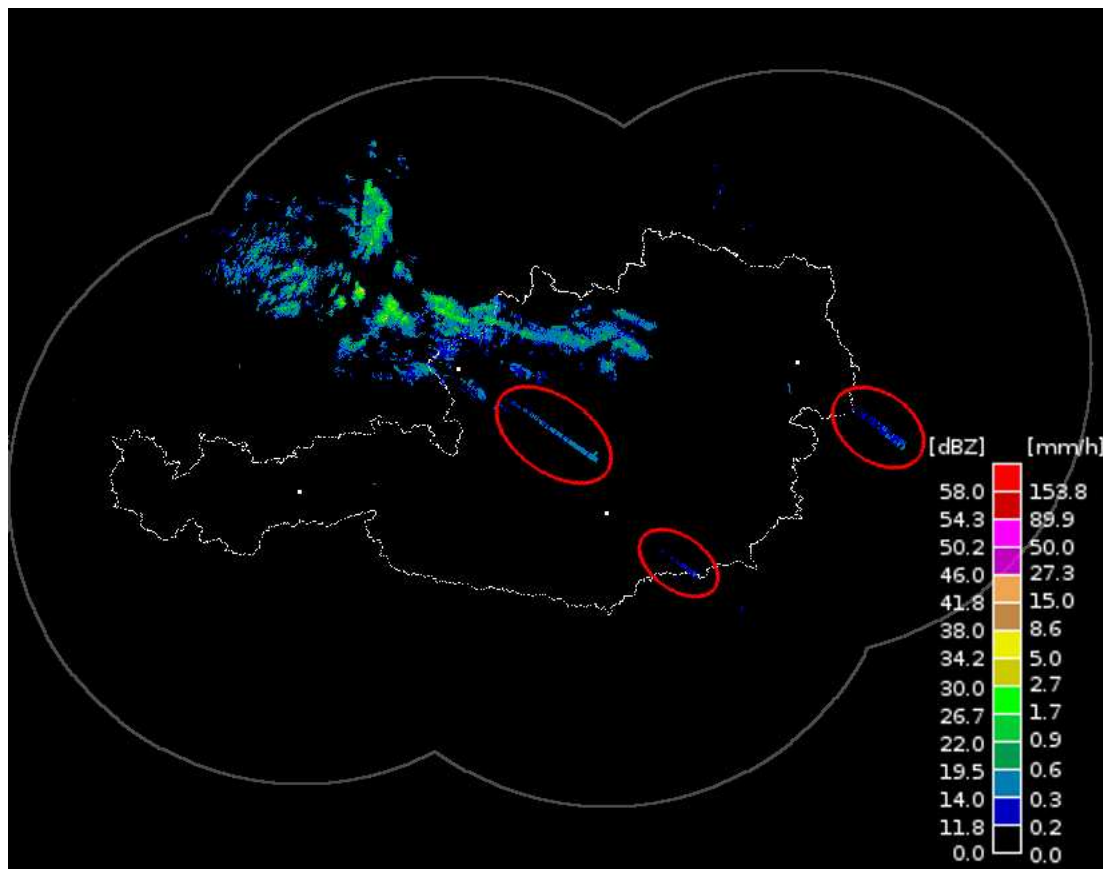


Abbildung 3.6:

Sonnenechos, die beim Sonnenaufgang entstanden sind. Die Sonne steht hier also im Südosten der Radaranlage Salzburg/Feldkirchen, zu der die Fehlerstreifen zeigen. (Kaltenböck et al., 2010a)

- *Sonnenechos*: Befindet sich der Ausbreitungsweg des Sonnenlichts genau in der Hauptrichtung des Radarstrahls, kommt es ab einer gewissen Stärke der Sonneneinstrahlung zu elektromagnetischen Interferenzen (da auch die Sonne im Wellenlängenbereich des Radars elektromagnetische Wellen aussendet) und somit zu verstärktem Rauschen im Radarempfänger und in weiterer Folge zu Störungen im Radarbild (Hailzl, 2006; Gysi, 2010). Dieser Effekt tritt bei sehr niedrigem Sonnenstand, also vorrangig im Winter am Morgen und am Abend auf. Die entstehenden Fehler sind schmale Streifen, mit konstantem Azimutwinkel, die zu einer Radarstation zeigen (Siehe Abbildung 3.6).
- *Echos durch dielektrische Inhomogenitäten*: Diese Echos entstehen an scharfen und räumlich begrenzten Sprüngen des Brechungsindex. Sie treten z.B. an

Thermikblasen, zwischen Wolke und Umgebungsluft oder bei Inversionen im vertikalen Temperaturprofil mit starkem Feuchtegradienten auf (Hailzl, 2006).

- *Second Trip Echos*: Ein Radar hat eine eindeutige maximale Reichweite, die durch die Impulswiederholungsfrequenz gegeben ist. Für Echos, die im Intervall bis zum nächsten Impuls zurückkehren, kann die Entfernung eindeutig bestimmt werden. Echos aus Entfernungen außerhalb der maximalen Reichweite kehren erst nach dem nächsten Impuls zurück, werden dann diesem zugeordnet und erscheinen daher in viel geringerer Entfernung. Dies ist zum Beispiel bei sehr hoch reichenden, starken Gewittern außerhalb der maximalen Reichweite problematisch. Wenn das Radar mit einer von Puls zu Puls wechselnden Pulswiederholungsfrequenz (staggered PRF (Pulse Repetition Frequency)) arbeitet, so wird die (fehlerhafte) Entfernungszuordnung der Second Trip Echos abwechselnd an verschiedenen Stellen geschehen. Somit könnte man die Second Trip Echos anhand ihres radialen Streifenmusters erkennen (Gysi, 2010). Dieses Verfahren wird in Österreich allerdings (außer beim WXR auf der Valluga, das aber nicht in das WXR-Austria-Composite eingebunden ist) nicht verwendet. Es besteht nur die Möglichkeit zwischen 450Hz und 600Hz umzuschalten.
- *Seitenkeulenechos*: Die vom Radar ausgestrahlte Energie entlang des Hauptstrahls (Hauptkeule) hat eine typische Strahlbreite von 1° . Aufgrund der Antennen-Charakteristik, kommt es aber auch entlang der Nebenkeulen (siehe Abbildung 3.7), die um ein paar Grad zum Hauptstrahl verschoben sind, zu geringem Energieausstoß. Daraus können vor allem bei sehr stark reflektierenden Zellen, die sich nahe der Station befinden, zusätzliche Echos resultieren.
- *Kreisförmige / Spiralförmige Störung*: Oft sind auf Radarbildern kreisbogen-/spiralförmige Störungen zu erkennen (siehe Abbildung 3.8), die oft mit der Grenze des Nichterfassungsbereichs zusammenfallen oder radial nach außen verlaufen, deren Orientierung manchmal jedoch auch völlig willkürlich erscheint. Die Ursachen für diese Effekte dürften Nebenkeuleneffekte und die Wechselwirkung mit anderen Radarstationen sein. Im Detail ist dies jedoch noch unklar.
- *Störungen durch RLANs*: Einige lokale Funknetzwerke, RLANs (Radio Local Area Network) arbeiten mit demselben Frequenzbereich (5250-5725GHz), in dem auch die österreichischen Wetterradarstationen (5,625GHz) senden und

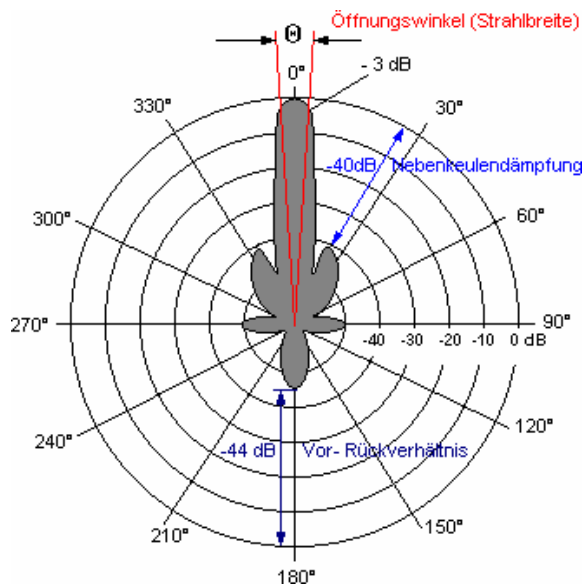


Abbildung 3.7:
Richt- oder Antennendiagramm (Hailzl, 2006).

empfangen. Dies führt zu Störungen, die sich als Streifen mit Orientierung zur Radarstation äußern. Die Streifenbreite kann dabei variieren. Weiters wurde beobachtet, dass die Streifen meist nicht ganz bis zur Station reichen (Siehe Abbildung 3.9).

- *Tiefflugerfassungsradar*: Es wurden Störungen beobachtet, die durch ihre Erscheinungsform keiner anderen hier aufgeführten Störungsart zugeordnet werden können. Diese kommen durch militärisch genutzte Tiefflugradars zu Stande. Diese Störung tritt naturgemäß eher selten auf und ihre Form ist stark von der Entfernung zur WXR Station abhängig (siehe Abbildung 3.10). In diesem Zusammenhang sei auch Düppel (Chaff) erwähnt. Das ist ein militärisches System der elektronischen Kampfführung. Es werden viele kleine, stark reflektierende Partikel (Aluminiumfolienstreifen oder metallisch bedampfte Glasfasern) in der Atmosphäre freigesetzt, diese verteilen sich zu ausgedehnten Wolken, stark reflektierenden Materials und schweben langsam in Richtung Boden. Diese Störform soll die Ortung von Kampfflugzeugen erschweren und verursacht naturgemäß massive Radarstörungen (Wolff, 2010). Aufgrund ihrer Seltenheit sind diese Störungsarten im Vergleich von eher geringer Bedeutung.

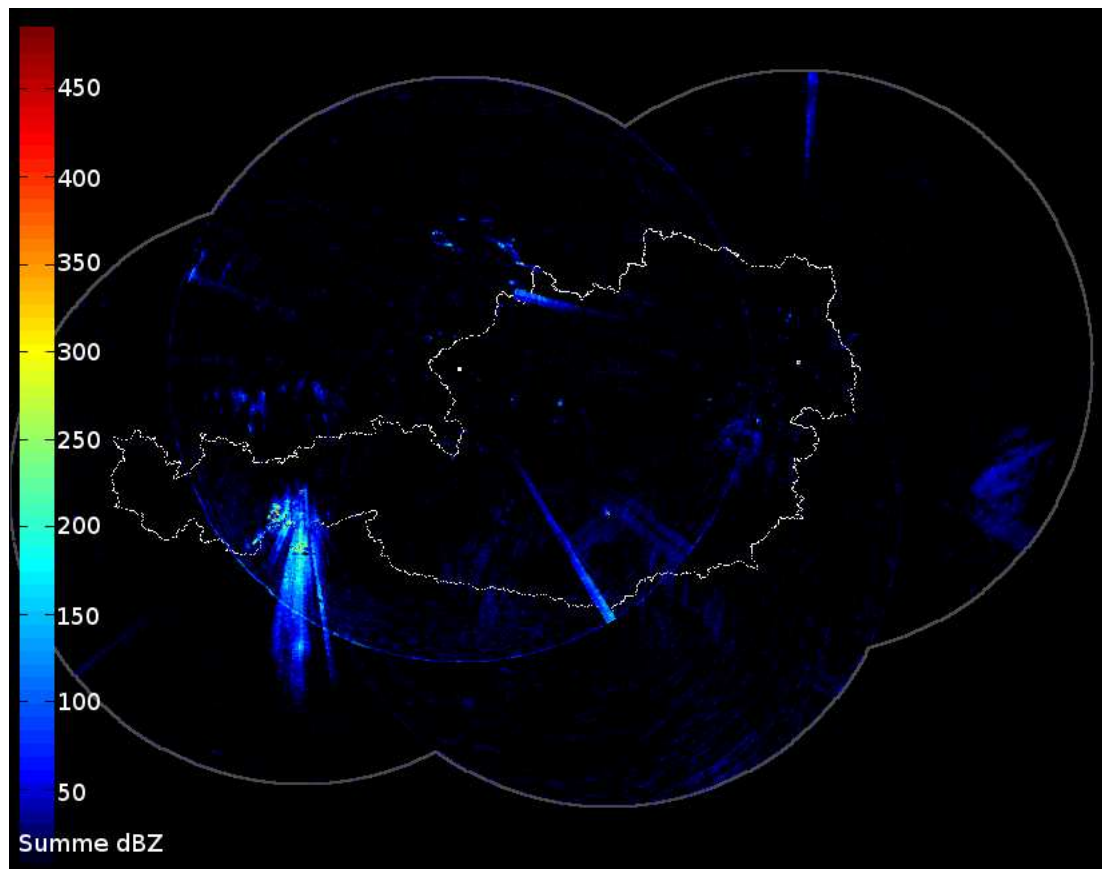


Abbildung 3.8:

Kreisförmige und spiralförmige Störungen im Tagesmittel: Durch das Summieren von Radarbildern über mindestens einige Stunden werden kreis- und spiralförmige Störungen noch deutlicher sichtbar. Hier wurden alle WXR-Austria-Composite Bilder vom 14. Dezember 2008 aufsummiert.

3.2.1.2.2 Erfassungslücken

Das zweite, aus meteorologischer Sicht noch größere Problem sind meteorologische Echos, die vom Radar nicht erfasst werden. Die folgenden Effekte verändern das WXR-Bild maßgeblich, sind aber oft nicht eindeutig identifizierbar. Durch summierte WXR-Bilder ist es am ehesten möglich, ihre Auswirkungen zumindest qualitativ zu zeigen.

- *Orographische Abschattungen*: Diese Form der Abschattung tritt ein, sobald kein direkter Sichtkontakt zwischen der Radarstation und dem Niederschlagsfeld gegeben ist, oder die Sicht zumindest teilweise durch orographische Hindernisse gestört ist. Vor allem im Bereich der Alpen kommt es in Österreich zu massiven Abschattungsgebieten. Das Radar am Patscherkofel bei Innsbruck kann

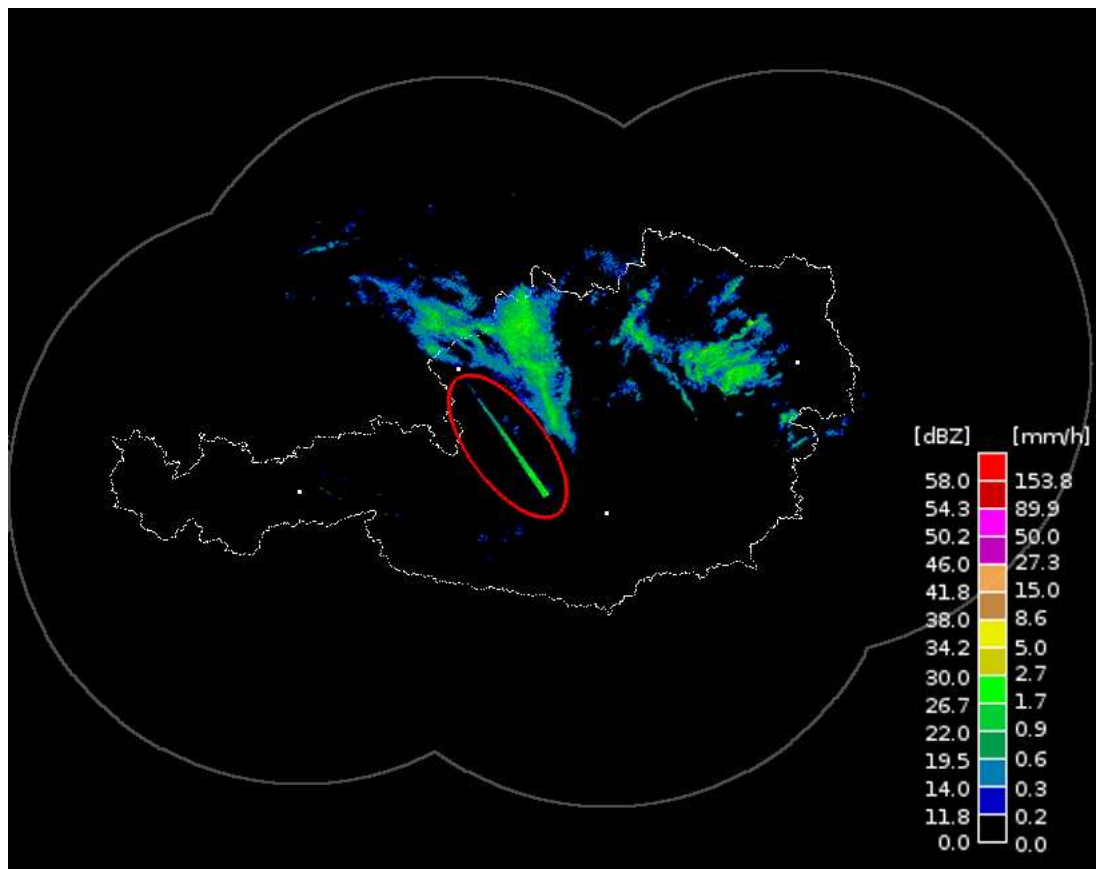


Abbildung 3.9:

RLAN-Störung des WXR Salzburg/Feldkirchen (Kaltenböck et al., 2010a)

zum Beispiel kaum Information über Niederschlag in Vorarlberg liefern. Zur Hilfestellung für den Meteorologen wurden Abschattungskarten für die einzelnen Radarstationen erstellt. Es ist jedoch zu bedenken, dass die Abschattungsgebiete sich mit der Elevation des Radarstrahls ändern. Die Höhe der Bewölkung spielt somit die entscheidende Rolle, ob ein Niederschlagsfeld weitgehend erfasst werden kann, man nur die obersten (eventuell nicht repräsentativen) Schichten der niederschlagsträchtigen Bewölkung sieht, oder gar kein Niederschlag gemessen werden kann. Gerade inneralpin ist dieses Problem als das dominierende anzusehen und verursacht massive Probleme aus meteorologischer Sicht, sowie im Zuge der Verifikation gegen Niederschlagsmessungen.

- *Ausbreitungsdämpfung (Niederschlagsdämpfung)*: Im Wellenlängenbereich der österreichischen WXR-Anlagen spielt nur die Ausbreitungsdämpfung im Nie-

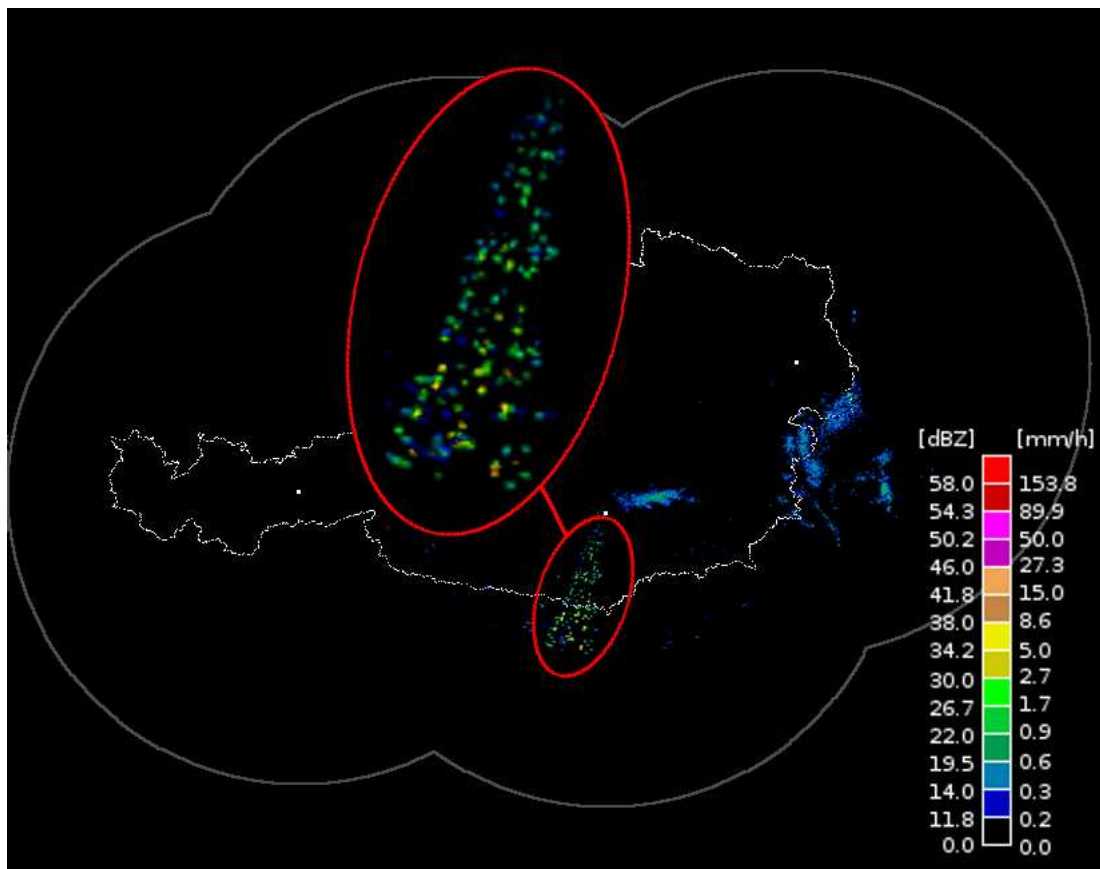


Abbildung 3.10:
WXR-Störung durch militärisches Tiefflugerfassungsradar (Kaltenböck et al., 2010a)

derschlag eine entscheidende Rolle. Das Radarsignal wird zwar theoretisch auch durch Luftteilchen und in Wolken gedämpft (Streuung, Absorption), dieser Anteil ist aber vernachlässigbar. Die Dämpfung im Niederschlag ist jedoch von entscheidender Bedeutung und variiert stark mit Art, Form und Größe der Niederschlagsteilchen. Für Regen mit durchschnittlicher Tröpfchengröße beträgt sie im G-Band (4-6GHz) rund $-0,0032\text{dB/km}$ pro 1mm/h Niederschlagsrate. Bei festem Niederschlag liegt sie weit darüber und ist außerdem aufgrund der unregelmäßigen Form und Konsistenz dieser Hydrometeore kaum zu berechnen (Hailzl, 2006). Die Ausbreitungsdämpfung hat zur Folge, dass ausgedehnte stratiforme Niederschläge vor allem im Winter nie in ihrer tatsächlichen Ausdehnung und Intensität erfasst werden. Auch Niederschläge hinter starken konvektiven Zellen können aufgrund dieses Problems nicht detektiert werden (Abschattung durch Starkniederschlag). Die Ausbreitungsdämpfung ist daher als eine der gravie-

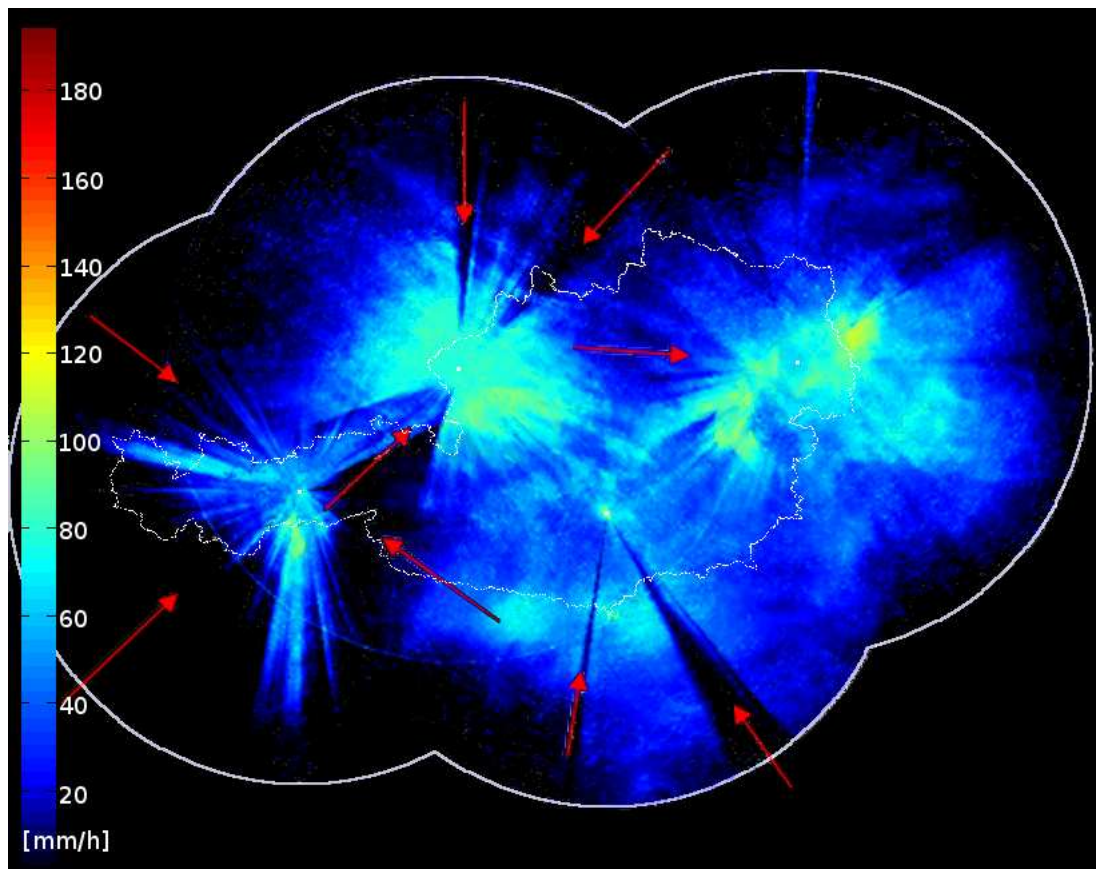


Abbildung 3.11:

Orographisch bedingte WXR-Abschattungsgebiete im WXR-Austria-Composite (Kaltenböck et al., 2010a): Die Summe mehrerer WXR-Bilder zeigt, auch im MAXCAPPI-Composite-Austria (siehe Abschnitt 2.2.1) deutliche, orographisch bedingte Abschattungsbereiche, die als Kreissektoren erkennbar sind. In der Graphik wurden nur einige, eindeutige Bereiche markiert.

rendsten Problemstellungen der Radarmeteorologie anzusehen.

- *Entfernungsdämpfung*: Darunter versteht man die entfernungsbedingte Abschwächung des Radarstrahls. Sie weist eine exponentielle Entfernungsabhängigkeit auf: Von 0-50km beträgt sie -34dB, von 51-100km -6dB und von 101-150km nur mehr -5dB (Hailzl, 2006). So kann es dazu kommen, dass sich ein zur Station ziehendes Niederschlagsfeld, obwohl es sich in Wirklichkeit nicht verändert, immer mehr verstärkt, je näher es der Station kommt. Auch dies birgt Potential für Fehlinterpretationen.
- *Unvollständige Strahlausfüllung*: Die unvollständige Strahlerfüllung sei nur der

Vollständigkeit halber erwähnt. Da die detektierten Volumenelemente nicht konstant sind, sondern aufgrund der Strahlverbreiterung mit zunehmender Entfernung größer werden, kann es dazu kommen, dass Niederschlag in großer Entfernung ein Volumenelement nur teilweise ausfüllt und daher unterschätzt wird.

3.2.1.3 Weitere Eigenheiten der radiometrischen Messung

Neben der Umrechnungsproblematik (Abschnitt 3.2.1.1) sowie den Messfehlern (Abschnitt 3.2.1.2) hat die radiometrische Messung weitere Eigenheiten, die beim Vergleich mit anderen Datenquellen eine Rolle spielen würden.

- Erstens ist zu beachten, dass der Radarstrahl mit zunehmender Entfernung an Höhe gewinnt. Vor allem in der MAXCAPPI-Darstellung (siehe Abschnitt 2.2.1) ist dies nicht offensichtlich. Dies ist im Wesentlichen auf zwei Ursachen zurück zu führen (Dotzek, 2010):
 1. *Geometrie*: Der Radarstrahl breitet sich grundsätzlich geradlinig mit Lichtgeschwindigkeit aus. Daher werden aufgrund der Erdkrümmung in großer Entfernung nur mehr sehr hochreichende Objekte getroffen. So hat zum Beispiel der Radarstrahl bei 0° Elevation in 200 km Entfernung bereits eine Höhe von etwa 3,5km (Hailzl, 2006). Somit wird auch in orographisch eigentlich gut einsehbaren Gebieten, die aber weiter von der WXR-Station entfernt sind, nur der obere Teil einer Niederschlagszelle erfasst.
 2. *Thermodynamik*: In der Atmosphäre kommt außerdem noch die Brechung zwischen optisch unterschiedlich dichten Luftschichten zum Tragen. (Beim Übergang vom optisch dichteren zum optisch dünneren Medium erfolgt eine Brechung zum Lot. Umgekehrt wird vom Lot gebrochen.) So kommt es zu zwei unterschiedlichen Phänomenen (Hailzl, 2006):
 - (a) Subrefraktion: Der Brechungsindex nimmt mit der Höhe zu. Der Radarstrahl wird dabei weniger stark zur optisch dichteren Schicht gebrochen und somit von der gekrümmten Erdoberfläche weggebrochen. Dies führt zu einer weiteren Verringerung der Radarreichweite und von weiter entfernten Niederschlagsgebieten wird somit noch weniger oder sogar nichts mehr erfasst.
 - (b) Superrefraktion: Die Atmosphäre wird nach oben hin optisch dünner. Hier tritt nun der gegenteilige Fall auf, der Radarhorizont vergrößert

sich, weil der Radarstrahl, relativ zum Lot, zur Erdoberfläche gebrochen wird. Sie tritt bei stabilen Luftschichtungen, insbesondere Inversionen auf. Im Extremfall ist dieser Effekt so stark, dass der Radarstrahl den Erdboden erreicht und Bodenziele verstärkt als Echos auf dem Radarbild zu sehen sind.

- Weiter spielt mit zunehmender Höhe auch die Verdunstung in den Luftschichten unterhalb der Wolke eine entscheidende Rolle. Je trockener diese Luftschichten sind, desto größer fällt dieser Effekt natürlich aus. Laut Hailzl (2006) kann in Radarbildern generell eine Abnahme der Reflektivität beobachtet werden, sobald Niederschlag aus der Wolke fällt. Dies ist auf eine Verringerung der Tropfengrößen durch Verdunstung (und bei großen Tropfen durch zusätzliches Zersplittern) zurückzuführen.
- Außerdem darf die Niederschlagsverfrachtung durch den Wind nicht vernachlässigt werden. Bei einer realistischen Endfallgeschwindigkeit von 3,5 m/s fällt ein Regentropfen aus einer Höhe von 1000m über Grund immerhin zirka 4 Minuten und 45 Sekunden. Eine mittlere Luftströmung von 10 m/s reicht dann aus, um den Niederschlag um 2,9km zu verfrachten (Hailzl, 2006).

3.2.2 Fertige Anforderungen- Schlussfolgerungen

Wie schon erwähnt soll es das Ziel dieser Arbeit sein, ein Verifikationsschema zu finden, dass die Qualität einer beliebigen WXR-Extrapolation erfassen kann. Dazu benötigt man als Erstes, wie schon in Kapitel 3.1 beschrieben, einen geeigneten Vergleichsdatensatz. Im Abschnitt 3.2.1 wurden nun alle wesentlichen Probleme behandelt, die auftreten, wenn man versucht WXR-Daten mit Bodenniederschlagsmessungen, beziehungsweise daraus gewonnen Produkten zu vergleichen. Die bedeutensten davon für den Synoptiker sind mit Sicherheit orographische Abschattungen, die Niederschlagsdämpfung (vor allem im Winter), sowie der Brightband-Effekt. Im Zuge einer Verifikation fallen aber auch einige Festechos, auf Grund ihrer Häufigkeit und Ausdehnung stark ins Gewicht. Hier seien vor allem kreisförmige Störungen, Sonnenechos und Seitenkeuleneffekte hervorgehoben. Da all diese im zu behandelnden Kontext auf keinen Fall vernachlässigt werden dürfen, liegt es nahe, in erster Instanz Wetterradarbilder als Vergleichsdaten zu verwenden. Viele der im Abschnitt 3.2.1 beschriebenen Probleme werden dadurch gelöst. An dieser Stelle sei nochmals betont,

dass jede Extrapolation natürlich stark von der Güte der Eingangsdaten abhängig ist. Durch die Verwendung von gemessenen WXR-Bildern als Vergleichdatensatz werden einige Fehler der Eingangsdaten zwar im Zuge der Verifikation berücksichtigt, sodass es besser gelingt, die Qualität der Extrapolation an sich zu erfassen, es muss einem aber bewusst sein, dass jede Vorhersage immer nur so „gut“ sein wird, wie die Qualität, der ihr zu Grunde liegenden Daten. Außerdem gibt es auch einige Phänomene, die vor allem aufgrund ihrer zeitlichen Variabilität auch beim Vergleich mit Wetterradaarbildern die Verifikation maßgeblich beeinflussen können. Dies trifft zum Beispiel auf Festechos wie Flugzeugechos oder Sonnenechos zu. Es gilt aber auch für orographische Abschattungsgebiete, die zwar natürlich an sich konstant bleiben, aber Niederschlagsfelder, die sich in den beziehungsweise aus dem Abschattungsbereich hinein/heraus bewegen, können im zeitlichen Verlauf nicht durchgehend detektiert werden. Dieses Erkenntnis muss vorrangig bei der Interpretation der Verifikationsergebnisse berücksichtigt werden. Trotz allem liegt der Vorteil der Verwendung von Wetterradaarbildern als Vergleichsdaten auf der Hand. Um verschiedene WXR-Extrapolationen (siehe Abschnitt 2) vergleichbar zu machen, ist es notwendig, ein Verifikationsschema zu finden, dass auf alle Extrapolationsmethoden gleichwertig anwendbar ist. Das bedeutet weiter, dass keinem Algorithmus Vorteile oder Nachteile gegenüber anderen entstehen dürfen. Das Schema muss absolut objektive Ergebnisse liefern. Alle weiteren Anforderungen konnten bereits zu Beginn des Kapitels 3.2 festgelegt werden und sollen zusammen mit den neuen Erkenntnissen zusammengefasst werden:

1. Die Verifikation soll wirklich nur die Qualität der WXR-Extrapolation wiedergeben, nicht die der Radarmessung an sich.
2. Das Verifikationsschema soll auf alle Arten von WXR-Extrapolationen (siehe Abschnitt 2) in gleichem Maße anwendbar sein um verschiedene Algorithmen vergleichbar zu machen.
3. Die Qualität einer WXR-Extrapolation soll möglichst vollständig beschrieben werden.
4. Das Verifikationsschema soll auch in Echtzeit verwendbar sein um die aktuelle Qualität zu überwachen und muss daher möglichst einfach aufgebaut und „überschaubar“ sein.
5. Die Ergebnisse sollen übersichtlich dargestellt und schnell interpretierbar sein.

4 Verifikationsmethodik

Nun ist es an der Zeit, aufgrund der zuvor definierten Anforderungen (Abschnitt 3.2.2) an ein Verifikationsschema, dieses auch näher festzulegen.

Der Forderung Nummer 1 (Es soll wirklich nur die Extrapolation verifiziert werden) soll in erster Linie durch die Auswahl geeigneter Vergleichsdaten nachgekommen werden. Es werden daher im ersten Ansatz ausschließlich gemessene WXR-Bilder zum vorhergesagten Zeitpunkt mit den Extrapolationen verglichen. Außerdem soll, wie schon erwähnt, im Rahmen der Interpretation der Ergebnisse auf diverse Effekte von Fehlern in den Ausgangsdaten geachtet werden.

Die nächste Forderung soll gewährleisten, dass unterschiedlichste Extrapolationalgorithmen (siehe Abschnitt 2) bezüglich ihrer Qualität vergleichbar gemacht werden können. Um dies zu ermöglichen, wird in weiterer Folge ausschließlich mit pixelbasierten Verifikationsmethoden gearbeitet werden. Es liegt die Vermutung nahe, dass bei jeder objektbasierten Herangehensweise die Ergebnisse stark von der Definition der Objekte abhängig sind. Der pixelbasierte Ansatz erscheint dem Autor daher weit objektiver und gerechter, wenn es darum geht, ein allgemein gültiges Verifikationssystem für WXR-Extrapolationen zu entwerfen.

Die dritte Forderung verlangt eine möglichst umfangreiche Beschreibung der Qualität der Extrapolationen in all ihren Aspekten. Zu diesem Zweck soll auf eine Einteilung von Murphy (1993) zurückgegriffen werden. Er charakterisiert zehn verschiedene Qualitätsattribute (siehe Tabelle 4.2), um die Qualität einer Vorhersage weitreichend erfassen zu können. Definiert werden diese Attribute durch die, ihnen zu Grunde liegenden Verteilungsfunktionen. Murphy verwendet dabei also einen verteilungsorientierten Verifikationsansatz (distribution oriented approach). Dieser Ansatz wird unter anderem bei Wilks (2006) näher erläutert. Brooks und Doswell-III (1996) greifen diesen ebenfalls auf und vergleichen ihn mit dem maßzahlorientierten Ansatz (measures-orientated approach). Der große Vorteil des verteilungsorientierten Ansatzes liegt eindeutig darin, alle Qualitätsaspekte beschreiben zu können, andererseits nimmt die Komplexität und Dimensionalität stark zu. Der traditionelle, maßzahlori-

enterte Ansatz zielt darauf ab, die Qualität mit einem, beziehungsweise zwei Aspekten (zum Beispiel Accuracy und Skill) pauschal zu beschreiben (Murphy, 1993).

Aufgrund der Forderung, die Qualität möglichst vollständig zu erfassen, liegt die Aufgabe nun darin, Methoden zu finden um diese, in Tabelle 4.2 theoretisch formulierten Qualitätsattribute, praxisgerecht bestimmen zu können. Dabei soll des weiteren darauf geachtet werden, dass das System eine gewisse Komplexibilität nicht übersteigt und somit den Forderungen Nummer 3 (Echtzeitanwendbarkeit), sowie Nummer 4 (Die Ergebnisse sollen schnell interpretierbar sein) Rechnung getragen werden. Im Folgenden sollen die einzelnen Qualitätsmerkmale, inklusive Lösungsansätzen diskutiert werden.

Zunächst soll noch die Beschaffenheit eines WXR-Bildes, beziehungsweise einer WXR-Extrapolation erläutert werden: Ein (gemessenes oder vorhergesagtes) WXR-Bild besteht aus Reflektivitätsinformation für jeden Bildpunkt. Das heißt jeder Bildpunkt wird mit einer Intensitätsklasse belegt. Nähere Information zur Intensitätsskala findet sich unter Kapitel 2.2.1. (Diese Klassen lassen sich, nach der in Abschnitt 3.2.1.1 erläuterten Systematik in Niederschlagsraten umrechnen.) Somit werden im Rahmen der Verifikation ausschließlich diskrete, kategorisierte Wertepaare verwendet. Die Vorhersagewerte enthalten keinerlei Unsicherheitsinformation („non-probabilistic“). Wenn man als Vergleichdaten, anstatt WXR-Bilder zum Beispiel Niederschlags-Summenbilder verwendet, so müssen diese, um sie vergleichbar zu machen, zuvor kategorisiert werden. Es werden also wieder diskrete Klassen kreiert.

Eine klassische Möglichkeit dieser Problematik zu begegnen, ist es, mit Hilfe eines Schwellwertes, die Dimensionalität zu beschränken. Wird nur ein Schwellwert verwendet, so entsteht ein dichotomes System, das heißt Vorhersage, sowie Beobachtung können je nur zwei verschiedene Werte annehmen.

In unserem Fall sind dies:

1. Niederschlag
2. kein Niederschlag

Somit kann jeder Bildpunkt, durch die entsprechende Kombination des Vorhersage- und Messwertes einer von insgesamt nur vier Situationen zugeordnet werden. Es entsteht eine 2x2 Kontingenztafel (siehe Abbildung 4.1). Zwei dieser Möglichkeiten kennzeichnen eine korrekte Vorhersage. Dies ist der Fall, wenn entweder Niederschlag vorhergesagt und gemessen wurde („hit“) beziehungsweise wenn er weder vorhergesagt

Aspekt	Definition	dafür relevante Verteilung(en)
Bias	Übereinstimmung zwischen mittlerer Vorhersage und mittlerer Beobachtung	$p(y)$ und $p(o)$
Accuracy	Mittlere Übereinstimmung zwischen Vorhersage und Beobachtungswertepaaren	$p(y, o)$
Skill	Accuracy der zu beurteilenden Vorhersage in Relation zur Accuracy durch eine Referenzprognose (Zufall, Klimatologie, Persistenz, etc.)	$p(y, o)$
Reliability	Übereinstimmung zwischen bedingter mittlerer Beobachtung und dazu passender Vorhersage, gemittelt über alle Vorhersagen	$p(o/y)$ und $p(y)$
Resolution	Differenz zwischen bedingter mittlerer Beobachtung und mittlerer Beobachtung, gemittelt über alle Vorhersagen	$p(o/y)$ und $p(o)$
Sharpness	Variabilität der Vorhersagen, beschrieben durch die Verteilung der Vorhersagen	$p(y)$
Uncertainty	Variabilität der Beobachtungen, beschrieben durch die Verteilung der Beobachtungen	$p(o)$
Discrimination	Übereinstimmung zwischen bedingter mittlerer Vorhersage und dazu passender Beobachtung, gemittelt über alle Beobachtungen	$p(y/o)$ und $p(o)$

Tabelle 4.2:

Ausgewählte, relevante Verteilungen für die Vorhersagequalität (Murphy, 1993)

		Beobachtung		
		Ja	Nein	
Vorhersage	Ja	a	b	a+b
	Nein	c	d	c+d
		a+c	b+d	n= a+b+c+d

Abbildung 4.1:

2x2 Kontingenztabelle, nachgebildet nach Wilks (2006): In den Zellen mit den Buchstaben „a“ bis „d“ stehen für die vier verschiedenen Kombinationsmöglichkeiten der Wertepaare die jeweiligen absoluten Anzahlen. Tritt der Fall „a“ ein, so wurde Niederschlag vorhergesagt und gemessen („*hit*“), im Fall „b“ wurde er zwar vorhergesagt, trat aber nicht ein („*false alarm*“). Im Fall „c“ verhält es sich genau umgekehrt, hier wurde kein Niederschlag prognostiziert, aber dafür gemessen („*missed event*“). Wenn Niederschlag weder vorhergesagt, noch gemessen wurde so liegt der Fall „d“ vor („*correct rejection*“). Die Gesamtanzahl der Wertepunkte („n“) berechnet sich durch die Addition der Summen aller vier möglichen Fälle, also der Stichprobenumfang.

noch gemessen wurde („*correct rejection*“) Letzterer tritt in Situationen, bei denen im Vergleich zur gesamten Wertepaaranzahl nur wenige niederschlagsbehaftete Pixel (in Messung oder Vorhersage) vorkommen, sehr häufig auf. Dies führt bei vielen klassischen Verifikationsmaßen, die man aus dieser Tabelle ableiten kann, zu Problemen bei der Interpretation dieser Maße.

Diese 2x2 Kontingenztabelle fasst somit die Verifikationsdaten auf möglichst einfache Weise zusammen, indem jeweils die Beobachtung sowie die Vorhersage nur zwei verschiedene Werte annehmen kann. Das Verifikationsproblem wird auf Dimension *drei* reduziert (Wilks, 2006). Die Dimension ergibt sich laut Formel 4.1, wobei I die Anzahl der möglichen Werte der Vorhersage und J jene der Beobachtung angibt.

$$dim = (I \times J) - 1 \quad (4.1)$$

Im Falle einer Reduktion auf ein dictionomes Problem, ergibt sich $I = J = 2$ und ein 3-dimensionales Verifikationsproblem. Das bedeutet weiter, dass mindestens drei verschiedene Parameter notwendig sind, um den Informationsgehalt der 2x2 Kontingenztabelle

tabelle vollständig wiederzugeben (Wilks, 2006). Die Liste der gängigen Parameter, die man aus dieser Tabelle gewinnen kann, ist lang. Der verteilungsorientierte Verifikationsansatz baut darauf auf, dass sämtliche Qualitätsinformation aus der gemeinsamen Wahrscheinlichkeitsverteilung der Vorhersagen und Beobachtungen gewonnen werden kann. Diese schreibt sich nach Wilks (2006) als:

$$p(y_i, o_j) = Pr \{y_i, o_j\} = Pr \{y_i \cap o_j\}; i = 1, \dots, I; j = 1, \dots, J. \quad (4.2)$$

In unserem dichtenomen Problem ist $I = J = 2$. Aber selbst in diesem vereinfachten Fall ist es schwierig, diese gemeinsame Verteilung direkt zu verwenden. Daher ist es üblich auf zwei Faktorisierungen zurückzugreifen, die von Murphy und Winkler (1987) entwickelt wurden und über die einzelnen Qualitätsaspekte leichter Aufschluss geben können. Die erste nennt sich „*Calibration-Refinement Factorization*“.

$$p(y_i, o_j) = p(o_j | y_i) p(y_i); i = 1, \dots, I; j = 1, \dots, J. \quad (4.3)$$

Der erste Teil besteht aus I bedingten Verteilungen. Diese geben an, wie oft jedes Ereignis eintrat, wenn ein bestimmtes vorhergesagt wurde. Dies nennt sich auch die Kalibrierung der Vorhersage („*calibration*“). Der zweite Teil ist die Randverteilung, die die relative Häufigkeit der vorhergesagten Ereignisse angibt (Wilks, 2006). Sie nennt sich auch „*refinement distribution*“.

Die zweite Möglichkeit, die gemeinsame Verteilung laut Murphy und Winkler zu faktorisieren, ist die „*likelihood-base rate factorization*“:

$$p(y_i, o_j) = p(y_i | o_j) p(o_j); i = 1, \dots, I; j = 1, \dots, J. \quad (4.4)$$

Hier zeigen die bedingten Wahrscheinlichkeiten die Wahrscheinlichkeit („*likelihoods*“), mit der jeder mögliche Vorhersagewert y_i bei einer bestimmten Beobachtung o_j eintritt. Weiters beinhaltet diese Faktorisierung die Randverteilung der Beobachtungen, die man auch die klimatologische Verteilung nennt (Wilks, 2006).

Genauere Informationen zu den Eigenschaften dieser Faktorisierungen sind in der erwähnten Literatur reichlich vorhanden, daher konzentriert sich diese Arbeit in weiterer Folge auf ihre Anwendbarkeit für das dichtenome Verifikationsproblem, beschrieben durch die Kontingenztafel aus Abbildung 4.1.

Die einzelnen Verteilungen, aus denen die Faktorisierungen bestehen, lassen sich

laut Wilks, für das vereinfachte Problem folgendermaßen aus den Einträgen der obigen Kontingenztafel berechnen: Als erstes die Einträge der „calibration-refinement-factorization“:

$$p(y_1) = \frac{a+b}{n} \quad (4.5a)$$

$$p(y_2) = \frac{c+d}{n} \quad (4.5b)$$

$$p(o_1 | y_1) = \frac{a}{a+b} \quad (4.6a)$$

$$p(o_2 | y_1) = \frac{b}{a+b} = FAR \quad (4.6b)$$

$$p(o_1 | y_2) = \frac{c}{c+d} \quad (4.6c)$$

$$p(o_2 | y_2) = \frac{d}{c+d} \quad (4.6d)$$

Analog dazu lassen sich die Verteilungen der „likelihood-base rate factorization“ folgendermaßen schreiben:

$$p(o_1) = \frac{a+c}{n} \quad (4.7a)$$

$$p(o_2) = \frac{b+d}{n} \quad (4.7b)$$

$$p(y_1 | o_1) = \frac{a}{a+c} = POD \quad (4.8a)$$

$$p(y_2 | o_1) = \frac{c}{a+c} \quad (4.8b)$$

$$p(y_1 | o_2) = \frac{b}{b+d} = POFD \quad (4.8c)$$

$$p(y_2 | o_2) = \frac{d}{b+d} \quad (4.8d)$$

Die Gleichungen 4.5a bis 4.8d machen es somit möglich, sämtliche Bestandteile der Faktorisierungen der gemeinsamen Verteilungsfunktion von Vorhersagen und Beob-

achtungen für das dichotome Verifikationsproblem aus der 2x2 Kontingenztabelle zu gewinnen und somit die Qualität dieses vereinfachten Systems vollständig zu beschreiben. Einige dieser Verteilungen können in weiterer Folge direkt verwendet werden, um einen bestimmten Qualitätsaspekt zu beschreiben. Daher wurden sie in obigen Gleichungen eigens benannt. Ergänzend zu dieser verteilungsorientierten Herangehensweise wurden im Laufe der Zeit einige zusätzliche Maßzahlen ermittelt und verwendet, um die Qualität dieses klassischen Verifikationsproblems umfangreich beschreiben zu können. Auf diese wird nun anschließend in den jeweiligen Abschnitten, in denen die einzelnen Qualitätsmerkmale näher untersucht werden sollen, eingegangen.

4.1 Bias

„Bias“ bezeichnet den *mittleren Fehler* (ME) einer Vorhersage. Er lässt sich für das dichotome Problem folgendermaßen berechnen (Atger et al., 2010):

$$ME_{dichotom} = \frac{a + b}{a + c} \quad 0 \leq 1(\text{perfect score}) < \infty \quad (4.9)$$

Der Bias für die 2x2 Kontingenztabelle ergibt sich somit aus dem Verhältnis der Summe von Punkten, an denen Niederschlag vorhergesagt wurde, zu jenen Punkten, an denen Niederschlag beobachtet wurde. Er enthält keinerlei Informationen über den Anteil der richtigen beziehungsweise falschen Prognosen. Er ist ausschließlich ein Maß dafür, ob im Mittel eher zu viel oder zu wenig Niederschlag prognostiziert wird.

- $ME < 1$: Es wird im Mittel zu wenig Niederschlag vorhergesagt („underforecast“)
- $ME > 1$: Es wird im Mittel zu viel/oft Niederschlag prognostiziert („overforecast“)

Der mittlere Fehler lässt sich jedoch auch recht einfach für das allgemeine (nicht probabilistische) Verifikationsproblem berechnen, womit zusätzliche Information gewonnen werden kann:

$$ME_{allg} = \frac{1}{n} \sum_{i=1}^n (f_i - o_i) \quad -\infty < 0(\text{perfect score}) < \infty \quad (4.10)$$

In dieser Formel repräsentieren die f_i die einzelnen Vorhersagen, die o_i die dazugehörigen Beobachtungen und n steht, wie schon beim dichotomen Fall für die

Gesamtzahl der Mess/Bildpunkte. Im Unterschied zu dem, aus der Kontingenztabelle berechneten Bias, werden hier Fehler gemäß ihrer Intensität (zum Beispiel in dBZ, siehe auch Abbildung 2.5) gewichtet. Das heißt eine Abweichung von drei Intensitätsstufen hat dreimal so viel Gewicht, wie eine Abweichung von einer Intensitätsstufe.

Für beide Maße bedeutet aber ein Wert nahe 0 nicht zwangsläufig, dass die Vorhersage keinen systematischen Fehler hat, sondern nur, dass die mittlere Abweichung sehr klein ist. Fehler mit positivem und Fehler mit negativem Vorzeichen könnten sich gegenseitig aufgehoben haben. Aus diesem Grund muss man den Bias mit anderen Maßen, wie zum Beispiel dem mittleren absoluten Fehler (siehe Abschnitt 4.2) vergleichen.

Weiters sei erwähnt, dass man qualitative Information über den systematischen Fehler (Bias) auch aus dem Reliability-Diagramm gewinnen kann. An dieser Stelle sei aber auf Abschnitt 4.3 verwiesen. Dort wird das Reliability-Diagramm ausführlich beschrieben.

4.2 Accuracy

Die Accuracy spiegelt die mittlere Übereinstimmung zwischen den Vorhersagen und den dazugehörenden Beobachtungen wider (Wilks, 2006). Sie berechnet sich mit den Einträgen der Kontingenztabelle nach folgender Formel:

$$PC = \frac{a + d}{n} \qquad 0 \leq PC \leq 1(\text{perfect score}) \qquad (4.11)$$

PC steht für „“, es wird hier also der Anteil der richtigen Vorhersagen ermittelt, wobei richtig vorhergesagter Niederschlag gleich stark in die Berechnung eingeht wie richtig prognostizierter Nicht-Niederschlag. Wenn sich auf den zu vergleichenden WXR-Produkten, im Verhältnis zur gesamten Pixelanzahl, nur wenig niederschlagsbehaftete Bildpunkte finden, wird man mit PC, durch die Gleichgewichtung der richtigen Vorhersagen („hits“ und „correct rejection“), in den meisten Fällen sehr gute Ergebnisse erzielen. Dies tritt bei Radarbildern recht häufig auf.

Um in solchen Fällen eine Alternative zur Verfügung zu haben, wurde von Gilbert (1884) der „threat score“ (TS) oder „critical success index“ (CSI) entwickelt.

$$TS = CSI = \frac{a}{a + b + c} \qquad 0 \leq CSI \leq 1(\text{perfect score}) \qquad (4.12)$$

Im Gegensatz zu PC werden beim CSI die korrekten Nicht-Niederschlagsprognosen nicht berücksichtigt. Beide falsch vorhergesagten Ereignisse („false alarms“ und „misses“) werden sowohl bei PC, als auch bei CSI gleich gewichtet. Auch der Wertebereich der beiden Maße ist identisch. Er erstreckt sich von $PC = CSI = 0$ als schlechtest mögliches Ergebnis bis zur perfekten Vorhersage mit $PC = CSI = 1$.

Aber auch klassische, statistische Momente können wichtige Aufschlüsse über die mittlere Übereinstimmung von Vorhersagen und Beobachtungen liefern. Auch sie kommen, wie schon der unter Abschnitt 4.1 erwähnte mittlere Fehler ME_{allg} , ohne Vereinfachung des Verifikationsproblems aus und können allgemein definiert werden:

$$MAE = \frac{1}{n} \sum_{i=1}^n | (f_i - o_i) | \quad 0(\text{perfect score}) \leq MAE < \infty \quad (4.13)$$

Der mittlere absolute Fehler (MAE) summiert im Gegensatz zum ME (Abschnitt 4.1) alle Abweichungen zwischen Vorhersage und Beobachtung, unabhängig vom Vorzeichen, auf. Es können sich somit keine Fehler gegenseitig aufheben, andererseits gewinnt man aber auch keine Information über die Richtung der Abweichungen (Atger et al., 2010). Er stellt eine gute Ergänzung zum ME dar. Die Fehler werden ebenfalls gemäß ihrer Größe gewichtet. Wenn es darum geht, große Abweichungen noch stärker, nämlich quadratisch, zu berücksichtigen, so muss man den „mean squared error“, also den mittleren quadratischen Fehler (MSE), beziehungsweise seine Wurzel, den „root mean squared error“ (RMSE) heranziehen.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2 \quad 0(\text{perfect score}) \leq MSE < \infty \quad (4.14a)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad 0(\text{perfect score}) \leq RMSE < \infty \quad (4.14b)$$

Für genauere Betrachtungen zum MSE inklusive Zerlegungen, um mehr Information über einzelne Fehlerquellen zu gewinnen, sei auf Murphy und Winkler (1987) verwiesen.

Bedingte Verteilung	Name	Abkürzung
$p(o_1 y_1) = \frac{a}{a+b}$	Frequency of Hits	FOH
$p(o_2 y_1) = \frac{b}{a+b}$	False Alarm Ratio	FAR
$p(o_1 y_2) = \frac{c}{c+d}$	Detection Failure Ratio	DFR
$p(o_2 y_2) = \frac{d}{c+d}$	Frequency of correct Null Forecasts	FOCN

Tabelle 4.3:

Definition der „Reliability“-Maße nach Doswell-III et al. (1990): Diese Maße erhält man aus der „calibration-refinement factorization“ der gemeinsamen Wahrscheinlichkeitsverteilung von Vorhersagen und Beobachtungen. Das gängigste Maß ist eindeutig FAR.

4.3 Reliability

Die „Reliability“ beschreibt die Zuverlässigkeit der Vorhersage. Für das dichotome System lässt sich diese Glaubwürdigkeit vollständig durch die Gleichungen 4.6a bis 4.6d beschreiben (Wilks, 2006). Diese vier Gleichungen decken alle möglichen Fälle ab, sie beschreiben die relative, bedingte Häufigkeit, zu der ein Ereignis (Niederschlag oder Nicht-Niederschlag) beobachtet wurde unter der Bedingung einer bestimmten Vorhersage (wiederum Niederschlag oder Nicht-Niederschlag vorhergesagt). Von Doswell-III et al. (1990) werden alle vier skalaren „Reliability“-Maße der 2x2 Kontingenztafel näher behandelt und auch benannt (siehe Tabelle 4.3).

Das mit Abstand am häufigsten verwendete Maß dieser vier, ist das „False Alarm Ratio“ (FAR). Es gibt an, welcher Teil aller vorhergesagten Niederschlagsereignisse sich als falsch erwiesen hat.

$$FAR = \frac{b}{a+b} \qquad 0(\text{perfect score}) \leq FAR \leq 1 \qquad (4.15)$$

Aber gerade im Bereich der „Reliability“ wäre es wünschenswert auch Informationen über das nicht vereinfachte, vollständige Verifikationsproblem zu erhalten, besonders um die Glaubwürdigkeit von Vorhersagen mit hoher Intensität zu beurteilen. Zu diesem Zweck soll in dieser Arbeit eine für diskrete Prediktanden neue Herangehensweise gewählt werden, die aber im Bereich der probabilistischen Prognosen mit binärem Prediktanden („Ensemble-Ja/Nein-Vorhersagen“), sowie für nicht probabilistische Vorhersagen mit kontinuierlichem Prediktanden eine Standardmethode darstellt. Im Bereich der probabilistischen Prognosen wird zur umfassenden Beurteilung der „Reliability“ ein sogenanntes Reliability-Diagramm erstellt (Persson und Grazi-

ni, 2007), für deterministische Vorhersagen mit kontinuierlichem Prediktanden laut Wilks (2006) sogenannte „Conditional Quantile Plots“.

Beim klassischen Reliability-Diagramm wird die Wahrscheinlichkeit, mit der ein Ereignis vorhergesagt wird (auf der Abszisse), gegen die entsprechende Beobachtungsfrequenz (auf der Ordinate) geplottet.

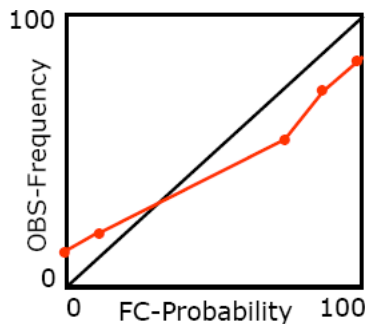


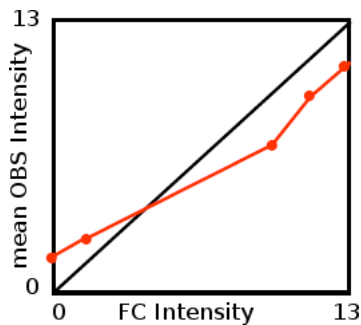
Abbildung 4.2:

Layout eines klassischen Reliability-Diagramms (Hagedorn, 2009): In dieser Konfiguration wird es zur Verifikation probabilistischer Prognosen mit binärem Prediktanden (zum Beispiel: „Ensemble-Ja/Nein-Niederschlagsvorhersage“) am Europäischen Zentrum für mittelfristige Wettervorhersage (ECMWF) laut Persson und Grazzini (2007); Hagedorn (2009) verwendet.

Dieses Konzept ist natürlich nicht eins zu eins auf WXR-Extrapolationen übertragbar, da es sich bei ihnen um diskrete, deterministische Prognosen handelt, die naturgemäß keinerlei Wahrscheinlichkeitsinformation liefern. Daher soll die Konfiguration, des für unsere Zwecke verwendbaren Reliability-Diagramms der der „Conditional Quantile Plots“ entsprechen. Auf der Abszisse soll die Vorhersageintensität aufsteigend aufgetragen werden, auf der Ordinate die entsprechenden, gemittelten Beobachtungsintensitäten (siehe Abbildung 4.3). Im weiteren Dokument soll diese Darstellung trotzdem als Reliability-Diagramm (für deterministische Prognosen mit diskretem Prediktanden) bezeichnet werden.

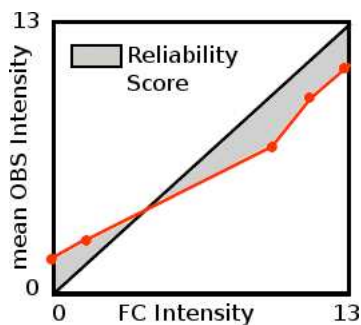
Wie kann man nun aus diesem Diagramm die Glaubwürdigkeit der Prognose gewinnen? Persson und Grazzini (2007) beschreibt den „Reliability Score“ als Fläche zwischen der Datenpunktskurve und der Diagramm-Diagonale von links unten nach rechts oben (siehe Abbildung 4.4). Je kleiner diese Fläche ausfällt, desto größer ist die Reliability.

Wie schon unter Abschnitt 4.1 angedeutet, kann man aus diesem Reliability-Diagramm auch (zumindest qualitative) Information über den systematischen Fehler



Abbildungung 4.3:

Layout des adaptierten Reliability-Diagramms: In dieser Konfiguration ist das Prinzip des Reliability-Diagramms nun auf nicht probabilistische Vorhersagen anwendbar.



Abbildungung 4.4:

Ermittlung des Reliability-Scores nach Hagedorn (2009)

(Bias) ableiten.

Eine zu „trockene“ Vorhersage, das heißt es wird im Mittel mehr Niederschlag gemessen als vorhergesagt („underforecasting“), erkennt man an Datenpunkten der Reliability-Kurve (Abbildung 4.5a), die oberhalb der 45°-Diagonalen zu liegen kommen. Das Gegenteil, eine zu „nasse“ Vorhersage („overforecasting“) führt andererseits zu einer Kurve unterhalb der Diagonalen (Abbildung 4.5b).

Eine andere Möglichkeit, mit der die bedingte Verteilung der Beobachtungen bei gegebener Vorhersage, also $p(o/y)$, noch genauer untersucht werden kann, sind Histogramme, in denen für jede beobachtete Intensität die absolute oder relative Häufigkeit aufgetragen wird (siehe Abbildung 4.6). Für jede Vorhersageintensität muss ein eigenes Histogramm erstellt werden. Im Falle von 14 Vorhersageklassen kann die volle Information der bedingten Verteilungen also in 14 Histogrammen dargestellt werden.

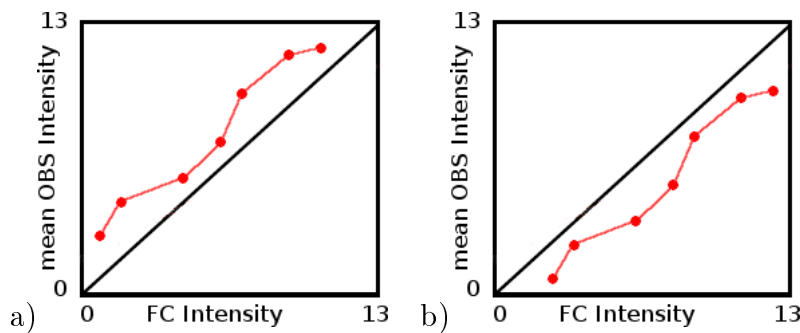


Abbildung 4.5:

Das Reliability-Diagramm liefert auch Bias Information. Anhand der Lage der Datenpunktskurve, gegenüber der 45°-Diagonalen, kann man auf den systematischen Fehler (Bias) schließen. Liegt die Kurve oberhalb der Diagonalen spricht man von einer zu „trockenen“ Vorhersage, liegt sie darunter handelt es sich um eine zu „nasse“ Prognose.

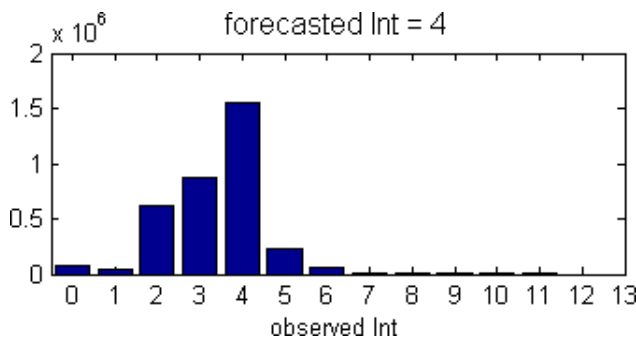


Abbildung 4.6:

Histogramm der bedingten Beobachtungen: Bei gegebener Vorhersageintensität 4 werden hier die absoluten Häufigkeiten der dazugehörigen Beobachtungen im Histogramm aufgetragen.

4.4 Resolution & Sharpness

Das zuvor vorgestellte Reliability-Diagramm (siehe Abschnitt 4.3) liefert neben der schon gezeigten Reliability und dem Bias noch mehr Information über die gemeinsame Wahrscheinlichkeitsverteilung unserer Vorhersage-Beobachtungs-Wertepaare. Auch die Resolution (Auflösung) der Prognose kann daraus gewonnen werden. Die Auflösung definiert die Schärfe des Vorhersagesystems (Haimberger, 2008). Das Maß für die Auflösung ist die Fläche zwischen der Verteilungsfunktion und der Linie der klimatologischen Verteilung. Je größer diese ausfällt, desto höher ist die Auflösung des Prognosesystems (siehe Abbildung 4.7).

Laut Haimberger (2008) ist die Auflösung im Vergleich zur Zuverlässigkeit (Re-

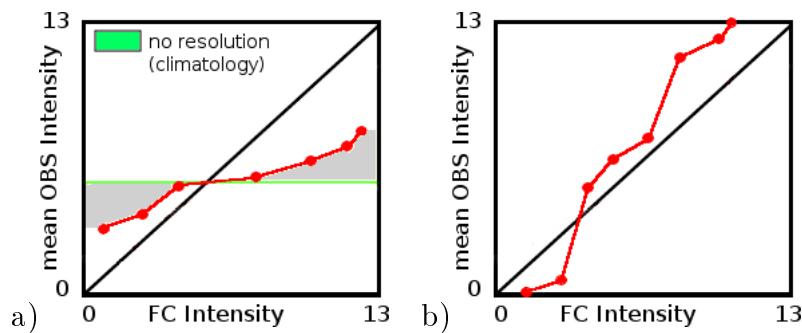


Abbildung 4.7:

Allgemeine Reliability-Diagramme unterschiedlicher Resolution: Im Teil a) der Abbildung wurde eine Verteilungsfunktion mit schlechter Auflösung geplottet (overconfident). Die grüne Linie kennzeichnet die Grenze, ab der die Vorhersage keinerlei Auflösung aufweist. Die Vorhersagen entsprechen dann der Klimatologie des Ereignisses. Die Größe der grauen Fläche ist das Maß für die Auflösung des Systems. Je größer die Fläche, desto höher die Resolution. Im Teil b) sieht man eine Verteilungsfunktion mit guter Resolution (underconfident).

liability) das wichtigere Maß, weil die Zuverlässigkeit optimiert werden kann. Diesen Prozess nennt man Kalibrierung.

In diesem Zusammenhang sei noch erwähnt, dass man aus einem Reliability-Diagramm für probabilistische Vorhersagen zusätzlich Information über die „Sharpness“ des Prognosesystems gewinnen kann. Die „Sharpness“ beschreibt die Fähigkeit des Systems, sehr hohe beziehungsweise niedrige Wahrscheinlichkeiten vorherzusagen (Atger et al., 2010). Diese Eigenschaft lässt sich aber auf das nicht probabilistische System kaum anwenden. Die, der „Sharpness“ zu Grunde liegende Information, nämlich, wie viele Wertepaare die einzelnen Datenpunkte im Diagramm repräsentieren, ist aber auch für unsere Zwecke von großer Bedeutung. Je mehr Information ein Datenpunkt enthält, desto besser gilt er als abgesichert. Somit kann man daraus eine Art Unsicherheit für jeden Datenpunkt ableiten.

4.5 Discrimination

Der Aspekt der „Discrimination“ (dt.: Abgrenzung) untersucht die Fähigkeit des Vorhersagesystems zwischen den verschiedenen Ereignissen zu unterscheiden (Persson und Grazzini, 2007). Dazu muss man untersuchen, welche Prognosen das System bei gegebenen Beobachtungen liefert. Dies ist ein bedeutender Unterschied zu den bisherigen Untersuchungen. Bis jetzt wurde vorhersage-bedingt vorgegangen, das heißt es wurde

bei gegebenem Vorhersageergebnis, die Verteilung der bedingten Beobachtungen untersucht. An dieser Stelle soll genau umgekehrt vorgegangen werden. Es wird nun die bedingte Verteilung der Vorhersagen bei gegebener Beobachtung untersucht. Die folgenden Untersuchungen bauen dabei auf der „likelihood-base rate factorization“ nach Murphy und Winkler (1987) auf. Der Einfachheit halber soll an dieser Stelle wieder auf das dichtenome Verifikationsproblem reduziert werden. Man untersucht also die bedingten Verteilungen $p(y_1|o_1)$ sowie $p(y_1|o_2)$, die für beide möglichen Beobachtungsergebnisse die Häufigkeit einer dazu gehörenden Niederschlagsprognose zeigen (Wilks, 2006). Diese Verteilungen wurden bereits in den Formeln 4.8a, sowie 4.8c beschrieben.

Dabei handelt es sich um zwei der bedeutendsten Maße, die aus der 2x2 Kontingenztafel berechenbar sind.

$$POD = \frac{a}{a + c} \qquad 0 \leq POD \leq 1(\text{perfect score}) \qquad (4.16)$$

$$POFD = \frac{b}{b + d} \qquad 0(\text{perfect score}) \leq POFD \leq 1 \qquad (4.17)$$

Die POD (probability of detection) oder auch Entdeckungswahrscheinlichkeit zeigt, mit welcher Wahrscheinlichkeit Niederschlag, der gemessen wurde, auch vorhergesagt wurde. Dieses Maß ist gut geeignet um Vorhersagen von „seltenen“ Ereignissen zu bewerten. Es ist aber darauf zu achten, dass falsch vorhergesagter Niederschlag das Ergebnis nicht negativ beeinflusst. Somit kann eine gute POD auch durch generelle Überschätzung der Niederschlagsfelder zustande kommen. Um aussagekräftige Ergebnisse zu erhalten, sei es empfohlen, POD mit einer Abschätzung der Zuverlässigkeit (von vorhergesagten Niederschlagsereignissen), üblicherweise FAR (siehe Gleichung 4.15) oder eben mit dem zweiten hier vorgestellten Maß, der POFD („probability of false detection“) zu kombinieren (Atger et al., 2010). POFD sagt aus, welcher Anteil des gemessenen Nicht-Niederschlags fälschlicherweise als Niederschlag prognostiziert wurde. POFD findet man in der Literatur auch häufig unter dem Namen „false alarm rate“, abgekürzt mit F, wieder. POD wird auch „hit rate“ (H) genannt.

Eine sehr anschauliche Kombination von POD und POFD stellt das ROC-Diagramm („relativ operating characteristic“) dar. Es wurde erstmals von Mason (1982) für meteorologische Zwecke vorgeschlagen. Dabei wird POFD auf der Abszisse gegen POD

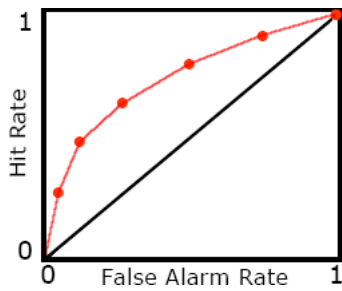


Abbildung 4.8:

Schematisches ROC-Diagramm nach Hagedorn (2009): Die Fläche unterhalb der ROC-Kurve (rot) ist ein Maß für die Fähigkeit des Prognosesystems zur Discrimination (Abgrenzung) der Ereignisse.

auf der Ordinate geplottet (siehe Abbildung 4.8).

Für die Kurve im ROC-Diagramm gilt, je mehr sie sich dem linken oberen Eckpunkt ($POD = 1$ und $POFD = 0$) annähert, desto besser ausgeprägt ist die Fähigkeit des Vorhersagesystems, die beiden möglichen Ereignisse (Niederschlag und Nicht-Niederschlag) voneinander abzugrenzen. Als Maß dafür kann man die Fläche unterhalb der ROC-Kurve heranziehen (Wilks, 2006). Je größer diese ausfällt, um so besser. Ein Vorhersagesystem mit perfekt ausgeprägter Discrimination würde zu $A = A_{perf} = 1$ führen. Würde man ein ROC-Diagramm für Zufallsprognosen auswerten, so würden die einzelnen Werte in der Nähe der Diagonale (von links unten nach rechts oben) zu liegen kommen. Die Fläche unterhalb der ROC-Kurve würde sich somit $A = A_{rand} = 0,5$ annähern. Somit kann man sagen, dass ein Vorhersagesystem, das zu einer Fläche unterhalb der ROC-Kurve, die kleiner oder gleich 0,5 ist, führt, keinen Vorteil gegenüber einer Zufallsprognose liefert, somit keine Fähigkeit zur Discrimination der Ereignisse besitzt. Es sei auch erwähnt, dass ein ROC-Diagramm keine Auskunft bezüglich systematischer Fehler (Bias) geben kann (Persson und Grazzini, 2007).

Sollte es notwendig sein, die Information aus dem ROC-Diagramm in einem skalaren Maß zusammenzufassen, so schlägt Wilks (2006) zu diesem Zweck den ROC-Skill-Score vor. Die genaue Definition eines Maßes, mit dem der Skill einer Prognose gemessen werden kann, wird anschließend in Abschnitt 4.6 erläutert werden. Sozusagen als kleine Überleitung soll an dieser Stelle nur die Berechnung dieses einen Maßes

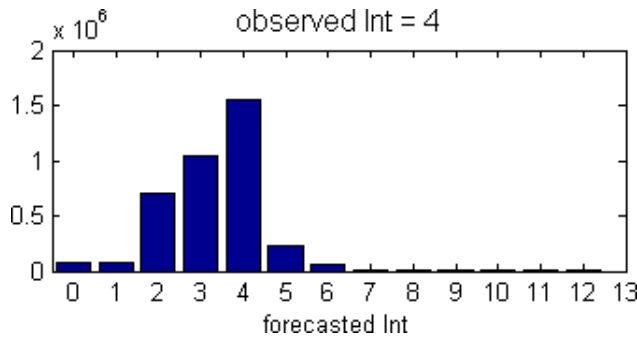


Abbildung 4.9:

Histogramm der bedingten Vorhersagen: Bei gegebener Beobachtungsintensität 4 werden die absoluten Häufigkeiten der dazugehörigen Vorhersagen im Histogramm aufgetragen.

Erwähnung finden.

$$ROC_{SS} = \frac{A - A_{rand}}{A_{perf} - A_{rand}} = \frac{A - \frac{1}{2}}{1 - \frac{1}{2}} = 2A - 1 \quad (4.18)$$

Ergänzend gibt es noch die Möglichkeit, die bedingte Verteilung der Vorhersagen bei gegebener Beobachtung, $p(y/o)$, des nicht vereinfachten (nicht dichten) Systems in Histogrammen darzustellen (siehe Abbildung 4.9). Dies kann analog zu den Betrachtungen zur Reliability (siehe Abbildung 4.6) behandelt werden. Der Vorteil liegt auf der Hand. Man verliert keine Information der ursprünglichen Verteilung durch die Vereinfachung des Systems. Andererseits sind mit dieser Methode mehrere Diagramme (je eines für jede beobachtete Intensität) nötig.

4.6 Skill

Den „Skill“ einer Prognose kann man laut Wilks (2006) auch als „relative Accuracy“ bezeichnen. „Relativ“ deshalb, weil der Skill einer Prognose durch den Vergleich mit Referenzprognosen gewonnen wird. Man berechnet also die Accuracy im Vergleich zu Referenzprognosen, wie zum Beispiel der Klimatologie, Persistenz oder Zufallsprognose. Der Vergleich mit anderen Prognosen hat den großen Vorteil, einen Anstieg der Accuracy, der nur durch für das Vorhersagesystem „günstige Wetterbedingungen“ zustande kommt, also im Allgemeinen durch Zufall, berücksichtigen zu können. Skill

Scores berechnen sich nach Wilks (2006) nach folgendem Schema:

$$SS_{allg} = \frac{A_{fc} - A_{reference}}{A_{perfect} - A_{reference}} \quad (4.19)$$

Wobei A ein Maß für die Accuracy darstellt, zum Beispiel PC (Gleichung 4.11), CSI (Gleichung 4.12) oder MSE (Gleichung 4.14a). Wenn man als Maß für die Accuracy den MSE heranzieht, kann man unter anderem folgende Referenz nach Wilks (2006) leicht berechnen:

$$MSE_{persistence} = \frac{1}{n} \sum_{i=1}^n (o_{i,t-1} - o_{i,t})^2 \quad (4.20)$$

Der Index t in Gleichung 4.20 repräsentiert die Zeitinformation. Mit $t-1$ wird das letzte WXR-Bild vor dem Bild zum Zeitpunkt t gekennzeichnet.

Nun sind im Laufe der Zeit eine ganze Reihe von Skill Scores entstanden, von denen die bekanntesten hier vorgestellt werden sollen:

4.6.1 Gilbert Skill Score (Equitable Threat Score)

$$CSI = \frac{a}{a+b+c} \implies ETS = \frac{a - a_{random}}{(a - a_{random}) + b + c} \quad -\frac{1}{3} \leq ETS \leq 1 (perfect\ score) \quad (4.21)$$

$$a_{random} = \frac{(a+b)(a+c)}{n} \quad (4.22a)$$

Man sieht in Gleichung 4.21, dass der Gilbert Skill Score, oder auch Equitable Threat Score (ETS) genannt, eine Modifikation des CSI aus Gleichung 4.12 darstellt. Es wird versucht, die nur auf Zufall basierenden richtig vorhergesagten Niederschlagsereignisse („hits“) zu eliminieren (Atger et al., 2010). Ein Wert von 0 bedeutet, dass eine Prognose keinen Vorteil gegenüber der Zufallsprognose liefert, man sagt, sie hat keinen Skill. Der ETS eignet sich laut Atger et al. (2010) gut um Niederschlagvorhersagen zu bewerten, weil er unterschiedlich große Niederschlagsregime, durch die Elimination des Zufalls, auch unterschiedlich bewertet. Analog zum CSI geht richtig vorhergesagter Nicht-Niederschlag nicht in die Berechnung ein und beide Fehlerarten („misses“ und „false alarms“) werden gleich gewichtet. Bei sehr seltenen Ereignissen strebt der ETS gegen den herkömmlichen CSI.

4.6.2 Heidke Skill Score

Auch der Heidke Skill Score (HSS) nach Heidke (1926) versucht, ähnlich dem ETS, durch Zufall entstandene richtige Prognosen zu eliminieren. HSS wurde speziell für die Verifikation von „seltenen Ereignissen“ entwickelt. Als Referenz wird hier PC (Gleichung 4.11) verwendet, das aus Zufallsvorhersagen gewonnen wird. Somit gilt:

$$A_{fc} = \frac{a + d}{n} \quad \text{und} \quad A_{perfect} = 1 \quad (4.23)$$

$$A_{reference} = Prob_{pos,corr} + Prob_{neg,corr} \quad (4.24)$$

$$Prob_{pos,corr} = \frac{(a + b)(a + c)}{n} \quad Prob_{neg,corr} = \frac{(b + d)(c + d)}{n} \quad (4.25a)$$

Damit kann man nun die allgemeine Gleichung für Skill Scores (Gleichung 4.19) auswerten. Es ergibt sich (Wilks, 2006):

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)} \quad -\infty \leq HSS \leq 1(\text{perfect score}) \quad (4.26)$$

Schaefer (1990) zeigte, dass der ETS und der HSS über folgende Beziehung (Gleichung 4.27) miteinander verknüpft sind:

$$ETS = \frac{HSS}{2 - HSS} \quad (4.27)$$

4.6.3 True Skill Statistics

Die „True Skill Statistics“ (TSS) wurden ursprünglich von Peirce (1884) unter dem Namen „Peirce’s skill score“ (PSS) eingeführt und im Laufe der Jahre einige Male wieder entdeckt und unter neuem Namen publiziert, zum Beispiel auch als „Hanssen-Kruipers discriminant“ von Hanssen und Kruipers (1965). Der Name „True Skill Statistics“ stammt letztendlich von Flueck (1987). TSS ist ähnlich aufgebaut wie HSS. Als Referenzprognose wird wieder PC (Gleichung 4.11) verwendet. Der Unterschied liegt laut Atger et al. (2010) oder Wilks (2006) in den Termen im Nenner, die die Wahrscheinlichkeit einer korrekten Niederschlagsvorhersage, beziehungsweise einer korrekten Nicht-Niederschlagsvorhersage beschreiben. Beide sollen bei TSS frei von systema-

tischen Abweichungen („unbiased“) sein. Diese Terme haben somit folgende Gestalt:

$$Prob_{pos,corr} = \frac{(a+c)^2}{n} \quad Prob_{neg,corr} = \frac{(b+d)^2}{n} \quad (4.28)$$

Die Gleichung 4.23 kann für TSS übernommen werden. Einsetzen in die allgemeine Definition, Gleichung 4.19, liefert dann:

$$TSS = \frac{ad - bc}{(a+c)(b+d)} \quad -1 \leq TSS \leq 1(\text{perfect score}) \quad (4.29)$$

4.6.4 Vergleich der Skill Scores

Die drei verschiedenen, in den Abschnitten 4.6.1 bis 4.6.3 angeführten Skill Scores stellen nur einen kleinen Auszug aller, in der Literatur vorhandenen Möglichkeiten dar, den Skill einer Vorhersage zu bestimmen. Für genauere Informationen sei auf Wilks (2006) verwiesen, dort findet sich eine weit umfangreichere Auflistung.

Wie bereits erwähnt, zeigte Schaefer (1990), dass der ETS mit dem HSS verknüpft ist (siehe Gleichung 4.27), wobei der Unterschied der beiden darin besteht, dass für den ETS in erster Linie die korrekten Vorhersagen von Niederschlagsereignissen von Belang sind. Beim HSS geht die korrekte Vorhersage von Nicht-Niederschlagsereignissen stärker ins Ergebnis ein. Auch TSS gibt sowohl Niederschlags-, als auch Nicht-Niederschlagsereignissen entsprechendes Gewicht. Doswell-III et al. (1990) kritisiert jedoch, dass TSS, die sich laut Atger et al. (2010) auch als $TSS = POD - POFD$ schreiben lässt, bei sehr seltenen Ereignissen zu sehr an POD annähert. Somit ist es möglich, durch eine zu „nasse“ Vorhersage (Niederschlagsgebiete werden überschätzt) den TSS zu schönen. Doswell-III et al. schlägt daher den HSS für die Vorhersage von „seltenen Ereignissen“ als passender vor, als TSS. Wilks (2006) kann dieser Eigenschaft jedoch auch Positives abgewinnen, da ein korrekt prognostiziertes Ereignis (Niederschlag oder Nicht-Niederschlag), wenn es seltener auftritt, größeren Einfluss auf die TSS hat.

4.7 Festlegen des Verifikationsschemas

Um der Forderung einer umfassenden Qualitätsbeschreibung nachzukommen, muss das Schema so aufgebaut sein, dass jeder der obigen Qualitätsaspekte durch mindestens ein Maß, Diagramm oder ähnliches repräsentiert wird. Um eine möglichst einfache und übersichtliche Aufbereitung der Ergebnisse gewährleisten zu können, soll

der Grundstock des entstehenden Schemas aus Diagrammen bestehen. Dazu eignen sich in erster Linie das Reliability-Diagramm und das ROC-Diagramm. Allein mit diesen zwei Informationsquellen sind bereits die Qualitätsaspekte Reliability, Resolution, Sharpness und Bias (durch das Reliability-Diagramm), sowie die Discrimination (ROC-Diagramm) erfasst. Ergänzend kann mit geeigneten Maßzahlen wie CSI, MAE oder RMSE die Accuracy und mittels eines Skill Scores, zum Beispiel TSS der Skill der Prognosen ermittelt werden. Welche Maßzahlen und Scores für unsere Zwecke am geeignetsten sind, soll sich noch im Rahmen der Prototyp-Testphase zeigen. Allein auf Grund theoretischer Überlegungen fällt es schwer, auch durch widersprüchliche Meinungen in der gängigen Literatur bedingt, eine fundierte Entscheidung zu treffen.

Der Grundstock des gewählten Verifikationsschemas besteht also aus:

- Reliability-Diagramm
- ROC-Diagramm
- Maß für die Accuracy (zum Beispiel: CSI)
- Maß für den Skill (zum Beispiel: TSS)

In Abschnitt 1.2 wurden zu Beginn die Qualitätsanforderungen an eine WXR-Extrapolation definiert. Im Wesentlichen soll eine Vorhersage von WXR-Bildern die Position von Niederschlagszellen (vor allem bei konvektiven Systemen), sowie ihre Intensität richtig prognostizieren. Das beschriebene Verifikationsschema soll die Qualität einer WXR-Extrapolation umfassend beschreiben. Dabei ist es aber wichtig, dass neben den theoretisch definierten Qualitätskriterien nach Murphy (1993) (siehe Tabelle 4.2) auch die, aus der Praxis entstandenen Kriterien an eine WXR-Extrapolation durch das gefundene Schema erfassbar sind. Es soll somit erläutert werden, welche Information die einzelnen Maße (oder die aus ihnen abgeleiteten Qualitätsattribute) in Bezug auf die Orts-, beziehungsweise Intensitätsvorhersage der Niederschlagsgebiete liefern können.

Bei der Betrachtung des Bias (Abschnitt 4.1, Gleichung 4.10) wurde bereits erwähnt, dass ein Wert nahe 0 nicht zwangsläufig auf eine Prognose ohne systematischen Fehler hinweist. Fehler mit positivem und negativem Vorzeichen könnten sich gegenseitig aufgehoben haben. Aber er liefert in jedem Fall Information darüber, ob Niederschlagsfelder im Mittel über- oder unterschätzt wurden, also über die Qualität der Intensitätsvorhersage.

Besonders interessant ist aber der Vergleich des Bias mit Maßen für die Accuracy, zum Beispiel MAE (Gleichung 4.13). Bei der Berechnung des MAE werden, im Gegensatz zum Bias alle Intensitätsabweichungen zwischen Beobachtung und Prognose als Betrag aufsummiert. Wenn somit der Bias sehr kleine Werte liefert, der MAE im Gegensatz dazu durch hohe Werte aber auf einen großen Fehler schließen lässt, liegt es auf der Hand, dass in diesem Fall die Ortsvorhersage versagt hat, die Intensität der Zelle aber gut prognostiziert wurde. Liefern beide Maßzahlen hohe Werte, so muss man von einer unzureichenden Intensitätsvorhersage ausgehen. Über die Beschaffenheit des Fehlers in der Intensitätsprognose gibt der Vergleich mit dem RMSE (Gleichung 4.14b) weiteren Aufschluss. Er hat die Eigenschaft, große Abweichungen stärker zu gewichten als kleinere.

Maße, die sich aus dem vereinfachten, dichten System ableiten, sich also aus der 2x2 Kontingenztabelle (Abbildung 4.1) berechnen, geben naturgemäß nur bedingt Information über die Intensitätsvorhersage wieder, einzig das generelle Über-, beziehungsweise Unterschätzen der Größe von Niederschlagsfeldern kann erfasst werden. Hierfür eignet sich unter anderem der Bias, berechnet aus der Kontingenztabelle (Gleichung 4.9). Im Vordergrund steht hier aber eindeutig die Bestimmung der Qualität der Ortsvorhersage, wofür sich zum Beispiel der CSI (Gleichung 4.12) anbietet.

Ergänzend dazu bieten das Reliability-Diagramm, sowie die beiden Arten der bereits vorgestellten Histogramme zur Beschreibung der Reliability (siehe Abbildung 4.6) und der Discrimination (siehe Abbildung 4.9) eine genaue Darlegung der Intensitätsvorhersage.

Somit ist gewährleistet, dass beide, praktisch begründeten Qualitätsmerkmale ausreichend und aus mehreren Standpunkten beurteilt werden können. Darüber hinaus aber ermöglicht die Betrachtung der allgemein gültigen Qualitätsaspekte für Vorhersagen eine noch viel detailliertere Analyse des Prognosesystems. Diese zusätzlichen Informationen können vor allem, wenn es um die Optimierung des Extrapolationsalgorithmus geht, von großer Bedeutung sein. Immerhin soll das Verifikationssystem die wissenschaftliche Verifikation nach Brier und Allen (1951), also jene Verifikation, die die kontinuierliche Verbesserung des Systems zum Ziel hat, genauso ermöglichen, wie die administrative (Echtzeit-) Verifikation zur steten Überwachung der aktuellen Qualität.

5 Verifikationssoftware

Das im Kapitel 4 entwickelte Verifikationsschema soll im weiteren Verlauf anhand eines Prototyps umgesetzt und getestet werden. Zuvor müssen jedoch, die den Prototyp betreffenden Anforderungen an das Verifikationssystem besprochen werden, um ihnen im Zuge der Umsetzung gerecht zu werden.

5.1 Anforderungen

Um ein Programm längerfristig nutzbar zu machen, ist einerseits eine ausführliche Dokumentation von Nöten, andererseits erleichtert eine intuitive Bedienung den Umgang erheblich. Zu diesem Zweck wurde beschlossen eine einfache und kleine, aber leicht bedienbare graphische Benutzeroberfläche zu schaffen.

Da es sich bei diesem System um einen Prototyp handelt, der unter anderem auch dazu verwendet werden soll, das vorgeschlagene Verifikationsschema zu testen, sollen zusätzlich zu den essentiellen Bestandteilen, die in Abschnitt 4.7 definiert wurden, einige weitere der besprochenen Methoden (in Kapitel 4) implementiert werden. Dies sind zum Beispiel die erwähnten Histogramme zur detaillierten Ermittlung der Reliability und Discrimination, sowie mehrere Maße des dichotomen Problems.

Um trotz der Vielfalt der daraus entstehenden Produkte und Maße das System einigermaßen übersichtlich zu gestalten, sollen möglichst alle Bestandteile graphisch und soweit möglich (zum Beispiel einzelne dichotome Maße) kombiniert darstellbar sein. Somit sollte auch die Möglichkeit einer schnellen Interpretation für fachkundige Personen gegeben sein.

Die Problematik der Vergleichdaten (siehe Kapitel 3.2.1) wurde ausführlich erläutert. Es hat sich gezeigt, dass der Vergleich mit gemessenen WXR-Bildern am sinnvollsten erscheint. Der zu entwickelnde Prototyp muss also in der Lage, sein extrapolierte Bilder, sowie gemessene WXR-Bilder zu verarbeiten. All diese Daten liegen (siehe Abschnitt 2.2.1) als Bilder im PNG-Format vor. Die Dateinamenskonvention wurde ebenfalls bereits an entsprechender Stelle behandelt.

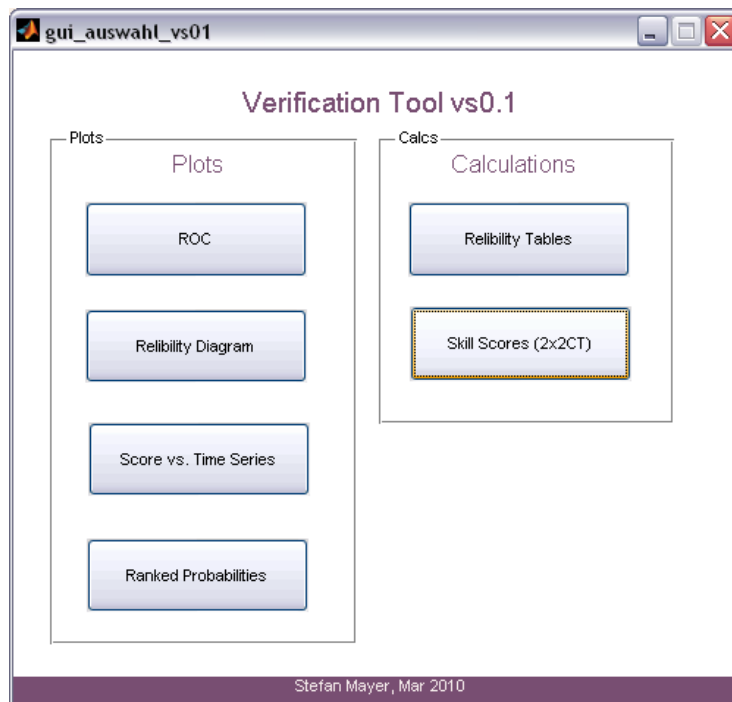


Abbildung 5.1:

Verification Tool: Startfenster des MatLab Prototyps. Man erkennt zwei Funktionsgruppen, auf der rechten Seite jene, die für die Berechnung der Maße und Verteilungen zuständig sind, sowie die Darstellungsfunktionen auf der linken Seite des Fensters.

5.2 Umsetzung

Aufgrund der Flexibilität beim Umgang mit Bilddateiformaten (PNG), sowie der gut anwendbaren vektorbasierten Rechenoperationen (Bilder kann man im Allgemeinen als Matrix betrachten) und den programmierspezifischen Vorkenntnissen des Autors, wurde der Prototyp mittels MatLab (Version R2008a) entwickelt. Abbildung 5.1 zeigt das Ausgangsfenster des Programms, dass mit dem Befehl *gui_auswahl_vs01* im MatLab Command Window zu starten ist. Grundsätzlich beinhaltet es zwei unterschiedliche Gruppen von Funktionen. Auf der rechten Seite des Fensters (Calculations) jene, die die Datengrundlage (Dichotome Maße, Reliability-Matrizen, etc.) für etwaige Darstellungen wie zum Beispiel dem ROC-, oder dem Reliability-Diagramm liefern und auf der linken Seite (Plots) solche, die diese graphischen Darstellungen dann tatsächlich zeichnen.

Das System wurde entwickelt um Analysen über längere Betrachtungszeiträume (zum Beispiel ein Monat) zu erlauben. Zu diesem Zweck wurde das Programm so

aufgebaut, dass man die, in diesem Fall länger andauernden Berechnungen der einzelnen Qualitätsparameter (Calculations) zuerst startet. Nach abgeschlossener Berechnungsphase stehen dann die Darstellungsfunktionen auf der linken Seite (Plots) zur Verfügung. Sie verwenden die zuvor berechneten Daten. Sollte eine Darstellungsfunktion trotzdem gestartet werden, obwohl die benötigten Daten noch nicht berechnet wurden, so weist das Programm auf das Fehlen hin. Man hat dann die Möglichkeit, die Berechnung nachzuholen oder den Plotbefehl abubrechen. Auf die genaue Bedienung wird nun im Zuge der einzelnen Funktionen näher eingegangen. Der Source-Code sowie ein Prozessbaum liegt im Anhang bei.

5.2.1 Calculations

Wie bereits erwähnt, bewerkstelligen die beiden Funktionen im „calculations“-Block sämtliche Berechnungen. Das entwickelte Verifikationschema beruht, wie in Kapitel 4 gezeigt, auf zwei unterschiedlichen Ansätzen, für die je eine eigene Berechnungsfunktion angelegt wurde. Der „Skill Scores“-Button öffnet die Benutzeroberfläche zur Berechnung sämtlicher dichotomen Maße, sowie der klassischen, statistischen Momente (ME, MAE, etc.). Sollen hingegen bedingte Wahrscheinlichkeiten des allgemeinen Systems betrachtet werden (Reliability-Diagramm, Ranked-Probability-Histogramme), so gelangt man mit Hilfe des „Reliability Tables“-Buttons zur benötigten Benutzeroberfläche.

5.2.1.1 Reliability Tables

Abbildung 5.2 zeigt die Bedienungsumgebung zur Berechnung der „Reliability Tables“. Hier erkennt man nun, dass der Prototyp auf den Testdatensatz (siehe Abschnitt 2.2.2) zugeschnitten ist. Daher kann man als Erstes entscheiden, welche(s) der vier möglichen Parameterkombinationen (Parametersets) verifiziert werden soll(en). Um die Software universell nutzen zu können, müsste man an dieser Stelle simple Modifikationen vornehmen. Weiter ist es notwendig, den/die Extrapolationsschritt(e) auszuwählen, um somit die Art der Extrapolation vollständig zu bestimmen. Zur näheren Bestimmung der Extrapolation wird also der linke Teil der Benutzeroberfläche verwendet.

In den mittleren Fenstern werden Verifikationsdetails sowie Dateipfade festgelegt. Der Zielordner ist jener, in dem die Ergebnisse gespeichert werden sollen, der Datenordner bezeichnet den Ort, der zu verarbeitenden Rohdaten. Beide Pfadangaben

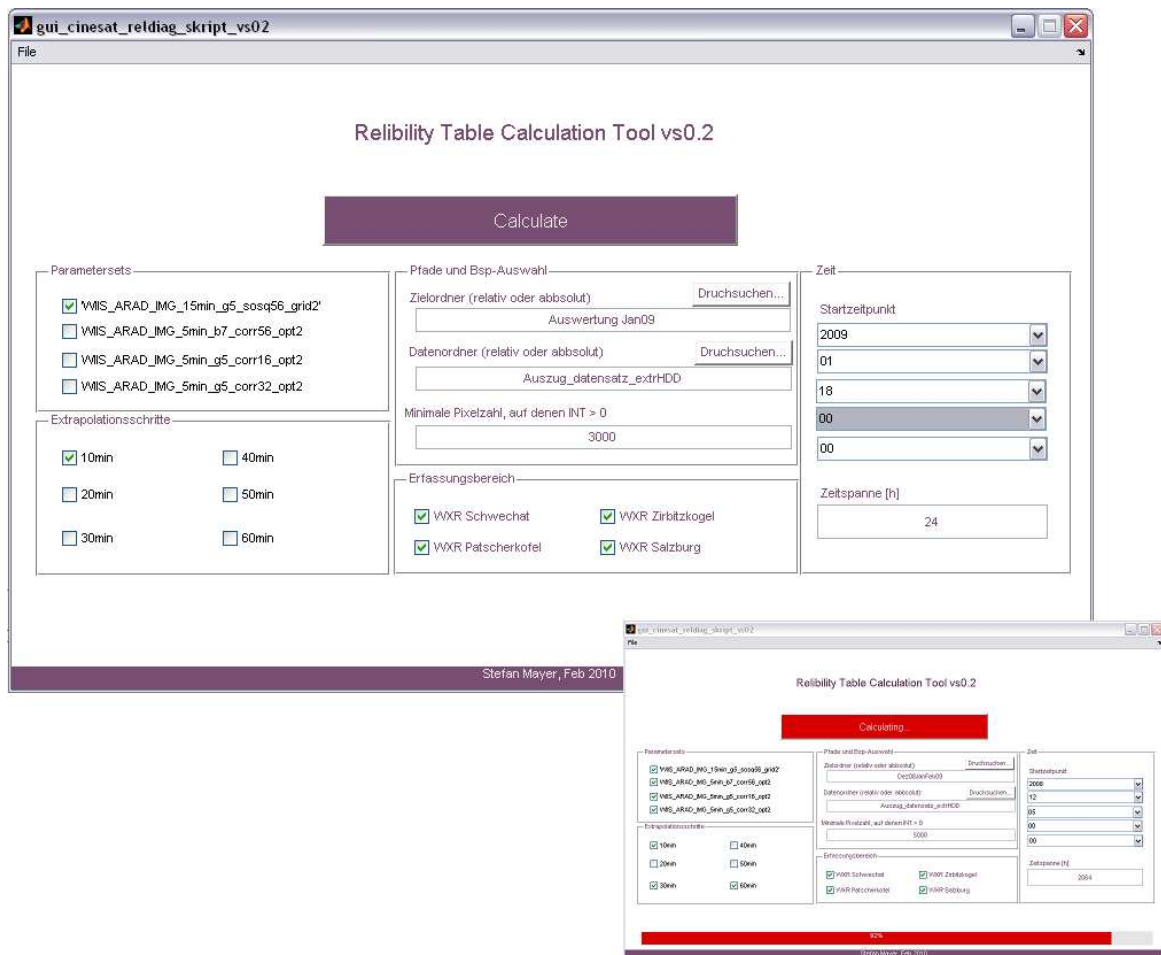


Abbildung 5.2:

Reliability Table Calculation Tool: Benutzeroberfläche zur Berechnung der bedingten Häufigkeiten des allgemeinen Systems. Die große Abbildung zeigt den Grundzustand, in dem die Berechnungseinstellungen vorgenommen werden können. Das kleine, untere Bild beschreibt den „Busy“-Zustand. Am roten Balken erkennt man den Berechnungsfortschritt, der in diesem Beispiel bereits 92% beträgt.

können absolut oder relativ zum Ordner, in dem sich die MatLab-Routine befindet, erfolgen. Als nächstes muss die Verifikation näher beschrieben werden, genauer gesagt geht es um den Ausschnitt des WXR-Austria-Composite, der betrachtet werden soll, sowie um die minimale Anzahl jener Pixel, die in diesem Ausschnitt einen Intensitätswert größer Null aufweisen. Durch den Ausschluss von Extrapolationen, die diese Bedingung nicht erfüllen, wird der störende Einfluss von etwaigen Artefakten (das sind Festechos, siehe Abschnitt 3.2.1.2) auf das Verifikationsergebnis verkleinert und zusätzlich die Rechenzeit optimiert. Die auswählbaren Verifikationsgebiete spiegeln die Erfassungsbereiche der österreichischen WXR-Stationen wieder.

Bevor die Berechnungen nun endlich starten können ist es nur noch notwendig, den Startzeitpunkt und die Betrachtungszeitspanne einzutragen. In Abbildung 5.2 startet das Programm mit den Extrapolationen die am 18. Jänner 2009 um Mitternacht erstellt wurden und endet mit jenen, die am 19. Jänner um Mitternacht berechnet wurden, also genau 24 Stunden (Zeitspanne) später.

Jetzt kann die Berechnung mittels des „Calculate“-Buttons gestartet werden. Daraufhin ändert der Button seine Farbe zu Rot und seine Inschrift in „Calculating...“. Außerdem erscheint im unteren Bereich des Fensters ein ebenfalls in rot gehaltener Balken, dem der Berechnungsfortschritt zu entnehmen ist (siehe Abbildung 5.2). Das Ende der Berechnung wird durch das Verschwinden des Balkens sowie durch einen wieder in violett erscheinenden „Calculate“-Button angezeigt. Ab diesem Zeitpunkt ist es möglich, Einstellungen zu verändern und neue Berechnungen zu starten. Die Reliability-Tables werden als MAT-Datei abgespeichert. Der Dateiname gibt Auskunft über alle einstellbaren Größen. Er könnte zum Beispiel folgendermaßen aussehen: „reltab_set2_step10_eb1234_mpz5000_200812050000.123840.mat“. Hier wurde also das Parameterset Nummer 2 mit einem Extrapolationsschritt von 10 Minuten für die Erfassungsbereiche 1 bis 4 (bei einer minimalen Pixelanzahl von 5000) vom 5. Dezember 2008 bis Ende Februar 2009 (86 Tage a 24 Stunden a 60 Minuten = 123840 Minuten) ausgewertet.

Anmerkung: Während des Berechnungsvorganges darf das MatLab-Arbeitsverzeichnis nicht gewechselt werden, das Ausführen anderer MatLab-Programme in dieser Zeit ist nicht möglich.

5.2.1.2 Skill Scores

Das zweite Berechnungsprogramm deckt, wie bereits zu Beginn dieses Kapitels er-

wähnt, alle Maße des dichotomen Systems sowie die klassischen statistischen Momente, wie ME, MAE oder RMSE ab. Die meisten Einstellmöglichkeiten decken sich mit dem zuvor beschriebenen Programm zur Berechnung der Reliability-Tables. Parameterset, Extrapolationschritt, die einzelnen Dateipfade sowie der Erfassungsbereich- und Zeiteinstellungen müssen analog zu Abschnitt 5.2.1.1 eingegeben werden. Ergänzend dazu ist hier aber die Auswahl der Schwellwerte zu beachten. Es gibt zwei davon. Als Erstes soll der „klassische“ Schwellwert erwähnt werden, der das allgemeine System in ein dichotomes transformiert (siehe Abschnitt 4). Durch diesen Schwellwert wird Ereignis von Nicht-Ereignis, beziehungsweise im Speziellen Niederschlag von Nicht-Niederschlag unterschieden. Er ist durch die mit „Intensität 1“ bis „Intensität 4“ beschrifteten Kästchen auszuwählen (siehe Abbildung 5.3), wobei auch eine Mehrfachauswahl möglich ist. Der zweite Schwellwert, der mit dem Kürzel „SW0“ bezeichnet werden soll, ist ein Versuch, um eine ausgewogenere Kontingenztafel zu erhalten. Durch ihn wird festgelegt, ob bestimmte Pixel gar nicht betrachtet werden. Alle Bildpunkte, die in beiden zu vergleichenden Bildern einen kleineren Wert als diesen zweiten Schwellwert aufweisen, werden bei der Berechnung ignoriert. Damit ergibt sich ein Verifikationsproblem, in dem korrekt vorhergesagte Nicht-Niederschlagsereignisse nicht mehr dominieren und laut Theorie viele dichotome Maßzahlen besser anwendbar sein sollten. Natürlich muss der klassische Schwellwert mindestens eine Intensitätsstufe höher gesetzt werden als dieser Zusätzliche. Dies wird vom Berechnungsprogramm erzwungen.

Gestartet wird die Berechnung auch hier über den „Calculate-Button“, der während der Berechnung in roter Farbe erscheint und die Meldung „Calculating...“ beinhaltet. Zusätzlich erscheint ebenfalls wieder ein Balken, der den Berechnungsfortschritt anzeigt. Im Unterschied zu 5.2.1.1 werden die Ergebnisse des Skill Score Calculation Tool als Textdateien gespeichert. Die Dateinamensgebung verläuft nach dem selben Muster wie zuvor. „WIIS_ARAD_IMG_5min_b7_corr56_opt2__step010_sw2_0_eb1234__mpz5000_200812050000.123840.txt“ kennzeichnet somit wieder eine Berechnung vom 5. Dezember 2008 bis Ende Februar 2009, bei der das Parameterset 2 (das entspricht: 5min_b7_corr56_opt2; siehe Tabelle 2.2) mit einem Extrapolationschritt von 10 Minuten für den gesamten Erfassungsbereich aller Radarstationen ausgewertet wurde. Der klassische Schwellwert wurde auf Intensität 2 gesetzt, der zweite, ergänzende Schwellwert (SW0) wurde hier nicht berücksichtigt. Sein Wert beträgt in diesem Beispiel Intensität 0.

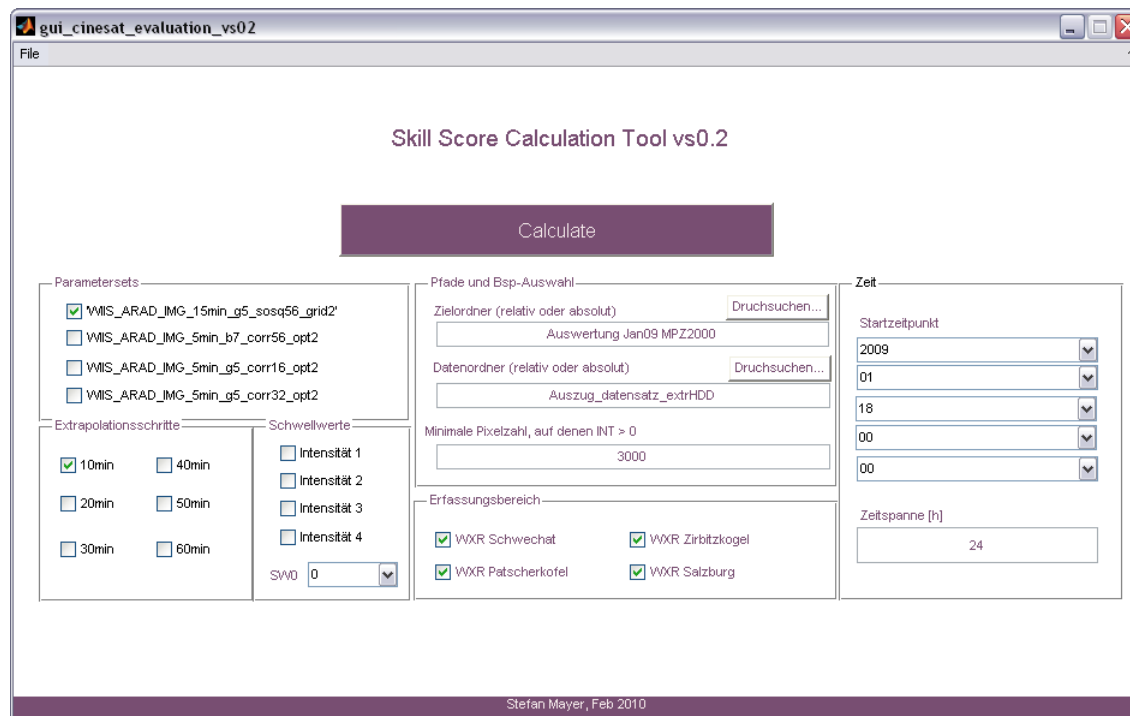


Abbildung 5.3:

Skill Score Calculation Tool: Benutzeroberfläche zur Berechnung der klassischen statistischen Momente, sowie der Maße des dichotomen Systems. Im Unterschied zum Berechnungsprogramm der Reliability-Tables aus Abbildung 5.2 müssen hier zwei Schwellwerte spezifiziert werden.

5.2.2 Plots

Nun soll die zweite Funktionsgruppe behandelt werden. Es handelt sich um jene, die für die Darstellung der Berechnungsergebnisse verantwortlich ist. Je zwei bauen auf die Ergebnisse eines Berechnungsprogramms auf. Das Programm „Score vs. Time Series“ plottet einzelne statistische Momente sowie Maße des dichotomen Systems gegen den Extrapolationsschritt und baut daher genauso wie das „ROC“-Programm auf den Ergebnissen aus dem „Skill Score Calculation Tool“ auf. Um andererseits ein „Reliability-Diagramm“ oder ein Histogramm der bedingten Wahrscheinlichkeiten des allgemeinen Systems („Ranked Probabilities“) darzustellen, muss auf den Berechnungen des „Reliability Table Calculation Tool“ aufgebaut werden.

5.2.2.1 ROC

Beim ROC-Diagramm wird POFD auf der Abszisse gegen POD auf der Ordi-

nate geplottet. Diese Begriffe sowie die Adaption des ROC-Diagramms für WXR-Extrapolationen wurden bereits ausführlich unter Abschnitt 4.5 erläutert. Der Aufbau des „ROC Plotting Tools“ wurde wie alle Darstellungsfunktionen ähnlich zu den bereits behandelten Berechnungsprogrammen gewählt. Somit decken sich hier einige Einstellmöglichkeiten mit denen vom „Skill Score Calculation Tool“. Somit sind Startzeitpunkt und Zeitspanne sowie der Datenordner (der jetzt natürlich den Output des Skill Score Calculation Tools beinhalten muss), Extrapolationsschritt und Erfassungsbereich sowie die minimale Pixelanzahl (mit Intensitäten größer 0) analog zu Abschnitt 5.2.1.2 zu behandeln. Um einen Vergleich zwischen Berechnungen mit verschiedenen Schwellwerten zu ermöglichen, wurde bei diesem Programm allerdings die Auswahl von Parameterset und klassischem Schwellwert gekoppelt. Somit kann jedes Parameterset mit jedem Schwellwert beliebig kombiniert und dargestellt werden. Der zusätzliche Schwellwert SW0 (siehe Abschnitt 5.2.1.2) kann nur global gewählt werden.

Um das ROC-Diagramm nun plotten zu können, ist es notwendig, die Wertebereiche (Bins) zu definieren, deren Daten zu einem Punkt der ROC-Kurve gemittelt werden. Dies kann im Automatikmodus erfolgen, in dem das Programm diese Bins so definiert, dass in jedem Bereich gleich viele Wertepaare zu liegen kommen. Man hat nur die Möglichkeit, die Anzahl der Bins zu steuern. Es gibt aber auch die Möglichkeit, die Intervalle per Hand, also manuell festzulegen. Im Manuellmodus muss somit die Anzahl der Bins sowie deren Unter- und Obergrenze im unteren Fenster der Benutzeroberfläche eingestellt werden. Im Automatikmodus verschwindet dieses Eingabefenster. Das Ergebnis beider Möglichkeiten ist in den Abbildungen 5.4b,c ersichtlich. In beiden Fällen wird ein ROC-Diagramm mit zusätzlichem Histogramm, aus dem die Anzahl der Wertepaare pro Bin ersichtlich ist, geplottet. Auf der Abszisse sind zusätzlich die Mittelwerte der Bins mit kleinen Strichen markiert.

5.2.2.2 Score vs. Time Series

Als nächstes soll gleich auf die zweite graphische Anwendung der Werte des dichotomen Systems eingegangen werden. Dabei handelt es sich um „Score vs. Time“-Diagramme. Das heißt, es wird ein beliebiges Maß gegen den Extrapolationsschritt aufgetragen. Im Testdatensatz sind Extrapolationen mit einer Vorhersagezeit von 10 bis 60 Minuten enthalten. Es ist auch möglich, zwei verschiedene Maße in einem Diagramm darzustellen. In diesem Fall erscheint im Diagramm eine zweite Ordinate

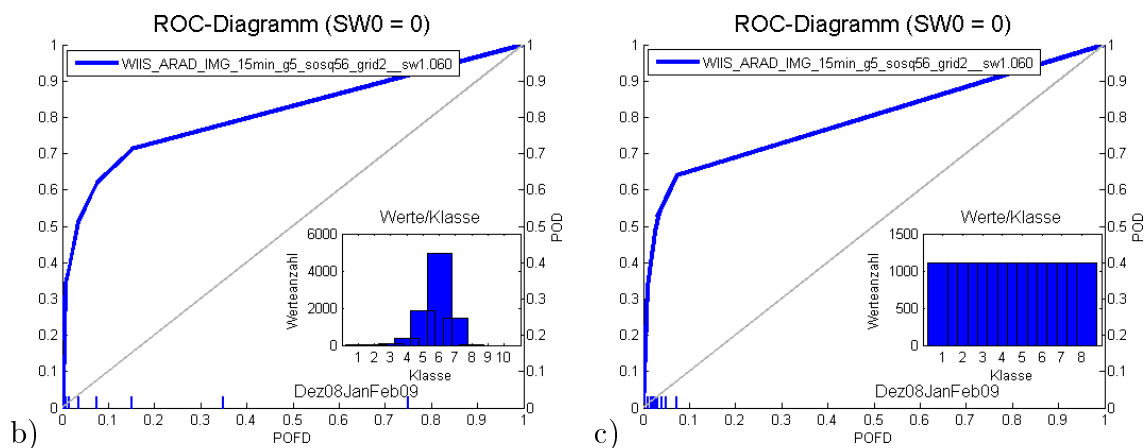
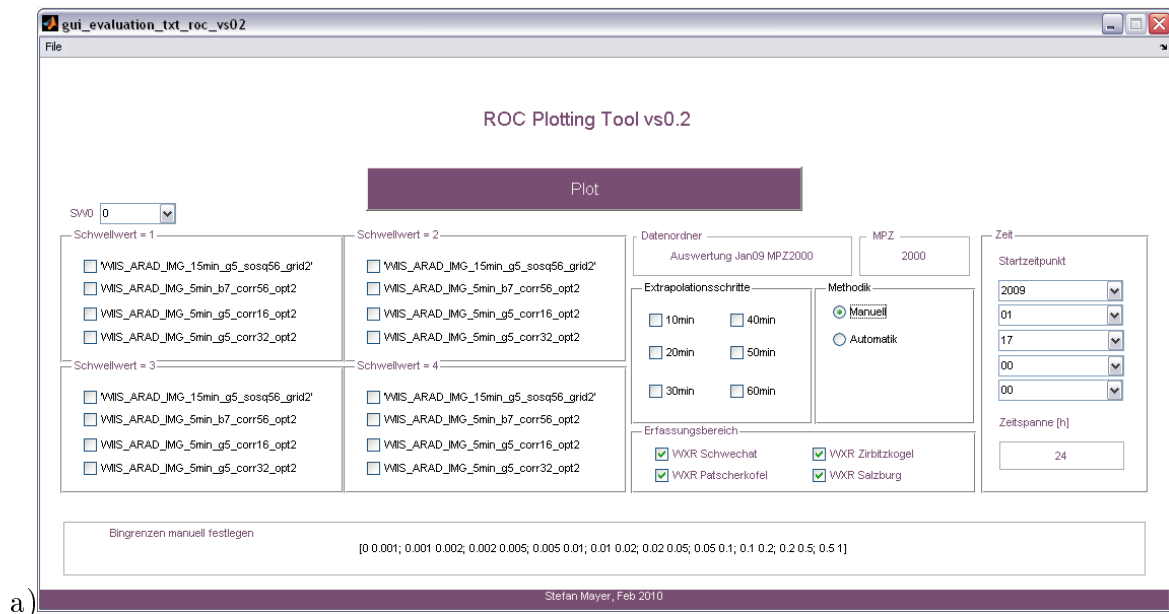


Abbildung 5.4:

ROC Plotting Tool: (a) Für mehr Flexibilität wurden die Parametersets an die Schwellwerteinstellung gebunden. Es sind somit freie Kombinationen zwischen Berechnungen mit verschiedenen Schwellwerten möglich. Zusätzlich zu den klassischen Eingabeoptionen müssen beim ROC-Diagramms die Bins definiert werden. Dies kann automatisch oder manuell erfolgen. (b) zeigt ein ROC-Diagramm, das im manuellen Modus erstellt wurde. Die Anzahl der Wertepaare in den Bins variiert stark. Fallen in eine Klasse nur sehr wenig Daten, so wird das Ergebnis zunehmend unsicher. (c) Aufgrund dieses Problems wurde der Automatikmodus geschaffen. Er setzt die Intervalle so, dass eine gleichmäßige Werteanzahlverteilung gewährleistet ist. In diesem Modus ist nur die Anzahl der Bins einstellbar.

auf der rechten Seite (siehe Abbildung 5.5 unten). Die einzelnen Parametersets und Schwellwerte können wieder beliebig kombiniert werden. Die einzige neue Einstelloption befindet sich, wie in Abbildung 5.5 ersichtlich, im Feld „Scores“. Hier sind alle möglichen Maßzahlen aufgelistet, wobei es sich bei den Auswahlkästchen der ersten Spalte nur um die Einträge in den Zellen der zugrunde liegenden Kontingenztafel handelt. Die zweite Spalte enthält die klassischen statistischen Momente, die dritte Maße der Kontingenztafel und durch die vierte Spalte erfolgt die Auswahl etwaiger Skill Scores (siehe Abschnitt 4.6). Um alle Werte in einer Kurve plotten zu können, ist es notwendig, sie innerhalb eines Extrapolationsschrittes zusammenzufassen. Das Programm bietet die Möglichkeit, den Mittelwert oder den Median der Werte zu berechnen.

5.2.2.3 Reliability-Diagramm

Für den Aufbau eines Reliability-Diagramms, das für die Verifikation einer WXR-Extrapolation, also einer deterministischen Vorhersage mit kategorischem Predikanden geeignet ist, sei auf Abschnitt 4.3 verwiesen. Auf Grund der Fülle an Informationen, die man daraus gewinnen kann, ist es für das Verifikationsschema unverzichtbar. Die Benutzeroberfläche ist in Abbildung 5.6a dargestellt. Sie enthält keine neuen Eingabeoptionen und konnte relativ kompakt und übersichtlich gestaltet werden. Zur Auswahl stehen die vier Parametersets, die Extrapolationsschritte sowie die gewohnten Optionen bezüglich Datenverwaltung, Erfassungsbereich und Verifikationszeitraum. Es sei erwähnt, dass auch bei der Auswahl von mehreren Parametersets oder Extrapolationsschritten nur eine Kurve geplottet wird. In solchen Fällen werden die Kombinationen zusammengefasst. Zusätzlich enthält das Diagramm (siehe Abbildung 5.6b) Information über die Anzahl der Werte pro Datenpunkt. Dies ist bei Intensität 0 durch einen gelben, ansonsten durch grüne Kreise dargestellt. Die Größe der Kreise ist proportional zu der Anzahl der Werte in diesem Punkt. Aufgrund der großen Anzahl an Bildpunkten, die mit Intensität 0 behaftet sind, wurde der gelbe Kreis im Vergleich zu den anderen verkleinert. Er ist daher symbolisch zu sehen.

5.2.2.4 Ranked Probabilities

Das letzte noch nicht besprochene Darstellungsprogramm liefert die Histogramme der bedingten Wahrscheinlichkeiten des allgemeinen Systems. Die Bedienoberfläche ist identisch aufgebaut wie beim „Reliability Diagram Plotting Tool“. Nachdem das

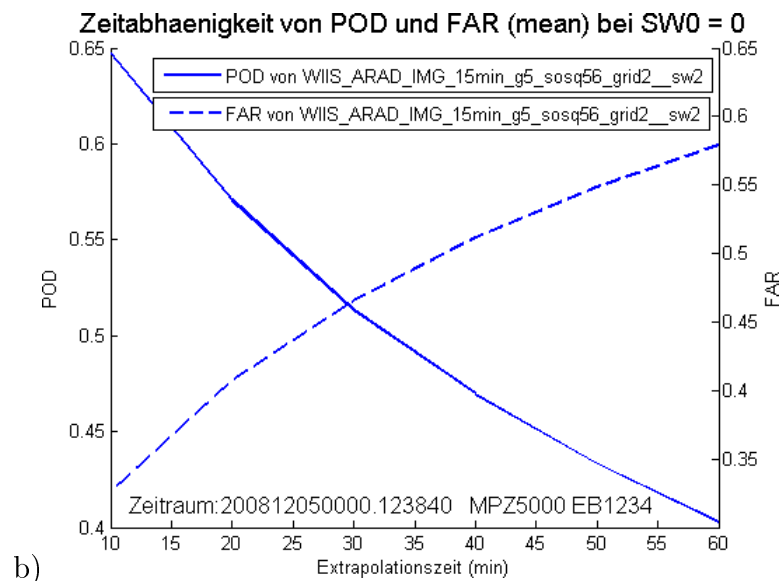
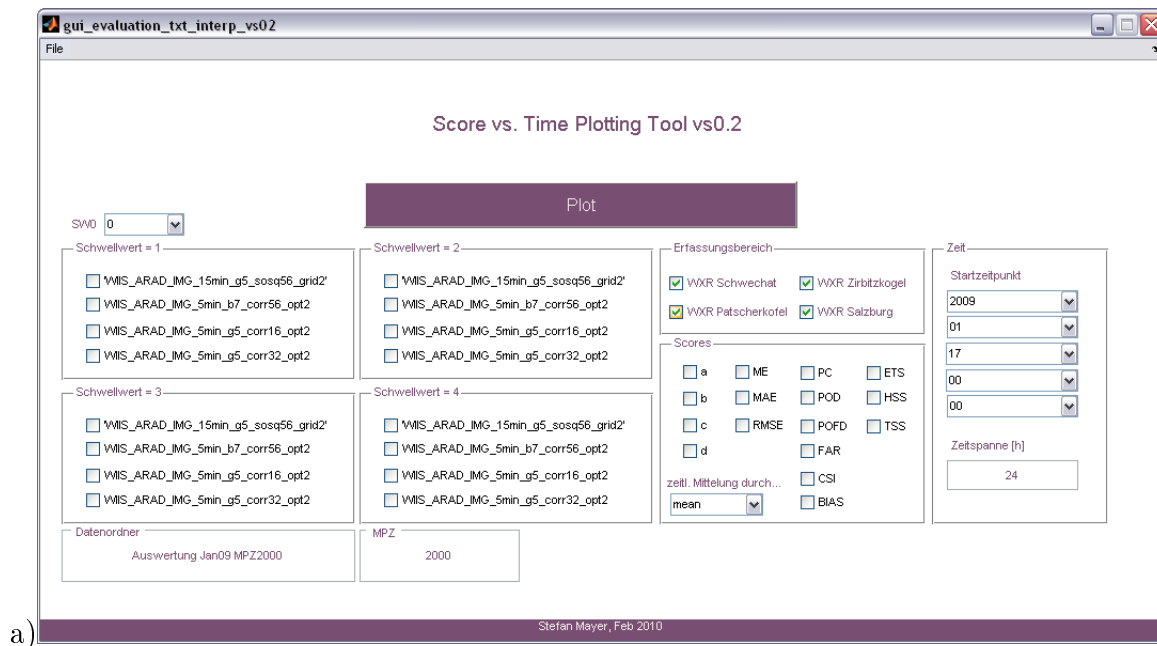


Abbildung 5.5:

Score vs. Time Plotting Tool: (a) Die Benutzeroberfläche des Time vs. Score Plotting Tools enthält neben den bekannten Einstelloptionen das Scores-Menü. So ist es möglich ein oder zwei Maße gegen die Extrapolationszeit zu plotten. (b) zeigt ein Beispiel für ein entstehendes Diagramm.

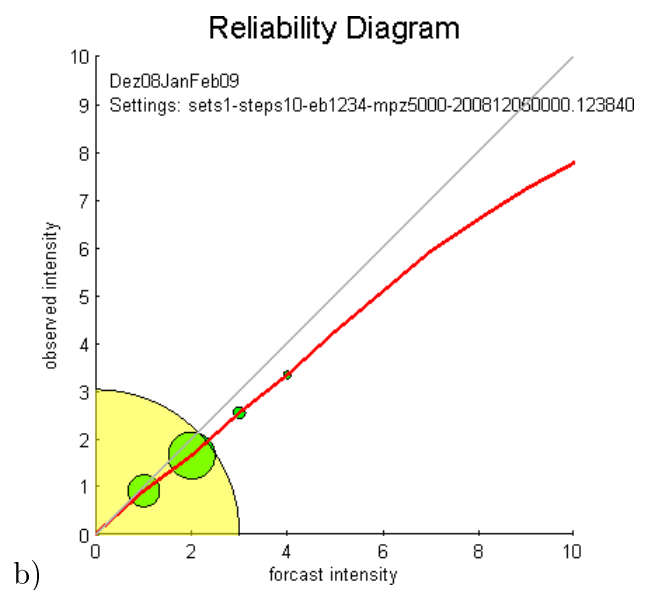
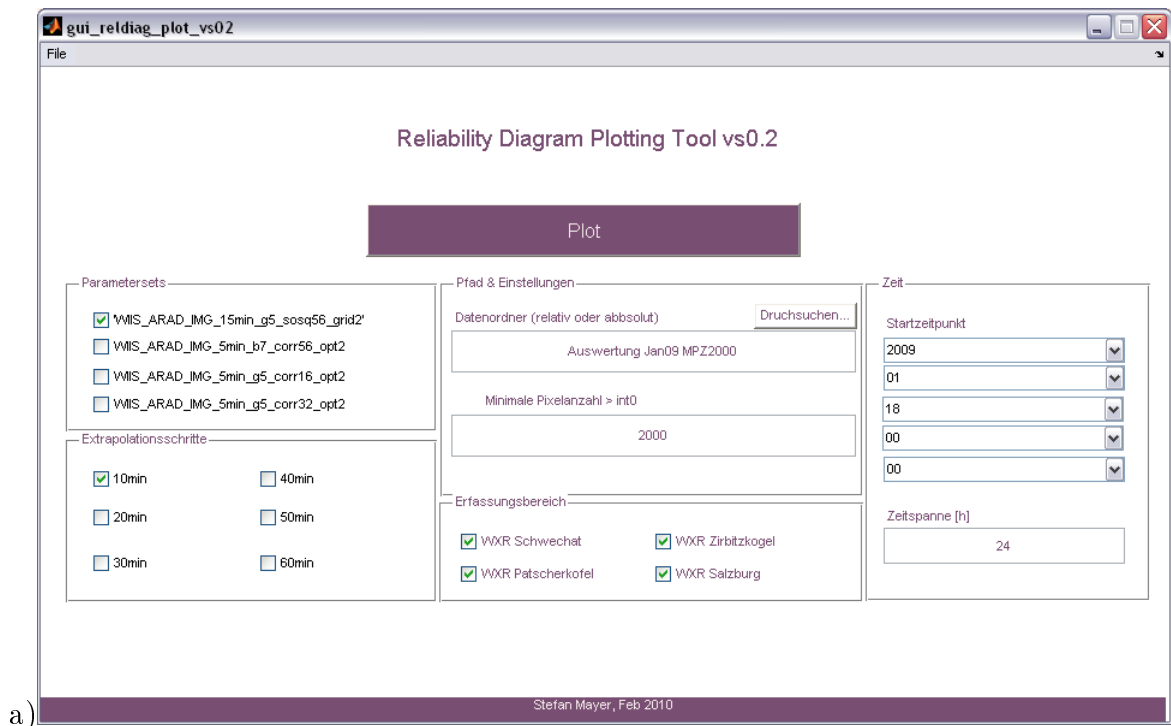
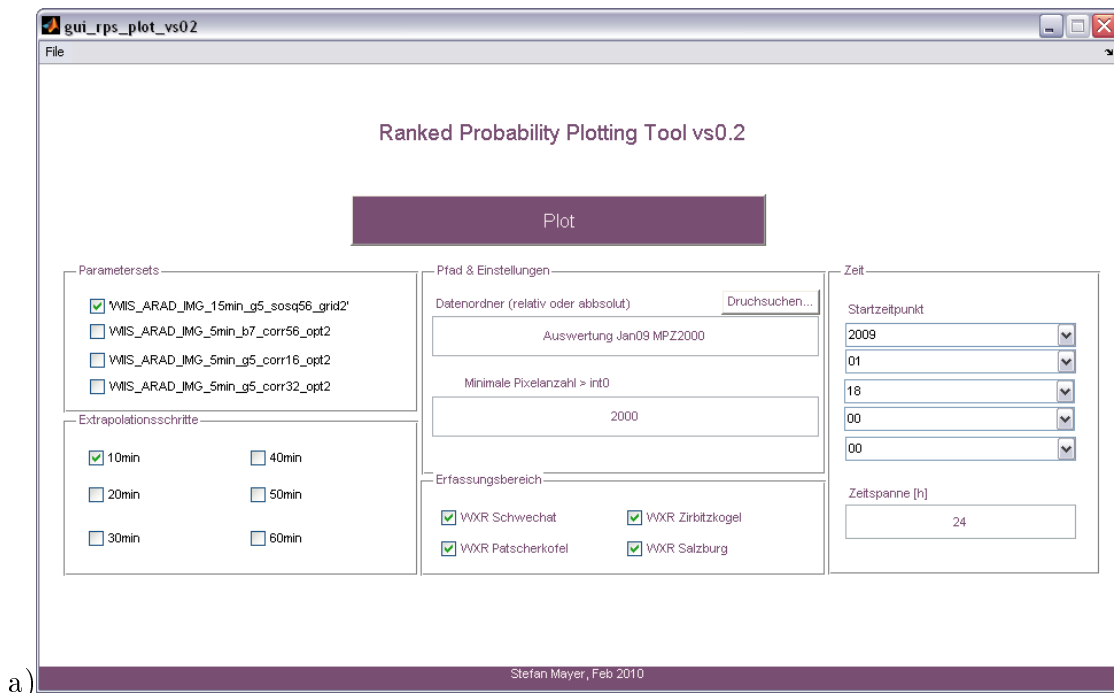
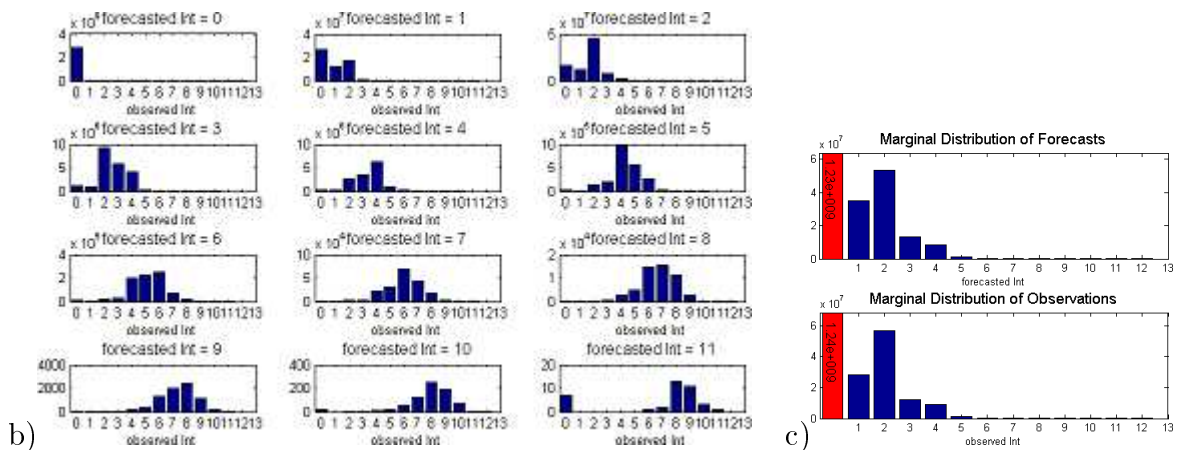


Abbildung 5.6: Reliability Diagram Plotting Tool

(a) Die Bedienoberfläche enthält die klassischen Elemente, die aus den anderen Programmen bereits bekannt sein sollten. (b) Das entstehende Reliability-Diagramm besteht zum einen aus der Reliability-Kurve und zum anderen aus den farbigen Kreisen, die Information über die Anzahl der Werte pro Datenpunkt geben. Je größer sie sind, desto mehr Bildpunkte wurden für den entsprechenden Datenpunkt zusammengefasst.



a)



b)

c)

Abbildung 5.7:

Ranked Probabilities Plotting Tool: (a) Das Darstellungsprogramm zu den bedingten Wahrscheinlichkeiten des allgemeinen Systems (ranked probabilities) ist analog zum „Reliability Diagram Plotting Tool“ aufgebaut. (b) Es entstehen immer zwei Plots zu je 12 Histogrammen. Somit könnten bedingte Beobachtungen genauso betrachtet werden wie bedingte Vorhersagen, es wird also Information zur Reliability und zur Discrimination des Vorhersagesystems geliefert (siehe Abschnitte 4.3 und 4.5). An dieser Stelle wurden nur die Verteilungen der bedingten Vorhersagen geplottet. Das Layout des zweiten Produktes, der bedingten Beobachtungen ist analog dazu. (c) Zusätzlich werden auch die marginalen Verteilungen der Beobachtungen und Vorhersagen dargestellt. Die Intensitätsklasse 0 wurde dabei farblich hervorgehoben, weil die Höhe ihres Histogrammbalkens reduziert dargestellt wird, um die andern Klassen besser auflösen zu können.

Programm mittels des „Plot-Buttons“ gestartet wurde, werden immer zwei Produkte generiert. Erstens werden Histogramme für die Vorhersagen bei gegebener Beobachtung, $p(y/o)$, generiert und zweitens auch für die umgekehrte Betrachtungsweise, also die Beobachtungen bei gegebener Vorhersage, $p(o/y)$. Der zweite Fall ist in Abbildung 5.7 (unten) dargestellt. Jedes Produkt enthält ein Histogramm für jede Intensitätsstufe (je nach dem von Vorhersage oder Beobachtung), somit insgesamt 14 Diagramme, wobei die beiden höchsten Intensitäten (Stufe 12 und 13) wegen ihrer geringen Aussagekraft, aufgrund der wenigen Werte in diesen Bereichen, nicht dargestellt werden.

6 Ergebnisse

Die in Kapitel 5 beschriebene Verifikationssoftware wurde nun verwendet, um jene WXR-Extrapolationen, die beim österreichischen Flugwetterdienst AustroControl GmbH zur Zeit routinemäßig berechnet werden, zu vergleichen und untersuchen. Zu diesem Zweck wurde ein Datensatz von 5.Dezember 2008 bis 30.September 2009 herangezogen. Er enthält vier verschiedene TREC-Extrapolationen für jeden Zeitschritt, denen unterschiedliche Parametereinstellungen zu Grunde liegen. Die Vorhersagedauer beträgt 10 bis 60 Minuten. Da täglich um Mitternacht die Ground-Clutter-Filter neu initialisiert werden, weisen die entsprechenden WXR-Bilder starke Störechos auf. Sie wurden daher nach den ersten Versuchen aussortiert. Der TREC-Algorithmus an sich sowie die einzelnen Parameter wurden bereits in Kapitel 2 näher beschrieben. Weiterführende Informationen zu den verwendeten Daten, insbesondere die Dateinamenskonventionen sind dem Anhang 7 zu entnehmen.

Nummer	Parameterkombination (PS) und Farbgebung
1	WIIS_ARAD_IMG_15min_g5_sosq56_grid2_*.png
2	WIIS_ARAD_IMG_5min_b7_corr56_opt2_*.png
3	WIIS_ARAD_IMG_5min_g5_corr16_opt2_*.png
4	WIIS_ARAD_IMG_5min_g5_corr32_opt2_*.png

Tabelle 6.1:

Farbgebung und Nummerierung der vier Parameterkombinationen in Kapitel 6: Um die Übersichtlichkeit zu gewährleisten, bleibt die Farbgebung und Nummerierung der vier zu vergleichenden Parametersets innerhalb der Auswertung konstant.

Der Datensatz wurde zur näheren Betrachtung in eine winterliche Periode mit hauptsächlich advektiv dominierten Prozessen (Dezember 2008 bis Februar 2009) sowie in eine sommerlich konvektive Periode (Juni 2009 bis August 2009) unterteilt. Diese Zeiträume zu je drei Monaten wurden daraufhin jeweils für den gemeinsamen Erfassungsbereich aller vier WXR-Stationen des WXR-Austria-Composite (siehe Abschnitt 2.2.1) sowie nur für die WXR-Station Wien/Raichenwarth getrennt analysiert. Die daraus resultierenden Ergebnisse sollen nun zusammengefasst vorgestellt werden.

Um den Vergleich der verschiedenen Parameterkombinationen zu vereinfachen, wurden die einzelnen Parametersets (PS) in den Diagrammen in unterschiedlichen Farben dargesellt, diese bleiben innerhalb dieses Kapitels unverändert und sind Tabelle 6.1 zu entnehmen.

6.1 Auswertung Winter 2008-2009

Als erstes sollen die Produkte des „Score vs. Time Series“-Plotprogramms, beginnend mit den klassischen, statistischen Momenten betrachtet werden. Abbildung 6.1 zeigt den mittleren Fehler ($ME_{\text{allgemein}}$) im Vergleich zum mittleren absoluten Fehler (MAE), jeweils aus dem nicht vereinfachten (allgemeinen) System berechnet (siehe auch Abschnitt 4.1, sowie Abschnitt 4.2).

Abbildung 6.1a beinhaltet Information aus dem gesamten Erfassungsbereich aller vier WXR-Stationen, die in das WXR-Austria-Composite eingehen. Abbildung 6.1b hingegen behandelt nur den Erfassungsbereich des WXR Wien/Rauchenwarth. Man erkennt sofort, dass sowohl der mittlere Fehler (ME) als auch der mittlere absolute Fehler (MAE) im Bereich des WXR Schwechat weitgehend über dem Austria-Composite Durchschnitt liegen. Die naheliegendste Erklärung hierfür dürfte die große Echodichte in diesem Gebiet sein. Der Erfassungsbereich des WXR Schwechat ist jener, der am geringsten durch orographische Hindernisse durchsetzt ist und daher im Vergleich zu den anderen drei Erfassungsbereichen die geringsten Erfassungslücken (siehe Abbildung 3.11) aufweist. Daraus resultiert bei ungefähr gleicher Niederschlagshäufigkeit (die zumindest im Bezug auf die Größenordnung gegeben ist) die größte Echodichte. Somit ermöglicht dieser Vergleich keine Aussage bezüglich geographisch bedingter Qualitätsunterschiede in den Extrapolationen, er zeigt jedoch nochmals, dass jede Extrapolation uneingeschränkt auf qualitativ hochwertige, vollständige WXR-Bilder angewiesen ist.

Betrachtet man die Entwicklung des Bias mit zunehmender Extrapolationsdauer (Abbildung 6.1), so sticht die Parameterkombination 1 heraus. Sie ist die einzige, die im Schnitt einen positiven mittleren Fehler produziert, somit die Niederschlagsfelder tendenziell etwas zu groß, beziehungsweise mit leicht überschätzten Intensitäten vorhersagt. Die anderen Extrapolationstypen prognostizieren in ihrer Intensität und Ausdehnung leicht unterschätzte Niederschlagsregime. Um diese deutlichen Unterschiede zwischen den einzelnen Extrapolationstypen detaillierter zeigen zu können, wurden in Abbildung 6.2 die marginalen Verteilungen der 60 minütigen Vorhersagen von PS

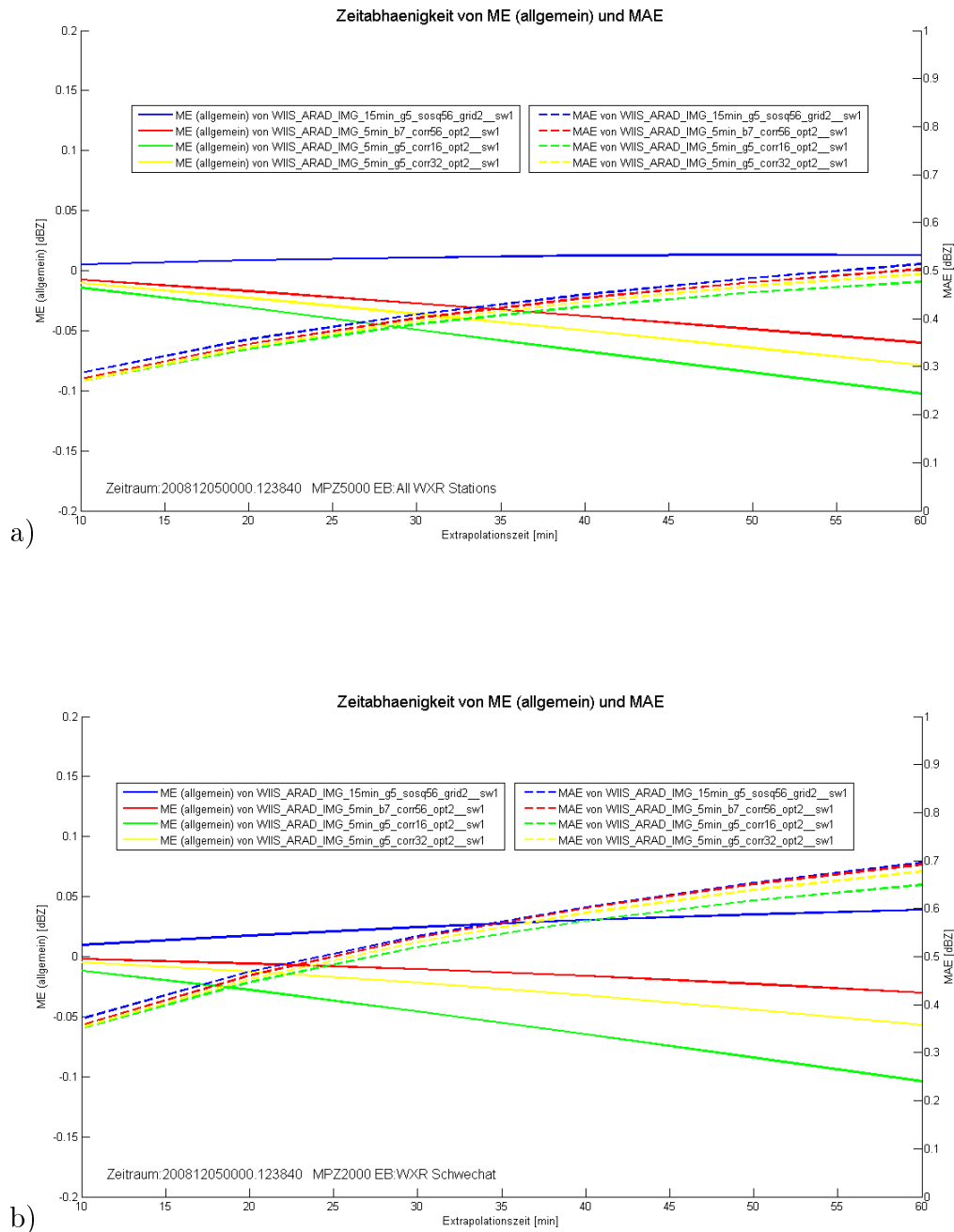


Abbildung 6.1:

Vergleich von ME und MAE aller vier Parameterkombinationen (Winter): Vor allem der ME zeigt deutliche Unterschiede zwischen den Extrapolationstypen auf. PS 1 f"uhrt als einzige Extrapolation zu einem positiven ME, PS 2 bis PS 4 unterscheiden sich durch die Gr"o"e des Fehlers, jedoch nicht durch dessen Vorzeichen. (a) F"ur den Vergleich wurde das gesamte Gebiet des WXR-Austria-Composite (alle vier Stationen) herangezogen. Dieses Gebiet beinhaltet viele (orographische) Abschattungsgebiete. Die Echodichte ist daher recht gering. (b) Hier wurde nur der Erfassungsgebiet des WXR Wien/Rauchenwarth betrachtet. Aufgrund der h"oheren Echodichte ergeben sich etwas h"oherer Fehlerma"e, die jedoch keine Aussage "uber die regionale Extrapolationsqualit"at im Vergleich zur durchschnittlichen Qualit"at im gesamten WXR-Austria-Composite Gebiet zulassen.

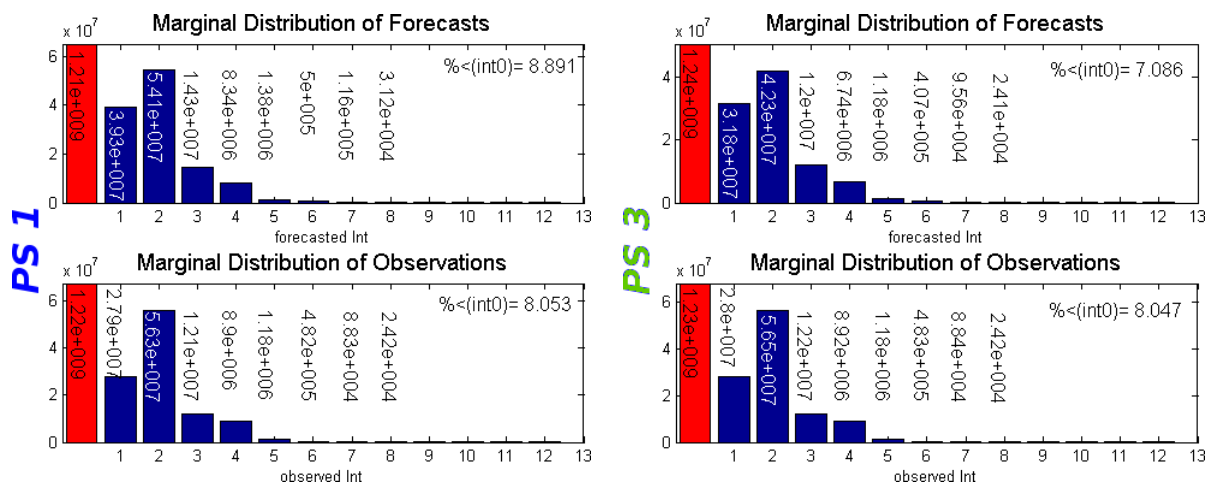


Abbildung 6.2:

Marginale Verteilungen von 60 minütiger Vorhersage und Beobachtung (Winter): Vergleich der marginalen Verteilungen der Vorhersagen mit PS 1 und PS 3 (für den Erfassungsbereich des WXR Wien/Rauchenwarth) mit den dazugehörigen marginalen Beobachtungsverteilungen. Die geringen Unterschiede in den Beobachtungsdaten entstehen, weil nicht jede Extrapolationsart zu jenem Zeitpunkt berechnet wurde und sich somit der Vergleichsdatensatz zwischen den Extrapolationstypen leicht unterscheidet. Man erkennt die leichte Überschätzung der Niederschlagsregime („nasse“ Vorhersage) durch das PS 1, sowie eine leichte Unterschätzung durch das PS 3 („trockene“ Vorhersage). Außerdem fällt auf, dass die Häufigkeit der Intensitätsklasse 1 bei beiden Extrapolationstypen überschätzt wird.

1 und PS 3 für den Erfassungsbereich des WXR Wien/Rauchenwarth im Vergleich zu den passenden Beobachtungsverteilungen geplottet. Man erkennt, dass die Anzahl jener Pixel, die Intensitätsstufe 1 aufweisen, bei beiden gezeigten Extrapolationstypen überschätzt wird, wobei dies bei PS 3 wesentlich geringer ausfällt. Bei höheren Intensitäten werden die Unterschiede noch deutlicher sichtbar. PS 1 neigt dazu, wie auch durch den ME in Abbildung 6.1 gezeigt, zu „nass“ zu prognostizieren. PS 3 hingegen, ist jener Extrapolationstyp, der den größten negativen Bias aufweist und somit tendenziell zu „trockene“ Vorhersagen liefert.

In Abbildung 6.1 erkennt man außerdem, vor allem wenn man den gesamten Erfassungsbereich aller vier WXR-Stationen (Abbildung 6.1a) heranzieht, dass der mittlere Fehler des PS 1, im Gegensatz zu den anderen mit zunehmender Vorhersagedauer kaum anwächst. Dies lässt den Schluss zu, dass dieser Extrapolationstyp die grundsätzliche zeitliche Entwicklung der Struktur und Intensitätsverteilung der Niederschlagsregime am besten erfasst. Jedoch unter der zuvor besprochenen Annahme,

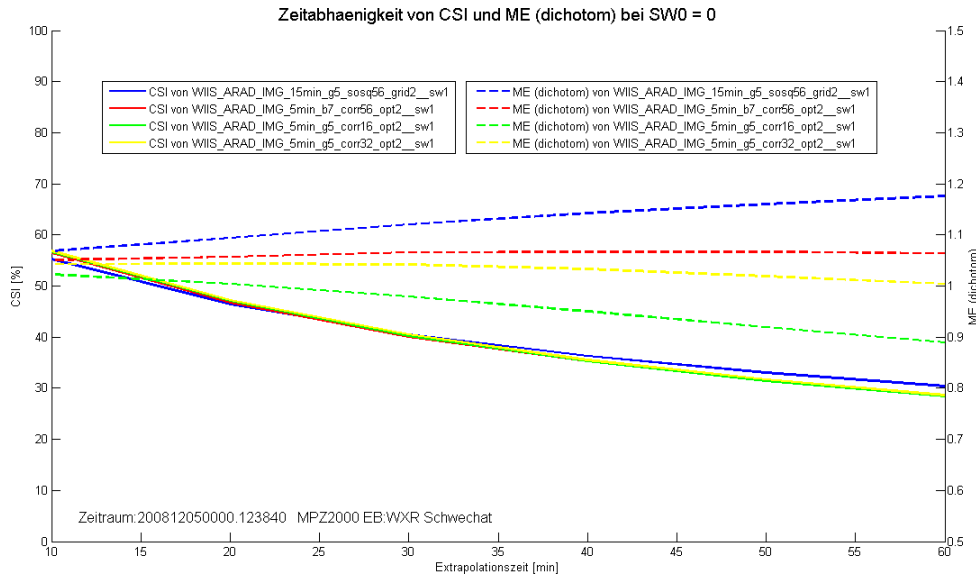


Abbildung 6.3:

CSI und ME des dichotomen Systems (Winter): PS 1 weist zufolge des CSI die höchste Accuracy, vor allem bei längeren Extrapolationen auf. Dass dieser Vorteil jedoch mit einer „nassen“ Prognosecharakteristik einhergeht, zeigt sich in den Kurven des mittleren Fehlers des dichotomen Systems ME(dichotom). Die Accuracy aller anderen Parameterkombinationen liegt auf ähnlichem Niveau. Somit wäre von diesen in Summe PS 4 zu bevorzugen, da es den geringsten mittleren Fehler zeigt.

dass die Betrachtung des Erfassungsbereichs des WXR Wien/Rauchenwarth für sich, aufgrund der größeren Echodichte repräsentativere Ergebnisse liefern kann, sei an dieser Stelle auch erwähnt, dass in diesem Gebiet der Betrag des mittleren Fehlers von PS 2 über alle Vorhersageschritte kleiner ausfällt als jener des PS 1. Vor allem bei kurzen Extrapolationsschritten (bis 20 Minuten) wären laut Abbildung 6.1b PS 2 oder auch PS 3 zu bevorzugen. Bei kurzer Extrapolationsdauer weisen diese außerdem, im Vergleich zu PS 1, einen leicht geringeren mittleren absoluten Fehler (MAE) auf. Der MAE liegt in beiden Diagrammen bei allen Extrapolationstypen deutlich über dem Betrag des jeweiligen ME. Dies deutet darauf hin, dass ein großer Teil der Abweichung zwischen Extrapolation und Beobachtung durch fehlerhafte Ortsvorhersagen zu Stande kommt.

Ergänzend zu dem in Abbildung 6.1 geplotteten ME des allgemeinen Systems findet man in Abbildung 6.3 den ME des vereinfachten, dichotomen Systems. Durch die marginalen Verteilungen in Abbildung 6.2 wurde erkannt, dass alle Extrapolationstypen die Häufigkeit der Intensitätsklasse 1 überschätzen. Diese Eigenschaft gewinnt

durch eine Reduzierung des Systems auf dichotome Verhältnisse an Gewicht, weil die falsche Vorhersage hoher Intensitäten wenig oder gar nicht berücksichtigt wird. Somit weist im vereinfachten System ausschließlich PS 3 einen mittleren Fehler kleiner 1 auf (siehe Abbildung 6.3). Der dichotom berechnete ME stellt ein Maß für die Ausdehnung der vorhergesagten Niederschlagsregime dar. Liegt er unter dem Wert 1 (PS 3), so wird die Niederschlagsausdehnung im Mittel unterschätzt, bei Werten größer als 1 (PS 1 und PS 2) werden die Niederschlagsgrenzen zu weitläufig prognostiziert. Ist also die Ausdehnung eines Niederschlagsgebietes von Bedeutung, so ist PS 4 zu bevorzugen. Die entsprechende Kurve in Abbildung 6.3 liegt für alle Vorhersageschritte nahe dem Wert 1, die Größe der Niederschlagsregime wird also weitgehend richtig prognostiziert.

Weiters wurde in Abbildung 6.3 der CSI (siehe Abschnitt 4.2) geplottet. Dieser ist das wohl gängigste Maß für die Accuracy eines dichotomen Systems und kann vergleichend zum MAE aus Abbildung 6.1 betrachtet werden. Die kombinierte Darstellung zusammen mit dem ME des dichotomen Systems wurde gewählt, weil eine generelle Über- oder Unterschätzung der Niederschlagsregime im CSI nicht berücksichtigt wird. Die einzelnen Extrapolationstypen unterscheiden sich aber wie schon beim MAE des allgemeinen Systems wenig. Einzig PS 1 ist hervorzuheben, dessen CSI bei sehr kurzen Extrapolationen leicht unter den anderen liegt. Mit zunehmender Extrapolationsdauer schneidet diese Parameterkombination relativ zu den anderen aber immer besser ab und ist im Falle einer 50 oder 60 minütigen Extrapolation leicht zu bevorzugen. Diese Erkenntnis kann gut durch die Zeitspanne zwischen den für die Extrapolation genutzten Ausgangsbildern erklärt werden. Sie beträgt im Fall von PS 1 15 Minuten, bei allen anderen nur 5 (siehe Abschnitt 2.2). Somit werden bei PS 1 kleinskalige, zeitliche Fluktuationen besser gefiltert und damit die Grundbewegung der Systeme besser erfasst und in weiterer Folge das Ergebnis mit zunehmender Extrapolationsdauer verbessert. Bei sehr kurzen Extrapolationen (10 bis 20 Minuten) wirkt sich dieser Informationsverlust jedoch auch leicht negativ aus.

Bis jetzt wurden in diesem Kapitel ausschließlich die Gütemerkmale Accuracy (CSI, MAE) und Bias (ME) behandelt. Abbildung 6.4 versucht nun die Zuverlässigkeit (Reliability, siehe Abschnitt 4.3) und Unterscheidungskompetenz (Discrimination, siehe Abschnitt 4.5), anhand des dichotomen Systems zu beschreiben. Eine klassische Methode hierfür ist die gemeinsame Betrachtung der POD und FAR. Die ausführliche Beschreibung dieser Maße ist wiederum den Abschnitten 4.3 sowie 4.5 zu entnehmen.

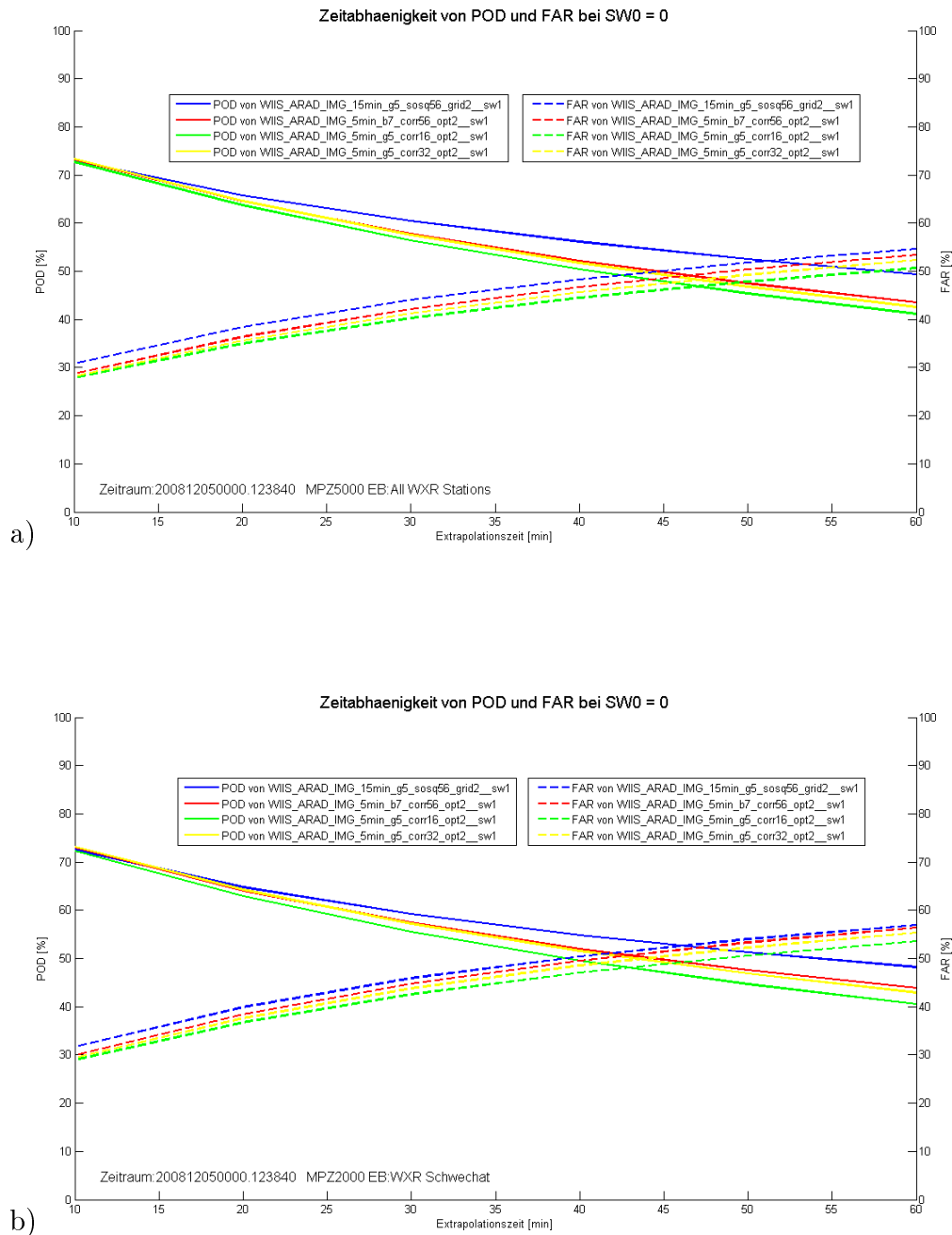


Abbildung 6.4:

POD und FAR Entwicklung (Winter): Für das gesamte WXR-Austria-Composite (a) sowie für den Erfassungsbereich des WXR Wien/Raasdorf (b) wurde die Reliability (FAR) und die Discrimination (POD) des dichotomen Systems berechnet. Zwischen den betrachteten Gebieten zeigen sich nur wenig Unterschiede. Aus den vier Parameterkombinationen sticht PS 1 hervor. Es zeigt im Vergleich eine deutliche höhere POD. Leider aber auch die höchsten FAR-Werte. Zusammen deutet dies auf eine zu „nasse“ Prognosecharakteristik hin.

Auch an dieser Stelle soll wieder der Erfassungsbereich des WXR Wien/Rauchenwarth zum gesamten WXR-Austria_Composite Gebiet verglichen werden. Im Gegensatz zu bisherigen Betrachtungen der Accuracy und des Bias sind bezüglich der Reliability sowie der Discrimination kaum Unterschiede zwischen dem WXR Wien/Rauchenwarth und dem Durchschnitt über alle vier WXR-Stationen auszumachen. Dies liegt hauptsächlich daran, dass hier die Gruppe der korrekt prognostizierten Nicht-Niederschläge (correct rejections, siehe Abbildung 4.1) und auch die Gesamtpixelzahl des betrachteten Gebiets nicht in die Berechnung von POD und FAR (Gleichungen 4.8a und 4.15) eingeht. Somit spielt die Echodichte keine wesentliche Rolle.

Die, wie bereits festgestellt, leicht nasse Prognosecharakteristik des PS 1 spiegelt sich in den Kurven in Abbildung 6.4 wieder. Diese Parameterkombination weist einerseits die mit Abstand höchste POD, andererseits aber auch die höchsten FAR Werte auf. Trotzdem dürfte vor allem bei längeren Extrapolationszeiten (40 bis 60 Minuten) der Vorteil der höheren POD den Ausschlag zu Gunsten dieses Extrapolationstyps geben, da hier die POD deutlich über den der anderen Extrapolationen liegt. Der Unterschied bezüglich der Reliability (FAR) fällt eindeutig geringer aus.

Betrachtet man zusätzlich das ROC-Diagramm für den Erfassungsbereich der WXR-Station Wien/Rauchenwarth für 60 minütige Extrapolationen in Abbildung 6.5 so bestätigen sich die Ergebnisse bezüglich der Discrimination aus Abbildung 6.4b. Bei dieser Darstellung wird die Entdeckungswahrscheinlichkeit für Niederschlagsereignisse (POD) der Zuverlässigkeit von prognostiziertem Nicht-Niederschlagsereignissen (POFD) gegenübergestellt (siehe Abschnitt 4.5). Somit werden beide wesentlichen Maße zur Beschreibung der Discrimination eines dichotomen Systems kombiniert. Je größer die Fläche unterhalb der ROC-Kurve, desto besser ist die Fähigkeit des Extrapolationstyps, Niederschlag und Nicht-Niederschlag voneinander abzugrenzen, ausgeprägt. Auch hier zeigt PS 1 die besten Ergebnisse und ist den anderen Parameterkombinationen vor allem durch höhere POD-Werte (bei ungefähr gleicher POFD) überlegen. Daher soll in Abbildung 6.6 genau dieser Extrapolationstyp für unterschiedliche Extrapolationsschritte weiter untersucht werden:

Es stellt sich das zu erwartende Ergebnis ein. Die ROC-Kurve nähert sich mit zunehmender Extrapolationszeit etwas mehr der 45°-Diagonale an, das heißt die Fähigkeit zur Abgrenzung der Ereignisse nimmt ab. Abbildung 6.6 zeigt den Vergleich von 10, 30, sowie 60 minütigen PS 1-Extrapolationen für den Erfassungsbereich des WXR Wien/Rauchenwarth.

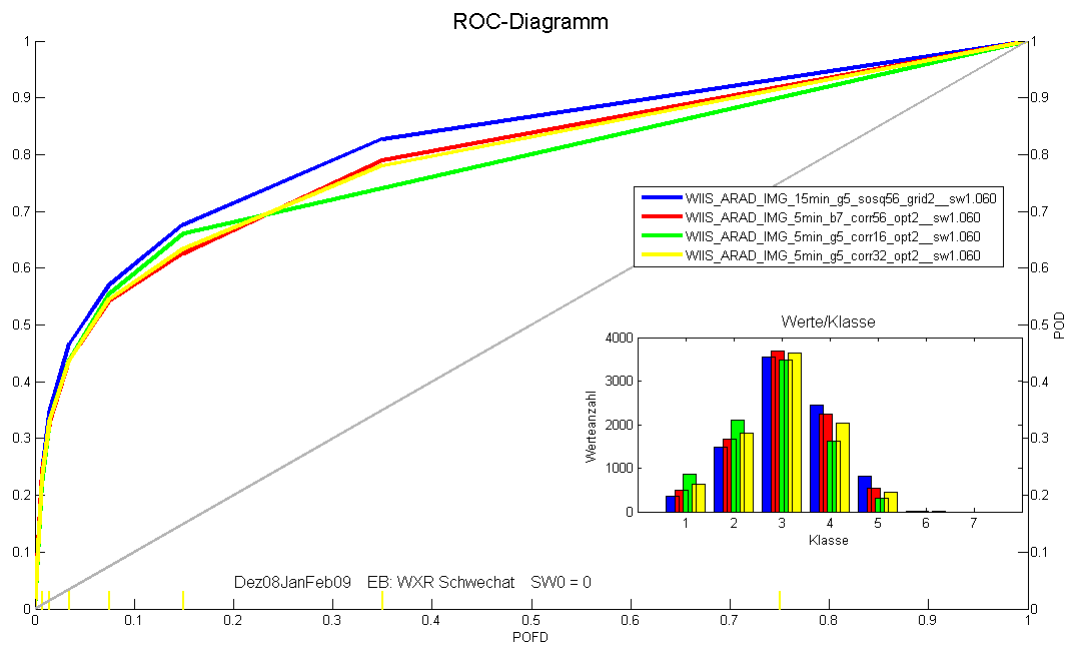


Abbildung 6.5:

ROC-Diagramm: PS-Vergleich (Winter): Vergleich der ROC-Kurven aller vier Parameterkombinationen für 60 minütige Extrapolationen. Das Diagramm wurde im manuellen Modus erstellt, daher ist die Werteanzahl in den einzelnen Klassen unterschiedlich. Wieder zeigt sich, dass PS 1 die beste Discrimination aufweist. Im Subdiagramm ist ergänzend die Anzahl der Werte je Klasse dargestellt, die Klassenmittel kann man durch gelbe Markierungen an der Abszisse erkennen.

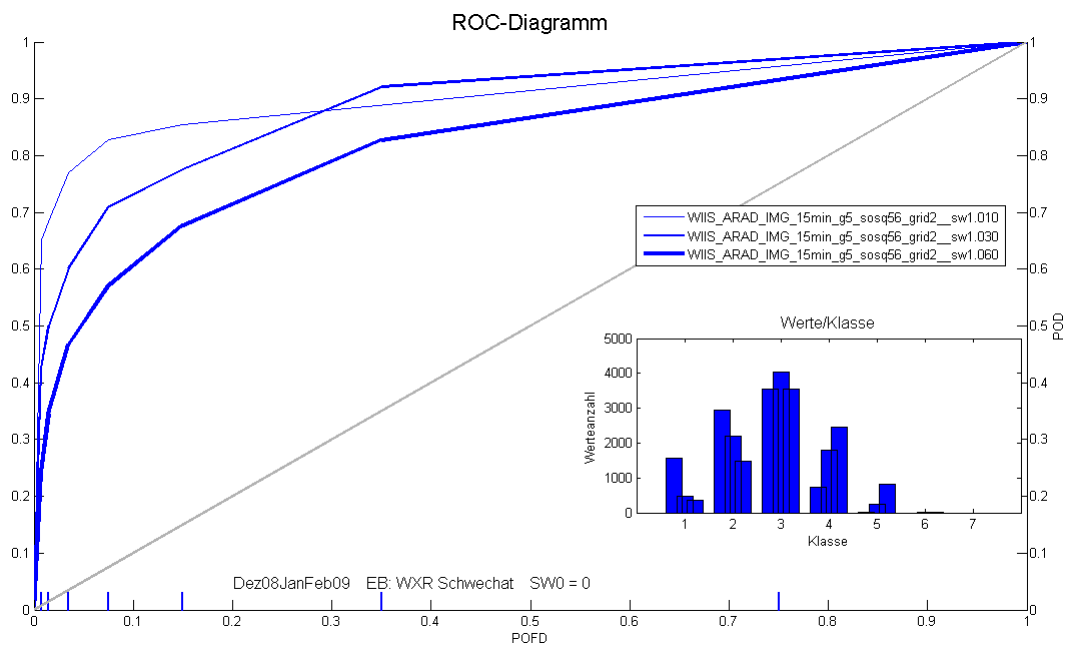


Abbildung 6.6:

ROC-Diagramm für PS 1 (Winter): Vergleich von ROC-Kurven für unterschiedliche Extrapolationsschritte: Für das PS 1 (im Erfassungsbereich des WXR Wien/Raichenwarth) werden die 10, 30 und 60 minütigen Extrapolationen bezüglich ihrer Discrimination verglichen. Es zeigt sich der erwartungsgemäße Rückgang des Qualitätsmerkmals mit zunehmender Vorhersagedauer.

Um nun auch die Reliability der einzelnen Extrapolationstypen näher zu untersuchen, sollen an dieser Stelle einige Reliability-Diagramme geplottet werden:

Abbildung 6.7 beinhaltet die Reliability-Diagramme des PS 1 für unterschiedliche Vorhersageschritte. Man erkennt eine abnehmende Zuverlässigkeit der Vorhersagen mit zunehmender Vorhersagedauer. Für alle Vorhersageschritte liegt die Reliability-Kurve ab der Intensitätsstufe 1 generell unterhalb der 45°-Diagonale. Dies deutet auf eine „nasse“-Vorhersagecharakteristik (Überschätzung des Niederschlags, $ME > 0$) hin und trifft auf alle Parameterkombinationen zu (siehe Abbildung 6.8). Es wurde aber bereits zu Beginn dieses Kapitel in der Abbildung 6.1 gezeigt, dass der mittlere Fehler (ME) keineswegs immer im positiven Bereich liegt. Bei den Parameterkombinationen PS2, PS3 und PS4 fällt er sogar deutlich negativ aus.

Die Ergebnisse aus Abbildung 6.7 und Abbildung 6.8 kann daher nur die schiefe Werteverteilung zwischen den Niederschlagsklassen erklären. Während die Intensitätsstufe 0 im Radarbild in der Regel dominiert, kommen Intensitäten größer Stufe 8 kaum vor. Die Reliabilitykurve des PS1 (Abbildung 6.7) liegt bei Intensität 0 knapp oberhalb der Diagonale. Dieser Datenpunkt beinhaltet einen Großteil aller zu hoch vorhergesagten Intensitäten, die jedoch durch die gewaltige Anzahl der richtig prognostizierten Nichtniederschläge überschattet werden. Die zu nasse Charakteristik ab Intensität 1 erklärt sich durch unzureichende Ortsvorhersagen der Niederschlagsregime. Da im Reliability-Diagramm die bedingten Wahrscheinlichkeiten der Beobachtungen bei gegebener Vorhersageintensität ($p(o/y)$) betrachtet werden, ergibt sich durch eine „Verschiebung“ zwischen Vorhersage- und Beobachtungsniederschlagsfeldern im Reliability-Diagramm das Bild einer „nassen“ Vorhersagecharakteristik. Dieser Eindruck wird auch durch die Plots der bedingten Beobachtungen und Vorhersagen für PS 1 in Abbildung 6.9 bestätigt. Aber, wie bereits in Abbildung 6.1 gezeigt wurde, ist ausschließlich beim PS 1 eine eindeutig „nasse“ Charakteristik zu beobachten. Aufgrund der schiefen Werteverteilung liefert das Reliability-Diagramm keine glaubwürdige Information über den Bias der Vorhersagen. Es können aber sehr wohl wertvolle Aussagen über die Reliability der einzelnen Parameterkombinationen gewonnen werden. Wie schon beim dichotomen System (Abbildung 6.4) weist die PS 3 die besten Reliability-Werte auf. Diese sind aber hauptsächlich durch die „trockene“ Vorhersagecharakteristik zu erklären. Die tendenziell schlechtesten Ergebnisse liefert erwartungsgemäß PS 1, vor allem im Bereich der Intensitätsstufen 2 bis 5.

Noch genauere Erkenntnisse über die Reliability und Discrimination der Extra-

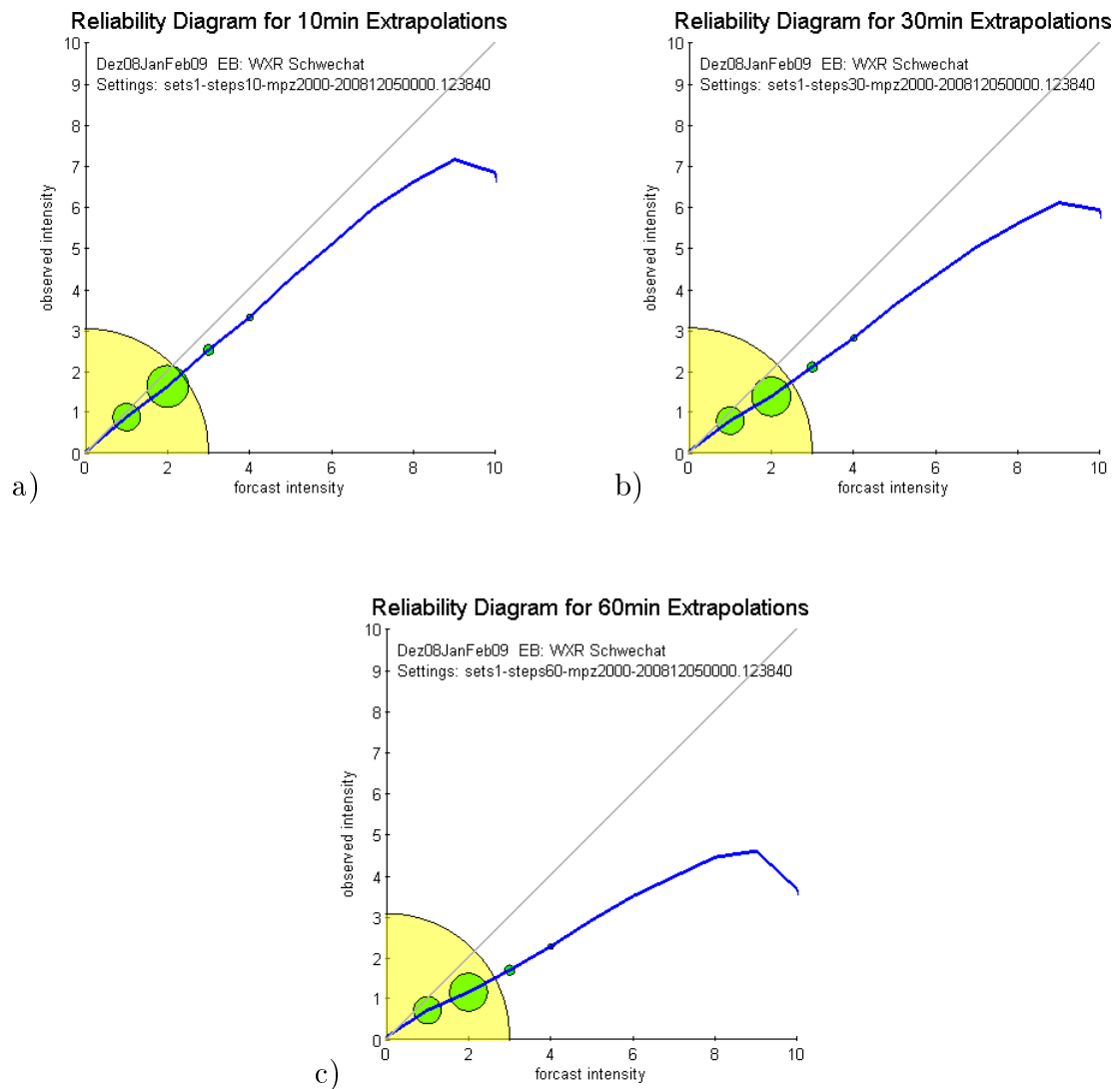


Abbildung 6.7:

Reliability bei unterschiedlichen Extrapolationsschritten (Winter): Reliability Diagramme des PS1 für die Vorhersageschritte: a) 10 Minuten; b) 30 Minuten; c) 60 Minuten. Die Reliability nimmt erwartungsgemäß mit zunehmender Vorhersagedauer ab. Die Diagramme lassen für alle Zeitschritte eine zu „nasse“ Vorhersage vermuten.

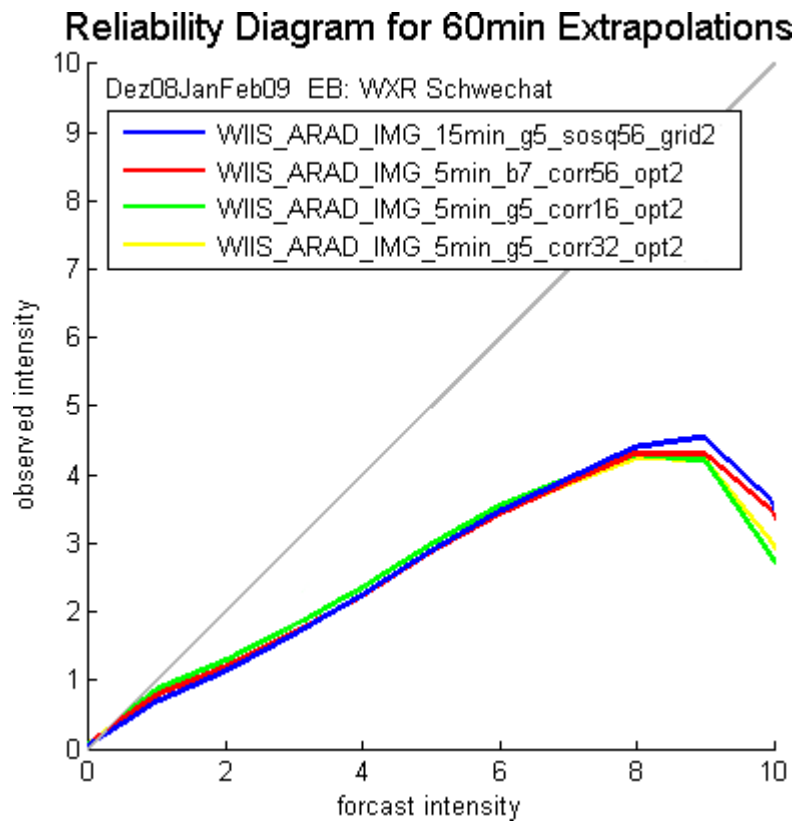


Abbildung 6.8:

Reliability-Vergleich der Parameterkombinationen (Winter): Es werden die 60 minütigen Extrapolationen der unterschiedlichen Parameterkombinationen für den Erfassungsbereich des WXR Wien/Rauchenwarth bezüglich ihrer Reliability verglichen. Man erkennt nur kleinste Unterschiede zwischen den Extrapolationstypen. PS 3 weist bei mittleren Intensitäten im Vergleich eine leicht höhere Reliability auf. Die Werte bei den Intensitäten größer als 8 sind aufgrund der geringen Werteanzahl nicht repräsentativ.

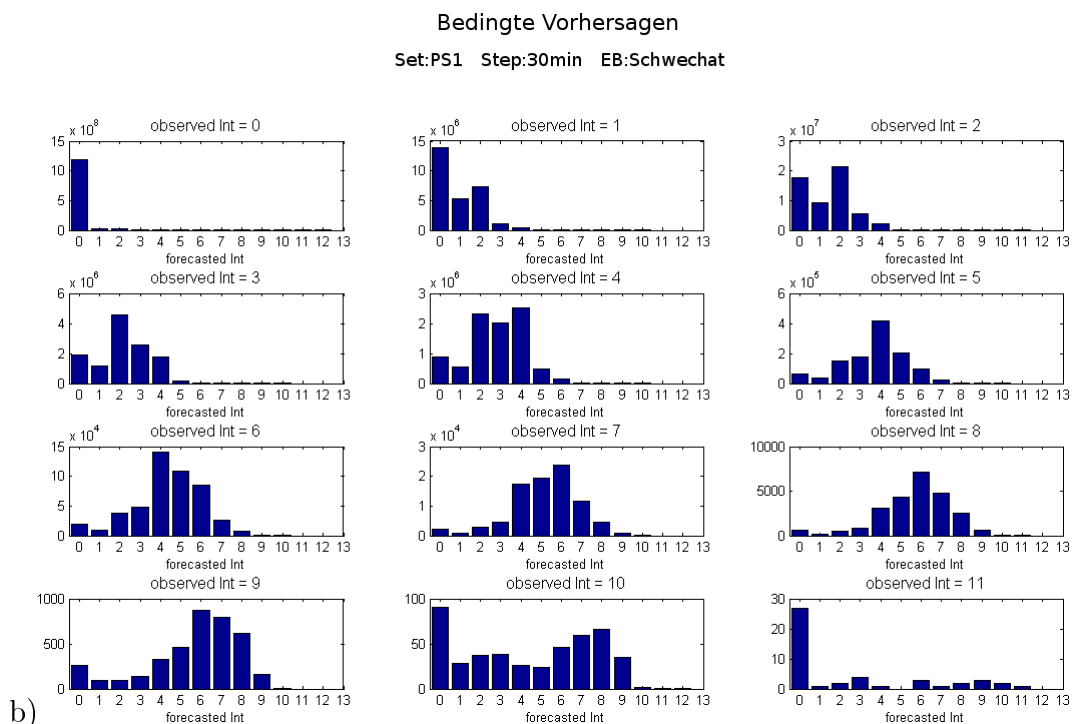
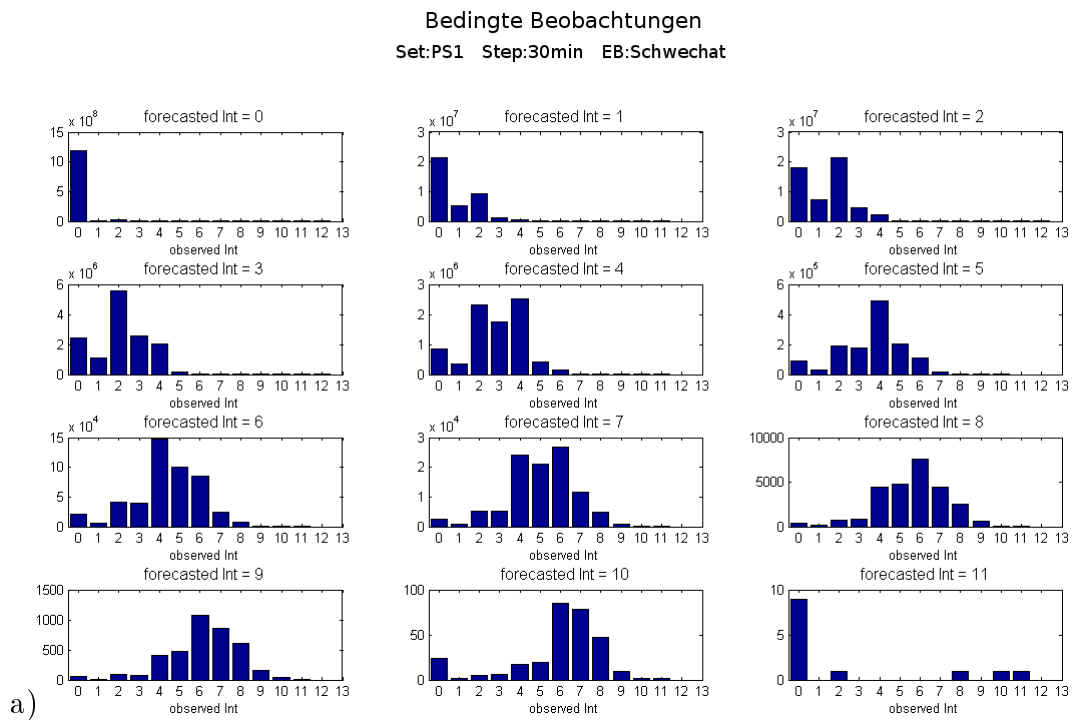


Abbildung 6.9:

Bedingte Beobachtungen und Vorhersagen von PS 1 (Winter): Anhand der bedingten Häufigkeiten zeigt sich die Problematik der genauen Ortsvorhersage am besten. In Abbildung 6.2 sieht man, dass sich die marginalen Häufigkeiten der Beobachtungen und Vorhersagen recht ähnlich sind. Hier jedoch liegen sowohl die bedingten Beobachtungen, als auch die bedingten Vorhersagen im Intensitätsmittel zu tief. Das spricht dafür, dass die Niederschlagsregime mitunter gut erfasst, aber nicht an die exakt richtige Position extrapoliert werden.

polationen lassen sich durch die bereits erwähnten, bedingten Häufigkeiten der Beobachtungen und Vorhersagen gewinnen. In Abbildung 6.9 sind diese als Histogramme für 30 minütige Extrapolationen von PS1 im Erfassungsbereich des WXR Wien/Raichenwarth dargestellt. An den marginalen Verteilungen in Abbildung 6.2 wurde gezeigt, dass sich die Verteilungen von Beobachtungen und Vorhersagen durchaus ähnlich sind. Je nach Parameterkombination treten dort leicht Abweichungen in beide Richtungen (Über- oder Unterschätzung) auf. Die Interpretation der bedingten Häufigkeiten ist dagegen ungleich schwieriger. In Abbildung 6.9a sind die bedingten Beobachtungen dargestellt. Diese spiegeln die Ergebnisse aus dem Reliability-Diagrammen (Abbildung 6.8 und 6.7) wider. Sie sind ein Maß für die Reliability der Extrapolationen (siehe Abschnitt 4.3). Man erkennt die bereits diskutierte, tendenziell zu hohe Vorhersageintensität, die durch eine schlechte Ortsvorhersage zustande kommt. Diese Vermutung wird durch die bedingten Vorhersagen in Abbildung 6.9b sehr gut bestätigt. Diese alleine betrachtet, würden nämlich auf ein gänzlich anderes Verhalten der Vorhersagen schließen lassen. Bei gegebener Beobachtungsintensität wird im Mittel eine zu niedrige Intensität prognostiziert. Somit könnte man daraus eine zu „trockene“ Vorhersagecharakteristik ableiten. Durch die Kombination beider bedingter Verteilungen, sowie der marginalen Verteilungen von Beobachtung und Vorhersage aus Abbildung 6.2 bestätigt sich jedoch die bereits formulierte Vermutung der unzureichenden Ortsvorhersage. Diese stellt somit das Hauptproblem der Prognosequalität dar.

Da besonders bei hohen Intensitäten die Reliability sowie die Discrimination der Extrapolationen dramatisch abnimmt, wurden diese noch näher untersucht. Es wurden alle Beobachtungsfelder einzeln betrachtet, die eine Vorhersage mit mindestens einen Pixel der Intensitätsstufe 11 zur Folge haben. Das sind im winterlichen Betrachtungszeitraum für das gesamte WXR-Austria-Composite insgesamt 44 Radarbilder, die sich auf nur 8 Tage verteilen. 35 dieser Bilder liegen sogar innerhalb von 19 Stunden, beginnend am 4. Jänner 2009 um 17:30. In diesem Zeitraum erkennt man, dass für das WXR Salzburg/Feldkirchen der Ground-Clutter-Filter (siehe Abschnitt 3.2.1.2.1) ausgefallen war und somit Bodenziele Störechos hoher Intensität verursachen (siehe Abbildung 6.10). Diese Störungen fluktuieren zeitlich stark, sind somit nicht extrapolierbar. Beim Vergleich der bedingten Beobachtungen mit den bedingten Vorhersagen, jeweils für hohe Intensitäten, fällt weiter auf, dass die Anzahl der Beobachtungen deutlich über der der Vorhersagen liegt. Dies ist gut durch die Filterung der Daten im Zuge der Extrapolationsmethodik zu erklären, durch die einzelne hohe Intensitäten

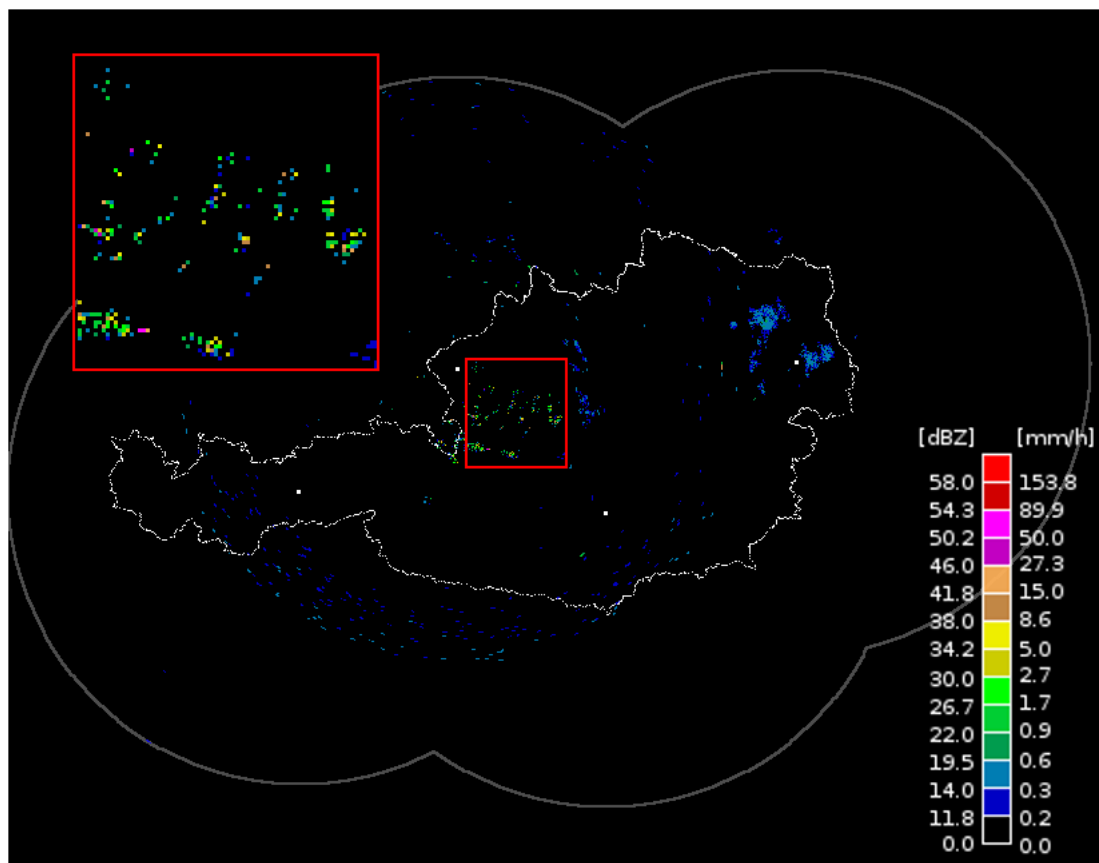


Abbildung 6.10

Das WXR-Austria-Compsite vom 4.Jänner 2009 um 19:50 zeigt die Auswirkungen eines Ausfalls des Ground-Clutter-Filters deutlich. Im Bereich des WXR Salzburg/Feldkirchen sind Störechos sehr hoher Intensität sichtbar. Diese erschweren die automatische Verifikation ungemein.

abgeschwächt oder gar entfernt werden.

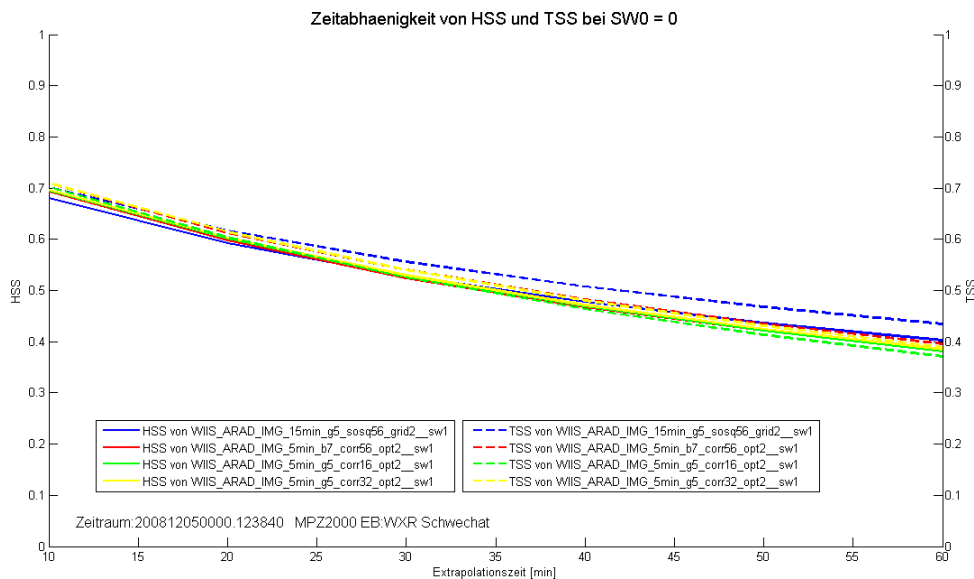


Abbildung 6.11:

Vergleich der Skills der Extrapolationstypen (Winter): Vergleich der Skill-Scores TSS und HSS für alle vier Parameterkombinationen. Die Berechnungen wurden wieder für den Erfassungsbereich des WXR Wien/Rauchenwarth gemacht. HSS zeigt sehr ähnliche Ergebnisse für alle Extrapolationstypen. TSS bevorzugt „nasse“ Prognosen leicht, somit schneidet PS1, vor allem bei längeren Extrapolationsschritten am besten ab.

Als letztes Qualitätskriterium soll nun anhand Abbildung 6.11 der Prognose-Skill besprochen werden. Für den Erfassungsbereich des WXR Wien/Rauchenwarth werden die vier Extrapolationstypen durch die Skill-Scores HSS und TSS (siehe Abschnitt 4.6) miteinander verglichen. Der ETS wurde aufgrund seiner engen Verknüpfung mit dem HSS nicht geplottet. Die beiden zeigen weitgehend idente Ergebnisse. Alle vier Extrapolationstypen schneiden, gemessen nach HSS, sehr ähnlich ab, wobei mit zunehmender Extrapolationsdauer PS1 wieder leicht zu bevorzugen sein dürfte. Etwas größere Unterschiede zwischen den Parameterkombinationen ergeben sich bei der Berechnung des TSS. Hier wird PS1 der mit Abstand beste Skill bescheinigt. PS3 schneidet am schlechtesten ab. Im Abschnitt 4.6.4 wurde aber bereits besprochen, dass laut Doswell-III et al. (1990) TSS sich bei seltenen Ereignissen zu sehr an POD annähert, somit „nasse“ Prognosen etwas geschönt dargestellt werden. Wilks (2006) wiederum hebt den positiven Effekt dieser Eigenschaft hervor: Ein korrekt prognostiziertes Ereignis (Niederschlag oder Nicht-Niederschlag) hat größeren Einfluss auf die TSS, wenn

es seltener auftritt. Außerdem ist zu bemerken, dass PS1 sich erst bei längeren Extrapolationsschritten von den anderen Extrapolationstypen absetzen kann. Dies ist ein Indiz für den positiven Einfluss des 15 minütigen Abstandes zwischen den Ausgangsbildern bei längeren Extrapolationen.

6.2 Auswertung Sommer 2009

Aufbauend auf den Erkenntnissen aus Abschnitt 6.1, in dem der Winter 2008-2009 ausgewertet wurde, sollen nun die Sommermonate (Juni bis August) des Jahres 2009 untersucht werden. Es wird versucht in erster Linie die Unterschiede zwischen den zwei Stichproben zu zeigen. Da in den Wintermonaten keine qualitativen Unterschiede zwischen dem Erfassungsbereich Wien/Rauchenwarth und dem gesamten WXR-Austria-Composite festgestellt wurden, wird nun ausschließlich die WXR-Station Schwechat-Feldkirchen untersucht.

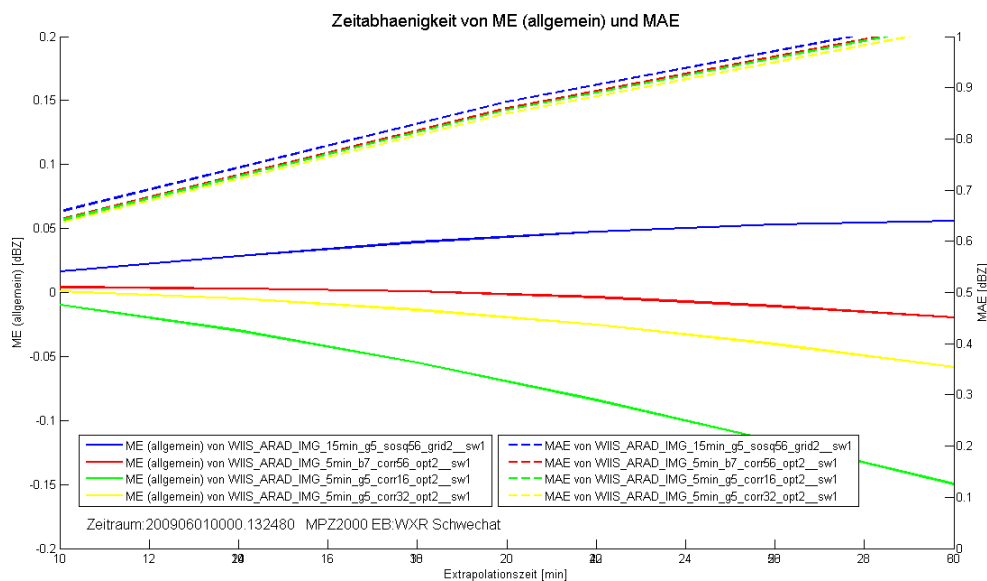


Abbildung 6.12:

Vergleich von ME und MAE aller Parameterkombinationen (Sommer): Ähnlich wie bei der Auswertung der Wintermonate zeigt auch in der Sommerperiode vor allem der ME deutliche Unterschiede zwischen den Extrapolationstypen auf. PS 1 weist wieder einen positiven ME auf, PS 2 und PS 3 einen leicht negativen und PS 4 einen stark negativen. Die MAE-Werte liegen deutlich über den winterlichen Werten, zeigen aber nur sehr geringe Unterschiede zwischen den Extrapolationstypen auf. Es wurde nur das Erfassungsgebiet des WXR Wien/Rauchenwarth ausgewertet.

Analog zur Auswertung der Wintermonate soll anhand von Abbildung 6.12 mit

der Betrachtung der klassischen, statistischen Momente, dem mittleren Fehler (ME), sowie dem mittleren absoluten Fehler (MAE), begonnen werden. Beim Vergleich mit Abbildung 6.1 erkennt man, dass der MAE aller Parameterkombinationen deutlich höher liegt, als dies im Winter der Fall war. (Um solche Unterschiede besser heraus zu arbeiten, wurden die Achsenskalierungen aus den Abbildungen in Abschnitt 6.1 beibehalten.) Da der Unterschied zwischen ME und MAE als Maß für die Qualität der Ortsvorhersagen herangezogen werden kann und dieser hier noch größer ausfällt als im winterlichen Vergleichszeitraum, bestätigt sich, dass die Verlagerungsvektoren konvektiver Systeme aufgrund ihrer, im Mittel kleineren räumlichen Ausdehnungen, komplexeren Zugbahnen, sowie kürzeren Lebenszyklen und der damit verbundenen hohen, zeitlichen Variabilität schwieriger zu extrapolieren sind. Andererseits kommt ein großer Teil dieses markant höheren MAE sicher durch die höheren Intensitäten in konvektiven Systemen und deren räumlicher und zeitlicher Fluktuation zu Stande, die nicht genau vorhersagbar ist. Dieser Teil des Fehlers wäre somit einer fehlerhaften Intensitätsvorhersage zuzuordnen.

Die ME-Werte der einzelnen Parameterkombinationen weisen im Vergleich zu den Wintermonaten eine etwas höhere Streuung auf. Die leichte Überschätzung der Niederschlagsfelder durch PS 1 ist genauso stärker ausgeprägt wie die zu trockene Prognosecharakteristik des PS 3. PS 2 und PS 4 zeigen kaum Unterschiede bezüglich Sommer und Winter. Allein aufgrund der Abbildung 6.12 wäre somit PS 2 zu bevorzugen, da es den betragsmäßig kleinsten ME aufweist und die Unterschiede der MAE-Werte zwischen den vier Parameterkombinationen äußerst gering ausfallen.

Vergleicht man nun aber die CSI-Werte (siehe Abschnitt 4.2) für die Sommermonate (Abbildung 6.13) mit denen des Winters, so erkennt man etwas überraschend eine höhere Accuracy der Prognosen der konvektiven Jahreszeit. Der CSI aller Parameterkombinationen aus Abbildung 6.13 liegt leicht über den entsprechenden Werten aus Abbildung 6.3, wobei PS 1, PS 2 und PS 4 weitgehend gleichwertige Ergebnisse liefern, PS 3 hingegen bei längeren Extrapolationen relativ gesehen etwas an Qualität verliert. Da der CSI aus dem dichotomen System berechnet wird und somit nur die Ausdehnung der Niederschlagsfelder (und deren Position) entscheidend ist, könnte man die geringfügig besseren Ergebnisse im Sommer auf eindeutigere Niederschlagsgrenzen bei konvektiven Systemen zurückführen. Auf jeden Fall spricht das recht gute Abschneiden der vereinfachten (dichotomen) Extrapolationen dafür, dass ein großer Teil des, nach Abbildung 6.12 im Vergleich zu den Wintermonaten, höheren MAE in

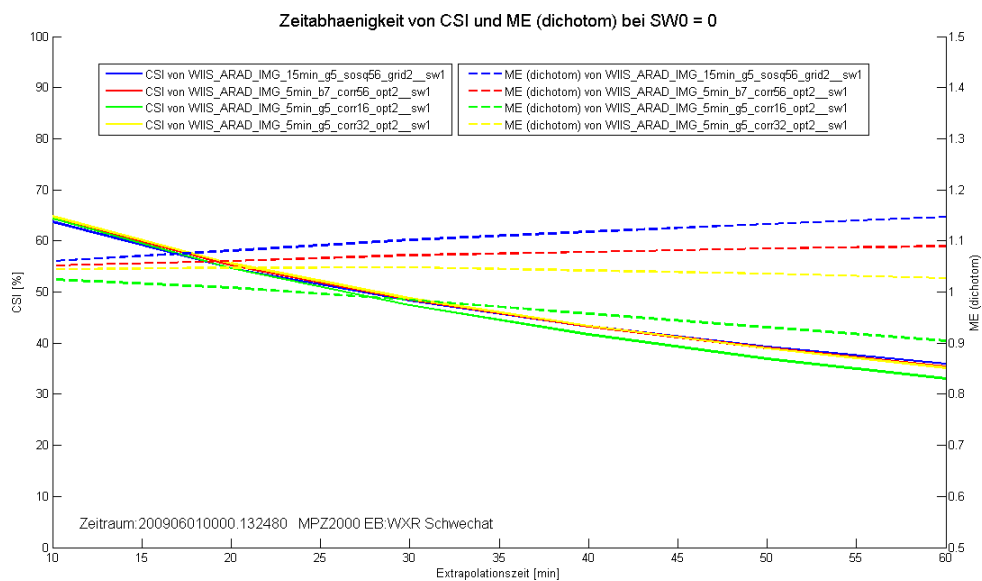


Abbildung 6.13:

CSI und ME des dichotomen Systems (Sommer): Der CSI bescheinigt den Parameterkombinationen PS 1, PS 2 und PS 4 ähnlich Accuracy. PS 3 liegt hier deutlich zurück. Ergänzend dazu wurde der Bias des dichotomen Systems ME(dichotom) geplottet, um zu erkennen, ob die gewonnene Accuracy auf zu „nasse“ Prognosecharakteristika zurückzuführen ist. Somit ist anhand dieser Graphik in Summe PS 3 zu bevorzugen. Bei annähernd gleicher Accuracy weist es den mit Abstand geringsten mittleren Fehler auf.

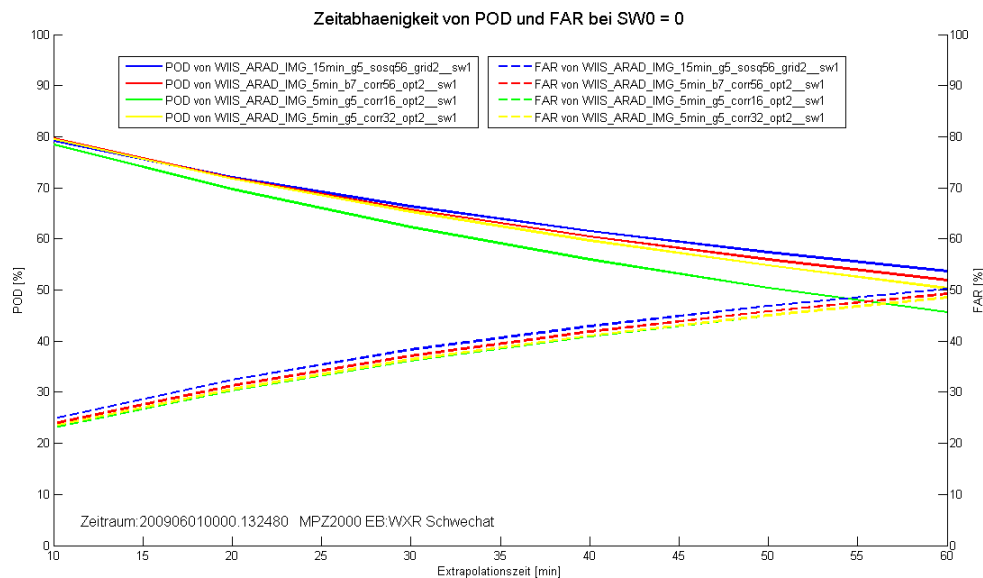


Abbildung 6.14:

POD und FAR Vergleich (Sommer): Für den Erfassungsbereich des WXR Wien/Rauchenwarth wurden POD und FAR aller vier Parameterkombinationen geplottet. Die Extrapolationstypen PS 1, PS 2 und PS 4 liegen eng beieinander, aufgrund seiner leicht „nassen“ Prognosecharakteristik weist PS 1 etwas höhere POD-, aber auch leicht erhöhte FAR-Werte auf. Negativ macht nur PS 3 auf sich aufmerksam. Bei, im Vergleich zu den anderen Extrapolationstypen, gleichwertigen FAR-Werten weist PS 3 deutlich niedrigere POD-Werte auf.

der Intensitätsvorhersage begründet ist. Die Accuracy des dichotomen Systems liegt im Winter und Sommer auf ähnlichem Niveau, die des allgemeinen Systems liegt im Winter deutlich höher als im Sommer.

Dies gilt, laut Abbildung 6.14 auch für die Reliability (FAR, siehe Abschnitt 4.3) und die Discrimination (POD, siehe Abschnitt 4.5) des dichotomen Systems. Wobei sowohl POD, als auch FAR im Vergleich zu der Auswertung der Wintermonate in Abbildung 6.4 in der konvektiven Jahreszeit sogar leicht bessere Werte liefern. Auffällig ist auch, dass PS 3 eine offensichtlich schlechtere Discrimination als alle anderen, die sich nur wenig unterscheiden, aufweist. Zum Vergleich: In den Wintermonaten (siehe Abbildung 6.4) war PS 1 als positiver „Ausreißer“ erkennbar, PS 2 bis PS 4 lieferten ähnliche Werte.

Zum Abschluss der Betrachtungen des dichotomen Systems soll das ROC-Diagramm aller Parameterkombinationen für 60 minütige Extrapolationen in Abbildung 6.15 dienen. Der markanteste Unterschied zu der Auswertung der Wintermonate liegt in der

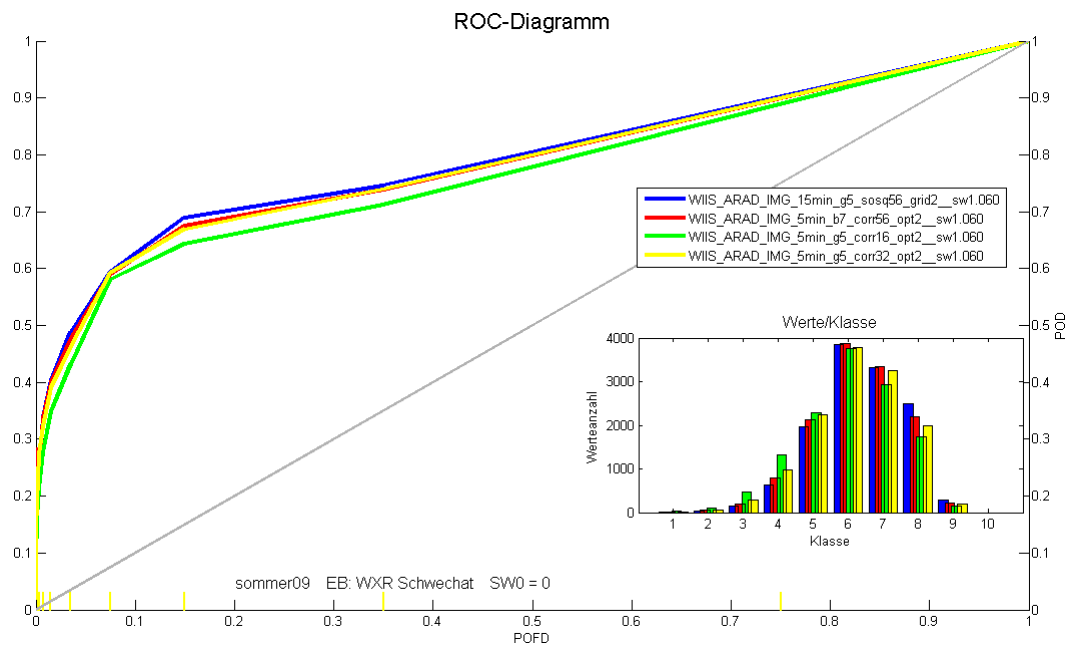


Abbildung 6.15:

ROC-Diagramm (Sommer): Vergleich der 60 minütigen Extrapolationen aller vier Parameterkombinationen mittels ihrer ROC-Kurven. Die Klassenaufteilung wurde gegenüber der Betrachtung der Wintermonate in Abbildung 6.5 nicht verändert. Durch die „sommerliche“ Intensitätsverteilung (siehe auch Abbildung 6.17) sind nun die oberen Klassen zahlreicher besetzt, daher auch besser abgesichert. Bis auf PS 3 weisen alle Parameterkombinationen ähnliche Ergebnisse auf. PS 3 liegt bezüglich der Discrimination des dichotomen Systems hinter den anderen zurück.

Anzahl der Werte in den höheren Klassen, bei gleicher Klassenwahl. Daher besitzen hier die Datenpunkte bei höheren POFD-Werten etwas mehr Glaubwürdigkeit als in Abbildung 6.5. Die Discrimination liegt in den beiden Vergleichszeiträumen insgesamt auf ähnlichem Niveau. Das ROC-Diagramm in Abbildung 6.15 bestätigt den Eindruck aus Abbildung 6.14: PS 3 hebt sich in negativer Hinsicht von den anderen Extrapolationstypen ab.

Bei näherer Betrachtung der Reliability des allgemeinen Systems mittels der Reliability-Diagramme für 60 minütige Extrapolationen mit PS 1 in Abbildung 6.16, erkennt man wie schon in der winterlichen Periode die scheinbar zu „nasse“ Prognosecharakteristik, die höchstwahrscheinlich durch die schiefe Werteverteilung zwischen den Intensitätsklassen zustande kommt (siehe Abschnitt 6.1). Außerdem fällt eine gleichmäßigere Intensitätsverteilung auf. Höhere Intensitäten sind weitaus öfter vertreten als im Winter, somit können hier die Datenpunkte zwischen Intensitätsstufe 1 bis 6 im Sommer als repräsentativer betrachtet werden.

Ergänzend sind in Abbildung 6.17 die marginalen Verteilungen der Beobachtungen und Vorhersagen dargestellt. Beim Vergleich mit Abbildung 6.2 erkennt man die leicht zu höheren Intensitäten verschobene Verteilung.

Nun sollen noch die bedingten Beobachtungen und Vorhersagen betrachtet werden. Aufgrund des geringsten mittleren Fehlers wurde an dieser Stelle PS 2 ausgewählt und alle 30 minütigen Extrapolationen ausgewertet. Im Gegensatz zu den Wintermonaten (siehe Abbildung 6.9) erkennt man die bei weitem größere Anzahl von höheren Intensitäten. Die größere Stichprobe wirkt sich vor allem bei der Darstellung der bedingten Beobachtungen (Abbildung 6.18a) positiv aus. Bei einer hohen Vorhersageintensität (in etwa größer als Intensitätsstufe 4) ist die Wahrscheinlichkeit groß, dass eine solche auch eintritt. Etwas anders verhalten sich die bedingten Vorhersagen in Abbildung 6.18b. Hier überwiegt bei Punkten, auf denen hohe Intensitäten gemessen werden, die Wahrscheinlichkeit, dass kein Niederschlag prognostiziert wurde. (PS 2 darf als repräsentativ für alle vier Extrapolationstypen gelten, sie verhalten sich sehr ähnlich.) Im Unterschied zu den Wintermonaten (siehe Abbildung 6.9) unterscheiden sich also im Sommer die bedingten Beobachtungen und die bedingten Vorhersagen bezüglich ihrer Qualität bei hohen Intensitäten deutlich. Die Situation ist aber auch ungleich komplexer zu analysieren. Bei einer Beobachtungsintensität von 11 gilt es insgesamt 10321 Radarbilder zu betrachten (im Winter waren es nur 44 Bilder). Daher wurde an dieser Stelle nur eine Stichprobe manuell untersucht. Diese reicht jedoch aus, um

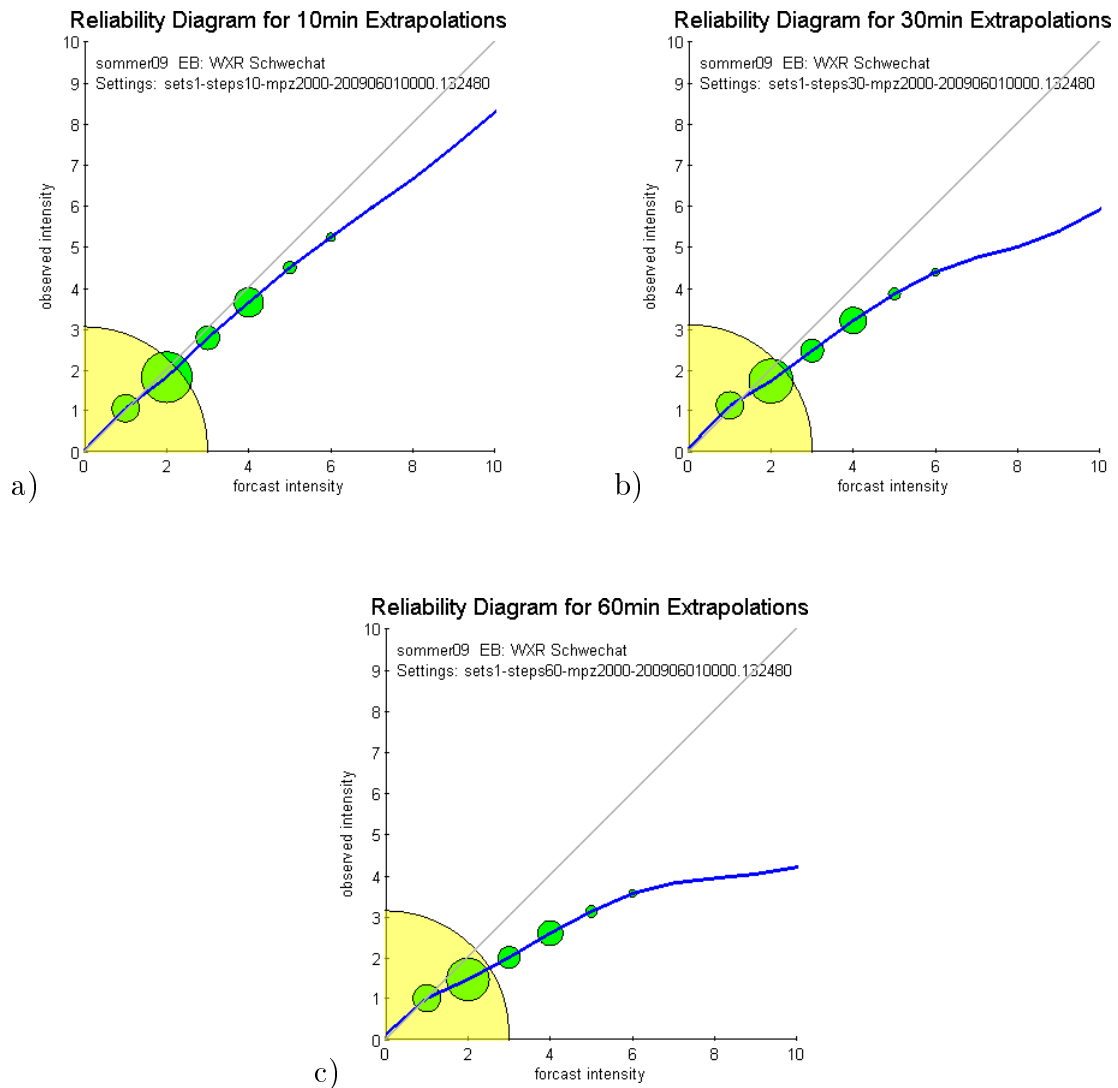


Abbildung 6.16:

Reliability bei unterschiedlichen Extrapolationsschritten (Sommer): Reliability Diagramme des PS1 für die Vorhersageschritte: a) 10 Minuten; b) 30 Minuten; c) 60 Minuten. Die Reliability nimmt wieder erwartungsgemäß mit zunehmender Vorhersagedauer ab. Durch die schiefe Verteilung der Häufigkeit der Intensitäten ergeben sich ab Intensitätsstufe 2 scheinbar „nasse“ Prognosecharakteristika.

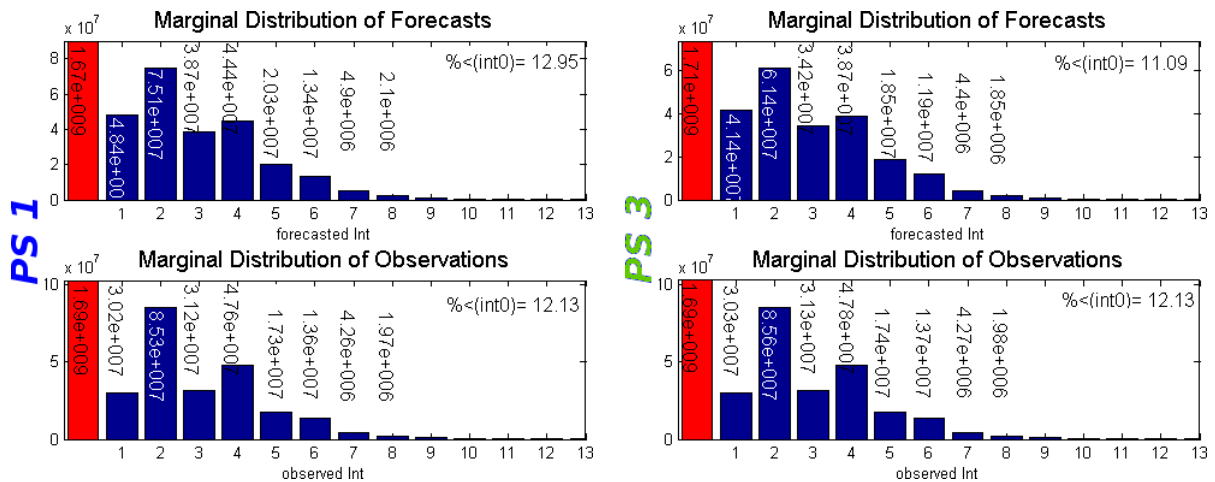


Abbildung 6.17:

Vergleich der marginalen Verteilungen der Vorhersagen mit PS 1 und PS 3 (für den Erfassungsbereich des WXR Wien/Rauchenwarth) mit den dazugehörigen marginalen Beobachtungsverteilungen. Man erkennt wie schon in den Wintermonaten die leichte Überschätzung der Niederschlagsregime („nasse“ Vorhersage) durch das PS 1, sowie eine leichte Unterschätzung durch das PS 3 („trockene“ Vorhersage). Außerdem fällt auf, dass die Häufigkeit der Intensitätsklasse 1, wieder bei beiden Extrapolationstypen deutlich überschätzt wird.

zu erkennen, dass auch die Ursachen vielfältiger ausfallen als im winterlichen Vergleichszeitraum, in dem der Ausfall eines Ground-Clutter-Filters einen Großteil der beobachteten, hohen Intensitäten verursachte (siehe Abbildung 6.10). Auch im Sommer sind Störsignale vermutlich eine der Hauptursachen der geringen Discrimination der Extrapolationen bei hohen Intensitäten. Die gefundenen Störungen umfassen neben einigen Ground-Clutter-Ausfällen jedoch auch die meisten anderen Störungsarten aus Kapitel 3.2.1.2, wie zum Beispiel Second-Trip-Echos oder Abschattungen hinter konvektiven Zellen (siehe Abbildung 6.19).

Zusätzlich ist zu bemerken, dass auch im Sommer die Zahl der beobachteten, hohen Intensitäten (in 10321 WXR-Bildern gegenüber 7659 Extrapolationen bei Intensität 11) überwiegt. Störechos, die zeitlich stark fluktuieren, beziehungsweise sogar nur in einem Bild auftreten (siehe auch Abbildung 6.20), werden nicht extrapoliert oder wenn es sich nur um einzelne, räumlich isolierte Pixel handelt (zum Beispiel beim Ground-Clutter), durch die Filterung im Zuge der Extrapolation entfernt. Daher kommen diese Störungen hauptsächlich bei der Betrachtung der bedingten Vorhersagen (zum Beispiel bei gegebener Beobachtungsintensität 11) zum Tragen. Dies ist auf jeden

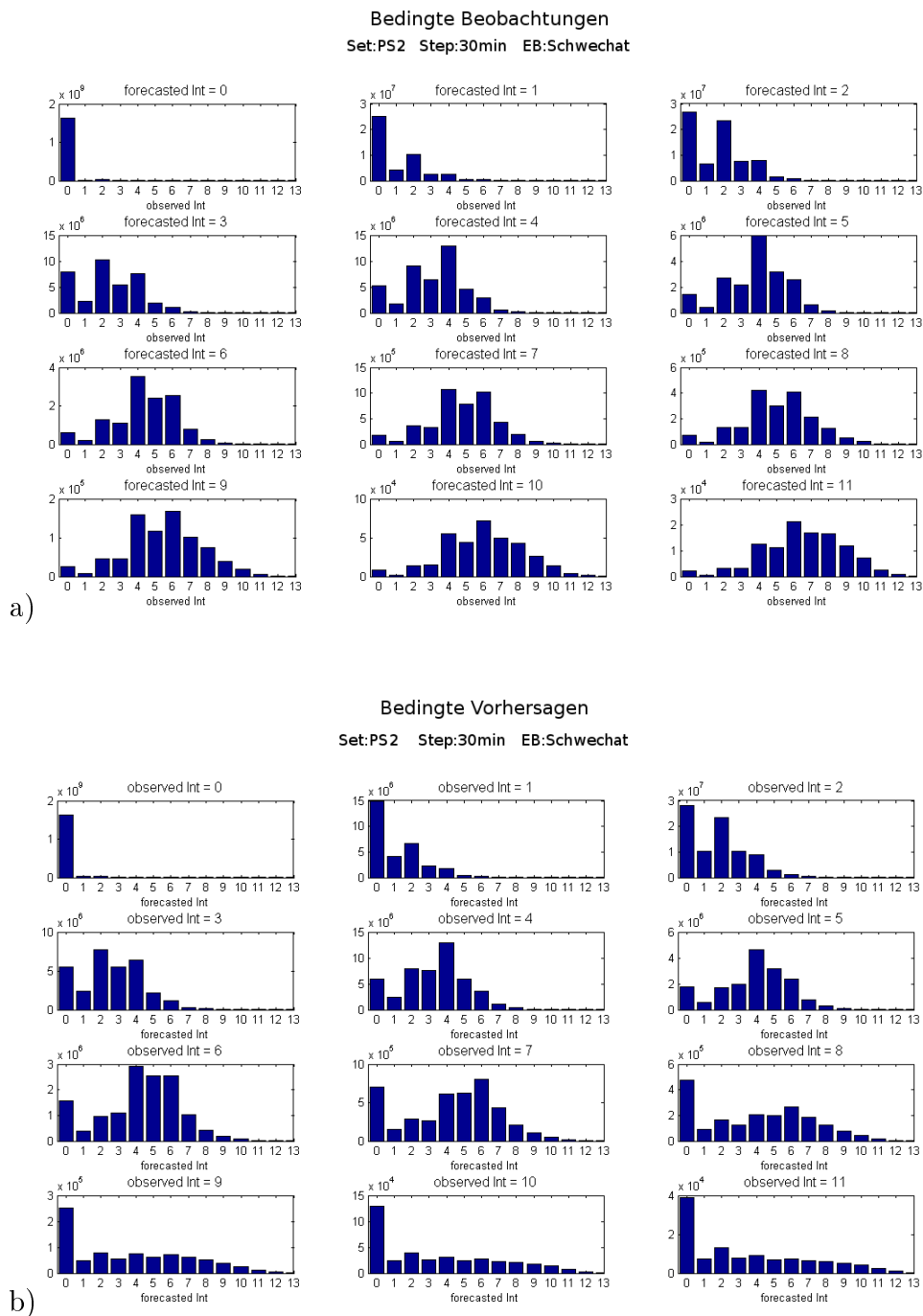


Abbildung 6.18:

Bedingte Beobachtungen und Vorhersagen von PS 2 (Sommer): Anhand der bedingten Häufigkeiten für die Sommermonate zeigt sich nochmals die Problematik der genauen Ortsvorhersage. Wieder liegen sowohl die bedingten Beobachtungen, als auch die bedingten Vorhersagen im Intensitätsmittel zu tief. Das spricht zusammen mit den marginalen Verteilungen aus Abbildung 6.17 dafür, dass die Niederschlagsregime mitunter gut erfasst, aber nicht an die exakt richtige Position extrapoliert werden. Bei hohen Intensitäten kann man hier von glaubwürdigeren Ergebnissen als im Winter ausgehen, weil diese bei weitem öfter auftreten. Interessant ist auch, dass die bedingten Beobachtungen bei hohen Intensitäten wesentlich besser abschneiden als die bedingten Vorhersagen bei gegebener hoher Intensität.

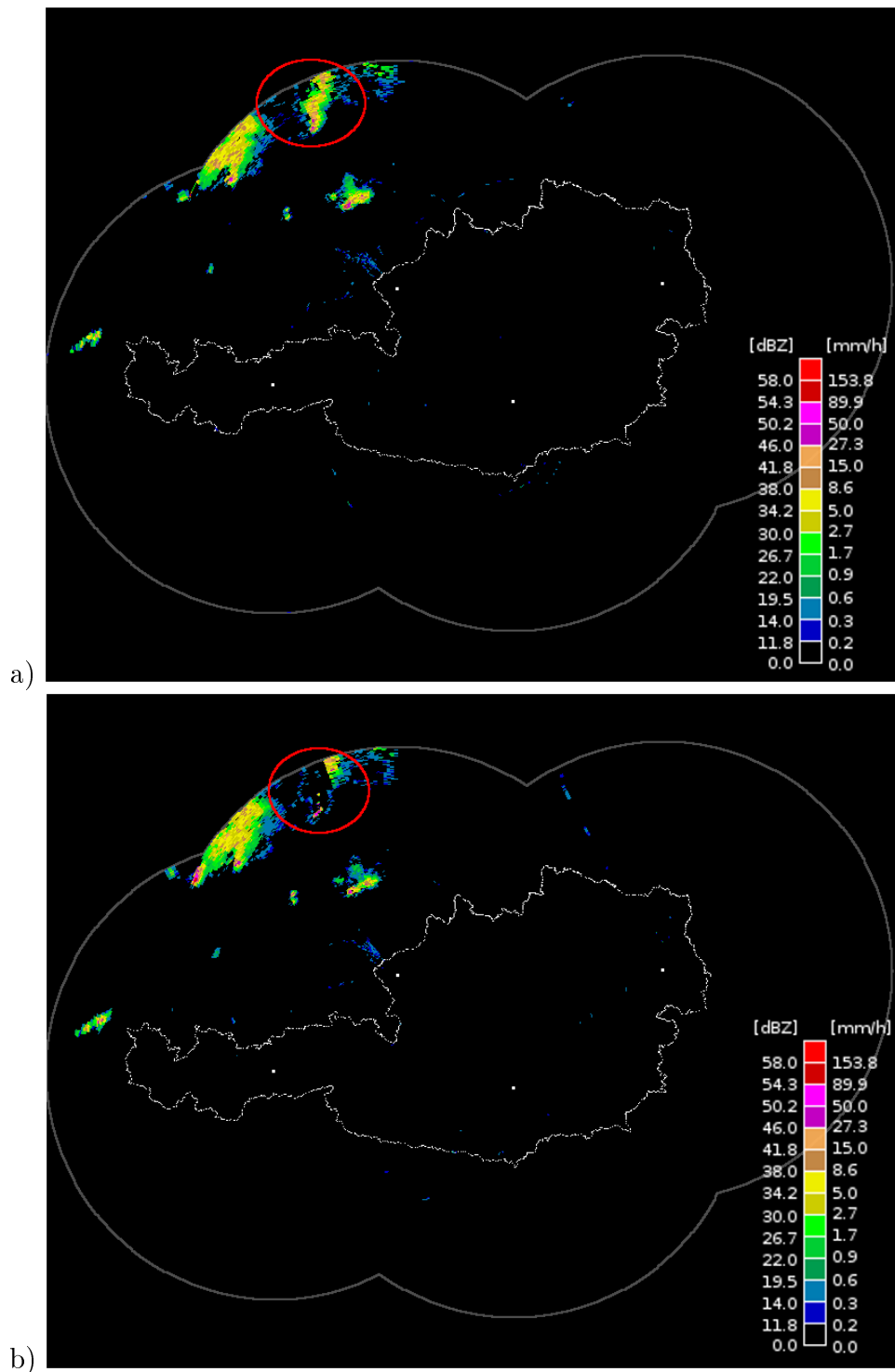


Abbildung 6.19:

Abschattung von Radarechos hinter einer konvektiven Zelle am 14. Juli 2009: (a) Um 02:20 treten zwar bereits Dämpfungseffekte in dem Niederschlagsregime nordnordwestlich des WXR Salzburg/Feldkirchen auf, die durch die unterbrochene Struktur erkennbar sind, aber die rot markierte Zelle ist noch einigermaßen gut erkennbar. (b) Im nächsten WXR-Bild um 02:25 verschwindet die markierte Zelle weitgehend, sie wird von der etwas weiter südlich liegenden Zelle vom WXR Salzburg/Feldkirchen abgeschattet.

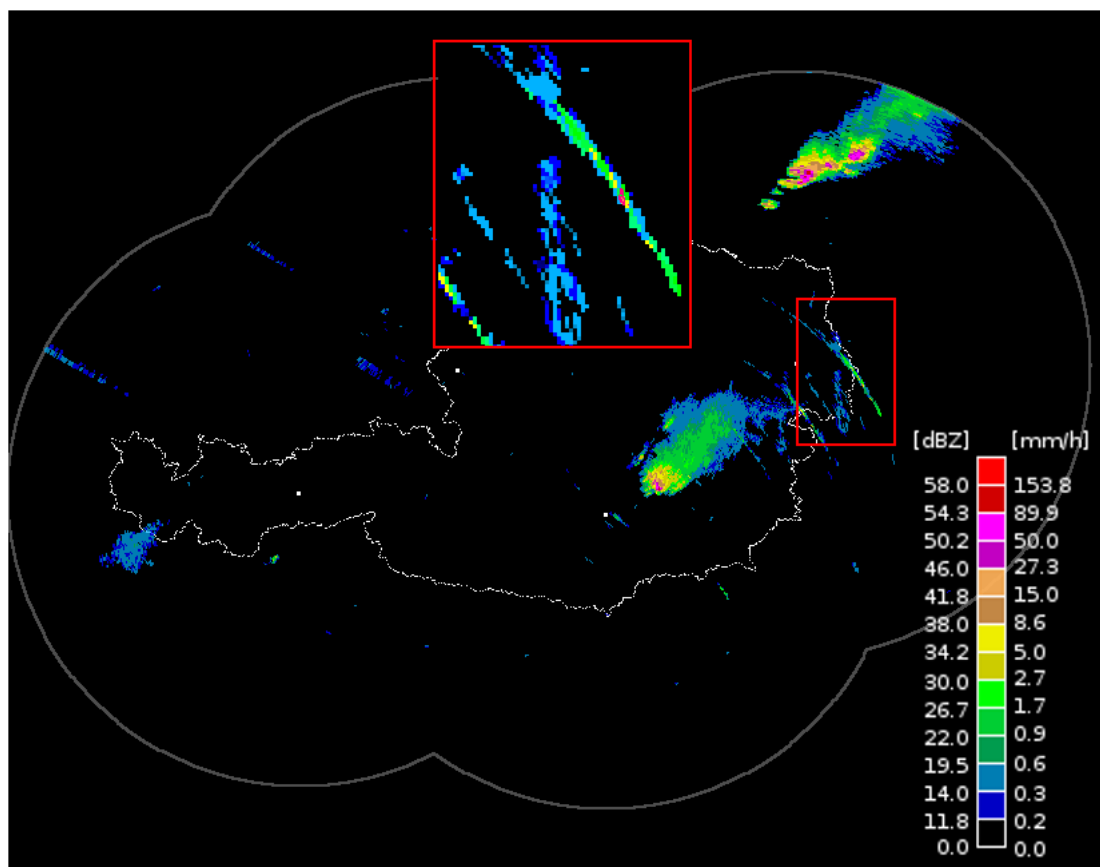


Abbildung 6.20:

Eine kreisförmige Störung, gemessen vom WXR Zirbitzkogel am 14. Juli um 18:45 fällt durch ihre recht hohen Intensitäten auf. Sie trat nur in einem einzigen WXR-Bild auf.

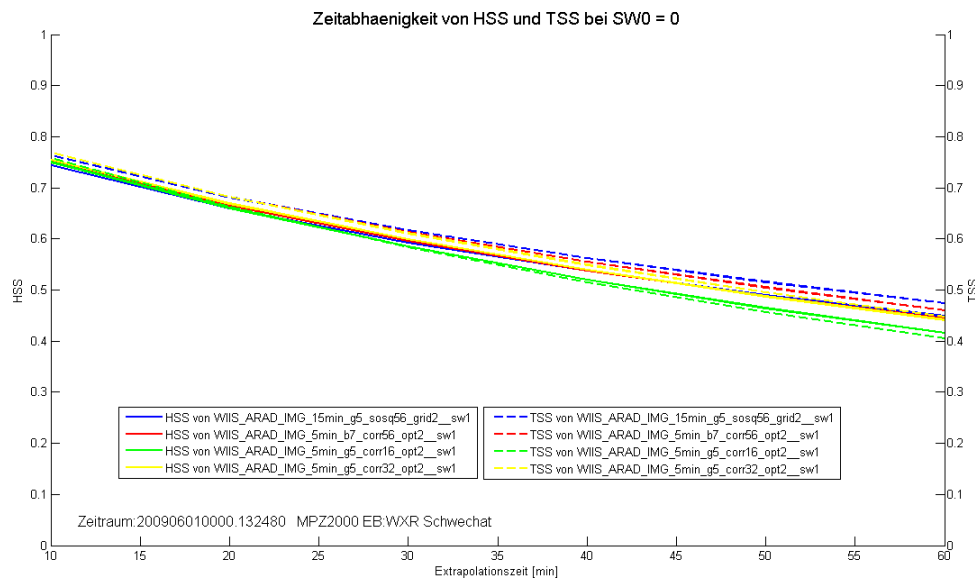


Abbildung 6.21:

Skill-Vergleich (Sommer): Vergleich der Skill-Scores TSS und HSS für alle vier Parameterkombinationen. Die Berechnungen wurden wieder für den Erfassungsbereich des WXR Wien/Rauchenwarth gemacht. PS 1 ist laut beiden Maßen leicht zu bevorzugen, PS 3 schneidet eindeutig am schlechtesten ab. Die Werte liegen insgesamt leicht über denen der Wintermonate.

Fall eine Ursache für die ungleich schlechtere Qualität der bedingten Vorhersagen bei hohen Intensitäten.

Es sei aber auch erwähnt, dass die Häufigkeiten der bedingten Beobachtungen bei hohen Intensitäten eindeutig für eine ansprechende Reliability des Extrapolationssystems sprechen. Sie zeigen, dass hohe Intensitäten im Schnitt etwas überschätzt vorhergesagt werden, aber in diesen Fällen mit sehr hoher Wahrscheinlichkeit auch ein starkniederschlagsträchtiges Echo zu erwarten ist.

Zum Abschluss soll noch der Skill der vier Parameterkombinationen mittels HSS und TSS (siehe Abschnitt 4.6) betrachtet werden. Ähnlich zur Auswertung der Winterperiode zeigen sich nur kleine Unterschiede zwischen den Extrapolationstypen. PS 1 ist gegenüber den anderen leicht zu bevorzugen, vor allem bei längeren Extrapolationsschritten. PS 3 fällt relativ gesehen eindeutig ab. Der Skill aller Parameterkombinationen liegt auch hier leicht über jenem der Wintermonate. Da sich auch diese Maße (wie schon der CSI und ME(dichotom) in Abbildung 6.13) aus dem dichotomen System berechnen, liegt der Schluss nahe, dass auch hier eindeutigere Niederschlagsgrenzen dafür verantwortlich sind.

7 Ergebnis und Ausblick

Grundsätzlich zeigen alle vier Extrapolationstypen ähnliche Stärken und Schwächen. Positiv ist zu erwähnen, dass im Allgemeinen die marginalen Intensitätsverteilungen der Extrapolationen (Abbildung 6.2 und 6.17) der Gemessenen sehr nahe kommen. Dies gilt insbesondere für das PS 2, das in den Sommer- und Wintermonaten den vergleichsweise geringsten mittleren Fehler aufweist (siehe Abbildung 6.1 und 6.12). PS 1 hebt sich durch die beste Discrimination (höchste POD-Werte) von den anderen Extrapolationsvarianten ab. Diese kommt aber durch eine leicht zu „nasse“ Prognosecharakteristik zustande, wodurch sich Nachteile in Bezug auf die Reliability ergeben. Der Vergleich mittels Reliability-Diagramm in Abbildung 6.8 zeigt jedoch, dass dieser Nachteil sehr gering ausfällt, somit eindeutig der Nutzen der höheren Discrimination überwiegt (siehe Abbildung 6.4 und 6.14), der im Winter vor allem im Vergleich zu PS 2 und PS 4 noch größer ausfällt.

Besonders durch die Histogramme der bedingten Vorhersagen und Beobachtungen (Abbildung 6.9 und 6.18) erkennt man, dass exakte Ortsvorhersagen die Schwachstelle dieser Extrapolationen darstellen, wobei diese Schwäche im Sommer etwas stärker ausgeprägt ist als im Winter. Das ist durch den Vergleich der mittleren, absoluten Fehler für Winter und Sommer (Abbildung 6.1 und 6.12) gut ersichtlich. Bezüglich ihres Skills (Abbildung 6.11 und 6.21) liegen ebenfalls alle vier Parameterkombinationen eng beieinander, wobei PS 1 leichte Vorteile hat, die aber erst bei größeren Extrapolationsschritten zum Tragen kommen.

Zusammenfassend kann man sagen, dass bei (in erster Linie) stratiformen Niederschlagsformen in den Wintermonaten der Vorteil einer höheren Accuracy, Skill und Discrimination bei nur leicht abfallender Reliability den Ausschlag zu Gunsten von PS 1 geben sollte. In den konvektiv dominierten Sommermonaten fällt dieser Vorteil wesentlich geringer aus. PS 2 und PS 4 weisen hier eine nur wenig geringere Discrimination und Accuracy auf, liegen außerdem bezüglich des mittleren Fehlers etwas besser als PS 1. Bei konvektiven Systemen scheint der Vorteil der großzügigeren, zeitlichen Mittelung durch das 15 minütige Ausgangsintervall von PS 1 deutlich weniger positive

Effekte zu haben, als in den Wintermonaten. Dies erscheint glaubwürdig, wenn man den kürzeren Lebenszyklus, sowie die zeitlich stärker variierenden Bewegungen konvektiver Systeme bedenkt. Somit liegen PS 1 und PS 4 im betrachteten Sommer auf ähnlichem Niveau, dicht gefolgt von PS 2. PS 3 erwies sich sowohl im Winter als auch im Sommer als die ungünstigste Extrapolationsvariante, die weit hinter den anderen zurück blieb. Welche Parameterkombination nun tatsächlich für einen bestimmten Anwendungsbereich am besten geeignet ist, hängt aber natürlich auch stark von der Anwendung selbst ab. An dieser Stelle sei nochmals betont, dass die Qualität der WXR-Daten bei allen Extrapolationsarten maßgebend für die Qualität der Extrapolationen ist und es daher das Ziel sein muss, diese weiter zu verbessern. Ein Großteil der Probleme, die während der Verifikation auftraten, sind auf fehlerbehaftete Ausgangsdaten zurückzuführen.

Zum entwickelten und getesteten Verifikationssystem kann man sagen, dass es in der Lage ist, alle entscheidenden Qualitätsmerkmale, sei es aus theoretischer (siehe Kapitel 4) oder aus praktischer Sicht, (Orts- und Intensitätsvorhersage) zu erfassen. Die klassischen, statistischen Momente sowie die Maße des dichotomen Systems, die in Kapitel 4 ausführlich beschrieben wurden, konnten uneingeschränkt für die Verifikation von WXR-Extrapolationen verwendet werden. Auch die hier ergänzend eingeführten Methoden, die grundsätzlich der Verifikation probabilistischer Prognosen entstammen (ROC- und Reliability-Diagramm), liefern wertvolle Erkenntnisse und können daher als geeignet betrachtet werden. Allerdings muss die Interpretation der Ergebnisse auf die schiefe Häufigkeitsverteilung der Intensitäten angepasst und teilweise überdacht werden. Das Reliability-Diagramm kann zum Beispiel nur relative Information über den Bias einer Vorhersage zu einer anderen liefern (siehe unter anderem Abbildung 6.8). Als besonders wertvoll stellten sich die Histogramme der marginalen und bedingten Häufigkeiten von Prognose und Beobachtung heraus. Sie zeigen sehr schön, dass die Intensitätsverteilung im Großen und Ganzen gut vorhergesagt, eine leicht fehlerbehaftete Ortsvorhersage aber schnell zu großen Abweichungen zwischen Vorhersage und Beobachtung und somit zu etwas schlechteren Verifikationsergebnissen führen kann. Dies ist besonders in den konvektiv dominierten Sommermonaten augenscheinlich.

Zusammenfassend darf man sagen, dass es anhand dieses, rein auf bewährten, klassischen Methoden aufbauenden Verifikationssystems möglich ist, WXR-Extrapolationen ausführlich zu analysieren und vor allem untereinander zu vergleichen. Ein nächster sinnvoller Schritt wäre, die in Abschnitt 2.2.1 kurz erwähnte optimierte Extrapo-

lation der ACG zu verifizieren und mit den bisherigen Ergebnissen zu vergleichen. Außerdem sei nochmals angemerkt, dass hier ausschließlich TREC-Extrapolationen betrachtet wurden. Auch die Anwendung auf objektbasierte Extrapolationsverfahren (siehe Kapitel 2) wäre sehr interessant.

Literatur

- Atger, F. et al. Forecast Verification: Issues, Methods and FAQ, März 2010. URL <http://www.cawcr.gov.au/projects/verification/>.
- Battan, L. J. *Radar Observation of the Atmosphere*. University of Chicago Press, Rev. Edition, 1973.
- Blahak, U. *Analyse des Extinktionseffektes bei Niederschlagsmessungen mit einem C-Band Radar anhand von Simulation und Messung*. Dissertation, Universität Karlsruhe, 2004.
- Brier, G. W. und Allen, R.A. Verification of weather forecasts. *American Meteorological Society*, Compendium of Meteorology:841–848, 1951.
- Brooks, H. E. und Doswell-III, C. A. A comparison of measure-oriented and distributions-oriented approaches to forecast verification. *Weather and Forecasting*, 11:288–303, 1996.
- Dixon, M. und Wiener, G. Titan: Thunderstorm identification, tracking, analysis, and nowcasting - a radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10:785–797, 1993.
- Doswell-III, C. A., Davies-Jones, R. P., und Keller, D. L. On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, 5:576–585, 1990.
- Dotzek, N. Radarmeteorologie. Blockvorlesung, FU Berlin, März 2010. URL <http://www.essl.org/people/dotzek/lectures/WS/Vorlesung-WS200910-03.pdf>.
- Flueck, J. A. A study of some measures of forecast verification. *American Meteorological Society*, Preprints, Tenth Conference on Probability and Statistics in Atmospheric Sciences:69–73, 1987.

- Gilbert, G. K. Finley's tornado predictions. *American Meteorological Journal*, 1: 166–172, 1884.
- Grimbacher, T. Das Distrometer, 2002. URL <http://iacweb.ethz.ch/en/groups/richter/cd/doc/disdrometer/>.
- Gysi, H., 2010. URL <http://radar-info.fzk.de/abc.html>.
- Hagedorn, R. Ensemble Verification 1, 2009. URL http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/PR/EnsVer1.pdf. ECMWF Trainingscourse.
- Hailzl, G. Skriptum zur Radarmeteorologie. 2006. LV an der Universität Wien, Kontakt: AustroControl GmbH, 1030 Wien, Schnirchgasse 11a.
- Haimberger, L. Vorhersagbarkeit und Ensemblevorhersage. 2008. LV an der Universität Wien, verfügbar vom Autor (Universität Wien, Institut für Meteorologie und Geophysik, 1090 Wien, Althanstrasse 14).
- Handwerker, J. Cell tracking with TRACE3D - a new algorithm. *Atmospheric Research*, 61:15–34, 2002.
- Hanssen, A. W. und Kruipers, W. J. A. On the relationship between the frequency of rain and various meteorological parameters. *Mededeelingen en Verhandelingen*, 81:2–15, 1965.
- Heidke, P. Berechnung des Erfolges und der Güte der Windstärkenvorhersagen im Sturmwarnungsdienst. *Geografika Annaler*, 8:301–349, 1926.
- Hering, A. M., Morel, C., Galli, G., Sényesi, S., Ambrosetti, P., und Boscacci, M. Nowcasting thunderstorms in the alpine region using a radar based adaptive thresholding scheme. *Proceedings of ERAD*, Seiten 206–211, 2004.
- Heuel, E. M. *Quantitative Niederschlagsbestimmung aus Radardaten*. Dissertation, Rheinischen Friedrich Wilhelms Universität Bonn, April 2004.
- Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. DeWayne, Stumpf, G. J., Eilts, M. D., und Thomas, K. W. The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. *Weather and Forecasting*, 13(2):263–276, 1998.

- Kaltenböck, R. Applications of weather radar in aviation weather service provision in austria. *Metalliance Aviation Training Seminar. DWD, Langen, Germany*, 2007.
- Kaltenböck, R., Hennermann, K., Kerschbaum, M., und Mayer, S. Wetterradar- state of the art. 2010a. Projektinterner Bericht (DIBMETSAT)- D2.2.1 state of the art WXR.v2.0, Kontakt: AustroControl GmbH, 1030 Wien, Schnirchgasse 11a.
- Kaltenböck, R., Hennermann, K., Kerschbaum, M., und Mayer, S. Applications and visualisation of ensemble weather radar extrapolation in austrian aviation weather service provision. *ERAD 2010 - The sixth european conference on radar in meteorology and hydrology*, 6, 2010b.
- Kessler, E. und Russo, J. A. J. A programm for the assembly and display of the radar-echo distributions. *Journal of Applied Meteorology*, 5:582–593, 1963.
- Kyznarová, H. und Novák, P. Development of cell-tracking algorithm in the czech hydrometeorological institute. *Proceedings of the World Weather Research Programme Symposium on Nowcasting and very short Range Forecasting (WSN05), Toulouse, France*, 5-9 September 2005.
- Li, L., Schmid, W., und Joss, J. Nowcasting of motion and growth of precipitation with radar over a complex orography. *Journal of Applied Meteorology*, 34:1286–1300, 1995.
- Marshall, J. S. und Palmer, W. McK. The distribution of raindrops with size. *Journal of Meteorology*, 5:165–166, 1948.
- Mason, I. A model for assessment of weather forecasts. *Australian Meteorological Magazine*, 30:291–303, 1982.
- Mecklenburg, S. *Nowcasting precipitation in an Alpine region with a radar echo tracking algorithm*. Dissertation, ETH Zürich, 2000.
- Murphy, A. H. What is a good forecast? An essay of the nature of goodness in weather forecasting. *Weather and Forecasting*, 8:281–293, 1993.
- Murphy, A. H. und Katz, R. W. Economic value of weather and climate forecasts. *Cambridge University Press*, Seite 222 pp., 1997a.

- Murphy, A. H. und Winkler, R. L. A general framework for forecast verification. *Monthly Weather Review*, 115:1330–1338, 1987.
- Novák, P. The czech hydrometeorological institute’s severe storm nowcasting system. *Atmospheric Research*, 83:450–457, 2007.
- Ohtake, T. Factors affecting the size distribution of raindrops and snowflakes. *Journal of Atmospheric Sciences*, 27:804 – 813, 1970.
- Peirce, C. S. The numerical measure of the success of predictions. *Science*, 4:453–454, 1884.
- Persson, A. und Grazzini, F. *User Guide to ECMWF forecast products*. ECMWF, 4.0 Edition, März 2007. URL http://www.ecmwf.int/products/forecasts/guide/user_guide.pdf.
- Peura, M. und Hohti, H. Optical flow in radar images. *Proceedings of the Third European Conference on Radar Meteorology (ERAD)*, Visby, Sweden, Seiten 454–458, 6-10 September 2004.
- Randeu, W. L., Schönhuber, M., und Lammer, G. Real-time measurements and analyses of precipitation micro-structure and dynamics. In *Proceedings of ERAD 2002*, Seite 78 ff, 2002.
- Reyniers, M. Quantitative precipitation forecasts based on radar observations: principles, algorithms and operational systems, 2008. URL <http://radar.meteo.be/meteo/download/en/3412634/pdf/qpfatrmirev.pdf>.
- Rinehart, R. E. und Garvey, E. T. Three-dimensional storm motion detection by conventional weather radar. *Nature*, 273:287–289, 1978.
- Schaefer, J. T. The critical success index as an indicator of warning skill. *Weather and Forecasting*, 5:570–575, 1990.
- Schneider, S. *Analyse von Niederschlagsereignissen unter Verwendung von Boden- und Radardaten*. Dissertation, Universität Wien, Oktober 2008.
- Steinacker, R. Diagnose und Prognose der Schneefallgrenze. *Wetter und Leben*, 35: 81–90, 1983.

UCAR. TITAN -Thunderstorm Identification Tracking Analysis and Nowcasting-Homepage, September 2010. URL <http://www.rap.ucar.edu/projects/titan/home/>.

Wilks, Daniel S. *Statistical Methods in the atmospheric sciences*, Kapitel 7, Seiten 255ff. Academic Press, Second Edition, 2006.

Wolff, C. Radartutorial: Wetterradar-Störechos, 2010. URL <http://www.radartutorial.eu/15.weather/wx24.de.html>.

Wuest, M. Atmosphärenphysik iii (iac eth), 2005. URL http://www.iac.ethz.ch/people/wueest/atmospheric_physics_iii/apiii_streuung_radargleichung.pdf.

Abbildungsverzeichnis

2.1	Arbeitsweise: TREC-Algorithmus nach Rinehart und Garvey (1978) . .	9
2.2	Möglichkeiten der Zellendetektion aus Handwerker (2002)	11
2.3	TITAN: Beschreibung der Zellenform (Quelle: TITAN-Homepage) . . .	13
2.4	Österreichisches Wetterradarnetz aus Kaltenböck (2007)	16
2.5	Vergleich der gängigen Einheiten für WXR-Reflektivitäten	18
2.6	Überblick: WXR-Austria-Composite MAXCAPPI	19
2.7	Beispiele für TREC-Extrapolationen berechnet bei AustroControl GmbH	23
2.8	Optimierungsschema nach Kaltenböck et al. (2010b)	24
3.1	Schematische Darstellung des Pulsvolumens (Heuel, 2004)	28
3.2	Brightband-Effekt (Hailzl, 2006)	30
3.3	Exponentialverteilung nach Marshall und Palmer (1948)	32
3.4	Wirkung des Ground-Clutter-Filters	36
3.5	Brightband-Effekt in Summenbild	37
3.6	Sonnenecho (Kaltenböck et al., 2010a)	38
3.7	Richt- oder Antennendiagramm (Hailzl, 2006)	40
3.8	Kreisförmige und spiralförmige Störungen im Tagesmittel	41
3.9	RLAN-Störung (Kaltenböck et al., 2010a)	42
3.10	WXR-Störung durch Tiefflugerfassungsradar (Kaltenböck et al., 2010a)	43
3.11	WXR-Abschattungsgebiete (Kaltenböck et al., 2010a)	44
4.1	2x2 Kontingenztafel, nachgebildet nach Wilks (2006)	51
4.2	Layout eines klassischen Reliability-Diagramms (Hagedorn, 2009) . . .	58
4.3	Layout des adaptierten Reliability-Diagramms	59
4.4	Reliability-Scores nach Hagedorn (2009)	59
4.5	Bias Information aus dem Reliability-Diagramm	60
4.6	Histogramm der bedingten Beobachtungen	60
4.7	Reliability-Diagramme unterschiedlicher Resolution	61
4.8	Schematisches ROC-Diagramm nach Hagedorn (2009)	63
4.9	Histogramm der bedingten Vorhersagen	64

5.1	Verification Tool: Startfenster	71
5.2	Reliability Table Calculation Tool	73
5.3	Skill Score Calculation Tool	76
5.4	ROC Plotting Tool	78
5.5	Score vs. Time Plotting Tool	80
5.6	Reliability Diagram Plotting Tool	81
5.7	Ranked Probabilities Plotting Tool	82
6.1	Vergleich von ME und MAE der vier Parameterkombinationen (Winter)	86
6.2	Marginale Verteilungen von Vorhersage und Beobachtung (Winter) . .	87
6.3	CSI und ME des dichotomen Systems (Winter)	88
6.4	POD und FAR Entwicklung (Winter)	90
6.5	ROC-Diagramm: PS-Vergleich (Winter)	92
6.6	ROC-Diagramm für PS 1 (Winter)	93
6.7	Reliability bei unterschiedlichen Extrapolationsschritten (Winter) . . .	95
6.8	Reliability-Vergleich der Parameterkombinationen (Winter)	96
6.9	Bedingte Beobachtungen und Vorhersagen von PS 1 (Winter)	97
6.10	Ausfall des Ground-Clutter-Filters des WXR Salzburg/Feldkirchen . .	99
6.11	Vergleich der Skills der Extrapolationstypen (Winter)	100
6.12	Vergleich von ME und MAE aller Parameterkombinationen (Sommer) .	101
6.13	CSI und ME des dichotomen Systems (Sommer)	103
6.14	POD und FAR Vergleich (Sommer)	104
6.15	ROC-Diagramm (Sommer)	105
6.16	Reliability bei unterschiedlichen Extrapolationsschritten (Sommer) . . .	107
6.17	Marginale Verteilungen von 60 minütiger Vorhersage und Beobachtung (Sommer)	108
6.18	Bedingte Beobachtungen und Vorhersagen von PS 2 (Sommer)	109
6.19	Abschattung hinter einer Starkniederschlagszelle	110
6.20	Kreisförmige Störung hoher Intensität	111
6.21	Skill-Vergleich (Sommer)	112
7.1	Prozessbaum-Verifikationssoftware	128

Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.

Tabellenverzeichnis

1.1	Kennzeichen guter Vorhersagen (Murphy, 1993)	4
2.1	Koordinaten der österreichischen Wetterradarstationen aus Kaltenböck et al. (2010a)	15
2.2	Parametergruppen der TREC-Extrapolation	21
3.1	Potenzgesetzparameter für Tropfen-Endgeschwindigkeit (Hailzl, 2006) .	33
3.2	Gängige ZR-Beziehungen aus Hailzl (2006)	33
4.2	Verteilungen für die Vorhersagequalität (Murphy, 1993)	50
4.3	Definition der „Reliability“-Maße nach Doswell-III et al. (1990)	57
6.1	Farbgebung und Nummerierung der Parameterkombinationen	84

Abkürzungsverzeichnis

ACG Österreichische Flugwetterdienst (AustroControl GmbH)

CHMI Tschechischer Wetterdienst

COTREC Continuity of TREC Vectors

CSI Critical Success Index

CT Cell Tracking

dBZ logarithmische Einheit für den Radarreflektivitätsfaktor

ETS Equitable Threat Score

FAR False Alarm Ratio

HSS Heidke Skill Score

MAE Mean Absolute Error

MAXCAPPI Maximum Constant Altitude Plan Position Indicator

ME Mean Error

NWP Numeric Weather Prediction

PC Proportion Correct

POD Probability of Detection

POFD Probability of False Detection

PS Parameterset

RMSE Root Mean Square Error

ROC Relative Operating Characteristics

SCIT Storm Cell Identification and Tracking

TAWES Teilautomatisches Wetter Erfassungs System

TITAN Thunderstorm Identification, Tracking, Analysis and Nowcasting

TREC Tracking Echos by Correlation

TRT Thunderstorm Radar Tracking

TSS True Skill Statistics

VVP Volume Velocity Processing

WXR Wetterradar

Z Radarreflektivitätsfaktor

Anhang

Dateinamenskonventionen - TREC

Die Qualität einer TREC-Extrapolation ist, wie sich in der Praxis gezeigt hat, stark von einzelnen Parametereinstellungen abhängig. Daher sind diese im jeweiligen Dateinamen vermerkt. Außerdem enthält er noch Informationen bezüglich des Ausgangszeitpunktes der Extrapolation, sowie die Vorhersagezeitspanne. Ein gültiger Dateiname könnte wie folgt aussehen:

WIIS_ARAD_IMG_5min_b7_corr56_opt2__0901010605.010.png

Zur weiteren Betrachtung wurde der Namen in drei Gruppen unterteilt, die durch die unterschiedliche Farbgebung erkennbar sind. Der erste, hier rote Bestandteil des Dateinamens bezeichnet die Art der Bilder, aus denen die Extrapolation berechnet wurde. Es handelt sich dabei um den MAXCAPPI Grundriss des WXR-Austria-Composite-Bildes (siehe Abschnitt 2.2.1). Alle hier verwendeten Extrapolationen entstammen derselben Art von Ausgangsbildern, dieser Teil des Namens bleibt daher für alle Extrapolationen gleich. Der zweite, blaue Bereich des Dateinamen ist der wohl interessanteste. Er beinhaltet die Parametereinstellungen, deren Kombination die Güte der Extrapolation maßgeblich beeinflusst. Genau betrachtet handelt es sich um fünf Einstellungsmöglichkeiten, deren genauere Bedeutung aus Tabelle 2.2 ersichtlich ist. Es sei an dieser Stelle nochmals erwähnt, dass nicht alle möglichen Parameterkombinationen im Testdatensatz auch wirklich vorhanden sind. Um den Ressourcenaufwand in Grenzen zu halten, wurden im Vorfeld vier Kombinationen ausgewählt. Diese lauten wie folgt:

- WIIS_ARAD_IMG_5min_b7_corr56_opt2__*.png
- WIIS_ARAD_IMG_5min_g5_corr16_opt2__*.png
- WIIS_ARAD_IMG_5min_g5_corr32_opt2__*.png
- WIIS_ARAD_IMG_15min_g5_sosq56_grid2__*.png

Der letzte, grün markierte Abschnitt in obiger Dateinamenkonvention beinhaltet die Zeitinformation. Die zehnstellige Ziffernfolge zu Beginn ist der Zeitstempel des späteren Ausgangsbildes und somit (ungefähr) der Zeitpunkt, zu dem die Extrapolation berechnet wurde. Im Dateinamen-Beispiel findet man die Ziffernkette „0901010605“. Diese Extrapolation hat also das WXR-Bild vom 1. Jänner 2009 um 06:05h zur Grundlage. Der zweite Teil des grün eingefärbten Abschnitts steht für den Vorhersageschritt (in Minuten). „010“ kennzeichnet also eine 10-minütige Vorhersage. Aus dem Dateinamen ist weiter ersichtlich, dass alle Extrapolationsergebnisse als PNG-Dateien abgespeichert werden.

Prozessbaum-Verifikationssoftware

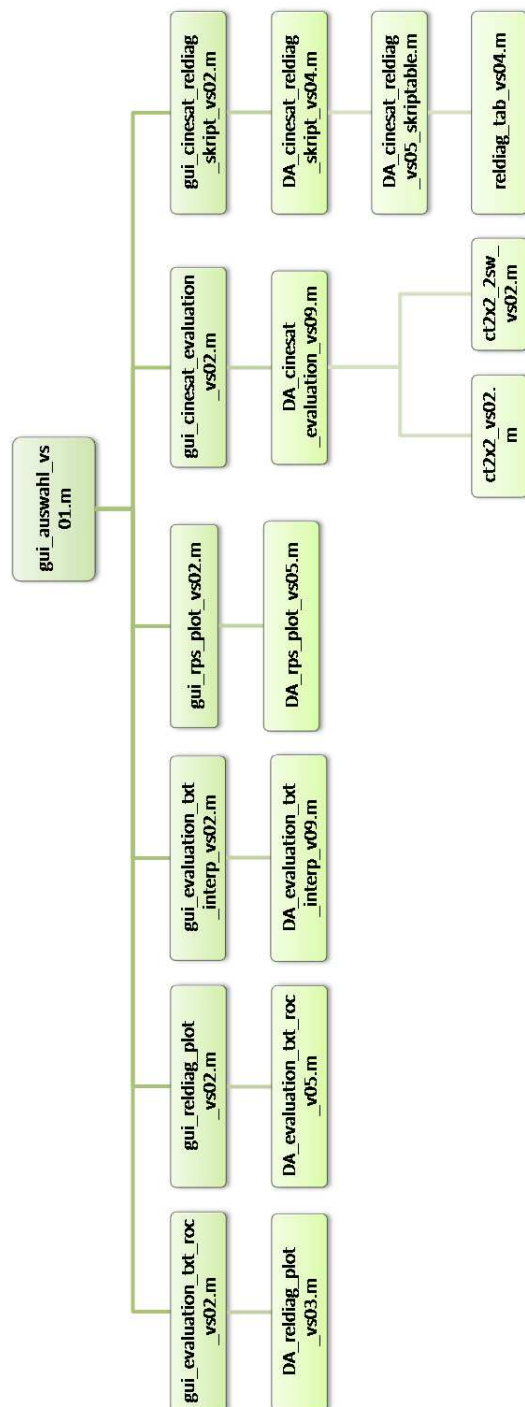


Abbildung 7.1:

Der Prozessbaum beinhaltet die entscheidenden Programme die im Rahmen des getesteten Verifikationssystems verwendet wurden sowie ihre logische Verknüpfung. Rückkopplungen, die der Ausfallssicherheit dienen sind nicht berücksichtigt.

Danksagung

An dieser Stelle möchte ich all jenen danken, die mich während der Entstehung dieser Arbeit und meines Studiums unterstützten. Zu allererst meinem Betreuer Prof.Dr. Leopold Haimberger, der sehr geduldig mit mir war und sich immer wieder, oft auch kurzfristig Zeit nahm, mir meine Fragen zu beantworten und dessen Ideen maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Weiter möchte ich Mag. Rudolf Kaltenböck und Dr. Markus Kerschbaum von Austro-Control GmbH danken, die mich immer wieder motivierten und mir auch tatkräftig zur Seite standen.

Besonderer Dank gebührt meinen Eltern, ohne deren Unterstützung ich diese Ausbildung nie absolvieren hätte können. Zum Schluss danke ich noch all meinen Freunden, Studienkollegen und speziell meiner Freundin für eine unvergessliche und schöne Studienzeit, aber auch für alle aufmunternden Worte in den „zähen“ Abschnitten.

Curriculum Vitae

PERSÖNLICHE DATEN

Name: Stefan Mayer

Staatsbürgerschaft: Österreich

Geburtstag: 03.06.1985

Geburtsort: Neunkirchen

Familienstand: ledig

AUSBILDUNG:

1995-1999 Gymnasium: Unterstufe BRG Neunkirchen

1999-2004 HTBLuVA Wiener Neustadt (Automatisierungstechnik)

2004 Matura mit ausgezeichnetem Erfolg

2004-2005 Ableistung des Präsenzdienstes

2005-2007 Studium der Meteorologie und Geophysik an der Universität Wien

seit 2007 Studium der Meteorologie an der Universität Wien

ARBEITSPRAXIS:

seit Mai 2007 Meteorologe bei der MeteoServe Wetterdienst GmbH

Februar 2007 Praktikum bei Austro Control GmbH in der Abteilung Meteorologie

Dezember 2006-Mai 2007 Werkvertrag mit Semperit Wimpassing, Physik. Labor

September 2006 Praktikum bei Semperit Wimpassing (Physikalisches Labor)

Sommer 2002-2004 insgesamt drei Monate Praktikum bei Hoerbiger Origa Pneumatik

Juli 2000 Praktikum bei Schoeller-Bleckmann Edelstahlrohr GmbH