

**Entwicklung eines automatisierten
Wolkendetektions- und
Wolkenklassifizierungsverfahrens
mit Hilfe von Support Vector Machines
angewendet auf METEOSAT-SEVIRI-Daten
für den Raum Deutschland**

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Rainer Vogel

aus

Bonn-Duisdorf

Bonn 2011

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Gunter Menz

2. Gutachter: Prof. Dr. Klaus Greve

Tag der Promotion: 27.08.2012

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn publiziert:

<http://hss.ulb.uni-bonn.de>

Erscheinungsjahr: 2013

Danksagung

Vorliegende Arbeit entstand in Verbindung mit der Betreuung der METEOSAT-Empfangsanlage am Zentrum für Fernerkundung der Landoberfläche (ZFL). Schon während meiner Diplomarbeit ging es um die Problematik der Wolkenableitung aus METEOSAT-Daten, weshalb eine weitere Auseinandersetzung mit diesem Themenbereich in einer nachfolgenden Arbeit nahe lag.

Daher gilt in erster Linie mein Dank Herrn Prof. Dr. Gunter Menz für die Betreuung und Unterstützung während des gesamten Ausarbeitungsprozesses. In zahlreichen Gesprächen und Anregungen konnte so der Themenkomplex eingegrenzt und effektiv gestaltet werden. Zudem konnte diese Arbeit nur durch die umfangreichen technischen Mittel realisiert werden, welche ausnahmslos zur Verfügung gestellt wurden. Dies gilt insbesondere für das teilweise kostenintensive Prozessierungs- und Archivierungssystem der Satellitenempfangsanlage.

Weiterer Dank gilt meinem Vorgänger Dr. Torsten Welle, der den Aufbau der Empfangsanlage mit initiiert, und mir in der Einarbeitungsphase die Übernahme erleichtert hat. Aufgrund der vielfältigen Hilfestellungen bei technischen Problemen bin ich auch dem Support der Firma VCS AG zu Dank verpflichtet. Gleiches gilt für die Systemadministratoren der Arbeitsgruppe Fernerkundung (RSRG) und des ZFL, Tomasz Dobrzeniecki und Andreas Mähler.

Mein besonderer Dank gilt zudem allen Mitgliedern und ehemaligen Mitgliedern von RSRG und ZFL, die mich durch vielfältige Gespräche, Diskussionen, Anregungen und Hilfestellungen unterstützt haben. Dies betrifft in Bezug auf Archivierung und Nachprozessierung der METEOSAT-Daten vor allem Susanne Meyfarth, Raphael Knoth und Andreas Rienow. Ingo Elbertzhagen danke ich für seine Hilfestellung bei der Justierung der Empfangsantenne. Für ihre Anregungen in Sachen Fernerkundungsverfahren, Programmierung und Geographische Informationssysteme danke ich Dr. Doris Klein, Roland Götzke, Dr. Pierre Fritzsche, Dr. Thorsten Mewes, Henryk Hodam, Dr. Côme Linsoussi und Konrad Hentze.

Für die Bereitstellung der Wetter-Kamera-Daten des Jahres 2008 danke ich Thomas Dümmler (Meteorologisches Institut FU Berlin), Michael Zuther (Mineralogie Universität Bremen), John Kalisch und Andreas Macke (IFM-GEOMAR Kiel), Michael Kaes (WDR Köln), Gerhard Rabitsch (Info Wetter Mannheim) sowie Robert Goler (Meteorologisches Institut Universität München). Zudem danke ich dem SAFNWC-Help-Desk-Team für ihre zügigen Hilfestellungen in Bezug auf die Prozessierung mit der SAFNWC-Software. Ebenfalls gilt mein Dank dem

Geomatik Department der Humboldt-Universität zu Berlin für die Anfertigung und Bereitstellung von imageSVM, inklusive Quellcode (Bereitgestellt über Torsten Welle). Zudem danke ich Horst Nitschmann und Annette Ortwein für ihre umfangreiche Suche nach Rechtschreibfehlern.

Abschließender Dank gilt meinen Eltern ebenfalls für ausgiebiges Korrekturlesen und ihre sonstige Unterstützung.

Inhaltsverzeichnis

Abbildungsverzeichnis	8
Tabellenverzeichnis	12
Abkürzungsverzeichnis	15
Zusammenfassung	17
Summary	19
1 Einleitung	21
1.1 Wolken als Untersuchungsgegenstand	21
1.2 Wolken in der Satellitenfernerkundung	22
1.3 Untersuchungsgebiet Deutschland	25
1.4 Ziel der Arbeit und Fragestellungen	26
2 Wolken als atmosphärisches Phänomen	30
2.1 Wolkenbildung	30
2.2 Wolkeneinteilung nach Höhe und Gestalt	31
2.3 Wolken und atmosphärische Dynamik	35
2.3.1 Zyklogenese als Ursache der Bewölkungsverhältnisse in Europa	35
2.3.2 Einflussfaktor Orographie	39
2.3.3 Zirkulationsdynamische Variabilität und Wolkenbedeckung	40
2.4 Die Rolle der Wolken im Strahlungshaushalt	42
2.4.1 Atmosphärischer Einfluss auf Strahlungsvorgänge	42
2.4.2 Wirkungsweise von Wolken auf Strahlungsvorgänge	45
2.5 Wolken und Klima in Deutschland	48
3 Wolkendetektionsverfahren basierend auf Schwellenwertalgorithmen	51
3.1 APOLLO	51
3.2 SAFNWC	54
3.2.1 Cloud Mask	55

3.2.2	Cloud Type	61
4	Datengrundlage METEOSAT-SEVIRI	63
4.1	METEOSAT Second Generation	63
4.1.1	SEVIRI	64
4.2	Vorverarbeitung und Übertragung der METEOSAT-Daten	66
4.3	Kalibrierungsverfahren	67
4.3.1	Kalibrierung der thermalen Kanäle	67
4.3.2	Kalibrierung der solaren Kanäle	69
4.3.3	Geometrische Genauigkeit	70
4.4	Datenempfang und Prozessierung am Zentrum für Fernerkundung der Land- oberfläche	71
4.5	Erstellung der SAFNWC-Produkte	72
4.6	DWDSAT als Validierungsgrundlage	75
5	Automatisierter Ansatz der satellitenbasierten Wolkendetektion und Wolkenklas- sifizierung mit Support Vector Machines	77
5.1	Support Vector Machines als Klassifikator	82
5.2	Umsetzung mit imageSVM	86
5.3	Erste Schritte für die Gewinnung von Trainingsdaten	88
5.3.1	Wetter-Kameras zur Erfassung von Wolkeninformationen	89
5.3.2	Abschätzung Grenzbereich „Wolkenkontaminiert - Wolkenfrei“	94
5.4	Generierung weiterer Trainingsdaten und Anwendung der SVM-Wolkenmaske	102
5.5	Anwendung der weiter differenzierten SVM-Wolkenklassifizierung	110
6	Validierung der Wolkendetektion mit synoptischen Beobachtungsdaten	117
6.1	Umsetzung des Validierungsverfahrens	118
6.2	Validierungsergebnisse SVM-Wolkenmaske im Vergleich mit SAFNWC-Cloud- mask	123
7	Ergebnisse und Anwendungsbeispiele für das Jahr 2008	134
7.1	Anwendungsbeispiele der SVM-Wolkenklassifizierung	134
7.2	Ergebnisdarstellung anhand des Wolkenbedeckungsgrades	136
7.2.1	SVM-Wolkenmaske im Vergleich mit SAFNWC-Cloudmask	137
7.2.2	SVM-Wolkenklassifizierung im Vergleich mit SAFNWC-Cloudtype	145
8	Einordnung der Ergebnisse mit Hilfe von Großwetterlagen	155

9 Diskussion und Ausblick	162
9.1 Herausforderungen für die Zukunft	165
Literatur	177

Abbildungsverzeichnis

1.1	Satellitenaufnahme von Europa und Nordafrika mit METEOSAT-9 (MSG-2), RGB-Falschfarben.	23
2.1	Schematische Darstellung (inklusive Wettersvorgänge) einer Zyklone in Verbindung mit Warm- und Kaltfront.	37
2.2	Wolkenbildung bei der Überströmung von Erhebungen (Gebirgen), inklusive Temperatureffekten im Lee- und Luv-Bereich.	39
2.3	Wolkenbedeckungsgrad (%) während positiver NAO-Phasen von Dezember bis März (oben). Abgeleitet aus Satellitenbeobachtungen für den Zeitraum 1983 - 1990. Differenz des Bedeckungsgrades (%) zwischen positiver und negativer NAO-Phase (unten).	41
2.4	Das elektromagnetische Spektrum in Verbindung mit der Strahlungsenergie der Sonne und der Durchlässigkeit der Atmosphäre.	43
2.5	Strahlungshaushalt des Systems Erde-Atmosphäre.	44
2.6	Cloud Forcing in Abhängigkeit der Wolkenhöhe und optischen Dicke. Solare Strahlung (SOL), Infrarotstrahlung (IR).	47
2.7	Räumliche Differenzierung des Cloud Forcing (Beispiel: Juli 1985). Fehlende Werte sind grau dargestellt.	48
3.1	APOLLO Tag-Testreihe. Abgrenzung wolkenfreier und wolkenkontaminierter Pixel (First Run) und Bestimmung völlig bewölkter Pixel (Second Run). T4 und T5 = Strahlungstemperaturen (Kanal 4 und 5), R1 und R2 = Reflektanzen (Kanal 1 und 2), SD = Standardabweichung.	53
3.2	SAFNWC-Cloudmask (CMA), Teilausschnitt Europa, 06.09.2010 13:00 UTC.	57
3.3	Testsequenz der SAFNWC-Cloudmask (CMA) über Landflächen. Die Beschreibung der Einzeltests befindet sich im Text. θ_{sat} = Satelliten-Zenitwinkel.	58
3.4	SAFNWC-Cloudtype (CT), Teilausschnitt Europa, 06.09.2010 13:00 UTC.	62
4.1	Aufnahme der „Erdscheibe“. Abnehmende Auflösung vom Sub-Satelliten-Punkt nach außen. Die Pixel-Auflösung beträgt im inneren Kreis 3 km, gefolgt von 4, 5, 6, 8 und 11 km (äußerer Kreis).	64

4.2	MSG-Aufnahmeverfahren.	66
4.3	Datenübertragungswege von METEOSAT Second Generation.	67
4.4	Schematische Darstellung der METEOSAT-Empfangsanlage am ZFL, inklusive Datenarchivierung.	73
5.1	Kombinierte Wolkenerkennung und Wolkenklassifizierung basierend auf Support Vector Machines (SVM) für jeden METEOSAT-Zeitschnitt, einschließlich Validierungsverfahren und Vergleich mit SAFNWC-Produkten.	78
5.2	Untersuchungsgebiet extrahiert aus dem gesamten METEOSAT-Aufnahmebereich. Ausschnitt Deutschland bzw. Zentraleuropa (327 x 271 Pixel) dargestellt als RGB. Weitere Eingrenzung durch zusätzliche Maske, inklusive Ausmaskierung der Meeresflächen.	80
5.3	Merkmalsraum der SVM-Klassifizierung, bestehend aus wolkenrelevanten Infrarotkanälen und Kanal-Differenzen. Ausmaskierter Ausschnitt Deutschland.	81
5.4	Verschiedene Möglichkeiten der Trennung zweier Klassen (zweidimensionaler Raum).	83
5.5	Optimale Einbindung der Hyperebene mit entsprechenden Stützvektoren (Support Vectors).	84
5.6	Überführung in einen höher dimensionalen Raum, zur Lösung eines nicht linear trennbaren Klassifizierungsproblems.	85
5.7	Wetter-Kameras - Räumliche Verteilung der Standorte.	88
5.8	Wetter-Kameras - Ausmaskierte Himmelsfläche.	90
5.9	Unterschiedliche Aufnahmen der Wetter-Kamera München.	93
5.10	Abschätzung eines ungefähren Grenzwertes/Grenzbereiches anhand des Strahlungstemperaturbereiches von Kanal IR120. Grenzwert bzw. Grenzbereich dienen der späteren Abgrenzung generierter Trainingsdaten.	96
5.11	Histogramme (Zeitschnitte aus April und Juli), inklusive Cluster-Analyse und Abschätzung eines ungefähren Grenzwertes nach den in Tabelle 5.2 aufgeführten Bedingungen. Cluster 1 (Rot), Cluster 2 (Grün), Cluster 3 (Blau) und Cluster 4 (Gelb). Die rote Linie im Histogramm stellt den gleitenden Mittelwert über 20 Datenpunkte dar.	100
5.12	Histogramme (Zeitschnitte aus Oktober und Dezember), inklusive Cluster-Analyse und Abschätzung eines ungefähren Grenzwertes nach den in Tabelle 5.2 aufgeführten Bedingungen. Cluster 1 (Rot), Cluster 2 (Grün), Cluster 3 (Blau) und Cluster 4 (Gelb). Die rote Linie im Histogramm stellt den gleitenden Mittelwert über 20 Datenpunkte dar.	101

5.13	Ablaufschema des Wolkendetektions- und Wolkenklassifizierungsverfahrens, inklusive Trainingsdatengewinnung für jeden METEOSAT-Zeitschnitt	103
5.14	Erzeugung weiterer Trainingsdaten mit Hilfe logischer Verknüpfungen in Bezug auf bekannte Schwellenwertalgorithmen, inklusive Zuordnung über Regions of Interest (ROIs). Umfang der niedrigsten und höchsten Werte und die Maximale Anzahl der Trainingspixel sind auf den Gesamtumfang der Szene Deutschland abgestimmt.	106
5.15	Beispielhafte Darstellung verwendeter Trainingspixel der SVM-Wolkenmaske .	109
5.16	Erzeugung von Trainingsdaten für die SVM-Klassifizierung mit Hilfe verknüpfter Schwellenwertalgorithmen. Zuordnung der vier weiteren Klassen erfolgt über Regions of Interest (ROIs). Umfang der kleinsten und größten Werte sowie die maximale Anzahl an Trainingspixeln sind auf den Gesamtumfang der Szene Deutschland abgestimmt.	112
5.17	Beispielhafte Darstellung verwendeter Trainingspixel der weiter differenzierten SVM-Wolkenklassifizierung.	114
6.1	Standorte der durch DWDSAT verfügbaren WMO-Stationsdaten (Deutschland), für die Validierung des Wolkendetektionsverfahrens.	119
6.2	Vergleich der Validierungsergebnisse als grafische Darstellung.	125
6.3	Effekt der SVM-Überschätzung.	126
7.1	Vergleich zwischen SVM-Wolkenklassifizierung und SAFNWC-Cloudtype anhand verschiedener Einzelbeispiele.	135
7.2	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Januar und Februar 2008.	139
7.3	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. März und April 2008.	140
7.4	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Mai und Juni 2008.	141
7.5	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Juli und August 2008.	142
7.6	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. September und Oktober 2008.	143

7.7	Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. November und Dezember 2008.	144
7.8	Wolkenbedeckungsgrad der Klasse „Hoch Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008.	147
7.9	Wolkenbedeckungsgrad der Klasse „Hoch Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008.	148
7.10	Wolkenbedeckungsgrad der Klasse „Hoch Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008.	149
7.11	Wolkenbedeckungsgrad der Klasse „Hoch Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008.	150
7.12	Wolkenbedeckungsgrad der Klasse „Niedrig Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008.	151
7.13	Wolkenbedeckungsgrad der Klasse „Niedrig Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008.	152
7.14	Wolkenbedeckungsgrad der Klasse „Niedrig Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008.	153
7.15	Wolkenbedeckungsgrad der Klasse „Niedrig Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008.	154
8.1	Vergleich des durchschnittlichen Wolkenbedeckungsgrades über der ausmasierten Fläche Deutschland (vgl. Abbildung 5.2) mit Kennzahlen (vgl. Tabelle 8.1) der objektiven Wetterlagenklassifikation des DWD. Bedeckungsgrad berechnet für jeweils 10 Tage.	158

Tabellenverzeichnis

2.1	Wolkenstockwerke nach der internationalen Klassifikation.	32
2.2	Hauptwolkengattungen. Hohe und mittelhohe Wolken (a). Tiefe und vertikale Wolken (b).	33
3.1	NOAA-AVHRR-Kanäle die zur Wolkendetektion genutzt werden.	52
3.2	Liste der Input-Daten für SAFNWC-Cloudmask (CMa) und -Cloudtype (CT). .	56
3.3	Output SAFNWC-Cloudmask (CMa) und -Cloudtype (CT).	61
4.1	METEOSAT Second Generation - Kanalübersicht.	65
4.2	Werte der Korrekturkoeffizienten für die jeweiligen Infrarotkanäle.	69
4.3	Verschiedene Werte von I_m zur Kalibrierung der solaren Kanäle.	70
4.4	Archivierte Datensätze am ZFL in Bonn.	74
5.1	Standorte der Wetter-Kameras und ihre jeweiligen Eigenschaften. Aufnahmeintervalle beziehen sich auf die letztendlich genutzte zeitliche Auflösung. Berlin, Bremen und Kiel (a). Köln, Mannheim und München (b).	91
5.1	Begutachtung der automatisierten Wolkenableitung durch Wetter-Kameras basierend auf visueller Inspektion. Mit drei (oben) und mit zwei (unten) Kategorien. Beobachtete Werte sind in den Spalten, detektierte Werte in den Zeilen dargestellt.	95
5.2	Histogramm-Bedingungen für die nachfolgende Auswahl des ungefähren Grenzwertes. T = Strahlungstemperatur Kanal IR120.	98
5.3	Zuweisung der Cluster (Grenzwert/Grenzbereich) durch Interpretation des Histogramms (Bedingungen). Wenn Bedingung 2 erfüllt ist, muss Bedingung 1 ebenfalls erfüllt sein, wodurch die Anzahl möglicher Kombinationen reduziert wird.	99
5.4	Tag-Zeitfenster indem Kanäle des sichtbaren Wellenlängenbereiches für die Trainingsdatengewinnung eingesetzt werden können. Die Auswahl erfolgt durch die Betrachtung des verwendeten Ausschnitts Deutschland (vgl. Abbildung 5.2), welcher voll ausgeleuchtet sein muss.	105

5.5	Durchschnittliche Rechenzeiten der SVM-Wolkendetektion und SVM-Wolkenklassifizierung (ohne Validierung) für einen Zeitschnitt, abgeleitet aus rückwirkender Berechnung der ersten Hälfte des Jahres 2008.	116
6.1	Beobachtungsdaten an den WMO-Stationen. Wolkenbedeckung (Total Cloud Cover) in Achtern.	120
6.2	Aufbau der Konfusionsmatrix für die Validierung der SVM-Wolkenmaske und der SAFNWC-Cloudmask.	122
6.3	Validierungsergebnisse der automatisierten SVM-Wolkenmaske mit WMO-Stationen, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenfrei“ angesehen.	128
6.4	Validierungsergebnisse der SAFNWC-Cloudmask mit WMO-Stationen, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenfrei“ angesehen.	129
6.5	Validierungsergebnisse der automatisierten SVM-Wolkenmaske mit WMO-Stationen, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenkontaminiert“ angesehen.	130
6.6	Validierungsergebnisse der SAFNWC-Cloudmask mit WMO-Stationen, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenkontaminiert“ angesehen.	131
6.7	Validierungsergebnisse der SVM-Wolkenmaske mit WMO-Stationen, Oktober 2008. Mit und ohne Trainingsinformation abgeleitet aus Wetter-Kameras.	132
6.8	Validierungsergebnisse der SVM-Wolkenmaske und der SAFNWC-Cloudmask mit WMO-Stationen, Validierungszeitraum Juli bis Dezember 2008. Obere vier Zeilen: Unterscheidung zwischen Tag- und Nacht-Testsequenzen bei der Trainingsdatengewinnung (bezieht sich auf SVM-Anwendung, SAFNWC dient nur als Vergleich). Untere vier Zeilen: Aufteilung der Validierung in unterschiedliche Höhenlagen (Höhe der jeweiligen WMO-Station).	133
7.1	Datenverfügbarkeit der METEOSAT-Daten (alle Kanäle) und der daraus erstellten Produkte für das Jahr 2008. SVM-Wolkenmaske (SVM-WM) und SVM-Wolkenklassifizierung (SVM-WK), sowie SAFNWC-CMa und SAFNWC-CT. Die Daten setzen sich aus 96 Zeitschnitten pro Tag zusammen, alle vorhandenen METEOSAT-Kanäle werden zu einem Wert zusammengefasst. Vom 14.05.2008 00:00 UTC bis 19.05.2008 12:00 UTC und vom 01.12.2008 12:00 UTC bis 09.12.2008 12:00 UTC wurde jeweils auf METEOSAT-8 umgeschaltet.	137

- 8.1 Einteilung der objektiven Wetterlagenklassifikation des DWD. Kennung setzt sich wie folgt zusammen. Stelle 1 und 2 ergeben die Anströmungsrichtung (NO = Nordost, NW = Nordwest, SO = Südost, SW = Südwest, XX = nicht definiert). Stelle 3 bezeichnet die Zyklonalität bei 950 hPa und Stelle 4 die Zyklonalität bei 500 hPa (A = Antizyklonal, Z = Zyklonal). Stelle 5 beschreibt die Feuchte (T = Trocken, F = Feucht) 157

Abkürzungsverzeichnis

AEMET	Agencia Estatal de Meteorologia
APOLLO	AVHRR Processing scheme Over cLOUDs, Land and Ocean
AVHRR	Advanced Very High Resolution Radiometer
C _{Ma}	Cloud Mask
CT	Cloud Type
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DN	Digital Number
DVB	Digital Video Broadcast
DWD	Deutscher Wetterdienst
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FTP	File Transfer Protocol
GERB	Geostationary Earth Radiation Budget
GLT	Geographic Lookup Table
HDF	Hierarchical Data Format
HRIT	High Rate Information Transmission
HRV	High Resolution Visible
IDL	Interactive Data Language
IPCC	Intergovernmental Panel on Climate Change
K	Kelvin
LCL	Lifting Condensation Level
LRIT	Low Rate Information Transmission
LSASAF	Land Surface Analysis Satellite Application Facility
METEOSAT	Meteorological Satellite
MPEF	Meteorological Product Extraction Facility
MSG	METEOSAT Second Generation
NaN	Not a Number
NAO	Nordatlantische Oszillation
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction

OAA	<u>O</u> ne <u>A</u> gainst <u>A</u> ll
OA	<u>O</u> ne <u>A</u> gainst <u>O</u> ne
PGE	<u>P</u> roduct <u>G</u> enerator <u>E</u> lement
RGB	Farbraum <u>R</u> ot <u>G</u> rün <u>B</u> lau
ROI	<u>R</u> egion of <u>I</u> nterest
RSRG	<u>R</u> emote <u>S</u> ensing <u>R</u> esearch <u>G</u> roup
RSS	<u>R</u> apid <u>S</u> can <u>S</u> ervice
SAFNWC	<u>S</u> atellite <u>A</u> pplication <u>F</u> acility on support to <u>N</u> owcasting and Very Short- Range Forecasting
SEVIRI	<u>S</u> pinning <u>E</u> nhanced <u>V</u> isible and <u>I</u> nfrared <u>I</u> mager
SMHI	<u>S</u> wedish <u>M</u> eteorological and <u>H</u> ydrological <u>I</u> nstitute
SST	<u>S</u> ea <u>S</u> urface <u>T</u> emperature
SVM	<u>S</u> upport <u>V</u> ector <u>M</u> achines
TOA	<u>T</u> op of the <u>A</u> tmosphere
UTC	<u>U</u> niversal <u>T</u> ime <u>C</u> oordinated
WMO	<u>W</u> orld <u>M</u> eteorological <u>O</u> rganization
ZAMG	<u>Z</u> entral <u>A</u> nstalt für <u>M</u> eteorologie und <u>G</u> eodynamik
ZFL	<u>Z</u> entrum für <u>F</u> ernerkundung der <u>L</u> andoberfläche

Zusammenfassung

Wolken stellen im Klimasystem der Erde einen zentralen Faktor dar, nicht nur in Bezug auf die Niederschlagsverteilung, sondern auch in Verbindung mit Effekten auf Strahlungsvorgänge innerhalb der Atmosphäre. Eine genaue Abschätzung der komplexen Wechselwirkungen innerhalb des Systems Erde-Atmosphäre ist jedoch immer noch mit Problemen verbunden. Dies betrifft vor allem den vielfach diskutierten Klimawandel. Gerade durch die hohe zeitliche und räumliche Variabilität von Wolken ist ein differenzierter Informationsgewinn über diese von enormer Relevanz. Satellitendaten haben sich hier als Mittel der ersten Wahl herausgestellt. Dies gilt insbesondere für geostationäre Satelliten, die im Gegensatz zu polarumlaufenden Satelliten, bei vergleichbarer spektraler Information, eine höhere zeitliche Auflösung zur Verfügung stellen. Gängige Wolkenerkennungsverfahren beziehen sich vor allem auf Schwellenwertansätze und benötigen in der Regel Zusatzdaten über den Zustand der Atmosphäre. Die Schwierigkeit dieser Verfahren liegt bei der exakten Bestimmung des jeweiligen Schwellenwertes. Andere Bildverarbeitungsverfahren wie Neuronale Netze, Cluster-Analysen oder Support Vector Machines (SVM) wurden zwar schon experimentell durchgeführt, jedoch meist nicht im Sinne einer automatisierten Anwendung auf zeitlich hoch aufgelöste Datenreihen.

Ziel dieser Arbeit war demnach die Entwicklung einer Wolkendetektion bzw. Wolkenklassifizierung mit Hilfe von Support Vector Machines, angewendet auf METEOSAT-SEVIRI-Daten für den Raum Deutschland. Die Umsetzung sollte hierbei auf der reinen Bildinformation (wolkenrelevante METEOSAT-Kanäle), ergänzt durch notwendige Trainingsdaten (Ground Truth), basieren. Support Vector Machines stellen in diesem Zusammenhang einen relativ modernen Klassifikator dar, der auch mit wenigen Trainingsdaten effektiv Klassifizierungsprobleme lösen kann. Die Anwendung auf das Gebiet Deutschland ist mit den vielfältigen Wolkengegebenheiten in dieser Region begründet. Eine effektive Wolkendetektion stellt durch die relativ nördliche Lage und den Einfluss des Nordatlantiks eine besondere Anforderung dar.

Die größte Herausforderung bei einem Verfahren wie Support Vector Machines ist die exakte Auswahl repräsentativer Trainingsdaten auf deren Basis der Klassifikator lernt. Eine erste Idee war, zu diesem Zweck sogenannte Wetter-Kameras einzusetzen, die an vielen Standorten Teile des Himmels aufnehmen. Rückwirkend stellte sich jedoch die Zusammenstellung eines aussagekräftigen Datensatzes als schwierig heraus. Vielerorts werden die Aufnahmen nicht archiviert, wodurch letztendlich nur sechs Standorte für das Jahr 2008 zur Verfügung standen, die

aber dennoch innerhalb des Prozessierungsverfahrens Verwendung finden. Insgesamt wird der Trainingsdatenumfang durch die Anwendung bekannter Schwellenwertalgorithmen erweitert in dem Sinne, dass nur solche Pixel, die mit hoher Sicherheit einer Klasse angehören, einsetzbar sind. Neben der Erstellung einer Wolkenmaske erfolgt die weitere Abtrennung in vier Wolkenklassen, wobei zwischen Höhe und Strahlungsdurchlässigkeit der Wolken unterschieden wird.

Zur Überprüfung der Detektion bzw. Klassifizierung wurde, neben dem Vergleich mit den Wolkenabtrennungsverfahren der Satellite Application Facility on support to Nowcasting and Very Short Range Forecasting (SAFNWC), eine Validierung mit synoptischen Beobachtungsdaten durchgeführt. Dies bezieht sich jedoch nur auf die Wolkenmaske mit Hilfe der sogenannten Achtel-Skala. Anhand der Validierungsergebnisse und der Betrachtung des Wolkenbedeckungsgrades für das Jahr 2008 wurde deutlich, dass mit einem automatisierten SVM-Wolkendetektionsverfahren bzw. -Wolkenklassifizierungsverfahren gute Ergebnisse erzielt werden können. Dies betrifft vor allem Situationen, in denen Wolken ausgeprägt bzw. im Vergleich mit wolkenfreien Flächen ausgeglichen vorkommen. Hochdruckwetterlagen mit umfangreichen wolkenfreien Flächen führen dagegen zu einer verstärkten Wolkenüberschätzung durch den SVM-Klassifikator. Demgegenüber kommt es bei der Anwendung der SAFNWC-Cloudmask (CMA) ohne zusätzliche Informationen aus Wettervorhersagemodellen zu einer Unterschätzung, deren Ausprägung sich jedoch nicht auf bestimmte Wetterlagen zurückführen lässt.

Zudem konnte durch das Verfahren eine zeitliche und räumliche Differenzierung dargestellt werden, die sich in das allgemeine Wettergeschehen des Jahres 2008 einordnen lässt. Beispielsweise ließ sich anhand der objektiven Großwetterlagen-Klassifikation des Deutschen Wetterdienstes (DWD) eine zeitliche Verknüpfung mit dem Wolkenbedeckungsgrad darstellen. Regionale Unterschiede zeigten sich vor allem in Bezug auf die Höhenlagen der Mittelgebirge, die im Allgemeinen auch höhere Bedeckungsgrade aufweisen. Im Gegensatz dazu sind Teile Nord- und Süddeutschlands, sowie das Rheintal häufig mit geringerer Bedeckung gekennzeichnet.

Anhand der vorliegenden Arbeit konnte das Potential von Support Vector Machines, bezogen auf die automatisierte Anwendung einer Wolkendetektion, dargelegt werden. Obwohl es sich um ein sehr rechenintensives Verfahren handelt, lässt sich die gesamte Prozessierung inklusive Training für jeden METEOSAT-Aufnahmezeitpunkt in einem adäquaten zeitlichen Rahmen realisieren. Nach weiterer Optimierung, wäre eine operationelle Anwendung durchaus vorstellbar.

Summary

Clouds are an important part of the earth's climate system, not only because of their connection to precipitation, but also due to their effects on radiation. The accurate estimation of the complex interactions within the earth-atmosphere-system is still a major challenge. Particularly if related to the issues of the continuously discussed climate change. Because of the high temporal and spatial variability in conjunction with clouds, gaining additional information is absolutely necessary. Satellite data presents itself as the first choice. Especially geostationary satellites embodying high temporal resolution and comparable spectral information, can meet the demands of effective cloud detection. Methods of continuous cloud detection are primarily based on threshold techniques and typically require additional data on the atmospheric state. Regarding these methods, determining an appropriate threshold accurately, is still a problem to be solved. Other approaches of image analysis, like Neural Networks, Cluster-Analysis or Support Vector Machines (SVM), have indeed been carried out experimentally but mostly not in terms of an automated application for high temporal resolution data series.

Therefore this study's aim is to develop a cloud detection and cloud classification, applied by using Support Vector Machines on METEOSAT-SEVIRI-data covering the area of Germany. The implementation should only rely on genuine image information (cloud-related METEOSAT-channels), supplemented by the required training data (ground truth). In this context, Support Vector Machines represent a comparatively modern classifier, which is able to solve classification problems effectively, even with only a small amount of training data. The choice of Germany as the main investigation area has been based on the multiple cloud variations and conditions found here. Especially the relatively northern location and the influence of the North Atlantic are challenging aspects regarding effective cloud detection.

The main difficulty using Support Vector Machines is the precise selection of representative training data, by which the classifier learns. A first idea for this purpose has been the application of so-called webcams or weather cameras taking pictures of the sky in many places. Retrospectively, the compilation of a meaningful data set turned out to be difficult. In many places, images will not be stored, thus in the end, only six locations for the year 2008 were available. Although this is not enough for a proper classification, the data is used in the processing scheme. To get additional training data, the data set has been expanded through the use of known threshold algorithms. But only those pixels, having a high probability to belong to a certain class, were

selected for training. After creating a cloud mask, the clouds were subdivided into four classes, distinguishing between height and radiation transparency.

Besides comparing the results to the cloud products of the Satellite Application Facility on support to Nowcasting and Very Short Range Forecasting (SAFNWC), an accuracy assessment has been carried out. This validation procedure has been applied with the help of synoptic observations. However, this has only been done with the cloud mask, using the so-called octa-scale. Based on the validation results and the additional observation of the cloud cover for the year 2008, it was possible to show that accurate results can be achieved with an automated SVM-cloud-detection, respectively SVM-cloud-classification. This refers mainly to situations where you have a lot of clouds, or where clouds are equally distributed, compared to cloud free areas. In contrast, high-pressure weather conditions with extensive cloud free areas lead to increased overestimation of clouds by the SVM-classifier. But on the other side, the results of the SAFNWC-Cloudmask used without additional information from numerical weather prediction (NWP) reveal an underestimation of clouds. This underestimation cannot be reduced to certain weather conditions.

In addition, a temporal and spatial differentiation of cloud cover could be shown for the year 2008. In this context, it was possible to link the results with the general weather patterns of 2008. For example, temporal variations have been presented by the connection between the cloud amount and the results of an objective weather type classification established by the German Meteorological Service (DWD). Regional differences were mainly in conjunction to the low mountain ranges, which are generally connected to high cloud coverage. On the other hand, parts of northern and southern Germany and the Rhine Valley are often combined with less coverage.

With this work it was possible to demonstrate the potential of Support Vector Machines as an automated application of cloud detection on temporal high resolution data. Although it is a very computationally intensive procedure, the entire processing, including training of the classifier for each METEOSAT-timeslot, can be realized in an appropriate timeframe. After further optimization, an operational application would be quite conceivable.

1 Einleitung

1.1 Wolken als Untersuchungsgegenstand

Bei der Betrachtung des allgemeinen Wettergeschehens sind Wolken eines der eindrucksvollsten Phänomene. Es wird schnell klar, dass sie im Klimasystem der Erde einen zentralen Faktor darstellen. Dies betrifft nicht nur den Zusammenhang zur Niederschlagsverteilung und zum globalen Wasserkreislauf, sondern auch ihren Einfluss auf Strahlungsvorgänge in der Atmosphäre. Aufgrund der hohen zeitlichen und räumlichen Variabilität sind Wolken jedoch nur schwierig zu erfassen, und eine genaue Abschätzung der komplexen Wechselwirkungen innerhalb des Klimasystems ist immer noch mit Problemen verbunden. Die detaillierte Auseinandersetzung mit dem Thema Wolken ist daher sowohl in der Meteorologie, als auch in der Klimatologie bzw. Klimageographie von besonderem Interesse (Simmer 2000, Häckel 2005, Bridgman und Oliver 2006, Heyer et al. 2005, Schönwiese 2008, Kraus 2009).

Vor allem Fragestellungen, die den gegenwärtig vielfach diskutierten Klimawandel betreffen, stehen hier im Vordergrund. Welche Veränderungen sind bezogen auf Wolkenbildung und Wolkenverteilung durch eine globale Erwärmung zu erwarten? Inwieweit werden sich beispielsweise Veränderungen der Bewölkungsverhältnisse auf globaler und regionaler Ebene einstellen, und welche Auswirkungen auf die Niederschlagsverhältnisse sind damit verbunden? Dies betrifft in erster Linie die Tatsache, dass durch allgemein höhere Temperaturen mehr Feuchtigkeit in die Atmosphäre aufgenommen werden kann, mit möglichem Einfluss auf globale, als auch auf regionale Bewölkungsgegebenheiten. Aber auch langfristige Veränderungen der zirkulationsdynamischen Variabilität können zu regionalen Änderungen der Wolkenbedeckung führen, mit erheblichen Auswirkungen auf die Niederschlagsverteilung und auf die damit verbundene Wasserverfügbarkeit (Yao und Del Genio 1999, IPCC 2001, IPCC 2007).

Gleichzeitig ist offen, wie Wolken das Klima rückwirkend beeinflussen. Vor allem die Rückkopplungseffekte zwischen Wolkenausprägung, Strahlung und Zirkulationsdynamik sind nur schwer abzuschätzen. Zwar sind Wolken eher als Begleiterscheinung der atmosphärischen Zirkulation anzusehen, dennoch haben sie einen wesentlichen Einfluss auf Strahlungsvorgänge, die die atmosphärische Zirkulation antreiben. Einerseits sind Wolken ein wesentlicher Bestandteil des natürlichen Treibhauseffektes, andererseits halten sie solare Strahlung ab. Welche Auswirkungen eine Zu- oder Abnahme der globalen Bewölkung in dieser Hinsicht hätte, ist noch

weitestgehend unklar. Klimamodelle sollen hier zu mehr Erkenntnissen führen, jedoch ist die Parametrisierung von Wolken in globalen Modellen (Gitter 100 km x 100 km) mit Schwierigkeiten verbunden. Vor allem die hohe räumliche und zeitliche Variabilität von Wolken auf kleineren Skalen ist problematisch (Kärner und Keevallik 1993, Simmer 2000).

An den beschriebenen Problemen und möglichen Veränderungen wird deutlich, dass jeglicher Informationsgewinn über den Zustand von Wolken bzw. über die Bewölkungsgegebenheiten eine enorme Bedeutung hat, insbesondere um das Wirkungsgefüge innerhalb des Klimasystems noch besser verstehen zu können. Da sich Wolken aus Wassertropfen oder Eiskristallen unterschiedlicher Größe zusammensetzen und ein sehr dynamisches Verhalten vorweisen, sind direkte Messungen jedoch äußerst schwer umzusetzen. Zur Untersuchung von Wolken bezieht man sich daher vorrangig auf *Fernerkundungsverfahren*¹. Hierbei handelt es sich sowohl um bodengestützte Verfahren (vgl. z.B. Feister et al. 2000, Seiz et al. 2007), als auch um Verfahren der *Satellitenfernerkundung*. Mit Hilfe von Satellitendaten können Wolken quasi synoptisch erfasst werden. In Bezug auf die Veranschaulichung und Auswertung des täglichen Wettergeschehens geschieht dies schon seit mehreren Jahrzehnten. Beispielsweise wird anhand von Wetter-Satellitenbildern (vgl. Abbildung 1.1) deutlich, wie sich Wolken innerhalb der atmosphärischen Zirkulation verteilen und darstellen. Wolkenstrukturen, wie sie im Bereich eines Tiefdruckgebietes vorkommen, lassen sich deutlich erfassen, und können in Bezug auf die Wettervorhersage analysiert und interpretiert werden. Daher liegt es nahe, auch für langfristige Untersuchungen die benötigten Informationen aus Satellitendaten abzuleiten (Carleton 1991, S. 76 - 111, Kidder und Vonder Haar 1995, König und Bissolli 2001, Mannstein et al. 2004, Meerkötter et al. 2004).

1.2 Wolken in der Satellitenfernerkundung

Innerhalb der Fernerkundung von Landoberflächen ist einer der wesentlichen Vorverarbeitungsschritte, störende Effekte der Atmosphäre, wie z.B. durch Wolken, zu reduzieren und wolkenlose Aufnahmen bzw. Teilaufnahmen zu erhalten (siehe Gomez-Chova et al. 2010). Andererseits spielen Wolken, sowie deren Verteilung, für klimageographische und meteorologische Fragestellungen eine entscheidende Rolle. Das heißt, ein sicheres Analyse-Schema von Wolken, mit Hilfe von Fernerkundungsdaten, kann für vielfältige wissenschaftliche Zwecke von Nutzen sein, wie in diesem Fall die Untersuchung der Wolkengegebenheiten über Deutschland (Carleton 1991, Kidder und Vonder Haar 1995, Bendix 2002).

Um sich auf größeren Skalen mit dem Thema Wolken zu beschäftigen, sind Satellitendaten daher das Mittel der ersten Wahl. Mit Hilfe dieser Daten können die Wolkenverhältnisse großflächig analysiert werden. Eine Vielzahl von Satelliten und Sensoren stehen hierfür zur Verfügung.

¹Messung über eine gewisse Distanz, ohne Kontakt zum zu erfassenden Körper.

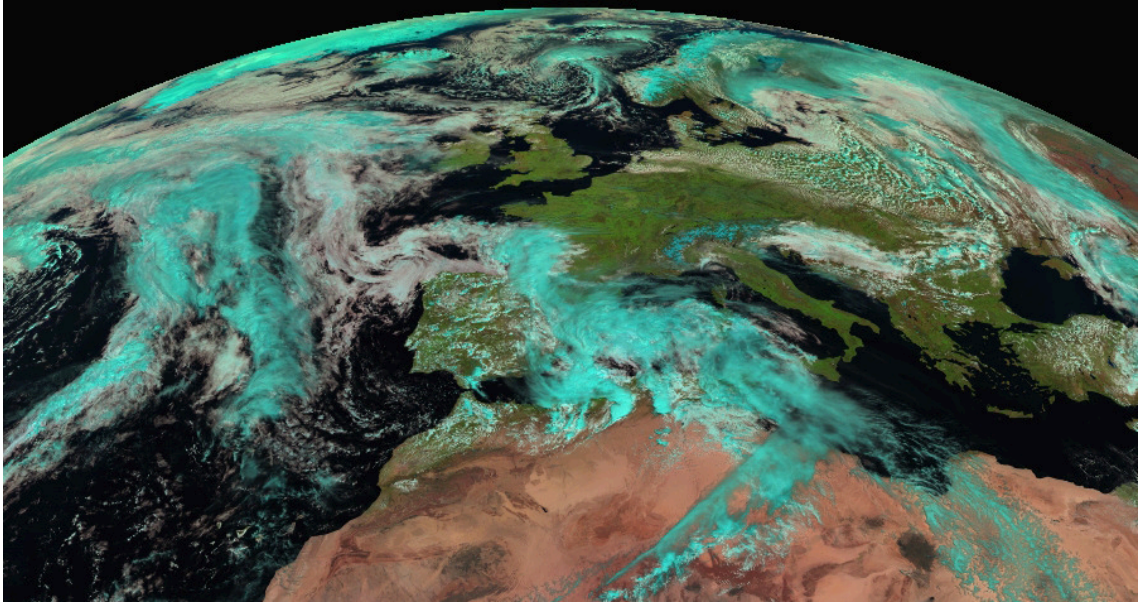


Abb. 1.1: Satellitenaufnahme von Europa und Nordafrika mit METEOSAT-9 (MSG-2), RGB-Falschfarben. *Quelle:* <http://www.nwcsaf.org> [12.03.2009]

Da es in der vorliegenden Arbeit vorrangig um die Wolkendetektion, basierend auf den Daten des METEOSAT-SEVIRI-Sensors geht (vgl. Kapitel 4), soll im Weiteren auch nur auf Verfahren der passiven Satellitenfernerkundung eingegangen werden. Weiterführende Informationen über andere Verfahren, wie z.B. die der Radarfernerkundung sind in der einschlägigen Literatur zu finden (Kidder und Vonder Haar 1995).

Bei passiven Verfahren wird das nach oben gerichtete Strahlungsfeld mit Hilfe der Satellitensensoren (Imager) erfasst, die dann entsprechende Daten in verschiedenen Wellenlängenbereichen bzw. Kanälen zur Verfügung stellen. Die Betrachtung erfolgt von oben, und im Gegensatz zur Radartechnologie wird die vom System Erde-Atmosphäre emittierte und reflektierte elektromagnetische Strahlung aufgenommen. Es handelt sich in der Regel um Wellenlängen in dem für das menschliche Auge sichtbaren Bereich (ca. $0,4 - 0,7 \mu\text{m}$) bis zu Wellenlängen im Infrarot- bzw. Thermalbereich (z.B. $10 \mu\text{m}$). Im Endeffekt erhält man Bilddaten (Rasterdaten bestehend aus Bildpunkten bzw. Pixeln^2) die, abhängig von dem durch den Satelliten aufgenommenen Kanal, entweder Reflektanzen oder Strahlungstemperaturwerte repräsentieren. Eine Vielzahl solcher multispektralen Sensoren stehen für wissenschaftliche Zwecke zur Verfügung (z.B. NOAA-AVHRR, MODIS oder METEOSAT-SEVIRI) (Carleton 1991, Kidder und Vonder Haar 1995, S. 87 ff.).

Weiterhin unterscheidet man zwischen den Umlaufbahnen der Satelliten, auf denen sich die Sensoren befinden. Beispielsweise dreht sich die Erde unterhalb polarumlaufender Satelliten

²Picture Element

hinweg, wobei nur einzelne Streifen aufgenommen werden. Die Abtastrate der gesamten Erde liegt bei polarumlaufenden Wetter-Satelliten bei ca. 12 Stunden (14 Umläufe pro Tag). Im Gegensatz zu der geringen Bildwiederholfrequenz (zeitliche Auflösung) ist der Vorteil dieser Systeme eine vergleichsweise hohe räumliche Auflösung, die bei Wetter-Satelliten meist 1 km beträgt. Polarumlaufende Satellitensysteme sind z.B. die NOAA-Serie, EOS-Terra und -Aqua oder MetOp. Im Bereich der Wetter-Satelliten sind jedoch die geostationären Satelliten von besonderer Bedeutung. Sie bewegen sich in einer Umlaufbahn mit dem bewegenden Zentralkörper, also der Erde, und können somit immer denselben Ausschnitt erfassen. Mit diesem Prinzip lassen sich hohe Aufnahmefrequenzen realisieren. Der Ausschnitt Europa/Afrika wird in diesem Zusammenhang durch die METEOSAT-Satelliten abgedeckt (vgl. Kapitel 4), welche seit der zweiten Generation (METEOSAT Second Generation, MSG) ein Abtastintervall von 15 Minuten zur Verfügung stellen. Auch bei vergleichsweise geringer räumlicher Auflösung (vgl. Abbildung 4.1) bietet dieser Umstand bei der Wolkenanalyse einen enormen Vorteil. Dies betrifft vor allem die hohe zeitliche und räumliche Variabilität von Wolken (Kidder und Vonder Haar 1995, S. 87 ff., EUMETSAT 2000, EUMETSAT 2001).

Um nun anhand der zur Verfügung stehenden Satellitendaten Wolkengegebenheiten oder Wolkeneigenschaften abschätzen zu können, gibt es eine Reihe von möglichen Ansätzen der Digitalen Bildverarbeitung. Bezogen auf spektrale Bilddaten steht an erster Stelle die Erkennung solcher Bildpunkte/Pixel, die mit Wolken behaftet sind (cloud-contaminated oder wolkenkontaminiert) und sich von wolkenfreien (cloud free) Pixeln abgrenzen lassen. Anhand der durch den Sensor zur Verfügung stehenden spektralen Information können aber auch, mittels verschiedener Algorithmen, weitere Informationen gewonnen werden. Beispielsweise erlauben unterschiedliche Kombinationen von Infrarotkanälen die Erfassung verschiedener Wolkenarten oder die Höhe der Wolkenoberkante. Auch der Flüssigwasseranteil von Wolken kann durch Informationen der Spektralkanäle abgeleitet werden (vgl. Kapitel 3.2 und Watts et al. 1998). Durch den Einfluss von Wolken auf den Strahlungshaushalt oder die Niederschlagsverteilung ist eine *Klassifizierung* der Bilddaten in unterschiedliche Wolkenarten sicherlich von besonderer Bedeutung. Sie kann nachfolgend als Grundlage für weitere wissenschaftliche Fragestellungen herangezogen werden. Verschiedene Klassifizierungsverfahren werden zu diesem Zweck bereits eingesetzt (vgl. Carleton 1991, S. 97 ff.). Neben Neuronalen Netzen, Cluster-Analysen oder anderen statistischen Bildverarbeitungserfahren (vgl. z.B. Azimi-Sadjadi et al. 1996, Ameur et al. 2004, Ghosh et al. 2006, Amato et al. 2008), werden hierzu vor allem *Schwellenwert- bzw. Grenzwertalgorithmen* verwendet (vgl. hierzu z.B. Saunders und Kriebel 1988, Kriebel 1990, Thiermann und Ruprecht 1992, Kriebel et al. 2003). Schon seit einigen Jahrzehnten wird mit dieser Methode versucht, wolkenkontaminierte von wolkenfreien Flächen abzugrenzen, wobei das grundlegende Prinzip relativ einfach ist:

- *Wolkenkontaminiert* - Der gemessene Wert an einem Pixel ist je nach Kanal „heller“ bzw. „kälter“ als ein bestimmter Wert
- *Wolkenfrei* - Der gemessene Wert an einem Pixel ist je nach Kanal „dunkler“ bzw. „wärmer“ als ein bestimmter Wert

In diesem Zusammenhang ist vor allem das Analyse-Tool APOLLO (AVHRR Processing scheme Over cLOUDs, Land and Ocean) zu nennen (siehe Kapitel 3.1), welches sich aus einer Kombination von Schwellenwertalgorithmen zusammensetzt. Basierend auf den Arbeiten von Saunders (1986), Saunders und Kriebel (1988) und Kriebel et al. (2003) wurde es am Deutschen Zentrum für Luft- und Raumfahrt (DLR) entwickelt und in erster Linie auf Daten des NOAA-AVHRR-Sensors (Advanced Very High Resolution Radiometer) angewendet. Die Schwierigkeit solcher Verfahren liegt jedoch in der exakten Bestimmung des jeweiligen Grenzwertes (vgl. Kidder und Vonder Haar 1995, S. 98 ff.). Wird dieser falsch abgeschätzt, kommt es zu Ungenauigkeiten. Für eine exakte Bestimmung werden daher Zusatzdaten benötigt, mit deren Hilfe der jeweilige Zustand der Atmosphäre berücksichtigt wird (vgl. Kapitel 3).

Klassifizierungsansätze die, wie *Support Vector Machines*, auf dem Prinzip des maschinellen Lernens basieren, beziehen sich zwar auf die reine Bildinformation, benötigen jedoch Trainingsdaten (z.B. Ground Truth oder Expertenauswertung, vgl. Kapitel 5.1) aus denen der Klassifikator lernt. Es muss also abgewogen werden welches Verfahren mit welchem Aufwand, welche Ergebnisse liefert. Bei Support Vector Machines handelt es sich um einen relativ modernen Klassifikator, der auch mit geringen Mengen an Trainingsdaten zurechtkommt, und der bei ersten Klassifizierungsversuchen von Wolken schon vielversprechende Ergebnisse geliefert hat (Han et al. 2006, Zibert et al. 2006, Gomez-Chova et al. 2010).

1.3 Untersuchungsgebiet Deutschland

Die Detektion bzw. Klassifizierung von Wolken mit Hilfe von Satellitendaten ist, vor allem in den mittleren bis hohen Breiten, eine Herausforderung (vgl. Dybbroe et al. 2000, Dybbroe 2001). Besonders in Europa, durch den Einfluss des Nordatlantiks und im Bereich der damit verbundenen Wetterfronten kommt es zu unterschiedlichsten Ausprägungen von Wolken. Wolken treten in verschiedenen Höhen auf, sie überlagern sich oder erscheinen nur als einzelne Fragmente. Eine sichere Wolkenerkennung sollte demnach besonders im Bereich Zentraleuropa bzw. Deutschland von Nutzen sein, und eine eventuelle Übertragung auf andere Regionen erleichtern (Wanner et al. 2001, Häckel 2005, S. 183 ff., Müller-Westermeier 2006).

Neben der rein methodischen Auseinandersetzung, sind aber auch klimatologische bzw. meteorologische Gesichtspunkte der Bewölkung im Raum Deutschland von Interesse. Obwohl

Klimastudien häufiger Parameter wie Temperatur oder Niederschlag behandeln, sollten Wolken keinesfalls außer Acht gelassen werden. Hierbei geht es vor allem um die Verteilung von Wolken innerhalb Deutschlands oder den engen Zusammenhang zum Niederschlag. Auch die Betrachtung auf verschiedenen Raum- und Zeitskalen ist von Bedeutung. Gibt es beispielsweise eine Beziehung zwischen der regionalen Verteilung und gleichzeitiger Änderung großräumiger Strömungsmuster? Inwieweit hat z.B. die Nordatlantische Oszillation (NAO, vgl. Kapitel 2.3.3) Einfluss auf regionale Bewölkungsmuster? Gerade in Bezug auf die regionale Differenzierung der Bewölkung stellt Deutschland ein äußerst interessantes Untersuchungsgebiet dar (Beck 2000, Simmer 2000, Wanner et al. 2001).

Küstenbereiche im Norden und die westliche Nähe zum Atlantik sorgen für maritim geprägten Einfluss, während weiter östlich und im Landesinneren mit kontinentalen Bedingungen zu rechnen ist. Hinzu kommt die nördliche Lage zu den Alpen, durch die sich eine natürliche Abgrenzung zum Mittelmeerraum ergibt. Aber auch die Orographie im Landesinneren spielt eine Rolle. Je nach Anströmung kann es an den Mittelgebirgen zu Staulagen oder Föhnwinden kommen (vgl. Kapitel 2.3.2 und Müller-Westermeier 2006, Schönwiese 2008). In Verbindung mit diesen Gegebenheiten sorgen die Druck- und Strömungsmuster im Großraum Europa/Nordatlantik für charakteristische Bewölkungsmuster. Tiefdruckwirbel, die über dem Atlantik entstehen, bewegen sich in verschiedenen Bahnen über Europa hinweg und sorgen meist für eine, mit viel Feuchtigkeit angereicherte, westliche Grundströmung. Auf der anderen Seite kann ein ausgeprägtes Hoch über Mitteleuropa zu einer wolkenlosen Zeit führen (Kurz 1990).

Allgemein sollte für Deutschland ein relativ hoher *Wolkenbedeckungsgrad* zu erwarten sein, jedoch mit regionalen Unterschieden, die es aufzudecken gilt. Natürlich macht auch die enge Verbindung zwischen Wolken und Niederschlag die Ableitung der Bewölkungsverhältnisse zu einem interessanten Untersuchungsgegenstand. Hier stellt sich die Frage, ob Gebiete mit allgemein höheren Niederschlägen auch höchste Werte des Wolkenbedeckungsgrades aufweisen. Andererseits können trockenere Bereiche ebenfalls mit einem gewissen Grad an Bedeckung gekoppelt sein. Mit Hilfe eines effektiven Wolkenerkennungs-Schemas, angewendet auf Satellitendaten, sollte demnach ein besseres Verständnis der komplexen Wechselwirkungen im Klimasystem möglich sein. Dies betrifft im Endeffekt nicht nur den begrenzten Raum Deutschland, sondern auch die Verknüpfung mit Prozessen auf größeren Skalen (Simmer 2000).

1.4 Ziel der Arbeit und Fragestellungen

Da zur Abschätzung der Bewölkung auf größerer räumlicher Ebene direkte Messungen kaum möglich sind, bleibt als einzige Alternative nur die Nutzung von Fernerkundungsmethoden bzw. Satellitendaten. Eine großflächige Erfassung von Wolken wird hierdurch ermöglicht. Schon seit mehr als 20 Jahren werden in dieser Hinsicht Arbeiten durchgeführt und das Bestreben liegt

darin für bestimmte Gebiete eine operationelle Wolkendetektion durchzuführen (vgl. hierzu z.B. Saunders und Kriebel 1988, Kriebel et al. 2003, Derrien und Le Gléau 2005a). Für eine sinnvolle Untersuchung der Bewölkungsverhältnisse mit Satellitendaten müssen jedoch einige Voraussetzungen erfüllt sein. Neben der spektralen Information erfordert die hohe Variabilität von Wolken ein gewisses Maß an räumlicher und zeitlicher Auflösung. In diesem Zusammenhang haben sich vor allem die polarumlaufenden NOAA-Satelliten mit ihrem Sensor AVHRR hervorgetan (vgl. z.B. Berger 1995). Mit bis zu 5 Kanälen, in den für Wolkenuntersuchungen wichtigen Wellenlängenbereichen (vgl. Tabelle 3.1), können mit diesem Messinstrument sehr gute Ergebnisse erzielt werden. Der große Nachteil ist jedoch die relativ geringe zeitliche Auflösung von nur einer globalen Abdeckung pro Tag (vgl. Saunders und Kriebel 1988). Hier liegt der Vorteil klar bei den geostationären Satelliten. Ältere Modelle, wie die erste Generation von METEOSAT oder frühe Generationen vom Geostationary Operational Environmental Satellite (GOES), waren zwar in ihrem Angebot an Spektralkanälen begrenzt (teilweise nur 2 bis 3 Kanäle, vgl. EUMETSAT 2000), die Grundprinzipien der Wolkenerkennung konnten aber dennoch durchgeführt werden. Beispielsweise wurde dies innerhalb des International Satellite Cloud Climatology Projects (ISCCP) umgesetzt (vgl. Kidder und Vonder Haar 1995, S. 282 - 284, Bissolli und Gelhardt 2002, Bissolli 2004). Bei geostationären Satelliten neuerer Generation ergibt sich diese Problematik nicht mehr. METEOSAT Second Generation (MSG) bietet eine Erweiterung auf insgesamt 12 Kanäle (vgl. EUMETSAT 2001). Das heißt, alle in Bezug auf Wolken relevante Wellenlängen können mit Aufnahmeintervallen von 15 Minuten erfasst werden.

Für eine Wolkendetektion und eine weiter differenzierte Klassifizierung verschiedener Wolkentypen bzw. Wolkenarten scheint der METEOSAT-Sensor demnach ideal (vgl. hierzu auch Roebeling et al. 2006). Dies sollte auch für das ausgewählte Untersuchungsgebiet Deutschland gelten. Obwohl mit zunehmender geographischer Breite die räumliche Auflösung geringer wird (vgl. Abbildung 4.1), können Untersuchungen auf dieser Ebene bei der Einordnung kleinskaliger Phänomene beitragen. Im Rahmen der Satellite Application Facility on support to Nowcasting and Very Short Range Forecasting (SAFNWC) wird dies, ähnlich wie bei APOLLO, bereits mit verschiedenen Schwellenwertalgorithmen durchgeführt (vgl. Kapitel 3.2). Versuche mit anderen Klassifikatoren sind dagegen nur vereinzelt zu finden. Insbesondere die Anwendung von Support Vector Machines im Bereich der Wolkendetektion bezieht sich meist auf Einzelversuche im Bereich polarumlaufender Satelliten (vgl. z.B. Lee et al. 2004, Han et al. 2006). Eine Anwendung dieses Klassifikators auf METEOSAT-Datenreihen wäre also durchaus reizvoll und kann als Alternative bzw. Ergänzung zu den operationell betriebenen Schwellenwertalgorithmen gesehen werden.

Das Ziel dieser Arbeit ist demnach die Entwicklung eines effektiven Wolkendetektions- bzw. Wolkenklassifizierungsverfahrens mit Hilfe von Support Vector Machines, angewendet auf Da-

ten des METEOSAT-Sensors. Seit 2006 besteht am Zentrum für Fernerkundung der Landoberfläche (ZFL) in Zusammenarbeit mit der Arbeitsgruppe Fernerkundung (Remote Sensing Research Group, RSRG) des Geographischen Instituts der Universität Bonn die Möglichkeit, METEOSAT-Daten über die eigene Empfangsanlage zu empfangen und zu archivieren (vgl. Kapitel 4.4). Die Archivierung des Ausschnitts Deutschland bzw. Zentraleuropa erfolgt seit Mitte 2007 und ein vollständiger Datensatz liegt für das Jahr 2008 vor. Obwohl eine fortlaufende Archivierung erfolgt, bezieht sich die vorliegende Arbeit, auch auf Grund zeitintensiver Testläufe, nur auf Daten des Jahres 2008. Rückwirkend soll aber dennoch geprüft werden, ob die Klassifizierung in einem operationellen Rahmen möglich wäre. Zwar sind alle Berechnungen und Programme auf den Zeitraum 2008 ausgelegt, die Rechenzeit der Einzelaufnahmen soll aber im Endeffekt Werte von ca. 10 Minuten nicht überschreiten. Wie die gesamte Prozessierung bzw. Klassifizierung innerhalb eines solchen Zeitrahmens umsetzbar ist, soll anhand dieser Arbeit dargelegt werden. Gleiches gilt für die Auswahl oder Gewinnung notwendiger Trainingsdaten. Solche Daten müssen vorhanden sein, damit bestimmte Bildelemente für das maschinelle Lernen eindeutig einer Klasse zugeordnet werden können (vgl. Canty 2006, Abe 2010). Unter anderem sollen in diesem Zusammenhang Bilder sogenannter Wetter-Kameras (vgl. Kapitel 5.3) als Grundlage verwendet werden. Ob und inwieweit solche Daten für das Klassifizierungsverfahren sinnvoll sind, soll ebenfalls Bestandteil vorliegender Arbeit sein. Dies bezieht sich insbesondere auf die Problematik zweier gegensätzlicher Betrachtungsgeometrien (Blickrichtung von oben und von unten bzw. seitlicher Blickwinkel). Um den Klassifikator in seiner Genauigkeit zu testen, soll eine Validierung anhand unabhängiger synoptischer Beobachtungsdaten erfolgen. Abschließend ist ein Vergleich mit gängigen Wolkendetektionsverfahren, wie denen der SAFNWC, angestrebt.

Obwohl sich der Schwerpunkt der Arbeit auf die methodische Entwicklung und Umsetzung des Detektionsverfahrens bezieht, sollen klimageographische Gesichtspunkte, besonders in Bezug auf die Ergebnisse, ebenfalls Berücksichtigung finden. Zwar kann man bei der Betrachtung der Daten des Jahres 2008 nicht von einem klimatologischen Zeitraum sprechen, dennoch sollten sich die Ergebnisse in einen größeren zeitlichen Rahmen einbetten lassen. Spiegeln sich beispielsweise allgemeine Tendenzen einer Klimaveränderung in dem betrachteten Zeitraum wider? Auch in Anlehnung an eine regionale Differenzierung. Zusammenfassend lassen sich folgende Fragestellungen formulieren, die in den nachfolgenden Kapiteln abgearbeitet werden.

- Sind Support Vector Machines für eine effektive Wolkendetektion bzw. Wolkenklassifizierung geeignet?
- Kann das Verfahren auf große Datenreihen angewendet werden und welche Herangehensweise ist in diesem Zusammenhang die günstigste?

- Ist die Anwendung auf Daten des METEOSAT-Sensors in den mittleren und hohen Breiten sinnvoll?
- Wie lassen sich notwendige Trainingsdaten gewinnen und in welchem Umfang sind sie erforderlich?
- Lässt sich der gesamte Prozess in einem adäquaten zeitlichen Rahmen realisieren?
- Sind die Ergebnisse vergleichbar mit gängigen Wolkenerkennungsverfahren und welcher Grad der Genauigkeit ist zu erreichen?
- Wie können die Ergebnisse in das Wettergeschehen des Jahres 2008 eingeordnet werden?
- Ergibt sich eine regionale Differenzierung für den Raum Deutschland, eingebettet in großräumige Strukturen und Strömungsmuster?

Der Aufbau der Arbeit setzt sich aus drei Hauptteilen zusammen. Theoretische Hintergrundinformationen zum Thema Wolken, aber auch zu den gängigen Verfahren der Wolkendetektion mit Hilfe von Satellitendaten, werden im ersten Teil der Arbeit ausführlich beschrieben (Kapitel 2 und 3). Anschließend werden im zweiten Teil die Datengrundlage sowie das methodische Verfahren der Wolkendetektion und -klassifizierung selbst dargelegt (Kapitel 4 und 5). Alle notwendigen Schritte werden einzeln behandelt und ausgeführt. Abschließend erfolgt die Beschreibung des Validierungsverfahrens gefolgt von einer Einordnung der Ergebnisse (Kapitel 6, 7 und 8). Am Ende der Arbeit stehen allgemeine Diskussion und Ausblick, in denen die zuvor aufgeführten Fragestellungen abschließend beurteilt werden (Kapitel 9).

2 Wolken als atmosphärisches Phänomen

2.1 Wolkenbildung

Bei Wolken handelt es sich im Wesentlichen um kondensierten Wasserdampf in Form von Tropfen mit einer Größe von 5 bis 20 μm . Neben Wassertropfen können aber auch vielgestaltige Eiskristalle mit einer Ausdehnung von bis zu einem Millimeter zur Wolkenentstehung beitragen. Wolken bedecken im globalen Mittel ca. 70 % der Erdoberfläche und sie entstehen innerhalb der Atmosphäre, wenn wasserdampfhaltige Luft bis zum Taupunkt abgekühlt wird. Gleichzeitig müssen Kondensationskerne in ausreichender Menge vorhanden sein, an denen sich die Wasserdampfmoleküle bei der Kondensation anlagern können. Die Abkühlung erfolgt neben thermischer Ausstrahlung in der Regel durch *Vertikalbewegungen* (Simmer 2000, S. 8, Häckel 2005, S. 110 ff.).

Den Vorgang der Wolkenbildung bei einem aufsteigenden Luftpaket, kann man anhand der trocken- und feuchtadiabatischen Abkühlung erklären. Adiabatisch bedeutet, dass kein Energieaustausch zwischen dem Luftpaket und seiner Umgebung stattfindet. Dies ist zwar in der Atmosphäre nicht unbedingt der Fall, sollte jedoch für das Verständnis der wolkenbildenden Vorgänge bei Vertikalbewegungen von Vorteil sein. Wenn sich ein in Bodennähe befindliches Luftpaket stärker erwärmt als seine Umgebung, kommt es zu einer Volumen- bzw. Dichteänderung. Durch die Ausdehnung wird das nun leichtere Luftpaket zum Aufstieg gezwungen (Auftrieb). Mit zunehmender Höhe schreitet die Ausdehnung fort, und durch die geleistete Ausdehnungsarbeit (auf Kosten der inneren Energie) kommt es zu einer Abkühlung. Man spricht auch von trockenadiabatischer Abkühlung (Häckel 2005, S. 93, Heyer et al. 2005, S. 74 ff., Kraus 2009, S. 182).

Durch die Temperaturabnahme mit zunehmender Höhe, steigt die relative Feuchte auf 100 % und es kommt zur Sättigung (Dampfdruck = Sättigungsdampfdruck). In der sich daraus ergebenden Höhe, dem Kondensationsniveau oder Lifting Condensation Level (LCL), beginnen sich kleinste Wolkentropfen zu bilden. Die Entstehung der Wolke. Oberhalb des Kondensationsniveaus folgt, aufgrund der freiwerdenden latenten Wärme, eine langsamere Abkühlung. Ab dieser Stelle spricht man von feuchtadiabatischer Abkühlung (vgl. Häckel 2005, S. 90 ff.). Bei Temperaturen zwischen 0 und $-15\text{ }^{\circ}\text{C}$ bleiben die Wolkentropfen in der Regel noch im flüssigen Zustand. Es handelt sich um den unterkühlten Zustand der Wasserwolke, verursacht durch die

gegebenen Druckverhältnisse. Ab etwa -12 °C gefrieren die ersten Wolkentropfen und es entstehen Mischwolken. Erst bei Temperaturen von ca. -35 °C bestehen die Wolken überwiegend aus Eiskristallen (Simmer 2000, S. 8).

Inwieweit ein Luftpaket aufsteigen kann, hängt im Wesentlichen vom Gleichgewicht zur umgebenden Luft ab. Man spricht von Stabilität bzw. Labilität. Nimmt die Umgebungstemperatur mit der Höhe langsamer ab, als die Temperatur des Luftpaketes, so handelt es sich um eine stabile Schichtung. Ein forciertes Aufsteigen ist hier nicht möglich, womit einer Wolkenbildung entgegengewirkt wird. Gleiches gilt bei einer Inversion, die durch Temperaturzunahme mit der Höhe gekennzeichnet ist. Umgekehrt spricht man von Labilität, wenn die Temperatur der Umgebung schneller abnimmt als die des Luftpaketes. Solche Temperaturverhältnisse führen zu einer starken Vertikalbewegung, die je nach Wasserdampfgehalt des Luftpaketes mit starker Wolkenbildung verbunden sein kann. Generell bilden sich Wolken in diesem Sinne nur bis zur Tropopause (Grenzschicht zwischen Troposphäre und Stratosphäre), da in der unteren Stratosphäre Inversion bzw. Isothermie vorherrscht. Die Höhe der Tropopause variiert jedoch. In den Tropen befindet sie sich in einer größeren Höhe als in den mittleren oder hohen Breiten (vgl. Kraus 2009, S. 163 - 166). Bei diesen modellhaften Erklärungsansätzen muss jedoch berücksichtigt werden, dass in der Atmosphäre Misch- und Wärmeübergangsvorgänge ablaufen, und die Prozesse der Wolkenbildung komplexer sind, als es hier dargestellt ist. Für Vertikalbewegungen von Luftmassen, die eine Hauptursache der Wolkenbildung darstellen, gibt es neben dem konvektivem Auftrieb noch weitere Ursachen (vgl. Kraus 2009, S. 184 ff.):

- *Erzwungener Aufstieg an Gebirgen bzw. Erhebungen* - Luftmassen werden innerhalb ihrer Strömung an einem Gebirge zum Aufstieg gezwungen (vgl. Kapitel 2.3.2)
- *Aufgleiten im Bereich von Fronten* - Aufgleiten von wärmerer Luft auf keilförmig angeordnete kältere Luft (vgl. Abbildung 2.1)
- *Positiver Vertikalwind bei Konvergenz des horizontalen Windfeldes* - Luftaufstieg durch Konvergenz im horizontalen Windfeld (z.B. in Tiefdruckgebieten oder in der innertropischen Konvergenzzone). Divergenz führt zu absteigenden Vorgängen, mit Wolkenauflösung (z.B. im subtropischen Hochdruckgürtel)

2.2 Wolkeneinteilung nach Höhe und Gestalt

Da es in dieser Arbeit neben der reinen Wolkendetektion auch um eine weitere Differenzierung verschiedener Wolkenarten geht, ist eine Beschreibung der unterschiedlichen Wolkenformen in verschiedenen Höhen durchaus nötig. Dies geschieht auch, um die einzelnen Entstehungsmechanismen, die Wolkenverteilung innerhalb der atmosphärischen Zirkulation (vgl. Kapitel 2.3) und den Wolkeneinfluss auf Strahlungsvorgänge (vgl. Kapitel 2.4) besser verstehen zu können.

Tab. 2.1: Wolkenstockwerke nach der internationalen Klassifikation. *Quelle: Häckel 2005*

Stockwerk	Polargebiete	Gemäßigte Zone	Tropische Zone
Oberes (Hoch)	3 - 8 km	5 - 13 km	6 - 18 km
Mittleres (Mittel)	2 - 4 km	2 - 7 km	2 - 8 km
Unteres (Tief)	Erdboden bis 2 km	Erdboden bis 2 km	Erdboden bis 2 km






Das grundlegende Prinzip der Wolkenbildung lässt sich zwar in einfacher Weise an einem aufsteigenden Luftpaket darstellen, die Vielfalt der Wolken zeigt jedoch wie umfangreich die Einflüsse bei der Wolkenentstehung sein müssen. Zur Beschreibung dieser Mannigfaltigkeit bedarf es zunächst eines Ordnungssystems, einer Klassifikation der Wolken (nicht zu verwechseln mit der Klassifizierung der Satellitendaten). Man unterscheidet genetische (nach Entstehungsprozessen) von phänologischen (nach Aussehen und Erscheinungsform) Klassifikationen. Nachfolgend wird die phänologische Einteilung genauer beschrieben (Heyer et al. 2005, S. 94, Schönwiese 2008, S. 80 - 82, Kraus 2009, S. 214 - 218).

Bei der phänologischen Klassifikation handelt es sich um eine international vereinbarte Differenzierung der Wolken nach Höhe und Gestalt. Sie ist im *International Cloud Atlas* der World Meteorological Organization (WMO) ausführlich beschrieben (vgl. WMO 1975). Ein erstes Unterscheidungsmerkmal ist die Höhe über dem Erdboden (nicht über NN), wobei eine grobe Einteilung in drei Stockwerke (Tief, Mittel und Hoch) vorgenommen wird. Diese Höhenstufen überlappen sich teilweise und ihre Grenzen variieren mit der geographischen Breite (vgl. Tabelle 2.1). Durch den Bezug zum Erdboden, ist eine solche Einteilung im Hochgebirge mit Schwierigkeiten verbunden (Häckel 2005, S. 111 ff.).



Neben der Höhenunterteilung, erfolgt die Einteilung nach Form und Gestalt. Hierbei unterscheidet man 10 Wolkengattungen, die sich gegenseitig ausschließen, sehr verschieden aussehen und spezifische Zuordnungen zu den einzelnen Stockwerken besitzen. Entsprechend Gestalt und interner Struktur, wird innerhalb der Gattungen nach Arten differenziert. Weitere Unterarten beziehen sich auf die Anordnung der makroskopischen Elemente und den Grad der Durchsichtigkeit. Neben der Einteilung nach Gattungen, Arten und Unterarten können noch zusätzliche spezifische Erscheinungsformen angegeben werden. Da sich manche Wolken nicht unbedingt aus einem wolkenlosen Zustand entwickeln, sondern aus anderen Gattungen hervorgehen, weist die internationale Wolkenklassifikation sogenannte Mutterwolken aus. Beispielsweise kann eine Cumulus-Wolke zu einem Cumulonimbus aufwachsen. Zur Veranschaulichung sind in Tabelle 2.2 die wesentlichen Hauptgattungen aufgeführt (Häckel 2005, S. 113 - 119, Schönwiese 2008, S. 80 - 82).

Eine erste Abschätzung der komplexen Entstehungsprozesse einzelner Wolkenformen, wird durch das in Tabelle 2.2 dargestellte Einteilungsschema und die zuvor aufgeführten Grundla-

Tab. 2.2: Hauptwolkengattungen. Hohe und mittelhohe Wolken (a). Tiefe und vertikale Wolken (b). *Quelle: Kraus 2009***(a)** Hohe und mittelhohe Wolken.

Bezeichnung	Beschreibung	Foto
Cirrus	Ci Isolierte, meist seidig schimmernde, oft faserige Eiswolken verschiedenster Formen in großer Höhe. (Foto: Przemyslaw Idzkiewicz)	 A photograph showing wispy, feathery cirrus clouds against a clear blue sky.
Cirrocumulus	Cc Cumuluswolken im oberen Stockwerk. (Foto: Simon Eugster)	 A photograph showing a layer of small, rounded cumulus clouds in the upper atmosphere.
Cirrostratus	Cs Schichtwolke im oberen Stockwerk. (Foto: Simon Eugster)	 A photograph showing a thin, uniform layer of cirrostratus clouds covering the sky.
Altostratus	As Schichtwolke vor allem im mittleren Stockwerk, aber oft auch ins Obere hineinragend. (Foto: LivingShadow)	 A photograph showing a dense, greyish layer of altostratus clouds.
Altostratus	As Schichtwolke vor allem im mittleren Stockwerk, aber oft auch ins Obere hineinragend. (Foto: LivingShadow)	 A photograph showing a layer of altostratus clouds with the sun visible through a gap, creating a bright spot.

(b) Tiefe und vertikale Wolken.

Bezeichnung	Beschreibung	Foto
Stratocumulus Sc	Schichtwolke im unteren Niveau mit konvektiv bedingter, von Ort zu Ort unterschiedlicher Dicke, oder Felder von flachen Walzen und Ballen. <i>(Foto: LivingShadow)</i>	
Stratus St	Horizontal sehr homogene Schichtwolke im unteren Niveau, aus ihr kann Niesel und Schneegriesel fallen. <i>(Foto: LivingShadow)</i>	
Cumulus Cu	Haufenwolke mit beachtlicher Vertikalerstreckung, die ihre Basis im unteren Niveau besitzt, aber bis ins mittlere und obere Stockwerk ragen kann. <i>(Foto: Michael Jastremski)</i>	
Nimbostratus Ns	Schichtwolke aus der Regen oder Schnee fällt, vor allem im mittleren Niveau, aber auch ins obere oder untere Stockwerk ragend. <i>(Foto: LivingShadow)</i>	
Cumulonimbus Cb	Gewitterwolke. <i>(Foto: Simon Eugster)</i>	

gen der Wolkenbildung bereits möglich. Beispielsweise sind großräumige Vorgänge, wie die Hebungen flächenhaft verteilter Luftmassen verantwortlich für eine Schichtbewölkung (St, As, Ns, Cs), während vergleichsweise kleinskalige Schwankungen des Vertikalwindes (z.B. durch Konvektion) in der Bildung von Haufenwolken (Cu, Cb, Ac, Cc) resultieren. Einzelne Entstehungsprozesse bzw. verantwortliche Strukturelemente der atmosphärischen Turbulenz können sozusagen innerhalb eines Wolkenfeldes erkannt werden. Dennoch ist die Entstehung unterschiedlicher Wolkenformen durch solch eine Vielzahl von Einflüssen geprägt, dass eine vollständige und detaillierte Abschätzung der einzelnen Entstehungsmechanismen überaus schwierig ist. Schon eine einzelne Cumulus-Wolke besteht aus einer sehr komplexen internen Struktur, mit kleinen Wirbeln, die meist an den Rändern sichtbar werden, wo sich eine Mischung von Wolke und wolkenloser Luft vollzieht (Häckel 2005, S. 113 - 119, Kraus 2009, S. 214 - 218).

2.3 Wolken und atmosphärische Dynamik

2.3.1 Zyklonogenese als Ursache der Bewölkungsverhältnisse in Europa

Trotz der vielfältigen und komplexen Vorgänge im Zusammenhang mit Wolkenbildung, die von kleinräumiger Konvektion bis zu großflächigen Hebungsvorgängen reichen, lassen sich bestimmte Vorgänge eingrenzen, die im Wesentlichen für die Wolkenbildung bzw. Wolkenverteilung in Zentraleuropa verantwortlich sind. Vor allem die Druckverteilung im Bereich Nordatlantik/Europa und die damit einhergehenden Strömungsmuster, spielen für das Wettergeschehen und die Wolkenverteilung eine entscheidende Rolle. Im Bereich der planetaren Frontalzone³ kommt es in der westlichen Höhenströmung (Jet-Stream) zu Divergenz- und Konvergenzprozessen, die zu Auf- bzw. Abstieg von Luftmassen führen. Unter Einbeziehung von Drehimpuls und weiteren dynamischen Vorgängen, führt dies zur Bildung von Tief- (Zyklone) und Hochdruckgebieten (Antizyklone) (Häckel 2005, S. 280 ff.).

Für die Wolkenbildung über Deutschland ist vor allem die Zyklonogenese und die gleichzeitige Entstehung wetterwirksamer Fronten entscheidend (vgl. Abbildung 2.1). Die meist in der Nähe von Island entstehenden dynamischen Tiefdruckgebiete bewegen sich in Richtung Osten und gelangen so nach Europa, bevor sie sich in kontinental gelegeneren Bereichen auflösen. Je nach Durchzugsbahn und Lage zu anderen Druckgebilden entstehen charakteristische Strömungsmuster, die für den Transport unterschiedlicher Luftmassen verantwortlich sind. Das heißt, innerhalb des drehenden Systems kommt es zu einer Verfrachtung warmer und kalter (trocken oder feucht) Luft. Trifft hierbei Kaltluft auf rückseitige Warmluft, entsteht eine Kaltfront verbunden mit Hebungsprozessen und charakteristischer Wolkenbildung (vgl. Abbildung 2.1). Das Strömungsverhalten im Zusammenhang mit der Zyklonogenese hat daher einen erheblichen Ein-

³Frontalzone in den mittleren bis hohen Breiten mit höchsten Temperaturgradienten.

fluss auf die Wolkenausprägung im Raum Deutschland. Entscheidend ist auch, inwieweit durch die jeweilige Strömungsrichtung eine verstärkte Feuchtigkeitszufuhr über dem Ozean stattgefunden hat. Westliche Strömungen sind in diesem Zusammenhang begünstigt (Häckel 2005, S. 288 ff.).

Um die verschiedenen Strömungsmuster bzw. Wetterlagen besser überblicken und zuordnen zu können, bedient man sich einer sogenannten Großwetterlagen-Klassifikation, wie z.B. die von Hess-Brezowski (vgl. Beck 2000, S. 106 - 108, Heyer et al. 2005, S. 164). Anhand der Druckverteilung im Nordatlantik und in Europa wird die Hauptströmung ermittelt, die mit Hilfe der entsprechenden Himmelsrichtung und des entscheidenden Druckgebildes gekennzeichnet wird. Eine erste grobe Einteilung erfolgt durch die Strömungsrichtung (zonal, meridional oder gemischt) in Großwettertypen, wie z.B. West (zonal), Süd (meridional) oder Südwest (gemischt). Die genauere Aufteilung erfolgt je nach Druckzentrum in Großwetterlagen, wie z.B. West-Antizyklonal (WA), Südwest-Zyklonal (SWZ) oder Zentrales Hoch über Mitteleuropa (HM). Mit Hilfe dieser Klassifikation können 29 Wetterlagen differenziert werden, die einen Vergleich des Wettergeschehens in Europa, auch in Bezug auf die Wolkenverteilung, vereinfachen. Da insbesondere das Wettergeschehen im Bereich von Wetterfronten zu einer unterschiedlichen Wolkenbedeckung führen kann, wird in den nächsten Abschnitten näher darauf eingegangen. Zur Veranschaulichung sind modellhafte Bilder einer typischen Warm- und Kaltfront dargestellt. Innerhalb der einzelnen Wetterlagen kann es natürlich zu komplexeren Ausprägungen und Mischformen kommen (Sönning und Keidel 2007).

Im Bereich einer Warmfront (vgl. Abbildung 2.1) trifft warme Luft auf rückseitig kalte Luft. Als Folge kommt es zu einem Aufgleiten der warmen Luft. Bei genügender Feuchte führt dies zu einem umfangreichen und nach vorn aufwärts orientierten Wolkensystem. Die Hebung resultiert in tiefer Bewölkung, die innerhalb des Warmfrontsystems zu einem dichten Nimbostratus verschmelzen kann, der im Extremfall als geschlossene stratiforme Wolkenmasse eine vertikale Mächtigkeit von 6 - 9 km erreicht. In horizontaler Richtung kann es zu einer Breite von 200 - 300 km kommen. Dort ist gewöhnlich mit anhaltenden kräftigen Niederschlägen in Form von Regen oder Schnee zu rechnen. Vor dem Schlechtwettergebiet befindet sich eine gleichmäßige Zone mit AsCs-Bewölkung (vgl. Tabelle 2.2). In der noch weiter vorgelagerten Kaltluft kommt es durch Absinkprozesse zu freundlichen Wetterbedingungen. Erste Anzeichen der herannahenden Warmfront sind Cirrusbänder, die etwa frontparallel orientiert sind und 500 - 800 km vor der Bodenfront auftauchen. Im Endeffekt hängt die Ausprägung der Wolken von der ursprünglichen Feuchteverteilung und der Intensität der Vertikalbewegung ab. Wolkenschichten und wolkenlose Räume können dadurch in vielseitiger Form abwechseln. Die Neigung der Warmfront ist aufgrund der Reibungseffekte relativ gering. Für die mittlere Troposphäre sind Neigungswerte von 1:100 charakteristisch. Abschließend sei erwähnt, dass eine Warmfront auch mit einer Abkühlung verbunden sein kann, wenn im Sommer maritime, wolkenreiche Warmluftmassen,

wolkenarme und dadurch in Bodennähe stark aufgeheizte Polarluft ersetzen (Maskierte Warmfront) (Kurz 1990, S. 60 - 61).

Bei einer Kaltfront (vgl. Abbildung 2.1) trifft die kalte Luft auf die rückseitige warme Luft und zwingt diese durch zunehmende Labilisierung zu einem abrupten Aufstieg, was auch hier zu charakteristischen Wolkenformen führt. Man unterscheidet zwei Typen der Kaltfront. Die Ana-Kaltfront (Kaltfront erster Art) und die Kata-Kaltfront (Kaltfront zweiter Art). Die Ana-Kaltfront ist charakteristisch für den Randbereich einer Zyklone. Ihre Verlagerungsgeschwindigkeit ist meist nur langsam, weshalb sie auch als „slow moving coldfront“ bezeichnet wird. An solchen Fronten bilden sich häufig Wellen, die den ersten Ansatz für die Entstehung von Tiefdruckgebieten darstellen. Die Erscheinungsform dieser Front ähnelt sehr der einer Warmfront, jedoch sind Bewölkung und Niederschlag postfrontal angeordnet. Das Wolkensystem ist schmaler und reicht im Allgemeinen nicht so hoch. Ein Cirrus-Schirm fehlt in vielen Fällen vollständig. Der unterste Teil der Kaltfront steht sehr steil, und die Hebung der Warmluft ist dort maximiert. Bei potentieller Instabilität der Warmluft können konvektive Umlagerungen ausgelöst werden, so dass die Frontpassage meist von Schauern oder Gewittern begleitet wird, ehe der gleichförmige Niederschlag beginnt. Im Gegensatz zur Warmfront kommt es bei der Kaltfront innerhalb der zyklonalen Entwicklung zu einer raschen Änderung der Strömungs- und Bewölkungsstruktur. Von der Rückseite her setzt sich in der Warmluft immer mehr eine Relativbewegung in Verlagerungsrichtung, die Front nach vorn kreuzend, durch. Das Transportband warmer Luft richtet sich auf und neigt sich immer weiter nach vorn. Als Folge davon ist die Bewölkung direkt an und vor der Bodenfront, mit Höhen von 3 - 5 km relativ niedrig, während sie mit zunehmender Entfernung anwächst. In der Höhe herrscht eine Relativströmung von hinten nach vorn über die Front hinweg, rückseitig absinkend, vorderseitig aufsteigend. Durch die Absinkbewegung wird eher trockene Luft herangeführt. Im Endeffekt nimmt die Ana-Kaltfront einen Kata-Kaltfront-Charakter an. Eine solche Kaltfront verlagert sich gewöhnlich sehr rasch und wird auch als „fast running cold front“ bezeichnet. Ihre Passage ist mit kräftigen Böen verknüpft, die Sturm und Orkanstärke erreichen können. Entsprechend stark ist der Hebungsantrieb vorderseitig des Kaltluftkeils. Niederschlag fällt fast ausschließlich präfrontal, während nach der Frontpassage ein rasches Aufklären erfolgt, das über mehrere Stunden anhalten kann (Kurz 1990, S. 61 - 63).

Da sich im Allgemeinen die Warmfront langsamer bewegt als die Kaltfront, kommt es in Bodennähe zu einer Schrumpfung des zwischen beiden Fronten aufgespannten Warmluftsektors (vgl. Abbildung 2.1). Dies führt im Laufe der Zeit zu einer Verschmelzung der Fronten, die sogenannte Okklusion. Hierbei kommt es zu einer Abhebung der Warmluft und die vorderseitige Kaltluft trifft am Boden auf die rückseitige Kaltluft. Da beide Kaltluftbereiche jedoch meist unterschiedliche Temperaturen aufweisen, bleibt der Frontcharakter bestehen (Schönwiese 2008, S. 181 - 182).

2.3.2 Einflussfaktor Orographie

Da neben diversen anderen Hebungsvorgängen von Luftmassen (vgl. Kapitel 2.1) der erzwungene Auf- und Abstieg an Gebirgen ebenso in einer entsprechenden Wolkenbildung bzw. Wolkenauflösung resultiert, ist an dieser Stelle die Besonderheit des Föhnprozesses zu erwähnen. Es handelt sich um einen warmen Fallwind im Lee eines Gebirges, der seine höhere Temperatur (verglichen mit gleichem Niveau im Luv) aus der latenten Wärme des Wasserdampfes bezieht, die beim Aufsteigen im Luv freigeworden ist. Aus diesem Grund ist der Föhn trockener als die Luft im Luv bei gleicher Höhe. Er hat seinen Namen aus dem Alpengebiet, kommt jedoch auch am skandinavischen Bergland, an den Rocky Mountains und vielen anderen Gebirgen der Erde vor. Seine Entstehung ist mit thermodynamischen Zustandsänderungen verknüpft, die die Luft erfährt, wenn sie ein Gebirge überströmt (vgl. Abbildung 2.2) (Häckel 2005, S. 93, Schönwiese 2008, S. 175 - 176).

Am Beispiel der Alpen wird, in Verbindung mit einer entsprechenden Großwetterlage, feuchte Luft von Süden gegen die Alpen geführt und zum Aufstieg gezwungen. Beim Aufsteigen kühlt sich die Luft an der Alpensüdseite zunächst trockenadiabatisch um ca. 1 K/100m ab, bis Sättigung erreicht ist und es zur Wolkenbildung kommt. Ab hier setzt feuchtadiabatische Abkühlung mit Werten von ca. 0,5 - 0,8 K/100m ein. Durch Niederschläge sinkt bei diesem Vorgang der Feuchtegehalt der Luft. Beim Überströmen des Gebirges steigt die mit weniger Feuchte angereicherte Luft ab und erwärmt sich. Hierbei löst sich die überhängende Wolke (Föhnkamm oder Föhnmauer), die dem Gipfel aufliegt, schon nach wenigen hundert Metern auf, und der weitere Abstieg der Luft erfolgt trockenadiabatisch. Aus der Tatsache, dass sich die Luft beim Aufsteigen überwiegend feuchtadiabatisch abkühlt, beim Absteigen jedoch vorrangig trockenadiabatisch erwärmt, kommt es zu einer höheren Temperatur im Lee, als in vergleichbarer Höhe im Luv. Gleichzeitig ist das Wetter auf der Luvseite des Gebirges stark bewölkt und regnerisch (Stauniederschlag), während auf der Leeseite durch den Föhn trotz Luftdruckabfalls mit freundlichem und wolkenarmem Wetter zu rechnen ist.

Da der Föhnprozess in gewisser Weise an allen Erhebungen bzw. Gebirgen, je nach Strömungsrichtung vorkommen kann, spielt er auch bei der regionalen Wolkenausprägung über

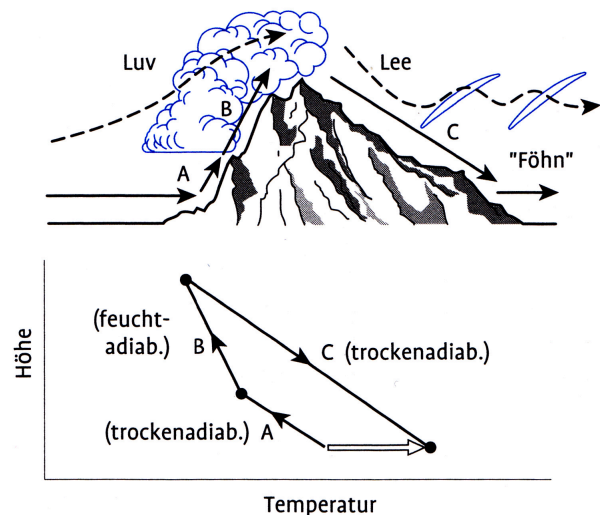


Abb. 2.2: Wolkenbildung bei der Überströmung von Erhebungen (Gebirgen), inklusive Temperatureffekten im Lee- und Luv-Bereich. *Quelle: Schönwiese 2008*

Deutschland eine entscheidende Rolle. Aufgrund verschiedener Großwetterlagen kommt es zu unterschiedlichen Strömungsrichtungen, wodurch insbesondere die Mittelgebirge verschiedenen Anströmungen ausgesetzt sind. Dies kann einerseits zu einer Staulage, andererseits aber auch zu einem gewissen Föhn Effekt führen. Man kann vermuten, dass sich diese Effekte in einer regionalen Differenzierung des Wolkenbedeckungsgrades widerspiegeln (Heyer et al. 2005, S. 361 - 365).

2.3.3 Zirkulationsdynamische Variabilität und Wolkenbedeckung

In den vorherigen Abschnitten wurde deutlich, dass das allgemeine Wettergeschehen und damit auch die Wolkenverteilung über Deutschland, vor allem von den Druckverhältnissen im Nordatlantik und Europa abhängen. Hierbei kann es auf kleinen Zeitskalen (Stunden oder Tage) zu einer enorm hohen Variabilität kommen. Der Durchzug einer Kaltfront benötigt teilweise nur wenige Stunden. Um die Variabilität der Bewölkungsverhältnisse auf größeren Zeitskalen abschätzen zu können, muss man sich mit den durchschnittlichen Druckgegebenheiten bzw. den durchschnittlichen Zirkulationsmustern im Bereich Nordatlantik auseinandersetzen. Insbesondere die Luftdruckschaukel zwischen Azorenhoch und Islandtief, die sogenannte Nordatlantische Oszillation (NAO) ist hier von besonderer Bedeutung. Es handelt sich um eine negative Luftdruckkompensation zwischen beiden Druckgebilden. Obwohl die einzelnen Druckgebiete (vor allem dynamische Tiefdruckgebiete) eine gewisse räumliche Variabilität aufweisen, lassen sich die Druckzentren über den Azoren und über Island gut lokalisieren. Vor allem bei einer langfristigen Betrachtung über mehrere Monate oder Jahre. Während bei der Einordnung von Großwetterlagen (vgl. Kapitel 2.3.1) versucht wird, die einzelnen Strömungsmuster zu jeweiligen Hauptströmungsrichtungen zusammenzufassen, liegt hier das Bestreben darin, ein durchschnittliches Strömungsverhalten über einen ausgewählten Zeitraum abzuschätzen. Beispielsweise über die Wintermonate, in denen das Phänomen am stärksten ausgeprägt ist. Vom Prinzip her ergibt sich folgendes Verhalten. Eine starke Ausprägung des Azorenhochs ist mit einer ebenfalls starken Ausprägung des Islandtiefs verbunden und umgekehrt (Wanner et al. 2001, Chaboureaud und Claud 2006).

Mit Hilfe eines NAO-Index-Wertes lässt sich so die zirkulationsdynamische Variabilität im Nordatlantik für einen bestimmten Zeitraum abschätzen. Als Grundlage der Indexerstellung dient die Luftdruckdifferenz zwischen den Azoren und Island. In der Regel bezieht man sich auf die Wintermonate, es gibt aber auch Index-Werte für ganze Jahre oder für nur einen Monat. Positive (hoher Indexwert) und negative (niedriger Indexwert) NAO-Phasen sind das Resultat dieser Betrachtung. In einer positiven Phase kommt es zu einer starken zonalen bzw. westlichen Strömung über Zentraleuropa, verbunden mit milden und feuchten Wintern. Negative NAO-Phasen sind dagegen mit einer verstärkt meridionalen Strömung verbunden, was zu kalten und

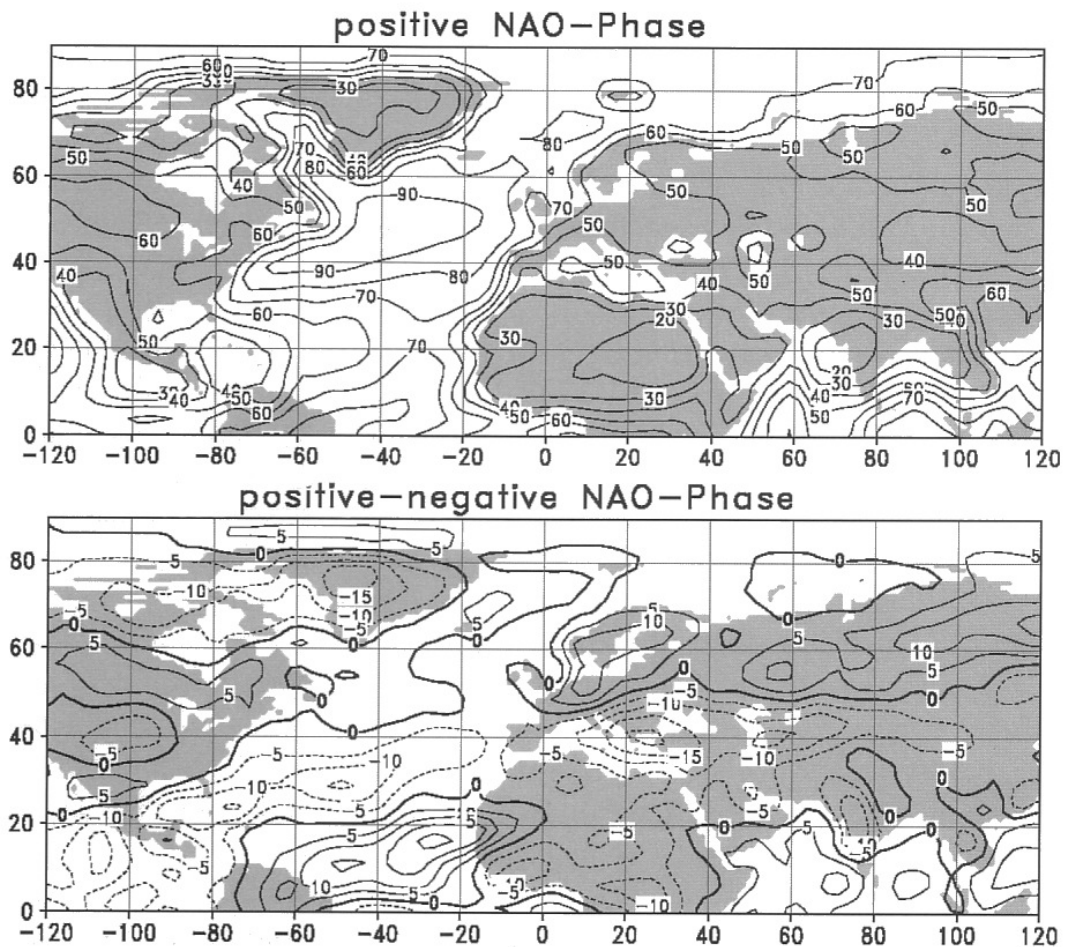


Abb. 2.3: Wolkenbedeckungsgrad (%) während positiver NAO-Phasen von Dezember bis März (oben). Abgeleitet aus Satellitenbeobachtungen für den Zeitraum 1983 - 1990. Differenz des Bedeckungsgrades (%) zwischen positiver und negativer NAO-Phase (unten). *Quelle: Simmer 2000*

trockenen Wintern führt. Das heißt, negative und positive NAO-Phasen sollten auch mit typischen Mustern der Bewölkung verbunden sein (Wanner et al. 2001, Schönwiese 2008, S. 183 - 185).

Bei der Betrachtung von Wolkenbedeckungsgraden (vgl. Abbildung 2.3) während positiver NAO-Phasen (Wintermonate) erkennt man Gebiete mit höchster Bewölkung quer durch den Atlantik von Südwesten nach Nordosten. Selbst über Nordwesteuropa liegt die Wolkenbedeckung noch bei über 70 %. Eine Begründung liegt darin, dass die starken Westwinde mit einer erhöhten Verdunstung einhergehen, die zu einer intensiven Wolkenbildung während dieser Phase führt. Unterschiede in der Bewölkung zwischen positiver und negativer Phase sind, neben Bereichen des tropischen Ostatlantik, vor allem in Nord- und Südeuropa zu erkennen. In Nordeuropa steigt der Bewölkungsgrad bei einer positiven Phase um etwa 10 % an, während es

in mediterranen Gebieten zu einer starken Abnahme der Bewölkung, verbunden mit Einbußen beim Niederschlag, kommt (vgl. Abbildung 2.3) (Simmer 2000, S. 30 - 33).

Obwohl es in dieser Arbeit verstärkt um die methodische Ausarbeitung der Detektions- bzw. Klassifizierungsverfahren geht und die Datengrundlage auf ein Jahr begrenzt ist, sollten die zuvor beschriebenen Gegebenheiten dennoch berücksichtigt werden. Insbesondere bei der exemplarischen Beurteilung bzw. Einordnung der Ergebnisse.

2.4 Die Rolle der Wolken im Strahlungshaushalt

Innerhalb des dynamischen Systems entstehen Wolken, wie zuvor beschrieben, meist durch Vertikalbewegungen. Wenn man sich z.B. ein Tiefdruckgebiet, mit entsprechenden Fronten, auf einem Wetter-Satellitenbild anschaut (vgl. Abbildung 1.1), stellt man schnell fest, dass sich Wolken als Ausdruck der dynamischen Gegebenheiten darstellen. Sie sind sozusagen eine Begleiterscheinung der atmosphärischen Zirkulation. Auf der anderen Seite greifen Wolken unmittelbar in den Prozess ein, der die atmosphärische Zirkulation antreibt, den Strahlungshaushalt des Systems Erde-Atmosphäre⁴ (siehe Simmer 2000, S. 10 - 13). Aus diesem Grund und auch zum besseren Verständnis späterer Untersuchungsmethoden wird in den folgenden Abschnitten der Strahlungseinfluss von Atmosphäre und Wolken näher beschrieben.

2.4.1 Atmosphärischer Einfluss auf Strahlungsvorgänge

Die Atmosphäre der Erde hat durch ihre spezifische Zusammensetzung einen gewissen Einfluss auf Strahlungsvorgänge, sowohl auf solare Strahlung als auch auf langwellige Wärmestrahlung. Obwohl es sich bei einem solaren Strahlenbündel um ein kontinuierliches Spektrum handelt, in dem alle Wellenlängen enthalten sind, nimmt man eine Unterscheidung in Kurz- und Langwellig vor, da innerhalb der solaren Strahlung lange Wellenlängen sehr geringe Strahlungswerte annehmen. Der Wellenlängenbereich von 0,2 - 4 μm wird deswegen als kurzwellige solare Strahlung bezeichnet, und lässt sich in weitere Bereiche unterteilen (vgl. Abbildung 2.4):

- Sichtbares Licht (0,36 - 0,76 μm)
- Ultraviolett (< 0,36 μm)
- Infrarot (> 0,8 μm)

Da die Sonne mit einer Temperatur von ca. 5800 K (Sonnenoberfläche) ausstrahlt, ergibt sich ein Strahlungsmaximum bei Wellenlängen des sichtbaren Lichts (vgl. Abbildung 2.4). Die Erde (inklusive Atmosphäre) strahlt nur mit einer Temperatur von ca. 300 K und hat deshalb

⁴Durch unterschiedliche regionale Nettostrahlungsflüsse wird die Zirkulation angetrieben.

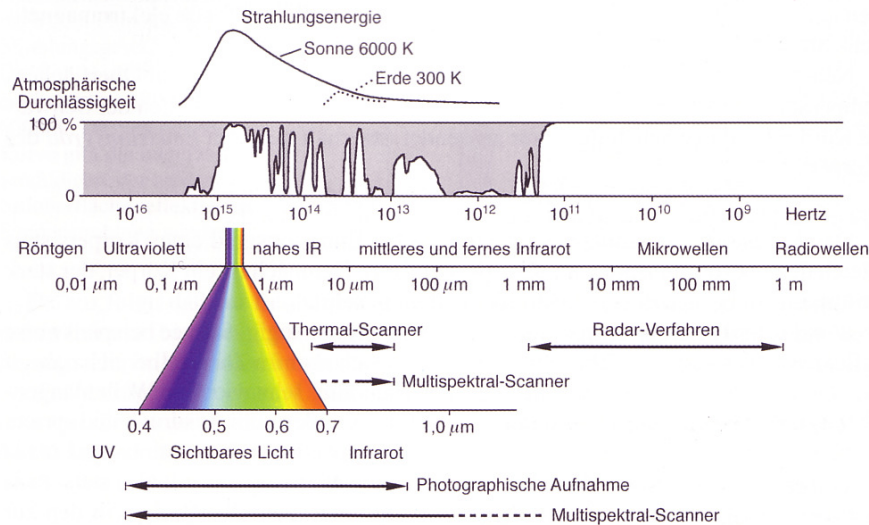


Abb. 2.4: Das elektromagnetische Spektrum in Verbindung mit der Strahlungsenergie der Sonne und der Durchlässigkeit der Atmosphäre. *Quelle: Albertz 2001*

ihr Strahlungsmaximum im Bereich um etwa $10 \mu\text{m}$. Zwar wird auch hier in einem kontinuierlichen Spektrum ausgestrahlt, jedoch nehmen in diesem Fall die kurzen Wellenlängen verschwindend geringe Werte an. Daher bezeichnet man den Infrarot-Wellenlängenbereich von ca. $3,5 - 100 \mu\text{m}$ auch als langwellige Wärmestrahlung bzw. terrestrische Strahlung (vgl. ebenfalls Abbildung 2.4). Beide Strahlungsflüsse sind die wesentlichen Bestandteile innerhalb des Strahlungshaushaltes (vgl. Abbildung 2.5), und spielen für klimatologische und meteorologische Vorgänge eine zentrale Rolle (Häckel 2005, S. 161 ff.).

Von dem am Oberrand der Atmosphäre ankommenden Strahlungsfluss (ca. 1370 W/m^2) steht durch die Kugelgestalt der Erde und die tägliche Rotation im Mittel nur ein Viertel zur Verfügung. Hinzu kommt, dass 30 % dieser Strahlung ungenutzt ins All zurück reflektiert werden (planetare Albedo), der Rest kann sowohl an der Erdoberfläche als auch an Bestandteilen der Atmosphäre absorbiert und in Wärme umgewandelt werden (vgl. Abbildung 2.5). Die Reflexion läuft diesbezüglich direkt und indirekt ab, wobei man von direkter Reflexion spricht, wenn Strahlung sowohl an der Erdoberfläche (verstärkt an Schneeflächen oder Wüsten) als auch an der Atmosphäre (z.B. durch Wolken) ohne Umweg ins All zurück reflektiert wird. Bei einer indirekten oder diffusen Reflexion kommt es zu einer Streuung an Bestandteilen der Atmosphäre. Luftmoleküle, Spurengase und Staub (Aerosol) sind neben atmosphärischer Feuchtigkeit (Wolken, Nebel oder Eiskristalle), die wesentlichen Substanzen in der Atmosphäre, die Strahlung reflektieren bzw. streuen, aber auch absorbieren. Gase weisen in Bezug auf ihr Absorptionsverhalten eine Besonderheit auf. Die Absorption vollzieht sich nur in bestimmten Wellenlängenbereichen, den sogenannten Banden. Das heißt, aus einem Strahlenbündel wird nur in entsprechenden Wellenlängen Strahlungsenergie absorbiert. Gleiches gilt für die Emission

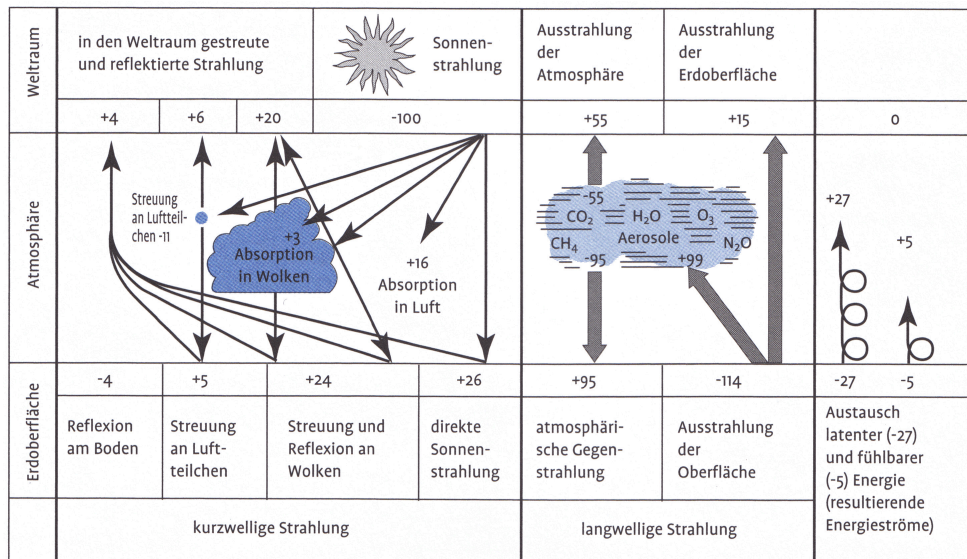


Abb. 2.5: Strahlungshaushalt des Systems Erde-Atmosphäre. *Quelle: Häckel 2005*

der Wärmestrahlung, die bei Gasen ebenfalls nur in einzelnen Wellenlängenbereichen stattfindet. Ozon (O_3), Wasserdampf (H_2O) oder Kohlendioxid (CO_2) sind in diesem Zusammenhang vorrangig zu nennen (Simmer 2000, Häckel 2005, S. 161 ff.).

Anhand von Abbildung 2.5 und der beschriebenen Vorgänge wird deutlich, dass Strahlung beim Durchgang der Atmosphäre selektiv geschwächt wird (Extinktion). In Bezug auf solare Strahlung ist vor allem die Sonnenhöhe entscheidend, die die Weglänge der Strahlen durch die Atmosphäre bestimmt. Mit zunehmender Wegstrecke durch die Atmosphäre, steigt der Strahlungsverlust durch diffuse Reflexion und selektive Absorption (Häckel 2005, S. 165).

Neben der Absorption und Reflexion kurzwelliger Strahlung in der Atmosphäre, findet auch eine Absorption der nach oben gerichteten terrestrischen Wärmestrahlung statt. Durch Bestandteile der Atmosphäre wird diese Strahlung je nach Wellenlängenbereich absorbiert und gleichzeitig emittiert. Man spricht vom sogenannten Treibhauseffekt. Verantwortlich hierfür sind, neben Wolken (flüssiges Wasser bzw. Eiskristalle) und Wasserdampf, vor allem Spurengase, wie Kohlendioxid (CO_2), Methan (CH_4) und troposphärisches Ozon (O_3). Solche Spurengase absorbieren langwellige Wärmestrahlung in besagten Absorptionsbanden, und strahlen nach Stefan-Boltzmann proportional zur vierten Potenz der Temperatur (Temperatur des ausstrahlenden Körpers) wieder zurück. Dies führt zu einer Erwärmung an der Erdoberfläche und macht das Klima, wie wir es kennen, erst möglich (natürlicher Treibhauseffekt, Häckel 2005, S. 170 - 171, Simmer 2000).

Zusammenfassend kann man sagen, dass nur ein gewisser Teil der solaren Strahlung die Erdoberfläche erreicht, während ein anderer Teil an der Atmosphäre reflektiert oder absorbiert wird. Gleichzeitig absorbiert die Atmosphäre terrestrische Ausstrahlung und emittiert selbst

Wärmestrahlung. Wolken sind bei diesen Vorgängen ein wesentlicher Faktor, deren Einfluss nachfolgend dargestellt wird.

2.4.2 Wirkungsweise von Wolken auf Strahlungsvorgänge

Das Wolken einen Effekt auf die Strahlungsflüsse haben, wird an zwei Beispielen deutlich. Zum einen halten sie die solare Strahlung teilweise ab, was sich z.B. zeigt, wenn an einem generell wolkenlosen Tag aufkommende Wolken die Sonne verdecken. Eine spürbare Abkühlung ist die Folge. Ähnliches gilt in umgekehrter Weise, Wolken absorbieren die Wärmestrahlung der Erde und strahlen selbst zurück. Dies ist dann zu spüren, wenn in einer klaren Nacht keine Wolken vorhanden sind und eine verstärkte Wärmeabstrahlung stattfindet. Die Minimumtemperaturen in solchen Nächten sind wesentlich niedriger als in bewölkten Nächten. Verantwortlich hierfür ist ein breites Spektralband in der Atmosphäre, in dem diese fast transparent ist. Wolken schließen dieses Fenster sehr effektiv und eine verstärkte Abkühlung der Erdoberfläche wird gehemmt. Dies ist Bestandteil des zuvor bereits erwähnten Treibhauseffektes. Neben anderen Treibhausgasen sind Wolken in diesem Zusammenhang von zentraler Bedeutung. Sie sind die stärksten Treibhauseffekt-Erzeuger, die wir kennen (vgl. Simmer 2000, S. 10 - 13). Die Reflexion und Streuung kurzwelliger Strahlung an Wolken hängt von verschiedenen Faktoren ab (vgl. Kärner und Keevallik 1993, Häckel 2005, S. 183 ff.):

- Physikalischer Zustand der Wolken
- Vertikale Mächtigkeit
- Wolkenhöhe und Bedeckung
- Einfallswinkel der Sonnenstrahlung

Aufgrund der schier unerschöpflichen Kombinationsmöglichkeiten dieser Einflussfaktoren lassen sich daher nur sehr allgemein gefasste Grundregeln ableiten, die den Einfluss von Wolken auf kurzwellige Strahlung beschreiben können. Beispielsweise bewirken ausgeprägte Schichtwolken eine erhebliche Vergrößerung des diffusen Himmelslichts. Ähnliches gilt für eine teilweise Bewölkung. Die diffuse Strahlung wird vergrößert, bei gleichzeitiger Verringerung der Globalstrahlung⁵ (im Vergleich zu einem wolkenlosen Himmel). Im speziellen Fall eines dünnen Besatzes des Himmels mit vertikal mächtigen, dichten Quellwolken kann in einer von Wasserdampf und Trübung freien Atmosphäre die Globalstrahlung merkbar größere Werte annehmen, als bei völliger Wolkenlosigkeit. Wolken würden in diesem Fall die Einstrahlung am Boden erhöhen (Blüthgen und Weischet 1980, S. 93).

⁵Direkte und diffuse solare Strahlung die die Erdoberfläche erreicht und in Wärme umgewandelt werden kann.

$$I_v^{clear} = \tau_0^H(\nu, \psi, \theta) \varepsilon_s(\nu, \theta, \psi) B_\nu(T_s) + \int_0^H B_\nu(T(H)) \frac{\delta \tau_h^H(\nu, \psi, \theta)}{\delta h} dh + \tau_0^H(\nu, \psi, \theta) I_R(\nu, \psi, \theta, \phi) \quad (2.1)$$

$$I_v = (1 - \varepsilon_c(\nu, \theta, \psi)) I_v^{clear} + \varepsilon_c(\nu, \theta, \psi) B_\nu(T_c) \quad (2.2)$$

$B_\nu(T)$ = Planckfunktion bei Frequenz ν und Temperatur T

I_v = Wärmestrahlung, die den Satellitensensor erreicht (clear = wolkenfreie Bedingungen)

I_R = Reflektierte atmosphärische Gegenstrahlung der Erdoberfläche (im thermalen Infrarotbereich meist gering)

τ_0^H = Strahlungstransmission der Atmosphäre (H = Höhe der Atmosphärenoberkante)

ε_s = Emissivität der Erdoberfläche

ε_c = Emissivität der Wolken

ν = Frequenz

ψ = Variable atmosphärischen Zustandsgrößen (Temperatur, Feuchte usw.)

θ = Satelliten-Zenit-Winkel

ϕ = Satelliten-Azimut-Winkel

Quelle: Saunders und Kriebel 1988

Insgesamt werden ca. zwei Drittel der planetaren Albedo durch Wolken verursacht. Gleichzeitig geschieht mehr als die Hälfte der an Wolken emittierten Wärmestrahlung an den Wolkenoberkannten. Das heißt, Wolken strahlen nicht nur im Sinne des Treibhauseffektes zur Erde zurück, sondern sie strahlen genauso Wärme ins All ab. Die Wärmestrahlungsverhältnisse, die einen Satellitensensor in Abhängigkeit von der Wellenlänge (bzw. Frequenz) erreichen, sind in Gleichung 2.1 (ohne Wolken) und 2.2 (mit Wolken) dargestellt. Neben den Betrachtungswinkeln des Satelliten und der unterschiedlichen Emmissivität von Wolken und wolkenfreien Flächen, ist auch der unterschiedliche Einfluss von Wolken auf die Wärmestrahlungstransmission der Atmosphäre wesentlich (Saunders und Kriebel 1988).

An all diesen Vorgängen kann man erkennen, dass die Wirkungen im solaren und infraroten Spektralbereich, bezüglich der Nettostrahlungsbilanz, teilweise gegensätzlich sind. Das heißt, Wolken kühlen und erwärmen gleichzeitig. Hieran wird deutlich, dass ein vermehrter Informationsgewinn über das Vorkommen von Wolken eine enorme Bedeutung hat. Den Effekt von Wolken auf die Strahlungsbilanz des Gesamtsystems Erde-Atmosphäre bezeichnet man als wolkeninduzierten Strahlungsantrieb (Cloud Radiative Forcing, CRF). Hierbei handelt es sich um die Differenz der beobachteten Nettostrahlungsbilanz Q am Oberrand der Atmosphäre, zur Nettostrahlungsbilanz Q_0 , die herrschen würde, wenn man die Wolken aus dem System entfernt. Wenn Q über den Globus gemittelt wird, sollte ungefähr Null herauskommen. Das Gesamtsystem ist im Strahlungsgleichgewicht. Mittelt man Q_0 , muss dies nicht der Fall sein, da diese Situation (Temperatur und Feuchte wie bisher, aber wolkenlose Atmosphäre) nicht real

ist, und keinem Gleichgewichtszustand entsprechen muss. Satellitenmessungen haben ergeben, dass man mit ca. -20 W/m^2 von einem negativen wolken-induziertem Strahlungsantrieb ausgehen muss (Kärner und Keevallik 1993, S. 77 ff., Simmer 2000, S. 15).

$$CRF = Q - Q_0 \quad (2.3)$$

CRF = Cloud Radiative Forcing

Q = Nettostrahlungsbilanz mit beobachteten Wolken

Q_0 = Nettostrahlungsbilanz ohne Wolken

Quelle: Simmer 2000

Die globale Betrachtung zeigt, dass Wolken eine kühlende Wirkung auf die Erde haben. Man kann sich jedoch vorstellen, dass das Cloud Forcing nicht überall auf der Erde -20 W/m^2 beträgt, sondern regional variiert, und dass es durch verschiedene Wolkenarten unterschiedlich ausfällt. So müssen niedrige, dichte Wolken ein starkes negatives Forcing besitzen, da sie einen großen Anteil der solaren Strahlung reflektieren und somit stark kühlen. Doch ihre Wärmestrahlung unterscheidet sich kaum von der wolkenlosen Erdoberfläche, da ihre Obergrenzen niedrig und sie somit relativ warm sind (vgl. Abbildung 2.6). Umgekehrtes gilt für hohe und dünne Wolken.

Dazu muss man wissen, dass Wolken für Infrarotstrahlung viel eher optisch dicht sind, als für solare Strahlung. Dünne Wolken lassen also noch viel Sonnenlicht bis zum Boden durch, im Infrarotbereich verhalten sie sich dagegen annähernd wie schwarze Körper⁶. Sie strahlen also mit ihrer sehr kalten Obergrenzen-Temperatur (bis zu -70 °C) relativ wenig Energie ins All ab. In Bezug auf Infrarotstrahlung können Wolken daher anhand niedriger Werte erkannt werden, was sich die verschiedenen Wolkendetektionsverfahren zu Nutze machen.

Bei der Betrachtung einzelner Regionen kann man erkennen, dass besonders Wüsten und zentrale Gebiete der Kontinente meist durch einen negativen, wolkeninduzierten Strahlungsantrieb auffallen (vgl. Abbildung 2.7). In den mittleren und auch höheren Breiten der jeweiligen Winterhemisphäre ergibt sich ein durch Wolken verursachter positiver Strahlungsantrieb. Dies

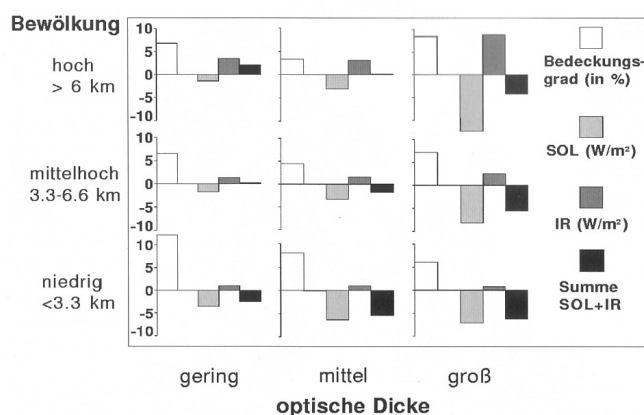


Abb. 2.6: Cloud Forcing in Abhängigkeit der Wolkenhöhe und optischen Dicke. Solare Strahlung (SOL), Infrarotstrahlung (IR).
Quelle: Simmer 2000

⁶Idealisierter physikalischer Körper mit frequenzunabhängiger Absorption und Emission (vgl. Häckel 2005, S. 172).

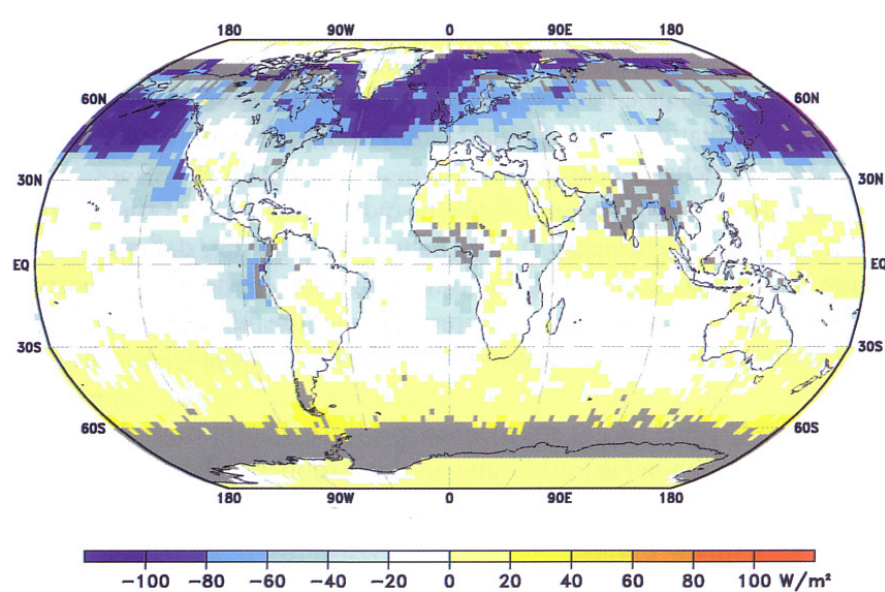


Abb. 2.7: Räumliche Differenzierung des Cloud Forcing (Beispiel: Juli 1985). Fehlende Werte sind grau dargestellt. *Quelle: Simmer 2000*

liegt zum einen an der geringeren solaren Einstrahlung und zum anderen an der gleichzeitig bestehenden isolierenden Wirkung der Wolken. Auf der Sommerhemisphäre ergibt sich dagegen, in den ozeanisch geprägten Gebieten der mittleren Breiten, die von niedrigen und mächtigen Wolken dominiert werden, ein verstärkt negatives Cloud Forcing (vgl. Abbildung 2.7). Die Verteilung des wolkeninduzierten Strahlungsantriebs zeigt, dass Wolken eine regelnde Wirkung im Klimasystem zukommt. Schwache meridionale Strahlungsgegensätze auf der Sommerhemisphäre werden vergrößert, während starke Strahlungsgegensätze auf der Winterhemisphäre reduziert werden. Gegensätze zwischen Ozean und Kontinent werden ebenfalls reduziert (Simmer 2000, S. 13 - 18).

Um den Einfluss von Wolken auf Strahlungsvorgänge noch besser verstehen zu können, ist eine genaue Abschätzung der Bewölkungssituation sicherlich von Vorteil. Dies gilt sowohl auf regionaler als auch auf globaler Ebene. Mit einer höheren räumlichen Auflösung, bezogen auf die Wolkenbedeckung, können einzelne Effekte auf Strahlungsvorgänge gegebenenfalls noch besser abgeschätzt werden.

2.5 Wolken und Klima in Deutschland

In den vorherigen Kapiteln wurde schon auf die Zyklogense und die Druckverhältnisse über dem Atlantik als Einflussfaktor auf die Bewölkungsverhältnisse im Raum Deutschland hingewiesen. Um sich eine grobe Vorstellung über die klimatischen Verhältnisse Deutschlands, auch

in Verbindung mit Wolken, machen zu können, wird in den nächsten Abschnitten, unter Berücksichtigung des Jahresganges, kurz darauf eingegangen. Dies ist auch notwendig, damit eine Einordnung der Ergebnisse in das allgemeine Wettergeschehen besser möglich ist. Ob sich die langjährigen Gegebenheiten in den Ergebnissen widerspiegeln, wird entsprechend zu diskutieren sein. Die klimatologische Betrachtung bzw. Beschreibung erfolgt für den Referenzzeitraum 1961 - 1990.

Januar, Februar, März

Als kältester Monat des Jahres weist der Januar in Deutschland im Gebietsmittel eine Durchschnittstemperatur von $-0,5\text{ °C}$ auf, während im Februar die Durchschnittstemperaturen schon um ca. 1 °C höher liegen. Nur im Bereich der norddeutschen Inseln oder in höheren Lagen kann sich der Februar kälter präsentieren als der Januar. Trotzdem können sich im Februar durch Nordost-Wetterlagen immer noch hochwinterliche Perioden ergeben. Im Gegensatz dazu stellt sich im März die Brücke von Winter zu Frühling ein und die Durchschnittstemperatur liegt bei $3,5\text{ °C}$, wobei die Temperaturen auf bis zu -10 °C absinken, als auch Tageshöchsttemperaturen von $15 - 20\text{ °C}$ erreicht werden können. Die Niederschläge in den Monaten Januar, Februar und März ergeben im vieljährigen Mittel jeweils $15 - 16$ Niederschlagstage bei einer Wassermenge von $49 - 62$ Litern pro Quadratmeter (mm) im Gebietsmittel. In den Monaten Januar und Februar kann die Hälfte dieser Tage durch Schneefall gekennzeichnet sein. Im März ist nur noch bei einem Drittel der Niederschlagsereignisse mit Schneefall zu rechnen. Charakteristisch für den März sind stark wechselhafte Witterungseigenschaften. Die durchschnittliche Wolkenbedeckung beträgt in allen drei Monaten über 70% , wobei ausgeprägte Nebeltage nicht ungewöhnlich sind (Müller-Westermeier 2006, S. 16 - 30).

April, Mai, Juni

Ähnlich wie der März ist der April durch wechselhaftes Wetter gekennzeichnet. Häufig bildet sich hier eine ausgeprägte, von Nordwest nach Südost verlaufende, mit entsprechenden Strömungsmustern verbundene Tiefdruckrinne. Bezogen auf die Temperaturen kommt es zu einem starken Anstieg, wobei ein Monatsmittelwert (Gebietsmittel Deutschland) von $7,4\text{ °C}$ erreicht wird. Im Mai und Juni steigen die Durchschnittstemperaturen weiter von 12 °C auf $15,4\text{ °C}$, und es kommt meist schon zu einer sommerlichen Prägung. Die Niederschläge nehmen im April an Intensität zu und im Bergland werden durchschnittliche Werte von bis zu 90 mm erreicht. Mai und Juni stellen sich hier ähnlich dar. Es kann aber zu Schwankungen zwischen 10 und 200 mm kommen. In allen drei Monaten sind ca. 15 Niederschlagstage zu erwarten und der Grad der Wolkenbedeckung liegt etwas niedriger als in den ersten drei Monaten. Aber auch

hier kann es innerhalb der langfristigen Betrachtung zu einer erheblichen Variabilität kommen (Müller-Westermeier 2006, S. 31 - 42).

Juli, August, September

Die Witterung im Juli ist geprägt durch die jahreszeitlich bedingte Erwärmung. Es ergibt sich eine Durchschnittstemperatur für das Gebietsmittel Deutschland von 16,9 °C. Im August stellt sich dagegen der schon ausklingende Hochsommer ein und durch den vollzogenen Wärmeausgleich zwischen Kontinent und Meer sind stärkere Temperaturschwankungen seltener. Die Durchschnittstemperatur beträgt 16,5 °C, während sie im spätsommerlich geprägten September auf 13,3 °C absinkt. Im September können zwar noch Temperaturwerte von 25 °C und höher erreicht werden, gleichzeitig können nächtliche Minimumtemperaturen in der Nähe des Gefrierpunktes liegen. Die durchschnittliche Niederschlagsmenge des Julis beträgt 78 mm und er ist neben dem Juni der niederschlagsreichste Monat des Jahres. Im August ist die Niederschlags-situation mit 77 mm ähnlich. Je nach Wetterlage (Druckverteilung) kann es in diesen Monaten aber auch zu ausgeprägten Trockenperioden kommen. Im Vergleich zu Juli und August stellt sich der September, mit 61 mm Niederschlagsmenge im Gebietsmittel, allgemein trockener dar. Der Bewölkungsgrad liegt in den drei Monaten zwischen 60 und 70 % (Müller-Westermeier 2006, S. 43 - 54).

Oktober, November, Dezember

Im Oktober ergeben sich bereits herbstliche Verhältnisse, mit einer durchschnittlichen Temperatur von 9 °C. Besonders in den Nächten kann es empfindlich kühl werden. Der nachfolgende November ist dagegen durch den Übergang von Herbst zu Winter gekennzeichnet. Die Durchschnittstemperatur sinkt hier auf 4 °C. Vollständig winterlich zeigt sich der Dezember mit einer durchschnittlichen Temperatur von 0,8 °C (Gebietsmittel). Ähnlich wie der September, weist der Oktober im Vergleich zu November und Dezember relativ geringe Niederschlagswerte auf (56 mm im Landesdurchschnitt). In extrem trockenen Oktobern kann es auch zu Niederschlagswerten von unter 10 mm kommen. Im November steigen die Niederschläge wieder auf 66 mm, und an durchschnittlich 4 Tagen kann es bereits zu Schneefall kommen. Dieser Trend setzt sich im Dezember fort. Es ergeben sich Niederschlagswerte von 70 mm. Je nach Höhenlage erfolgt der Niederschlag als Schnee oder Regen. Der Wolkenbedeckungsgrad steigt in den drei Monaten von knapp unter 70 % bis auf Werte um die 80 %. In jedem der drei Monate kann es zu verstärkter Nebelbildung kommen (Müller-Westermeier 2006, S. 55 - 70).

3 Wolkendetektionsverfahren basierend auf Schwellenwertalgorithmen

Verschiedene Verfahren der Digitalen Bildverarbeitung können eingesetzt werden, um Wolken innerhalb von Satellitenaufnahmen abzugrenzen. Neben dem störenden Einfluss von Wolken bei der Analyse von Landoberflächen, ist das Interesse an den Wolken selbst, vor allem im Rahmen klimatologischer Fragestellungen, von besonderer Bedeutung. Bei den optischen Satellitensensoren, die sowohl mit sichtbaren als auch mit thermalen Kanälen ausgestattet sind, haben sich in den letzten Jahrzehnten vor allem Schwellenwertalgorithmen als Mittel der Wolkendetektion durchgesetzt. In erster Linie ist hierbei das bereits erwähnte Analyse-Tool APOLLO zu nennen. Bezogen auf METEOSAT-Daten sind aber auch die Wolkenprodukte⁷ der Satellite Application Facility on support to Nowcasting and Very Short Range Forecasting (SAFNWC) von entscheidender Relevanz. Da die SAFNWC-Produkte in dieser Arbeit als Vergleichsgrundlage herangezogen werden und Schwellenwertalgorithmen bei der Gewinnung von Trainingsdaten zum Einsatz kommen (vgl. Kapitel 5.3), erfolgt in den folgenden Abschnitten eine kurze Beschreibung dieser Verfahren.

3.1 APOLLO

Mit Hilfe des Analyse-Tools APOLLO werden anhand spektraler Informationen Wolken von Land- und Meeresflächen getrennt. Die erste Entwicklung der Algorithmen erfolgte schon vor mehr als 20 Jahren durch Saunders (1986) bzw. Saunders und Kriebel (1988). Das Tool bezieht sich auf den AVHRR-Sensor, der sich unter anderem auf den polarumlaufenden NOAA-Satelliten befindet. Er bietet Aufnahmen mit einer räumlichen Auflösung von bis zu einem Kilometer bei einer globalen Abdeckung von einem Tag. Durch die Datenverwendung mehrerer NOAA-Satelliten kann die Abdeckung erhöht werden. Die spektrale Information des Sensors setzt sich aus bis zu 5 Kanälen zusammen, mit deren Hilfe die Wolkendetektion durchgeführt werden kann (vgl. Tabelle 3.1 und Kriebel et al. 2003, S. 2389 - 2391).

⁷Meteorologische Produkte erstellt aus den spektralen Eigenschaften der Satellitendaten, hier Wolkeneigenschaften.

Tab. 3.1: NOAA-AVHRR-Kanäle die zur Wolkendetektion genutzt werden. *Quelle: Saunders und Kriebel 1988*

Kanal	Wellenlänge [μm]	Signal	Bezeichnung
1	0,58 - 0,68 (Sichtbar)	Reflektanz	R ₁
2	0,72 - 1,10 (Nahes Infrarot)	Reflektanz	R ₂
3	3,55 - 3,93 (Infrarot)	Temperatur	T _{3,7} oder T3
4	10,3 - 11,3 (Infrarot)	Temperatur	T ₁₁ oder T4
5	11,5 - 12,5 (Infrarot)	Temperatur	T ₁₂ oder T5

Die Detektion mit APOLLO basiert auf einer Reihe von Schwellenwert-Tests, wobei alle 5 AVHRR-Kanäle einbezogen werden. Das vorrangige Ziel besteht in der Einteilung wolkenfreier, vollständig bewölkter und teilweise bewölkter Pixel. Die Testreihen beziehen sich im Wesentlichen auf die Strahlungstemperaturen der Infrarotkanäle (vgl. Abbildung 3.1 und Tabelle 3.1). Tagsüber können aber auch Reflektanzen der Kanäle des sichtbaren Bereiches verwendet werden. APOLLO setzt sich aus zwei Haupt-Testreihen zusammen, dem First Run und dem Second Run. Im ersten Durchlauf werden wolkenfreie (cloud free) und wolkenkontaminierte (cloud-contaminated) Pixel voneinander abgegrenzt, während in der zweiten Testreihe eine weitere Einteilung der wolkenkontaminierten Pixel in „Völlig Bewölkt“ und „Teilweise Bewölkt“ erfolgt.

Da bestimmte Tests nur tagsüber angewendet werden können, ergibt sich eine weitere Einteilung der Testreihen in Tag und Nacht. Während der Tag-Tests können sowohl Reflektanzen solarer Strahlung als auch Strahlungstemperaturen der Infrarotkanäle zur Wolkendetektion beitragen. Es kommen demnach die Kanäle 1 und 2 (Sichtbar und Nahes Infrarot), sowie die Kanäle 4 und 5 (Thermale Kanäle) zum Einsatz. Nachts bedient man sich ausschließlich solcher Schwellenwert-Tests, die nur auf Strahlungstemperaturen basieren (Kanal 3, 4 und 5). Ab einer Sonnenhöhe von 10° werden die Tag-Testreihen verwendet, ansonsten muss auf die Nachtabfolge zurückgegriffen werden (Saunders und Kriebel 1988).

Bei der Betrachtung der Infrarotkanäle können auch *Strahlungstemperatur-Differenzen* (Differenz zweier Infrarotkanäle) eingesetzt werden. Da es sich bei der gemessenen Strahlungstemperatur nicht um die eines schwarzen Körpers handelt, werden an den unterschiedlichen Infrarotkanälen (unterschiedliche Wellenlänge) abweichende Temperaturen gemessen. Dieser Effekt kann bei der Wolkendetektion verwendet werden, da Wolken bei solchen Strahlungstemperatur-Differenzen andere Werte aufweisen als wolkenfreie Flächen (vgl. Abbildung 3.1). Insbesondere bei teilweiser Bewölkung kann es am entsprechenden Pixel zu Unterschieden kommen, z.B. bei Kanal 3 ($3,7 \mu\text{m}$) und Kanal 4 ($11 \mu\text{m}$) (Saunders 1986).

An dieser Stelle muss auf das besondere Strahlungsverhalten, welches sich tagsüber im $3,7 \mu\text{m}$ -Wellenlängenbereich einstellt, hingewiesen werden. Die an Kanal 3 ($3,7 \mu\text{m}$) ankomen-

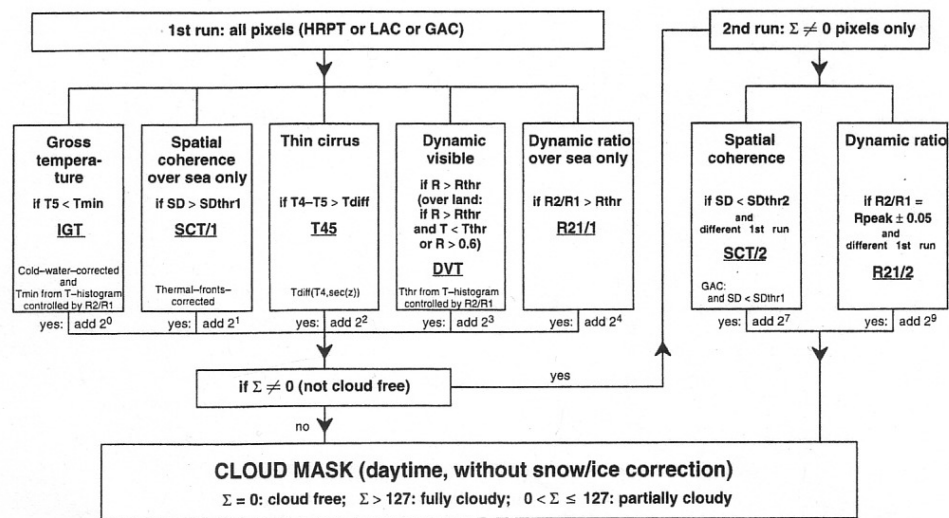


Abb. 3.1: APOLLO Tag-Testreihe. Abgrenzung wolkenfreier und wolkenkontaminierter Pixel (First Run) und Bestimmung völlig bewölkter Pixel (Second Run). T4 und T5 = Strahlungstemperaturen (Kanal 4 und 5), R1 und R2 = Reflektanzen (Kanal 1 und 2), SD = Standardabweichung. *Quelle: Kriebel et al. 2003*

den Strahlungswerte bestehen tagsüber nicht nur aus Wärmestrahlungsemissionen, sondern teilweise auch aus solarer Reflexion (Mischsignal). Die Reflexion und das Emissionsvermögen(ϵ)⁸ hängt unter anderem von den mikrophysikalischen Eigenschaften der Wolke ab. Das unbekannte Emissionsvermögen ($\epsilon_{tag} = ?$) und der Anteil der Reflexion an Kanal 3 muss demnach tagsüber beachtet werden. Nachts findet wiederum eine Angleichung mit den anderen Infrarotkanälen (Kanal 4 und 5) statt (Saunders 1986, S. 875 - 876, Saunders und Kriebel 1988, S. 129).

In Abbildung 3.1 ist exemplarisch die Tag-Testreihe mit der weiteren Einteilung in First und Second Run dargestellt. Zur Abgrenzung von wolkenkontaminierten Pixeln werden verschiedene Schwellenwert-Tests durchlaufen. Hierbei handelt es sich beispielsweise um einen einfachen Strahlungstemperatur-Grenzwert ($T_{12} < T_{min}$) mit dessen Hilfe schon ein Großteil der wolkenkontaminierten Pixel erfasst werden kann. Aber auch Schwellenwerte an Kanal-Differenzen kommen zum Einsatz (z.B. $T_{11} - T_{12} > T_{diff}$), mit deren Hilfe weitere Wolkenarten bzw. teilweise bewölkte Pixel abgegrenzt werden können. Neben Grenzwerten, die sich auf Strahlungstemperaturen oder Reflektanzen beziehen, wird auch auf Homogenität getestet, wodurch völlige von teilweise bewölkten Pixeln getrennt werden können. Hierbei läuft ein bewegter Bildausschnitt (Moving-Window) über die entsprechende Szene (z.B. 3 x 3 Pixel). Dabei wird die Standardabweichung aller Bildpunkte dieses Fensters errechnet und dem zentralen Pixel zugeordnet. So lässt sich eine Aussage über die Unterschiede der Strahlungsverhältnisse benachbarter Pixel treffen.

⁸Emissionsvermögen (ϵ) = Bruchteil der Emission eines schwarzen Körpers (vgl. Häckel 2005, S. 172).

Da die APOLLO-Einzeltests ähnlich gestaltet sind wie die der SAFNWC-Wolkenprodukte, wird an dieser Stelle nicht detaillierter darauf eingegangen. Die genauere Beschreibung verwendbarer Einzeltests geschieht aus diesem Grund anhand der SAFNWC-Testsequenzen in Kapitel 3.2. Zwar lässt sich die APOLLO-Wolkenmaske auf METEOSAT-Daten anwenden (die Wellenlängenbereiche der NOAA-AVHRR Kanäle sind in den MSG-Kanälen enthalten, vgl. Tabelle 3.1 und 4.1), bei den SAFNWC-Wolkenprodukten handelt es sich jedoch um eine Optimierung speziell auf Daten des METEOSAT-Sensors. Eine ausführlich Beschreibung aller von APOLLO verwendeten Einzeltests lässt sich bei Saunders und Kriebel (1988) und Kriebel et al. (2003) nachschlagen.

3.2 SAFNWC

Im Rahmen von EUMETSAT (Europäische Organisation zur Nutzung meteorologischer Satelliten und Betreiber der METEOSAT-Satelliten) gibt es mehrere Einrichtungen, die sich mit der Erstellung bestimmter Produkte aus Satellitendaten beschäftigen. Es handelt sich um die sogenannten Satellite Application Facilities, kurz SAFs. Eine Vielzahl dieser Einrichtungen, jeweils zusammengestellt aus verschiedenen Expertengruppen der einzelnen Mitgliedstaaten, setzen sich mit unterschiedlichsten Fragestellungen auseinander. Diese reichen von rein meteorologischen Fragestellungen (Numerical Weather Prediction, NWP-SAF) bis zu Untersuchungen der Landoberfläche (Land Surface Analysis, LSA-SAF) oder der chemischen Zusammensetzung der Atmosphäre (Ozone and Atmospheric Chemistry Monitoring, O3M-SAF). Auch klimatologische, ozeanographische oder hydrologische Analysen (Climate Monitoring, CM-SAF; Ocean and Sea Ice, OSI-SAF; Hydrology and Water Management, H-SAF) finden ihre Berücksichtigung⁹.

Innerhalb der Satellite Application Facility on Support to Nowcasting and Very Short Range Forecasting (SAFNWC) geht es um die Entwicklung eines Softwarepaketes zur Erstellung meteorologischer Satellitenprodukte. Beteiligt sind hierbei der spanische (AEMET), der französische (Météo-France), der schwedische (SMHI) und der österreichische Wetterdienst (ZAMG). Die Software besteht seit Juni 2004 und ist im Rahmen von EUMETSAT nutzbar. Mit ihrer Hilfe können 12 verschiedene Produkte aus METEOSAT-Daten erstellt werden. Neben den in dieser Arbeit genutzten Wolkenprodukten lassen sich weitere Produkte über Niederschlag, Wind oder Luftmasseigenschaften generieren. Eine Zusammenstellung aller Produkte ist nachfolgend aufgeführt¹⁰.

⁹Weitere Details unter <http://www.eumetsat.int/Home/Main/Satellites/GroundNetwork/ApplicationGroundSegment/SAFs/SAFProjects> [12.10.2010].

¹⁰Siehe auch unter <http://www.nwcsaf.org> [12.10.2010].

- *Cloud Products* - Cloud Mask, Cloud Type, Cloud Top Temperature and Height
- *Precipitation and Convection Products* - Precipitating Clouds, Convective Rainfall Rate, Rapid Development Thunderstorms
- *Clear Air Products* - Total Precipitable Water, Layer Precipitable Water, Stability Analysis Imagery
- *Conceptual Model and Wind Products* - Air Mass Analysis, Automatic Satellite Image Interpretation, High Resolution Winds

Bezogen auf Wolkendetektion bzw. Wolkenbedeckung sind jedoch in erster Linie die beiden Produkte Cloud Mask (CMA, später auch SAFNWC-Cloudmask) und Cloud Type (CT, SAFNWC-Cloudtype) von besonderem Interesse. Ähnlich wie bei APOLLO (vgl. Kapitel 3.1) basieren die Algorithmen von CMA und CT ebenfalls auf multispektralen Schwellenwertverfahren. Auch hier liegt das Bestreben darin, erst wolkenkontaminierte von wolkenfreien Pixeln anhand einer Maske zu unterscheiden, um dann in einem nächsten Schritt weitere Wolkenarten bzw. Wolkenhöhen zu unterscheiden. Da die Software speziell auf METEOSAT-Daten angepasst ist, dienen die Kanäle dieses Sensors im HRIT-Datenformat (High Rate Information Transmission, vgl. Kapitel 4 und Tabelle 4.1) als Input. Für die Erstellung sind jedoch nicht alle Kanäle zwingend erforderlich. Zudem können numerische Wettervorhersagemodell-Daten (Numerical Weather Prediction, NWP) als zusätzliche Information genutzt werden, um die dynamische Schwellenwertbestimmung zu verbessern. Weitere Daten, die zur Berechnung notwendig sind, wie Land-Meer-Verteilung, Höhenprofile und klimatologische Daten sind in der Software implementiert und fließen automatisch in die Berechnungen ein. Stehen beispielsweise keine NWP-Daten zur Verfügung werden klimatologische Daten zur Bestimmung der dynamischen Schwellenwerte herangezogen, verbunden mit einer eventuell geringeren Genauigkeit. Tabelle 3.2 zeigt einen Überblick aller für die Berechnung der Produkte notwendigen Daten (Dybbroe et al. 2000, Dybbroe 2001, Dybbroe et al. 2002, Dybbroe et al. 2004, Dybbroe et al. 2005, Derrien und Le Gléau 2005a, Derrien und Le Gléau 2005b, SAFNWC 2007a, SAFNWC 2007c, SAFNWC 2007e, SAFNWC 2007g, SAFNWC 2007k, SAFNWC 2007m, SAFNWC 2008b).

3.2.1 Cloud Mask

Um die Wolkemaske erstellen zu können, wird für jedes Pixel der entsprechenden METEOSAT-Szene eine Kombination von Schwellenwert-Tests herangezogen. Vor der Wolkenabtrennung erfolgt eine Abgrenzung schnee- und eisbehafteter Pixel. Zudem gibt es eine Unterscheidung

Tab. 3.2: Liste der Input-Daten für SAFNWC-Cloudmask (CMa) und -Cloudtype (CT). *Quelle: Derrien und Le Gléau 2005a*

Input-Daten für CMa und CT Algorithmen	
Satellitendaten	R _{0,6} (VIS006) - zwingend erforderlich
- Reflektanz (%)	R _{0,8} (VIS008)
- Strahlungstemperatur (K)	R _{1,6} (NIR016)
	T _{3,9} (IR039) - zwingend erforderlich
	T _{8,7} (IR087)
	T _{10,8} (IR108) - zwingend erforderlich
	T _{12,0} (IR120) - zwingend erforderlich
NWP-Parameter	Oberflächentemperatur
nicht zwingend erforderlich	Lufttemperatur bei 950 hPa (alternativ bei 925 hPa), 850 hPa, 700 hPa, 500 hPa und Tropopauseniveau Wasserdampfgehalt der Atmosphäre Höhe des NWP-Modell-Gitterpunktes (alternativ Oberflächengeopotential)
Atlas und Klimatologie	Land/Meer Verteilung Höhenmodell Minimum Sea Surface Temperature (SST)* Mittlere 0,6 μ m Oberflächenreflektanz (nur über Land)* Integrierter atmosphärischer Wasserdampfgehalt* Durchschnittliche Lufttemperatur bei 1000 hPa, 850 hPa, 700 hPa und 500 hPa* * Klimatologie auf Basis von Monaten

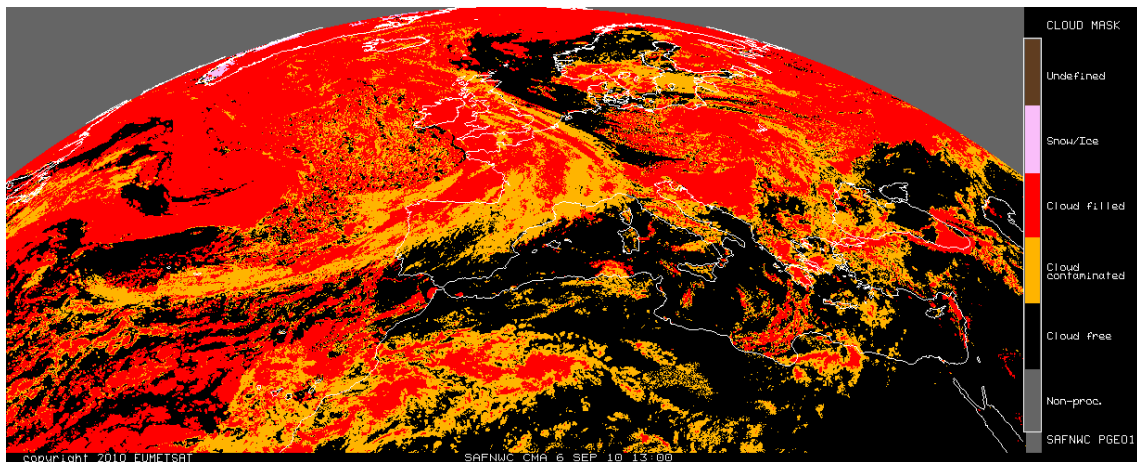


Abb. 3.2: SAFNWC-Cloudmask (CMA), Teilausschnitt Europa, 06.09.2010 13:00 UTC. *Quelle:* <http://www.nwcsaf.org> [12.10.2010]

der Testreihen über Land- und Meeresflächen, kombiniert mit der Berücksichtigung des Sonnenstandes in Bezug auf die Auswahl der Einzeltests. Man unterscheidet zwischen Nacht (Sonnenhöhe $< -3^\circ$), Dämmerung ($-3^\circ < \text{Sonnenhöhe} < 10^\circ$) und Tag (Sonnenhöhe $> 10^\circ$). Gleichzeitig werden Spiegeleffekte (Sunlint) über Meeresflächen berücksichtigt und isolierte Pixel innerhalb der entsprechenden Szene reklassifiziert (Derrien und Le Gléau 2005a).

Die Bestimmung der dynamischen Schwellenwerte erfolgt über Lookup-Table (wenn vorhanden in Kombination mit NWP-Parametern) in Abhängigkeit von der Betrachtungsgeometrie (Sonnen- und Satellitenbetrachtungswinkel). Solche Lookup-Table wurden für die thermalen Kanäle bzw. Kanal-Differenzen anhand von Strahlungstransfermodellen (RTTOV-6) und mit Hilfe von 3000 Radiosondenmessungen abgeleitet. Für die Berechnung der solaren Kanäle werden simulierte Top of the Atmosphere (TOA) Oberflächenreflektanzen herangezogen. Auch hier kommen Lookup-Table zum Einsatz, die ein großes Spektrum an Wasserdampf- und Ozonmengen, sowie bestimmte Betrachtungswinkel berücksichtigen. Eine detaillierte Beschreibung der Schwellenwertbestimmungen findet sich bei Derrien und Le Gléau (2005a), S. 4717 - 4720.

Der CMA-Output erfolgt anhand des HDF5-Dateiformates (vgl. SAFNWC 2007i, SAFNWC 2008a). Neben den eigentlichen Daten der Klassifizierung erhält man zusätzliche Informationen über die Beleuchtungssituation, die jeweilige Verfügbarkeit der Input-Daten (SEVIRI, NWP) und die Qualität der Prozessierung. Die Wolkenmaske besteht aus sechs Kategorien, die mit 3 Bit kodiert sind (vgl. Tabelle 3.3). Abbildung 3.2 zeigt die SAFNWC-Cloudmask (CMA) für den europäischen Raum (Derrien und Le Gléau 2005a, S. 4710).

Im Folgenden werden die verschiedenen Einzeltests und ihr jeweiliger Hintergrund genauer beschrieben (vgl. hierzu auch Derrien und Le Gléau 2005a, S. 4713 - 4717). Da es in der vorliegenden Arbeit um die Landfläche von Deutschland geht, bezieht sich die Beschreibung im Wesentlichen auf die für Landflächen vorgesehenen Schwellenwert-Tests. Die Grundlagen, auf

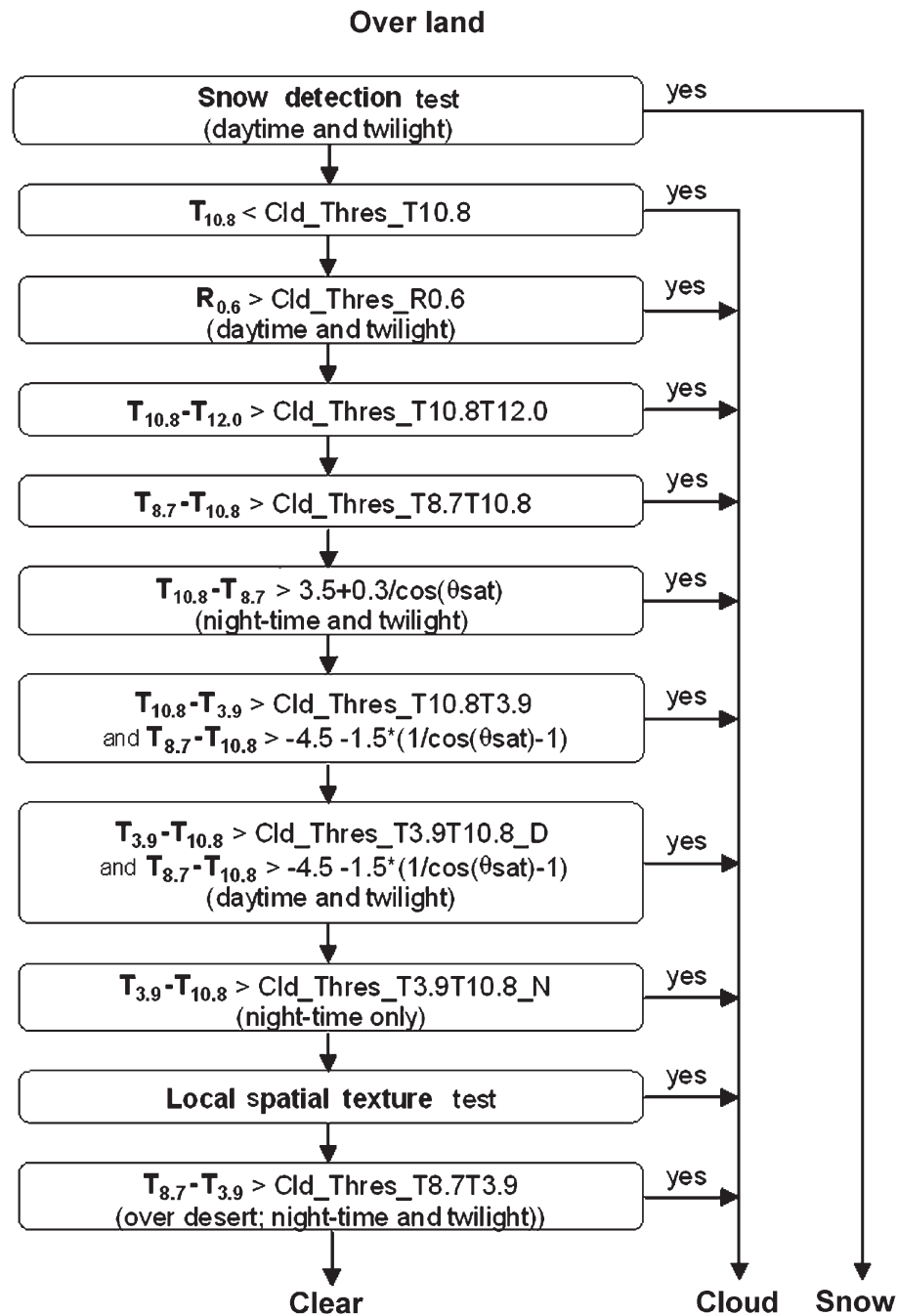


Abb. 3.3: Testsequenz der SAFNWC-Cloudmask (CMA) über Landflächen. Die Beschreibung der Einzeltests befindet sich im Text. θ_{sat} = Satelliten-Zenitwinkel. *Quelle: Derrien und Le Gléau 2005a*

denen die Tests basieren, sind über Meeresflächen jedoch ähnlich. Die gesamte Testreihe für Landoberflächen ist in Abbildung 3.3 dargestellt.

Schwellenwert-Test bei $T_{10,8}$ (Nacht und Tag)

Mit diesem einfachen Schwellenwertansatz können Großteile der Bewölkung abgegrenzt werden. Geringere Strahlungstemperaturwerte, als der entsprechende Schwellenwert, verweisen auf den Einfluss von Wolken. Die Festlegung des Schwellenwertes erfolgt wie zuvor beschrieben. Wolkenkontaminierte Pixel, die sich nicht durch diesen Test erfassen lassen, werden gegebenenfalls durch nachfolgende Tests erfasst.

Test bezogen auf $T_{10,8} - T_{12,0}$ (Nacht und Tag)

Mit Hilfe dieser Strahlungstemperatur-Differenz lassen sich hohe und dünne Cirrus-Wolken detektieren. Die Werte sollten im Vergleich zu wolkenfreien Flächen höher sein. Der Test lässt sich bei allen Beleuchtungsverhältnissen und allen Oberflächen durchführen.

Test bezogen auf $T_{8,7} - T_{10,8}$ (Nacht und Tag)

Auch mit dieser Differenz lassen sich in ähnlicher Weise wie mit der vorherigen Differenz hohe und dünne Cirrus-Wolken abgrenzen.

Test bezogen auf $T_{3,9} - T_{10,8}$ (Nacht)

Dieser Test kommt nachts zum Einsatz, um hohe semitransparente Wolken detektieren zu können. Der Strahlungstemperaturanteil der wärmeren Erdoberfläche ist bei $3,9 \mu\text{m}$ gewöhnlich höher als bei $10,8 \mu\text{m}$ (Durchlässigkeit bei Eiswolken), wodurch sich hohe semitransparente Wolken über relativ warmen Flächen erfassen lassen. Aufgrund des speziellen Verhaltens im Bereich von $3,9 \mu\text{m}$ am Tag (vgl. Seite 53) wird dieser Test nur nachts angewendet.

Test bezogen auf $T_{10,8} - T_{3,9}$ (Nacht)

Mit dieser Differenz können nachts niedrige Wasserwolken detektiert werden. Die Emissivität von Wasserwolken ist bei $3,9 \mu\text{m}$ niedriger als bei $10,8 \mu\text{m}$.

Test bezogen auf $T_{8,7} - T_{3,9}$ (Nacht)

Dieser Test lässt sich für die Erkennung niedriger Wasserwolken über Wüsten verwenden. Normalerweise bezieht man sich hierfür auf die Differenz $T_{10,8} - T_{3,9}$, jedoch funktioniert dies nachts über Wüstenflächen nur schlecht, weil der Kontrast zwischen niedrigen Wolken und

Wüste nicht groß genug ist. Wüsten weisen sowohl bei $3,9 \mu\text{m}$, als auch bei $8,7 \mu\text{m}$ eine niedrige Emissivität auf. Niedrige Wasserwolken dagegen nur bei $3,9 \mu\text{m}$, und nicht bei $8,7 \mu\text{m}$.

Test bezogen auf $T_{10,8} - T_{8,7}$ (Nacht und Dämmerung)

Bei der Betrachtung von Vegetationsflächen mit einem hohen Satelliten-Zenitwinkel lassen sich nachts und bei niedriger Sonnenhöhe mit dieser Differenz tiefliegende Wolken abgrenzen. Im Vergleich zu der Differenz $T_{10,8} - T_{3,9}$, welche eigentlich für die Erkennung von niedrigen Wolken genutzt wird, scheint die Differenz $T_{10,8} - T_{8,7}$ unter den beschriebenen Gegebenheiten besser zu funktionieren. Wolkenfreie Flächen ergeben bei großen Satelliten-Zenitwinkeln für die Differenz $T_{10,8} - T_{3,9}$ ebenso hohe Werte, wie wolkenkontaminierte Flächen.

Test bezogen auf $T_{3,9} - T_{10,8}$ (Tag und Dämmerung)

Durch die Umkehr der Strahlungsverhältnisse bei $3,9 \mu\text{m}$ im Zusammenhang mit niedrigen Wolken am Tag, kann diese Differenz tagsüber zur Detektion niedriger Wasserwolken eingesetzt werden.

Test bezogen auf $R_{0,6}$ (Tag)

Tagsüber können die Reflektanzen der sichtbaren Kanäle, sowie des nahen Infrarots zur Erkennung von Wolken eingesetzt werden. Ist der Wert an einem Pixel „heller“ als die wolkenfreie Erdoberfläche, kann man dieses als wolkenkontaminiert einstufen. Über Landflächen bezieht man sich hierbei gewöhnlich nur auf die Reflektanz $R_{0,6}$. Der Bereich des Nahen Infrarots kommt ausschließlich über Meeresflächen zum Einsatz, auch um den Einfluss von Spiegeleffekten zu verhindern.

Schnee- und Eisflächenerkennung (Tag)

Vor der eigentlichen Wolkendetektion steht ein Schnee- und Eiserkennungstest. Fällt dieser positiv aus, findet keine weitere Wolkenerkennung statt. Schnee- und Eisflächen verhalten sich, bei auf Grenzwerten basierenden Erkennungsverfahren, ähnlich wie Wolken. Sie stellen ebenfalls sehr kalte oder helle Flächen dar, wodurch es vor allem zu einer falschen Abschätzung im Vergleich mit niedrigen Wolken kommen kann. Um Schneeflächen von Wolken differenzieren zu können, betrachtet man vornehmlich die Reflektanz bei $1,6 \mu\text{m}$. Schnee- und Eisflächen weisen hier niedrigere Werte auf, weshalb eine Abgrenzung nur tagsüber möglich ist. Eine genaue Beschreibung des Testverfahrens ist bei Derrien und Le Gléau (2005a), S. 4716 - 4717 aufgeführt.

Tab. 3.3: Output SAFNWC-Cloudmask (CMA) und -Cloudtype (CT). *Quelle: Derrien und Le Gléau 2005a*

CMA-Output	CT-Output
0 - non-processed	0 - non-processed
1 - cloud free	1 - cloud free land
2 - cloud contaminated	2 - cloud free sea
3 - cloud filled	3 - land contaminated by snow
4 - snow/ice contaminated	4 - sea contaminated by snow/ice
5 - undefined	5 - very low and cumuliform clouds
	6 - very low and stratiform clouds
	7 - low and cumuliform clouds
	8 - low and stratiform clouds
	9 - medium and cumuliform clouds
	10 - medium and stratiform clouds
	11 - high opaque and cumuliform clouds
	12 - high opaque and stratiform clouds
	13 - very high opaque and cumuliform clouds
	14 - very high opaque and stratiform clouds
	15 - high semitransparent thin clouds
	16 - high semitransparent meanly thick clouds
	17 - high semitransparent thick clouds
	18 - high semitransparent above low or medium clouds
	19 - fractional clouds (sub-pixel water clouds)
	20 - undefined (undefined by CMA)

Lokale Homogenität

Anhand des Vergleichs mit Nachbarpixeln erfolgt abschließend eine Überprüfung auf lokale Homogenität. Über Landflächen dient dies vor allem zur Abgrenzung teilweise bewölkter Pixel bzw. Wolkenkanten innerhalb eines Pixels. Die Vorgehensweise bezieht sich bei Landflächen auf die Strahlungstemperatur $T_{10,8}$ und die Strahlungstemperatur-Differenz $T_{10,8} - T_{3,9}$. Über die Werte dieser Kanäle bzw. Kanal-Differenzen wird die Standardabweichung benachbarter Pixel berechnet und mit Grenzwerten (1,0 K bzw. 2,0 K am Tag) verglichen. Liegt entsprechend eine verstärkte Heterogenität vor, kann von teilweise bewölkten Pixeln ausgegangen werden (Derrien und Le Gléau 2005a, S. 4717).

3.2.2 Cloud Type

Für die weitere Differenzierung von Wolkenarten und Wolkenhöhen werden die zuvor detektierten Wolken weiter klassifiziert, vor allem um meteorologische Auswirkungen im Rahmen von kurzfristigen Vorhersagen noch besser abschätzen zu können. Auch hierbei wird ein mul-

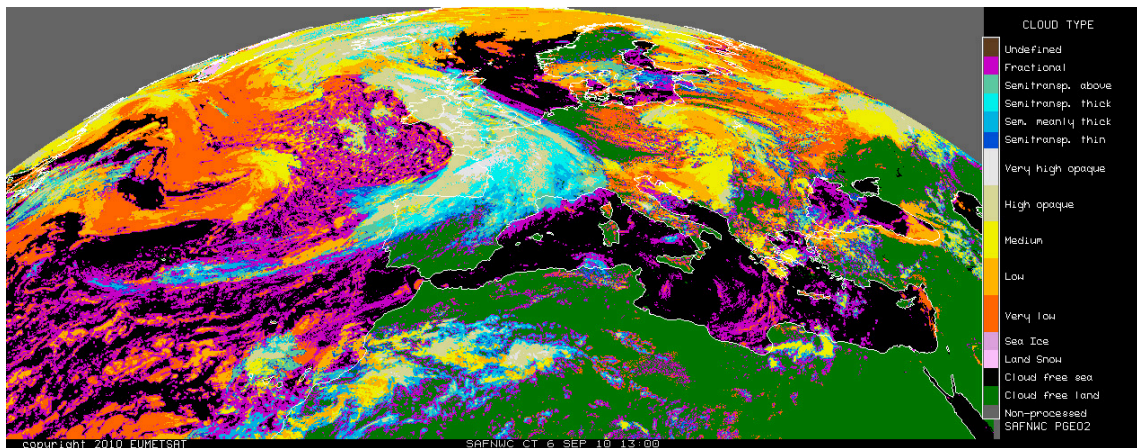


Abb. 3.4: SAFNWC-Cloudtype (CT), Teilausschnitt Europa, 06.09.2010 13:00 UTC. *Quelle:* <http://www.nwcsaf.org> [12.10.2010]

tispektraler Schwellenwertansatz durchgeführt. Die Anwendung erfolgt an jedem Pixel und ist abhängig von Beleuchtungs- und Betrachtungsgeometrie. Wie bei der SAFNWC-Cloudmask (C_{Ma}) beziehen sich die Schwellenwerte auf den vertikalen Zustand der Atmosphäre am entsprechenden Punkt, und die Berechnung erfolgt wie auf Seite 57 beschrieben. Nachfolgend wird die Vorgehensweise der SAFNWC-Cloudtype-Testreihe kurz vorgestellt. Eine ausführliche Beschreibung findet sich bei Derrien und Le Gléau (2005a), S. 4720 - 4721.

- Trennung zwischen teilweise bewölkten Pixeln bzw. Pixeln mit hoch semitransparenten Wolken, von solchen mit niedrigen, mittleren oder hohen kompakten Wolken. Dies wird nachts durch die Betrachtung der Differenzen $T_{10,8} - T_{12,0}$ und $T_{3,9} - T_{10,8}$ erreicht. Tagsüber erfolgt die Trennung durch Werte der Reflektanz $R_{0,6}$, sowie mit Hilfe des Homogenitäts-Test bei $T_{10,8}$ und $R_{0,6}$.
- Die Trennung von teilweise bewölkten Pixeln und hohen semitransparenten Wolken wird durch die Verwendung der Differenz $T_{8,7} - T_{10,8}$ erreicht.
- Die weitere Einteilung erfolgt durch den Vergleich von $T_{10,8}$ mit Lufttemperaturen auf unterschiedlichen Luftdruckniveaus (850 hPa, 700 hPa, 500 hPa und Tropopause).
- Eine Unterscheidung zwischen kumuliformen und stratiformen Wolken ist zurzeit noch nicht möglich.

Ähnlich wie bei der SAFNWC-Cloudmask (C_{Ma}), erfolgt der CT-Output über eine HDF5-Datei. Man erhält 21 mit 5 Bit kodierte Wolken-Kategorien (vgl. Tabelle 3.3). Abbildung 3.4 zeigt beispielhaft ein Ergebnis von SAFNWC-Cloudtype (CT) (vgl. auch Derrien und Le Gléau 2005a, S. 4710).

4 Datengrundlage METEOSAT-SEVIRI

Da der Ansatz der Wolkenerkennung auf den Daten von METEOSAT-SEVIRI basiert, wird in den nächsten Abschnitten ausführlich auf diesen Sensor und seine Funktionalität eingegangen. Die genauen Hintergründe über Vorverarbeitung und Verbreitung der Satellitendaten werden beschrieben und dargelegt. Dies betrifft im Wesentlichen die Kalibrierungsverfahren, mit deren Hilfe Strahlungstemperaturen und Reflektanzen der jeweiligen Kanäle abgeschätzt werden.

4.1 METEOSAT Second Generation

METEOSAT (Meteorological Satellite) ist die Bezeichnung für geostationäre Wettersatelliten, die von der europäischen Organisation EUMETSAT in enger Zusammenarbeit mit der European Space Agency (ESA) betrieben werden. Mit dem Start von METEOSAT-1 (1977) und den folgenden METEOSAT-Satelliten werden nunmehr seit fast 30 Jahren Wetterinformationen bereitgestellt, die zur Wettervorhersage und zu weiteren Forschungszwecken genutzt werden können. Die geostationäre Position¹¹ am Nullmeridian (0° geographische Länge), mit einer Entfernung zum Satelliten von 36000 km über dem Äquator, erlaubt eine optimale Erfassung von Wetterphänomenen über Afrika, Europa und Teilen des Atlantiks.

Auf METEOSAT-1 folgten, bis METEOSAT-7 (1997), sechs weitere Satelliten. Diese geostationären Satelliten, die auch als Satelliten der ersten Generation bezeichnet werden, besaßen (bzw. besitzen) einen Sensor, der in 3 Kanälen Strahlung erfassen kann. Hierbei handelt es sich um Kanal 1 ($0,45 \mu\text{m} - 1,0 \mu\text{m}$) im sichtbaren Bereich, Kanal 2 ($5,7 \mu\text{m} - 7,1 \mu\text{m}$) im infraroten Wasserdampf-Absorptionsbereich und Kanal 3 ($10,5 \mu\text{m} - 12,5 \mu\text{m}$) im thermischen Infrarotbereich. Diese Reihe von METEOSAT-Satelliten nahmen in halbstündigen Abständen Bilder auf. Mit der Einführung der zweiten Generation (METEOSAT Second Generation, MSG), also der Satelliten METEOSAT-8 und METEOSAT-9, wurde METEOSAT-7 versetzt und befindet sich jetzt an der Position 57° Ost. Durch diese Maßnahme wurde es möglich, Teile Asiens und des Indischen Ozeans besser zu erfassen.

Aufgrund des technologischen Fortschritts begann man schon in den frühen 1990er Jahren mit Überlegungen für eine neue Generation an METEOSAT-Satelliten, die eine umfangreichere spektrale Information zur Verfügung stellen sollten. Es war jedoch klar, dass eine wirkliche

¹¹Drehung des Satelliten mit der Erdrotation, so dass immer derselbe Bildausschnitt erfasst wird.

Umsetzung erst mit Ablauf der mindestens fünfjährigen Lebensdauer vorheriger METEOSAT-Satelliten beginnen konnte. Der erste Satellit zweiter Generation (METEOSAT-8 bzw. MSG-1) wurde mit dem Start von Ariane 5, am 29. August 2002 in seine Umlaufbahn gebracht. Nach einer ausführlichen Testphase, begann Anfang 2004 der operationelle Betrieb. Am 21. Dezember 2005 wurde ein weiterer MSG-Satellit (METEOSAT-9 bzw. MSG-2) in den Orbit gebracht, der ab 2006 seinen Betrieb aufnahm. Weitere Satelliten der dritten Generation sollen 2015 - 2018 folgen. Zurzeit stellt METEOSAT-9 den operationellen Hauptsatelliten dar und nur bei Problem- oder Wartungsfällen wird auf METEOSAT-8 ausgewichen. Außerhalb solcher Ausnahmesituationen wird METEOSAT-8 ($9,5^\circ$ Ost) für den sogenannten Rapid-Scan-Service (RSS) eingesetzt. Hierbei wird nur der nördliche Teil der „Aufnahmescheibe“ (ca. 15° bis 70° Nord, geographische Breite) mit Europa und Nordafrika erfasst. Dieser Umstand macht Aufnahmeintervalle von fünf Minuten möglich¹² (EUMETSAT 2000, EUMETSAT 2001, EUMETSAT 2004a, EUMETSAT 2005a, EUMETSAT 2005b, EUMETSAT 2005d).

4.1.1 SEVIRI

Der enorme Vorteil von METEOSAT Second Generation, gegenüber den Satelliten der ersten Generation, ist die höhere spektrale Information, verbunden mit einer ebenso höheren zeitlichen Auflösung von 15 Minuten. Der an Bord befindliche Spinning Enhanced Visible and InfraRed Imager (SEVIRI) bietet Aufnahmen in 12 Kanälen, wovon 11 den sichtbaren und infraroten Wellenlängenbereich abdecken. Ein zusätzlicher Kanal erfasst räumlich hochauflösende Bilddaten im sichtbaren Bereich (High Resolution Visible, HRV). Tabelle 4.1 zeigt die Zusammensetzung der Kanäle mit ihren jeweiligen Wellenlängen. Mit Hilfe von SEVIRI wird die vollständige „Erdscheibe“ in einem fünfzehnminütigen Zyklus erfasst. Das Aufnahmeverfahren wird durch Umdrehungen des Satelliten realisiert. Hundert Umdrehungen pro Minuten sorgen für einen zeilenweisen Ost-West-Scan, schrittweise von Süden nach Norden. Im Bereich des Sub-Satelliten-Punktes beträgt das Abtastintervall 3 km für die 11 sichtbaren und thermalen Kanäle, und 1 km für den HRV-Kanal. Jeder der 11

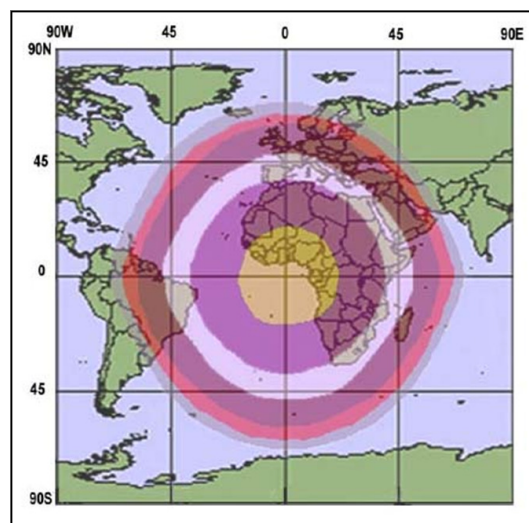


Abb. 4.1: Aufnahme der „Erdscheibe“. Abnehmende Auflösung vom Sub-Satelliten-Punkt nach außen. Die Pixel-Auflösung beträgt im inneren Kreis 3 km, gefolgt von 4, 5, 6, 8 und 11 km (äußerer Kreis). *Quelle: EUMETSAT 2001*

¹²Detaillierte Informationen unter <http://www.eumetsat.int> [12.01.2011].

Tab. 4.1: METEOSAT Second Generation - Kanalübersicht. *Quelle: EUMETSAT 2001*

Kanal	Bezeichnung	Zentrale Wellenlänge	Untersuchungsgegenstand
1	VIS006 (R _{0,6})	0,635 μm	surface, clouds, wind fields
2	VIS008 (R _{0,8})	0,81 μm	surface, clouds, wind fields
3	NIR016 (R _{1,6})	1,64 μm	surface, cloud phase
4	IR039 (T _{3,9})	3,90 μm	surface, clouds, wind fields
5	WV062 (T _{6,2})	6,25 μm	water vapor, high level clouds, atmospheric instability
6	WV073 (T _{7,3})	7,35 μm	water vapor, atmospheric instability
7	IR087 (T _{8,7})	8,70 μm	surface, clouds, atmospheric instability
8	IR097 (T _{9,7})	9,66 μm	ozone
9	IR108 (T _{10,8})	10,80 μm	surface, clouds, wind fields, atmospheric instability
10	IR120 (T _{12,0})	12,00 μm	surface, clouds, atmospheric instability
11	IR134 (T _{13,4})	13,40 μm	cirrus cloud height, atmospheric instability
12	HRV	0,4 - 1,1 μm	surface, clouds

Kanäle besitzt 3 Detektoren, so dass aus 1250 Abtastlinien 3750 Bildlinien werden (vgl. Abbildung 4.1 und 4.2). Die vollständige Abtastung der „Scheibe“ benötigt 12 Minuten, gefolgt von einer dreiminütigen Kalibrierungsphase der Infrarotkanäle, einhergehend mit der Ausrichtung des Aufnahmespiegels zum Startpunkt. Beim HRV-Kanal erfolgt der Scan mit 9 Detektoren pro Abtastlinie, resultierend in einem Bild von 11136 x 5568 Pixeln. Die Bilder dieses Kanals bestehen jedoch letztendlich nicht aus der vollständigen Scheibe, sondern, je nach Sonnenstand, nur aus Europa und einzelnen Teilen des afrikanischen Kontinents. Neben SEVIRI existieren weitere Instrumente an Bord der MSG-Satelliten, wie z.B. GERB (Geostationary Earth Radiation Budget), mit dem eine Abschätzung der globalen Strahlungsbilanz ermöglicht wird. Andere Zusatzgeräte dienen der Kommunikation und eventuellen Rettungsmaßnahmen (EUMETSAT 2000, EUMETSAT 2001).

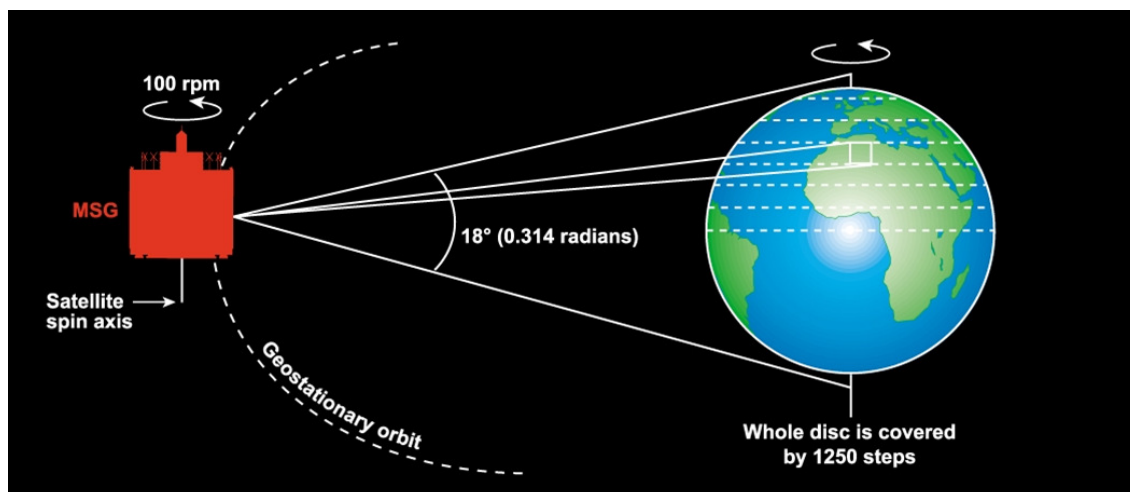


Abb. 4.2: MSG-Aufnahmeverfahren. *Quelle: EUMETSAT 2001*

4.2 Vorverarbeitung und Übertragung der METEOSAT-Daten

Da die Daten so schnell wie möglich die jeweiligen Anwender erreichen sollen, enthält der Satellit einen Transponder, von dem die Daten zu verschiedenen Empfangsstationen im Satellitenblickfeld übertragen werden können. Für Empfang, Verarbeitung und Weiterverteilung der Daten besteht am Boden eine umfangreiche Infrastruktur (MSG Ground Segment). Der Kern dieser Infrastruktur stellt das EUMETSAT-Kontrollzentrum in Darmstadt dar, welches mit mehreren Hauptempfangsstationen (Primary Ground Stations) verbunden ist. Das heißt, die von SEVIRI aufgenommenen Daten erreichen in digitaler Form die Empfangsstationen und werden zum Kontrollzentrum weitergeleitet. Hier erfolgen die ersten Vorverarbeitungsschritte. Anhand geometrischer und radiometrischer Korrekturen werden aus sogenannten Level 1.0 Rohdaten, Level 1.5 Daten erzeugt (EUMETSAT 2005a, EUMETSAT 2006).

Die weitere Verteilung dieser Level 1.5 Daten zu den verschiedenen Anwendern erfolgt über das MSG-Kommunikationsmodul (funktioniert bei MSG-1 jedoch nicht), oder via Digital Video Broadcast (DVB über EUMETCast¹³) mittels kommerzieller Satelliten, wie HotBird bzw. EuroBird. Die Übertragung geschieht auf zwei Wegen, entweder durch eine Low Rate Information Transmission (LRIT, nur 5 SEVIRI-Kanäle verlustbehaftet alle 30 Minuten), oder durch eine High Rate Information Transmission (HRIT, verlustfrei alle 12 SEVIRI-Kanäle mit einer Wiederholrate von 15 Minuten, siehe auch Wolf 1999).

Jeder, der hierfür eine entsprechende Empfangsstation eingerichtet hat, kann die Daten in Absprache mit EUMETSAT empfangen und nutzen. Man unterscheidet bei den Empfangs-

¹³<http://www.eumetsat.int/Home/Main/DataAccess/EUMETCast> [12.01.2011]

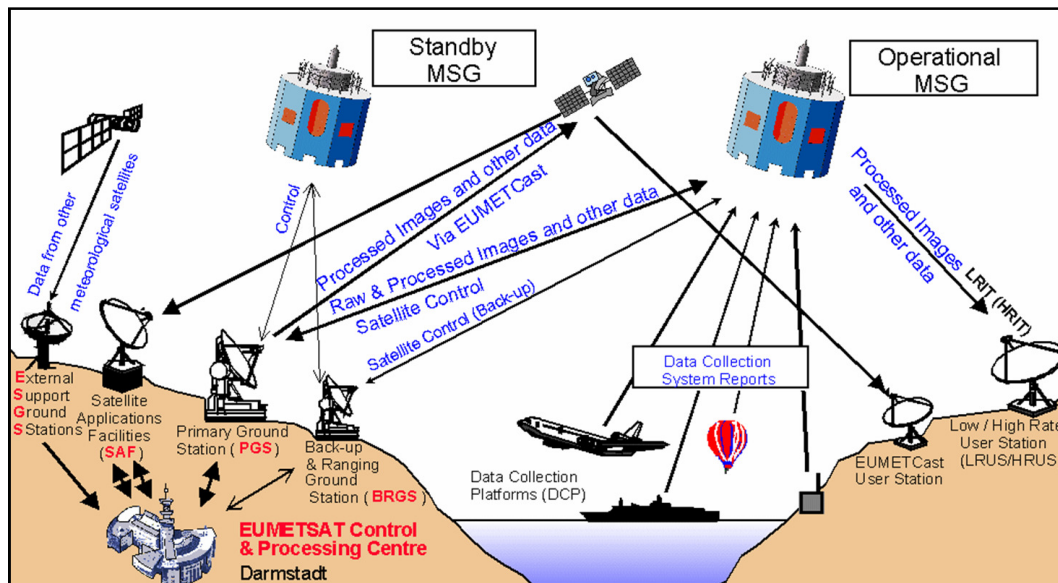


Abb. 4.3: Datenübertragungswege von METEOSAT Second Generation. Quelle: EUMETSAT 2000

stationen zwischen High-Rate User Station für HRIT-Daten und Low-Rate User Station für LRIT-Daten. Es besteht aber die Möglichkeit, innerhalb einer High-Rate User Station den LRIT-Datenstrom kombiniert mit dem HRIT-Datenstrom zu empfangen. Zusätzlich können den METEOSAT-Nutzern innerhalb der Datenströme Daten anderer geostationärer Satelliten, z.B. aus dem Pazifik- und Westatlantikraum, zur Verfügung gestellt werden. Die von EUMETSAT in Verbindung mit anderen Instituten erstellten meteorologischen und klimatologischen Produkte, sowie die von sogenannten Data Collection Platforms erfassten zusätzlichen Umweltdaten, werden ebenfalls über das MSG-Datenfernnetz verbreitet bzw. weitergeleitet. Die einzelnen Übertragungswege zwischen Satellit und Bodeninfrastruktur sind in Abbildung 4.3 dargestellt (EUMETSAT 2000, EUMETSAT 2001).

4.3 Kalibrierungsverfahren

4.3.1 Kalibrierung der thermalen Kanäle

Die Kalibrierung der infraroten bzw. thermalen Kanäle erfolgt mit jedem Aufnahmezyklus durch einen an Bord befindlichen Schwarzkörper. Um die Strahlungsspanne abzudecken, kommt bei der Kalibrierung eine kalte und eine warme Quelle zum Einsatz. Der kalte Endpunkt wird durch den Blick ins All bestimmt, während die warme Seite durch Betrachtung des Schwarzkörpers festgelegt wird. Diese Vorgehensweise ist möglich, da die von SEVIRI ausgegebenen digitalen Pixelwerte in einer annähernd linearen Beziehung zu den gemessenen Strahlungswerten stehen (vgl. Gleichung 4.1) (Schmetz et al. 2002).

$$C(L) = gR(\lambda, T) + C_0 \quad (4.1)$$

$C(L)$ = Digitale Pixelwerte (10 Bit = 0 - 1023)

g = Faktor (bzw. Inverse = Kalibrierungskoeffizient)

$R(\lambda, T)$ = Gemessene Strahlung [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

λ = Wellenlänge

T = Effektive Schwarzkörpertemperatur einer observierten Szene [K]

C_0 = Offset

Quelle: Schmetz et al. 2002

Durch den Detektor verursachte nicht-lineare Effekte werden mit entsprechenden Kalibrierungskoeffizienten behoben. Zur Umrechnung der SEVIRI-Strahlungswerte in Strahlungstemperaturen, geht man bei EUMETSAT unter Einbeziehung der Planck-Funktion von einer analytischen Beziehung zwischen *äquivalenter Strahlungstemperatur* und gemessenen Strahlungswerten aus. Unter der äquivalenten Strahlungstemperatur versteht man die Temperatur, mit der ein schwarzer Körper im gleichen Umfang der gemessenen Strahlungswerte emittieren würde. Aufgrund des unterschiedlichen Emissionsvermögens der jeweilig erfassten Oberflächen ist diese Annahme bei der Umrechnung erforderlich. Man erhält folgende Gleichung.

$$T_b = \left(\frac{c_2 \nu_c}{\log(1 + c_1 \nu_c^3 / R)} - B \right) / A \quad (4.2)$$

$R(\lambda, T)$ = Gemessene Strahlung [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

T_b = Äquivalente Strahlungstemperatur [K]

ν_c = Zentrale Wellenzahl des jeweiligen Kanals [cm^{-1}]

$c_1 = 2hc^2$ mit Lichtgeschwindigkeit c und Planckkonstante h

$c_2 = hc/k$ mit Boltzmannkonstante k

$c = 299792458 \text{ ms}^{-1}$

$h = 6,62606876(52) \cdot 10^{-34} \text{ Js}$

$k = 1,3806503(24) \cdot 10^{-23} \text{ JK}^{-1}$

A und B = Korrekturkoeffizienten

Quelle: EUMETSAT 2005e

Die Korrekturkoeffizienten A und B wurden vorab für jeden Kanal bestimmt (vgl. Tabelle 4.2). Hierdurch wird für die Beziehung zwischen Strahlungswerten und Strahlungstemperaturen für den Temperaturbereich von 150 - 350 K eine Genauigkeit von 0,05 K erreicht. Insgesamt ist für das Kalibrierungsverfahren jedoch eine Genauigkeit von ca. 0,5 - 1 K zu erwarten (Schmetz et al. 2002, EUMETSAT 2005e, Tjemkes 2005).

Die hier beschriebenen annähernd linearen Beziehungen sind möglich, da man sich auf eine spektrale Schwarzkörperstrahlung (Spectral Blackbody Radiance) bezieht, also die Werte der zentralen Wellenlänge des jeweiligen Kanals. Seit Mai 2008 ist man jedoch dazu übergegangen

Tab. 4.2: Werte der Korrekturkoeffizienten für die jeweiligen Infrarotkanäle. *Quelle: EUMETSAT 2005e*

Kanal	ν_c	A	B
IR039	2569,094	0,9959	3,471
WV062	1598,566	0,9963	2,219
WV072	1362,142	0,9991	0,485
IR087	1149,083	0,9996	0,181
IR097	1034,345	0,9999	0,060
IR108	930,659	0,9983	0,627
IR120	839,661	0,9988	0,397
IR134	752,381	0,9981	0,576

nicht mehr die spektrale, sondern die effektive Schwarzkörperstrahlung (Effective Blackbody Radiance) als Definition zu verwenden. Schon vor diesem Zeitpunkt bezog man sich innerhalb der ebenfalls von EUMETSAT betriebenen Meteorological Product Extraction Facility (MPEF, Teil des MSG Ground Segment) auf die effektive Strahlung, wobei über den Wellenlängenbereich des entsprechenden Kanals integriert wird. Um in diesem Zusammenhang Konsistenz zu schaffen, einigte man sich auf die allgemeine Verwendung dieser Strahlungsdefinition, auch für Level 1.5 Daten. Das gesamte Kalibrierungsverfahren bleibt demnach erhalten, es wird jedoch nicht die zentrale Wellenlänge herangezogen, sondern der Wellenlängenbereich des entsprechenden Kanals (vgl. EUMETSAT 2007a, EUMETSAT 2008a, EUMETSAT 2008b).

$$R = \frac{\int R_\nu r_\nu d\nu}{\int r_\nu d\nu} \quad (4.3)$$

$R(\lambda, T)$ = Gemessene Strahlung [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

ν = Wellenzahl [cm^{-1}]

r_ν = Spektralbereich am jeweiligen Kanal

Quelle: EUMETSAT 2007a

4.3.2 Kalibrierung der solaren Kanäle

Da für die Kalibrierung der solaren Kanäle (VIS006, VIS008, NIR016 und HRV) keine an Bord befindliche Vorrichtung vorhanden ist, wurden in einer Testphase von MSG-1 Kalibrierungskoeffizienten abgeschätzt, die im operationellen Betrieb auf die SEVIRI-Messergebnisse angewendet werden. Die Abschätzung der Koeffizienten erfolgte über Wüsten- und Meeresflächen mit relativ stabilen Strahlungsbedingungen. Als Grundlage dienten hierfür Berechnungen anhand von Strahlungstransfermodellen. Die Umrechnung von digitalen Pixelwerten in Strahlungswerte ist nachfolgend aufgeführt. Werte von I_m ändern sich mit den Kanälen und sind in Tabelle 4.3 dargestellt (vgl. EUMETSAT 2004b, S. 17, Govaerts et al. 2004).

Tab. 4.3: Verschiedene Werte von I_m zur Kalibrierung der solaren Kanäle. *Quelle: EUMETSAT 2004b*

Kanal	Zentrale Wellenlänge	I_m [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]
VIS006	0,635 μm	65,2296
VIS008	0,815 μm	73,0127
NIR016	1,640 μm	62,3715
HRV	0,750 μm	78,8952

$$R_m = c_f K + R_0 \quad (4.4)$$

R_m = Strahlung [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

c_f = Kalibrierungskoeffizient [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$ /Digital Count]

K = Digitaler Pixelwert

R_0 = Kalibrierungskonstante [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

Quelle: EUMETSAT 2004b

$$r = \frac{\pi R_m d^2(t)}{I_m \cos(\theta_s(t, \vec{x}))} \quad (4.5)$$

r = Reflexionsfaktor

$d^2(t)$ = Abstand Erde-Sonne in astronomischen Einheiten zum Zeitpunkt t

I_m = Solare Strahlung für eine astronomische Einheit [$\text{mWm}^{-2}\text{sr}^{-1}(\text{cm}^{-1})^{-1}$]

$\theta_s(t, \vec{x})$ = Sonnen-Zenit-Winkel [rad] zum Zeitpunkt t am Ort \vec{x}

Quelle: EUMETSAT 2004b

4.3.3 Geometrische Genauigkeit

Bei der von EUMETSAT durchgeführten Erstellung der Level 1.5 Daten erfolgt neben radiometrischen Abschätzungen auch eine Zuordnung von geographischen Koordinaten für jedes Pixel (vgl. auch Seite 72). Hierbei handelt es sich sowohl um Längen- als auch Breitengradangaben. Die von MSG verwendete Projektion ist eine normalisierte geostationäre Projektion. Sie beschreibt den Blick von einem virtuellen Satelliten auf eine idealisierte Erde. In dieser Hinsicht befindet sich der virtuelle Satellit in einem geostationären Orbit exakt über dem Äquator und dem Nullmeridian. Details über die Zuordnung der Koordinaten und die sich hierbei ergebende Genauigkeit sind unter EUMETSAT (2006) und EUMETSAT (2007b) zu finden.

4.4 Datenempfang und Prozessierung am Zentrum für Fernerkundung der Landoberfläche

Die Beschaffung der notwendigen Daten erfolgt mit Hilfe der institutseigenen Empfangsanlage. Dies geschieht über EUMETCast, wie in Kapitel 4.2 und Abbildung 4.3 beschrieben. Die Empfangsanlage, bestehend aus einer Empfangsantenne sowie Empfangs- und Prozessierungsrechnern, befindet sich am Zentrum für Fernerkundung der Landoberfläche (ZFL) und wird in Zusammenarbeit mit der Arbeitsgruppe Fernerkundung des geographischen Instituts der Universität Bonn betreut. Im Wesentlichen basiert die Empfangs- und Prozessierungssystem auf den Softwareprodukten der Firma VCS AG Bochum¹⁴, mit deren Hilfe verschiedene Prozessierungsschritte und die Datenarchivierung durchgeführt werden kann. Das Softwarepaket setzt sich zusammen aus (vgl. VCS 2005):

- *2met! Acquisition LRIT/HRIT DVB* - Datenempfang über DVB
- *2met! Processing Advanced* - Prozessierung
- *2met! Dispatch* - Datenverteilung
- *2met! Limit* - Datenlimit innerhalb bestimmter Ordner
- *2met! Vision* - Visualisierung
- *2met! VCS Log* - Erstellung von Log-Dateien
- *2met! XPIF* - Internes Datenformat

Insgesamt besteht das Empfangs- und Prozessierungssystem aus einem Empfangsrechner (MSG-DVB), einem Prozessierungsrechner (MSG-Processing) und drei Archivierungsservern (vgl. Abbildung 4.4). Hinzu kommt die Anbindung des SAFNWC-Prozessierungsrechners, auf den in Kapitel 4.5 genauer eingegangen wird. Die Daten erreichen den Empfangsrechner im komprimierten HRIT-Format (alle 12 Kanäle in voller MSG Auflösung, vgl. Kapitel 4.2) und werden direkt zum Prozessierungsrechner weitergeleitet. Gleichzeitig werden über EUMETCast zusätzliche Daten empfangen und direkt zur Archivierung bereitgestellt (vgl. Tabelle 4.4). In diesem Fall handelt es sich um Daten der Land Surface Analysis SAF (LSASAF¹⁵) mit Oberflächentemperaturen von Europa und Afrika, sowie um synoptische Daten im Rahmen von DWDSAT¹⁶ (siehe auch Kapitel 4.6).

¹⁴<http://www.vcs.de> [12.01.2011]

¹⁵<http://landsaf.meteo.pt> [15.01.2011]

¹⁶<http://www.dwd.de/dwdsat> [15.01.2011]

Zur Weiterverarbeitung der eigentlichen HRIT-Daten müssen diese in einem ersten Schritt dekomprimiert werden. Nachfolgend kann die eigentliche Prozessierung erfolgen, die mit der Auswahl entsprechender Untersuchungsgebiete (Subsets) beginnt. Die Umrechnung der Level 1.5 Datenwerte (10 Bit) in Strahlungstemperaturen bzw. Reflektanzen ist ebenfalls Teil der Prozessierung. Abbildung 4.4 zeigt schematisch den Aufbau des Empfangs- und Prozessierungssystems am ZFL.

Bei den prozessierten Untersuchungsgebieten handelt es sich um Deutschland (Zentraleuropa), Kenia (Ostafrika) und den gesamten südlichen Teil Afrikas. Im Endeffekt erhält man für jedes Untersuchungsgebiet Rasterdaten der 12 MSG-Kanäle (vgl. Tabelle 4.1) im Flat-Binary-Format. Mit Hilfe der Bildverarbeitungssoftware ENVI/IDL (Interactive Data Language) können die Daten dann entsprechend weiterverarbeitet werden. Strahlungstemperaturen werden durch die VCS-Software in Grad Celsius ($^{\circ}\text{C}$) ausgegeben, während Reflektanzen als Digital Number (DN) dargestellt werden. Neben den reinen Bilddaten lassen sich zusätzliche Dateien mit entsprechender Geoinformation (Lat/Lon) erstellen. Das heißt, jedem Pixel des einzelnen Untersuchungsgebietes sind geographische Koordinaten zugeordnet. Da sich diese Dateien für die entsprechenden Untersuchungsgebiete nicht verändern, reicht eine einmalige Erstellung. Eine später notwendig werdende Georeferenzierung kann dann mit Hilfe solcher Daten (Geographic Lookup-Table, GLT) umgesetzt werden.

Auf die zur Archivierung eingerichteten Server werden die Daten mittels FTP verschoben und archiviert. Durch die fünfzehnminütigen Aufnahmeintervalle ergeben sich 96 Dateien pro Tag und Kanal. Diese werden in einer Jahr/Monat/Tag-Ordnerstruktur abgelegt. Nacht-Aufnahmen der sichtbaren Kanäle können entsprechend gelöscht werden. Hinzu kommt die Archivierung weiterer Daten, wie LSASAF und DWDSAT, sowie die vollständigen HRIT-Datensätze (12 MSG Kanäle) in komprimierter Form. Eine Übersicht der Datensätze ist in Tabelle 4.4 dargestellt.

4.5 Erstellung der SAFNWC-Produkte

Da die Wolkenprodukte des SAFNWC (vgl. Kapitel 3.2) in dieser Arbeit als Vergleich herangezogen werden, soll im Folgenden eine genauere Beschreibung der Prozessierung bzw. Erstellung dieser Produkte erfolgen. Zur Erstellung und Nutzung bedarf es zuerst einer Lizenzvereinbarung, die im Rahmen der SAFNWC¹⁷ angefordert werden kann. Hiernach besteht die Möglichkeit, das entsprechende Softwarepaket, inklusive Testdatensätze, herunterzuladen. Die Software ist nur auf einem Unix- bzw. Linux-System lauffähig und der Quellcode muss nach einigen vorbereitenden Maßnahmen (z.B. müssen C- und Fortran-Compiler installiert sein) kompiliert werden. Eine detaillierte Beschreibung befindet sich in der Dokumentation der jeweiligen

¹⁷<https://www.nwcsaf.org> [10.02.2011]

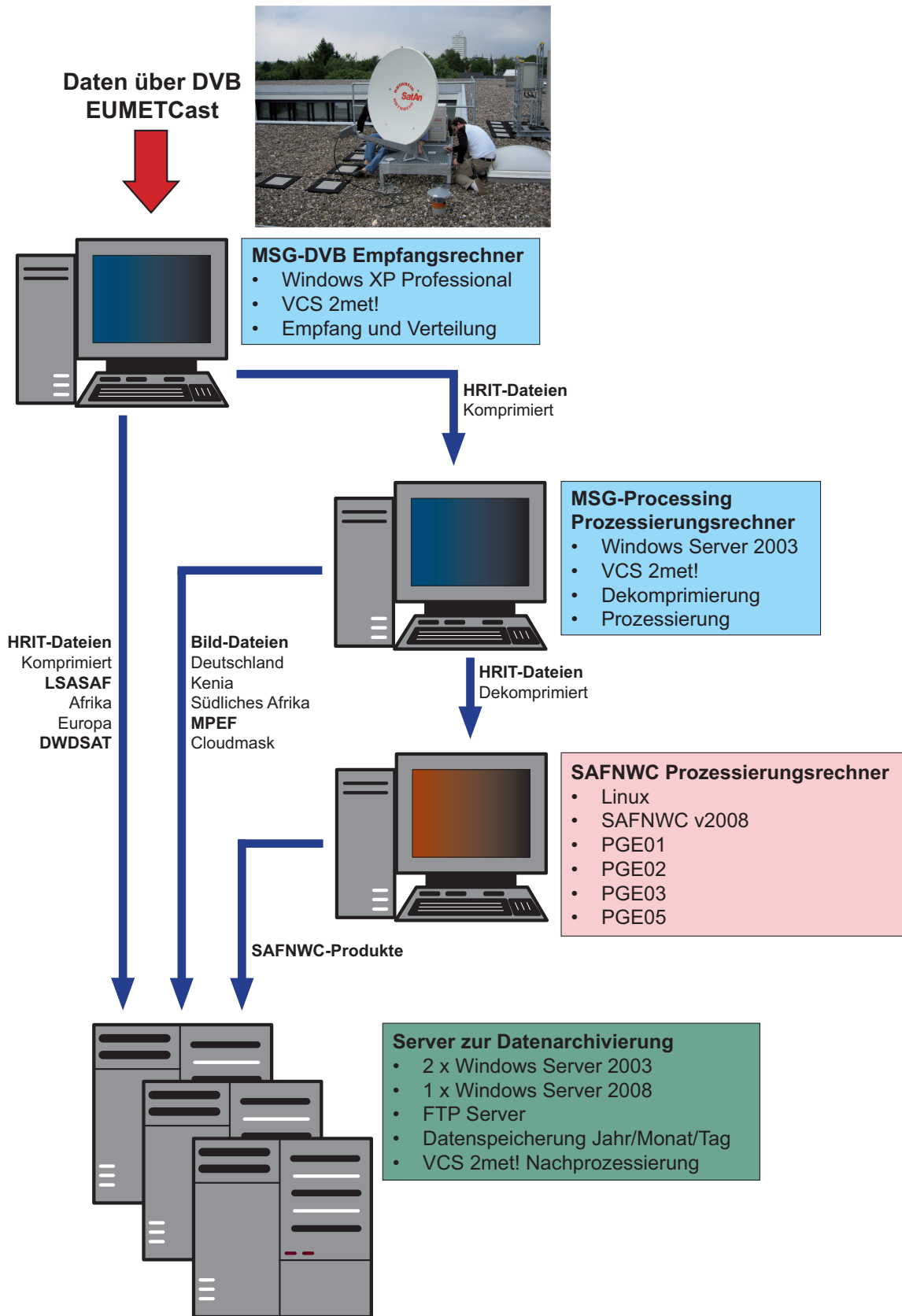


Abb. 4.4: Schematische Darstellung der METEOSAT-Empfangsanlage am ZFL, inklusive Datenarchivierung.
 Quelle: Eigener Entwurf

Tab. 4.4: Archivierte Datensätze am ZFL in Bonn. *Quelle: Eigene Aufstellung*

Datensatz	Zeitreihe		
Deutschland	327 x 271	12 Kanäle	seit Juli 2007
Kenia	525 x 470	12 Kanäle	seit Juli 2007
Südliches Afrika	1098 x 859	12 Kanäle	seit Mitte 2006
HRIT komprimiert	„Volle Scheibe“	12 Kanäle	seit Juli 2007
MPEF Cloudmask	„Volle Scheibe“		seit Juli 2007
DWDSAT	Kapitel 4.6		seit Juli 2007
LSASAF Surface Temperature	Afrika		seit Juli 2007
LSASAF Surface Temperature	Europa		seit Juli 2007
SAFNWC-Wolkenprodukte	Kapitel 4.5		seit Januar 2008

Softwareversion (vgl. SAFNWC 2007b, SAFNWC 2007g, SAFNWC 2007i, SAFNWC 2008a, SAFNWC 2008b). Im vorliegenden Fall handelt es sich um Version v2.0, sowie Version v2008. Ältere Versionen können bezogen auf aktuelle MSG-Daten, wegen der seit Mai 2008 eingeführten effektiven Schwarzkörperstrahlung (Kalibrierungsverfahren, vgl. Seite 68), nicht mehr eingesetzt werden. Mit jeder neuen Version kommt es zu Veränderungen, die neben besserer Bedienbarkeit und effektiverer Berechnung, auch die Genauigkeit der einzelnen Produkte erhöhen sollen. Bei der aktuellsten Version (Anfang 2011) handelt es sich um Version v2010¹⁸. Da die Produkterstellung auf einer Real-Time-Prozessierung der entsprechenden METEOSAT-Daten basiert, und in dieser Arbeit Daten des Jahres 2008 herangezogen werden, bezieht sich die Erstellung der Produkte bis einschließlich April 2008 auf Version v2.0 und danach auf Version v2008. Dies ist bei der späteren Auswertung von Genauigkeiten zu berücksichtigen.

Nach der Installation kann mit der Prozessierung begonnen werden. Über verschiedene Konfigurationsdateien werden, ähnlich wie bei VCS !2met, bestimmte Teilgebiete (Subsets) der „MSG-Scheibe“ ausgewählt. Ein Abgleich mit den von VCS !2met erstellten Teilgebieten (vgl. Kapitel 4.4 und Tabelle 4.4) Deutschland, Kenia und Südliches Afrika erfolgt über die Geoinformation die jedem Pixel zugeordnet ist (siehe hierzu Seite 72). Mit Hilfe der Koordinaten des zentralen Pixels und den Abmessungen der jeweiligen Region werden diese, in Bezug auf die „Volle Scheibe“, abgeglichen (SAFNWC 2007g, SAFNWC 2008b).

Gleichzeitig hängt die Generierung der Produkte nicht nur von der jeweiligen Fragestellung, sondern auch von den notwendigen Input-Daten ab (vgl. Tabelle 3.2). Da vorliegend keine Wettervorhersagemodell-Daten (NWP, vgl. Kapitel 3.2) zur Verfügung stehen, beschränkt sich die Erstellung auf nachfolgend aufgeführte Wolkenprodukte. Beispielsweise kann das Pro-

¹⁸Weitere Informationen unter <https://www.nwcsaf.org> [10.02.2011].

dukt Precipitating Clouds PGE04 (Product Generator Element) nicht erstellt werden (SAFNWC 2007d, SAFNWC 2007f, SAFNWC 2007h, SAFNWC 2008b).

- *PGE01* - Cloud Mask
- *PGE02* - Cloud Type
- *PGE03* - Cloud Top Temperature and Hight
- *PGE05* - Convective Rainfallrate

Stehen die METEOSAT-Daten im dekomprimierten HRIT-Format im Eingangsverzeichnis zur Verfügung (Zufuhr über FTP, vgl. Abbildung 4.4), wird die Prozessierung mittels integriertem Taskmanager gestartet. Dieser erfasst automatisch die notwendigen Zeitschnitte (Abgestimmt auf Universal Time Coordinated, UTC). Bei Bedarf kann das Eingangsverzeichnis nach der Prozessierung bereinigt werden. Die Produkte werden in Ausgangsverzeichnisse mit der Struktur Produkt/Region/Jahr/Monat/Tag als HDF5-Datei gespeichert, von wo aus sie für Vergleichs- oder Validierungszwecke weiterbearbeitet werden. Innerhalb der Konfigurationsmöglichkeiten steht zudem ein Offline-Betrieb zur Verfügung, mit dem bestimmte Zeitschnitte komfortabel nachprozessiert werden können (SAFNWC 2008b).

4.6 DWDSAT als Validierungsgrundlage

Wie bereits zuvor erwähnt, können im Rahmen von EUMETCast bzw. DVB neben den METEOSAT-SEVIRI-Daten eine Reihe weiterer Daten empfangen werden. Bei einem Teil dieser Daten handelt es sich um sogenannte DWDSAT-Daten, mit denen der Deutsche Wetterdienst (DWD) Außenstellen, Universitäten und anderen Nutzern meteorologische Informationen zur Verfügung stellt. Man unterscheidet zwischen einem Globalen Datensatz und solchen Produktgruppen, die mit einer Nutzungsbeschränkung versehen sind (kostenpflichtig)¹⁹.

Da für eine von Fernerkundungsmethoden unabhängige Validierung der Wolkendetektion bzw. Wolkenklassifizierung im Grunde nur synoptische Beobachtungsdaten in Frage kommen, können diese aus den DWDSAT-Datensätzen extrahiert werden. Innerhalb der Grundversorgung (Globaler Datensatz) werden synoptische Beobachtungsdaten²⁰ in sogenannten SYNOP-Bulletins zusammengefasst. Durch Setzen bestimmter Filter in den Eingangsverzeichnissen des Empfangsrechners können so SYNOP-Basis-Daten (gts06-SYNOP1_Bulletins) empfangen und archiviert werden. Nach Dekodierung der Daten ergibt sich ein lesbares Textformat, welches im

¹⁹Eine ausführliche Auflistung und Kurzbeschreibung aller erhältlichen Produkte und Daten ist unter <http://www.dwd.de/dwdsat> [15.01.2011] zu finden.

²⁰Ausführliche Informationen zur Synoptik, siehe Kurz (1990).

Rahmen der Validierung weiterverwendet werden kann. Die Archivierung erfolgt ähnlich wie bei den METEOSAT-Daten, in einer Jahr/Monat/Tag-Ordnerstruktur.

Innerhalb der SYNOP-Basis-Daten erhält man ausführliche Informationen aller Wetterstationen der World Meteorological Organization (WMO²¹) mit einer zeitlichen Auflösung von bis zu einer Stunde. Hierbei handelt es sich um ein weltweites Netz von Wetterstationen, bei denen im Rahmen der WMO, Standards für die Messung sowie den Austausch und die Verarbeitung meteorologischer Daten festgelegt sind. Jede Station ist über einen Zahlencode eindeutig identifizierbar und bietet umfassende Informationen einzelner meteorologischer Parameter, wie Temperatur, Luftdruck, Wolken, Niederschlag usw. Die für diese Arbeit verwendeten Stationen innerhalb von Deutschland sind mit ihrer jeweiligen Lokalisation in Abbildung 6.1 aufgeführt.

²¹<http://www.wmo.int> [18.02.2011]

5 Automatisierter Ansatz der satellitenbasierten Wolkendetektion und Wolkenklassifizierung mit Support Vector Machines

In den vorherigen Kapiteln wurde deutlich, dass eine satellitenbasierte Wolkendetektion und Wolkenklassifizierung mit einer Vielzahl von Bildverarbeitungsverfahren möglich ist. Bei operationell betriebenen Verfahren haben sich in diesem Zusammenhang vor allem Schwellenwertansätze (vgl. Kapitel 3) durchgesetzt. Die Schwierigkeit solcher Methoden liegt jedoch bei der exakten Bestimmung des jeweiligen Schwellenwertes (siehe hierzu auch Kriebel et al. 2003, Derrien und Le Gléau 2005a). Zusätzliche Daten über den Zustand der Atmosphäre (Daten aus Wettervorhersagemodellen) sind erforderlich, um genaue Ergebnisse erzielen zu können. Andere Klassifizierungsverfahren, wie beispielsweise Neuronale Netze oder Cluster-Analysen, sind entweder unüberwacht oder es sind erhebliche Mengen an unabhängigen Daten erforderlich, damit das entsprechende Verfahren eine Zuweisung durchführen kann. Support Vector Machines (SVM) basieren zwar auch auf der Grundlage bekannter Datenwerte (Ground Truth, Trainingsdaten) um eine Trennung der Klassen vollziehen zu können, sie kommen jedoch mit vergleichsweise wenigen dieser Daten aus. Gleichzeitig besteht in Bezug auf die Trainingsdaten keine Abhängigkeit zur Verteilung der Klassen, wie dies z.B. bei einer Maximum-Likelihood-Klassifizierung der Fall ist (Azimi-Sadjadi und Zekavat 2000, Cristianini und Shawe-Taylor 2000, Garay 2005, Canty 2006, Foody und Mathur 2006, Abe 2010).

Das Ziel der vorliegenden Arbeit ist demnach, mit Hilfe von Support Vector Machines eine vergleichsweise effektive Wolkenerkennung und anschließende Wolkenklassifizierung in Bezug auf METEOSAT-SEVIRI-Daten durchzuführen. Die Klassifizierung soll hierbei auf der Information der einzelnen Kanäle des METEOSAT-Sensors basieren (vgl. Tabelle 4.1). Das heißt, die Hauptdatengrundlage liegt bei den Bilddaten des Satelliten, während Zusatzdaten nur bedingt, im Rahmen der Trainingsdatengewinnung erforderlich sind (vgl. Kapitel 5.3 und 5.4).

Arbeiten die sich mit einer auf Support Vector Machines basierenden Wolkendetektion beschäftigen haben, wurden zwar schon durchgeführt, beziehen sich jedoch meist auf eine be-

grenzte Anzahl von Einzelszenen polarumlaufender Satelliten. Zudem erfolgt die Erhebung notwendiger Trainingsdaten meist auf Basis von Expertenwissen, verbunden mit einem erheblichen Arbeitsaufwand (Azimi-Sadjadi und Zekavat 2000, Mazzoni et al. 2005, Han et al. 2006, Zibert et al. 2006, Gomez-Chova et al. 2010). Ob und inwieweit eine Anwendung auf zeitlich hoch aufgelöste Datenreihen (fünfzehnminütige Aufnahmeintervalle von METEOSAT-SEVIRI) möglich ist, soll mit dieser Arbeit dargelegt werden.

Oft werden in der Fernerkundung Trainingsdatensätze aus multitemporalen Satellitendaten zusammengestellt, indem einzelnen Pixeln durch unabhängige Daten (Ground Truth) oder Expertenwissen ein Label (z.B. Wald, Acker oder aber Wolke, keine Wolke usw.) zugewiesen wird (vgl. Canty 2006). Aus diesen multitemporalen Trainingsdatensätzen lässt sich dann ein SVM-Klassifikator (SVM-Modell) erzeugen, der auf weitere Szenen angewendet werden kann. Meist wird zusätzlich ein durch Zufallsverfahren erzeugter Validierungsdatensatz von den Trainingsdaten abgegrenzt, mit dem die Genauigkeit des Klassifikators überprüft werden kann (vgl. Zibert et al. 2006, Gomez-Chova et al. 2010). Innerhalb der Untersuchung von Landbedeckungsveränderungen mag dies in Bezug auf unterschiedliche Zeitpunkte gut funktionieren. Bei der Auseinandersetzung mit Wolken in Verbindung mit zeitlich umfangreichen Datenreihen, müsste man jedoch einen enorm großen Trainingsdatensatz erzeugen, um alle möglichen Situationen im Zusammenhang mit Wolken zu berücksichtigen. Durch die hohe zeitliche und räumliche Variabilität von Wolken und den Einfluss der jeweiligen atmosphärischen Situation, ist dies wegen sich ändernder Strahlungsverhältnisse im Jahres- und auch Tagesgang extrem problematisch. Die durch den Satelliten gemessenen Strahlungsverhältnisse (insbesondere Strahlungstemperaturen der Infrarotkanäle) an einer Wolke können zu einem anderen Zeit-

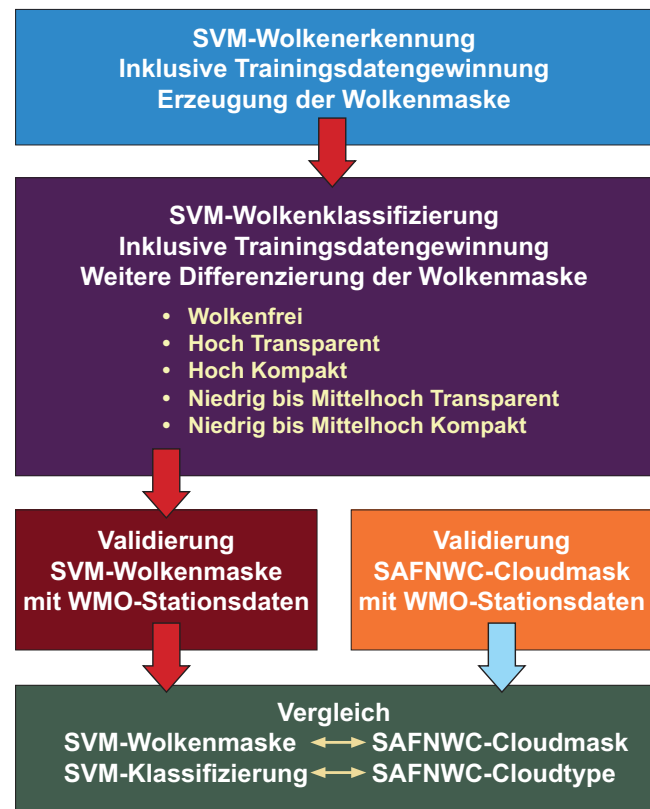


Abb. 5.1: Kombinierte Wolkenerkennung und Wolkenklassifizierung basierend auf Support Vector Machines (SVM), für jeden METEOSAT-Zeitschnitt, einschließlich Validierungsverfahren und Vergleich mit SAFNWC-Produkten. *Quelle: Eigener Entwurf*

raumlich unterschiedlichen Situationen, müsste man jedoch einen enorm großen Trainingsdatensatz erzeugen, um alle möglichen Situationen im Zusammenhang mit Wolken zu berücksichtigen. Durch die hohe zeitliche und räumliche Variabilität von Wolken und den Einfluss der jeweiligen atmosphärischen Situation, ist dies wegen sich ändernder Strahlungsverhältnisse im Jahres- und auch Tagesgang extrem problematisch. Die durch den Satelliten gemessenen Strahlungsverhältnisse (insbesondere Strahlungstemperaturen der Infrarotkanäle) an einer Wolke können zu einem anderen Zeit-

punkt eine wolkenfreie Fläche darstellen. Dies gilt sicherlich verstärkt für niedrige Wolken oder teilweise bewölkte Pixel. Insbesondere wolkenfreie Winternächte (Zentraleuropa) können durch nächtliche Ausstrahlung zu sehr „kalten“ Messwerten führen, was die Abgrenzung zu den relativ warmen, niedrigen Wolken schwierig macht. Würde man demnach einen Trainingsdatensatz mit Daten aus mehreren Einzelaufnahmen (Zeitschnitte) erzeugen, würde es dem zugehörigen SVM-Klassifikator wahrscheinlich nicht gelingen, bestimmte Bewölkungssituationen zu unterschiedlichen Zeitpunkten ausreichend genau widerzuspiegeln (Saunders 1986, Saunders und Kriebel 1988, Gomez-Chova et al. 2010).

Aus diesem Grund liegt die Entscheidung darin, sich bei der Klassifizierung immer nur auf eine Szene zu beziehen, wobei das komplette Verfahren inklusive angepasster Trainingsdatengewinnung durchgeführt wird. Dies geschieht folglich für jeden METEOSAT-Zeitschnitt, und sollte durch Optimierung in einem adäquaten zeitlichen Rahmen durchführbar sein (vgl. Tabelle 5.5). Obwohl solch eine Herangehensweise in Bezug auf maschinelles Lernen nicht unbedingt üblich ist, haben Testläufe gezeigt, dass eine effektive Berechnung auch im Rahmen eines operationellen und zeitnahen Betriebes möglich ist.

Die Eingrenzung des Untersuchungsgebietes erfolgt anhand der METEOSAT-Kanäle, bei dem der gewünschte Bereich (Subset) aus der Gesamtszene von METEOSAT („Full-Disc“, vgl. Kapitel 4), für jeden Aufnahmezeitpunkt, mit Hilfe der VCS 2met! Software ausgeschnitten wird (vgl. Kapitel 4.4). Zu Beginn der Datenarchivierung entschied man sich für einen Ausschnitt, der auch benachbarte Bereiche von Deutschland einschließt (Zentraleuropa, 327 x 271 Pixel, vgl. Abbildung 5.2). Da innerhalb des angestrebten Verfahrens versucht wird, notwendige Trainingsdaten aus der jeweiligen Bildinformation zu gewinnen (vgl. Kapitel 5.3 und 5.4), welche im Verhältnis zur Gesamtpixelzahl des Bildausschnitts so gering wie möglich gehalten werden sollten, wird eine weitere Eingrenzung (Maske) durchgeführt (vgl. Abbildung 5.2). Dies geschieht auch wegen der teilweise hohen Temperaturgradienten zwischen den südlich und nördlich der Alpen liegenden Teilregionen. Es schien daher sinnvoller, das gesamte Klassifizierungsverfahren auf einen kleineren Bildausschnitt anzuwenden. Niedrige Strahlungstemperaturen (z.B. Kanal IR108 oder IR120, vgl. Tabelle 4.1) können südlich der Alpen auf Wolken hindeuten, während es sich nördlich der Alpen um wolkenfreie Flächen handelt. Diese Problematik, ähnlich wie beim Vergleich über mehrere Zeitschnitte, erschwert es dem SVM-Klassifikator eine korrekte Trennung über die gesamte Fläche hinweg durchzuführen. Würde man das vorliegende Verfahren für ein größeres Gebiet anstreben, muss die Klassifizierung für mehrere Teilregionen angepasst abgearbeitet und anschließend zusammengefügt werden.

Da sich im Allgemeinen die Detektion von Wolken über Landflächen schwieriger darstellt als über großen Wasserflächen, wird in dieser Arbeit auch nur auf solche Flächen eingegangen. Unter Anpassung der Trainingsdaten bzw. Trainingsdatengewinnung (vgl. Kapitel 5.3 und 5.4) sollte eine Anwendung über großen Wasserflächen aber ohne weiteres durchführbar sein. Es ist

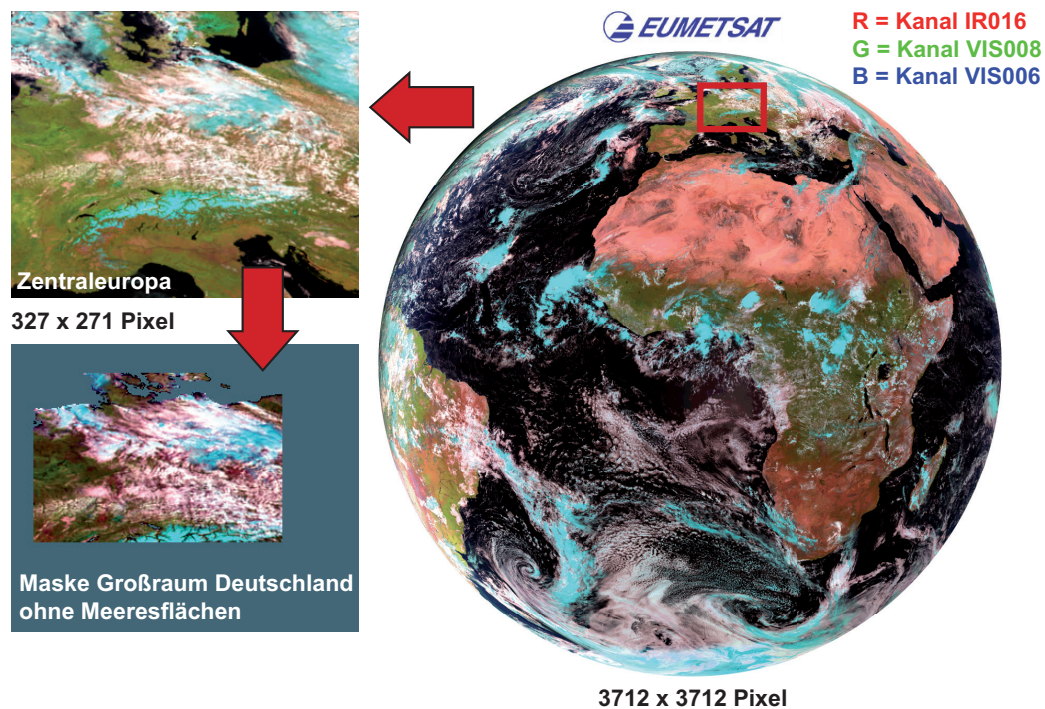


Abb. 5.2: Untersuchungsgebiet extrahiert aus dem gesamten METEOSAT-Aufnahmebereich. Ausschnitt Deutschland bzw. Zentraleuropa (327 x 271 Pixel) dargestellt als RGB. Weitere Eingrenzung durch zusätzliche Maske, inklusive Ausmaskierung der Meeresflächen. *Quelle: Eigener Entwurf*

zu erwarten, dass die Wolkendetektion über großräumig homogenen Wasserflächen noch effektiver umzusetzen ist (vgl. Saunders und Kriebel 1988, Derrien und Le Gléau 2005a). Um den Einfluss von Meeresflächen auf die Detektion von Landflächen, insbesondere in den Küstenbereichen, zu verhindern, erfolgt eine weitere Ausmaskierung. Alle Masken basieren auf Werten, die in der weiteren Berechnung keine Berücksichtigung finden (Not a Number, NaN).

Der für den SVM-Klassifikator erforderliche Merkmalsraum (vgl. Kapitel 5.1) setzt sich, der Fragestellung entsprechend, nur aus wolkenrelevanten Kanälen und Kanal-Differenzen (vgl. Kapitel 3.1 und Tabelle 4.1) des jeweiligen Zeitschnitts zusammen. Wertebereiche der verwendeten Kanäle spannen den Merkmalsraum auf, in dem die spätere Trennung der Klassen stattfindet. Die zeitliche Zuordnung erfolgt über den Dateinamen der Stapel-Datei (Layerstack). Gleichzeitig bezieht sich der Umfang notwendiger Trainingsdaten ebenfalls nur auf die jeweilige Einzelszene (vgl. Abbildung 5.2 und 5.3).

Da die Klassifizierung sowohl tagsüber als auch nachts angewendet werden soll, kommen zudem nur Infrarotkanäle (Strahlungstemperaturen) infrage die tagsüber nicht dem Einfluss kurzweiliger solarer Strahlung unterliegen (vgl. hierzu Seite 53). Abbildung 5.3 zeigt die Zusammensetzung des innerhalb der Klassifizierung verwendeten Merkmalsraumes.

Support Vector Machines stellen einen binären Klassifikator dar, der besonders effektiv bei der Lösung von Zwei-Klassen-Problemen einsetzbar ist (vgl. Kapitel 5.1). Obwohl mehrere

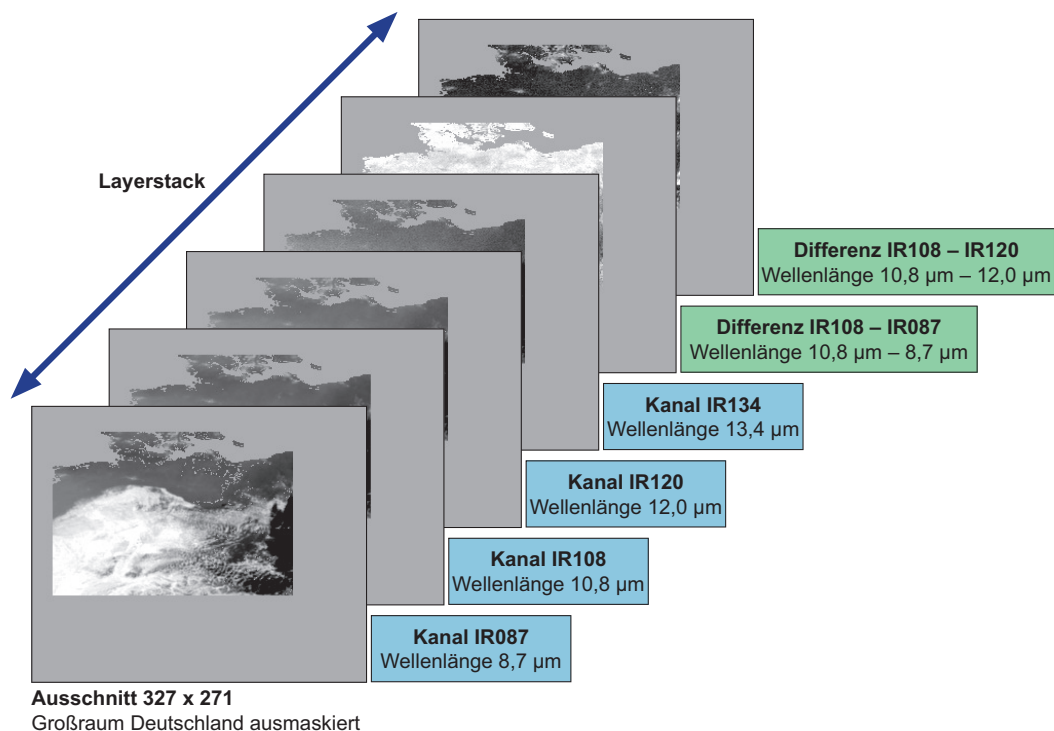


Abb. 5.3: Merkmalsraum der SVM-Klassifizierung, bestehend aus wolkenrelevanten Infrarotkanälen und Kanal-Differenzen. Ausmaskierter Ausschnitt Deutschland. *Quelle: Eigener Entwurf*

Klassen getrennt werden können (vgl. Kapitel 5.2), steht die Trennung von wolkenkontaminierten und wolkenfreien Flächen durch eine Wolkenmaske an erster Stelle. Die Einteilung in weitere Klassen (vier zusätzliche Wolkenklassen) erfolgt anschließend auf Basis der zuvor erstellten Maske. In verschiedenen Testläufen schien diese Herangehensweise vorteilhaft zu sein, sowohl bezogen auf die effektive Auswahl notwendiger Trainingsdaten, als auch in Bezug auf das spätere Validierungsverfahren. Eine separate Validierung der Wolkenmaske lässt sich hierdurch einfacher umsetzen.

Das gesamte Verfahren, für jede einzelne METEOSAT-Aufnahme, setzt sich wie in Abbildung 5.1 dargestellt, zusammen. Eine ausführliche Beschreibung erfolgt in den nachfolgenden Kapiteln. Da die ganze Konzeption rückwirkend auf den Datensatz von 2008 angewendet wird, kann die Validierung mit Beobachtungsdaten und der Vergleich mit den in Kapitel 3.2 und 4.5 beschriebenen SAFNWC-Wolkenprodukten ebenso für jeden Zeitschnitt durchgeführt werden (siehe Kapitel 6). Auf diese Weise erhält man, zusätzlich zur visuellen Begutachtung, einen quantitativen Vergleich.

5.1 Support Vector Machines als Klassifikator

Support Vector Machines (SVM) stellen ein relativ modernes Verfahren im Bereich des maschinellen Lernens dar. Hierunter versteht man die Fähigkeit eines Systems, anhand von Beispielen zu lernen und Gesetzmäßigkeiten zu erkennen (Lernphase bzw. Training). Aufgrund der gewonnenen Erfahrung können nachfolgend unbekannte Datenbestandteile beurteilt werden. Das Prinzip der SVM wurde von Vladimir Vapnik entwickelt und bietet die Möglichkeit binäre Klassifizierungsprobleme zu lösen (vgl. Vapnik 1998). Bestimmte Trainingsobjekte (in einem Merkmalsraum), bei denen bekannt ist welcher Klasse sie zugehören, werden als Vektoren in einem Vektorraum repräsentiert. Die Trennung der Klassen erfolgt hierbei durch die Einbindung einer Hyperebene, wobei der Abstand derjenigen Vektoren, die der Hyperebene am nächsten liegen, maximiert wird. Somit erhält man einen breiten Saum der dafür sorgt, dass auch Objekte die nicht annähernd den Trainingsobjekten entsprechen relativ genau klassifiziert werden können. Insbesondere die Trainingsvektoren die nah an der Hyperebene liegen, sind entscheidend für die exakte Position dieser. Die nächstliegenden Vektoren werden daher als Stützvektoren (Support Vectors) bezeichnet und sind namensgebend für das Verfahren.

Bei der Betrachtung eines zweidimensionalen Merkmalsraumes kann man beispielhaft deutlich machen wie die Trennung mittels Hyperebene funktioniert. Ausgehend von zwei mit +1 und -1 gekennzeichneten Klassen $y_i \in \{+1, -1\}$ und einer bestimmten Anzahl an Trainingspunkten $x_i \in X$, die diese Merkmalseigenschaften repräsentieren, besteht das Ziel darin, eine optimale Aufteilung des Merkmalsraumes zu gewährleisten, indem man negative von positiven Trainingspunkten abgrenzt. In Gleichung 5.1 ist dies dargestellt, wobei w und b so gewählt werden müssen, dass eine Trennung der Trainingsdaten durch die Hyperebene erfolgt (vgl. auch Abbildung 5.5). Diese kann dann als lineare Entscheidungsfunktion eingesetzt werden, wobei die Klassenzugehörigkeit auf dem Vorzeichen der Entscheidungsfunktion basiert.

$$H = \{x | \langle w, x \rangle + b = 0\} \quad (5.1)$$

H = Hyperebene

w = Normalvektor

x = Trainingsobjekte im Merkmalsraum

b = Verschiebung

$\langle \cdot, \cdot \rangle$ = Skalarprodukt

In Abbildung 5.4 sind verschiedene Möglichkeiten der Trennung dargestellt. Jedoch gibt es nur eine *optimale* Trennung mit breitem Rand (Hyperebene). Dies wird durch die Maximierung des kleinsten Abstands zur Hyperebene erreicht (Margin, siehe auch Abbildung 5.5), indem man eine Hyperebene mit minimaler quadratischer Norm $\|w\|^2$ auswählt. Gleichzeitig muss

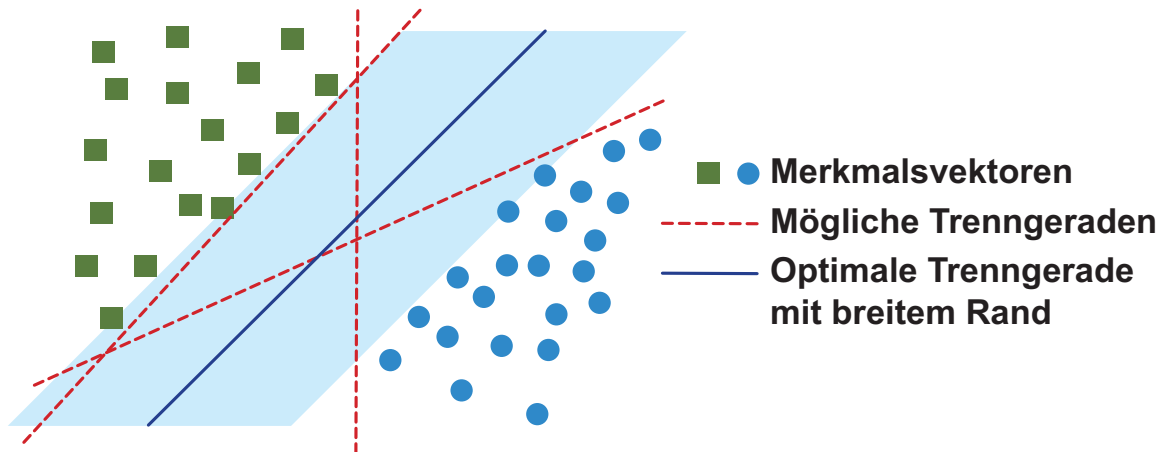


Abb. 5.4: Verschiedene Möglichkeiten der Trennung zweier Klassen (zweidimensionaler Raum). *Quelle: Verändert nach Markowitz unter http://lectures.molgen.mpg.de/statistik/docs/Kapitel_16.pdf [10.09.2010]*

die Nebenbedingung 5.3 für jeden Trainingspunkt x_i mit $1 \leq i \leq n$ erfüllt sein. Somit wird der Ausdruck 5.2 unter Berücksichtigung der Nebenbedingung 5.3 bezogen auf w und b minimiert.

$$\frac{1}{2} \|w\|^2 \quad (5.2)$$

$$y_i(\langle w, x_i \rangle + b) \geq 1 \quad (5.3)$$

Das beschriebene Optimierungsproblem lässt sich in seiner dualen Form mit der Lagrange-Funktion zusammenfassen. Der Normalvektor w ist hierbei eine Linearkombination aus den Trainingsdaten.

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (5.4)$$

Die duale Form lässt sich somit wie folgt darstellen.

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (5.5)$$

Innerhalb der Berechnung 5.5 erfolgt die Maximierung von α , wobei die Nebenbedingungen 5.6 erfüllt sein müssen.

$$\alpha_i \geq 0 \text{ und } \sum_{i=1}^n \alpha_i y_i = 0 \quad (5.6)$$

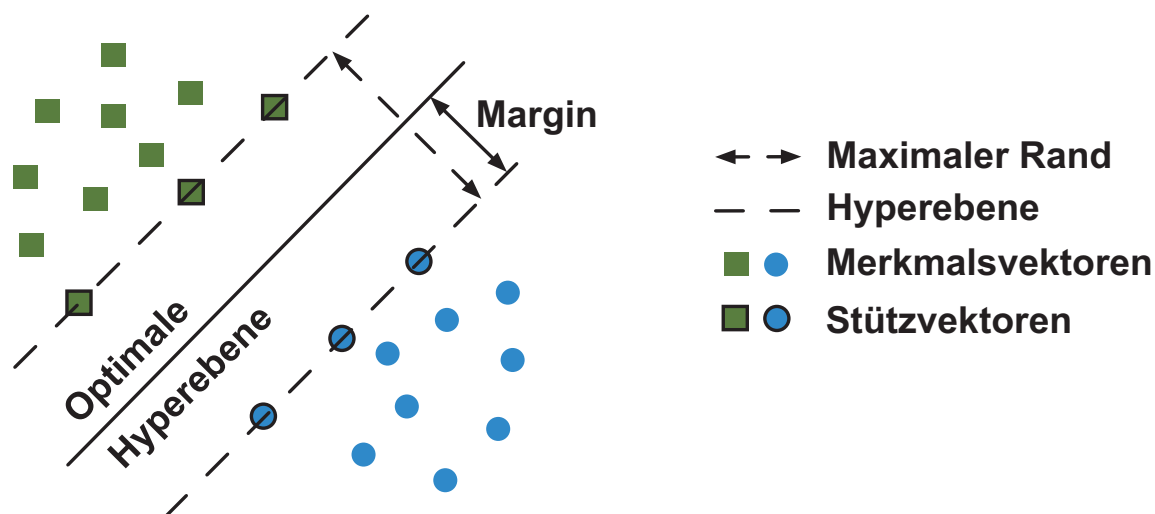


Abb. 5.5: Optimale Einbindung der Hyperebene mit entsprechenden Stützvektoren (Support Vectors). *Quelle:* *Verändert nach Markowitz unter* http://lectures.molgen.mpg.de/statistik/docs/Kapitel_16.pdf [10.09.2010]

Für die duale Form ergibt sich die Entscheidungsfunktion 5.7. Die Stützvektoren (Support Vectors, siehe auch Abbildung 5.5) sind hierbei eine spezielle Menge von Trainingspunkten deren Lagrangevariablen $\alpha_i \neq 0$ sind.

$$f(x) = \text{sgn}(\langle w, x_{\text{neu}} \rangle + b) = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i \langle x_i, x_{\text{neu}} \rangle + b\right) \quad (5.7)$$

In der Regel sieht es jedoch so aus, dass sich die Trainingsdaten nicht streng linear von einander separieren lassen. Häufig kann es bei der Verteilung der Trainingspunkte zu Überlappungen kommen. Aus diesem Grund werden bei der Trennung bzw. Optimierung Fehler zugelassen, die jedoch so gering wie möglich gehalten werden sollten. Durch Schlupfvariablen ξ_i , welche den Abstand des falsch positionierten Merkmalsvektors zur Hyperebene messen, wird die Nebenbedingung 5.3 verletzt. Da die Summe der Fehler minimal gehalten werden soll, wird diese der Zielfunktion hinzugefügt und ebenfalls minimiert. Hinzu kommt eine positive Konstante C die den Einfluss von Trainingsfehlern auf das Optimierungsproblem 5.2 reguliert. Das Optimierungsproblem stellt sich dann mit der Minimierung von w und b , und der Erfüllung der Nebenbedingung 5.9 wie folgt dar.

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (5.8)$$

$$y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \text{ für alle } 1 \leq i \leq n \quad (5.9)$$

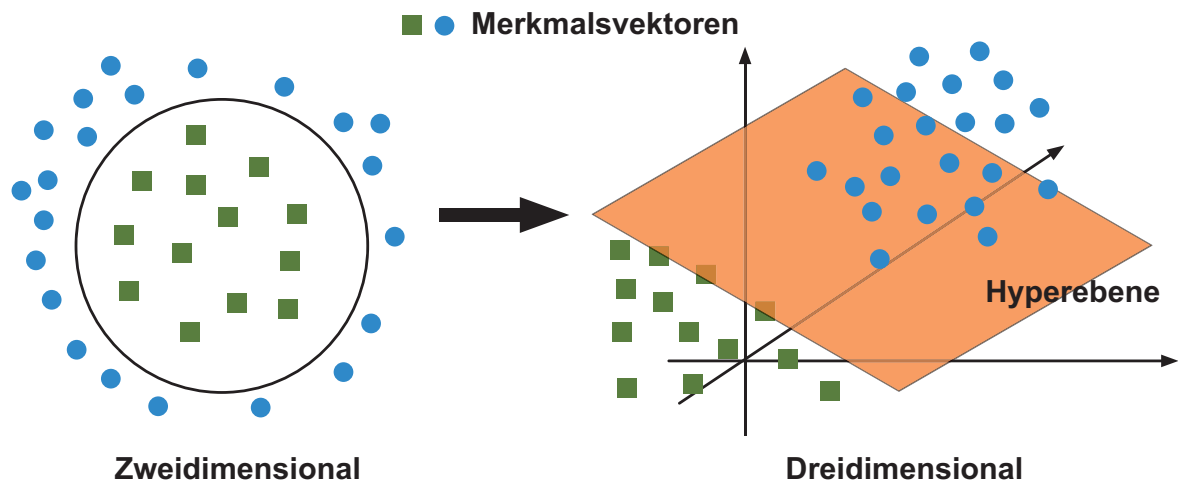


Abb. 5.6: Überführung in einen höher dimensionalen Raum, zur Lösung eines nicht linear trennbaren Klassifizierungsproblems. *Quelle: Verändert nach Markowetz unter http://lectures.molgen.mpg.de/statistik/docs/Kapitel_16.pdf [10.09.2010]*

Wenn das zu Grunde liegende Klassifizierungsproblem nicht linear trennbar ist, kann man dies mit den zuvor beschriebenen Trennverfahren nicht eindeutig lösen. Diese Problematik lässt sich jedoch beheben, indem man die Daten in einen Raum höherer Dimension überführt bzw. abbildet (vgl. Abbildung 5.6). Man bedient sich einer Funktion Φ wobei $d_1 < d_2$ gelten muss.

$$\Phi : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}, x \rightarrow \Phi(x) \quad (5.10)$$

Durch die Abbildung in einen höher dimensionalen Raum wird die Anzahl möglicher linearer Trennungen erhöht. Mit Support Vector Machines lässt sich dies relativ leicht umsetzen, da die Trainingspunkte x_i bezogen auf das Optimierungsproblem nur innerhalb von Skalarprodukten auftauchen. Das Skalarprodukt $\langle x_i, x_j \rangle$ lässt sich somit im höher dimensionalen Raum berechnen $\langle \Phi(x_i), \Phi(x_j) \rangle$. Um diese Rechnung effektiv durchführen zu können, setzt man zusätzlich eine Kernelfunktion K ein, den sogenannten Kernel-Trick.

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (5.11)$$

Durch die Kernelfunktion kann die Hyperebene im höher dimensionalen Raum eingebunden werden und man erhält folgende Entscheidungsfunktion.

$$f(x) = \text{sgn}(\langle w, \Phi(x) \rangle + b) = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i k(x_i, x_{neu}) + b\right) \quad (5.12)$$

Da durch die Nutzung der Kernelfunktion nicht direkt die Skalarprodukte berechnet werden, spart dies gerade in hochdimensionalen Vektorräumen Rechenaufwand und Rechenzeit. Die ge-

bräuchlichsten Kernelfunktionen sind neben der linearen Kernelfunktion 5.11, der polynomiale Kernel 5.13 und die Radial Basis Function 5.14 (Vapnik 1998, Cristianini und Shawe-Taylor 2000, Schölkopf und Smola 2001, Gomez-Chova et al. 2008, Gomez-Chova et al. 2010).

$$K(x,y) = (\gamma \langle x,y \rangle + c_0)^d \quad (5.13)$$

$$K(x,y) = \exp(-\gamma \|x - y\|^2) \quad (5.14)$$

5.2 Umsetzung mit imageSVM

Die Umsetzung der in dieser Arbeit notwendigen SVM-Berechnungen erfolgt mit Hilfe des auf IDL basierenden Tools imageSVM. ImageSVM wurde am Geographischen Institut der Humboldt-Universität zu Berlin, Abteilung Geomatik entwickelt (vgl. Janz et al. 2007) und kann in die Bildverarbeitungsumgebung ENVI/IDL eingebunden werden. Der Vorteil von imageSVM besteht darin, Support Vector Machines komfortabel auf Fernerkundungsdaten (Bilddaten) anwenden zu können. Sowohl Input als auch Output können über entsprechende Dateiformate (ENVI kompatible Formate) zugeführt bzw. gespeichert werden. Da es sich bei der SVM-Klassifizierung im eigentlichen Sinne um ein überwachtes Verfahren handelt, werden die notwendigen Trainingsdaten über sogenannte Regions of Interest (ROI, Bestandteil von ENVI) eingelesen. Gleiches gilt auch für die implementierte Möglichkeit der Validierung. Die eigentlichen SVM-Algorithmen sind über das von Chang und Lin (2001) entwickelte LIBSVM integriert. Sowohl imageSVM als auch LIBSVM werden in regelmäßigen Abständen aktualisiert, um noch effektivere Problemlösungen ermöglichen zu können. Die Berechnungen und Beschreibungen in dieser Arbeit beziehen sich jedoch auf die Versionen imageSVM 1.0 mit LIBSVM 2.84. Das Klassifizierungsverfahren von imageSVM setzt sich insgesamt aus drei Schritten zusammen.

- Grid-Search mit interner Cross-Validation sowie vorherige Auswahl der Kernel- und Regularisierungsparameter (vgl. Kapitel 5.1)
- SVM-Training basierend auf Grid-Search-Parametern
- Klassifizierung mit Hilfe des beim Training erstellten SVM-Modells (Klassifikator)

Durch die Aufteilung in drei Schritte wird das Verfahren transparenter. Beispielsweise können Regularisierungsparameter im Rahmen der Grid-Search verändert werden, um bessere Ergebnisse erzielen oder die Rechenzeit reduzieren zu können. Bei der in imageSVM verwendeten Kernelfunktion handelt es sich um eine Radial Basis Function (vgl. Kapitel 5.1), mit deren Hilfe das Klassifizierungsproblem in einen höher dimensional Raum überführt wird.

Da Support Vector Machines eigentlich einen binären Klassifikator darstellen, gibt es in imageSVM zwei verschiedene Ansätze, Klassifizierungsprobleme mit mehreren Klassen zu lösen. One Against One (OAO) und One Against All (OAA). Bei OAO werden alle möglichen Paare von Klassen betrachtet, um letztendlich die Klassenzugehörigkeit zu ermitteln. Der Ansatz OAA bezieht sich auf den Vergleich von einer Klasse zum Rest. In Bezug auf die Wolkenklassifizierung scheinen beide Herangehensweisen sehr ähnliche Ergebnisse zu liefern. Neben der Erstellung der Wolkenmaske (zwei Klassen), wird bei der weiter differenzierten Wolkenklassifizierung (vgl. Kapitel 5.5) auf OAA zurückgegriffen (Geomatics-Department 2007, Janz et al. 2007).

Alle Datensätze, die den Merkmalsraum bilden (Bildraten der unterschiedlichen Kanäle als Stapel bzw. Layerstack, vgl. Abbildung 5.3) sollten vor der Klassifizierung auf Werte zwischen 0 und 1 skaliert werden. Liegen alle Daten in einem ähnlichen Wertebereich, ist dies zwar nicht zwingend erforderlich, es hat sich aber während der Berechnungen gezeigt, dass die Durchführung des Trainings, mit skalierten Daten, erheblich schneller abläuft. In imageSVM ist eine Routine eingebunden mit deren Hilfe die Skalierung vorgenommen werden kann (Hsu et al. 2003, Janz et al. 2007).

Insgesamt werden bei jedem Klassifizierungsdurchgang mehrere Dateien mit unterschiedlichen Informationen erstellt. Nach Beendigung der Grid-Search erhält man zwei Dateien mit Werten der internen Cross-Validation²² (CV-Rate) und den damit verbundenen Kernel- und Regularisierungsparametern. Diese Informationen sind für das sich anschließende Training und die damit verbundene Erstellung des SVM-Modells (Klassifikator) erforderlich, mit dem nachfolgend klassifiziert oder auch validiert werden kann. Das eigentliche Klassifizierungsergebnis erhält man als Bild-Datei. Zudem lassen sich sogenannte Rule-Dateien erstellen, in denen die Distanz eines jeden Pixels zur Hyperebene dargestellt wird.

Normalerweise ist das in IDL geschriebene imageSVM an die ENVI-Benutzerumgebung angelehnt und Fernerkundungsdaten (alle möglichen durch ENVI lesbaren Bildformate) können in einfacher Art und Weise bearbeitet werden (vgl. Geomatics-Department 2007). Wenn jedoch, wie in dieser Arbeit, eine große Zeitreihe an Daten ausgewertet werden muss, ist die Nutzung eines User-Interfaces, wo alle Parameter immer wieder neu ausgewählt werden müssen, ausgesprochen ineffektiv. Aufgrund des vorliegenden Quellcodes von imageSVM werden demnach nur die notwendigen imageSVM-Einzelfunktionen (z.B. Grid-Search) in die für diese Arbeit entwickelten IDL-Routinen implementiert. Eine genauere Beschreibung der verwendeten Funktionen und Parameter erfolgt in den nachfolgenden Kapiteln.

²²Teile der Trainingspixel werden zur Einschätzung der Trennbarkeit herangezogen, es handelt sich hierbei nicht um eine unabhängige Überprüfung der Genauigkeit.

5.3 Erste Schritte für die Gewinnung von Trainingsdaten

Entscheidend für die erfolgreiche Klassifizierung mit einem Klassifikator, der wie Support Vector Machines auf maschinellem Lernen basiert, ist die Auswahl repräsentativer Trainingsinformationen. Damit eine sinnvolle Trennung durch die Hyperebene in Verbindung mit den Stützvektoren erfolgen kann, sollten solche Trainingsdaten aus einer unabhängigen Quelle stammen und die jeweilige Klasse nach Möglichkeit bis zu ihren Grenzbereichen widerspiegeln. Die Trainingsinformation soll also den wirklichen Zustand am entsprechenden Pixel beschreiben (Ground Truth). In der Praxis stellt sich dies jedoch mitunter als schwierig dar und hängt auch von der verfolgten Strategie der SVM-Klassifizierung ab. Oft stehen nicht genügend Trainingsdaten zur Verfügung oder die Datensätze sind so umfangreich, dass eine Auswertung und Zuteilung, beispielsweise durch Expertenwissen, nur mit erheblichem Aufwand möglich ist (Canty 2006, Gomez-Chova et al. 2010).

Wie bereits zuvor erklärt, soll aufgrund der hohen zeitlichen und räumlichen Variabilität von Wolken, die Klassifizierung für jeden METEOSAT-Aufnahmezeitpunkt inklusive Training separat durchgeführt werden. Da sich nur auf die am Satelliten gemessenen Strahlungseigenschaften bezogen wird, scheint eine andere Strategie mit Training und Klassifizierung über mehrere Zeitpunkte hinweg nicht sinnvoll. Aus diesem Grund erfolgen die Auswahl der Trainingspixel und die SVM-Klassifizierung nur bezogen auf die jeweilige Einzelszene. Da es sich bei der SVM-Klassifizierung um ein sehr rechenintensives Verfahren handelt, sollte der Umfang der notwendigen Trainingspixel nicht zu groß gewählt werden. Wird hierauf geachtet, vor allem bei Berücksichtigung der zunehmenden Rechenleistung moderner Computersysteme in den letzten Jahren, sollte eine effektive Durchführung des kompletten SVM-Verfahrens möglich sein (vgl. Abbildung 5.1).

Jetzt stellt sich die Frage, wie man für jede Szene eine repräsentative Menge an Trainingspixeln gewinnen kann. Die ursprüngliche Intention lag darin, hierzu Web-Cams bzw. Wetter-Kameras²³ einzusetzen, die auf einer großen Anzahl von Websites abrufbar sind. In der Re-

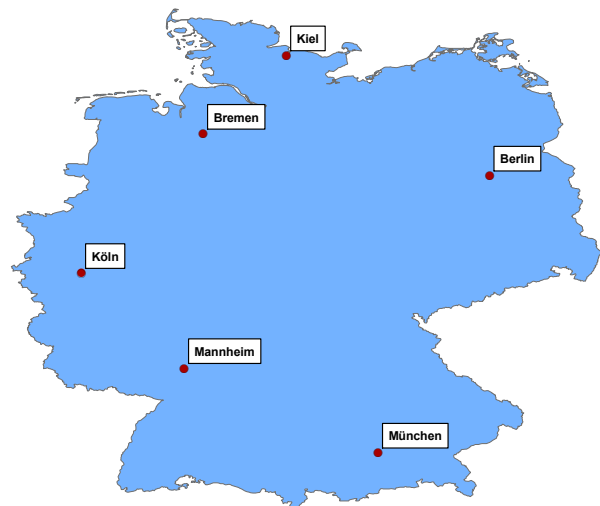


Abb. 5.7: Wetter-Kameras - Räumliche Verteilung der Standorte. *Quelle: Eigener Entwurf*

²³Mit einer Website verbundene Kamera die auf den Himmel bzw. Teile des Himmels gerichtet ist und in bestimmten Abständen Aufnahmen macht.

gel werden diese jedoch nicht lokal gespeichert und archiviert, weshalb die Zusammenstellung eines umfangreichen Datensatzes für das Jahr 2008 nicht mehr möglich war. Lediglich sechs Standorte (vgl. Abbildung 5.7 und Tabelle 5.1) mit unterschiedlicher zeitlicher Auflösung standen zur Verfügung (von minütigen bis zu stündlichen Aufnahmeintervallen). Eine SVM-Klassifizierung kann zwar mit einer geringen Anzahl an Trainingspixeln durchgeführt werden, bei nur sechs Trainingsstandorten lässt sich jedoch keine repräsentative Darstellung der jeweiligen Bewölkungssituation erzeugen. Daher stellen die Aufnahmen der Wetter-Kameras gezwungenermaßen nur eine ergänzende Rolle dar (vgl. Kapitel 5.3.1). Die eigentliche Gewinnung und Zuordnung von Trainingspixeln erfolgt mit Hilfe eines unüberwachten Verfahrens, basierend auf den in Kapitel 3 beschriebenen Erkenntnissen der Schwellenwertalgorithmen. In diesem Zusammenhang werden jedoch nur solche Pixel als Trainingspixel zugeordnet, bei denen mit relativ hoher Sicherheit davon ausgegangen werden kann, dass sie einer entsprechenden Klasse angehören (vgl. Abbildung 5.10). Hierzu wird in einem ersten Schritt ein ungefährender Grenzwert abgeschätzt, um dann jenseits dieses Grenzwertes Trainingspixel über Teile der in Kapitel 3 beschriebenen Algorithmen zuzuordnen. Zusammenfassend liegt also ein unüberwachter Klassifizierungsansatz vor, der mit überwachten Elementen (Informationen der Wetter-Kameras) ergänzt wird. Obwohl das angestrebte Verfahren größtenteils auf einer unüberwachten Herangehensweise basiert, sollte die Zuteilung der Trainingsdaten dennoch relativ genau funktionieren, da sich nur bedingt an den wirklichen Grenzbereich herangetastet wird. Unsicherheiten bei der Auswahl der notwendigen Trainingsinformation können jedoch nicht vollständig ausgeschlossen werden.

5.3.1 Wetter-Kameras zur Erfassung von Wolkeninformationen

Durch die immer einfacher werdende Einsatzmöglichkeit web-basierter Kameras (Web-Cams) besteht auch in zunehmendem Maße die Möglichkeit, den Himmel und das allgemeine Wettergeschehen zu erfassen. An vielen Standorten in Deutschland wird dies bereits durchgeführt und die Bilder können von Interessierten über die jeweilige Website betrachtet werden. Aktualisierungsraten von bis zu einer Minute sind in diesem Zusammenhang nicht selten. Zusätzlich werden die Bilder oft als Loop dargestellt, womit sich beispielsweise Wolkenbewegungen oder der Sonnengang verfolgen lassen.

Bei der Zusammenstellung einer großen Anzahl dieser Kameradaten ließe sich ein von Satellitendaten unabhängiges Bild der Bewölkungs- bzw. Wettersituation erzeugen, mit der Einschränkung, dass die Betrachtung von unten erfolgt. Beim Vergleich mit den an Wetterstationen durchgeführten synoptischen Beobachtungen (vgl. Kapitel 4.6) bieten Wetter-Kameras in der Regel eine höhere zeitliche Auflösung, was sie für Trainingszwecke, von ebenfalls zeitlich hoch

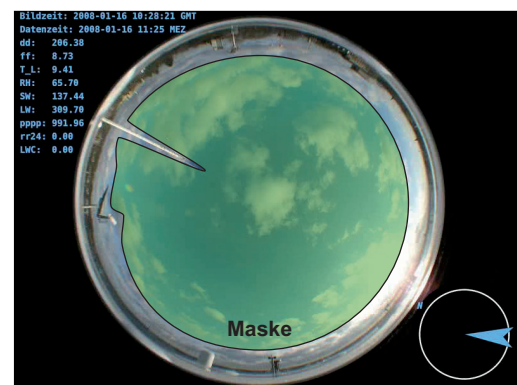
aufgelösten Satellitendaten, attraktiver macht. Dies gilt trotz der meist geringeren Himmelsfläche, die durch Wetter-Kameras erfasst werden kann.

Obwohl für das Jahr 2008 nur sechs Standorte mit unterschiedlichen Kameraeigenschaften zur Verfügung stehen (vgl. Tabelle 5.1), und die vielfältigen Facetten von Bewölkung durch eine so geringe Anzahl an Trainingspixeln nicht repräsentiert werden können, fließen die Daten ergänzend zum Trainingsprozess in das SVM-Klassifizierungsverfahren mit ein. Innerhalb der späteren Validierung soll dann festgestellt werden, wie weit sich ihr Anteil auf die Genauigkeit des Verfahrens auswirkt.

Da die Daten der sechs Kamera-Standorte (vgl. Abbildung 5.7 und Tabelle 5.1) nicht alle als Bildformat (JPEG oder BMP), sondern teilweise nur als Video zur Verfügung stehen, müssen für weitere Prozessierungszwecke die entsprechenden Einzelszenen extrahiert und als Einzelbilder gespeichert werden. Gleichzeitig erfolgt eine Angleichung mittels Dateinamen auf UTC. Dies ist erforderlich, da die METEOSAT-Aufnahmen ebenfalls in diesem Zeitsystem erfasst werden. Angepasst an die METEOSAT-Aufnahmeintervalle ergeben sich für das Jahr 2008 Datenreihen in fünfzehnminütigen Abständen. Einzige Ausnahme stellt Bremen dar, wo nur eine Aufnahme pro Stunde erstellt wird. Nachtaufnahmen oder solche die teilweise verdunkelt sind können für die Analyse nicht verwendet werden. Hier stellt Mannheim die Ausnahme dar, wo eine spezielle Nacht-Kamera den Einsatz solcher Aufnahmen möglich macht.

Die Abgrenzung zwischen Tag und Nacht erfolgt vorab durch visuelle Inspektion, wobei alle verwendbaren Szenen vollständig ausgeleuchtet sein müssen. Eine Berücksichtigung von Sonnenauf- und Sonnenuntergangszeiten war in diesem Zusammenhang nicht hilfreich, da die Beleuchtung je nach Standort und Betrachtungswinkel variiert. Hinzu kommt eine sich je nach Witterung anders darstellende Dämmerungsphase. Alle verdunkelten Aufnahmen werden während der Prozessierung herausgefiltert und tragen nicht zur weiteren Berechnungen bei.

Da jeder Kamera-Standort mit geographischen Koordinaten verknüpft ist, kann dieser einem Pixel der entsprechenden METEOSAT-Szene zugeordnet werden. Dies geschieht über die auf



Vollhimmelskamera des IFM-GEOMAR Kiel






Wetter-Kamera Mannheim

Abb. 5.8: Wetter-Kameras - Ausmaskierte Himmelsfläche. *Quelle: Eigener Entwurf*

Tab. 5.1: Standorte der Wetter-Kameras und ihre jeweiligen Eigenschaften. Aufnahmeintervalle beziehen sich auf die letztendlich genutzte zeitliche Auflösung. Berlin, Bremen und Kiel (a). Köln, Mannheim und München (b).

(a) Berlin, Bremen und Kiel.




Standort	Koordinaten	Abmessung	Aufnahmeintervall	Beschreibung	Bild
Berlin	52.4587° Nord 13.3104° Ost	384 x 288Pixel	15 Minuten	Blick vom Wasserturm des meteorologischen Instituts der FU Berlin (auf dem Fichtenberg in Steglitz) in Richtung Nord-West ^a	
Bremen	53.1091° Nord 8.8536° Ost	352 x 288Pixel	60 Minuten	Netzkamera der Mineralogie Universität Bremen - Blick in Richtung Osten ^b	
Kiel	54.3309° Nord 10.1481° Ost	800 x 600Pixel	15 Minuten	Vollhimmelskamera des IFM-GEOMAR - Blickrichtung nach oben ^c	

^a<http://www.met.fu-berlin.de/de/wetter/webcam/> [05.09.2010]

^b<http://www.min.uni-bremen.de/cgi-bin/Mincam.cgi> [05.09.2010]

^chttp://www.ifm-geomar.de/fileadmin/ifm-geomar/fuer_alle/fb1/ME/Kiel-Wetter/images/wcam/kiel-clouds-big.jpg [05.09.2010]

(b) Köln, Mannheim und München.

Standort	Koordinaten	Abmessung	Aufnahmeintervall	Beschreibung	Bild
Köln	50.9416° Nord 6.9581° Ost	400 x 300Pixel	15 Minuten	Kamera des WDR - Blick auf den Kölner Dom in Richtung Osten ^a	
Mannheim	49.4633° Nord 8.5572° Ost	640 x 480Pixel	15 Minuten	Kamera nahe des Wasserturms (Seckenheim) - Blickrichtung nach Norden - inklusive Nachtaufnahmen ^b	
München	48.1485° Nord 11.5741° Ost	640 x 480Pixel	15 Minuten	Kamera des meteorologischen Instituts der Universität München - Blickrichtung nach Norden (Olympiagelände) ^c	

^a<http://www.wdr.de/themen/global/webcams/domcam.jhtml> [05.09.2010]^b<http://cam.mannheim-wetter.info/> [05.09.2010]^c<http://www.meteo.physik.uni-muenchen.de/mesomikro/stadt/camera.jpg> [05.09.2010]



Abb. 5.9: Unterschiedliche Aufnahmen der Wetter-Kamera München. *Quelle: Meteorologisches Institut der Universität München* <http://www.meteo.physik.uni-muenchen.de/mesomikro/stadt/camera.jpg> [05.09.2010]

Seite 72 beschriebene Geoinformation, die für jedes Pixel abgerufen werden kann. Das Pixel, welches den Koordinaten des Kamera-Standortes am nächsten liegt, kann für Trainingszwecke eingesetzt werden. Aufgrund der Pixelgröße von ca. 6 km in den mittleren bis hohen Breiten, sollte dies in einem adäquaten Rahmen möglich sein. Dieser Umstand betrifft auch die Blickrichtung der jeweiligen Kamera, wobei nur ein Teil der Himmelsfläche über den zugeordneten Pixeln erfasst wird. Für die Erstellung der Wolkenmaske (Kamerainformationen kommen nur bei der Wolkenmaske zum Einsatz) sollten diese Teilflächen ausreichen, um die Bewölkungssituation relativ genau widerzuspiegeln. Je nachdem, ob sich der Kamera-Standort im Zentrum oder am Rand des nächsten Pixels befindet, kann es dennoch zu Ungenauigkeiten kommen.

Damit nicht jede Aufnahme einzeln auf ihre Bewölkungssituation analysiert werden muss, auch im Sinne eines angestrebten operationellen Betriebes, bedarf es einer Automatisierung. In diesem Zusammenhang gibt es bereits Bestrebungen, z.B. mittels Mustererkennung, Bilder auf bestimmte Wolkeneigenschaften zu analysieren (vgl. Yamashita et al. 2004). Da dies jedoch teilweise mit erheblichem Aufwand verbunden ist, und die Daten der Wetter-Kameras nur eine ergänzende Rolle spielen, basiert die Herangehensweise in dieser Arbeit nur auf der RGB-Zusammensetzung der Einzelbilder. Bei wolkenfreiem Himmel überwiegen Blauanteile gegenüber Grün und Rot. Unter Verwendung eines Schwellenwertes, angepasst an die Aufnahmen der unterschiedlichen Standorte, und in Verbindung der Flächenanteile (Berücksichtigung des Blickwinkels der Kamera) lassen sich völlige und teilweise bewölkte, sowie wolkenfreie Aufnahmen unterscheiden. Diese Information kann dem entsprechenden METEOSAT-Pixel zugeordnet werden. Für die Nachtaufnahmen am Standort Mannheim basiert die Erfassung lediglich auf Graustufen, da nachts keine RGB-Zusammensetzung möglich ist. Wolken zeigen sich hier heller als wolkenfreier Himmel. Teile der einzelnen Aufnahmen, die nicht die Himmelsfläche repräsentieren, werden ausmaskiert (vgl. Abbildung 5.8). Spiegeleffekte, Niederschlag, Vögel usw. können die Genauigkeit beeinträchtigen, lassen sich jedoch nur schwer vermeiden.

Zur Begutachtung der automatisierten Wolkenerfassung wurde eine stichprobenartige visuelle Inspektion der einzelnen Aufnahmen durchgeführt. Mit Hilfe eines Zufallsverfahrens erfolgte die Auswahl von 620 Zeitpunkten aus den Daten des Jahres 2008. Für jeden Monat und jeden Standort wurde die gleiche Anzahl an Zeitpunkten ausgewählt, um die räumliche Verteilung und den Jahresgang zu berücksichtigen. Die Begutachtung bezieht sich hierbei auf die Unterscheidung zwischen den drei Kategorien „Völlig Bewölkt“, „Teilweise Bewölkt“ und „Wolkenfrei“. Dies ist zwar für die spätere SVM-Wolkendetektion nicht notwendig, bei der nur zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ unterschieden wird, es lässt sich jedoch deutlicher abschätzen wie genau die automatisierte Erfassung durch Wetter-Kameras funktioniert.

In Bezug auf die Kategorie „Völlig Bewölkt“ wird davon ausgegangen, dass die komplette Himmelsfläche durch Wolken bedeckt ist, wohingegen eine teilweise bewölkte Situation durch wolkenfreie Stellen innerhalb der Gesamtfläche gekennzeichnet ist. Der seitliche Blickwinkel einiger Kameras macht es natürlich unmöglich, jede Lücke exakt erkennen zu können, weshalb die Kennzeichnung „Teilweise Bewölkt“ nur dann eindeutig gegeben ist, wenn größere Flächen (ca. ein Drittel) wolkenfrei sind (vgl. Abbildung 5.9a und 5.9b). Dünne Schleier (z.B. Cirrus-Wolken) werden sowohl bei dem automatisierten Verfahren als auch bei der visuellen Begutachtung nicht als wolkenfreie Flächen eingestuft. Eine Ausnahme bildet in diesem Zusammenhang der Horizont an völlig wolkenfreien Tagen. Auf vielen Aufnahmen ist unter solchen Bedingungen ein grauer Saum in diesem Bereich zu sehen, der jedoch nicht auf Wolken, sondern auf andere atmosphärische Effekte zurückzuführen ist (vgl. Abbildung 5.9c). Weiterhin muss angemerkt werden, dass auch die unterschiedliche Qualität der Aufnahmen Auswirkungen auf die automatisierte Abschätzung, sowie die Begutachtung hat. Insbesondere die Aufnahmen von Köln und Bremen sind oft von Gegenlicht beeinflusst, was die Unterscheidung zwischen Wolken und wolkenfreien Himmelsflächen teilweise schwer macht.

Die Ergebnisse der Begutachtung sind in Tabelle 5.1 dargestellt und zeigen, dass trotz der Beeinträchtigungen eine relativ hohe Genauigkeit erzielt werden kann. Insbesondere bei der Unterscheidung zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ (vgl. Tabelle 5.1 unten) stellt sich das Verfahren als sehr robust dar. Bei Berücksichtigung teilweiser Bewölkung zeigen sich einige Ungenauigkeiten, vor allem bezogen auf die Differenzierung zwischen „Völlig Bewölkt“ und „Teilweise Bewölkt“. Als Gütemaße dienen neben der Konfusionsmatrix, die Overall Accuracy und der Kappa-Koeffizient nach Cohen (vgl. Pontius 2000, Jensen 2004, S. 495 ff.).

5.3.2 Abschätzung Grenzbereich „Wolkenkontaminiert - Wolkenfrei“

Da die Trainingsinformationen der Wetter-Kameras für eine effektive Klassifizierung nicht ausreichen, müssen zusätzliche Trainingspixel gewonnen werden. Hierzu lassen sich die in Kapitel 3 aufgeführten Schwellenwertalgorithmen verwenden. Verschiedene dieser Algorithmen

Tab. 5.1: Begutachtung der automatisierten Wolkenableitung durch Wetter-Kameras basierend auf visueller Inspektion. Mit drei (oben) und mit zwei (unten) Kategorien. Beobachtete Werte sind in den Spalten, detektierte Werte in den Zeilen dargestellt. *Quelle: Eigene Erhebung*

Detektiert ↓ Beobachtet ⇒	Völlig Bewölkt	Teilweise Bewölkt	Wolkenfrei	Summe
Völlig Bewölkt	369	44	0	413
Teilweise Bewölkt	5	108	7	120
Wolkenfrei	1	21	65	87
Summe	375	173	72	620

Overall Accuracy: 87,42 %

Kappa: 0,76

Detektiert ↓ Beobachtet ⇒	Wolkenkontaminiert	Wolkenfrei	Summe
Wolkenkontaminiert	526	7	533
Wolkenfrei	22	65	87
Summe	548	72	620

Overall Accuracy: 95,32 %

Kappa: 0,79

werden auf Infrarotkanäle, aber auch auf Kanal-Differenzen angewendet, um unterschiedliche Wolkenarten und Wolkenhöhen erfassen zu können. Zu diesem Zweck werden Schwellenwerte dynamisch von Szene zu Szene bestimmt, um eine pixelbasierte Trennung durchführen zu können. Bei der vorliegenden Herangehensweise im Sinne einer Trainingsdatenauswahl muss jedoch gewährleistet sein, dass die ausgewählten Pixel wirklich der entsprechenden Klasse zuzuordnen sind. Da die Bestimmung der eigentlichen Schwellenwerte gewöhnlich mit Zuhilfenahme anderer Daten (Wettervorhersagemodell-Daten, NWP) erfolgt, und die genaue Position innerhalb eines Grenzbereiches mit Ungenauigkeiten verbunden ist, müssen die Trainingspixel eine gewisse Distanz zum Grenzwert/Grenzbereich aufweisen. Die eigentliche Trennung erfolgt später durch den SVM-Klassifikator. Verwendet man auf der anderen Seite nur Trainingswerte die in weiter Entfernung vom Grenzbereich vorkommen, so sind diese zwar mit sehr hoher Wahrscheinlichkeit einer bestimmten Klasse zuzuordnen, man erreicht jedoch nie die Nähe des Grenzbereiches, welche für eine exakte Trennung durch den Klassifikator notwendig ist.

Das Ziel sollte demnach sein, mit Hilfe bekannter Schwellenwertalgorithmen eine ausreichende Anzahl an repräsentativen Trainingspixeln zu generieren, die zu einem gewissen Grad auch in der Nähe des Grenzbereiches zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ liegen. Dies lässt sich realisieren, indem man, wie es bei Schwellenwertverfahren üblich ist, mehrere Tests kombiniert (vgl. Abbildung 3.3). Würde man beispielsweise nur einen einfachen Schwellenwert-Test bezogen auf einen Infrarotkanal heranziehen, erkennt dieser sehr gut große

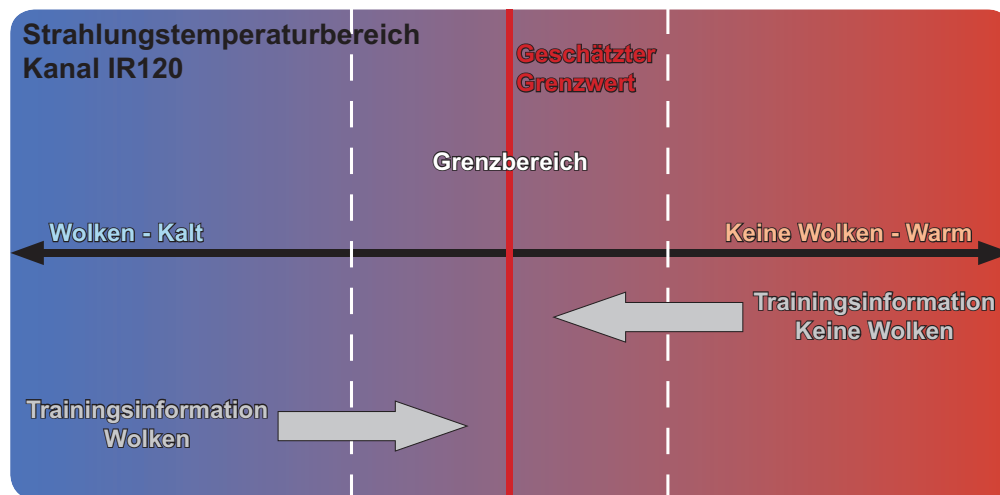


Abb. 5.10: Abschätzung eines ungefähren Grenzwertes/Grenzbereiches anhand des Strahlungstemperaturbereiches von Kanal IR120. Grenzwert bzw. Grenzbereich dienen der späteren Abgrenzung generierter Trainingsdaten. *Quelle: Eigener Entwurf*

Teile der im Vergleich zu Landflächen kälteren Bewölkung. Niedrige Wolken oder teilweise bewölkte Pixel, die den Grenzbereich zwischen Wolken und wolkenfreien Flächen mitbestimmen, werden jedoch nur bedingt berücksichtigt. Zur Bestimmung weiterer Trainingspixel ist deswegen, je nach Klasse, eine unterschiedliche Kombination verschiedener Einzeltests, wie sie in Kapitel 3 beschrieben werden, erforderlich.

Da in dieser Arbeit keine Wettervorhersagemodell-Daten verwendet werden und die Schwellenwerte der einzelnen Algorithmen nicht bekannt sind, muss das Verfahren von der anderen Seite betrachtet werden. Man nähert sich dem Grenzwert/Grenzbereich sozusagen aus anderer Richtung. Beispielsweise können die Pixel, die die kältesten Strahlungstemperaturen (Infrarotkanäle) aufweisen, definitiv als Trainingspixel der Klasse „Wolkenkontaminiert“ eingesetzt werden. Je nach betrachtetem Kanal bzw. betrachteter Kanal-Differenz bezieht man sich so auf die geringsten bzw. höchsten Datenwerte. Dennoch muss bei der Umsetzung, insbesondere für die erste Trennung zwischen „Wolkenkontaminiert“ und „Wolkenfrei“, ein Grenzbereich bzw. ungefähre Grenzwert definiert werden, damit falsche Zuweisungen von Trainingspixeln vermieden werden. Je nach Klasse darf es nicht zu einer Über- bzw. Unterschreitung kommen.

Eine grobe Abschätzung des Grenzbereiches bzw. Grenzwertes erfolgt durch die Betrachtung der Strahlungsverhältnisse des entsprechenden Einzelzeitschnittes. Insbesondere die Strahlungstemperaturen der Infrarotkanäle $10,8 \mu\text{m}$ (IR108) und $12 \mu\text{m}$ (IR120) lassen sich hierfür heranziehen (vgl. Tabelle 4.1). Bezogen auf eine ungefähre Grenzwertbestimmung sollte es bei

beiden Kanälen keine großen Unterschiede geben, weshalb in vorliegender Arbeit ausschließlich Kanal IR120 verwendet wird. Abbildung 5.10 zeigt exemplarisch das grundlegende Prinzip. Die eigentliche Generierung der Trainingsdaten durch die Kombination verschiedener Schwellenwertalgorithmen erfolgt in späteren Schritten (vgl. Kapitel 5.4 und 5.5).

Die Abschätzung des ungefähren Grenzwertes basiert in erster Linie auf dem Histogramm der jeweiligen IR120-Szene. Hierbei wird jedoch nur der Bereich zwischen -10 °C und dem Temperaturmaximum der Gesamtszene berücksichtigt. Innerhalb von Deutschland sollte selbst an kalten Wintertagen die gemessene Strahlungstemperatur (Kanal IR120) wolkenfreier Flächen den Wert -10 °C nicht erheblich unterschreiten. Eine stichprobenartige Begutachtung der Daten von 2008 hat dies bestätigt. Der Übergang von wolkenkontaminierten zu absolut wolkenfreien Flächen sollte demnach im Strahlungstemperaturbereich von -10 °C bis zum Temperaturmaximum der jeweiligen Szene liegen.

Für die weitere Eingrenzung kann man die Verteilung der Pixel (Histogramm) in diesem Temperaturbereich heranziehen. Je nach Ausprägung wolkenkontaminierter und wolkenfreier Flächen weist das Histogramm unterschiedliche Eigenschaften auf. Vor allem die Position des Maximums (Anzahl Pixel) kann weitere Hinweise geben. An dieser Stelle muss erwähnt werden, dass sprunghafte Datenwerte mittels eines gleitenden Mittelwertes geglättet werden. Dies ist in Abbildung 5.11 und 5.12 als rote Linie dargestellt und erleichtert die Analyse.

Bei der Betrachtung der einzelnen Histogramme lassen sich bestimmte Bewölkungssituationen ableiten (vgl. hierzu Saunders und Kriebel 1988). Beispielsweise deutet ein stark ausgeprägtes Maximum um einen eng begrenzten Wertebereich auf wolkenfreie Flächen hin, da dichte Bewölkung tendenziell ein breiteres Spektrum an Strahlungstemperaturen aufweist. Durch die Position des Maximums erhält man demnach einen ersten Grenzwert, um den sich ein grober Grenzbereich spannen lässt. Bezogen auf die Problematik sehr kalter wolkenfreier Flächen oder einer Verteilung von Wolken und wolkenfreien Flächen, die keine eindeutige Interpretation zulässt, kann es jedoch zu falschen Abschätzungen des Grenzwertes/Grenzbereiches kommen. In diesem Fall bedarf es eines Mechanismus, der den über das Histogramm-Maximum abgeschätzten Grenzwert verschiebt bzw. neu bestimmt (vgl. Abbildung 5.11 und 5.12).

Dies lässt sich lösen, indem man die Interpretation des Histogramms mit einem unüberwachten Klassifizierungsverfahren (nicht zu verwechseln mit dem eigentlichen SVM-Klassifizierungsverfahren) an Kanal IR120 kombiniert. Eine Einteilung in relativ homogene Strahlungstemperaturbereiche lässt sich in diesem Zusammenhang mit einer Cluster-Analyse nach dem K-Means Verfahren umsetzen (vgl. Canty 2006, S. 206 ff.). Durch die Trennung in vier Cluster erhält man entsprechende Temperaturbereiche, wobei einer den Übergang zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ repräsentiert (vgl. Abbildung 5.11 und 5.12). Testläufe haben ergeben, dass vier unterschiedliche Cluster ausreichen sollten, um einen ungefähren Grenzbereich festzulegen. Welcher der Strahlungstemperaturbereiche (Cluster) den Übergang darstellt,

lässt sich durch die Analyse des Histogramms unter Berücksichtigung mehrerer Bedingungen (vgl. Tabelle 5.2) festlegen. Je nach erfüllten Bedingungen ergibt sich ein abgeschätzter Grenzwert aus der durchschnittlichen Strahlungstemperatur des ausgewählten Clusters.

Tab. 5.2: Histogramm-Bedingungen für die nachfolgende Auswahl des ungefähren Grenzwertes. T = Strahlungstemperatur Kanal IR120. *Quelle: Eigene Aufstellung*

Nr.	Bedingung	Ausgabe
1	T-Wert bei Histogramm-Maximum $< 0\text{ °C}$	0 oder 1
2	T-Wert bei Histogramm-Maximum $< -5\text{ °C}$	0 oder 1
3	Durchschnittliche Histogramm-Sprünge (ohne Glättung) > 50	0 oder 1
4	Histogramm-Maximum (Anzahl Pixel) > 600	0 oder 1
5	T-Maximum – T-Wert bei Histogramm-Maximum $> 10\text{ °C}$	0 oder 1

Die Bedingungen für die Auswahl des wahrscheinlichen Grenzbereiches (Cluster) weisen Eigenschaften auf, die darauf hinweisen, wo sich der Übergang zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ befindet. Liegen beispielsweise die Strahlungstemperaturwerte des Histogramm-Maximums (Anzahl Pixel) unterhalb von 0 °C , oder sogar unterhalb von -5 °C , so spricht dies mit hoher Wahrscheinlichkeit für Bewölkung und der eigentliche Grenzbereich liegt bei Werten oberhalb 0 °C oder -5 °C . Auf der anderen Seite ist es möglich, dass das Histogramm-Maximum auf eine sehr kalte wolkenfreie Fläche deutet, wobei keine Verschiebung des Grenzbereiches notwendig wäre (vgl. Abbildung 5.12b). Aus diesem Grund erfolgt die Kombination mit weiteren Bedingungen, wie der Pixel-Anzahl des Histogramm-Maximums oder starken Sprüngen innerhalb des Histogramms. Liegt beispielsweise der Wert des Histogramm-Maximums, bezogen auf die ausmaskierte Szene Großraum Deutschland, bei über 600 Pixeln, spricht dies eher für wolkenfreie Verhältnisse, während niedrigere Werte oder eine hohe Anzahl an Sprüngen im Histogramm auf Wolken hinweisen (vgl. Abbildung 5.12a). Eine große Differenz zwischen dem Temperatur-Maximum der Gesamtszene und dem Temperaturwert am Histogramm-Maximum lässt wiederum eine allgemein erwärmte Situation der Landflächen und der Atmosphäre vermuten. Insbesondere niedrige Wolken können in solchen Situationen auch durch höhere Strahlungstemperaturen charakterisiert sein.

Je nach Kombination der Bedingungen (vgl. Tabelle 5.3) wird entsprechend die Durchschnittstemperatur eines anderen Clusters als Grenzwert herangezogen. Die durchschnittlichen Strahlungstemperaturwerte der Cluster werden in aufsteigender Reihenfolge betrachtet und entsprechend zugeordnet. Cluster 1 und 2 zeigen in der Regel die kältesten Werte und sind daher meist mit Bewölkung gekoppelt, weshalb nur Cluster 3 oder 4, je nach Kombination der Bedingungen als Grenzwert/Grenzbereich in Betracht kommen (vgl. Tabelle 5.3). Mit dieser Vorgehensweise wird eine Vielzahl an Situationen abgedeckt, und die Abschätzung des vorläufigen Grenzwertes sollte nicht zu übermäßigen Abweichungen führen. Einzig die Situationen

Tab. 5.3: Zuweisung der Cluster (Grenzwert/Grenzbereich) durch Interpretation des Histogramms (Bedingungen). Wenn Bedingung 2 erfüllt ist, muss Bedingung 1 ebenfalls erfüllt sein, wodurch die Anzahl möglicher Kombinationen reduziert wird. *Quelle: Eigene Aufstellung*

Histogramm-Bedingungen					Cluster
1	2	3	4	5	
0	0	0	0	0	3
1	0	0	0	0	3
0	0	1	0	0	3
1	1	1	1	1	4*
alle anderen Kombinationen					4

$$* + ((\text{Strahlungstemperatur-Maximum} - \text{Strahlungstemperatur-Mittelwert Cluster 4})/2)$$

einer annähernd vollkommenen Wolkenbedeckung bzw. einer absolut wolkenfreien Situation über Deutschland stellen die Ausnahme dar. Hier kann kein Grenzbereich/Grenzwert durch das zuvor beschriebene Verfahren abgeschätzt werden. Diese Gegebenheit muss bei der späteren SVM-Wolkendetektion bzw. -klassifizierung berücksichtigt werden. Zwar ist es für den Raum Deutschland bzw. Zentraleuropa überaus selten, dass eine der beiden Situationen in ihrer Absolutheit vorkommt, dennoch kann dies auch bei nicht vollständiger Ausprägung zu Ungenauigkeiten der Grenzwertabschätzung führen. Beispielsweise wenn bei einer ausgeprägten Hochdruckwetterlage vereinzelt Cumulus- oder Cirrus-Wolken auftreten, oder wenn bei großflächigen Front-Systemen die vollständige Wolkenbedeckung durch vereinzelt Wolkenlücken unterbrochen ist.

In den Abbildungen 5.11 und 5.12 sind einige Beispielsituationen der Abschätzung des ungefähren Grenzwertes dargestellt. Abbildung 5.11a zeigt eine Situation, bei der große Teile Deutschlands mit Wolken bedeckt sind. Die Landflächen sind bereits so erwärmt, dass sich Strahlungstemperaturen (Kanal IR120) von bis zu 20 °C ergeben. Obwohl das Histogramm-Maximum (Anzahl Pixel) im Bereich unter 0 °C liegt, befindet sich dort nicht der Grenzbereich zwischen „Wolkenkontaminiert“ und „Wolkenfrei“. Aufgrund eines zweiten Maximums bei ca. 11 °C, ist hier der eigentliche Grenzbereich zu vermuten. Bedingung 1 und 2 (vgl. Tabelle 5.2 und 5.3) sind zwar erfüllt, dies gilt aber auch für Bedingung 3 und 5. Eine Verschiebung bzw. Neubestimmung des ungefähren Grenzwertes, durch die Auswahl von Cluster 3, ist demnach erforderlich. Man erhält einen abgeschätzten Grenzwert von 11,12 °C. Im in Abbildung 5.12b gezeigten Beispiel sieht die Situation trotz ähnlicher Wolkengegebenheiten anders aus. Obwohl die Szene größtenteils wolkenfrei ist, erhält man überwiegend niedrige Strahlungstemperaturen. Hierdurch ergibt sich ein ausgeprägtes Maximum (Anzahl Pixel) im negativen Temperaturbe-

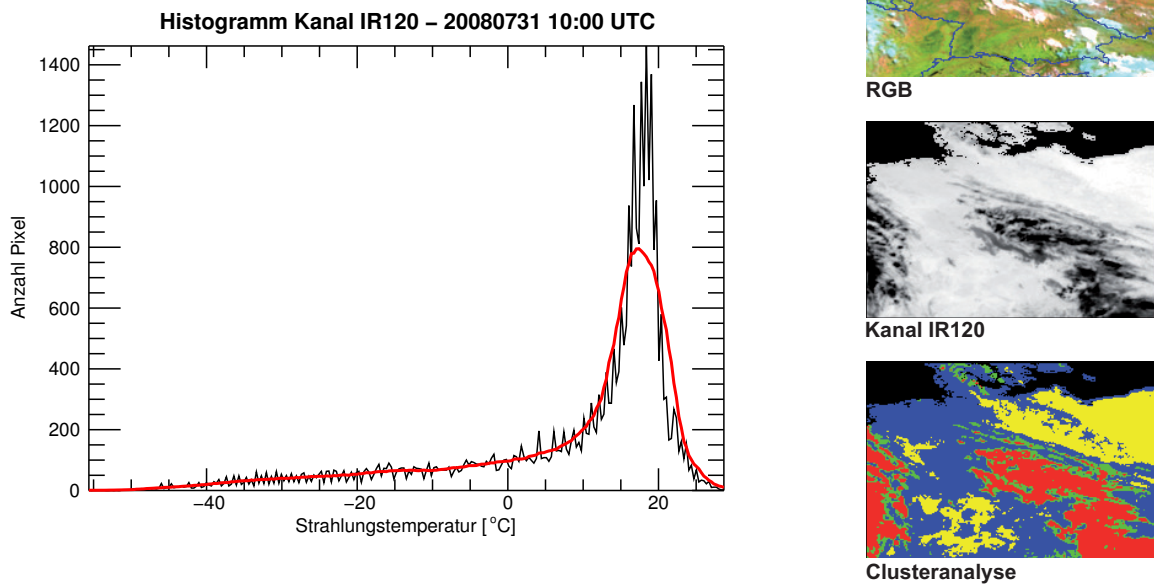
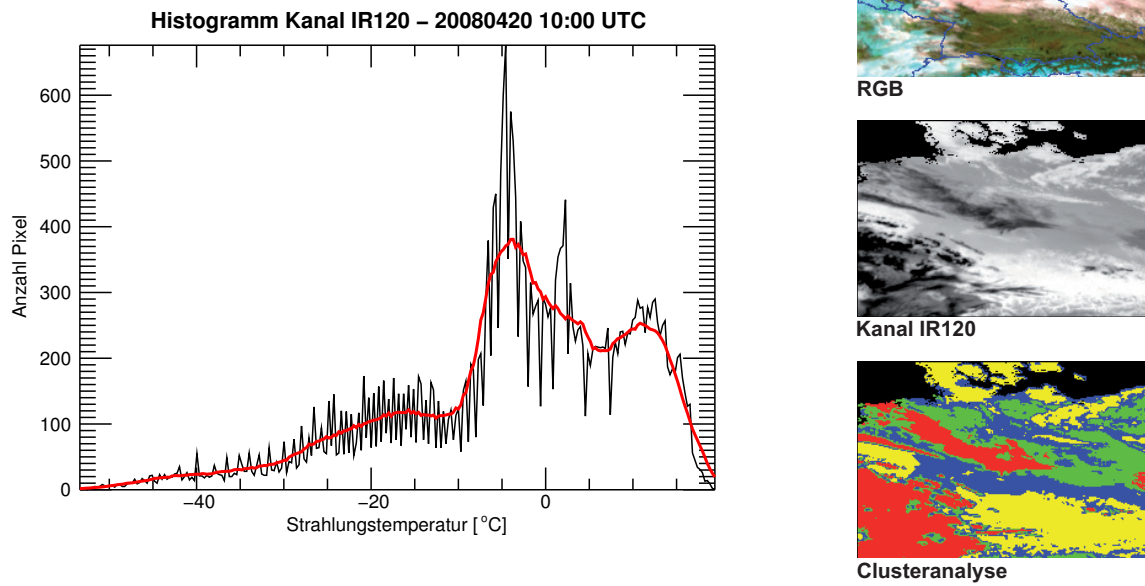
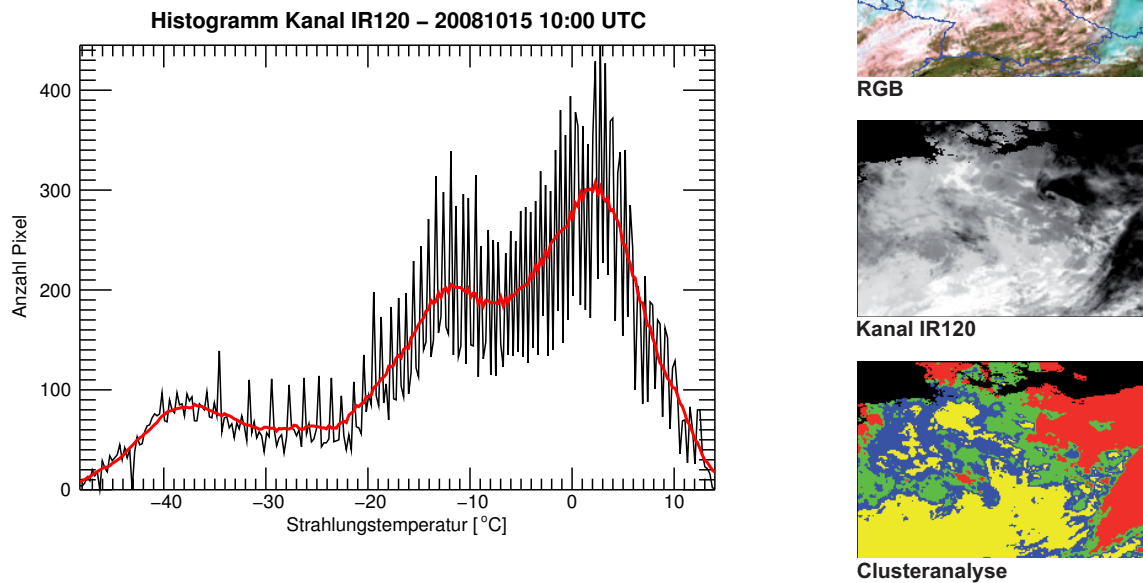
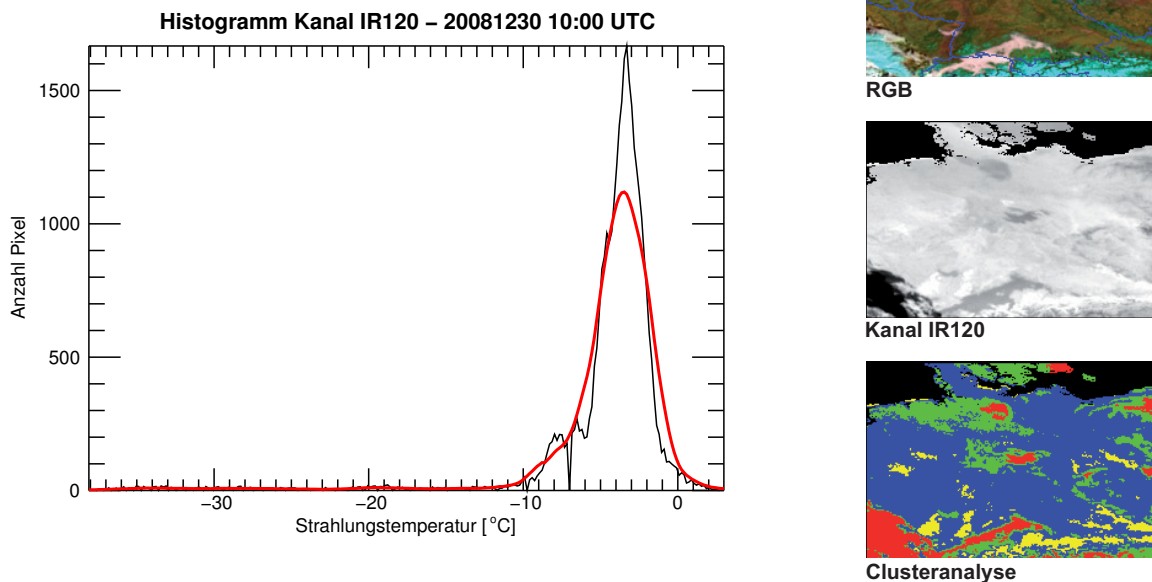


Abb. 5.11: Histogramme (Zeitschnitte aus April und Juli), inklusive Cluster-Analyse und Abschätzung eines ungefähren Grenzwertes nach den in Tabelle 5.2 aufgeführten Bedingungen. Cluster 1 (Rot), Cluster 2 (Grün), Cluster 3 (Blau) und Cluster 4 (Gelb). Die rote Linie im Histogramm stellt den gleitenden Mittelwert über 20 Datenpunkte dar. *Quelle: Eigene Berechnung*



(a) 15.10.2008 10:00 UTC - Geschätzter Grenzwert bei 4,85 °C



(b) 30.12.2008 10:00 UTC - Geschätzter Grenzwert bei -3,14 °C

Abb. 5.12: Histogramme (Zeitschnitte aus Oktober und Dezember), inklusive Cluster-Analyse und Abschätzung eines ungefähren Grenzwertes nach den in Tabelle 5.2 aufgeführten Bedingungen. Cluster 1 (Rot), Cluster 2 (Grün), Cluster 3 (Blau) und Cluster 4 (Gelb). Die rote Linie im Histogramm stellt den gleitenden Mittelwert über 20 Datenpunkte dar. *Quelle: Eigene Berechnung*

reich. Nur Bedingung 1 ist erfüllt. Cluster 3 wird demnach als Grenzbereich herangezogen. Dessen Durchschnittstemperatur liegt bei $-3,14\text{ }^{\circ}\text{C}$, nahe des eigentlichen Histogramm-Maximums.

Zusammenfassend sieht die Abschätzung des ungefähren Grenzwertes wie folgt aus. Zuerst wird das Histogramm von Kanal IR120 im Bereich von $-10\text{ }^{\circ}\text{C}$ bis zum Strahlungstemperaturmaximum betrachtet, wobei sich vor allem auf die Strahlungstemperatur des Histogramm-Maximums (Anzahl Pixel) in diesem Bereich konzentriert wird. In Verbindung mit einer K-Means Cluster-Analyse (4 Klassen) werden so, unter Berücksichtigung bestimmter Bedingungen, Cluster als Grenzwert/Grenzbereich zugeteilt. Letztendlich erhält man für jeden Zeitschnitt einen Grenzwert, der sich aus dem Mittelwert des zugeteilten Clusters ergibt. Dieser Grenzwert fließt in die nachfolgenden Berechnungen ein, und soll dafür sorgen, dass Trainingspixel nach Möglichkeit nicht falsch zugeordnet werden (vgl. Kapitel 5.4 und 5.5).

Die eigentliche Gewinnung der Trainingspixel erfolgt direkt vor der Berechnung der Wolkenmaske bzw. der weiter differenzierten Wolkenklassifizierung. Beide Verfahren werden in den nachfolgenden Kapiteln ausführlich beschrieben. Die aus den Wetter-Kameras abgeleiteten Trainingsinformationen und die Abschätzung des ungefähren Grenzwertes sind als Input für die weitere Prozessierung erforderlich (vgl. Abbildung 5.13).

5.4 Generierung weiterer Trainingsdaten und Anwendung der SVM-Wolkenmaske

Bevor eine differenzierte Wolkenklassifizierung durchgeführt werden kann, erfolgt die Unterscheidung zwischen wolkenkontaminierten und wolkenfreien Pixeln. Dies geschieht separat für jeden Zeitschnitt mit der Erzeugung einer auf Support Vector Machines basierenden Wolkenmaske. Hierzu werden in erster Linie die wolkenrelevanten Infrarotkanäle von METEOSAT-SEVIRI (vgl. Tabelle 4.1) herangezogen, welche den mehrdimensionalen, für den SVM-Klassifikator notwendigen, Merkmalsraum aufspannen (siehe Abbildung 5.3). Aufgrund der in Kapitel 3.1 beschriebenen Besonderheit in Bezug auf Strahlungstemperaturen im Wellenlängenbereich um $3,9\mu\text{m}$, wird der Kanal IR039 innerhalb des Merkmalsraumes nicht berücksichtigt. Alle Kanäle des Merkmalsraumes werden in einer Datei als Stapel (Layerstack) zusammengefasst und dem Wolkendetektionsverfahren zugeführt (vgl. Abbildung 5.13). Die Zuteilung notwendiger Trainingsdaten erfolgt über Bildpunkte innerhalb der in ENVI/IDL verfügbaren Regions of Interest (ROIs). Solche ROIs können über die Pixelposition mit dem jeweiligen Layerstack verbunden werden, was mit den Funktionen von ImageSVM umgesetzt werden kann. Nach der Definition der zwei ROIs „Wolkenkontaminiert“ und „Wolkenfrei“, fließen alle Pixel, die einem der beiden ROIs zugeordnet wurden, als Trainingspixel in das Klassifizierungsverfahren bzw. die Wolkendetektion ein.

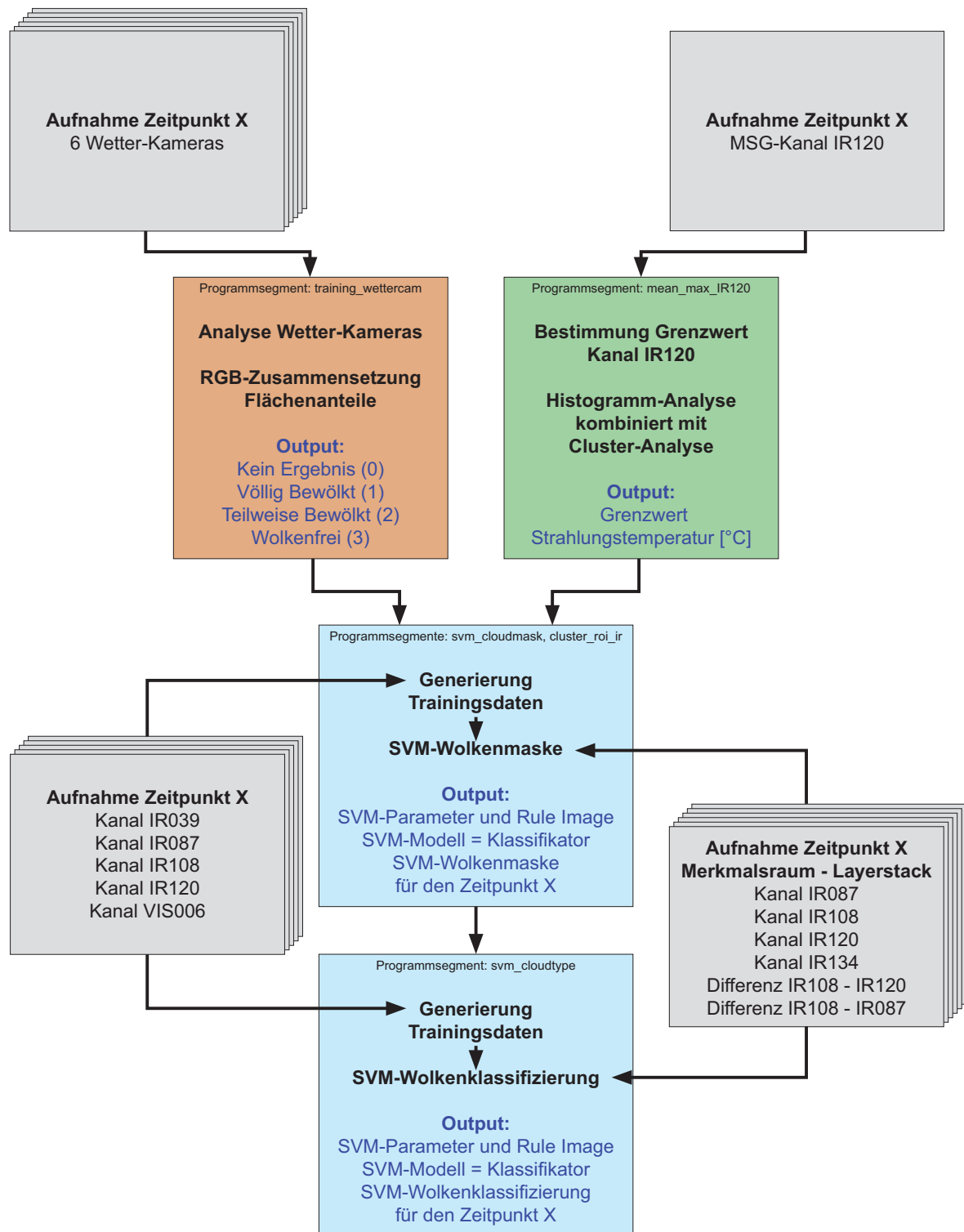


Abb. 5.13: Ablaufschema des Wolkendetektions- und Wolkenklassifizierungsverfahrens, inklusive Trainingsdatengewinnung für jeden METEOSAT-Zeitschnitt . *Quelle: Eigener Entwurf*

Die erste Zuweisung erfolgt über die Informationen der Wetter-Kameras (vgl. Kapitel 5.3.1 und Abbildung 5.13), welche über ihre Koordinaten eindeutig mit bestimmten Pixeln der Szene verknüpft werden können (vgl. Abbildung 5.7, Tabelle 5.1 und Seite 72). Das heißt, über die vorab durchgeführte Analyse der Wetter-Kameras lassen sich, anhand von Zahlenwerten (vgl. Abbildung 5.1 und 5.13), die entsprechenden Pixel dem jeweiligen ROI zuteilen. Um den Umfang dieser Trainingsinformation zusätzlich zu erhöhen, wird das Strahlungsverhalten des Pixels mit seinen benachbarten Pixeln verglichen. Hierzu wird eine Fläche von 11 x 11 Bildpunkten um das zentrale Pixel der entsprechenden Wetter-Kamera erzeugt. Weisen die benachbarten Bildpunkte ein ähnliches Strahlungsverhalten bei Kanal IR120 auf (Toleranz von 0,5 °C bzw. 0,5 K), so ist davon auszugehen, dass sie derselben Klasse angehören und als Trainingspixel eingesetzt werden können. Da es sich bei der Abschätzung durch die Wetter-Kameras, wie zuvor beschrieben, um ein automatisiertes Verfahren handelt, besteht natürlich auch hier die Gefahr einer Falschzuweisung. Einzelne komplett falsch zugeordnete Pixel können in der endgültigen Klassifizierung zu erheblichen Ungenauigkeiten führen. Um solche Falschzuweisungen weitestgehend zu vermeiden, werden die IR120-Werte an den Pixeln der Wetter-Kamera-Standorte mit dem zuvor abgeschätzten groben Grenzwert (vgl. Kapitel 5.3.2) verglichen. Falls es hier, je nach Klasse, zu Überschreitungen oder Unterschreitungen kommt, wird das Pixel nicht als Trainingspixel berücksichtigt. Sind Pixel nach dieser Überprüfung eindeutig durch die Wetter-Kameras zugeteilt, so wird dies im weiteren Verfahren nicht mehr geändert.

Trotz Erweiterung durch benachbarte Pixel ist die Menge der Trainingsdaten für eine effektive Detektion bzw. Klassifizierung nicht ausreichend. Daher bedarf es eines erweiterten Ansatzes, durch den zusätzliche Trainingsdaten für jede Einzelszene gewonnen werden können. Auch hierbei wird eine automatisierte Herangehensweise angestrebt. Eine Analyse der umfangreichen Datenreihen mittels Expertenwissen erscheint in dieser Hinsicht nicht praktikabel. Daher bezieht sich die Vorgehensweise, wie zuvor erwähnt, auf ein zusätzliches an Schwellenwertalgorithmen angelehntes Verfahren. Durch die Kombination bekannter Schwellenwertalgorithmen sollen weitere Trainingsdaten gewonnen werden (vgl. hierzu Kapitel 3). Normalerweise ist die größte Herausforderung solcher Verfahren die exakte Bestimmung der erforderlichen Schwellenwerte, was in Bezug auf eine SVM-Klassifizierung weniger relevant ist, da die eigentliche Trennung durch die spätere Einbindung der Hyperebene erfolgt (vgl. Kapitel 5.1). Die Trainingsdaten sollten jedoch so nah wie möglich an den Grenzbereich heran reichen und diesen nach Möglichkeit nicht in die andere Richtung überschreiten, auch wenn der SVM-Klassifikator mit einigen wenigen Überschreitungen zurecht kommen sollte (vgl. Foody und Mathur 2006).

Die Idee der Trainingsdatengewinnung durch Schwellenwertansätze bezieht sich daher nicht auf streng definierte Schwellenwerte. Nur der zuvor berechnete ungefähre Grenzwert/Grenzbereich wird berücksichtigt. Diesem Grenzwert nähert man sich durch die Kombination bekannter

Tab. 5.4: Tag-Zeitfenster in den Kanälen des sichtbaren Wellenlängenbereiches für die Trainingsdatengewinnung eingesetzt werden können. Die Auswahl erfolgt durch die Betrachtung des verwendeten Ausschnitts Deutschland (vgl. Abbildung 5.2), welcher voll ausgeleuchtet sein muss. *Quelle: Eigene Aufstellung*

Monat	Tag-Anfang (UTC)	Tag-Ende (UTC)
Jan/Dez	10:00	12:30
Feb/Nov	09:00	13:00
Mär/Okt	08:00	14:00
Apr/Sep	07:00	15:00
Mai/Jun/Jul/Aug	06:00	16:00

Schwellenwertalgorithmen (Test-Sequenzen, vgl. Abbildung 3.3), wobei eine Überschreitung vermieden wird, ähnlich wie es in Abbildung 5.10 für Kanal IR120 dargestellt ist.

Die Herangehensweise sieht wie folgt aus. Alle für die verwendeten Schwellenwertalgorithmen (Tests) erforderlichen Kanäle bzw. Kanal-Differenzen (nicht zu verwechseln mit dem Merkmalsraum, vgl. Abbildung 5.13) werden bezüglich ihrer Daten ausgelesen und aufsteigend sortiert, so dass sich ein Datenfeld (Array) mit aufsteigenden Strahlungstemperaturen oder Reflektanzen für jeden Kanal bzw. jede Kanal-Differenz ergibt. Auf Basis dieser sortierten Bilddaten können dann verschiedene Algorithmen angewendet werden, jedoch nicht bezogen auf einen festen Schwellenwert, sondern im Sinne der *Anzahl niedrigster oder höchster Werte*, je nach Klasse und verwendetem Test. Durch Verknüpfung der verschiedenen Tests in Kombination mit dem in Kapitel 5.3.2 abgeschätzten Grenzwert, lässt sich ein, für jede Einzelszene, relativ repräsentativer Trainingsdatensatz erzeugen. Dieser sollte auch in die Reichweite des eigentlichen Grenzbereiches vordringen. Gleichzeitig kann durch Festlegung einer bestimmten Anzahl von Trainingspixeln deren Umfang reduziert werden, was sich positiv auf die Rechenzeit auswirkt. Das Verfahren der zusätzlichen Trainingsdatengewinnung für die SVM-Wolkenmaske ist in Abbildung 5.14 dargestellt.

In Bezug auf die Klasse „Wolkenkontaminiert“, lassen sich durch die Kombination diverser Schwellenwert-Tests verschiedene Wolkenarten erfassen, die als Trainingsdaten eingesetzt werden können. Beispielsweise kann schon durch den allgemeinen Wolken-Test bei Kanal IR120 ein Großteil der Bewölkung erfasst werden (vgl. Kapitel 3 und Abbildung 5.14). Da die genaue Grenze zwischen „Wolkenkontaminiert“ und „Wolkenfrei“ nicht bekannt ist, werden jedoch nur die geringsten Werte verwendet, bei denen mit sehr hoher Sicherheit von Bewölkung ausgegangen werden kann. Man nähert sich sozusagen anhand der sortierten Bilddaten dem Grenzbereich von der „kalten“ Seite aus. Nun könnte man vermuten, dass dies ausreicht, wenn man alle Werte bis zu dem in Kapitel 5.3.2 abgeschätzten Grenzwert als Trainingsdaten für die Klasse „Wolkenkontaminiert“ heranzieht. Dies ist jedoch nicht zu empfehlen, da dieser Grenzwert nur ein grobes Maß darstellt, um Überschreitungen zur anderen Klasse zu vermeiden oder

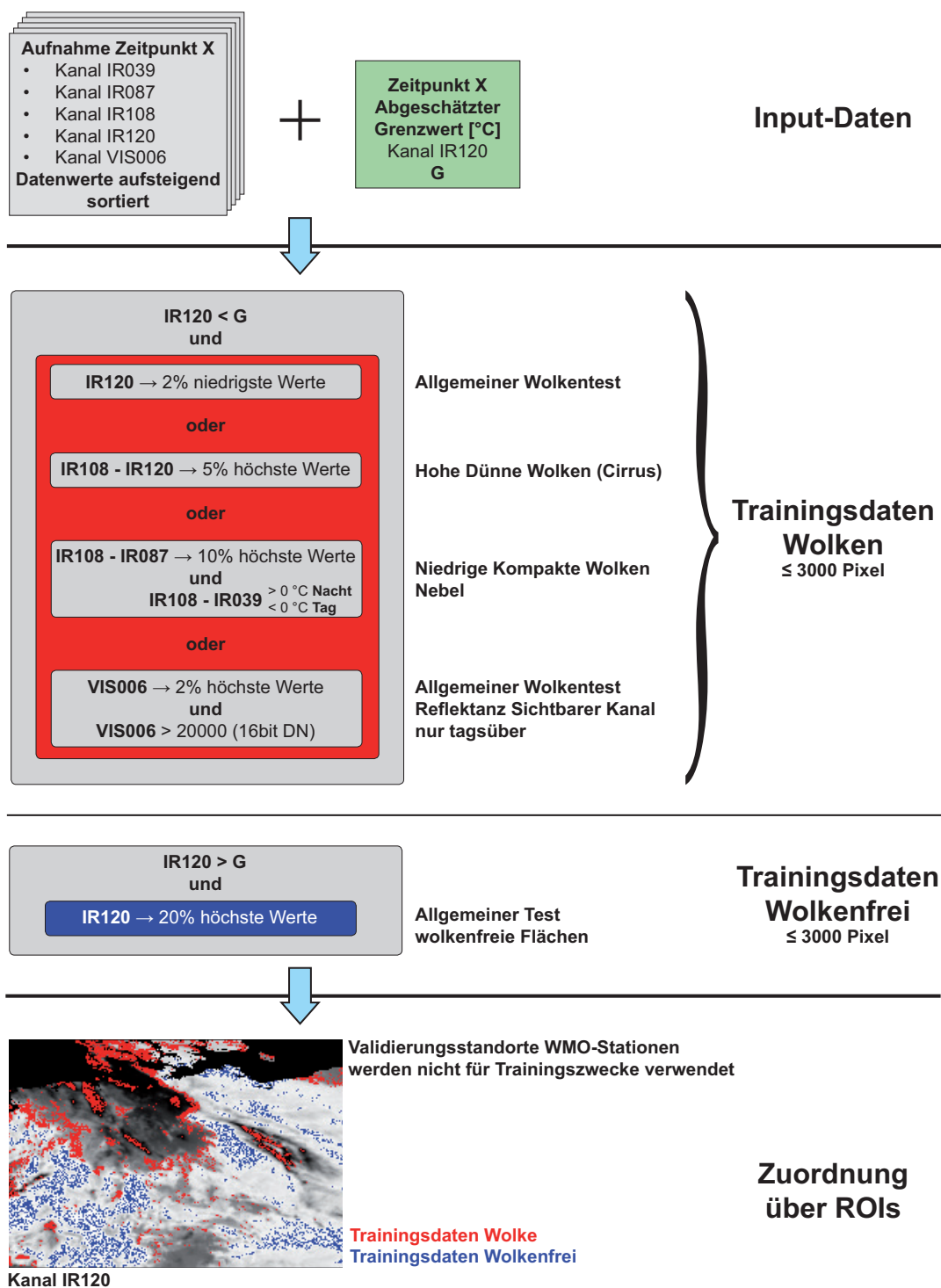


Abb. 5.14: Erzeugung weiterer Trainingsdaten mit Hilfe logischer Verknüpfungen in Bezug auf bekannte Schwellenwertalgorithmen, inklusive Zuordnung über Regions of Interest (ROIs). Umfang der niedrigsten und höchsten Werte und die Maximale Anzahl der Trainingspixel sind auf den Gesamtumfang der Szene Deutschland abgestimmt. *Quelle: Eigener Entwurf*

gering zu halten. Außerdem würde, je nach Umfang der Bewölkung die Zahl der Trainingspixel enorme Ausmaße annehmen. Aus diesem Grund findet, wie in Abbildung 5.14 aufgeführt, eine Verknüpfung mit weiteren Algorithmen (Tests) statt. Dies geschieht nicht nur, um die Trainingsdatenmenge für jeden Zeitschnitt gering zu halten, sondern auch um unterschiedlich vorkommende Wolkenarten abdecken und sich dem Grenzbereich besser nähern zu können. Das heißt, bestimmte Pixel werden nur dann als Trainingspixel herangezogen, wenn eine Kombination von Tests erfüllt ist, wobei sich der jeweilige „Schwellenwert“ auf eine bestimmte Anzahl niedrigster bzw. höchster Datenwerte bezieht. Beispielsweise lassen sich niedrige, relativ warme Wolken über die Differenz IR108 - IR087 oder IR108 - IR039 (Umkehr zwischen Tag und Nacht) erfassen, während hohe, dünne Cirrus-Wolken mit der Differenz IR108 - IR120 abgeschätzt werden können (vgl. hierzu Kapitel 3.2.1). Je nach Test Anzahl niedrigster bzw. höchster Datenwerte (Prozentwerte in Abbildung 5.14) an die Gesamtpixelmenge (je Kanal) der ausmaskierten Szene Deutschland angepasst. Wird die Anzahl der Trainingspixel für die Klasse „Wolkenkontaminiert“ (ohne Pixel der Wetter-Kameras) von 3000 dennoch überschritten, so erfolgt eine Begrenzung mit Hilfe einer Zufallsauswahl. Die Verwendung des sich im sichtbaren Wellenlängenbereich befindlichen Kanals VIS006 (siehe Abbildung 5.14) erfolgt nur tagsüber bei vollständig ausgeleuchteter Szene. Eine Einordnung der möglichen Zeitpunkte im Jahresverlauf wurde durch visuelle Begutachtung abgeschätzt und ist in Tabelle 5.4 dargestellt. Ähnlich wie bei den Bildern der Wetter-Kameras waren Sonnenauf- und Sonnenuntergangszeiten in diesem Zusammenhang nicht hilfreich, da die Krümmung der Erde zu Schatten über Teilen von Deutschland führt.

Im Vergleich mit der Klasse „Wolkenkontaminiert“ werden bei der Klasse „Wolkenfrei“ lediglich die höchsten Werte von Kanal IR120 herangezogen. Auch hier unter Berücksichtigung des in Kapitel 5.3.2 beschriebenen ungefähren Grenzwertes. In diesem Fall tritt genau dieselbe Problematik auf, wie zuvor bei der Klasse „Wolkenkontaminiert“. Es muss damit gerechnet werden, dass der Umfang der Trainingspixel enorm ansteigt. Da keine weitere Kombination von Schwellenwert-Tests in Bezug auf die Klasse „Wolkenfrei“ zur Verfügung steht, erfolgt auch hier eine Zufallsauswahl für den Fall, dass die Anzahl von 3000 Pixeln überschritten wird.

In Kapitel 5.3.2 wurde in Bezug auf die Abschätzung des ungefähren Grenzwertes schon auf die Problematik einer völlig bewölkten bzw. völlig wolkenfreien Situation über der Gesamtfläche Deutschlands hingewiesen, welche jedoch höchst selten vorkommt. Dieser Fall macht die Abschätzung eines Grenzbereiches im Sinne einer Wolkendetektion natürlich überflüssig. Schwierigkeiten entstehen aber auch dann, wenn eine Klasse innerhalb der Szene nur sehr gering ausgeprägt ist. Das Verfahren des „Ungefähren Grenzwertes“ ist zwar so ausgelegt, dass sich dieser sehr weit in eine Richtung verschieben bzw. neu bestimmen lässt (vgl. Abbildung 5.11 und 5.12), dennoch besteht die Möglichkeit einer Falschabschätzung. Insbesondere bei ausgeprägten wolkenfreien Flächen kann es zur falschen Zuweisung von Trainingspixeln der

Klasse „Wolkenkontaminiert“ kommen, da das Verfahren in erster Linie bestrebt ist, Wolkenpixel als Trainingsdaten zu erfassen. Aber auch in umgekehrter Richtung besteht diese Gefahr, bei wenigen Wolkenlücken in einer fast vollständig bewölkten Szene.

Weiterhin sei erwähnt, dass alle Pixel/Standorte, die später bei der Validierung zum Einsatz kommen, nicht als Trainingspixel verwendet werden (DWDSAT, WMO-Stationsdaten vgl. Kapitel 6 und Abbildung 6.1).

Für das eigentliche Klassifizierungsverfahren muss neben den Trainingsdaten, die über die zwei vorab definierten ROIs „Wolkenkontaminiert“ und „Wolkenfrei“ zugeordnet werden, der Merkmalsraum als Layerstack zur Verfügung stehen (wolkenrelevante Kanäle, siehe Abbildung 5.3). Um eine effektive Klassifizierung zu gewährleisten, sollte dies in skaliert Form (vgl. Kapitel 5.2 und Hsu et al. 2003) geschehen. Alle fehlerhaften und ausmaskierten Pixel werden mit NaN-Werten versehen und dürfen nicht zu Trainings- bzw. Klassifizierungszwecken herangezogen werden. Solche Werte bilden demnach die Kategorie „Unclassified“. Abschließend werden für die jeweilige Einzelszene die von imageSVM verfügbaren Funktionen und Prozeduren in sequentieller Reihenfolge durchlaufen.

An erster Stelle erfolgt die Grid-Search-Prozedur in Verbindung mit der Auswahl von Kernel- und Regularisierungsparametern. Die Auswahl dieser Parameter hängt unter anderem von der Verteilung der Klassen innerhalb des Merkmalsraumes ab und sollte nach Möglichkeit durch mehrfache Durchführung des Trainings (mit jeweils veränderten Parametern) für das entsprechende Klassifizierungsproblem angepasst werden. In Bezug auf eine Wolkendetektion bzw. Wolkenklassifizierung von immer neuen Einzelszenen ist dies jedoch vom Aufwand her nicht zu realisieren. Aus diesem Grund wird in dieser Arbeit auf die von imageSVM empfohlenen Parameter zurückgegriffen (siehe hierzu Janz et al. 2007, Geomatics-Department 2007). Im Rahmen der Kernel- und Regularisierungsparameter lassen sich zwei SVM-Typen auswählen, C und ν . Durch diese kommt es zu unterschiedlichem Einfluss der Trainingsdaten auf die Einbindung der Hyperebene. Eine stichprobenartige Begutachtung hat gezeigt, dass mit C-SVM gute Ergebnisse erzielt werden können (natürlich in Abhängigkeit der ausgewählten Trainingsdaten). Die erzeugten Daten der internen Cross-Validation (vgl. Kapitel 5.2) zeigen ebenfalls, dass eine sinnvolle Trennung mit Hilfe der empfohlenen Parameter möglich ist. Um die Rechenperformance zu erhöhen, wurde nur der sogenannte Multiplier innerhalb der Kernel-Parameter von 10 auf 100 gesetzt.

Im Anschluss erfolgt das Training des Klassifikators (SVM-Modell), und man erhält abschließend die Wolkenmaske als Bild-Datei mit der Einteilung „Wolkenkontaminiert“ (Zahlenwert 1), „Wolkenfrei“ (2) und „Unclassified“ (0). Sie dient als Grundlage für die weitere Differenzierung im Rahmen der automatisierten SVM-Wolkenklassifizierung.

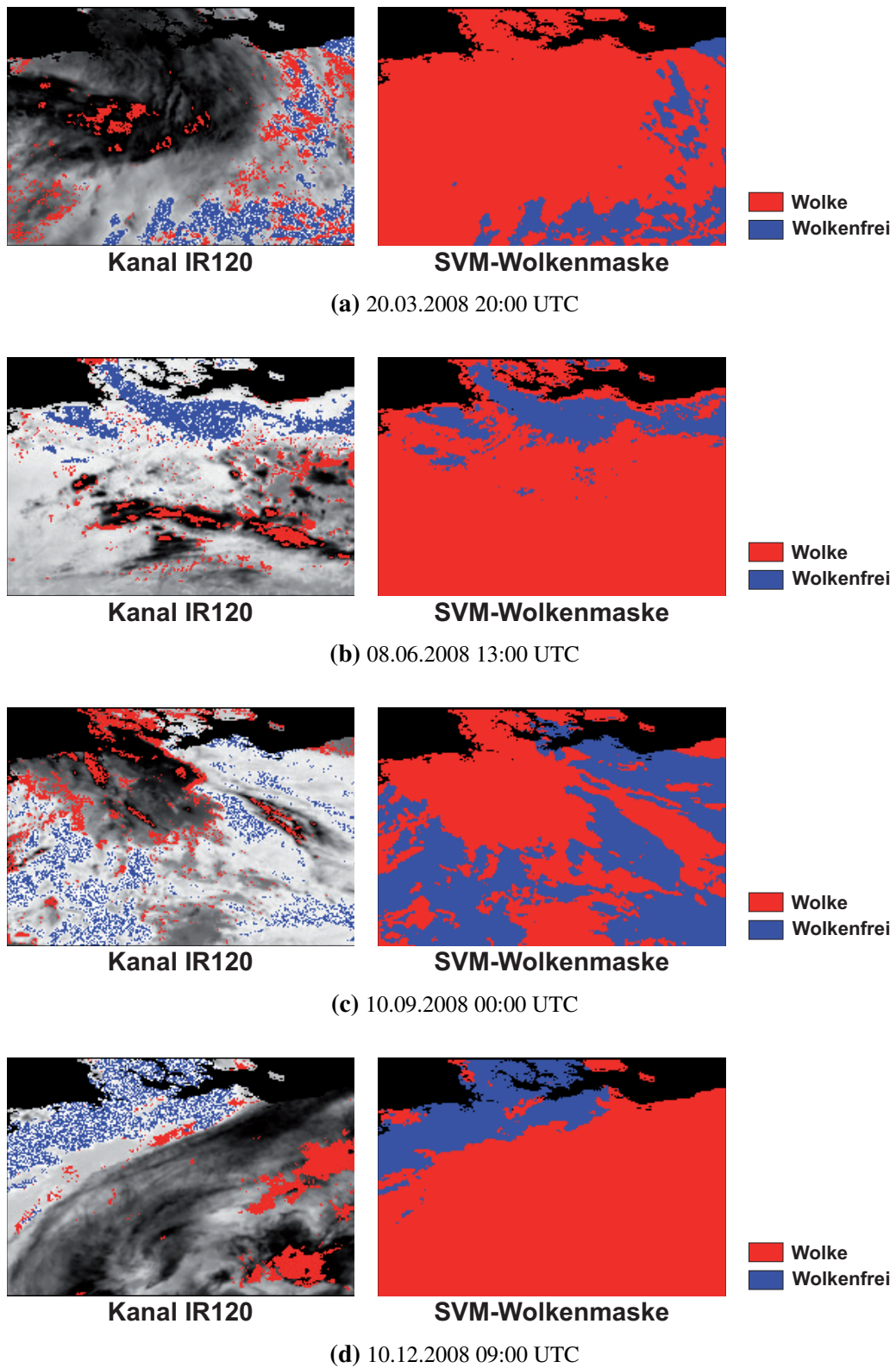


Abb. 5.15: Beispielhafte Darstellung verwendeter Trainingspixel der SVM-Wolkenmaske. *Quelle: Eigene Darstellung*

5.5 Anwendung der weiter differenzierten SVM-Wolkenklassifizierung

Die weiterführende Wolkenklassifizierung auf Basis der SVM-Wolkenmaske erfolgt für jeden Zeitschnitt mit einer ähnlichen Vorgehensweise wie bei der SVM-Wolkenmaske selbst. Zuerst werden die Trainingsdaten der einzelnen Klassen bestimmt und mit Hilfe von ROIs zugeordnet. Nachfolgend wird die SVM-Wolkenklassifizierung mit den durch imageSVM zur Verfügung stehenden Funktionen durchgeführt. Als Input der SVM-Klassifizierung dient, neben den Trainingsdaten, der sich aus den wolkenrelevanten Einzelkanälen und Kanal-Differenzen zusammensetzende Merkmalsraum (vgl. Abbildung 5.3). Auf alle Kanäle des Stapels (Layerstack) wird jedoch die zuvor erstellte Wolkenmaske angewendet, so dass nur Daten bezüglich der Wolken in die weitere Rechnung einfließen. Wolkenfreie Flächen werden nicht mehr berücksichtigt. Dies wird auch hier mit Hilfe von NaN-Werten realisiert.

Im Vergleich mit der Wolkenklassifizierung der SAFNWC (Cloud Type bzw. SAFNWC-Cloudtype, vgl. Kapitel 3.2.2), bei der zwischen 19 Wolken-Kategorien unterschieden wird, lässt sich mit dem SVM-Klassifizierungsverfahren, das auf der reinen Bildinformation der METEOSAT-Kanäle basiert, nur eine geringere Anzahl von Klassen abgrenzen. Trainingsdaten müssten in hohem Umfang und mit hoher Sicherheit für alle Klassen zur Verfügung stehen, was bei vorliegendem Verfahren nicht ohne weiteres möglich ist.

Durch die alleinige Betrachtung des oberen Teils der Bewölkung und die zur Verfügung stehenden Kanäle lässt sich dennoch eine sinnvolle Trennung in vier Wolkenklassen durchführen. Diese Klassen beziehen sich vor allem auf unterschiedliche Höhen und Strahlungsdurchlässigkeiten der Wolken. Zusammen mit der Klasse „Wolkenfrei“ erhält man fünf Klassen, die einen Überblick der Bewölkungssituation des jeweiligen Zeitschnitts darstellen.

- *Hoch Transparent* - Dünne durchlässige Wolken in großer Höhe (Cirrus-Niveau)
- *Hoch Kompakt* - Kompakte undurchlässige Wolken in großer Höhe (Cirrus-Niveau)
- *Niedrig Transparent* - Teilweise und transparente Bewölkung in niedrigen bis mittleren Höhen
- *Niedrig Kompakt* - Kompakte dichte Bewölkung in niedrigen bis mittleren Höhen
- *Wolkenfrei*

Eine Verwendung der Wetter-Kameras findet in Bezug auf die Trainingsdatengewinnung der differenzierten Wolkenklassifizierung nicht mehr statt. Versuche, Informationen unterschiedlicher Wolkenarten aus den Aufnahmen der Wetter-Kameras mittels Erkennungsverfahren zu extrahieren stellten sich als zu aufwendig bzw. fehlerhaft dar. Hinzu kommt die Problematik der

Betrachtungsgeometrie. Der Satellit erfasst Wolken von oben, während die Wetter-Kameras eine Blickrichtung seitlich bzw. von unten aufweisen. Bei der Unterscheidung zwischen wolkenkontaminierten und wolkenfreien Situationen scheint dies nicht so gravierend zu sein, bezogen auf eine weitere Differenzierung ist dies aber mit Problemen verbunden, da gewisse Wolkenarten nur schwer zu unterscheiden sind.

Die Gewinnung von Trainingsdaten erfolgt demnach ausschließlich auf Basis der Kombination verschiedener Schwellenwert-Tests (vgl. Kapitel 3 und Abbildung 5.16). Zu diesem Zweck werden wie bei der SVM-Wolkenmaske bestimmte Kanäle bzw. Kanal-Differenzen herangezogen, die als Grundlage für die Einzeltests dienen (vgl. hierzu Abbildung 5.13 und 5.16). Diese Daten werden vor der eigentlichen Berechnung aufsteigend sortiert, um die niedrigsten und höchsten Datenwerte verwenden zu können. Auch hier stellt die jeweilige Anzahl niedrigster und höchster Werte sozusagen den „Schwellenwert“ dar. Zusätzlich werden für die Trainingsdatengewinnung unterschiedlicher Höhengniveaus auch explizite Schwellenwerte festgelegt. Der in Kapitel 5.3.2 beschriebene ungefähre Grenzwert kommt nur noch bedingt zum Einsatz, da er sich vorrangig auf den Übergang zwischen wolkenkontaminierten und wolkenfreien Flächen bezieht.

Um bei der Trainingsdatengewinnung teilweise bewölkte Pixel von völlig bewölkten Pixeln unterscheiden zu können, bedarf es weiterer Informationen über die Homogenität der jeweiligen Einzelszene. Hierzu wird die Standardabweichung eines Pixels im Vergleich mit seinen benachbarten Bildpunkten berechnet. Dies geschieht mit Hilfe eines 3 x 3 Moving-Windows, angewendet auf die gesamte Szene des Kanals IR120, woraus abschließend ein neues Bild mit Standardabweichungswerten erzeugt wird. Jedes Pixel erhält somit einen Wert, der auf Homogenität schließen lässt. Diese Information fließt als separater Input-Kanal in den Prozess der Trainingsdatengewinnung mit ein (vgl. Abbildung 5.16).

Die Trainingsdaten der Klasse „Hoch Kompakt“ basieren im Wesentlichen auf der Strahlungstemperatur des Kanals IR120. Sehr geringe Werte weisen auf relativ hoch reichende Wolken hin. Damit die Trainingsdaten nicht in den Bereich des niedrigen Höhengniveaus gelangen, hat sich ein Schwellenwert von -20 °C als sinnvoll herausgestellt. Auch an kalten Wintertagen nehmen sehr niedrige Wolken in der Regel keine Strahlungstemperaturen unterhalb von -20 °C (Kanal IR120) an. Um den Umfang der Trainingsdaten gering zu halten und sich trotzdem dem Grenzbereich der anderen Klassen nähern zu können, kommen auch hier Verknüpfungen (bestimmte Bedingungen bzw. Tests müssen erfüllt sein) mit anderen Schwellenwert-Tests zum Einsatz (vgl. Abbildung 5.16). Beispielsweise lassen sich mit der Differenz IR108 - IR120 transparente Cirrus-Wolken erkennen (vgl. Kapitel 3), weshalb in Bezug auf die Klasse „Hoch Kompakt“ nur geringste Werte verwendet werden dürfen. Ähnlich verhält es sich mit der Differenz IR108 - IR087, die auf kompakte Wasserwolken hindeutet. Hier bezieht man sich entsprechend auf hohe Werte. Die in Verbindung mit der Algorithmen-Kombination verwendeten

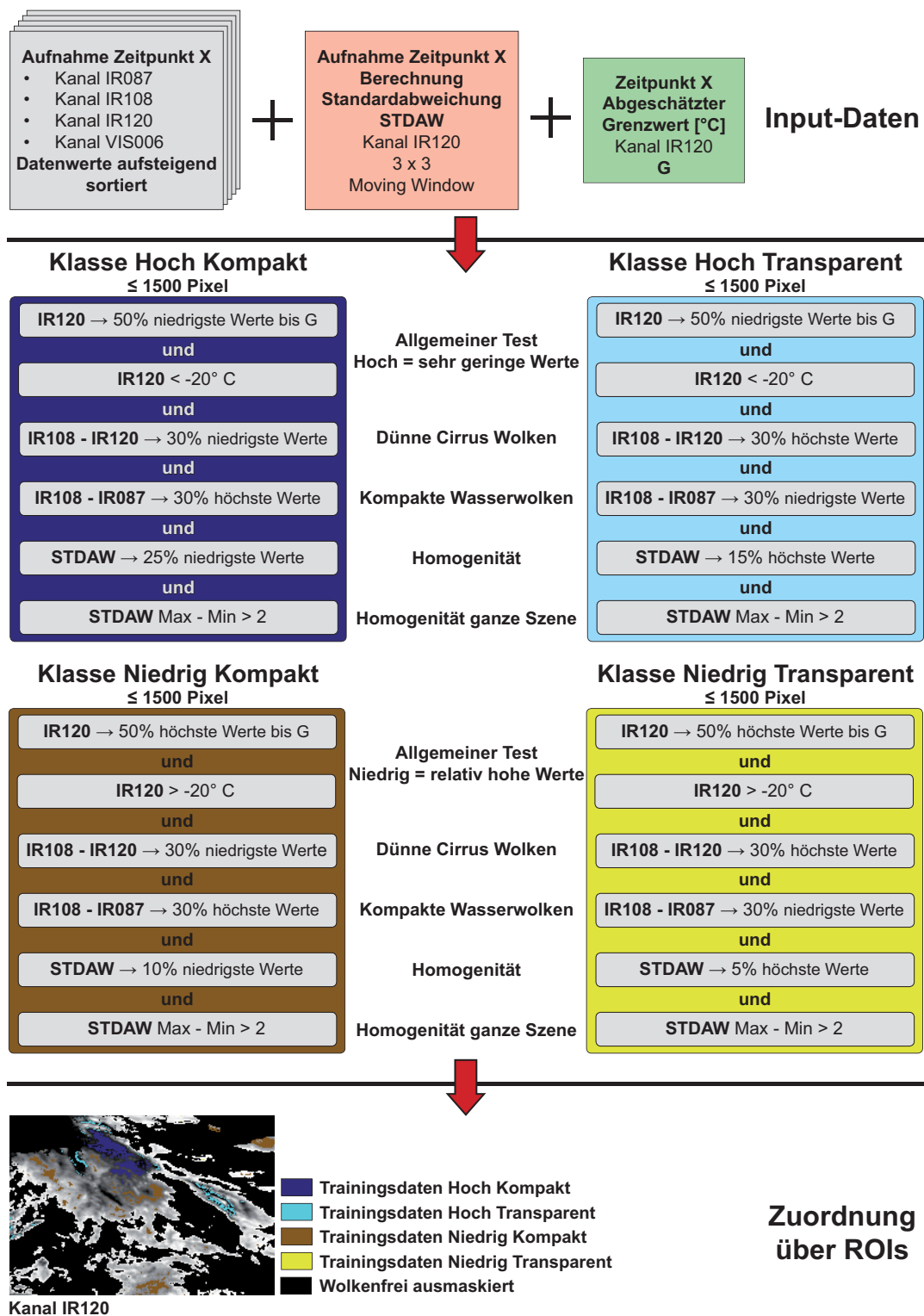


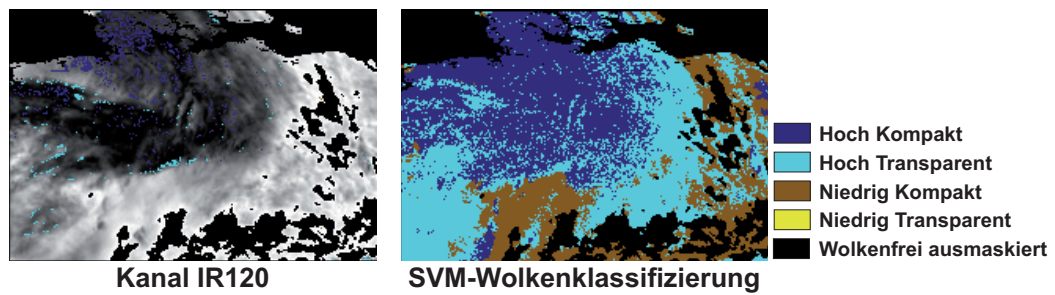
Abb. 5.16: Erzeugung von Trainingsdaten für die SVM-Klassifizierung mit Hilfe verknüpfter Schwellenwertalgorithmen. Zuordnung der vier weiteren Klassen erfolgt über Regions of Interest (ROIs). Umfang der kleinsten und größten Werte sowie die maximale Anzahl an Trainingspixeln sind auf den Gesamtumfang der Szene Deutschland abgestimmt. *Quelle: Eigener Entwurf*

„Schwellenwerte“, also im eigentlichen Sinne der Anteil niedrigster bzw. höchster Werte, sind auch hier, wie bei der SVM-Wolkenmaske, entsprechend angepasst. Sollte eine Klasse nur in geringem Maße oder gar nicht vorkommen, so wird durch die Kombination der Tests erreicht, dass auch nur wenige oder gar keine Pixel als Trainingsdaten erfasst werden. In Bezug auf die homogenen Eigenschaften der Klasse „Hoch Kompakt“ sind zudem die vorab berechneten Standardabweichungswerte von Interesse (siehe Seite 111). Geringe Werte sprechen für gleichmäßig kompakte Bewölkung.

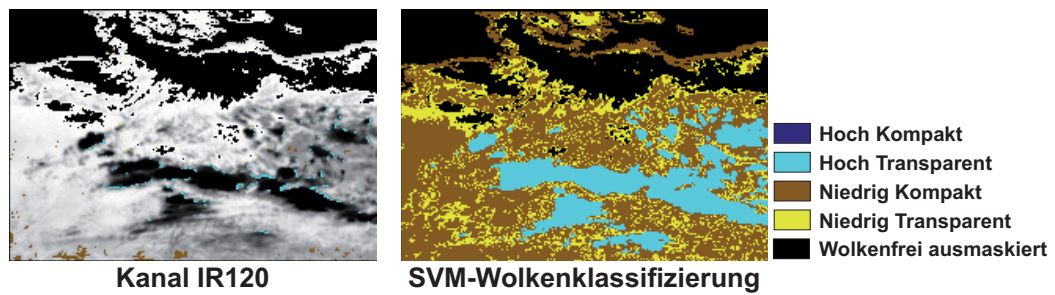
Bei den Trainingsdaten der Klasse „Hoch Transparent“ ist die Vorgehensweise ähnlich, sehr niedrige Werte, jenseits von -20 °C (Kanal IR120) lassen auf hohe Wolken schließen. Im Gegensatz zur Klasse „Hoch Kompakt“ zeigt die Betrachtung der Kanal-Differenzen jedoch eine andere Situation. Die Differenz IR108 - IR120 sollte für die Trainingsdatengewinnung dieser Klasse nur hohe Werte aufweisen, während die Differenz IR108 - IR087 niedrigste Werte voraussetzt. Auch die Betrachtung der Homogenität verhält sich entgegengesetzt. Hohe Standardabweichungswerte repräsentieren inhomogene Verhältnisse. Problematisch sind in diesem Zusammenhang Überlagerungen bestimmter Wolken in unterschiedlichen Höhen. Da der Satellit nur mit der Blickrichtung von oben arbeitet, können hohe dünne Wolken bzw. Wolkenfragmente, die niedrige kompakte Wolken überlagern, auch nur als solche erfasst werden, obwohl der Blick von unten eine kompakte Wolkendecke erkennen lässt. Dieses Problem wird im Rahmen der Validierung noch ausführlicher diskutiert.

Die Trainingsdatengewinnung der Klassen „Niedrig Kompakt“ und „Niedrig Transparent“ bezieht sich ebenfalls vorrangig auf Werte des Kanals IR120. Hohe Werte sprechen für niedrige Wolken. Durch den fixen Schwellenwert von -20 °C werden Überschreitungen in die Klassen der hohen Wolken weitestgehend vermieden. Eine Differenzierung zwischen transparenter bzw. teilweiser und kompakter Bewölkung erfolgt auch hier über Standardabweichungswerte und die gegensätzliche Verwendung der beiden Kanal-Differenzen IR108 - IR120 und IR108 - IR087. Die Differenz IR108 - IR120 wird zwar eigentlich nur bei hohen dünnen Wolken angewendet, sollte aber auch als Maß für transparente Wolken im mittleren Höhengniveau nutzbar sein (vgl. Kapitel 3), da hohe Wolken ohnehin durch den Schwellenwert -20 °C (Kanal IR120) ausgeschlossen sind. Wie zuvor schon dargestellt, sprechen hohe Werte der Differenz IR108 - IR087 für niedrige kompakte Wasserwolken.

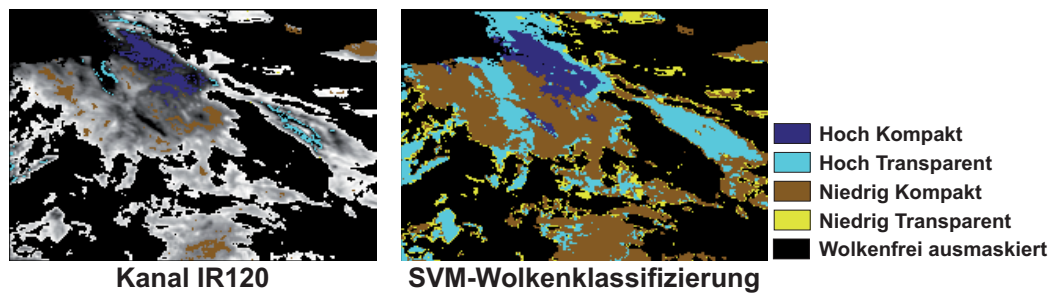
Sind die Wolken einer Einzelaufnahme so verteilt, dass bestimmte Klassen nicht vorkommen, so werden auch keine Trainingsdaten für diese Klassen generiert. Werden nur wenige Trainingspixel für eine Klasse erzeugt, sollte der SVM-Klassifikator dennoch zu einem Ergebnis kommen (vgl. Kapitel 5.1 und 5.2). Falls nur Trainingspixel einer Klasse vorhanden sind, kann keine weitere Differenzierung durch den SVM-Klassifikator vollzogen werden. Bei einer sehr heterogenen Bewölkungssituation scheint dies aber unwahrscheinlich, da im Bereich von teilweise bewölkten Pixeln in der Regel auch völlig bewölkte Pixel vorkommen. Ist je-



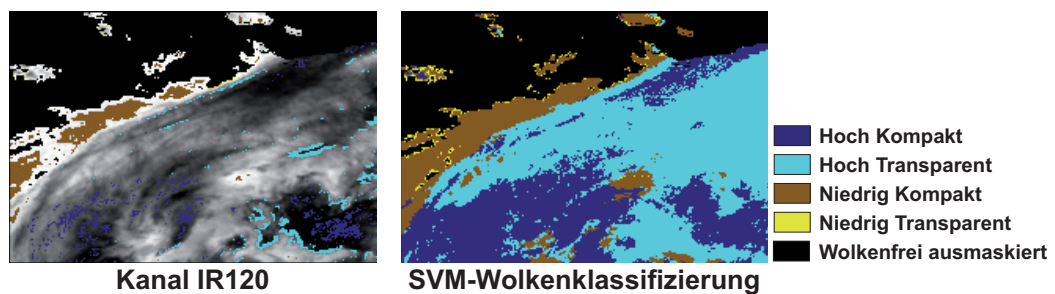
(a) 20.03.2008 20:00 UTC



(b) 08.06.2008 13:00 UTC



(c) 20.09.2008 00:00 UTC



(d) 10.12.2008 09:00 UTC

Abb. 5.17: Beispielhafte Darstellung verwendeter Trainingspixel der weiter differenzierten SVM-Wolkenklassifizierung. *Quelle: Eigene Darstellung*

doch der ganze Bildausschnitt mit kompakter Bewölkung überdeckt, besteht die Möglichkeit, dass die komplette Szene neben der Klasse „Wolkenfrei“ nur aus einer Wolkenklasse besteht (z.B. „Niedrig Kompakt“). Um zu vermeiden, dass nur Trainingsinformation einer Klasse in das weitere Klassifizierungsverfahren einfließt, wird bei der Generierung der Trainingspixel ein zusätzlicher Homogenitätstest durchgeführt. Ist die Differenz zwischen Standardabweichungsmaximum und Standardabweichungsminimum der ganzen Szene sehr gering, scheint dies für eine sehr einheitliche Bewölkungssituation zu sprechen. Wird hier der Wert 2 unterschritten, so wird keine weitere Klassifizierung durchgeführt (vgl. Abbildung 5.16).

Der Übergang zwischen völlig bewölkten und wolkenfreien Flächen ist oft charakterisiert durch Pixel mit teilweiser Bewölkung bzw. durch relativ transparente Wolken (z.B. Nebel), die jedoch nicht immer mit hohen Standardabweichungswerten verbunden sind. Die durchlässige Strahlung der Erdoberfläche kann hier zu vergleichsweise homogenen Verhältnissen führen, so dass fälschlicherweise Trainingspixel der Klasse „Niedrig Kompakt“ erzeugt werden. Zur Vermeidung werden relativ „warme“ Pixel (Kanal IR120) im Bereich des Übergangs „Völlig Bewölkt - Wolkenfrei“ noch einmal auf ihre Homogenität überprüft und entsprechend als Trainingspixel der Klasse „Niedrig Transparent“ zugeordnet.

Durch die Verknüpfungen der Schwellenwertalgorithmen (Einzeltests) in Bezug auf niedrigste bzw. höchste Wertebereiche (vgl. Abbildung 5.16), lässt sich die Menge der Trainingsdaten in einem für die Rechenzeit akzeptablen Rahmen halten. Um den Umfang der Trainingspixel trotzdem zu begrenzen, sollten 1500 Bildpunkte pro Klasse nicht überschritten werden. Andernfalls erfolgt auch hier eine Beschneidung mittels Zufallsverfahren.

Abschließend sei noch einmal erwähnt, dass alle Parameter, wie der Anteil höchster und niedrigster Werte innerhalb der Test-Verknüpfungen (vgl. Abbildung 5.14 und 5.16), maximaler Umfang von Trainingspixeln usw., auf den Ausschnitt Deutschland angepasst sind. Da die Berechnung der umfangreichen Datenreihen sehr zeitintensiv ist, besteht sicherlich immer noch die Möglichkeit der weiteren Optimierung dieser Parameter. Gleiches gilt bei einer Übertragung auf andere Gebiete.

Stehen für die einzelnen Klassen die notwendigen Trainingsdaten als ROI zur Verfügung, schließt sich das auf den Funktionen von imageSVM basierenden Klassifizierungsverfahren an. Die Trennung mehrerer Klassen erfolgt durch den Ansatz One Against All (OAA). In Bezug auf die für jedes Klassifizierungsproblem erzeugten Stützvektoren (Support Vectors) hat sich gezeigt, dass deren Umfang bei kleinerer Anzahl der Trainingspixel meist größer ist. Es ergeben sich jedoch teilweise enorme Unterschiede, da auch die Verteilung der Wolkenklassen von Szene zu Szene variiert. Neben den von imageSVM erzeugten Parameter-Dateien (vgl. Kapitel 5.2) erhält man die Klassifizierungsergebnisse in Form von Bilddaten (vgl. Abbildung 7.1).

Zusammenfassend ergibt sich für jeden Zeitschnitt ein Klassifizierungsergebnis mit vier Wolkenklassen ergänzt durch die zuvor erstellte Klasse „Wolkenfrei“. Das gesamte Verfahren in-

Tab. 5.5: Durchschnittliche Rechenzeiten der SVM-Wolkendetektion und SVM-Wolkenklassifizierung (ohne Validierung) für einen Zeitschnitt. Abgeleitet aus rückwirkender Berechnung der ersten Hälfte des Jahres 2008. *Quelle: Eigene Erhebung*

Zeitraum	Ø Rechenzeit in Minuten
Januar 2008	3:40
Februar 2008	3:55
März 2008	3:39
April 2008	3:42
Mai 2008	3:30
Juni 2008	3:57
Januar 2008 bis Juni 2008	3:44

klusive Erfassung von Wolken durch Wetter-Kameras, Abschätzung des ungefähren Grenzwertes, Erstellung der Wolkenmaske inklusive Gewinnung von Trainingspixeln und die Erzeugung der weiter differenzierten Wolkenklassifizierung kann für jede METEOSAT-Aufnahme komplett durchgeführt werden. Obwohl die Ausrichtung der Programme an eine Nachprozessierung der Datenreihe des Jahres 2008 angepasst sind, sollte nach den durchschnittlichen Rechenzeiten (vgl. Tabelle 5.5) ein operationeller Betrieb absolut möglich sein. Die Bilder der Wetter-Kameras müssten hierzu für den entsprechenden Zeitpunkt direkt nach ihrer Aufnahme ausgelesen werden, damit sie mit der entsprechenden METEOSAT-Szene in einem Real-Time-Modus abgeglichen werden können.

In den nachfolgenden Kapiteln wird das Validierungsverfahren der SVM-Wolkenmaske und der Vergleich mit den Ergebnissen der SAFNWC-Produkte detailliert beschrieben.

6 Validierung der Wolkendetektion mit synoptischen Beobachtungsdaten

Das Validierungsverfahren anhand von synoptischen Beobachtungsdaten ist so ausgerichtet, dass es, sofern Validierungsdaten zur Verfügung stehen, direkt für jede Einzelszene durchgeführt werden kann (vgl. Abbildung 5.1). Hierzu werden die Validierungsdaten aus dem DWD-SAT-Datensatz extrahiert (vgl. Kapitel 4.6). Das heißt, alle WMO-Stationen innerhalb von Deutschland werden bezüglich ihrer geographischen Koordinaten mit dem entsprechenden Pixel der METEOSAT-Szene verknüpft (vgl. Seite 72).

Aufgrund der räumlichen Auflösung von METEOSAT im Bereich von Deutschland (ca. 6 km, vgl. Abbildung 4.1), werden auch nur diese entsprechenden Pixel zu Validierungszwecken eingesetzt. Zwar wird durch die synoptische Beobachtung an der jeweiligen WMO-Station in der Regel ein größerer Teil der Himmelsfläche erfasst, dennoch sollten die Bewölkungsverhältnisse des METEOSAT-Bildpunktes hiermit gut darstellbar sein. Eine Hinzunahme benachbarter Pixel ist daher, bei vorliegender räumlicher Auflösung, nicht unbedingt notwendig. Innerhalb des Verfahrens der Trainingsdatengewinnung wurde zudem sicher gestellt, dass es zu keinen Überschneidungen zwischen Trainingsdaten und Validierungsdaten kommt.

Die Beobachtungsdaten der WMO-Stationen stehen für Deutschland in stündlichen Intervallen zur Verfügung, und die Bereitstellung der Daten erfolgt, im Vergleich zu den METEOSAT-Aufnahmen selbst, mit einer zeitlichen Verzögerung. Bei der rückwirkenden Bearbeitung der Daten des Jahres 2008 spielt dies keine Rolle, für eine Validierung im laufenden Betrieb muss der Versatz jedoch berücksichtigt werden.

Um einen Vergleich zwischen der SVM-Wolkendetektion und den Produkten des SAFNWC gewährleisten zu können (vgl. Kapitel 3.2), der nicht nur auf visueller Begutachtung beruht, erfolgt die Validierung in gleicher Weise für die SAFNWC-Cloudmask. Es werden demnach alle vorhandenen Validierungs-Bildpunkte für beide Detektionsverfahren herangezogen (vgl. Abbildung 5.1 und 6.1). Da die SAFNWC-Cloudmask innerhalb der Prozessierung auf den gleichen Ausschnitt angewendet wird (Abmessungen und Koordinaten bezogen auf die gesamte „MSG-Scheibe“, vgl. Kapitel 4.5 und Abbildung 5.2), kann für beide Verfahren identisch vorgegangen werden.

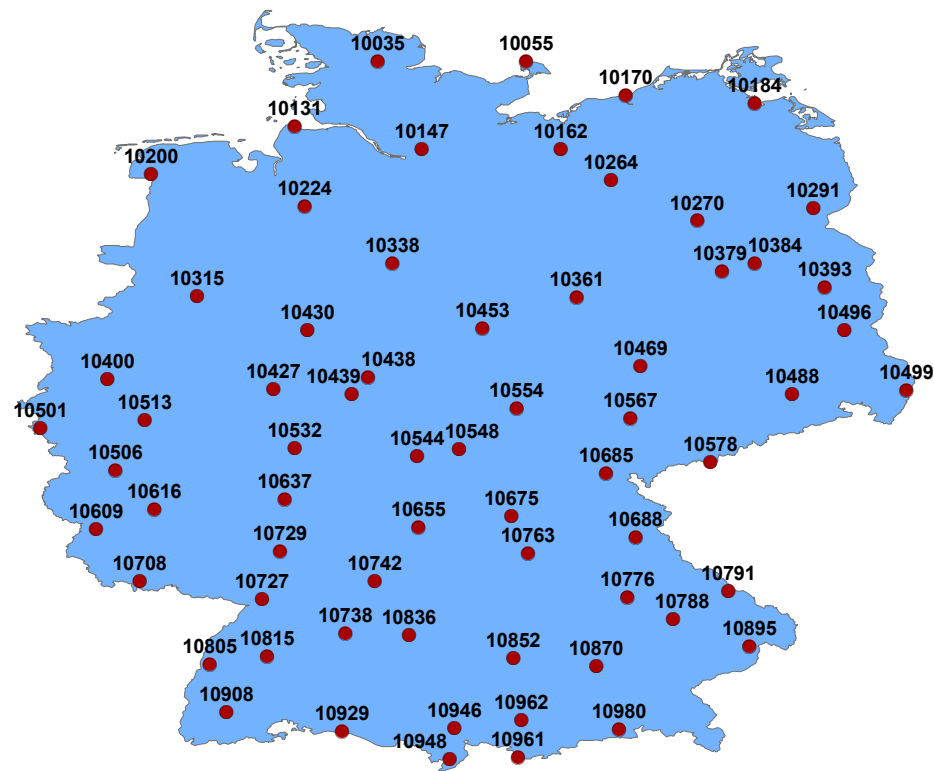
Ausführliche Validierungsansätze der SAFNWC-Wolkenprodukte, ebenfalls anhand von synoptischen Beobachtungsdaten, wurden zwar schon durchgeführt (vgl. Derrien und Le Gléau 2005a, SAFNWC 2007j, SAFNWC 2007l), basieren jedoch auf einer anderen Herangehensweise. Aus diesem Grund wird das Validierungsverfahren sowohl für die SVM-Wolkenmaske, als auch für die SAFNWC-Cloudmask mit gleichen Bedingungen durchgeführt.

Der Umfang des Validierungsdatensatzes setzt sich aus allen zur Verfügung stehenden Zeitpunkten mit ihren jeweiligen Validierungsstandorten (Bildpunkte der Stationen) zusammen. Insgesamt stehen WMO-Stationsdaten von Anfang Juli bis Ende Dezember des Jahres 2008 zur Verfügung. Da die Verteilung von Wolken und wolkenfreien Flächen in jeder Szene unterschiedlich ist, kann mit dieser Vorgehensweise gewährleistet werden, dass eine Vielzahl möglicher Bewölkungssituationen Teil des Validierungsverfahrens werden.

Eine auf synoptischen Beobachtungsdaten basierende Validierung der weiter differenzierten Wolkenklassifizierung stellte sich bei mehreren Testläufen als äußerst schwierig dar. Zwar stehen an den meisten Stationen auch Informationen über Wolkenart bzw. Wolkenhöhe zur Verfügung (vgl. Kapitel 6.1), diese beziehen sich jedoch auf die Betrachtung des Beobachters von unten. Je nach Bewölkungsgrad erhält man somit nur Informationen über den unteren Teil der Wolkengegebenheiten. Die Klassifizierung durch den Satelliten erfolgt jedoch von oben, wodurch ausschließlich Informationen von diesem Teil der Bewölkung gewonnen werden. Überlagerungen unterschiedlicher Wolkenarten oder die vertikale Mächtigkeit der Wolken lassen sich mit vorliegendem Verfahren nicht ableiten. Eine Validierung, wie sie im Sinne der SVM-Wolkenmaske durchgeführt wird, kann demnach nicht erfolgen. Andere Validierungsansätze, die z.B. auf der Interpretation durch Experten basieren, sind nur mit erheblichen Aufwand durchzuführen (vgl. Diamandi und Dybbroe 2001). Letztendlich bleibt nur ein Vergleich mit den Ergebnissen des Produkts SAFNWC-Cloudtype (vgl. Kapitel 3.2.2), welcher Hinweise auf die Genauigkeit der SVM-Wolkenklassifizierung geben kann.

6.1 Umsetzung des Validierungsverfahrens

Sobald die SVM-Wolkenmaske erstellt ist und die für die Validierung notwendigen Daten verfügbar sind, kann sich das Validierungsverfahren anschließen. Ähnlich wie bei der Trainingsdatenzuordnung, erfolgt die Validierung mit Hilfe von ROIs. Hierzu werden die mit WMO-Stationen verknüpften Pixel, in Abhängigkeit ihrer Klassenzugehörigkeit, als entsprechendes ROI definiert. Da die Validierung der SVM-Wolkenmaske mit Funktionen von imageSVM durchgeführt wird, bedarf es zusätzlich der für die SVM-Berechnung notwendigen Stapel-Datei (Layerstack, Merkmalsraum als Kanalkombination, siehe Abbildung 5.3) und der SVM-Modell-Datei (SVM-Klassifikator). Letztere wurde beim Training für den entsprechenden Zeitschnitt erstellt. Das eigentliche Klassifizierungsergebnis (SVM-Wolkenmaske als Bild-Datei)



Stationsnr.	Standort	Geo. Breite	Geo. Länge
10035	SCHLESWIG	54,533	9,550
10055	WESTERMARKELSDURF	54,533	11,067
10131	CUXHAVEN	53,867	8,700
10147	HAMBURG/FUHLBUTTEL	53,633	10,000
10162	SCHWERIN	53,633	11,417
10170	WARNEMUENDE	54,183	12,083
10184	GREIFSWALD	54,100	13,400
10200	EMDEN/FLUGPLATZ	53,383	7,233
10224	BREMEN	53,050	8,800
10264	MARNITZ (AUT)	53,317	11,933
10270	NEURUPPIN	52,900	12,817
10291	ANGERMUENDE	53,033	14,000
10315	MUNSTER/OSNABRUCK	52,133	7,700
10338	HANNOVER	52,467	9,700
10361	MAGDEBURG	52,117	11,583
10379	POTSDAM	52,383	13,067
10384	BERLIN/TEMPELHOF	52,467	13,400
10393	LINDENBERG	52,217	14,117
10400	DUSSELDORF	51,283	6,783
10427	KAHLER ASTEN(MOUNT)	51,183	8,483
10430	BAD LIPPSPRINGE	51,783	8,833
10438	KASSEL	51,300	9,450
10439	FRITZLAR (GER-ARMY)	51,133	9,283
10453	BROCKEN (PEAK)	51,800	10,617
10469	LEIPZIG/SCHKEUDITZ	51,417	12,233
10488	DRESDEN/KLOTZSCHE	51,133	13,783
10496	COTTBUS	51,783	14,317
10499	GOERLITZ	51,167	14,950
10501	AACHEN	50,783	6,100
10506	NUERBURG-BARWEILER	50,350	6,867
10513	KOLN/BONN (CIV/MIL)	50,867	7,167
10532	GIESSEN	50,583	8,700
10544	WASSERKUPPE (MOUNT)	50,500	9,950
10548	MEININGEN	50,567	10,383
10554	ERFURT/BINDERSLEBN&	50,983	10,967
10567	GERA/LEUMNITZ	50,883	12,133
10578	FICHELBERG MTN	50,433	12,950
10609	TRIER/PETRISBERG	49,750	6,667
10616	HAHN	49,950	7,267
10637	FRANKFURT MAIN ARPT	50,050	8,600
10655	WURZBURG (USA-AF)	49,767	9,967
10675	BAMBERG (USA-AF)	49,883	10,917
10685	HOF	50,317	11,883
10688	WEIDEN (MIL/MET)	49,667	12,183
10708	SAARBRUCKEN/ENSHEIM	49,217	7,117
10727	KARLSRUHE (USA-AF)	49,033	8,367
10729	MANNHEIM/NEUOSTHEIM	49,517	8,550
10738	STUTTGART/ECHTERDI	48,683	9,217
10742	OEHRINGEN	49,217	9,517
10763	NURNBERG	49,500	11,083
10776	REGENSBURG/OBERHUB	49,050	12,100
10788	STRAUBING(GER-ARMY)	48,833	12,567
10791	GROSSER ARBER (MTN)	49,117	13,133
10805	LAHR (CAN-AFB)	48,367	7,833
10815	FREUDENSTADT	48,450	8,417
10836	STOETTEN	48,667	9,867
10852	AUGSBERG/MULHAUSEN	48,433	10,933
10870	MUNICH ARPT	48,350	11,783
10895	FURSTENZELL	48,550	13,350
10908	FELDBERG/SCHWARZWAL	47,883	8,000
10929	KONSTANZ	47,683	9,183
10946	KEMPTEN/DURACH	47,717	10,333
10948	OBERSTDORF	47,400	10,283
10961	ZUGSPITZE MOUNTAIN	47,417	10,983
10962	HOHENPEIßENBERG	47,800	11,017
10980	WENDELSTEIN MTN	47,700	12,017

Abb. 6.1: Standorte der durch DWDSAT verfügbaren WMO-Stationsdaten (Deutschland), für die Validierung des Wolkendetektionsverfahrens. *Quelle: Eigener Entwurf, Stationsliste extrahiert aus DWDSAT-Synopdaten*

Tab. 6.1: Beobachtungsdaten an den WMO-Stationen. Wolkenbedeckung (Total Cloud Cover) in Achteln. *Quelle: Eigene Aufstellung abgeleitet aus DWDSAT-Synopdaten*

Einteilung Total Cloud Cover (Achtel)	
Wolkenfrei	sky clear
Wolken/Wolkenfrei	1/8th or less, but not sky clear
Wolken	2/8ths
	3/8ths
	4/8ths
	5/8ths
	6/8ths
	7/8ths or more, but not 8/8ths
	8/8ths
	sky obscured

ist bei dieser Vorgehensweise nicht unbedingt erforderlich, da sich die notwendige Information aus dem SVM-Modell, verbunden mit der verknüpften Stapel-Datei, ableiten lässt.

Um die Validierungsinformation der synoptischen Stationsdaten (gts06-SYNOPI_Bulletins, vgl. Kapitel 4.6) in tabellarischer Form verwenden zu können, müssen diese in einem ersten Schritt dekodiert werden. Anschließend erfolgt die Auswahl der für diese Arbeit notwendigen Stationen, welche anhand ihres Ländercodes (DL für Deutschland) extrahiert werden. In gleicher Weise können Datum und Zeit (UTC) mit der zugehörigen METEOSAT-Aufnahme abgeglichen werden. Insgesamt stehen 66 Stationen für den Raum Deutschland zur Verfügung (vgl. Abbildung 6.1). Stationen, die sich auf Inseln befinden, werden für die Validierung nicht verwendet, da sich das Detektionsverfahren nur auf Landflächen bezieht. Die Isolation durch Meeresflächen könnte zu Missinterpretationen führen. Alle an den Stationen aufgezeichneten, sich auf Bewölkung beziehende Beobachtungsdaten, sind nachfolgend aufgelistet:

- *Cloud Base* - Wolkenbasishöhe in Metern
- *Total Cloud Cover* - Wolkenbedeckung in Achteln (Achtel-Skala)
- *Low Cloud Type* - Wolkenart im niedrigen Stockwerk (vgl. Tabelle 2.1)
- *Middle Cloud Type* - Wolkenart im mittleren Stockwerk (vgl. Tabelle 2.1)
- *High Cloud Type* - Wolkenart im hohen Stockwerk (vgl. Tabelle 2.1)

Für die Validierung der SVM-Wolkenmaske wird die Wolkenbedeckung in Achteln herangezogen (vgl. Tabelle 6.1 und Kraus 2009, S. 305). Mit Hilfe der Achtel-Skala erfasst man die Gesamtbedeckung des Himmels, wobei sich die Himmelsfläche aus acht Teilen zusammensetzt.

Da zwischen wolkenkontaminierten und wolkenfreien Pixeln unterschieden wird, muss vorab geklärt werden, ab welchem Wert man von einer wolkenfreien Situation sprechen kann. Aus diesem Grund findet eine zweimalige Durchführung des Validierungsverfahrens statt. Zum einen wird angenommen, dass es sich bei dem Wert „1/8th or less, but not sky clear“ um Bewölkung handelt, während in einem zweiten Lauf von einer wolkenfreien Situation ausgegangen wird. Anhand der Ergebnisdarstellungen kann dies dann entsprechend verglichen werden.

Weiterhin ist zu erwähnen, dass Validierungspixel mit Schneebedeckung nicht innerhalb des Verfahrens eingesetzt werden. Im gesamten Prozess der Detektion werden Schneeflächen nicht berücksichtigt, da sich die Gewinnung entsprechender Trainingsdaten als zu fehlerhaft erwiesen hat. Von einer ausgeprägten Falschzuweisung durch Schneeflächen ist daher auszugehen. Da sich die Validierung aber auf die Trennung zwischen Wolken und wolkenfreien Landflächen bezieht, werden mit Schnee kontaminierte Pixel bei der Validierung außen vor gelassen. Innerhalb der einzelnen SVM-Wolkenmasken (je Zeitschnitt) muss daher einkalkuliert werden, dass Schnee- und Eisflächen zu größeren Ungenauigkeiten führen.

Liegt an einem Validierungspixel kein Detektionsergebnis vor („Unclassified“, NaN) so wird dieses auch nicht für die weitere Validierung herangezogen. Sind insgesamt keine Validierungsdaten für den Aufnahmezeitpunkt vorhanden, so findet hier keine Validierung statt.

Abschließend erhält man zwei ROIs („Wolkenkontaminiert“ und „Wolkenfrei“) deren Anzahl an Bildpunkten zusammen den Umfang der möglichen Validierungspixel ergibt. Mit den notwendigen Input-Daten (Merkmalsraum als Layerstack und SVM-Modell) wird dann, unter Verwendung der Validierungsfunktionen von imageSVM, die Bewertung begonnen.

Für die Validierung der SAFNWC-Cloudmask, die nach gleichem Muster erfolgt, können die imageSVM Funktionen nicht herangezogen werden, weshalb die entsprechenden Gütemaße des Accuracy Assessments eigenständig berechnet werden. Nachfolgend aufgeführte Parameter werden für jede Einzelszene, an der Validierungsdaten zur Verfügung stehen, erhoben (vgl. hierzu auch Jensen 2004, S. 495 ff.).

- Konfusionsmatrix
- Overall Accuracy
- Kappa-Koeffizient
- Producer Accuracy
- User Accuracy
- Clear Failure - Überschätzung Wolken
- Cloud Failure - Unterschätzung Wolken

Tab. 6.2: Aufbau der Konfusionsmatrix für die Validierung der SVM-Wolkenmaske und der SAFNWC-Cloudmask. *Quelle: Eigene Aufstellung*

	Wolke beobachtet	Wolkenfrei beobachtet	Summe
Wolke detektiert	A	B	A + B
Wolkenfrei detektiert	C	D	C + D
Summe	A + C	B + D	Σ

Der Aufbau der Konfusionsmatrix ist in Tabelle 6.2 dargestellt und die für die Wolkendetektion wesentlichen Genauigkeitsmaße werden wie folgt berechnet.

$$\text{Overall Accuracy} = \left(\frac{A+D}{\Sigma} \right) * 100 \quad (6.1)$$

$$\text{Cloud Failure} = \left(\frac{C}{A+C} \right) * 100 \quad (6.2)$$

$$\text{Clear Failure} = \left(\frac{B}{B+D} \right) * 100 \quad (6.3)$$

Da es sich bei Cloud und Clear Failure um den Gegensatz der jeweiligen Producer Accuracy handelt (vgl. Derrien et al. 2005), und wegen der im Umfang unterschiedlich großen Klassen, wird im weiteren nicht mehr näher auf Producer und User Accuracy eingegangen. Gleiches gilt für den Kappa-Koeffizienten (vgl. Pontius 2000), der zwar durch die Funktionen von imageSVM berechnet wird, jedoch für das vorliegende Klassifizierungsproblem (zwei Klassen mit unterschiedlicher Größe) kein sinnvolles Maß darstellt. Mit Hilfe der Overall Accuracy, sowie Cloud und Clear Failure, lassen sich in diesem Zusammenhang bessere Aussagen über die Genauigkeit treffen.

Liegen nur Validierungsdaten einer Klasse vor, beispielsweise wenn große Teile der Szene wolkenbedeckt sind, kann die Berechnung nicht mit den Funktionen von imageSVM durchgeführt werden. In diesem Fall erfolgt nur die eigenständige Berechnung der Overall Accuracy. Die Ausgabe der Werte wird mittels einer Log-Datei realisiert.

Da die Validierung für mehrere Zeitschnitte durchgeführt werden soll, müssen die Validierungsdaten der Einzelzeitschnitte zusammengefasst werden. Eine über alle Aufnahmezeitpunkte aufsummierte Konfusionsmatrix, aus der die Berechnung der Gütemaße über den Gesamtzeitraum erfolgt, wird erzeugt. Durch die Zusammenführung der Validierungspixel über einen größeren Zeitraum wird eine Vielzahl an Bewölkungssituationen abgedeckt, auch wenn für die jeweilige Einzelszene nur eine begrenzte Anzahl an Validierungspixeln zur Verfügung steht (vgl. Abbildung 6.1).

Damit die Validierung der SAFNWC-Cloudmask in gleicher Weise durchgeführt werden kann, müssen die Klassen dieser Maske angepasst werden. Im Ursprungsprodukt wird zusätzlich zwischen völlig bewölkten und teilweise bewölkten Flächen unterschieden (vgl. Abbildung 3.2). Zudem findet tagsüber eine Erfassung von Schneeflächen statt. Durch Auslesen des HDF5-Dateiformates lassen sich die verschiedenen Klassen zusammenfassen. Man erhält somit nur noch eine Differenzierung zwischen „Wolkenkontaminiert“ und „Wolkenfrei“. Schneeflächen werden als „Wolkenfrei“ eingestuft.

6.2 Validierungsergebnisse SVM-Wolkenmaske im Vergleich mit SAFNWC-Cloudmask

Im Folgenden sind die Validierungsergebnisse sowohl der SVM-Wolkenmaske als auch der SAFNWC-Cloudmask für den Zeitraum Juli bis Dezember 2008 dargestellt. Für jeden Monat und den Gesamtvalidierungszeitraum ergibt sich die entsprechende Konfusionsmatrix, Overall Accuracy, Cloud und Clear Failure (siehe Tabellen 6.3 bis 6.6). Abbildung 6.2 zeigt die Ergebnisse zusätzlich in grafischer Form. Bei der ersten Betrachtung wird schnell deutlich, dass beide Wolkendetektionsverfahren mit Genauigkeiten (Overall Accuracy) von über 80 % sehr solide funktionieren. Die SVM-Wolkendetektion kann hier, bis auf November und Dezember, sogar höhere Genauigkeiten vorweisen. Dies gilt für beide Ausgangslagen der Achtel-Skala (vgl. Kapitel 6.1). Ein-Achtel wird zum einen als „Wolkenkontaminiert“ (vgl. Tabelle 6.5 und 6.6), zum anderen als „Wolkenfrei“ eingestuft (vgl. Tabelle 6.3 und 6.4). Auffällig ist hierbei, dass bei der Annahme „Ein-Achtel entspricht wolkenkontaminierten Gegebenheiten“, beide Erkennungsverfahren allgemein geringere Genauigkeiten aufweisen. Man könnte vermuten, dass der Beobachter in diesen Situationen nur ganz geringfügige Bewölkung erfasst, die durch die Erkennungsverfahren (SVM und SAFNWC) nur bedingt richtig detektiert werden können. Bei der SVM-Wolkenmaske kommt es in diesem Unsicherheitsbereich verstärkt zu einer Überschätzung von Wolken, während die SAFNWC-Cloudmask eine übermäßige Unterschätzung zeigt. Dies wird noch deutlicher, wenn man sich die Werte von Cloud und Clear Failure der einzelnen Monate anschaut. Die SVM-Wolkenmaske verzeichnet hier, im Vergleich zur SAFNWC-Cloudmask, relativ hohe Clear Failure Werte, während es sich bei Cloud Failure umgekehrt verhält. Zwar ist die relative Differenz von Clear Failure zwischen den unterschiedlichen Verfahren höher als bei Cloud Failure, dies ist jedoch durch die unterschiedliche Größe der Klassen zu begründen. Innerhalb der jeweiligen Konfusionsmatrix lässt sich aber erkennen, wie sich die falsch detektierten Werte der jeweiligen Klassen in absoluten Zahlen zusammensetzen. Man muss trotzdem davon ausgehen, dass die Klasse „Wolkenfrei“, durch ihre geringere Größe, von

der SVM-Wolkenmaske in Teilen nur ungenau wiedergegeben werden kann. Innerhalb von Kapitel 7 wird dies weiter diskutiert.

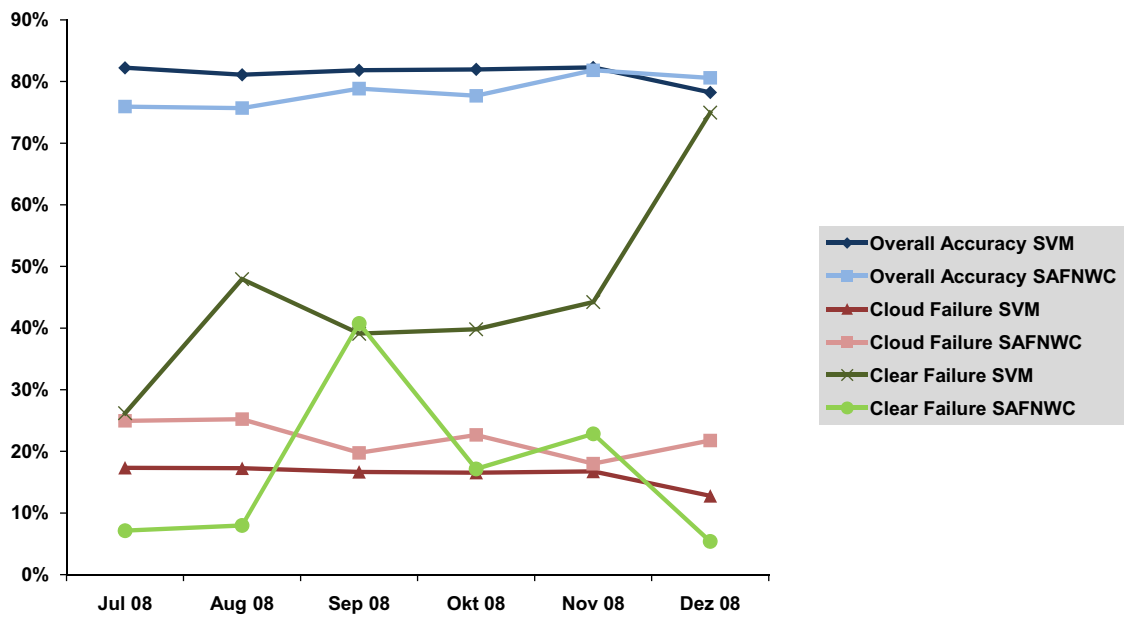
Auffällig ist vor allem die hohe Überschätzung der SVM-Wolkendetektion im Dezember. Diese lässt sich wahrscheinlich auf wolkenfreie Gegebenheiten in Verbindung mit ausgeprägten Schneeflächen in höheren Lagen zurückführen, die in diesem Monat zu verzeichnen waren²⁴. Wie bereits erwähnt, werden Schneeflächen zwar bei der Validierung nicht berücksichtigt, sie können jedoch als fehlerhafte Information bei der Trainingsdatengewinnung und Klassifizierung zu Ungenauigkeiten führen.

Die Problematik der Überschätzung durch den SVM-Klassifikator ist in Abbildung 6.3 anhand zweier Beispiele noch einmal genauer dargestellt. Zwar werden wolkenfreie Landflächen durch die SVM-Wolkendetektion fälschlicherweise als Wolke eingestuft, ob hier in Teilen dennoch von Bewölkung ausgegangen werden kann, liegt eventuell auch an Ungenauigkeiten durch die subjektive Einschätzung des Beobachters.

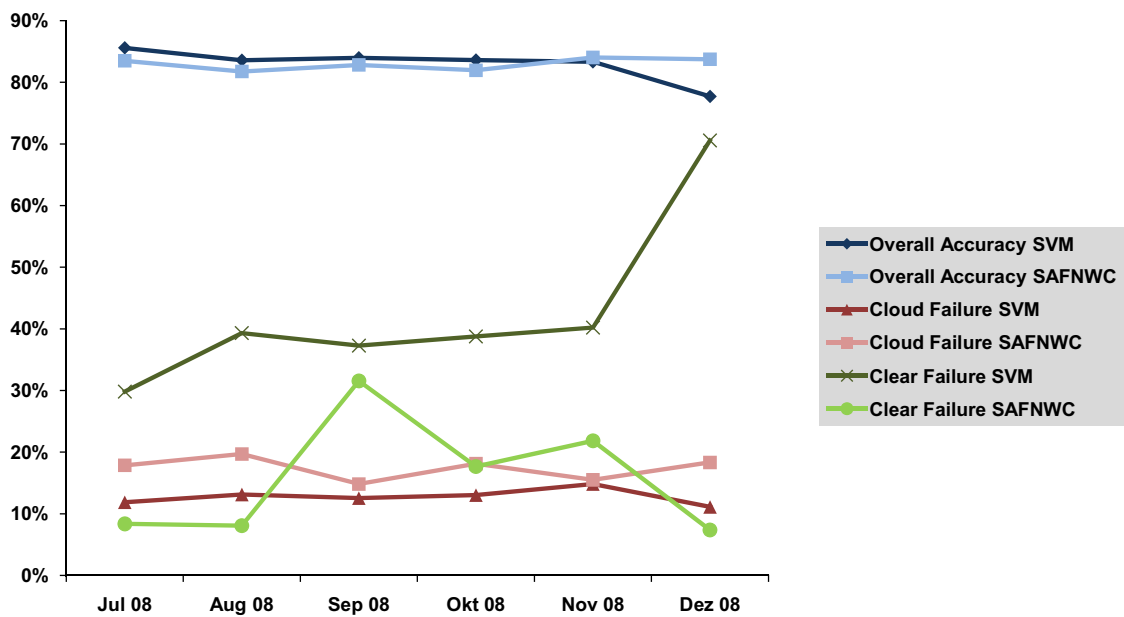
Betrachtet man die im Zusammenhang der SAFNWC bereits durchgeführten Validierungsversuche (vgl. hierzu Derrien und Le Gléau 2005a, S. 4721 - 4729) so sieht die Situation insbesondere für Regionen in den nördlicheren Breiten recht ähnlich aus. Zwar werden hier bei der Overall Accuracy höhere Werte erzielt (Nordic = 89,3 %; Mid-latitude = 91,7 %), dennoch kommt es auch hier vornehmlich zu Überschätzungen von Wolken (Clear Failure) vor allem bei den Nacht-Testreihen (Nordic = 25,5 %; Mid-latitude = 10,9 %; Nordic-Night = 33,6 %; Mid-latitude-Night = 15,0 %). Der Vergleich mit vorliegender Validierung ist jedoch mit Vorsicht zu betrachten, da die Herangehensweise bei Derrien und Le Gléau (2005a) eine andere ist. Derrien und Le Gléau (2005a) beziehen sich auf mehrere Pixel (auch benachbarte Pixel) und es wird nicht zwischen „Wolkenfrei“ und „Wolkenkontaminiert“, sondern zwischen „Wolkenfrei“ und „Völlig Bewölkt“ unterschieden. Zudem wurde eine andere Zeitspanne herangezogen (November 2003 bis April 2004) und die Stationsdaten beziehen sich auf den ganzen Bereich von Europa. In Bezug auf die in vorliegender Arbeit erkennbare Unterschätzung der SAFNWC-Cloudmask, kann man vermuten, dass hierfür die Berechnungen ohne zusätzliche Wettervorhersagemodell-Daten (NWP) verantwortlich sind (vgl. Kapitel 4.5). Die Abwesenheit dieser Zusatzdaten führt zwar zu einer veränderten Genauigkeit, ist aber erforderlich, um einen Vergleich zweier Verfahren zu gewährleisten, die beide vorrangig auf reinen Bilddaten basieren.

Über die Art der Bewölkung, die bei der SAFNWC-Cloudmask vornehmlich unterschätzt wird, lässt sich durch das vorliegende Validierungsverfahren keine Aussage treffen. Zu vermuten ist, dass vor allem Pixel mit dünnem Wolkenbesatz auf Subpixel-Ebene (z.B. Cirrus-Wolken, Kondensstreifen des Flugverkehrs usw.) hierfür verantwortlich sind. In Verbindung mit der Ergebnisdarstellung in Kapitel 7 wird genauer darauf eingegangen.

²⁴http://www.wetteronline.de/wotexte/redaktion/rueckblick/2008/12/30_dz.htm
[25.02.2011]



(a) „1/8th or less, but not sky clear“ wird als „Wolkenkontaminiert“ angesehen



(b) „1/8th or less, but not sky clear“ wird als „Wolkenfrei“ angesehen

Abb. 6.2: Vergleich der Validierungsergebnisse als grafische Darstellung. Quelle: Eigene Darstellung

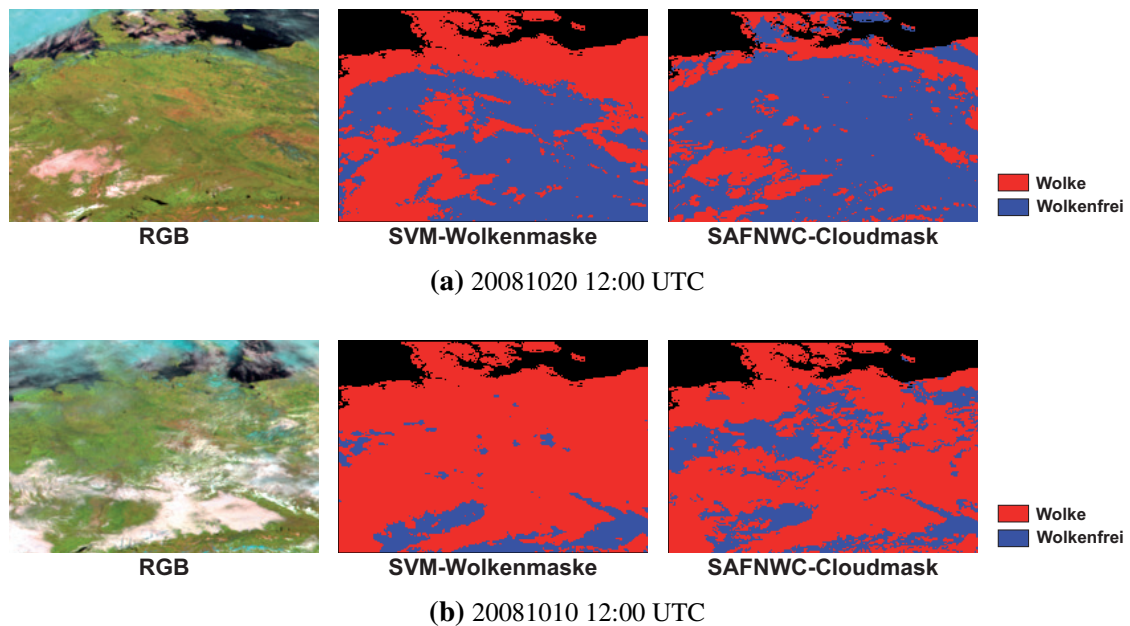


Abb. 6.3: Effekt der SVM-Überschätzung. *Quelle: Eigene Darstellung*

Um den Einfluss der Wetter-Kameras bei der Trainingsdatengewinnung abschätzen zu können, wurde ein separater Durchlauf der SVM-Wolkendetektion, ohne zusätzliche Information der Wetter-Kameras, für den Monat Oktober 2008 durchgeführt. Das Validierungsergebnis (Overall Accuracy, vgl. Tabelle 6.7) ist zwar nur geringfügig schlechter als bei ergänzender Verwendung der Wetter-Kameras, es wird jedoch deutlich, dass insbesondere die Überschätzung (Clear Failure) durch die Informationen der Kameras geringer ausfällt. In diesem Zusammenhang kommt es bei der Trainingsdatengewinnung zu einer Korrektur auf Basis der Wetter-Kameras.

Eine weitere Differenzierung des Validierungsverfahrens erfolgte in Bezug auf Tag- und Nacht-Testsequenzen bei der Trainingsdatengewinnung (vgl. hierzu Tabelle 5.4 und Abbildung 5.14). Diese Unterscheidung bezieht sich nur auf die Trainingsdatengewinnung der SVM-Wolkendetektion, bei der tagsüber (voll ausgeleuchtete Szene Deutschland) zusätzliche Test-Algorithmen basierend auf METEOSAT-Kanälen im sichtbaren Wellenlängenbereich zum Einsatz kommen. Der Vergleich ist in Tabelle 6.8 (obere vier Zeilen) aufgeführt und zeigt, dass durch die zusätzliche Berücksichtigung der sichtbaren Kanäle eine höhere Genauigkeit erzielt werden kann. Es ist zudem davon auszugehen, dass sich bei dieser Betrachtung ebenfalls der positive Effekt der Wetter-Kameras erkennen lässt. Bis auf den Standort Mannheim, steht nachts keine zusätzliche Trainingsinformation durch Kameras zur Verfügung. Beim Vergleich mit der SAFNWC-Cloudmask ist darauf zu achten, dass sich die Tag- und Nachteinteilung auf die SVM-Trainingsdatengewinnung bezieht. Bei der eigentlichen Erstellung der SAFNWC-

Cloudmask ist diese Einteilung eine andere (vgl. hierzu Derrien und Le Gléau 2005a, S. 4729). Abweichungen zwischen SVM-Wolkenmaske und SAFNWC-Cloudmask sind bei dieser Differenzierung jedoch nur in geringem Maße zu verzeichnen.

Abschließend wurde noch eine Abschätzung der Genauigkeit in Bezug auf unterschiedliche Höhenlagen der WMO-Stationen durchgeführt. Hierzu wurden Stationsdaten unterhalb und oberhalb von 300 Höhenmetern herangezogen. Mit dieser Betrachtung soll dargestellt werden, ob sich die Qualität des Detektionsverfahrens in den höheren Lagen der Mittelgebirge anders verhält, als dies in niederen Regionen der Fall ist (vgl. Tabelle 6.8, untere vier Zeilen). Die Overall Accuracy zeigt hier sowohl in der Höhe, als auch beim Vergleich SVM zu SAFNWC kaum Unterschiede. Lediglich die zuvor beschriebenen Phänomene der SVM-Überschätzung sowie der SAFNWC-Unterschätzung werden mit zunehmender Höhe leicht verstärkt.

In den nachfolgenden Kapiteln werden die Ergebnisse, sowohl der Wolkendetektion als auch der weiter differenzierten Wolkenklassifizierung, ausführlich dargestellt. Zudem sollen anhand der räumlichen und zeitlichen Differenzierung des Wolkenbedeckungsgrades weitere Ungenauigkeiten und ihre möglichen Ursachen abgeschätzt werden. Gleichzeitig werden die Ergebnisse in Verbindung mit dem allgemeinen Wettergeschehen des Jahres 2008 eingeordnet. Abschließend folgt dann in Kapitel 9 eine detaillierte Diskussion des durchgeführten Detektions- bzw. Klassifizierungsansatzes, bei der die zu Beginn der Arbeit definierten Fragestellungen aufgegriffen werden.

Tab. 6.3: Validierungsergebnisse der automatisierten SVM-Wolkenmaske mit WMO-Stationsdaten, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenfrei“ angesehen. *Quelle: Eigene Erhebung*

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Juli 2008	29300 3946	1633 3845	38724	85,59 %	11,87 %	29,81 %
August 2008	31169 4712	2028 3131	41040	83,58 %	13,13 %	39,31 %
September 2008	29787 4272	2087 3512	39658	83,97 %	12,54 %	37,27 %
Oktober 2008	30599 4580	2058 3251	40488	83,61 %	13,02 %	38,76 %
November 2008	24700 4299	922 1372	31293	83,32 %	14,82 %	40,19 %
Dezember 2008	20423 2552	3757 1566	28298	77,71 %	11,11 %	70,58 %
Juli bis Dezember 2008	165978 24361	12485 16677	219501	83,21 %	12,80 %	42,81 %

Tab. 6.4: Validierungsergebnisse von SAFNWC-Cloudmask mit WMO-Stationsdaten, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenfrei“ angesehen. *Quelle: Eigene Erhebung*

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Juli 2008	27310	458	38724	83,49 %	17,85 %	8,36 %
	5936	5020				
August 2008	28812	417	41040	81,76 %	19,70 %	8,08 %
	7069	4742				
September 2008	29013	1766	39658	82,82 %	14,82 %	31,54 %
	5046	3833				
Oktober 2008	28809	937	40488	81,95 %	18,12 %	17,65 %
	6370	4372				
November 2008	24503	501	31293	84,03 %	15,50 %	21,84 %
	4496	1793				
Dezember 2008	18764	392	28298	83,73 %	18,33 %	7,36 %
	4211	4931				
Juli bis Dezember 2008	157211	4471	219501	82,87 %	17,40 %	15,33 %
	33128	24691				

Tab. 6.5: Validierungsergebnisse der automatisierten SVM-Wolkenmaske mit WMO-Stationsdaten, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenkontaminiert“ angesehen. *Quelle: Eigene Erhebung*

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Juli 2008	30433 6383	500 1408	38724	82,23 %	17,34 %	26,20 %
August 2008	32150 6706	1047 1137	41040	81,11 %	17,26 %	47,94 %
September 2008	30839 6171	1035 1613	39658	81,83 %	16,67 %	39,09 %
Oktober 2008	31622 6266	1035 1565	40488	81,97 %	16,54 %	39,81 %
November 2008	25132 5053	490 618	31293	82,29 %	16,74 %	44,22 %
Dezember 2008	21113 3093	3067 1025	28298	78,23 %	12,78 %	74,95 %
Juli bis Dezember 2008	171289 33672	7174 7366	219501	81,39 %	16,43 %	49,34 %

Tab. 6.6: Validierungsergebnisse der SAFNWC-Cloudmask mit WMO-Stationsdaten, Juli bis Dezember 2008. „1/8th or less, but not sky clear“ wird als „Wolkenkontaminiert“ angesehen. *Quelle: Eigene Erhebung*

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Juli 2008	27632	136	38724	75,93 %	24,95 %	7,13 %
	9184	1772				
August 2008	29055	174	41040	75,69 %	25,22 %	7,97 %
	9801	2010				
September 2008	29700	1079	39658	78,85 %	19,75 %	40,75 %
	7310	1569				
Oktober 2008	29300	446	40488	77,69 %	22,67 %	17,15 %
	8588	2154				
November 2008	24751	253	31293	81,83 %	18,00 %	22,83 %
	5434	855				
Dezember 2008	18936	220	28298	80,60 %	21,77 %	5,38 %
	5270	3872				
Juli bis Dezember 2008	159374	2308	219501	78,18 %	22,24 %	15,87 %
	45587	12232				

Tab. 6.7: Validierungsergebnisse der SVM-Wolkenmaske mit WMO-Stationsdaten, Oktober 2008. Mit und ohne Trainingsinformation abgeleitet aus Wetter-Kameras.
Quelle: Eigene Erhebung

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Oktober 2008	30599	2058	40488	83,61 %	13,02 %	38,76 %
SVM - mit Wetter-Kameras	4580	3251				
Oktober 2008	30582	2137	40488	83,37 %	13,07 %	40,25 %
SVM - ohne Wetter-Kameras	4597	3172				

Tab. 6.8: Validierungsergebnisse der SVM-Wolkenmaske und der SAFNWC-Cloudmask mit WMO-Stationsdaten, Validierungszeitraum Juli bis Dezember 2008. Obere vier Zeilen: Unterscheidung zwischen Tag- und Nacht-Testsequenzen bei der Trainingsdatengewinnung (bezieht sich auf SVM-Anwendung, SAFNWC dient nur als Vergleich). Untere vier Zeilen: Aufteilung der Validierung in unterschiedliche Höhenlagen (Höhe der jeweiligen WMO-Station). *Quelle: Eigene Erhebung*

	Konfusionsmatrix		Σ	Overall Accuracy	Cloud Failure	Clear Failure
Juli bis Dezember 2008	67141	2785	82621	87,16 %	10,44 %	36,39 %
SVM - Tag	7827	4868				
Juli bis Dezember 2008	64305	665	82621	86,29 %	14,22 %	8,67 %
SAFNWC - Tag	10663	6988				
Juli bis Dezember 2008	98837	9700	136880	80,83 %	14,33 %	45,10 %
SVM - Nacht	16534	11809				
Juli bis Dezember 2008	92906	3816	136880	80,80 %	19,47 %	17,74 %
SAFNWC - Nacht	22465	17693				
Juli bis Dezember 2008	67700	5483	89362	83,12 %	12,42 %	45,45 %
SVM - > 300 m	9599	6580				
Juli bis Dezember 2008	63035	1692	89362	82,14 %	18,45 %	14,03 %
SAFNWC - > 300 m	14264	10371				
Juli bis Dezember 2008	98278	7002	130139	83,28 %	13,06 %	40,95 %
SVM - < 300 m	14762	10097				
Juli bis Dezember 2008	94176	2779	130139	83,37 %	16,69 %	16,25 %
SAFNWC - < 300 m	18864	14320				

7 Ergebnisse und Anwendungsbeispiele für das Jahr 2008

7.1 Anwendungsbeispiele der SVM-Wolkenklassifizierung

Obwohl eine Validierung der weiter differenzierten SVM-Wolkenklassifizierung mit Hilfe von synoptischen Beobachtungsdaten nicht ohne erhebliche Probleme durchführbar ist, soll dennoch im Rahmen der Ergebnisse, eine Begutachtung durch den visuellen Vergleich mit den SAFNWC-Produkten erfolgen. Um dies an Einzelbeispielen, aber auch bezogen auf den gesamten Datensatz umsetzen zu können, müssen die Klassen des Produkts SAFNWC-Cloudtype an die durch die SVM-Wolkenklassifizierung erkennbaren Klassen angepasst werden. Mit SAFNWC-Cloudtype können 20 Kategorien (vgl. Tabelle 3.3) unterschieden werden, welche auf die 5 Klassen (inklusive Klasse „Wolkenfrei“) der SVM-Klassifizierung zusammenzufassen sind. Hierbei werden alle „niedrigen“ Klassen („Very Low“ und „Low“, „Cumuliform“ und „Stratiform“) der Klasse „Niedrig Kompakt“ zugeordnet. Gleiches gilt für die „hohen“ Klassen („Medium“, „High“ und „Very High“), wobei zwischen „Opaque“ und „Semitransparent“ unterschieden wird, was zu den Klassen „Hoch Kompakt“ und „Hoch Transparent“ führt. Die Klasse „Niedrig Transparent“ bezieht sich bei der SVM-Klassifizierung im Wesentlichen auf teilweise bewölkte Pixel im niedrigen Höhengniveau, weshalb hier die SAFNWC-Klasse „Fractional Clouds“ zugeordnet wird. Zwar kann es sich bei Wolken im mittleren bis hohen Niveau auch um teilweise Bewölkung handeln, dies wird jedoch durch die „hoch transparenten“ Klassen besser widerspiegelt.

Abbildung 7.1 zeigt den Vergleich der beiden Klassifizierungsverfahren anhand mehrerer Beispiele, zu unterschiedlichen Jahreszeiten. Da sowohl die SVM- als auch die SAFNWC-Klassifizierung auf einer zuvor umgesetzten Wolkenmaske basieren, sind die wolkenfreien Flächen vorab ausmaskiert. Bei dem betrachteten Ausschnitt handelt es sich um den Großraum Deutschland ohne Meeresflächen (vgl. Abbildung 5.2). An den Anwendungsbeispielen wird deutlich, dass hohe transparente, als auch kompakte Wolken von beiden Verfahren relativ ähnlich eingestuft werden. Nur niedrige transparente, also teilweise Bewölkung wird durch den SVM-Klassifikator in höherem Maße erfasst, als dies beim SAFNWC-Verfahren der Fall ist, vor allem an den Randbereichen zwischen Wolken und wolkenfreien Flächen.

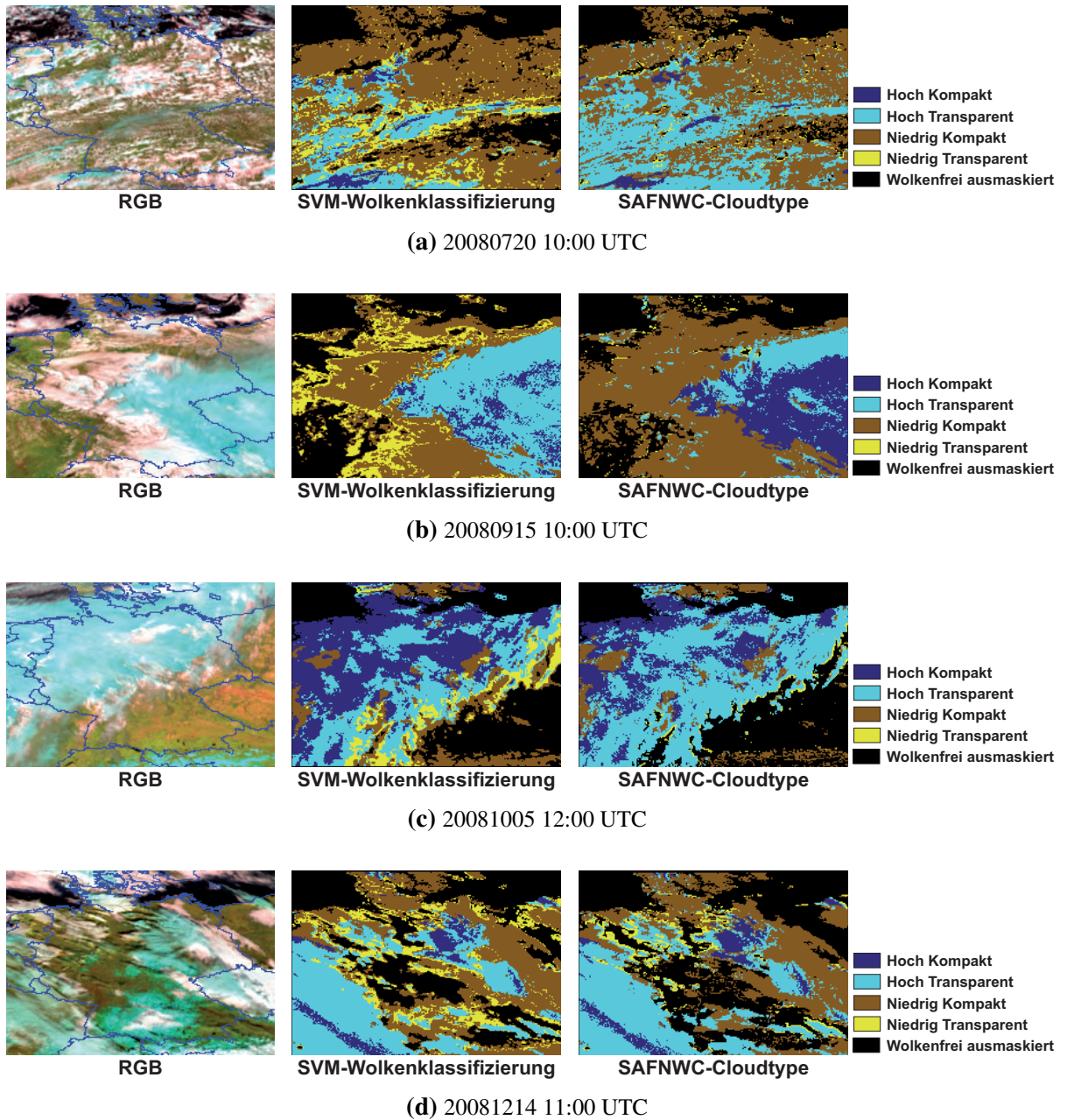


Abb. 7.1: Vergleich zwischen SVM-Wolkenklassifizierung und SAFNWC-Cloudtype anhand verschiedener Einzelbeispiele. *Quelle: Eigene Darstellung*

7.2 Ergebnisdarstellung anhand des Wolkenbedeckungsgrades

Um neben den Einzelbeispielen, die Ergebnisse des gesamten Zeitraumes 2008 darstellen zu können, wurde der Wolkenbedeckungsgrad über der Fläche von Deutschland berechnet. Beim Wolkenbedeckungsgrad²⁵ handelt es sich um den Anteil der Wolkenbedeckung über einer Fläche für einen bestimmten Zeitpunkt oder Zeitraum. Hiermit können die unterschiedlichen Detektions- bzw. Klassifizierungsverfahren zeitlich und räumlich verglichen werden. Zusätzlich sind mögliche Auswirkungen der Wolkenbedeckung, beispielsweise auf Strahlungsvorgänge regional besser abzuschätzen.

Die Berechnung basiert für die SVM-Wolkenmaske auf den wolkenkontaminierten Pixeln, während sie bei der weiter differenzierten SVM-Wolkenklassifizierung nur für die Pixel einer Klasse durchgeführt wird. Bei den Ergebnissen der SAFNWC-Produkte wird nach gleichem Muster verfahren.

In Bezug auf die Wolkenmasken erfolgt die Auswertung monatsweise, wobei alle zur Verfügung stehenden Aufnahmen eines Monats (96 x Tage des Monats) für den Wolkenbedeckungsgrad herangezogen werden. Anhand der Werte 1 („Wolkenkontaminiert“) und 0 („Wolkenfrei“) kann dann über alle Zeitschnitte aufsummiert und entsprechend der Bedeckungsgrad für jedes Pixel in Prozentwerte umgerechnet werden. Bei der Berechnung einzelner Wolkenklassen erhalten nur diese den Wert 1 und alle anderen Klassen den Wert 0. Hierdurch lässt sich die Bedeckung nur der jeweilig betrachteten Klasse ableiten.

Da die Bestimmung des Wolkenbedeckungsgrades auf der Verfügbarkeit der entsprechenden Produkte bzw. Klassifizierungsergebnisse des Jahres 2008 basiert, zeigt Tabelle 7.1, neben der Anzahl der METEOSAT-Daten (Ausschnitt Deutschland), alle vorhandenen Detektions- und Klassifizierungsdaten. Um in diesem Zusammenhang Vollständigkeit zu gewährleisten, wurden alle Daten, die nicht im Real-Time-Modus erstellt und abgespeichert werden konnten, nachprozessiert. Fehlende METEOSAT-Daten sind auf die Nicht-Verfügbarkeit vom Satelliten bzw. EUMETSAT selbst zurückzuführen. Zusätzlich fehlende Daten bei der SVM-Wolkenklassifizierung sind mit nicht klassifizierbaren Situationen des entsprechenden Zeitschnitts zu erklären. Beispielsweise wenn keine Trainingsdaten gewonnen werden konnten, oder wenn Trainingsdaten nur einer Klasse zur Verfügung standen. Fehlende Daten der SAFNWC-Produkte sind direkt mit der Verfügbarkeit der METEOSAT-Daten (HRIT Level 1.5) gekoppelt.

²⁵In der Literatur findet sich keine eindeutige Definition.

Tab. 7.1: Datenverfügbarkeit der METEOSAT-Daten (alle Kanäle) und der daraus erstellten Produkte für das Jahr 2008. SVM-Wolkenmaske (SVM-WM) und SVM-Wolkenklassifizierung (SVM-WK), sowie SAFNWC-CMa und SAFNWC-CT. Die Daten setzen sich aus 96 Zeitschnitten pro Tag zusammen, alle vorhandenen METEOSAT-Kanäle werden zu einem Wert zusammengefasst. Vom 14.05.2008 00:00 UTC bis 19.05.2008 12:00 UTC und vom 01.12.2008 12:00 UTC bis 09.12.2008 12:00 UTC wurde jeweils auf METEOSAT-8 umgeschaltet. *Quelle: Eigene Aufstellung*

Monat	MSG	SVM-WM	SVM-WK	SAFNWC-CMa	SAFNWC-CT
Jan (2976)	2976	2974	2974	2976	2976
Feb (2784)	2784	2776	2770	2784	2784
Mär (2976)	2972	2972	2972	2971	2971
Apr (2880)	2879	2879	2879	2880	2880
Mai (2976)	2955	2950	2950	2897	2897
Jun (2880)	2879	2879	2879	2879	2879
Jul (2976)	2976	2976	2974	2976	2976
Aug (2976)	2976	2976	2975	2976	2976
Sep (2880)	2880	2880	2880	2880	2880
Okt (2976)	2976	2976	2975	2976	2976
Nov (2880)	2879	2875	2875	2877	2877
Dez (2976)	2976	2969	2969	2975	2975

7.2.1 SVM-Wolkenmaske im Vergleich mit SAFNWC-Cloudmask

Die Ergebnisse des Wolkenbedeckungsgrades, errechnet aus der SVM-Wolkenmaske, sind für alle Monate des Jahres 2008 in den Abbildungen 7.2 bis 7.7 dargestellt. Jeweils im Vergleich mit den Daten der SAFNWC-Cloudmask. Zusätzlich wurde die Differenz zwischen beiden Verfahren berechnet, um Unterschiede auch auf räumlicher Ebene deutlich zu machen.

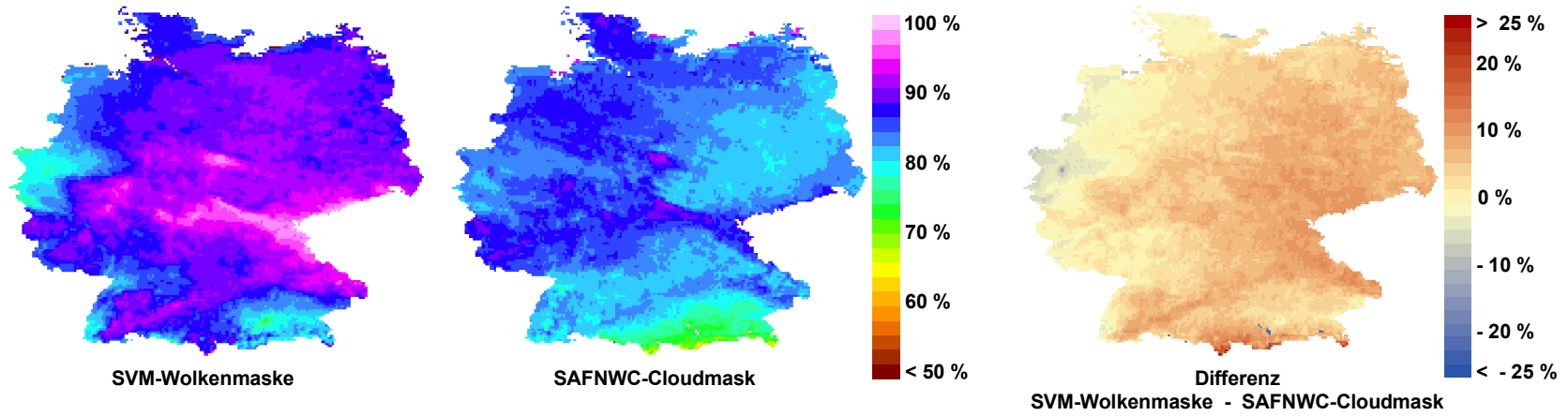
Auf den ersten Blick ist zu erkennen, dass die Werte der SVM-Wolkenmaske, im Vergleich zur SAFNWC-Cloudmask, meist höher sind. Von der räumlichen Ausprägung ergibt sich jedoch bei beiden Verfahren ein sehr ähnliches Muster. Vor allem die Mittelgebirge weisen im Allgemeinen einen höheren Bedeckungsgrad auf, was mit verstärkten Staulagen an diesen Erhebungen erklärt werden kann (vgl. Kapitel 2.3.2). Das Rheintal, sowie Teile Nord- und Süddeutschlands weisen dagegen meist geringere Werte auf.

Der Vergleich der Verfahren (Differenz) zeigt, dass insbesondere die Monate Februar, Mai und Dezember hohe Abweichungen aufweisen. Hier scheint die SVM-Wolkenmaske verstärkt zu überschätzen, während bei der SAFNWC-Cloudmask vorrangig eine Unterschätzung zu vermuten ist. Bei den Validierungsergebnissen des Monats Dezember wurde dies schon deutlich. Die Betrachtung des allgemeinen Wettergeschehens in besagten Monaten²⁶ scheint darauf hinzuweisen, dass die SVM-Wolkendetektion, vor allem bei ausgeprägten Hochdruckwetterlagen,

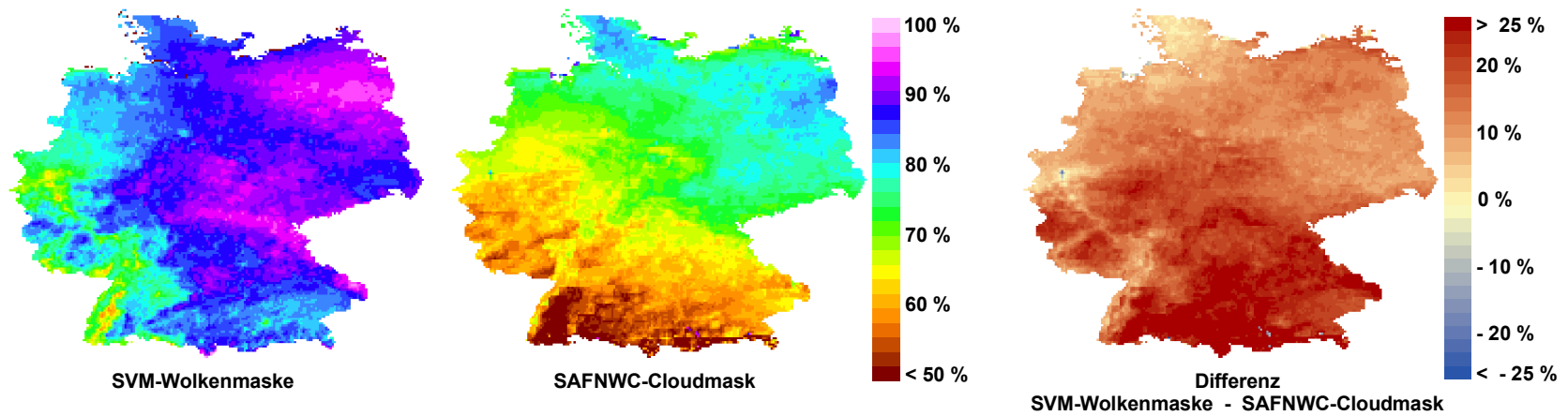
²⁶http://www.wetteronline.de/wotexte/redaktion/rueckblick/2008/12/31_ys.htm
[25.02.2011]

verbunden mit vielen wolkenfreien Situationen, mit Problemen behaftet ist (vgl. Abbildung 6.3). An dieser Stelle ist zu vermuten, dass bei der Abschätzung des ungefähren Grenzwertes und der Trainingsdatengewinnung, Teile ausgeprägter wolkenfreier Flächen, fälschlicherweise als Wolke eingestuft werden (vgl. Seite 107). Der SVM-Klassifikator sollte zwar mit wenigen Trainingspixeln, sowie mit geringfügigen Überschreitungen in Richtung der anderen Klasse zurecht kommen (vgl. hierzu Foody und Mathur 2006), dies kann jedoch in Zusammenhang mit großflächig wolkenfreien Situationen, zu einer ungenauen Trennung durch die Hyperebene führen. Obwohl die Validierungsergebnisse gezeigt haben, dass es im Rahmen der SVM-Wolkendetektion zu Überschätzungen kommt, besteht bei Hochdruckwetterlagen dennoch die Möglichkeit dünner Cirrus-Schleier in großer Höhe oder vereinzelte Quellwolken (Cumulus Humilis) auf Subpixel-Ebene, welche bei Betrachtung der sichtbaren Kanäle kaum zu erkennen sind (siehe Abbildung 6.3). In den Wintermonaten kann es zudem zu tiefliegendem Nebel kommen, der sich selbst im Tagesverlauf nur schwer auflöst. Solche Effekte können je nach Verfahren sowohl zur Über- als auch Unterschätzung führen.

In den weiteren Sommermonaten Juni, Juli und August ist die positive Differenz der beiden Detektionsverfahren weniger stark vertreten und in den restlichen Monaten verhalten sich beide Verfahren sehr ähnlich mit geringfügigen räumlichen Abweichungen in die eine oder andere Richtung (vgl. Abbildung 7.3, 7.6 und 7.7). Bei Wetterlagen mit ausgeprägter Bewölkung und vergleichsweise weniger wolkenfreier Flächen, wie im März und April (vgl. Kapitel 8), scheint die SVM-Wolkendetektion sehr gut zu funktionieren.

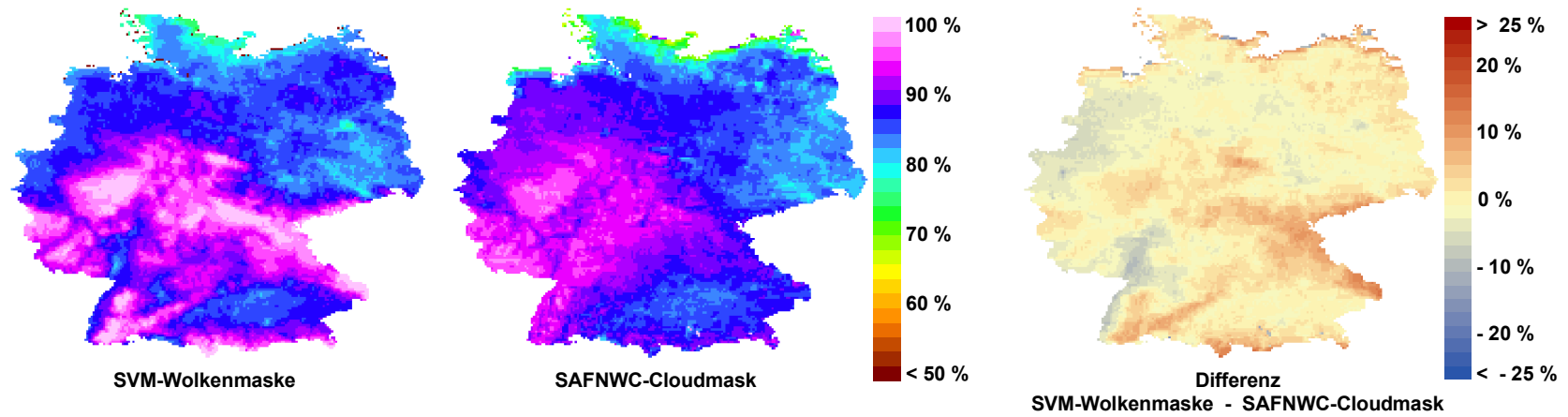


(a) Wolkenbedeckungsgrad Januar 2008

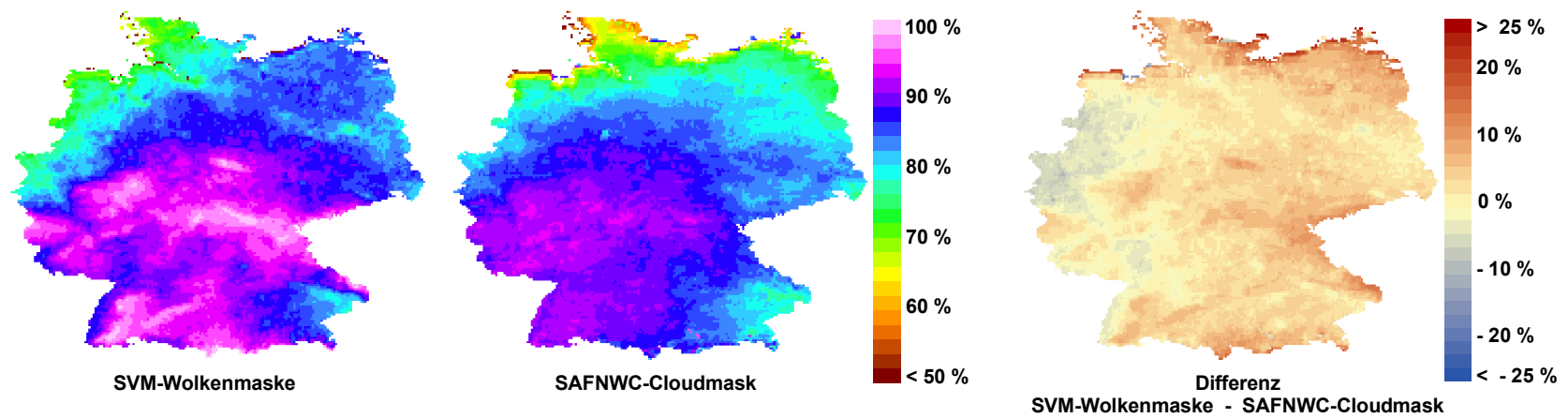


(b) Wolkenbedeckungsgrad Februar 2008

Abb. 7.2: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Januar und Februar 2008. *Quelle: Eigene Berechnung und Darstellung*

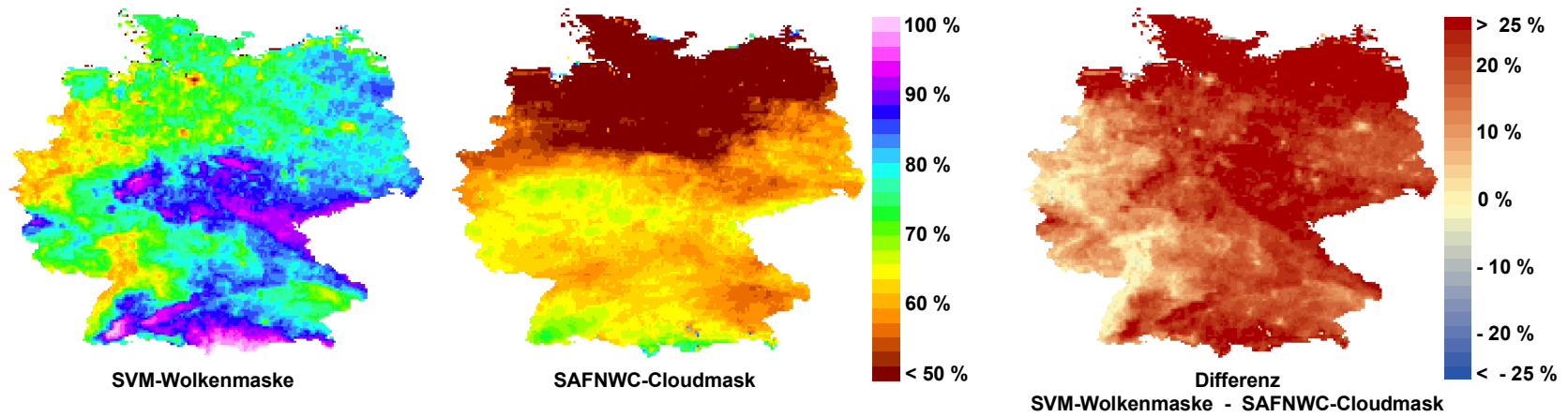


(a) Wolkenbedeckungsgrad März 2008

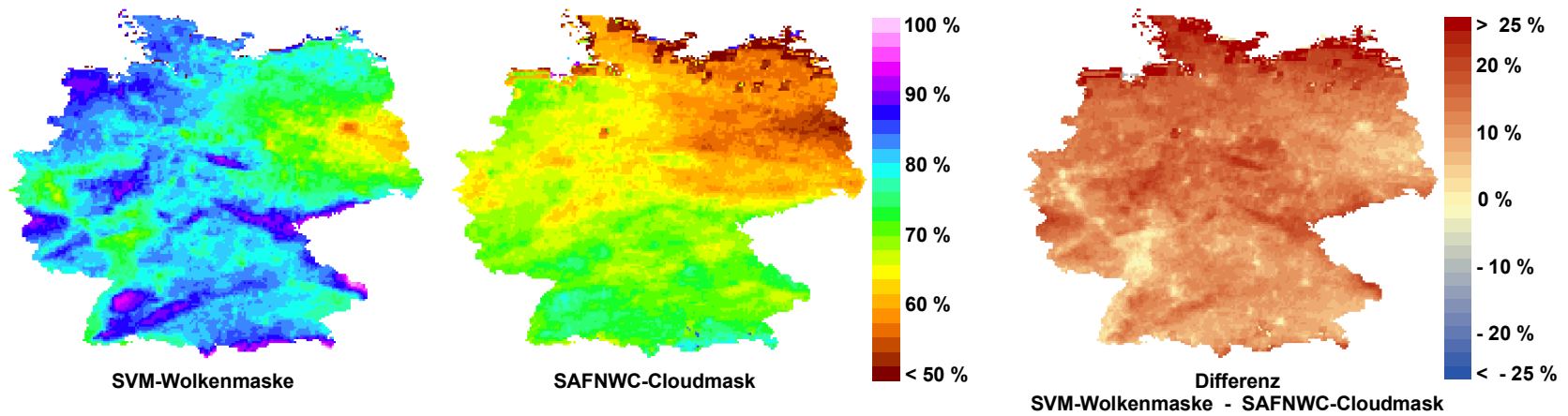


(b) Wolkenbedeckungsgrad April 2008

Abb. 7.3: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. März und April 2008. *Quelle: Eigene Berechnung und Darstellung*

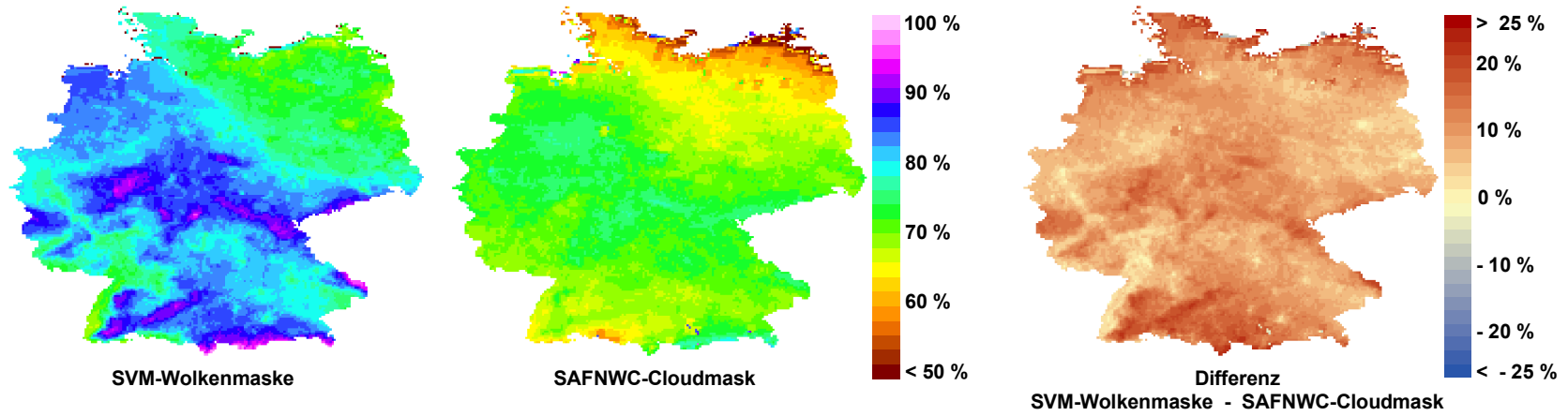


(a) Wolkenbedeckungsgrad Mai 2008

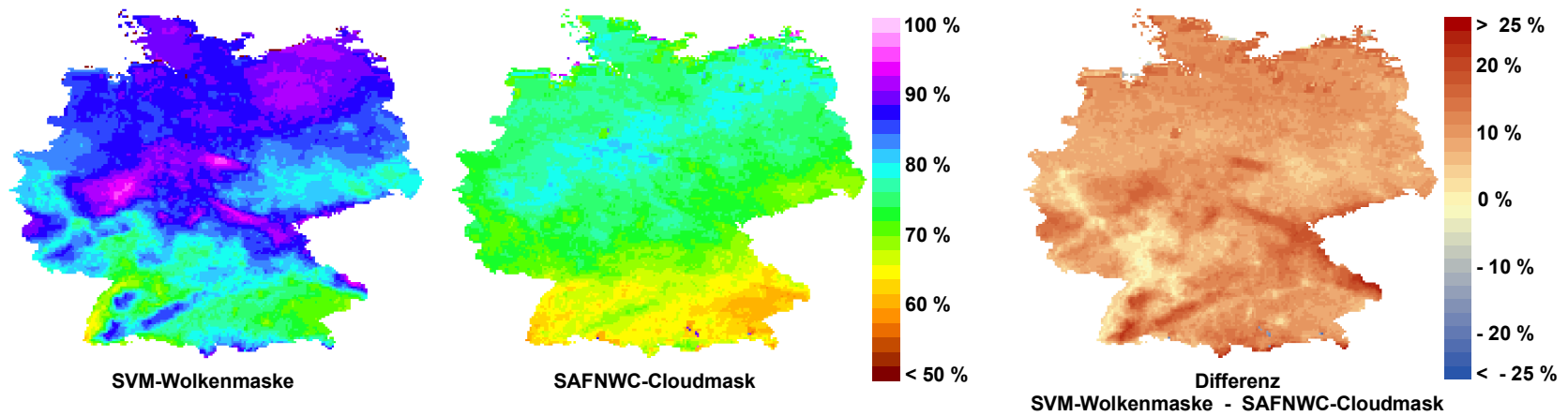


(b) Wolkenbedeckungsgrad Juni 2008

Abb. 7.4: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Mai und Juni 2008. *Quelle: Eigene Berechnung und Darstellung*

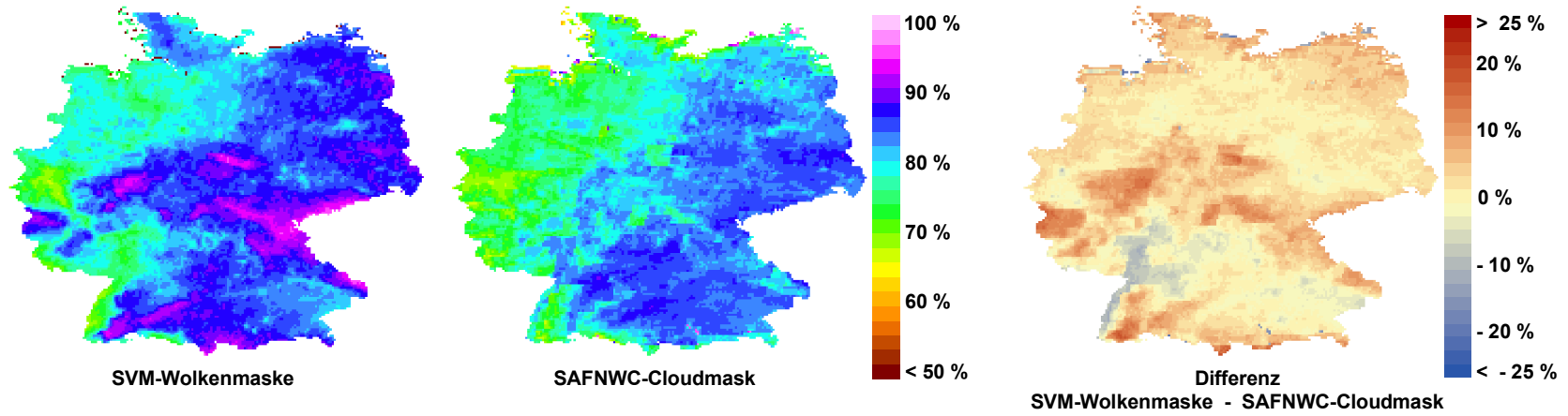


(a) Wolkenbedeckungsgrad Juli 2008

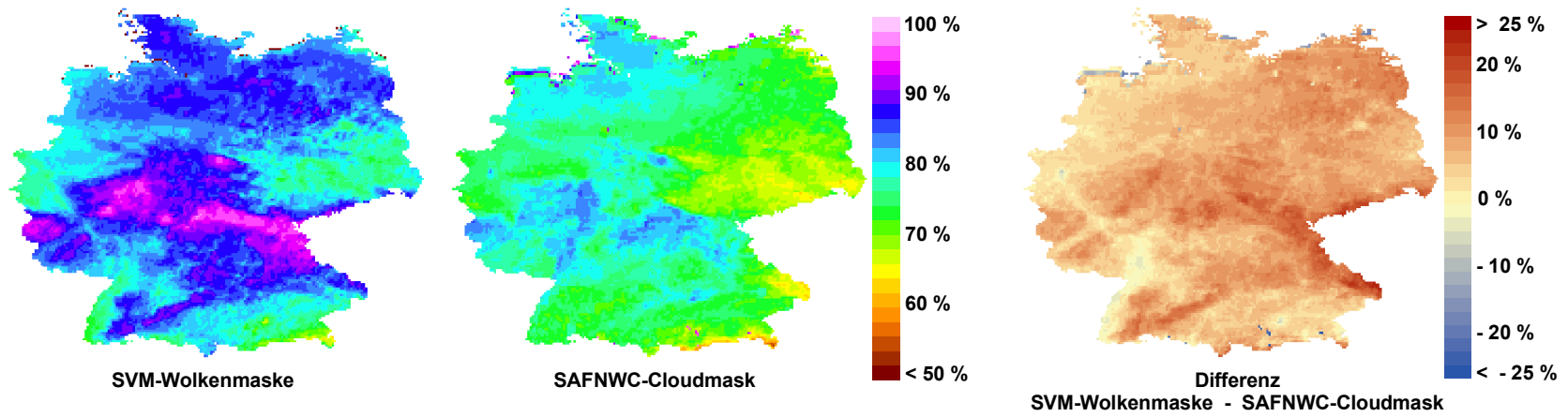


(b) Wolkenbedeckungsgrad August 2008

Abb. 7.5: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. Juli und August 2008. *Quelle: Eigene Berechnung und Darstellung*

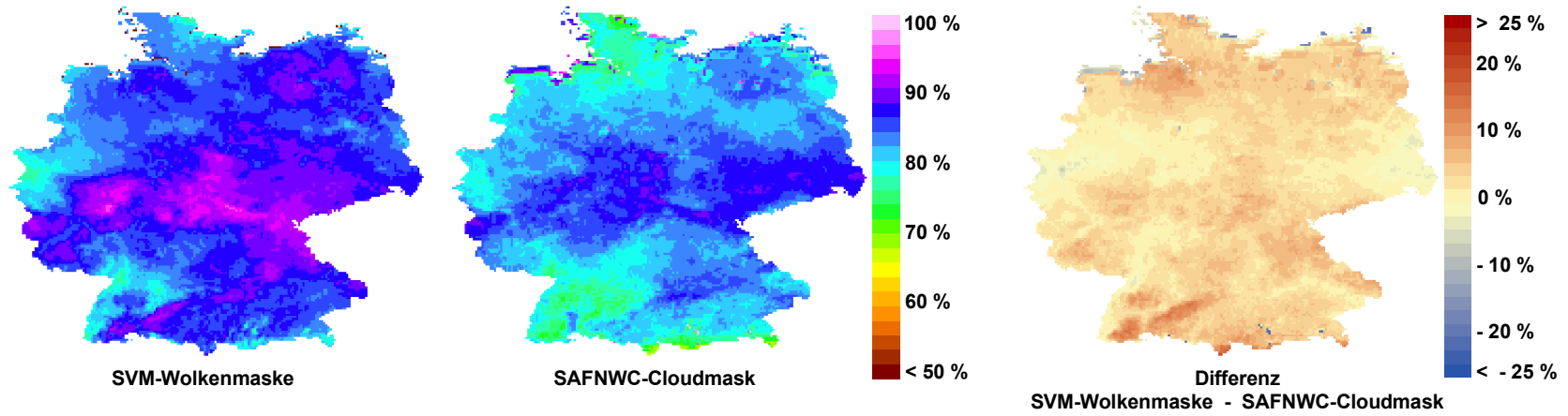


(a) Wolkenbedeckungsgrad September 2008

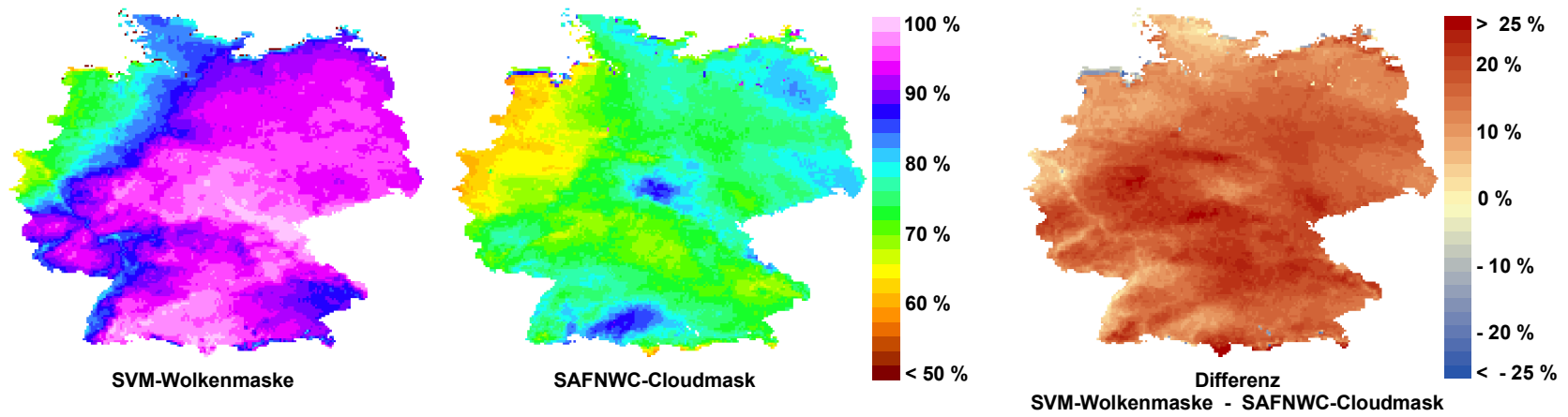


(b) Wolkenbedeckungsgrad Oktober 2008

Abb. 7.6: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. September und Oktober 2008. *Quelle: Eigene Berechnung und Darstellung*



(a) Wolkenbedeckungsgrad November 2008



(b) Wolkenbedeckungsgrad Dezember 2008

Abb. 7.7: Wolkenbedeckungsgrad errechnet aus der Summe aller vorhandener Einzelzeitschnitte. Nur Berücksichtigung der wolkenkontaminierten Pixel. November und Dezember 2008. *Quelle: Eigene Berechnung und Darstellung*

7.2.2 SVM-Wolkenklassifizierung im Vergleich mit SAFNWC-Cloudtype

Wie bei den Wolkenmasken, wird der Wolkenbedeckungsgrad auch für die einzelnen Klassen der SVM-Wolkenklassifizierung berechnet. Hierbei erfolgt nur die Betrachtung der jeweiligen Klasse, wobei Überlagerungen verschiedener Wolkenarten nicht berücksichtigt werden. Das heißt, wenn beispielsweise von hoher transparenter Bewölkung die Rede ist, kann nicht ausgeschlossen werden, dass sich darunter kompakte Bewölkung befindet. Der Bedeckungsgrad der Einzel-Klassen ist in den Abbildungen 7.8 bis 7.15 dargestellt. Wegen einer besseren Übersicht sind bei dieser Vorgehensweise jeweils drei Monate zusammengefasst.

Betrachtet man die Klasse „Hoch Kompakt“ (vgl. Abbildung 7.8 und 7.9), so ist in erster Linie zu erkennen, dass die Werte des Wolkenbedeckungsgrades relativ gering ausfallen und sich die Verteilung der Werte sowohl räumlich als auch zeitlich sehr homogen verhält. Höchste Werte sind im Frühjahr, Herbst und Winter zu verzeichnen, während die Sommermonate im Allgemeinen geringere Werte aufweisen. Beim Vergleich der Klassifizierungsverfahren zeigt die SVM-Wolkenklassifizierung über alle Monate hinweg höhere Werte. Die größten Abweichungen zwischen SVM und SAFNWC ergeben sich hierbei in den ersten drei Monaten des Jahres.

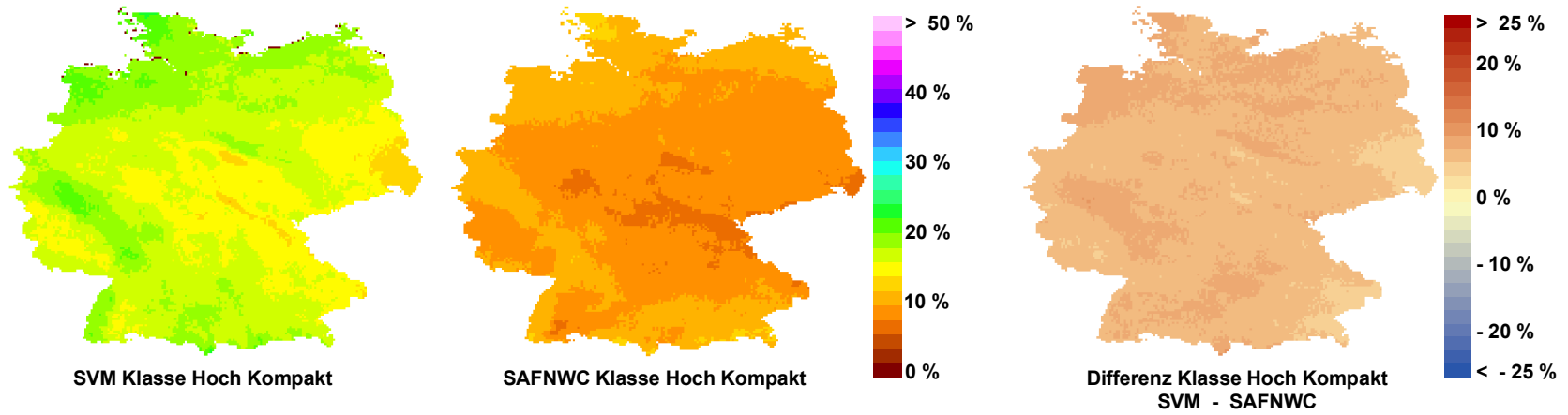
Die Klasse „Hoch Transparent“ (vgl. Abbildung 7.10 und 7.11) zeigt eine andere Situation; hier sind die Werte der SAFNWC-Cloudmask höher als die der SVM-Wolkenklassifizierung. Gleichzeitig liegen die Werte allgemein etwas höher, als die der Klasse „Hoch Kompakt“. Höchste Werte sind auch hier im Winter, Frühjahr und Herbst gegeben, während die Sommermonate vergleichsweise geringere Werte aufweisen. Räumlich betrachtet sind die höchsten Werte vor allem in den zentralen Mittelgebirgsregionen zu verzeichnen, während Teile Süd- und Norddeutschlands geringere Werte zeigen. Anhand der negativen Differenz zwischen den Verfahren ist zu vermuten, dass die SVM-Klassifizierung diese Klasse vornehmlich unterschätzt. Da es sich aber um einen Vergleich zwischen Fernerkundungsverfahren handelt, kann man genauso vermuten, dass das SAFNWC-Verfahren zur Überschätzung dieser Klasse neigt.

Von den vier Wolkenklassen zeigt die Klasse „Niedrig Kompakt“ (vgl. Abbildung 7.12 und 7.13) im Gesamtvergleich die höchsten Werte, was darauf schließen lässt, dass diese Art der Bewölkung im Raum Deutschland häufiger vorkommt. Beide Klassifizierungsverfahren spiegeln bei der räumlichen und zeitlichen Betrachtung ein relativ ähnliches Muster wider. Nur die Monate Januar/Februar/März zeigen bei der SVM-Wolkenklassifizierung verstärkt höhere Werte in den zentralen bis südlichen Mittelgebirgsbereichen. Vor allem die Mittelgebirge sind bezogen auf diese Klasse mit hoher Bedeckung gekoppelt. Im nördlichen Teil Deutschlands und in Teilen des Rheintals ergibt sich dagegen eine geringere Wolkenbedeckung. Dies kann, wie bereits erwähnt, mit Staulagen bzw. Föhn effekten (vgl. hierzu Kapitel 2.3.2) im Bereich der Mittelgebirge erklärt werden. Je nach Durchzug bestimmter Front-Systeme (meist aus West

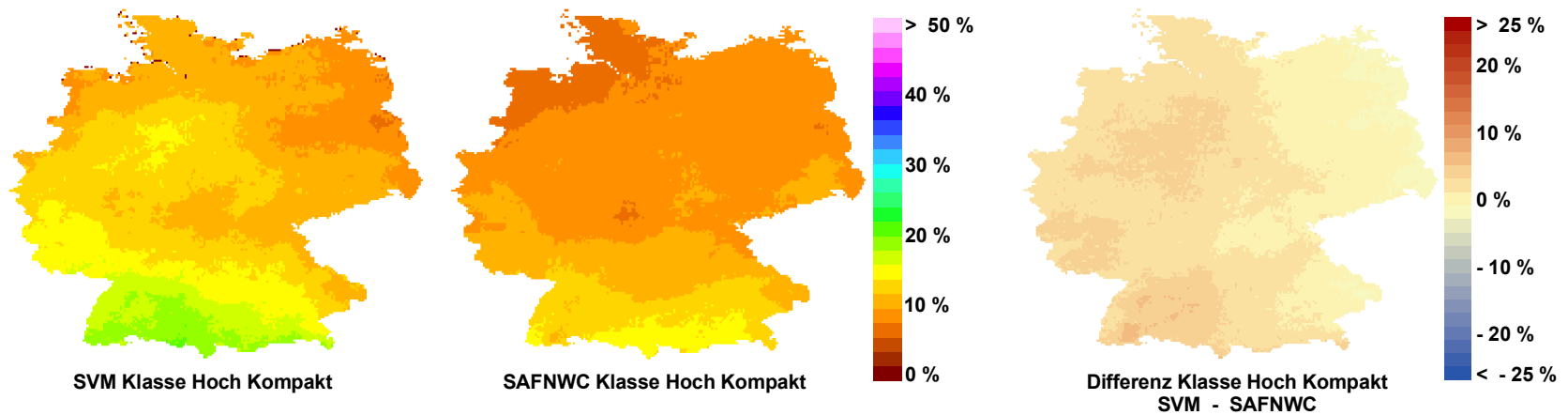
bzw. Süd-West) im Bereich von Tiefdruckgebieten (vgl. Kapitel 2.3.1), kommt es an den Mittelgebirgen zu verstärkter Ausprägung von dichten, meist stratiformen Bewölkungsverhältnissen.

Die Klasse „Niedrig Transparent“ (vgl. Abbildung 7.14 und 7.15) zeigt im Gegensatz zu der zuvor beschriebenen Klasse „Niedrig Kompakt“ wieder einen geringeren Wolkenbedeckungsgrad. Dennoch kommt es vornehmlich zu höheren Werten bei der SVM-Klassifizierung. Dies zeigt sich vor allem in den Monaten April bis September. In den Wintermonaten stellt sich die Situation beider Verfahren wieder ausgeglichener dar. Eine räumliche Differenzierung lässt sich in dieser Klasse nur schwer abschätzen, da beide Verfahren die Bedeckung mit unterschiedlicher regionaler Ausprägung darstellen. Die SAFNWC-Klassifizierung zeigt eine verstärkt homogene Verteilung der Werte, mit einer Tendenz zur stärkeren Bedeckung im Norden und Süd-Osten. Im Vergleich dazu zeigt die SVM-Klassifizierung höchste Bedeckungswerte im zentralen Bereich Deutschlands, aber auch in Teilen Norddeutschlands. Durch die hohe positive Differenz der Verfahren scheint die in Kapitel 6 beschriebene Überschätzung verstärkt auf Bewölkungsverhältnisse dieser Klasse zurückzuführen sein, also teilweise Bewölkung im Bereich des Übergangsbzw. Unsicherheitsbereiches zwischen wolkenkontaminierten und wolkenfreien Flächen. Da die einzelnen Wolkenklassen aber nicht, wie bei der Wolkenmaske, mit unabhängigen Beobachtungsdaten validiert wurden, ist diese Aussage mit Vorsicht zu betrachten.

Zusammenfassend kann man sagen, dass für den Raum Deutschland im Jahr 2008 insgesamt hohe Wolkenbedeckungsgrade erreicht wurden, wobei durch die SVM-Wolkendetektion (wolkenkontaminierte Pixel) in der Regel eine stärkere Bedeckung dargestellt wird, als dies bei der SAFNWC-Cloudmask der Fall ist. Vor allem ausgeprägte Hochdruckwetterlagen mit verbreiteten wolkenfreien Flächen scheinen in diesem Zusammenhang zu einer Überschätzung durch die SVM-Wolkendetektion zu führen (Februar, Mai und Dezember). Die regionale Ausprägung der Differenz zwischen den Verfahren macht zwar deutlich wo die Abweichungen am größten sind, ob es an diesen Stellen aber zu einer Über- oder Unterschätzung des jeweiligen Detektionsverfahrens kommt, kann nur bedingt beantwortet werden. Durch die gemeinsame Betrachtung des Wolkenbedeckungsgrades und der Validierungsergebnisse (vgl. Kapitel 6.2) können aber gewisse Hinweise abgeleitet werden. Insgesamt lässt sich bei beiden Verfahren feststellen, dass höchste Werte an den Höhenlagen der Mittelgebirge zu finden sind, während Teile Nord- und Süddeutschlands geringere Werte des Bedeckungsgrades verzeichnen. Bei der Einzelbetrachtung der Klasse „Niedrig Kompakt“ sieht dies sehr ähnlich aus, da diese den Hauptwolkenanteil im Raum Deutschland zu repräsentieren scheint. Die räumliche Verteilung des Bedeckungsgrades der anderen Wolkenklassen stellt sich eher homogen dar und weist im Vergleich wesentlich geringere Bedeckungswerte auf.

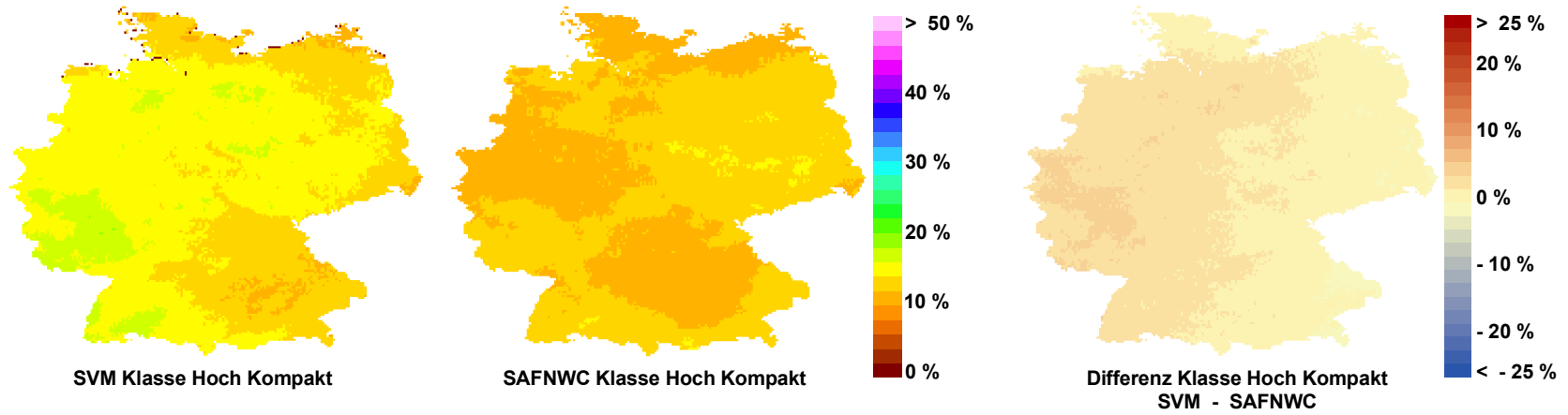


(a) Wolkenbedeckungsgrad Januar/Februar/März 2008

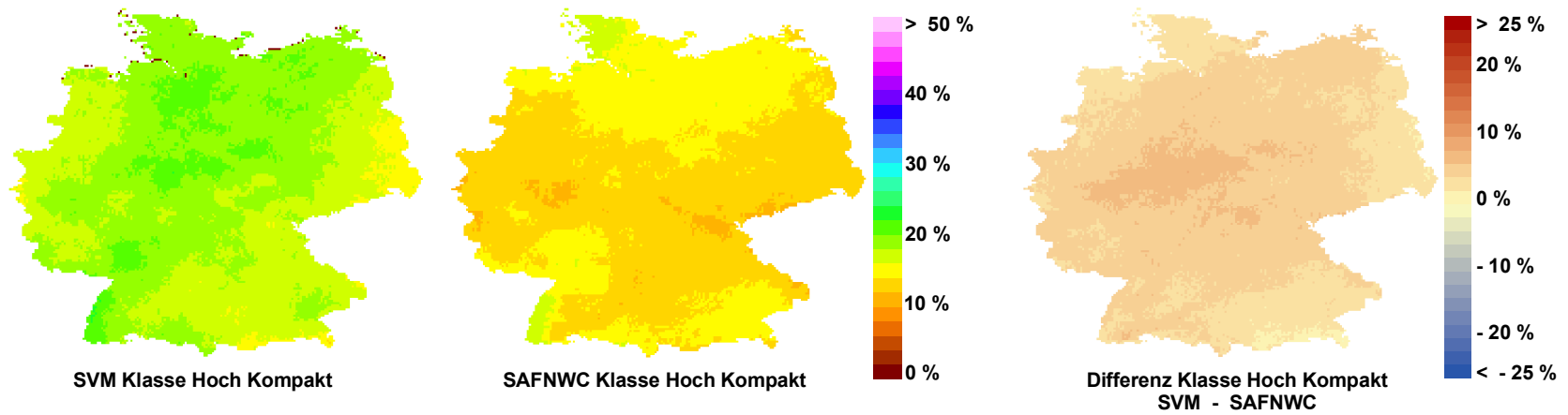


(b) Wolkenbedeckungsgrad April/Mai/Juni 2008

Abb. 7.8: Wolkenbedeckungsgrad der Klasse „Hoch Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008. *Quelle: Eigene Berechnung und Darstellung*

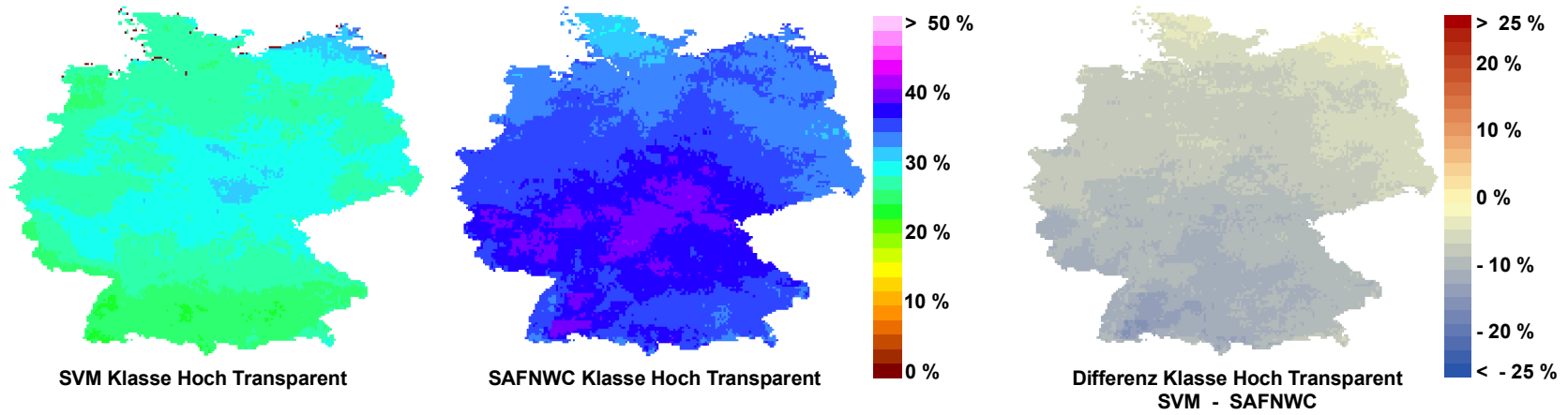


(a) Wolkenbedeckungsgrad Juli/August/September 2008

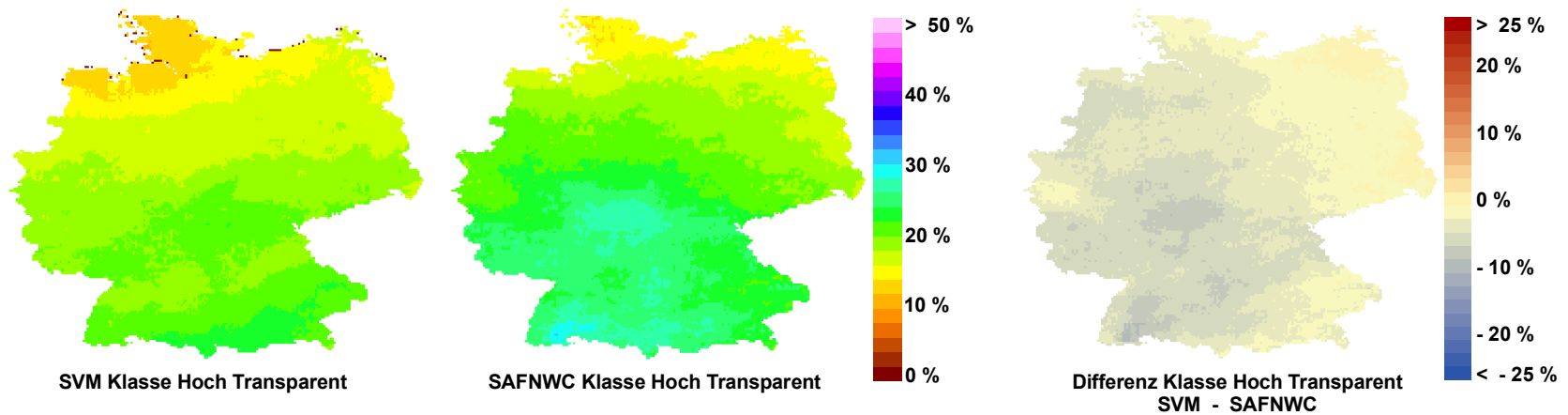


(b) Wolkenbedeckungsgrad Oktober/November/Dezember 2008

Abb. 7.9: Wolkenbedeckungsgrad der Klasse „Hoch Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008. *Quelle: Eigene Berechnung und Darstellung*

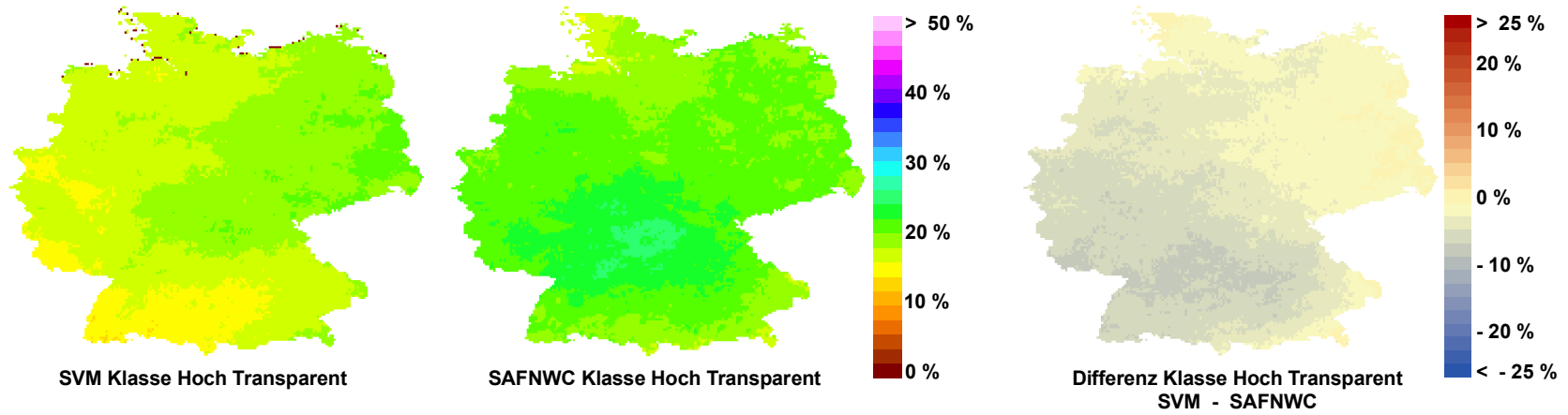


(a) Wolkenbedeckungsgrad Januar/Februar/März 2008

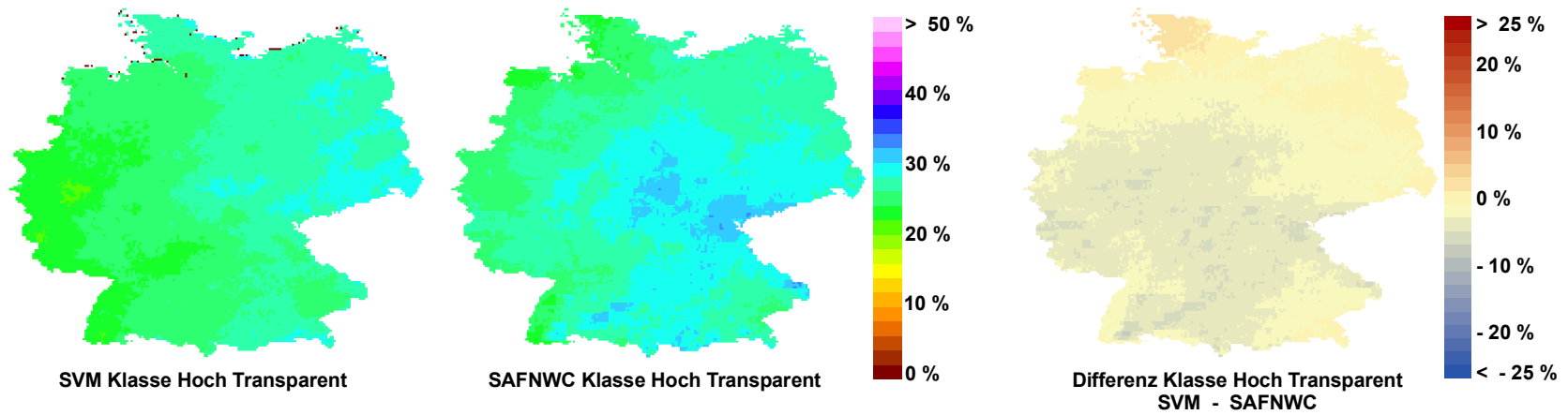


(b) Wolkenbedeckungsgrad April/Mai/Juni 2008

Abb. 7.10: Wolkenbedeckungsgrad der Klasse „Hoch Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008. *Quelle: Eigene Berechnung und Darstellung*

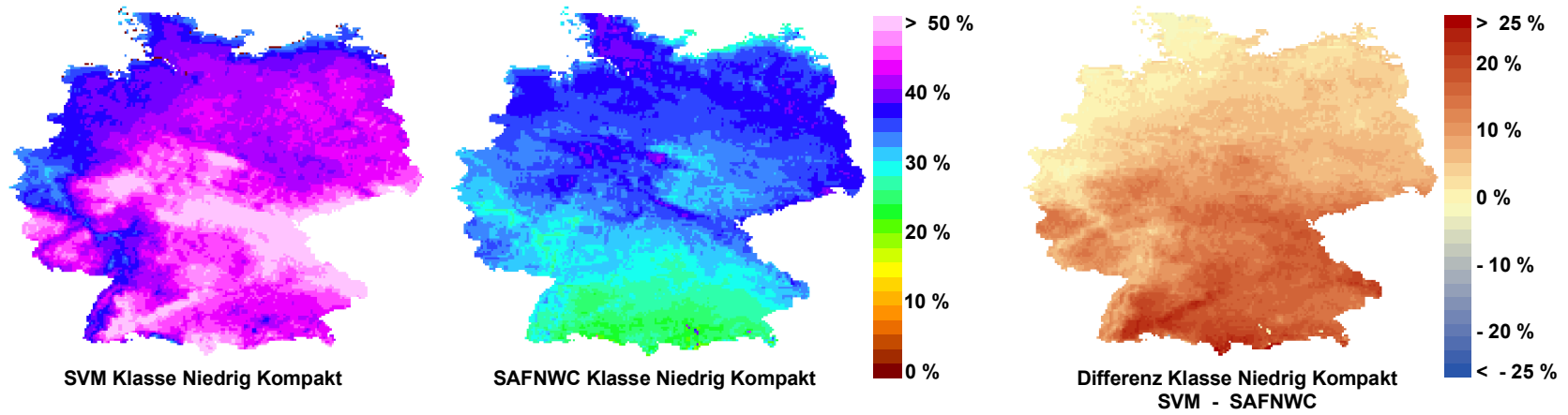


(a) Wolkenbedeckungsgrad Juli/August/September 2008

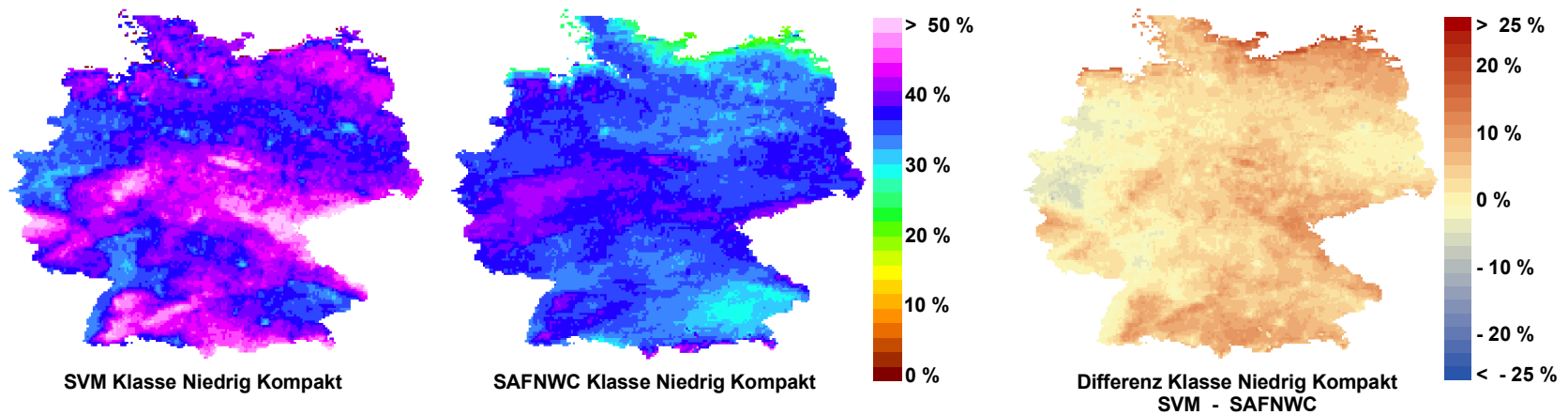


(b) Wolkenbedeckungsgrad Oktober/November/Dezember 2008

Abb. 7.11: Wolkenbedeckungsgrad der Klasse „Hoch Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008. *Quelle: Eigene Berechnung und Darstellung*

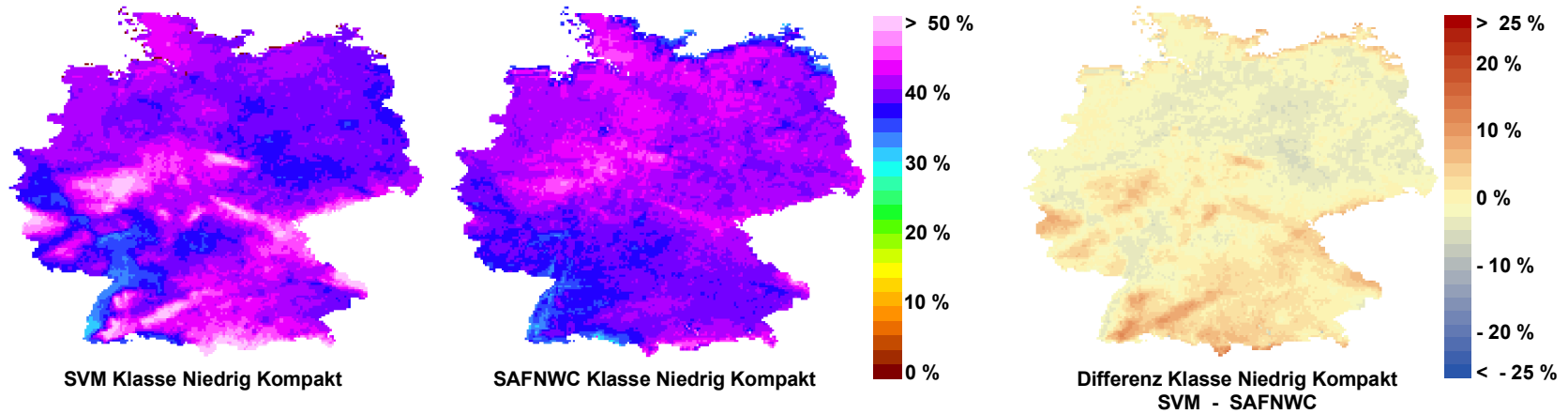


(a) Wolkenbedeckungsgrad Januar/Februar/März 2008

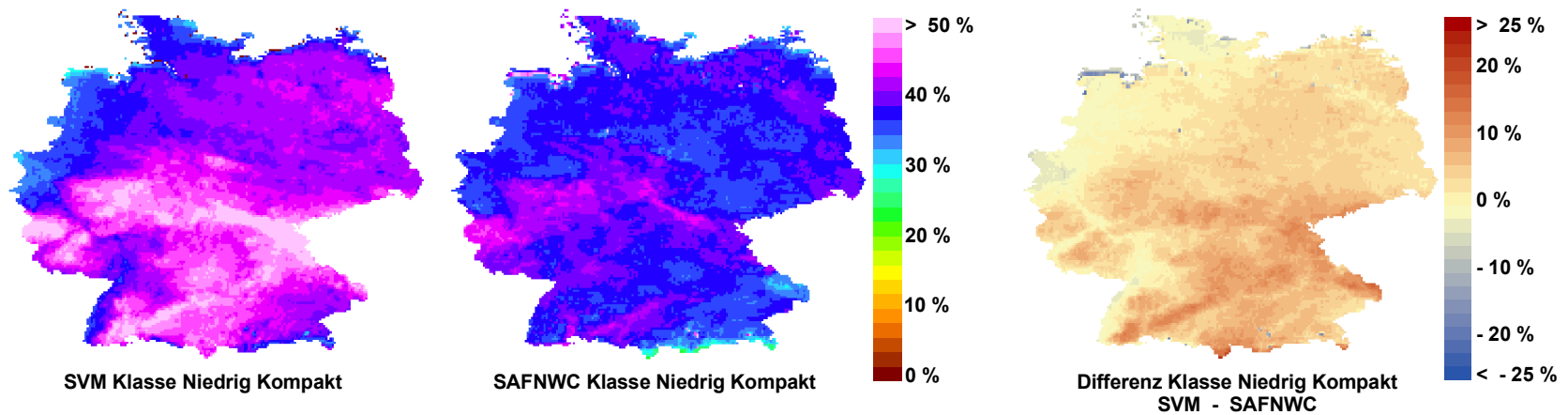


(b) Wolkenbedeckungsgrad April/Mai/Juni 2008

Abb. 7.12: Wolkenbedeckungsgrad der Klasse „Niedrig Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008. *Quelle: Eigene Berechnung und Darstellung*

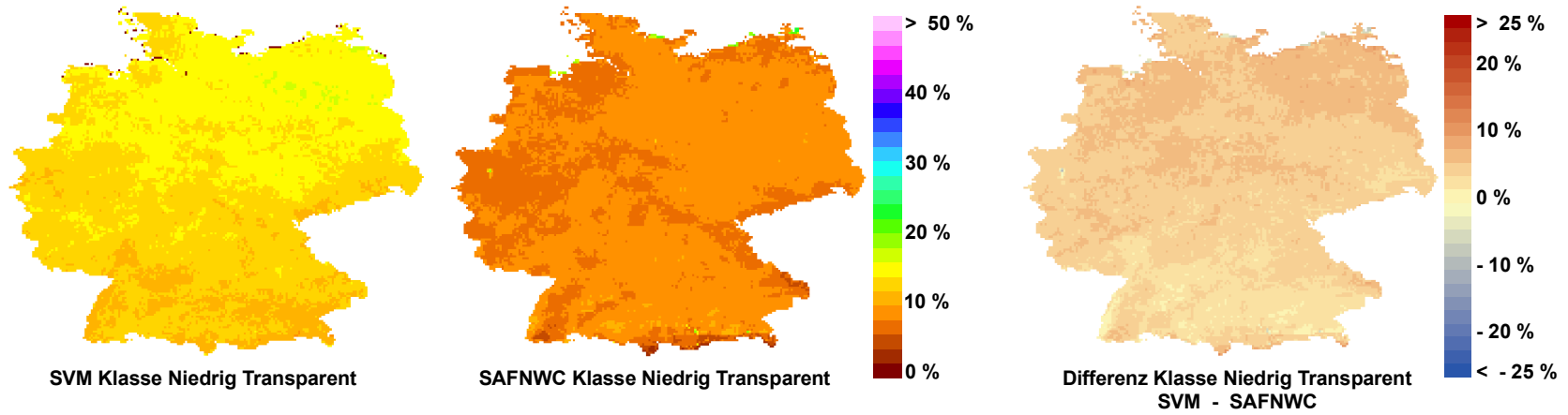


(a) Wolkenbedeckungsgrad Juli/August/September 2008

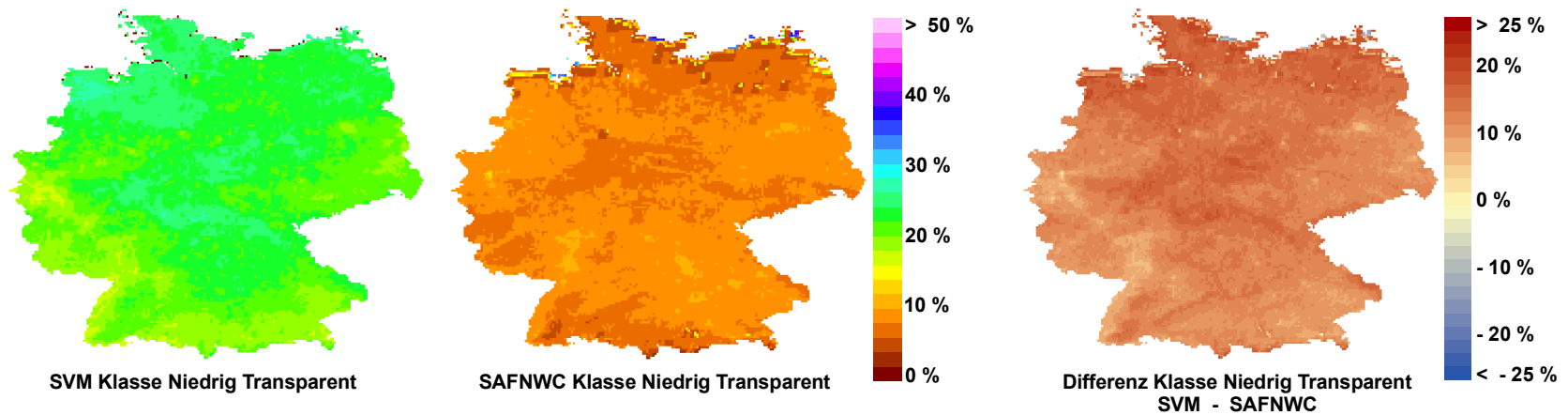


(b) Wolkenbedeckungsgrad Oktober/November/Dezember 2008

Abb. 7.13: Wolkenbedeckungsgrad der Klasse „Niedrig Kompakt“ errechnet aus der Summe der Einzelzeitschnitte. Juli bis Dezember 2008. *Quelle: Eigene Berechnung und Darstellung*

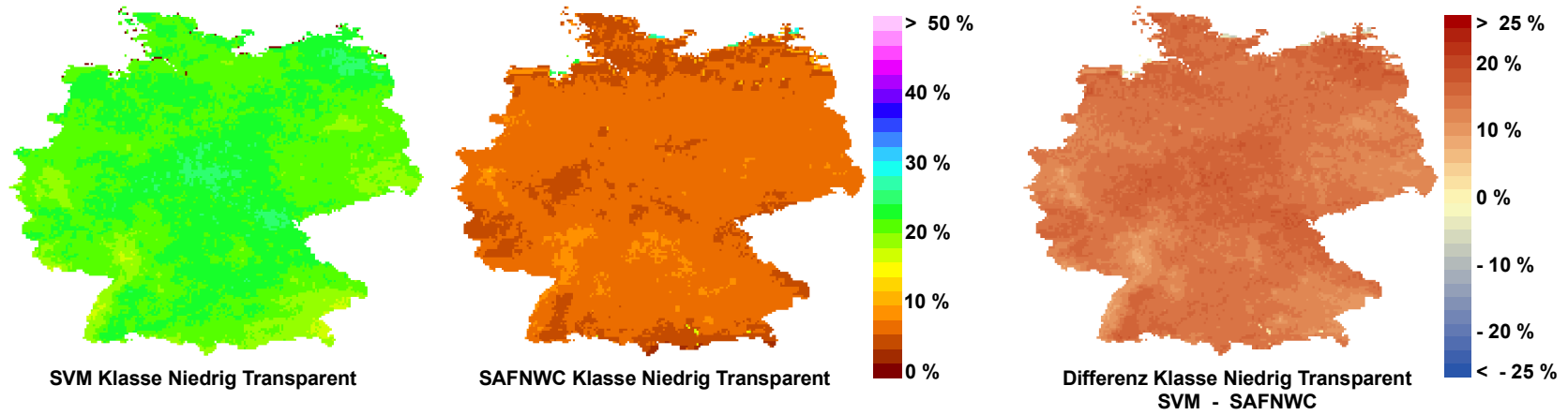


(a) Wolkenbedeckungsgrad Januar/Februar/März 2008

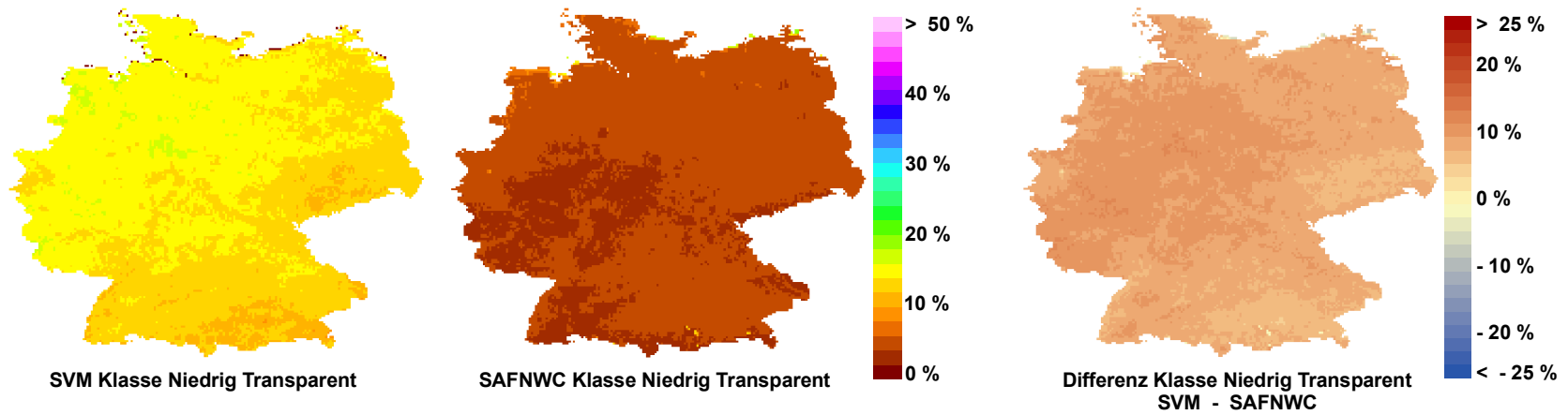


(b) Wolkenbedeckungsgrad April/Mai/Juni 2008

Abb. 7.14: Wolkenbedeckungsgrad der Klasse „Niedrig Transparent“ errechnet aus der Summe der Einzelzeitschnitte. Januar bis Juni 2008. *Quelle: Eigene Berechnung und Darstellung*



(a) Wolkenbedeckungsgrad Juli/August/September 2008



(b) Wolkenbedeckungsgrad Oktober/November/Dezember 2008

Abb. 7.15: Wolkenbedeckungsgrad der Klasse „Niedrig Transparent“ errechnet aus der Summe der Einzelzeitschnitte, Juli bis Dezember 2008. *Quelle: Eigene Berechnung und Darstellung*

8 Einordnung der Ergebnisse mit Hilfe von Großwetterlagen

Um die Ergebnisse noch besser im Zusammenhang mit Strömungsmustern und entsprechenden Wetterlagen einordnen zu können, soll nachfolgend ein Vergleich des Wolkenbedeckungsgrades mit Hilfe der Wetterlagenklassifikation des Deutschen Wetterdienstes (DWD) dargestellt werden. Beim Deutschen Wetterdienst erfolgt die Umsetzung solcher Klassifikationen mit zwei verschiedenen Verfahren²⁷, eine subjektive nach Hess-Brezowski (vgl. Kapitel 2.3.1, Beck 2000, S. 106 - 108), die sich auf das gesamte Gebiet Mitteleuropa bezieht, und eine objektive Klassifikation angewendet auf den Raum Deutschland. Die objektive Klassifikation ist eindeutig definiert und numerisch reproduzierbar, während bei dem subjektiven Verfahren durch verschiedene Fachleute unterschiedliche Ergebnisse entstehen können. Aus diesem Grund und auch wegen der einfachen Datenverfügbarkeit²⁸ bezieht sich der Vergleich alleine auf die objektive Wetterlagenklassifikation.

Die meteorologischen Kriterien dieser Wetterlagenklassifikation basieren auf Windrichtung (Advektionsrichtung der Luftmassen), Zyklonalität (Hoch bzw. Tiefdruckeinfluss) und Feuchtigkeit der Atmosphäre. Für die Einteilung der Wetterlagen werden Daten des operationellen, numerischen Wettersvorhersagesystems des Deutschen Wetterdienstes herangezogen. Hiermit können täglich Wetterlagen in 40 mögliche Klassen eingeordnet werden. Dies geschieht seit dem 1.7.1979. In den Jahren 1997 und 2002 wurde die Erfassung auf zusätzliche Vorhersagezeitpunkte erweitert. Die Berechnung erfolgt anhand von Indizes innerhalb des Modellgitters für den betrachteten Ausschnitt Deutschland. Man verwendet einen Zyklonalitätsindex bei 950 hPa und 500 hPa, sowie einen Windindex bei 700 hPa und einen Feuchteindex für den Bereich 950 hPa - 300 hPa. Durch die Kombination und Ausprägung dieser Indizes erfolgt abschließend die Klassenzuteilung. Weitere Details des Verfahrens finden sich bei Dittmann et al. (1995) und Bissolli und Dittmann (2001). Tabelle 8.1 zeigt die Aufteilung der einzelnen Wetterlagen und ihre zugehörigen Kennzahlen.

Durch Abruf der Daten²⁹ erhält man eine Tabelle mit der jeweiligen Wetterlagen-Kennzahl für alle Tage seit Juli 1979. Hieraus können entsprechend die Werte des Jahres 2008 entnom-

²⁷Siehe hierzu unter <http://www.dwd.de> [03.03.2011] Klima + Umwelt.

²⁸Als Tabelle abrufbar unter <http://www.dwd.de> [03.03.2011].

²⁹<http://www.dwd.de> [03.03.2011] Klima + Umwelt

men werden. Abbildung 8.1 zeigt in diesem Zusammenhang den gesamten Jahresverlauf 2008. Um einen Vergleich mit den berechneten Wolkenbedeckungsgraden durchführen zu können, wird in 10 Tages-Intervallen der durchschnittliche Bedeckungsgrad für die Fläche des Großraumes Deutschland (vgl. Abbildung 5.2) berechnet. Dies geschieht sowohl für die Ergebnisse der SVM-Wolkenmaske, als auch für die Ergebnisse der SAFNWC-Cloudmask (siehe Abbildung 8.1).

Bei der Betrachtung dieses Vergleichs wird, wie bei den zuvor aufgeführten Ergebnisdarstellungen schnell deutlich, dass der Wolkenbedeckungsgrad (durchschnittliche Werte für den Ausschnitt Deutschland) durch den SVM-Klassifikator generell etwas höher ausfällt, als dies bei der SAFNWC-Cloudmask der Fall ist. Innerhalb des Jahresganges zeigen beide Verfahren dennoch ein ähnliches Verhalten, und spiegeln den Verlauf der Großwetterlagenklassifikation relativ genau wider. Die SAFNWC-Werte sind jedoch stärkeren Schwankungen unterlegen. Dies betrifft vor allem die ausgeprägten antizyklonalen Wetterlagen im Februar, Mai und Ende Dezember, die durch die SVM-Detektion nicht so deutlich hervortreten (siehe hierzu Kapitel 6 und 7). Aber auch die Bedeckungswerte der SAFNWC-Cloudmask sind im Vergleich mit den Wetterlagen Abweichungen unterlegen, die nicht eindeutig mit Hilfe der Kennzahlen zu erklären sind, z.B. im August und Oktober.

Dennoch lässt sich der Bedeckungsgrad mit den Kennzahlen der Wetterlagenklassifikation verknüpfen, wodurch sich die Ergebnisse der Erkennungsverfahren in den Witterungsverlauf des Jahres 2008 einordnen lassen. Es ist jedoch darauf zu achten, dass die Werte des Wolkenbedeckungsgrades Durchschnittswerte von jeweils 10 Tagen repräsentieren und somit zeitlich höher aufgelöste Phänomene nicht berücksichtigt werden.

Um die Einordnung des detektierten Wolkenbedeckungsgrades in Verbindung mit Wetterlagen auch regional noch besser abschätzen zu können, wird nachfolgend das Wettergeschehen des Jahres 2008 etwas detaillierter beschrieben³⁰. Zudem lassen sich hiermit Abweichungen zu den in Kapitel 2.5 beschriebenen durchschnittlichen Verhältnissen des Referenzzeitraumes 1961 - 1990 besser abschätzen.

Seit Beginn der regelmäßigen Wetteraufzeichnungen zählt das Jahr 2008 mit einer Durchschnittstemperatur von 9,5 °C zu den wärmsten Jahren. Vor allem die Wintermonate Januar und Februar trugen hierzu bei. Wie in Abbildung 7.2b anhand der Detektionsergebnisse zu erkennen ist, war der Februar geprägt durch niederschlagsfreies Hochdruckwetter (vgl. auch Abbildung 8.1), bei dem es vor allem im Südwesten die meisten Sonnenstunden gab. Der Monat März war dagegen durch Wettergegensätze geprägt, verbunden mit einer Unwetterserie am Monatsanfang, gefolgt von einem neuerlichen Wintereinbruch und einem abschließend frühlingshaften Ausklang. Im Vergleich zum Referenzzeitraum 1961 - 1990 war die Niederschlagsmenge die-

³⁰http://www.wetteronline.de/wotexte/redaktion/rueckblick/2008/12/31_ys.htm
[25.02.2011]

Tab. 8.1: Einteilung der objektiven Wetterlagenklassifikation des DWD. Kennung setzt sich wie folgt zusammen. Stelle 1 und 2 ergeben die Anströmungsrichtung (NO = Nordost, NW = Nordwest, SO = Südost, SW = Südwest, XX = nicht definiert). Stelle 3 bezeichnet die Zyklonalität bei 950 hPa und Stelle 4 die Zyklonalität bei 500 hPa (A = Antizyklonal, Z = Zyklonal). Stelle 5 beschreibt die Feuchte (T = Trocken, F = Feucht). *Quelle:* <http://www.dwd.de> [03.03.2011]

Kennzahl	Kennung	Kennzahl	Kennung
1	XXAAT	21	XXZAT
2	NOAAT	22	NOZAT
3	SOAAT	23	SOZAT
4	SWAAT	24	SWZAT
5	NWAAT	25	NWZAT
6	XXAAF	26	XXZAF
7	NOAAF	27	NOZAF
8	SOAAF	28	SOZAF
9	SWAAF	29	SWZAF
10	NWAAF	30	NWZAF
11	XXAZT	31	XXZZT
12	NOAZT	32	NOZZT
13	SOAZT	33	SOZZT
14	SWAZT	34	SWZZT
15	NWAZT	35	NWZZT
16	XXAZF	36	XXZZF
17	NOAZF	37	NOZZF
18	SOAZF	38	SOZZF
19	SWAZF	39	SWZZF
20	NWAZF	40	NWZZF

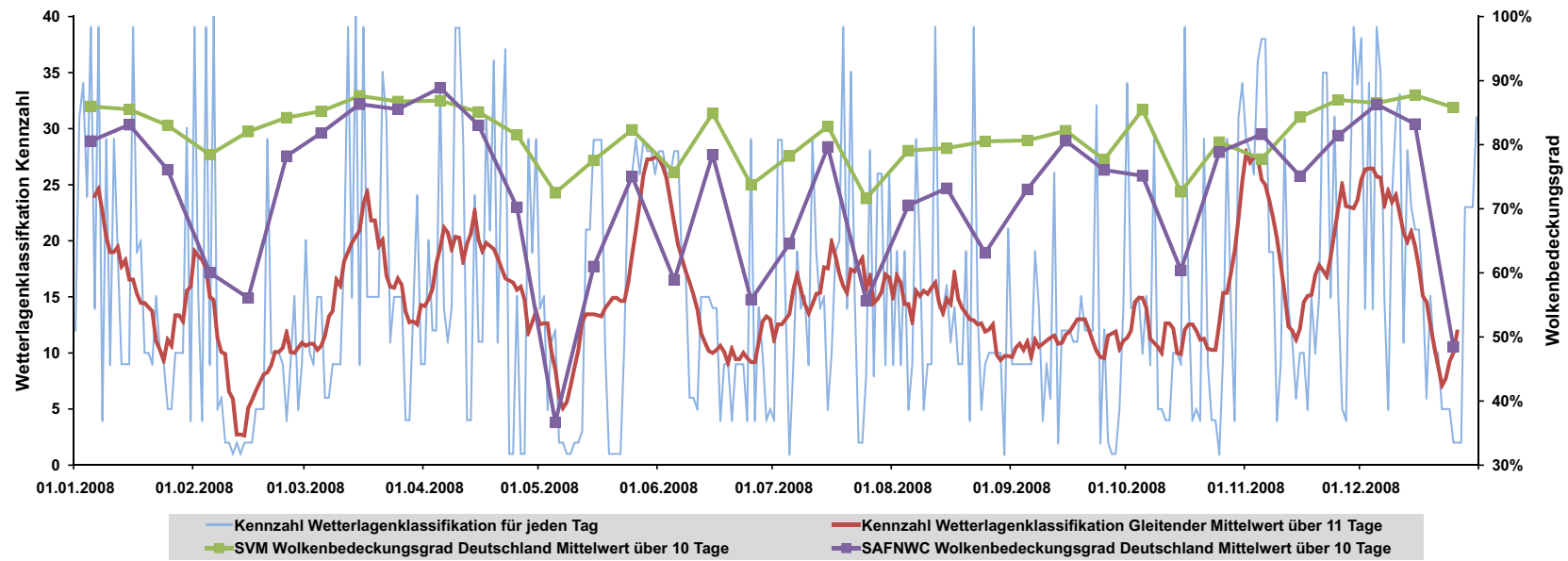


Abb. 8.1: Vergleich des durchschnittlichen Wolkenbedeckungsgrades über der ausmaskierten Fläche Deutschland (vgl. Abbildung 5.2) mit den Kennzahlen (vgl. Tabelle 8.1) der objektiven Wetterlagenklassifikation des DWD. Bedeckungsgrad berechnet für jeweils 10 Tage. *Quelle: Eigene Erhebung, Daten der Wetterlagenklassifikation abrufbar unter <http://www.dwd.de> Klima + Umwelt [03.03.2011]*

ses Monats vielerorts höher und die Sonnenscheindauer dementsprechend geringer, was sich auch in den Ergebnissen des Wolkenbedeckungsgrades dieses Monats widerspiegelt (vgl. Abbildung 7.3a). Sehr ähnlich sah es im April aus, wo es ebenfalls erst zum Ende des Monats freundlichere Bedingungen gab. Die Niederschlagsmengen waren demnach, mit Werten weit über dem Klimamittel, entsprechend hoch. Nur im Norden Deutschlands kam es zu trockeneren Bedingungen (vgl. Abbildung 7.3b). Bei der Betrachtung des Wettergeschehens dieser Monate wird deutlich, dass die jeweilige Bewölkungssituation durch die SVM-Wolkendetektion sehr gut wiedergegeben werden kann.

Mit einer durchschnittlichen Temperatur von $14,7\text{ °C}$ war der Mai um etwa $2,6\text{ °C}$ wärmer als das langjährige Mittel des Referenzzeitraumes. Gleichzeitig war er durch die ausgeprägte antizyklonale Wetterlage (vgl. Abbildung 8.1) einer der trockensten je gemessenen Maimonate. Die Sonnenscheindauer wies die höchsten Werte im Norden und Nordosten auf, während die Werte im Süden vergleichsweise geringer ausfielen. Dieser Umstand ist deutlich anhand der Detektionsergebnisse in Abbildung 7.4a zu erkennen, auch wenn davon auszugehen ist, dass die SVM-Detektion hier deutlich zur Überschätzung von Wolken neigt. Der Juni war dagegen geprägt durch zahlreiche Gewitter, wodurch es zu sehr unterschiedlichen Niederschlagsbilanzen kam. Nur im Norden blieb es vornehmlich bei trockeneren Gegebenheiten, was ebenfalls anhand der Detektionsergebnisse in Abbildung 7.4b zu erkennen ist. Trotz teilweise kräftiger Regenfälle wurden die Niederschlagswerte des langjährigen Mittels nicht erreicht. Der Juli zeigte nur am Monatsanfang und Monatsende trockene Verhältnisse. Im Rest des Monats kam es teilweise zu heftigen Niederschlagsereignissen, wodurch die Werte des langjährigen Mittels (1961 - 1990) teilweise deutlich überschritten wurden. Weniger Niederschlag gab es nur im Rheintal und im Nordosten Deutschlands, was sich ebenso bei der regionalen Verteilung des Wolkenbedeckungsgrades in Abbildung 7.5a zeigt. Auch der August war wolkenreich und wechselhaft. Nach einem sonnigen Beginn (mit Temperaturen bis zu 36 °C) kam es durch verstärkt südwestliche Strömungen (vgl. Abbildung 8.1) zu häufigerem Tiefdruckeinfluss mit starken Regenfällen und unwetterartigen Gewittern. Zu den höchsten Niederschlägen kam es hierbei im Norden und in den süddeutschen Mittelgebirgen. Trockener fiel es von der Mainlinie bis Südostbayern aus (vgl. Abbildung 7.5b). Die Witterungsbedingungen der Monate Mai bis Juni lassen sich sehr gut mit den Ergebnissen der Wolkendetektionsverfahren in Verbindung bringen, insbesondere im Zusammenhang mit hohen Niederschlagswerten. Wenn man vom Monat Mai absieht, zeichnen sich diese Muster bei der SVM-Wolkendetektion sogar noch etwas deutlicher ab, als dies bei der SAFNWC-Cloudmask der Fall ist.

Im September kam es in der zweiten Monathälfte zu einer deutlichen Abkühlung. Die Werte lagen jedoch noch im Bereich des Klimamittels. Bezogen auf den Niederschlag war der September im Vergleich zum deutschlandweiten Mittel eher trocken. Durch Tiefdruckeinfluss kam es nur im Osten und in den Mittelgebirgen zu verstärktem Niederschlag, was auch hier an-

hand der Ergebnisse des Wolkenbedeckungsgrades in Abbildung 7.6a deutlich erkennbar ist. Besonders die durch das SVM-Verfahren detektierten hohen Bedeckungsgrade der Mittelgebirge lassen sich sehr gut mit dieser Beschreibung koppeln. Der Monat Oktober war dagegen ein vergleichsweise nasser Monat, wobei auch hier die Temperaturen nur unwesentlich von den Werten des langjährigen Mittels abwichen. Die hohen Niederschläge wurden maßgeblich durch mehrere ausgeprägte Tiefdruckgebiete verursacht. An der Nordsee und im Schwarzwald wurden die höchsten Niederschläge verzeichnet. Dieser Umstand zeigt sich deutlich in Abbildung 7.6b anhand der SVM-Detektionsergebnisse. Im November kam es zu einer Zweiteilung mit milden Temperaturen in der ersten Hälfte und winterlichen Verhältnissen, verbunden mit Schneefall, in der zweiten Hälfte. Dennoch war der November vergleichsweise trocken, besonders entlang des Oberrheins und im Rhein-Main-Gebiet (vgl. Abbildung 7.7a). Höhere Niederschläge gab es nur im Bereich der Nordsee, sowie den bayrischen Alpen und den Hochlagen der südlichen Mittelgebirge. Der Dezember war dagegen geprägt von geringeren Niederschlagswerten, obwohl es zu Beginn stärkere Schneefälle bis in die Niederungen gab. Durch darauf folgendes Tauwetter blieb der Schnee jedoch nur in den Höhenlagen liegen, und es kam vielerorts zu „grünen“ Weihnachten. Erst am Ende des Monats folgten einige Tage, geprägt von Hochdruckeinfluss (vgl. Abbildung 7.7b und 8.1), mit sonnigem und kaltem Winterwetter. Bis auf diese Tage war der Dezember von Wolken und Hochnebel (Hochnebefelder vor allem im Nordosten) dominiert, wodurch es insgesamt zu einer vergleichsweise durchschnittlichen Sonnenscheindauer kam³¹. Der detektierte Wolkenbedeckungsgrad der Monate September bis November lässt sich sehr gut mit dem beschriebenen Wettergeschehen verknüpfen. Zwischen beiden Verfahren kommt es hierbei nur zu geringen Abweichungen. Im Dezember sieht dies anders aus, hier neigt die SVM-Detektion, wie bereits zuvor erwähnt zur Überschätzung, was sehr wahrscheinlich auf die Schneeflächen in den höheren Lagen der Mittelgebirge zurückzuführen ist.

Ob eine Einordnung der Ergebnisse in klimatologischer Hinsicht sinnvoll erscheint, soll abschließend kurz dargelegt werden. Da sich die Ergebnisse der Verfahren nur auf den Zeitraum des Jahres 2008 beziehen, scheint dies auf den ersten Blick nur schwer möglich. Dennoch sollte eine Einbettung in den klimatologischen Rahmen, beispielsweise durch den Vergleich mit NAO-Index-Werten, durchaus möglich sein. Der NAO-Index wird, wie in Kapitel 2.3.3 beschrieben, zur Abschätzung zirkulationsdynamischer Veränderungen auf klimatologischer Ebene eingesetzt. Im Jahr 2008 ergeben sich zusammengefasst sowohl für die Monate Januar/Februar/März, als auch für die Monate Oktober/November/Dezember positive Indexwerte³², verbunden mit milder und feuchter Luft, die nach Europa transportiert wird. Das heißt, die allgemein hohen Wolkenbedeckungsgrade, die anhand der Detektions- bzw. Klassifizierungser-

³¹http://www.wetteronline.de/wotexte/redaktion/rueckblick/2008/12/31_ys.htm [25.02.2011]

³²<http://www.cgd.ucar.edu/cas/jhurrell/indices.html> [10.03.2011]

gebnisse erfasst wurden (vgl. Abbildung 7.2 bis 7.13 und 8.1) stehen in Einklang mit den erhobenen NAO-Indexwerten. Diese Verknüpfung und die vergleichsweise milden Temperaturen des Jahres 2008³³ lassen zwar keine Aussage über eine allgemeine Klimaänderung zu, können aber innerhalb der Tendenz einer Klimaerwärmung eingebettet werden. Ob und wie zirkulationsdynamische Veränderungen im Zusammenhang mit einem Klimawandel bei langjähriger Betrachtung Einfluss auf den Wolkenbedeckungsgrad haben, kann natürlich ebenfalls mit vorliegenden Ergebnissen nicht beantwortet werden. Gleiches gilt für langfristige Veränderungen und Rückkopplungseffekte innerhalb des Strahlungshaushaltes. Daran wird aber deutlich, dass ein weiterer und detaillierter Informationsgewinn über Wolken für die Beantwortung solcher Fragestellungen von enormer Relevanz ist. Mit der Entwicklung neuer sowie der Optimierung bestehender Wolkendetektions- bzw. Wolkenklassifizierungsverfahren kann somit die Erzeugung langfristiger zeitlich und räumlich hoch aufgelöster Datenreihen vorangetrieben werden.

³³http://www.wetteronline.de/wotexte/redaktion/rueckblick/2008/12/31_ys.htm
[25.02.2011]

9 Diskussion und Ausblick

Wolken sind ein zentraler Bestandteil innerhalb der Atmosphäre, nicht nur bezogen auf die Niederschlagsverteilung, auch ihr Einfluss auf zirkulationsdynamische Vorgänge und den Strahlungshaushalt sind von entscheidender Bedeutung. Für die in diesem Zusammenhang teilweise noch offenen Fragestellungen ist eine umfangreiche Datengrundlage über den Zustand von Wolken und die damit verbundene Wolkenbedeckung von besonderem Interesse. Dies betrifft vor allem Änderungen im Rahmen des vielfach diskutierten Klimawandels (vgl. Simmer 2000, IPCC 2001, IPCC 2007). Für die Erkennung und Analyse der Bewölkungsgegebenheiten haben sich vor allem Verfahren der Satellitenfernerkundung als Mittel der ersten Wahl herausgestellt. Im Vergleich zu polarumlaufenden Satelliten bieten geostationäre Satelliten den enormen Vorteil einer hohen zeitlichen Auflösung (METEOSAT mit Aufnahmeintervallen alle 15 Minuten), womit die Variabilität von Wolken, sowohl auf räumlicher als auch zeitlicher Ebene, relativ genau abgeschätzt werden kann (siehe z.B. Derrien und Le Gléau 2005a).

Neben anderen Digitalen Bildverarbeitungsverfahren basieren die Wolkendetektions- bzw. Wolkenklassifizierungsverfahren in erster Linie auf Schwellenwert- bzw. Grenzwertansätzen (Tool APOLLO, SAFNWC-Wolkenprodukte), wobei die Bestimmung des korrekten Schwellenwertes als größte Herausforderung anzusehen ist (vgl. z.B. Kriebel et al. 2003). Gegenstand vorliegender Arbeit war daher die Entwicklung eines alternativen Verfahrens, mit dem Wolken innerhalb einer Satellitenszene vergleichsweise genau erkannt und klassifiziert werden können. Dieses Verfahren sollte vorwiegend auf der reinen Bildinformation des Satelliten basieren und nicht, wie es bei den meisten Schwellenwertverfahren der Fall ist, durch zusätzliche Informationen über den Zustand der Atmosphäre ergänzt werden. An für Klassifizierungsverfahren der digitalen Bildverarbeitung notwendigen Trainingsdaten (Ground Truth) kommt man jedoch nicht vorbei. Support Vector Machines stellen in diesem Zusammenhang einen modernen Klassifikator dar, der mit wenigen solcher Trainingsdaten eine effektive Trennung von Klassen durchführen kann. Diese wurden zwar schon auf experimenteller Ebene zu Zwecken der Wolkenerkennung eingesetzt, jedoch meist nur bezogen auf einzelne Szenen und auf Basis von Expertenwissen (vgl. Azimi-Sadjadi und Zekavat 2000, Han et al. 2006, Zibert et al. 2006, Gomez-Chova et al. 2010). Das vorrangige Ziel dieser Arbeit war demgegenüber die Entwicklung eines *automatisierten Verfahrens*, angewendet auf die zeitlich hoch aufgelösten Daten des METEOSAT-SEVIRI-Sensors.

Die Ausführung erfolgte zwar rückwirkend für jede Szene des Deutschland-Datensatzes 2008, ist aber so konzipiert, dass eine operationelle Durchführung vom zeitlichen Aspekt her durchaus in Erwägung gezogen werden kann.

Entscheidend für die erfolgreiche Detektion bzw. Klassifizierung ist die exakte Gewinnung repräsentativer Trainingsdaten für jede Einzelszene, bei der sogenannte Wetter-Kameras als Grundlage des Trainings verwendet werden sollten. Eine Vielzahl solcher Kameras erfasst den Himmel über Deutschland, mit einer teilweise extrem hohen zeitlichen Auflösung (Aufnahmeintervallen bis zu einer Minute). In dieser Hinsicht sind sie synoptischen Beobachtungsdaten überlegen, und können optimal an die zeitliche Auflösung des METEOSAT-SEVIRI-Sensors angepasst werden. Die Zusammenstellung eines repräsentativen Datensatzes für das Jahr 2008 stellte sich dennoch als problematisch dar, da die meisten Betreiber keine Archivierung durchführen und die Daten nach wenigen Tagen löschen. Durch diesen Umstand waren im Endeffekt nur sechs Kamera-Standorte verfügbar, mit denen eine effektive Wolkenklassifizierung, basierend auf Support Vector Machines, nur schwer umzusetzen ist. Die vielfältigen Strahlungsverhältnisse im Zusammenhang mit Wolken können so nur schlecht eingegrenzt werden.

Dennoch wurden die Daten der sechs Kamera-Standorte ergänzend zu Trainingszwecken verwendet, wobei die Extraktion der Trainingsinformationen auf der jeweiligen RGB-Zusammensetzung der Einzel-Aufnahmen basiert. Vor allem die Unterscheidung zwischen Wolken und wolkenfreien Himmelsflächen konnte so effektiv gelöst werden.

Um den Umfang der bis hierhin gewonnenen Trainingsdaten wesentlich zu erhöhen, wurden die aus anderen Arbeiten bereits bekannten Schwellenwertalgorithmen (APOLLO, SAFNWC) herangezogen. Die Abgrenzung bezieht sich hierbei jedoch nur auf solche Pixel, bei denen mit hoher Sicherheit von Bewölkung ausgegangen werden kann. Mit Hilfe eines ungefähren Grenzwertes bzw. Grenzbereiches und der Kombination mehrerer Schwellenwertalgorithmen lassen sich verschiedene Wolkenarten als Trainingsdaten erfassen, und die eigentliche Trennung der Klassen geschieht abschließend durch den SVM-Klassifikator.

Obwohl in großen Teilen der Arbeit von Trainingsdaten (eigentlich Ground Truth) die Rede ist, handelt es sich weitestgehend um einen unüberwachten Ansatz. Das gesamte Prozessierungsverfahren wurde so gestaltet, dass ein vollständig automatisierter Ablauf durchführbar ist. Wolkenrelevante Kanäle und Kanal-Differenzen spannen den Merkmalsraum auf, der neben den Trainingsdaten notwendig ist, um die Klassifizierung des jeweiligen Zeitschnitts durchzuführen.

Grundlage für die weiter differenzierte Wolkenklassifizierung stellt die im ersten Schritt erstellte SVM-Wolkenmaske dar. Wolkenmaske und Wolkenklassifizierung werden nach gleichem Prinzip, mit jeweils eigener Trainingsdatengewinnung erzeugt. Die vier zusätzlichen Wolkenklassen beziehen sich auf unterschiedliche Wolkenhöhen und Wolkentransparenz, und durch die Trennung von Maske und weiterer Klassifizierung wird eine einfachere und effektivere

Handhabung der zusätzlichen Trainingsdatengewinnung und der späteren Validierung ermöglicht.

Eine Validierung mit unabhängigen Daten erfolgte nur für die Wolkenmasken (SVM-Wolkenmaske, SAFNWC-Cloudmask), auf Grundlage der durch DWDSAT zur Verfügung stehenden synoptischen Beobachtungsdaten, für den Zeitraum Juli bis Dezember 2008.

Trotz der an den WMO-Stationen zur Verfügung stehenden Informationen über Bewölkungsverhältnisse auf unterschiedlichen Höhenniveaus, war eine Validierung der SVM-Wolkenklassifizierung nicht in einem sinnvollen Rahmen möglich. Im Gegensatz zur SVM-Wolkenklassifizierung, erfolgt die Betrachtung durch den Beobachter von unten, wodurch Überlagerungen verschiedener Wolkenarten nicht richtig zugeordnet werden können. Demnach wurde nur ein visueller Vergleich mit den Ergebnissen der SAFNWC-Produkte durchgeführt.

Anhand der Validierungsergebnisse konnte gezeigt werden, dass eine Wolkendetektion auf Basis von Support Vector Machines durchaus konkurrenzfähig ist. Verglichen mit den Ergebnissen der SAFNWC-Cloudmask ergaben sich in Teilen sogar höhere Genauigkeiten. Zwar zeigte sich innerhalb des Übergangsbereiches zwischen wolkenkontaminierten und wolkenfreien Flächen die Tendenz einer Wolkenüberschätzung, dem gegenüber steht aber eine Unterschätzung durch die SAFNWC-Cloudmask (ohne NWP). Welcher Faktor zur Überschätzung führt konnte nicht vollständig geklärt werden. Zu vermuten ist, dass der grob abgeschätzte Grenzwert in bestimmten Situationen zu weit in den Bereich der Klasse „Wolkenfrei“ hineinreicht, wodurch Trainingspixel falsch zugeordnet werden. Dies ist sicherlich vorrangig der Fall, wenn eine Szene mit ausgeprägten wolkenfreien Flächen (Hochdruckwetterlage) gegeben ist. Der SVM-Klassifikator ist in solchen Situationen immer bestrebt eine Trennung bzw. Entscheidung zu treffen. Wenn dann Trainingspixel als Wolke generiert werden, die eine ähnliche spektrale Signatur aufweisen wie wolkenfreie Flächen, z.B. durch geringfügige Bewölkung auf Subpixel-Ebene, kann dies zu falschen Klassifizierungsergebnissen führen. Unterschiedliche subjektive Einschätzungen der Bewölkung durch den Beobachter, mit Einfluss auf die Validierungsergebnisse, können auch nicht vollständig ausgeschlossen werden.

Die Durchführung der Wolkendetektion ohne die ergänzende Verwendung der Wetter-Kameras zeigte im Rahmen der Validierungsergebnisse vor allem einen verstärkenden Effekt auf das Phänomen der Überschätzung. Obwohl die Overall Accuracy bei Verwendung der Kameradaten nur geringfügig höhere Werte zeigt, kann man erwarten, dass eine Erweiterung durch zusätzliche Wetter-Kameras, mit Vorteilen für die Gesamtgenauigkeit des Verfahrens verbunden ist.

Anhand der Unterscheidung zwischen Tag- und Nacht-Testreihen innerhalb der Trainingsdatengewinnung war zudem festzustellen, dass zusätzliche Trainingsinformationen aus dem Kanal des sichtbaren Wellenlängenbereichs, zu einer höheren Overall Accuracy führen. Der Effekt der

höheren Genauigkeit am Tag lässt sich aber sicherlich auch auf den Einfluss der Wetter-Kameras zurückführen, da diese, bis auf den Standort Mannheim, nachts keine Daten liefern.

Bei der Trennung des Validierungsdatensatzes in unterschiedliche Höhenlagen der WMO-Stationen (> 300 m oder < 300 m) ergaben sich nur geringfügige Abweichungen der Genauigkeit. Die Tendenz der Überschätzung durch die SVM-Wolkenmaske und der Unterschätzung durch die SAFNWC-Cloudmask wird in höheren Lagen leicht verstärkt.

Um die Ergebnisse des auf Support Vector Machines basierenden Wolkendetektions- und Wolkenklassifizierungsverfahrens noch deutlicher darzustellen, erfolgte die Berechnung des Wolkenbedeckungsgrades aus der Summe der vorhandenen Einzelzeitschnitte. Durch Monatswerte konnte so die Effektivität bzw. das Potential des automatisierten Verfahrens, räumlich und zeitlich, veranschaulicht werden. In diesem Sinne zeigen die Muster der Bewölkung über Deutschland in der Regel große Ähnlichkeiten mit denen der SAFNWC-Cloudmask. Insbesondere bei ausgeprägter Bewölkung bzw. gleicher Verteilung von Wolken und wolkenfreien Flächen sind nur geringe Abweichungen zwischen beiden Verfahren zu erkennen. Hochdruckwetterlagen führen jedoch zu allgemein höheren Werten der SVM-Wolkendetektion. Durch den Vergleich zweier Fernerkundungsverfahren ist es jedoch schwierig, eindeutige Aussagen über die zeitliche und räumliche Ausprägung von jeweiliger Wolkenüber- bzw. Wolkenunterschätzung zu treffen.

Bezogen auf den Bedeckungsgrad der vier weiteren Wolkenklassen kam es zu ähnlichen Ergebnissen. Die räumlichen Muster weisen bei beiden Verfahren große Ähnlichkeit auf, und der Großteil der Bewölkung über Deutschland lässt sich der Klasse „Niedrig Kompakt“ zuordnen, also dichte Bewölkung im niedrigen bis mittleren Höhenniveau. Regional betrachtet, sind vor allem die Höhenlagen der Mittelgebirge mit höchsten Werten verknüpft, was sich auch bei der Einordnung der Ergebnisse in das allgemeine Wettergeschehen des Jahres 2008 darstellt. Je nach Strömungsrichtung bzw. Großwetterlage sind vor allem Mittelgebirge wie Eifel oder Schwarzwald mit hohen Wolkenbedeckungsgraden verknüpft. Die Betrachtung regionaler Niederschlagswerte zeigt hier eine ähnliche Tendenz. Das Rheintal bzw. das Rhein-Main-Gebiet, sowie Teile Nord- und Ostdeutschlands sind dagegen oft mit geringerer Bedeckung und geringerem Niederschlag gekoppelt. Dies zeigt sich auch bei der regionalen Betrachtung der Sonnenscheindauer innerhalb des Jahres 2008.

9.1 Herausforderungen für die Zukunft

Obwohl mit dieser Arbeit gezeigt werden konnte, dass eine Wolkendetektion bzw. -klassifizierung mit Hilfe von Support Vector Machines in einer mit den gängigen Methoden vergleichbaren Genauigkeit umgesetzt werden kann, bleiben gewisse Herausforderungen bestehen, die es zukünftig zu bewältigen gilt. Neben der Minimierung von Ungenauigkeiten, die sich im Rah-

men der Prozessierung ergeben, sollte das vorrangige Bestreben darin liegen, den Effekt der Überschätzung so weit wie möglich zu reduzieren. Hierzu bedarf es jedoch weiterer zeitintensiver Testläufe. Vor allem um mögliche Fehler bei der automatisierten Trainingsdatengewinnung einzugrenzen. Die Abschätzung des ungefähren Grenzwertes kann in diesem Zusammenhang sicher noch optimiert werden. Dies gilt besonders für Situationen in denen eine Klasse nur in geringem Umfang präsent ist.

Damit Trainingsdaten noch besser den Grenzbereich jeder einzelnen Szene zum Lernen des Klassifikators darstellen können, sollte für zukünftige Planungen der Umfang der Wetter-Kameras wesentlich erhöht werden. Anhand der Validierung konnte gezeigt werden, dass sich die Daten der Kameras positiv auf die Genauigkeit der Wolkendetektion auswirken. Ob anhand der Kameras auch Informationen über unterschiedliche Wolkenarten bzw. Wolkenhöhen abgeleitet werden können, wäre ebenfalls ein interessanter Aspekt zukünftiger Bearbeitungen.

Auch die weitere Anpassung der im Klassifizierungsverfahren verwendeten SVM-Parameter (Kernel- und Regularisierungsparameter) könnte einen zusätzlichen Ansatz der Optimierung darstellen. Je nach Verteilung von Wolken und wolkenfreien Flächen bzw. Wolkenklassen kann die Hyperebene eventuell besser eingebunden werden. Da hierzu jedoch das SVM-Training für jeden Zeitschnitt, mit hohem Zeitaufwand, erneut durchgeführt werden muss, wäre dies im Rahmen einer operationellen Anwendung aber nur bedingt praktikabel.

Eine Anwendung auf andere Regionen oder Teilgebiete der Erde ist sicherlich auch von Interesse. Web-Cams bzw. Wetter-Kameras gibt es mittlerweile an vielen Standorten und der Datenzugriff ist, nach Absprache mit den Betreibern, meist ohne Probleme möglich. Aufgrund der oft fehlenden Archivierung, sollte dies aber zeitnah durchgeführt werden. Mit vorliegendem Verfahren konnte gezeigt werden, dass auch in nördlichen Regionen (Ausschnitt Deutschland, bei METEOSAT hoher Satelliten-Zenitwinkel) mit unterschiedlichsten Wolkengegebenheiten durchaus gute Ergebnisse erzielt werden können. Es ist zu vermuten, dass die Anwendung z.B. in tropischen Regionen genauso gut oder gar noch besser funktioniert. Hoch aufsteigende, sehr kalte konvektive Zellen lassen sich besser von warmen Landflächen abgrenzen.

Will man hingegen einen großräumigen Bereich, wie beispielsweise Europa abdecken, entsteht die Problematik unterschiedlicher Strahlungstemperaturbereiche innerhalb der verschiedenen Teilregionen mit Einfluss auf die Trainingsdaten bzw. Trainingsdatengewinnung. Eine Trennung von Klassen wird hier schwierig. Strahlungstemperaturen kalter Landflächen im Norden können im Süden Wolken repräsentieren. Um diesem Phänomen gerecht zu werden, müssten mehrere Teilregionen einzeln bearbeitet und nachfolgend zusammengeführt werden. Eine zusätzliche Schwierigkeit, die sich in diesem Zusammenhang ergibt, ist die mit zunehmenden Trainingspixeln verbundene Erhöhung der Rechenzeit. Damit auch hier die Möglichkeit eines operationellen Betriebes erhalten bleibt, ist eine parallele Berechnung erforderlich. Durch die

Zunahme und einfachere Verfügbarkeit immer leistungsfähigerer Rechner sollte dies aber unproblematischer werden.

Bei einer Übertragung auf Sensoren anderer Satelliten sollte es bei Vorhandensein der notwendigen Kanäle keine Probleme geben. Die Anwendung auf polarumlaufende Wetter-Satelliten bringt zwar eine erhöhte räumliche Auflösung mit sich, der Vorteil der hohen zeitlichen Auflösung würde aber verloren gehen.

Um die Übertragung auf andere Gebiete oder andere Sensoren umsetzen zu können, ist eine Anpassung aller Parameter der Trainingsdatengewinnung (niedrigste/höchste Werte als „Schwellenwerte“, Anzahl Trainingspixel usw.) erforderlich. Dies ist jedoch nur einmalig durchzuführen.

Trotz beschriebener Probleme und Herausforderungen, die es noch zu lösen gilt, konnte mit vorliegender Arbeit das Potential von Support Vector Machines in Bezug auf eine automatisierte Wolkendetektion bzw. -klassifizierung an zeitlich hoch aufgelösten Daten dargelegt werden. Ob sich dieses Verfahren in einem operationellen Betrieb durchsetzen kann, bleibt abzuwarten. Gegebenenfalls können Teile des Ansatzes innerhalb der gängigen Verfahren implementiert werden oder als Erkenntnisgewinn zur Entwicklung neuer Verfahren beitragen.

Literatur

- Abe, S. (2010). *Support Vector Machines for Pattern Classification* (2. Aufl.). Springer.
- Albertz, J. (2001). *Einführung in die Fernerkundung - Grundlagen der Interpretation von Luft und Satellitenbildern* (2. Aufl.). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Amato, U., A. Antomadis, V. Cuomo, L. Cutillo, M. Franzese, L. Murino, und C. Serio (2008). Statistical cloud detection from SEVIRI multispectral images. *Remote Sensing of Environment* 112(3), 750–766.
- Ameur, Z., S. Ameur, A. Adane, H. Sauvageot, und K. Bara (2004). Cloud classification using the textural features of Meteosat images. *International Journal of Remote Sensing* 25(21), 4491–4503.
- Azimi-Sadjadi, M. R., M. A. Shaikh, B. Tian, K. E. Eis, und D. Reinke (1996). Neural network-based cloud detection/classification using textural and spectral features. In: *Geoscience and Remote Sensing Symposium, 1996 - IGARSS '96 - Remote Sensing for a Sustainable Future*, Volume 2, S. 1105–1107.
- Azimi-Sadjadi, M. R. und S. A. Zekavat (2000). Cloud classification using support vector machines. In: *Proceedings-IEEE-2000 (Hrsg.), Int. Geoscience and Remote Sensing Symp IGARSS 2000*, Volume 2, S. 669–671.
- Beck, C. (2000). *Zirkulationsdynamische Variabilität im Bereich Nordatlantik-Europa seit 1780*. Würzburg Selbstverlag des Instituts für Geographie der Universität Würzburg.
- Bendix, J. (2002). A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas. *Atmospheric Research* 64(1-4), 3–18.
- Berger, F. H. (1995). The variability of cloud cover and cloud forcing inferred from NOAA AVHRR data for the North Sea. *Advances in Space Research* 16(10), 29–32.

- Bissolli, P. (2004). SAT-KLIM: Überwachung der Wolkenbedeckung auf Basis von Satelliten- und Stationsdaten. In: D. Wetterdienst (Hrsg.), *DWD Klimastatusbericht 2004*, Offenbach am Main, S. 147–149. Online: http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU2/KU22/klimastatusbericht/einzelne__berichte/ksb2004__pdf/14__2004,templateId=raw,property=publicationFile.pdf/14_2004.pdf [12.11.2010].
- Bissolli, P. und E. Dittmann (2001). The objective weather types classification of the German Weather Service and its possibilities of application to environmental and meteorological investigations. *Meteorologische Zeitschrift* 10(4), 253–260.
- Bissolli, P. und U. Gelhardt (2002). *Klimatologische Validierung von ISCCP-Daten der Wolkengesamtbedeckung aus Satellitenquellen mit synoptischen Stationsdaten über Deutschland*. Offenbach am Main: Deutscher Wetterdienst.
- Blüthgen, J. und W. Weischet (1980). *Allgemeine Klimageographie*. Berlin: de Gruyter.
- Bridgman, H. A. und J. E. Oliver (2006). *The Global Climate System: Patterns, Processes, and Teleconnections*. Cambridge: Cambridge University Press.
- Canty, M. J. (2006). *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL*. Boca Raton: CRC Press.
- Carleton, A. (1991). *Satellite Remote Sensing in Climatology*. London: Belhaven Press.
- Chaboureau, J. P. und C. Claud (2006). Satellite-based climatology of Mediterranean cloud systems and their association with large-scale circulation. *Journal of Geophysical Research-Atmospheres* 111(D1). Online: http://mesonh.aero.obs-mip.fr/chaboureau/PUB/chaboureau_claud_jgr2006D01102.pdf [11.11.2010].
- Chang, C.-C. und C.-J. Lin (2001). *LIBSVM: a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm> [12.10.2010].
- Cristianini, N. und J. Shawe-Taylor (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (1. Aufl.). Cambridge: Cambridge University Press.
- Derrien, M. und H. Le Gléau (2005a). MSG/SEVIRI cloud mask and type from SAFNWC. *International Journal of Remote Sensing* 26(21), 4707–4732.
- Derrien, M. und H. Le Gléau (2005b). *SAFNWC/MSG SEVIRI Cloud Products*. Lannion.
- Derrien, M., H. Le Gléau, J. F. Daloze, und M. Haeffelin (2005). *Validation of SAFNWC/MSG Cloud Products with One Year of Seviri Data*. Lannion.

- Diamandi, A. und A. Dybbroe (2001). *Nowcasting SAF - Validation of AVHRR cloud products - Visiting scientist report*. Norrköping.
- Dittmann, E., S. Barth, G. Müller-Westermeier, und J. Lang (1995). Objektive Wetterlagenklassifikation. In: D. Wetterdienst (Hrsg.), *Berichte des Deutschen Wetterdienstes Nr. 197*, Offenbach am Main.
- Dybbroe, A. (2001). The retrieval of the SAFNWC Cloud Mask and Cloud Type from AVHRR and SEVIRI at high latitudes. In: SMHI (Hrsg.), *SMHI SAFNWC Action PT06-04*, Norrköping.
- Dybbroe, A., A. Henja, S. Hörnquist, K. G. Karlsson, C. Sverker, A. Thoss, V. Thy-rén, und A. Yong (2004). *The Nowcasting SAF Polar Platform System Software Package (PPS) - Current Status*. Norrköping. Online: http://www.eumetsat.int/Home/OldWhoWeAre/Conference_and_Workshop_Proceedings/groups/cps/documents/document/pdf_conf_p41_s2_dybbroe_p.pdf [11.11.2010].
- Dybbroe, A., K. G. Karlsson, M. Moberg, und A. Thoss (2000). Scientific Report for the SAFNWC Mid Term Review. In: SMHI (Hrsg.), *SMHI Scientific Report for the SAFNWC MTR, Issue 1.1*, Norrköping. Online: http://produkter.smhi.se/saf/smhi_sr_1.1.pdf [11.11.2010].
- Dybbroe, A., K. G. Karlsson, und A. Thoss (2002). *The AVHRR Cloud Mask Scheme of the SAFNWC*. Norrköping. Online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=?doi=10.1.1.138.3433> [11.11.2010].
- Dybbroe, A., A. Thoss, und S. Hörnquist (2005). Scientific user manual for the AVHRR/AM-SU cloud and precipitation products of the SAFNWC/PPS. In: E. SAFNWC (Hrsg.), *SAFNWC/PPS Scientific User Manual Issue 1.1*. Online: http://produkter.smhi.se/saf/SAF-NWC-IOP-SMHIPPS-SW-SUM-2_v1.1.0.pdf.
- EUMETSAT (2000). *The METEOSAT System, Revision 4*. Darmstadt.
- EUMETSAT (2001). *METEOSAT Second Generation MSG System Overview EUM TD 07 Issue 1.1*. Darmstadt.
- EUMETSAT (2004a). *Image News from Europes weather satellite organisation Issue 20*. Darmstadt.
- EUMETSAT (2004b). *MSG-1/SEVIRI Solar Channels Calibration Commissioning Activity Report, EUM/MSG/TEN/04/0024*. Darmstadt. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=pdf_ten_040024_seviri-solcalib&RevisionSelectionMethod=LatestReleased [22.11.2010].

- EUMETSAT (2005a). *Image News from Europes weather satellite organisation Issue 23*. Darmstadt.
- EUMETSAT (2005b). *METEOSAT Second Generation Satelliten - Ein bewährter Service noch besser*. Darmstadt.
- EUMETSAT (2005c). *MSG Level 1.5 Image Product - Quality Indicators, EUM/OPS-MSG/TEN/05/0421 v1*. Darmstadt. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=PDF_TEN_050421_MSG_LEV15_QUAL&RevisionSelectionMethod=LatestReleased [16.11.2010].
- EUMETSAT (2005d). *MSG METEOSAT Second Generation In Orbit - In Use*. Darmstadt.
- EUMETSAT (2005e). On the conversion from radiances to equivalent brightness temperature, EUM/OPS/TEN/05/2556 v2. Darmstadt.
- EUMETSAT (2006). *MSG Level 1.5 Image Data Format Description, EUM/MSG/ICD/105 V4*. Darmstadt.
- EUMETSAT (2007a). *A Planned Change to the MSG Level 1.5 Image Product Radiance Definition, EUM/OPS-MSG/TEN/06/0519 v1A*. Darmstadt. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=PDF_MSG_PLANNED_CHANGE_LEVEL15&RevisionSelectionMethod=LatestReleased [16.11.2010].
- EUMETSAT (2007b). *Typical Geometrical Accuracy for MSG-1/2, EUM/OPS/TEN/07/0313 v1*. Darmstadt. Online: http://www.eumetsat.int/groups/ops/documents/document/pdf_typ_geomet_acc_msg-1-2.pdf [16.11.2010].
- EUMETSAT (2008a). *Effective Radiance and Brightness Temperature Relation for Meteosat 8 and 9, EUM/OPS-MSG/TEN/08/0024 v1*. Darmstadt. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=PDF_TEN_080024_RAD_BRIGHT_TEMP&RevisionSelectionMethod=LatestReleased [16.11.2010].
- EUMETSAT (2008b). *A Simple Conversion from Effective Radiance back to Spectral Radiance for MSG Images, EUM/OPS-MSG/TEN/07/1053 v1*. Darmstadt. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=PDF_MSG_CONV_EFFRAD2SPECRAD&RevisionSelectionMethod=LatestReleased [16.11.2010].

- Feister, U., J. Shields, M. Karr, R. Johnson, K. Dehne, und M. Woldt (2000). Ground-Based Cloud Images and Sky Radiances in the Visible and Near Infrared Region from Whole Sky Imager Measurements. In: DWD/EUMETSAT/WMO (Hrsg.), *Proceedings of Climate Monitoring - Satellite Application Facility Training Workshop*, Dresden. Online: http://www-mpl.ucsd.edu/people/jshields/publications/dd_feist.pdf [12.11.2010].
- Foody, G. M. und A. Mathur (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103(2), 179–189.
- Garay, M. J. (2005). *The Application of Support Vector Machines to Analysis of Global Satellite Data Sets from MISR*. Pasadena. Online: <http://ml.jpl.nasa.gov/papers/mazzoni/garay-ams05.pdf> [12.11.2010].
- Geomatics-Department (2007). *ImageSVM, Support Vector Machine Classification for Remote Sensing Image Data, (preliminary) Manual*. Berlin: Humboldt-Universität zu Berlin.
- Ghosh, A., N. R. Pal, und J. Das (2006). A fuzzy rule based approach to cloud cover estimation. *Remote Sensing of Environment* 100(4), 531–549.
- Gomez-Chova, L., G. Camps-Valls, L. Bruzzone, und J. Calpe-Maravilla (2010). Mean Map Kernel Methods for Semisupervised Cloud Classification. *Ieee Transactions On Geoscience and Remote Sensing* 48(1), 207–220.
- Gomez-Chova, L., G. Camps-Valls, J. Munoz-Mari, und J. Calpe (2008). Semisupervised image classification with Laplacian support vector machines. *Ieee Geoscience and Remote Sensing Letters* 5(3), 336–340.
- Govaerts, Y. M., M. Clerici, und N. Clerbaux (2004). Operational calibration of the Meteosat radiometer VIS band. *IEEE Transactions on Geoscience and Remote Sensing* 42(9), 1900–1914. Online: www.meteo.be/meteo/download/de/519880/pdf/rmi_scpub-058.pdf [22.11.2010].
- Han, B., L. S. Kang, und H. Z. Song (2006). A fast cloud detection approach by integration of image segmentation and support vector machine. *Advances In Neural Networks - Isnn 2006, Pt 3, Proceedings* 3973, 1210–1215.
- Häckel, H. (2005). *Meteorologie* (5. Aufl.). Stuttgart: Ulmer.
- Heyer, E., F. M. Chmielewski, P. Hupfer, W. Kuttler, und H. Pethe (2005). *Witterung und Klima: Eine Einführung in die Meteorologie und Klimatologie* (11. Aufl.). Stuttgart: Teubner.

- Hsu, C.-W., C.-C. Chang, und C.-J. Lin (2003). *A Practical Guide to Support Vector Classification*. Taipei: National Taiwan University. Online: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> [16.11.2010].
- IPCC (2001). *Climate Change 2001: Synthesis Report*. Online: http://www.grida.no/publications/other/ipcc_tar [21.04.2010].
- IPCC (2007). *Climate Change 2007: Synthesis Report*. Online: http://www.ipcc.ch/publications_and_data/ar4/syr/en/contents.html [21.04.2010].
- Janz, A., S. van der Linden, B. Waske, und P. Hostert (2007). imageSVM - A User-Oriented Tool for Advanced Classification of Hyperspectral Data using Support Vector Machines. In: EARSeL (Hrsg.), *Proceedings of the 5th EARSeL workshop on Imaging Spectroscopy, Bruges, Belgium, 23-25 April, 2007, CD-Rom*.
- Jensen, J. R. (2004). *Introductory Digital Image Processing* (3. Aufl.). Prentice Hall.
- Kidder, S. Q. und T. H. Vonder Haar (1995). *Satellite Meteorology: An Introduction*. San Diego: Academic Press.
- König, M. und P. Bissolli (2001). Vergleichende Klimatologie Satelliten- und Bodengestützter Wolkenbeobachtungen über Deutschland. In: D. Wetterdienst (Hrsg.), *Deutscher Wetterdienst Forschung und Entwicklung Arbeitsergebnisse (Gelbe Reihe)*, Offenbach am Main.
- Kraus, H. (2009). *Die Atmosphäre der Erde: Eine Einführung in die Meteorologie* (3. Aufl.). Berlin: Springer.
- Kriebel, K. T. (1990). Wolkenanalyse mit Satellitendaten. *PROMET* 20(3/4), 86–91.
- Kriebel, K. T., G. Gesell, M. Kästner, und H. Mannstein (2003). The cloud analysis tool APOLLO: improvements and validations. *International Journal of Remote Sensing* 24(12), 2389–2408.
- Kärner, O. und S. Keevallik (1993). *Effective Cloud Cover Variations*. Hampton: A. Deepak Publishing.
- Kurz, M. (1990). *Synoptische Meteorologie* (2. Aufl.). Offenbach am Main: Deutscher Wetterdienst.
- Lee, Y., G. Wahba, und S. A. Ackerman (2004). Cloud classification of satellite radiance data by multicategory support vector machines. *Journal of Atmospheric and Oceanic Technology* 21(8), 1301–1301.

- Mannstein, H., R. Meerkötter, G. Gesell, V. Grewe, C. König, und S. Lohmann (2004). *A High Resolution European Cloud Climatology from 15 Years of NOAA/AVHRR Data*. Weßling: DLR. Online: http://www.eumetsat.int/Home/OldWhoWeAre/Conference_and_Workshop_Proceedings/groups/cps/documents/document/pdf_conf_p41_s5_mannstein_p.pdf [11.11.2010].
- Mazzoni, D., A. Horváth, M. Garay, B. Tang, und R. Davies (2005). A MISR cloud-type classifier using reduced Support Vector Machines. In: J. P. Laboratory (Hrsg.), *Proceedings of the Eighth Workshop on Mining Scientific and Engineering Datasets, 2005 SIAM International Conference on Data Mining*. Online: <http://ml.jpl.nasa.gov/papers/mazzoni/mazzoni-msd05-misr.pdf> [10.11.2010].
- Meerkötter, R., P. Bissolli, und G. Gesell (2004). Eine 15 jährige Wolkenklimatologie für Europa aus NOAA/AVHRR Daten im Vergleich zu Bodenbeobachtungen. In: D. Wetterdienst (Hrsg.), *Klimastatusbericht 2004*, S. 24–30. Online: http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU2/KU22/klimastatusbericht/einzelne__berichte/ksb2004__pdf/03__2004,templateId=raw,property=publicationFile.pdf/03_2004.pdf [11.11.2010].
- Müller-Westermeier, G. (2006). *Wetter und Klima in Deutschland* (4. Aufl.). Stuttgart: Hirzel.
- Pontius, R. G. (2000). Quantification Error Versus Location Error in Comparison of Categorical Maps. *Photogrammetric Engineering & Remote Sensing* 66(8), 1011 – 1016.
- Roebeling, R. A., A. J. Feijt, und P. Stammes (2006). Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *Journal of Geophysical Research-Atmospheres* 111(D20).
- SAFNWC (2007a). *Algorithm Theoretical Basis Document for "Cloud Products"(CMA-PGE01, CT-PGE02 & CTTH-PGE03 v1.4)*, SAF/NWC/CDOP/MFL/SCI/ATBD/01, Issue 1, Rev. 4.
- SAFNWC (2007b). *Architectural Design Document for the SAFNWC/MSG*, SAF/NWC/IOP/INM/SW/AD/1, Issue 2.
- SAFNWC (2007c). *Product User Manual for "Cloud Products"(CMA-PGE01, CT-PGE02 & CTTH-PGE03 v1.4)*, SAF/NWC/CDOP/MFL/SCI/PUM/01, Issue 1, Rev. 4.
- SAFNWC (2007d). *Scientific Report for version 2.0 of the PGE05 of SAFNWC/MSG (CRR - Convective Rainfall Rate)*, SAF/NWC/IOP/INM/SCI/RP/17, Issue 1.
- SAFNWC (2007e). *Scientific report on Additional Tuning of MSG PGE01-02-03 v1.2 (Cloud Products) following validation activities*, SAF/NWC/IOP/MFL/SCI/RP/02, Issue 0.

- SAFNWC (2007f). *Software User Manual for "Convective Rainfall Rate": Scientific Part (CRR-PGE05 v2.1)*, SAF/NWC/CDOP/INM/SCI/SUM/05, Issue 2, Rev. 1.
- SAFNWC (2007g). *Software User Manual for the SAFNWC/MSG Application: Software Part*, SAF/NWC/IOP/INM/SW/SUM/2, Issue 2.
- SAFNWC (2007h). *Software User Manual for the version 2.0 of the PGE05 of SAFNWC/MSG (CRR - Convective Rainfall Rate): Scientific part*, SAF/NWC/IOP/INM/SCI/SUM/05, Issue 2.
- SAFNWC (2007i). *Software Users Manual for the SAFNWC/MSG Application*, SAF/NWC/IOP/INM/MGT/SUM, Issue 2.
- SAFNWC (2007j). *Software Verification and Validation Plan for the SAFNWC/MSG*, SAF/NWC/CDOP/INM/SW/PL/2, Issue 1, Rev. 1.
- SAFNWC (2007k). *User Manual for the PGE01-02-03 v1.3 (Cloud Products) of the SAFNWC/MSG: Scientific part*, SAF/NWC/IOP/MFL/SCI/SUM/01, Issue 1, Rev. 3.
- SAFNWC (2007l). *Validation Report for "Cloud Products"(CMa-PGE01, CT-PGE02 & CTHH-PGE03 v1.4)*, SAF/NWC/CDOP/MFL/SCI/VR/02, Issue 1, Rev. 4.
- SAFNWC (2007m). *Validation Report for the PGE01-02-03 v1.2 (Cloud Products) of the SAFNWC/MSG*, SAF/NWC/IOP/MFL/SCI/VAL/01, Issue 1, Rev. 2.
- SAFNWC (2008a). *Software User Manual for the SAFNWC/MSG Application: Software Part*, SAF/NWC/CDOP/INM/SW/SUM/2, Issue 2, Rev. 1.
- SAFNWC (2008b). *Software Users Manual for the SAFNWC/MSG Application*, SAF/NWC/CDOP/INM/MGT/SUM, Issue 2, Rev. 1.
- Saunders, R. W. (1986). An automated scheme for the removal of cloud contamination from AVHRR radiances over western Europe. *International Journal of Remote Sensing* 7(7), 867–886.
- Saunders, R. W. und K. T. Kriebel (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing* 9(1), 123–150.
- Schölkopf, B. und A. J. Smola (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (1. Aufl.). The MIT Press.
- Schmetz, J., P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, und A. Ratier (2002). SEVIRI calibration. *Bulletin of the American Meteorological Society* 83.
- Schönwiese, C. D. (2008). *Klimatologie* (3. Aufl.). Stuttgart: UTB.

- Seiz, G., J. Shields, U. Feister, E. P. Baltsavias, und A. Gruen (2007). Cloud mapping with ground-based photogrammetric cameras. *International Journal of Remote Sensing* 28(9), 2001–2032.
- Simmer, C. (2000). Beeinflussen Wolken das Klima? In: N.-W. A. der Wissenschaften (Hrsg.), *Vorträge N 452*, Wiesbaden, S. 7–38.
- Sönning, W. und C. G. Keidel (2007). *Wolkenbilder, Wettervorhersage* (7. Aufl.). BLV.
- Thiermann, V. und E. Ruprecht (1992). A method for the detection of clouds using AVHRR infrared observations. *International Journal of Remote Sensing* 13(10), 1829–1841.
- Tjemkes, S. A. (2005). *On the Conversion from Radiances to Equivalent Brightness Temperatures*. Darmstadt: EUMETSAT. Online: http://www.eumetsat.int/groups/ops/documents/document/pdf_msg_seviri_rad2bright.pdf [12.11.2010].
- Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.
- VCS (2005). *VCS 2met! Software Documentation*. Bochum: VCS Aktiengesellschaft.
- Wanner, H., S. Bronnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, und E. Xoplaki (2001). North Atlantic Oscillation - Concepts and studies. *Surveys In Geophysics* 22(4), 321–382.
- Watts, P. D., C. T. Mutlow, A. J. Baran, und A. M. Zavody (1998). Study on Cloud Properties derived from Meteosat Second Generation Observations. In: R. A. Laboratory (Hrsg.), *Generation Observations: EUMETSAT ITT no. 97/181 Final Report*, Rutherford.
- WMO (1975). *International Cloud Atlas, Vol. 1*. World Meteorological Organization.
- Wolf, R. (1999). LRIT/HRIT Global Specification. In: EUMETSAT (Hrsg.), *Coordination Group for Meteorological Satellites, Doc. No.: CGMS 03, Issue: 2.6*, Darmstadt. Online: http://www.wmo.int/pages/prog/sat/documents/CGMS-03_HRIT-LRIT_v2-6.pdf [15.11.2010].
- Yamashita, M., M. Yoshimura, und T. Nakashizuka (2004). Cloud Cover Estimation Using Multitemporal Hemisphere Images. In: ISPRS (Hrsg.), *ISPRS Congress Istanbul 2004, Proceedings of Commission VII*, S. 826 ff. Online: <http://www.isprs.org/proceedings/XXXV/congress/comm7/papers/162.pdf> [12.11.2010].
- Yao, M.-S. und A. D. Del Genio (1999). Effects of Cloud Parameterization on the Simulation of Climate Changes in the GISS GCM. *Journal of Climate* 12(3), 761–779. Online: http://pubs.giss.nasa.gov/docs/1999/1999_Yao_DelGenio.pdf [23.11.2010].

Zibert, M. I., M. Derrien, und H. Le Gléau (2006). Automatic Cloud Classification by Supervised Learning on SEVIRI Data using Support Vector Machines Method. In: EUMETSAT (Hrsg.), *The 2005 EUMETSAT Meteorological Satellite Conference*, Darmstadt, S. 188–192. Online: http://www.eumetsat.int/idcplg?IdcService=GET_FILE&dDocName=PDF_CONF_P46_S2_17_ZIBERT_P&RevisionSelectionMethod=LatestReleased [12.11.2010].