

Geostatistische Analyse der Chlorophyllverteilung in der Nordsee basierend auf MERIS-Satellitendaten

Dissertation

Zur Erlangung des Doktorgrades
an der MIN-Fakultät, Department Geowissenschaften
der Universität Hamburg

vorgelegt von
Dagmar Müller
aus Hamm/Westf.

Hamburg 2010

Als Dissertation angenommen vom Department Geowissenschaften der
Universität Hamburg auf Grund der Gutachten von

Prof. Dr. Hartmut Grassl
und Dr. Roland Doerffer

Hamburg, den 15.12.2010

Prof. Dr. Jürgen Oßenbrügge
Leiter des Department Geowissenschaften

Inhaltsverzeichnis

Einleitung	1
I Das MERIS-Chlorophyllprodukt - Beschreibung der Daten und des Untersuchungsraumes	5
1 Einführung zur Satellitenozeanographie	8
1.1 Das bildgebende Spektrometer MERIS an Bord von ENVISAT	8
1.2 Begrifflichkeiten	8
1.3 MERIS Level 2-Produkte	9
1.3.1 Standardprozessor	10
1.3.2 Regionalprozessor meris_c2r	11
2 Ozeanographische Beschreibung der Nordsee	15
2.1 Geographische Definition der Nordsee	15
2.1.1 Untergliederung nach Wassermassen	16
2.2 Definition der Untersuchungsräume	18
2.2.1 Nordsee	18
2.2.2 ICES-Regionen	18
2.2.3 Gebiete 1-5	18
2.2.4 Weitere Areale	19
2.3 Physikalisch-ozeanographische Einflüsse	20
2.3.1 Tiden und Strömungen	20
2.3.2 Schichtung und Fronten	20
2.3.3 Vertikale Durchmischung durch Wind	22
3 Phytoplankton in der Nordsee	24
3.1 Jahresgang und Sukzession	24
3.1.1 Phytoplanktondynamik in Ökosystemmodellen	25
3.1.2 Beginn der Frühjahrsblüte	25
3.1.3 Langzeitentwicklungen - Continuous Plankton Recorder	26
3.2 Vertikalverteilung	28
3.3 Artspezifische optische Eigenschaften	28
3.4 Das "Phytoplanktonjahr"	30

II	Datenbearbeitung und Datenbasis	32
4	Datenbearbeitung	34
4.1	Einfache Filter	34
4.1.1	Sigma-Filter	34
4.1.2	Median-Filter	34
4.2	Filter für unerkannte Wolkenränder	35
4.2.1	Identifizierung über optische Dicke der Aerosole	35
4.2.2	Identifizierung über die Gesamtstrahldichte	35
4.2.3	Beispiel zur Identifizierung von Wolkenrändern	36
4.3	Standardgitter	37
4.4	Zeitangaben	38
5	Kritische Bewertung der Datenbasis	39
5.1	Normalverteilttheit	39
5.2	Separabilität	39
5.3	Überdeckungsstatistik	40
5.3.1	Datenüberdeckung	40
5.3.2	Wolkenüberdeckung	41
5.4	Frequenzanalyse unregelmäßiger Zeitreihen	43
5.4.1	Zeitreihen punktuell zusammengefasster Chlorophyllkonzentrationen	45
5.4.2	Frequenzanalyse unregelmäßiger Zeitreihen	48
5.4.3	Systematischer Fehler im Jahresgang	51
5.5	Systematische Fehler im Level 2-Produkt des Regionalprozessors	52
5.5.1	Instrumentenfehler und systematischer Fehler	53
5.5.2	Der Sonnenazimutwinkel	55
5.5.3	Theoretische Überlegung zum Sonnenazimutwinkel	57
5.5.4	Eine einfache Frequenzanalyse	60
5.5.5	Mittlerer Fehler	61
5.6	Anmerkung und Zusammenfassung	63
6	Phytoplankton der Nordsee in Beispielen	65
6.1	Klimatologie der Chlorophyllkonzentration nach Jahreszeiten	65
6.1.1	Überdeckungsstatistik	69
6.2	Jahresgang anhand von Monatsmitteln	69
6.3	Wiederkehrende Strukturen	71
6.3.1	Doggerbank	71
6.3.2	Schichtungsfront	72
6.3.3	Norwegische Rinne	74
III	Räumliche Methoden zur Aufbereitung von Datenlücken	77
7	Mittelungsverfahren	80

Inhaltsverzeichnis

7.1	Arithmetisches Mittel, Gewichtetes Mittel, Median	80
7.2	Mittelwert und Median größter Wahrscheinlichkeit - mit zeitlicher Gewichtung	81
7.3	Übersicht der Mittelungsverfahren	82
8	Kriging: ein exakter, linearer Schätzer	83
8.1	Ordinary Kriging in der Theorie	83
8.2	Variogramme	86
8.2.1	Experimentelles Variogramm	86
8.2.2	Theoretisches Variogramm (Variogramm-Modell)	86
8.2.3	Anisotropie	88
8.3	Cokriging	88
8.4	Übersicht der Krigingverfahren	89
9	Kriging in der Praxis I - Deutsche Bucht	90
9.1	Das experimentelle Variogramm	92
9.1.1	Globales Variogramm	92
9.1.2	Einfluss der Probenahme	93
9.1.3	Konvergenz des Medians von Variogrammrealisationen	95
9.1.4	Mittleres Variogramm aus unterschiedlichen Stichproben	98
9.1.5	Übertragbarkeit auf größere Datenmengen	98
9.2	Einfluss fehlender Daten auf das experimentelle Variogramm	100
9.2.1	Einfluss der Probenahme bei konstant fehlenden Daten	100
9.2.2	Einfluss der Positionierung der Wolkenmaske	101
9.3	Vom experimentellen Variogramm zum Variogramm-Modell	101
9.3.1	Akaike Informationskriterium	102
9.3.2	Automatische Wahl des Variogramm-Modells	103
9.4	Lokales Kriging	103
9.4.1	Konvergenz der relativen Fehler der Schätzwerte	104
9.4.2	Konvergenz der Krigingvarianz	104
9.4.3	Abhängigkeit der Schätzwerte vom Variogramm-Modell	105
9.5	Abhängigkeit der tatsächlichen Fehler von der Krigingmethode	106
9.5.1	Lokale und globale Genauigkeit der Schätzwerte	107
9.5.2	Beste Trendkorrektur im Ordinary Kriging der Residuen	110
9.5.3	Beste Krigingmethode bei konstanter Trendkorrektur	114
9.5.4	Beste Trendkorrektur in Abhängigkeit von der Variation der Wolkenposition	115
9.5.5	Beste Krigingmethode in Abhängigkeit von der Variation der Wolkenposition	117
9.5.6	Beste Trendkorrektur bei fester Wolkenposition und verschiedenen Datentagen	118
9.5.7	Beste Krigingmethode bei fester Wolkenposition und verschiedenen Datentagen	120
9.5.8	Beispiel für Schätzwerte aus verschiedenen Krigingmethoden	120

Inhaltsverzeichnis

9.6	Zusammenfassung und Ausblick	122
10	Kriging in der Praxis II - Nordsee	127
10.1	Künstliche Wolkenmasken	127
10.2	Beschreibung räumlicher Strukturen	128
10.2.1	Distanzklassenverteilung	129
10.2.2	Lückenhaftigkeit	131
10.3	Berechnung der Lückenhaftigkeit in der Praxis	132
10.4	Krigingexperiment mit künstlichen Wolkenmasken	135
10.4.1	Aufbereitung der Daten	135
10.4.2	Beste Trendkorrektur	137
10.4.3	Beste Krigingmethode	139
10.5	Diskussion: Beste Rekonstruktion mittels Kriging	143
11	Wann ist Kriging einsetzbar?	144
12	Iteratives Dateninterpolationsverfahren - DINEOF	148
12.1	Einführung zur Methode DINEOF	148
12.2	Beispiel DINEOF in der Deutschen Bucht	150
12.3	Beispiel DINEOF im Gebiet der Nordsee	151
IV	Räumliche und zeitliche Autokorrelation	159
13	Theoretische Betrachtungen zur Autokorrelation	161
13.1	Räumliche Autokorrelation	161
13.2	Zeitliche Autokorrelation	162
14	Räumliche oder zeitliche Autokorrelation von Chlorophylldaten	164
14.1	Räumliche Autokorrelation	165
14.1.1	Einfluss von Schätzwerten auf die räumliche Autokorrelation	165
14.1.2	Bestimmung von Korrelationslängen	168
14.1.3	Diskussion	170
14.2	Zeitliche Autokorrelation der Chlorophyllmonatsmittel	171
14.2.1	Nordsee	172
14.2.2	ICES-Regionen in der Nordsee	174
14.2.3	Kleinere Gebiete in der Nordsee	176
14.3	Zeitliche Autokorrelation der Chlorophylldatentage	176
V	Zusammenfassung und Ausblick	183
15	Zusammenfassung und Ausblick	184
15.1	Zusammenfassung	184
15.1.1	Fehler innerhalb des Chlorophyllproduktes	185

Inhaltsverzeichnis

15.1.2 Geostatistische Schätzwertverfahren	187
15.1.3 Ergebnisse der Analyseverfahren	189
15.2 Ausblick	190
Anhang	193
A Weiterführende Untersuchungen zum systematischen Fehler	195
A.1 Korrektur der Beobachtungsgeometrie - 1	195
A.2 Korrektur der Beobachtungsgeometrie - 2	199
A.3 Abweichung der Weglänge in planarem oder sphärischem Atmosphärenmodell	200
Literaturverzeichnis	201

Einleitung

Nach Jahrzehnten, in denen die Ozeanographie ihre Daten ausschließlich von schiffgestützten Untersuchungen gewann, eröffnete das Zeitalter der Raumfahrt gänzlich neue Erfassungsmethoden. 1978 gelang der NASA mit dem Coastal Zone Color Scanner CZCS der Startschuss zur Satelliten gestützte Meeresforschung. Diese und nachfolgende Missionen, wie z.B. SeaWiFS (Sea-viewing Wide Field-of-view Sensor) und MODIS (Moderate Resolution Imaging Spectroradiometer), schufen und schaffen ausgedehnte Datenbasen über den Zustand der Weltmeere (vgl. Bissett u. a. (2001) und IOCCG (2007)).

Die riesigen Areale der Meere und Ozeane konnten mit Hilfe der Satellitentechnik zum ersten Mal räumlich flächendeckend und, z.B. in Gebieten von der Größe der Nordsee, annähernd gleichzeitig betrachtet werden. Die Strukturen aus Wirbeln und Wirbelstraßen (entstanden aus Schwebstoffen oder Gelbstoffen in Flussausströmungen) sowie die Phytoplanktonblüten im Golfstrom wurden zum ersten Mal für das menschliche Auge im großen Zusammenhang sichtbar.

Mit MERIS an Bord des Umweltsatelliten ENVISAT brachte im März 2002 auch die Europäische Weltraumbehörde, ESA, ein bildgebendes Spektrometer zur Erfassung der Wasserfarbe in die Umlaufbahn.

Die verbreitete englische Bezeichnung *ocean colour* für diese Art der Satellitenozeanographie ist programmatisch zu verstehen: aus der Farbe des Meeres, d.h. dem sichtbaren Spektrum des vom Wasser zurückgestreuten Sonnenlichtes, werden die inhärenten optischen Eigenschaften - also die Streu- und Absorptionseigenschaften - des Wassers berechnet und daraus die wesentlichen Inhaltsstoffe (bzw. Stoffgruppen) abgeleitet. Seit Beginn dieser Satellitenmissionen gehört die Chlorophyll- bzw. Phytoplanktonkonzentration zu diesen Inhaltsstoffen. Als Basis des Nahrungsnetzes der Meere kommt dem Primärproduzent Phytoplankton größte ökologische Bedeutung zu. Es kann durch anomale Blütenbildung aber auch als Indikator für Klimaprozesse oder anthropogene Einflüsse in den Meeren dienen.

Auf dem Weg der Auswertung, beginnend bei den Spektren zurück gestreuten Lichtes im sichtbaren Wellenlängenbereich bis zur Bestimmung und Bewertung eines Produktes "Chlorophyllkonzentration" als Stellvertreter für das vorhandene Phytoplankton, sind allerdings einige Hindernisse zu überwinden:

Der Hauptanteil des Fernerkundungssignals besteht aus dem zurück gestreuten Licht der Atmosphäre. Nur etwa 10% der am Satelliten registrierten Intensität stammt tatsächlich aus dem Wasser und transportiert die gewünschte spektrale Information über dessen optisch wirksame Inhaltsstoffe, sodass eine genaue Atmosphärenkorrektur erfolgen muss. In küstennahen Gebieten ergibt sich darüber hinaus eine weitere Schwierigkeit. Hier ist die Konzentration des Phytoplanktons nicht allein ausschlaggebend für die Wasserfarbe, sondern es kommen weitere Substanzen hinzu, die in der Regel zu drei optisch wirksamen Komponenten zusammengefasst werden: Phytoplankton, Schwebstoff und Gelbstoff. Gerade die Küstenregionen sind aber wegen ihrer enormen ökologischen Bedeutung als Kinderstube vieler Meeresorganismen Bestandteil zahlreicher Beobachtungsprogramme. Gleichzeitig wird hier der anthropogene Einfluss direkt spürbar. Satellitendaten werden hier z.B. zur Frühwarnung vor toxischen Algenblüten herangezogen.

Wenn aus einer Kombination von Wasserinhaltsstoffkonzentrationen das erwartete modellierte mit dem tatsächlich registrierten Reflektanzspektrum in Einklang gebracht wer-

den kann und die Genauigkeit der abgeleiteten Chlorophyllkonzentration hinreichend erscheint, dann können folgende Fragestellungen untersucht werden:

- Wie entwickelt sich die Phytoplanktonverteilung im Jahresgang?
- Welche Einflussgrößen bestimmen die räumliche und zeitliche Verteilung des Phytoplanktons?
- Lassen sich Unterschiede in der Dynamik zwischen einzelnen Jahren feststellen und wie sind sie zu erklären?

Der Satellitendatensatz verspricht Antworten auf Fragestellungen, die sonst nur an Ökosystemmodelle gestellt werden. Konkret steht zu Beginn dieser Arbeit die Frage, ob die Satellitendaten hinreichende Genauigkeit besitzen, um neben interannualer Variation einen Trend festmachen zu können, der auf Einflüsse einer globalen klimatischen Veränderung zurückgeht. Dazu sollen verschiedene Einflussgrößen anhand weitere Datensätze hinzugezogen werden (vgl. Diagramm auf Seite 7).

Nach und nach habe ich die Fragestellung modifiziert. Neben der zu erwartenden Unvollständigkeit der Zeitreihen konnte ich einen systematischer Fehler im Chlorophyllprodukt entdecken, dessen genaue Charakterisierung vor der Auswertung notwendig war.

Die Zeitreihenanalyse wird vorrangig durch die Wolkenbedeckung und die damit einhergehende Lückenhaftigkeit des Datensatzes bzw. seine qualitative Minderung erschwert.

Wolken hindern das Spektrometer daran, die Wasseroberfläche im sichtbaren Spektralbereich beobachten zu können. Die optischen Charakteristika der Inhaltsstoffe sind jedoch ausschließlich im sichtbaren Spektralbereich zugänglich, da nur hier das Wasser hinreichend transparent ist.

Im Gebiet der Nordsee sind häufig große Wolkenbänder zu finden, die weite Bereiche der Wasserfläche bedecken. Um eine möglichst vollständige Zeitreihe aufstellen zu können, ist daher der Einsatz von geostatistischen Schätzwertverfahren zu bedenken. Unter welchen Annahmen ist die Erhaltung der räumlichen Strukturen bei Ergänzung mit Schätzwerten zu erwarten?

Letztlich habe ich die folgende Fragestellung bearbeitet, die sich auch im Aufbau dieser Arbeit widerspiegelt:

- Wie lässt sich der systematische Fehler charakterisieren und wo ist seine Quelle zu suchen?
- Welche Beobachtungen lassen sich trotz der bekannten Unsicherheit in den Daten bezüglich der räumlichen Dynamik machen?
- Mit welchen Fehlern ist bei der Vervollständigung der Datensätze durch Schätzwerte zu rechnen und welchen Einfluss haben verschiedene Schätzmethoden?
- Wie lassen sich die Strukturen fehlender Daten quantifizieren und wie hängen die Fehler der Schätzwerte von ihnen ab?

- Welche Aussagen lassen sich trotz des systematischen Fehlers bezüglich der Dynamik räumlicher und zeitlicher Strukturen treffen bzw. wie lassen sie sich quantifizieren?

Diese und andere Fragen sollen im Folgenden entwickelt und beantwortet werden.

Der erste Teil (Teil I) leitet ein, wie in der Satellitenozeanographie und speziell bei dem Spektrometer MERIS auf ENVISAT die Daten erhoben werden (Kapitel 1). In einem kurzen Abriss werden ozeanographische Aspekte des Untersuchungsraumes, der Nordsee, vorgestellt (Kapitel 2). Dazu gehört auch ein kurzer Einblick in die biologischen Prozesse des Phytoplanktons dieser Region. Aus *in situ* Beobachtungen und Ökosystemmodellierungen ist vieles über seine Zusammensetzung, jahreszeitliche Sukzession und über seine optische Eigenschaften (Kapitel 3) bekannt.

Eine Charakterisierung der Satellitendaten erfolgt im zweiten Teil (Teil II). Welche Hilfsmittel generell zur Aufbereitung der Daten verwendet werden (Kapitel 4), wird ebenso dokumentiert, wie einige spezielle Eigenschaften und im Rahmen dieser Arbeit aufgedeckte systematische Fehler innerhalb des Datensatzes (Kapitel 5). Kapitel 6 soll besondere Stärken und Schwächen des Datensatzes qualitativ illustrieren, bevor quantitative Bearbeitungsstrategien in nachfolgenden Teilen im Vordergrund stehen.

Der dritte Teil (Teil III) wendet sich verschiedenen Handhabungen von fehlenden Datenpunkten zu. Die einfachste Methode, um räumlich vollständige Karten zu erhalten, bleibt die zeitliche Mittelung von Datenpunkten verschiedener Beobachtungstage (Kapitel 7).

Um die hohe zeitliche Auflösung zu erhalten, empfiehlt sich allerdings die Berechnung von geostatistischen Schätzwerten. Basierend auf den mathematischen Grundlagen der Krigingmethode (Kapitel 8) werden verschiedene Techniken an einigen Beispielen erprobt und anhand der Güte der Rekonstruktion bewertet. Viele der praktischen Überlegungen sind dabei unabhängig von der genauen Art der Daten, sie treffen für Satellitenprodukte mit ihren vielen Datenpunkten allgemein zu. Während Kapitel 9 sich hauptsächlich mit dem Einfluss des Variogramms auf die Genauigkeit der Schätzwerte befasst, versucht Kapitel 10 eine Quantifizierung der Fehler der Schätzwerte in Abhängigkeit von der Struktur der fehlenden Daten.

Von weiteren möglichen Schätzwertverfahren wird das Dateninterpolationsverfahren *DINEOF* exemplarisch vorgestellt und mit den Krigingmethoden verglichen (Kapitel 12).

Die quantitative Untersuchung des Datensatzes mittels räumlicher und zeitlicher Autokorrelation bildet den methodischen Abschluss dieser Arbeit (Teil IV).

Teil I

Das MERIS-Chlorophyllprodukt - Beschreibung der Daten und des Untersuchungsraumes

Einleitung

Die Entwicklung der Satellitenozeanographie hat zum ersten Mal großflächige, annähernd gleichzeitige Untersuchungen der oberflächennahen Wasserschichten ermöglicht. In ihnen werden die Konzentrationen von denjenigen Wasserinhaltsstoffen ermittelt, die über die Farbe des Wassers abgeleitet werden können. Seit jeher wird dabei ein Algen- oder Chlorophyllprodukt bestimmt, um der herausragenden Stellung des Phytoplanktons im Meer als primärer Nahrungsquelle und Basis aller höheren Meeresorganismen gerecht zu werden.

Einen Einblick, wie aus einer Spektrometermessung von MERIS das Produkt "Chlorophyllkonzentration" gewonnen wird, gibt das Kapitel 1.

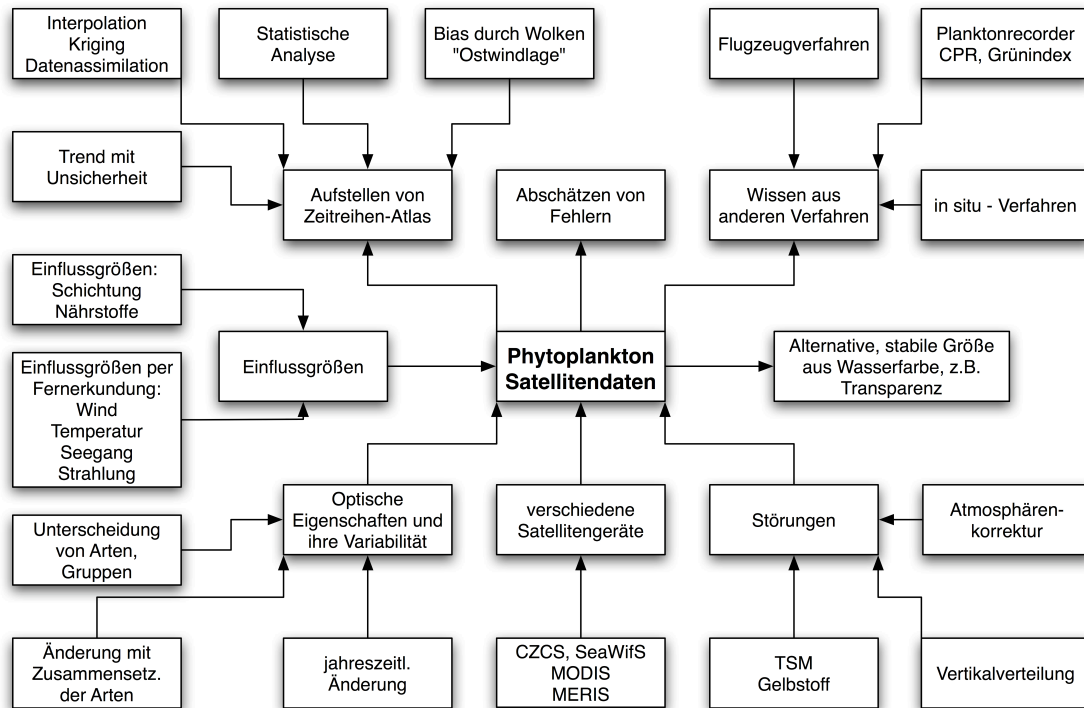
Alle nachfolgenden Betrachtungen legen ihr Augenmerk auf Daten aus dem Gebiet der Nordsee. Abhängig von Fragestellung werden die Beispieldaten aus dem Gesamtgebiet oder spezifischen Teilgebieten (Abschnitt 2.2) stammen.

Eine kurze physikalisch-ozeanographische Charakterisierung dieses Schelfmeeres (Kapitel 2) soll mit den wichtigsten Phänomenen bezüglich Bathymetrie, Wassermassen und deren jahreszeitlicher Dynamik, wie z.B. der Schichtung, vertraut machen.

Kapitel 3 behandelt die biologischen Aspekte des Phytoplanktons in der Nordsee, d.h. seine jahreszeitliche Wachstumsdynamik und die sie kontrollierenden Prozesse, wie Einstrahlung und Nährstoffe, sowie die charakteristische Artenzusammensetzung und ihre Sukzession, wie sie in der Literatur beschrieben werden. Typische optische Eigenschaften werden anhand artspezifischer Absorptionsspektren kurz erläutert.

Zur Meteorologie des Untersuchungsraumes sei lediglich auf die entsprechenden Übersichten in der Literatur (z.B. Holligan u. a. (1989); Huthnance (1991); Lee (1980); Siegismund und Schrum (2001)) verwiesen. Eine eingehende Beschäftigung mit Größen wie Wind und Einstrahlung wird es im Rahmen dieser Arbeit nicht geben, auch wenn sie zur kausalen Interpretation der auftretenden Phänomene sicherlich von größter Bedeutung sind.

Insbesondere der Wind besitzt einen hohen Einfluss auf die Strömung der oberen Wasserschicht, die vom Satelliten hauptsächlich erfasst werden kann. Insbesondere die Verlängerung der Perioden starken Windes im Jahresgang (im Rahmen des globalen Klimawandels, vgl. Siegismund und Schrum (2001)) lassen den vertikal durchmischten Zustand andauern und verhindern bzw. verzögern die thermische Stabilisierung. Diese Veränderung des Wind- bzw. Schichtungsjahresganges hat Auswirkungen auf den Beginn und die Ausprägung der Frühjahrsblüten des Phytoplankton (vgl. Ruardij u. a. (1997)). In zentral-nordöstlichen Bereichen der Nordsee kann der Einfluss des Windes auf die herrschende Strömungssituation wesentlich ausschlaggebender sein als die Tiden.



Die Übersicht von Einflussgrößen (vgl. Diagramm auf dieser Seite) verdeutlicht die Komplexität der Satellitenprodukte, die aus spektralen Beobachtungen abgeleitet werden können.

Insbesondere heißt dies, dass das abgeleitete Algenprodukt nicht absolut mit dem tatsächlichen Chlorophyllgehalt des Wassers gleichzusetzen ist. Dieser wäre im besten Fall eine Funktion sowohl im Raum, d.h. in der Fläche und der Tiefe, als auch in der Zeit. Doch Aussagen über die Vertikalverteilung lassen sich mittels des Satellitenproduktes nicht treffen, da dieses lediglich eine über die Sichttiefe gemittelte Größe liefert. Abhängig von weiteren Inhaltsstoffen variiert die Sichttiefe beträchtlich zwischen etwa einem halben Meter in den küstennahen Gewässern bis zu einigen Metern in der zentralen Nordsee. Zu diesen Ambivalenzen innerhalb des zu messenden Objektes kommen des Weiteren methodische und systematische Beschränkungen des Messprozesses selbst hinzu. So sind die Messgeometrie und der Zeitpunkt der Messung durch die Lage der Satellitenumlaufbahn festgelegt und beschränken die maximal mögliche zeitliche und räumliche Auflösung der Messung.

1 Einführung zur Satellitenozeanographie

1.1 Das bildgebende Spektrometer MERIS an Bord von ENVISAT

Gegenüber allen durchführbaren *in situ* Methoden haben Beobachtungen vom Satelliten den Vorteil, dass die Nordsee innerhalb von vier Minuten annähernd zeitgleich beprobt werden kann.

Seit März 2002 befindet sich der Umweltsatellit ENVISAT (Environmental Satellite) zur Überwachung von Klima, Ozeanen und Landflächen der Erde in polarer, sonnen-synchroner Umlaufbahn. Dem bildgebenden Spektrometer MERIS (*Medium Resolution Imaging Spectrometer*) fallen dabei Aufgaben der Meeres- und Küstenforschung, des Vegetationsmonitoring und der Atmosphärenforschung zu (ESA (2000)).¹

MERIS verfügt über fünfzehn Spektralbänder, die innerhalb des reflektierten Sonnenspektrums (sichtbarer Spektralbereich, nahes Infrarot) zwischen 390 nm und 1040 nm mit einer spektralen Breite zwischen 2.5 nm und 30 nm ausgewählt werden können. Das Instrument tastet die Erdoberfläche mit der so genannten "push-broom"-Methode ab. Lineare CCD-Anordnungen sorgen für die räumliche Abtastung in Richtung quer zur Flugbahn, während die Bewegung des Satelliten das flächige Bild ergibt. MERIS ist so konstruiert worden, dass das Spektrometer jederzeit Daten der Erdoberfläche aufzeichnen kann, sollte die Beleuchtung ausreichend sein. Mit einem Blickfeld von 68.5° um den Nadir überdeckt ein Schwad eine Breite von 1150 km. Das Blickfeld wird unterteilt von fünf gleichartigen Kameras aufgenommen. Die räumliche Auflösung beträgt 1040 m x 1200 m (reduzierte Auflösung), bzw. 260 m x 300 m (volle Auflösung). Nach 35 Tagen wiederholt MERIS den Durchlauf der 501 möglichen (relativen) Orbits. Die Nordsee wird aus etwa 24 Orbits gesehen.

1.2 Begrifflichkeiten

Eine Reihe von Begriffen, die in der Satellitenozeanographie geprägt worden sind und in dieser Arbeit verwendet werden, sollen kurz erläutert werden.

Der Begriff des (*relativen*) *Orbits* bezeichnet jeweils die Flugbahn und ihre Abbildung auf der Erdoberfläche von und bis zum dem Nordpol nächstgelegenen Punkt. Die Erdoberfläche ist dabei in 501 Orbits aufgeteilt, die innerhalb von 35 Tagen durchlaufen werden und sich danach exakt wiederholen.

¹vgl. Envisat-Homepage der ESA <http://earth.esa.int/>, dort unter instruments-> MERIS

Ein *Datentag* beinhaltet alle Datenpunkte in einer definierten Region aus dem Satellitenüberflug eines Tages. Da die Nordsee in der geographischen Länge nicht so ausgedehnt ist, ist der Datentag hier (meist) gleichbedeutend mit dem einen Orbit, der das Untersuchungsgebiet zur Mittagszeit erfasst hat.

Aus dem Datensatz eines Orbits wird eine so genannte *Szene* ausgeschnitten, die gerade den Untersuchungsraum abdeckt. Das Archiv von Szenen enthält insbesondere jene Ausschnitte aus Orbits, die den Punkt 54° N, 8° O in der Deutschen Bucht gesehen haben.

Satellitenprodukte können in verschiedenen Bearbeitungsstufen vorliegen: ein *Level 1-Produkt* ist ein georeferenziertes Radianzspektrum, während unter einem *Level 2-Produkt* jene abgeleiteten Größen (nach der Atmosphärenkorrektur) wie zum Beispiel Chlorophyll- oder Schwebstoffkonzentration zu verstehen sind. Als *Level 3* bezeichnet man weiterverarbeitete abgeleitete Größen z.B. ein Chlorophyllmonatsmittel.

Die Kombination von Atmosphärenkorrektur und biooptischem Modell zur pixelweisen Ableitung der Level 2-Produkte aus den Level 1-Radianzen wird zusammenfassend als *Prozessor* bezeichnet. Jede Weiterentwicklung der Prozessoren liefert neue Level 2-Produkte, sodass die Datenbasis einer beständigen Neuberechnung und Veränderung unterliegt.

1.3 MERIS Level 2-Produkte

Eine Berechnung der aus den Reflektanzen an der Wasseroberfläche (Level 2, nach Atmosphärenkorrektur) abgeleiteten Produkte wie die Algenprodukte oder die Konzentration von Schweb- und Gelbstoff, ist nicht Gegenstand dieser Arbeit. Die Prozessierung wird deshalb an dieser Stelle nur schematisch beschrieben.²

Die einfache Atmosphärenkorrektur basiert für Wasser, dessen Farbe in erster Linie durch den Chlorophyllgehalt bestimmt wird (so genanntes case 1-Wasser), auf der Beobachtung, dass Wasser im Infrarotbereich vollständig absorbiert und damit jegliches Signal in den diskreten Bändern dieses Spektralbereiches durch Streuung in der Atmosphäre zustande kommen muss.

Während einfache Chlorophyllalgorithmen (insbesondere für "grüne" Gewässer im Sinne von case 1-Wasser) nur auf die Verhältnisse von Reflektanzen bei $\lambda = 440$ nm (blau) und $\lambda = 560$ nm (grün), dem Absorptionsmaximum und -minimum des Pigments Chlorophyll (Morel (1974); Morel und Prieur (1977)), zurückgreifen, stützen sich komplexere Prozessoren auf Neuronale Netze. Mit ihnen lassen sich insbesondere in Gewässern, deren Farbe nicht allein vom Chlorophyllgehalt bestimmt wird (case 2), aus den Informationen des gesamten registrierten Spektrums die verschiedenen Inhaltsstoffe gleichzeitig ableiten.

Das Training eines Neuronalen Netzes setzt voraus, dass die optischen Eigenschaften (hauptsächlich Streuung und Absorption) des Wassers und der Kombination verschiedener Zusammensetzungen von Inhaltsstoffen durch Modelle adäquat beschrieben werden können. Die Wahl des Wertebereiches der Konzentrationen im Training und den zu ihren

²Ausführliche Dokumentationen finden sich in den "Algorithm Theoretical Basis Documents", vgl. <http://envisat.esa.int/instruments/meris/atbd/>.

Kombinationen gehörenden Spektren begrenzen die zuverlässig aus Messungen ableitbaren Konzentrationen. Gleichzeitig können gerade bei hohem Schwebstoffgehalt (starker Streuung) die Spektren uneindeutig werden und eine weite Bandbreite an Chlorophyllkonzentrationen zu kaum merklichen Veränderungen im Spektrum führen. Somit besitzt z.B. die Chlorophyllkonzentration eine variable Genauigkeit, die von den übrigen Wasserinhaltsstoffen abhängt.

Des Weiteren gehen die Modelle zur Beschreibung der Wassersäule von ihrer Homogenität aus und berücksichtigen nicht die Vertikalverteilung der Inhaltsstoffe. Gleichzeitig wird die Sichttiefe - und damit der beprobte Wasserkörper, über dessen optische Eigenschaften das Satellitenprodukt eine Aussage trifft - in Abhängigkeit von der Konzentration der Inhaltsstoffe stark räumlich und zeitlich schwanken. In stark getrübbten Gewässern mit hoher Absorption (vornehmlich durch Gelbstoff) wie z.B. den Küstengebieten liegt die Sichttiefe nur in der Größenordnung eines halben Meters, während sie im offenen Meer 10 m und mehr betragen kann.

Die Bestimmung der Wasserinhaltsstoffe anhand von Satellitendaten wird in der Literatur (Doerffer und Schiller (1997)) ausführlich behandelt. Im Allgemeinen basieren die biooptischen Modelle auf *in situ* Messungen der Streuspektren und der Laboranalyse von Wasserproben auf die definierten Inhaltsstoffe, sodass eine empirische Abhängigkeit zwischen den inhärenten optischen Eigenschaften des Wassers wie Streu- und Absorptionskoeffizient als Funktion der Wellenlänge zur (z.B.) Chlorophyllkonzentration hergestellt werden kann (vgl. Abschnitt 1.3.2).

Die Daten sind sowohl in Level 1- als auch in Level 2-Form über ein Internet-Archiv frei verfügbar.

1.3.1 Standardprozessor

Die Produkte (Level 2) des Standardprozessors, die in dieser Arbeit berücksichtigt werden, sind hauptsächlich die Algenprodukte *algal_1* und *algal_2*. Ihre Bezeichnung leitet sich von der optischen Einteilung der Wassermassen ab, deren Farbe von der Konzentration an Phytoplankton und damit dem Chlorophyll *a* dominiert wird (case 1-Wasser) bzw. aus dem Zusammenspiel mehrerer Inhaltsstoffe wie streuenden Schwebstoffen (*total_susp*, total suspended matter, mineralische bzw. anorganische Komponenten wie z.B. resuspendiertes Sediment) und absorbierenden Gelbstoffen (*yellow_subs*, organische Abbauprodukte der Phytoplanktons) entsteht (case 2-Wasser).

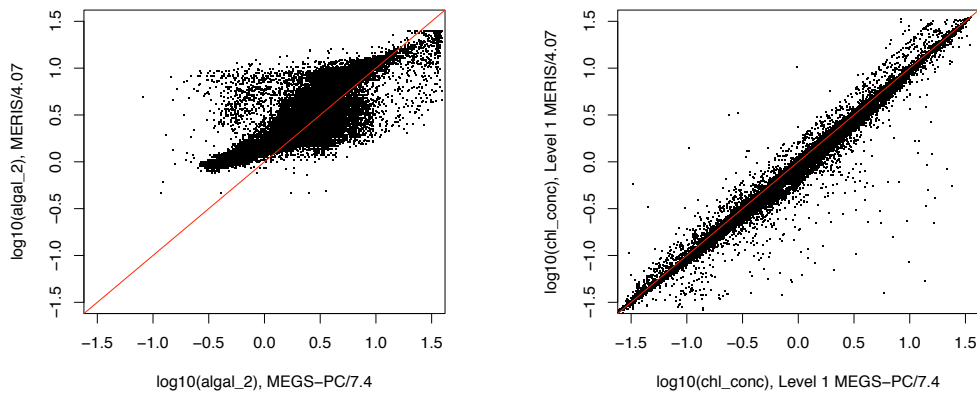
Eine Reihe von Qualitätsmerkmalen (*flags*) werden für jeden Datenpunkt während der Prozessierung überprüft und stehen auch danach zur Auswahl vertrauenswürdiger Datenpunkte zur Verfügung. Die Auswahl wird hier beschränkt auf Datenpunkte, die über Wasser und durch die Beobachtungsgeometrie nicht im Bereich direkter Reflektion an der Wasseroberfläche (*high_glint*) liegen. Ausgeschlossen werden des Weiteren alle Datenpunkte, die als Wolken (*cloud*) oder Dunst (*ice_haze*) identifiziert worden sind oder bei deren Prozessierung nach der Atmosphärenkorrektur unzulässige Spektren z.B. mit negativen Reflektanzen auftreten (*PCD_1_13*). Des Weiteren können Datenpunkte erkannt werden, deren Spektren bei Prozessierung zu Konzentrationen von Wasserinhaltsstoffen oder Aerosoldichten führen, die nicht im Trainingsdatensatz enthalten sind.

1 Einführung zur Satellitenozeanographie

Der Datensatz geht nicht nur auf eine Version des Standardprozessors zurück, so dass die Homogenität der Produkte nur mit Einschränkungen gewährleistet ist (vgl. Abb. 1.1a).

Level2-Daten der Jahre 2003 und 2004 sind mit MEGS-PC/7.4 berechnet worden, das Jahr 2005 mit MEGS-PC/7.4.1 bzw. MERIS/4.10, das Jahr 2006 bis zum 7. Juni mit MERIS/4.10, vom 8. Mai bis zum Ende des Jahres mit MERIS/5.02, die Jahre 2007, 2008 vollständig mit MERIS/5.02. Die Prozessorversionen und ihre Änderungen sind in Report des MERIS-SPPA-Team (2008) dokumentiert.

Inkonsistenz zwischen den einzelnen Prozessoren und ihren Produkten ist zu erwarten. Es wird empfohlen, bei Arbeiten basierend auf Level2-Produkten immer den neusten Prozessor zu verwenden.



(a) Level2-Produkte verschiedener Prozessoren im Vergleich

(b) Chlorophyllprodukt *chl_conc* des Regionalprozessors von verschiedenen prozessierten Level 1-Daten

Abb. 1.1: Level2-Standardprodukt *algal_2* von Prozessoren MEGS-PC/7.4 gegenüber MERIS/4.07, (a). Im Vergleich dazu unterscheidet sich das Level2-Produkt *chl_conc* des Regionalprozessors basierend auf Level 1-Daten unterschiedlicher Prozessoren nur marginal, (b). Datensatz vom 3. August 2004.

1.3.2 Regionalprozessor *meris_c2r*

Der Regionalprozessor (Doerffer und Schiller, 2007) ist ganz auf die Berechnung von Wasserinhaltsstoffen in der Nordsee ausgerichtet, deren Wasser insbesondere in den Küstenregionen in die Kategorie von case 2 fällt. Dabei sind die biooptischen Modelle speziell anhand von *in situ* Beobachtungen aus dieser Region formuliert worden.³

Aus den Level 1-Produkten der Jahre 2003 bis 2008 lässt sich mit dem Regionalprozessor ein konsistentes Algenprodukt ableiten. Die verbleibende Inkonsistenz durch verschiedene Level 1-Prozessierung (vgl. Abb. 1.1b) ist wesentlich geringer als diejenige zwischen

³c2r steht für “case 2 regional”.

1 Einführung zur Satellitenozeanographie

Level-2-Algenprodukten unterschiedlicher Prozessoren (Abb. 1.1a), sodass sich die Annahme eines homogenen Datensatzes rechtfertigen lässt.

Der Regionalprozessor bietet einige Verbesserung in Atmosphären- und Glint-Korrektur. Die Grenzen zwischen den fünf Kameras, aus denen sich MERIS zusammensetzt, sind hier wesentlich unscheinbarer (Abb. 1.2, vgl. später auch Abschnitt 5.5.1).

Die nachfolgenden Betrachtungen basieren hauptsächlich auf dem Chlorophyllprodukt chl_conc , das als abgeleitetes Produkt aus der Absorption der Pigmente bestimmt wird (vgl. Tabelle 1.1). Die empirische Beziehung

$$chl_conc [\mu g/l] = 21 \cdot a_pig^{1.04}$$

ist ein Teil des biooptischen Modells. Weitere wichtige Produkte sind z.B. die optische Dicke des Aerosols (tau_550), die Schwebstoffe (tsm) und die Sichttiefe ($Z90_max$), sowie die Absorption der Phytoplanktonpigmente (a_pig) bzw. die totale Absorption ($a_total = a_pig + a_gelb + a_part$) aus Absorption des Gelbstoffs und die Streuung der Schwebstoffe (b_tsm). Letztere ist die Streuung aller Partikel unter der Annahme, dass die Partikel nicht absorbieren bzw. Auswirkungen der Absorption auf die Volumenstreuungsfunktion bereits in der Absorption der Pigmente (a_pig) und der partikulären Absorption (a_part), d.h. die Absorption der Partikel ohne den Farbstoff der Pigmente, enthalten sind.

Inhärente optische Eigenschaften	Wertebereich	Abgeleitetes Produkt	Wertebereich
$\log(a_pig)$	-6.908 bis 0.6926	Chlorophyllkonzentration $chl_conc [\mu g/l] = 21 \cdot a_pig^{1.04}$	$1.37 \cdot 10^{-6} - 110.2867$ [$\mu g/l$]
$\log(a_gelb+a_part)$	-5.846 bis 2.222	-	-
$\log(b_part+b_whit)$	-4.571 bis 4.086	Schwebstoffkonzentration $tsm [mg/l] = 1.73 \cdot b_tsm$	$4.645 \cdot 10^{-5} - 21088.52$ [mg/l]

Tab. 1.1: Wertebereich der inhärenten optischen Eigenschaften und der daraus abgeleiteten Level-2-Produkte des Regionalprozessors

Vertrauenswürdige Datenpunkte werden nach den folgenden Qualitätskriterien (Level-2-Flags) ausgewählt: sie sollen über Wasser liegen und nicht zu den ungültigen Level-2-Punkten zählen (vgl. Doerffer und Peters (2006)), d.h. die Atmosphärenkorrektur und die Spektren der das Wasser verlassenden Strahldichte müssen innerhalb des Wertebereiches ihres Trainings bleiben.

Die Kriterien sind insgesamt weniger streng, sodass einige zuvor im Standardprozessor ungültige Datenpunkte nun zur Verfügung stehen, insbesondere diejenigen, die im Bereich der direkten Reflektion liegen.

Aus dem Maß für die Eindeutigkeit der Produkte (chi_square , ebenfalls als Produkt gegeben) lassen sich beliebige strengere Auswahlkriterien konstruieren. Chi-square repräsentiert dabei die Breite des Abweichungsminimums zwischen beobachtetem Spektrum und vorwärtsmodelliertem Spektrum bei den Kombinationen verschiedener Konzentrationen von Inhaltsstoffen. Gerade bei hohen Gelbstoffanteilen (hohe Absorption) kann die

1 Einführung zur Satellitenozeanographie

Chlorophyllkonzentration weite Variationen durchlaufen, ohne sich im vorwärtsmodellierten Spektrum bemerkbar zu machen. Diese Ungenauigkeit bzw. Vieldeutigkeit kann durch *chi_square* quantifiziert werden. Diese Möglichkeit für ein individuelles Fehlermaß jedes Bildpixel wird derzeit noch nicht genutzt.

Ergänzt wird die Auswahl durch die Identifizierung nicht erkannter Wolkenränder (Abschnitt 4.2).

Die kritische Begutachtung des Algenproduktes (Kapitel 5, Abschnitt 5.5) führte dazu, das Atmosphärenmodell innerhalb der -korrektur zu überarbeiten und Polarisationseffekte einzubeziehen. Diese vielversprechende Neuprozessierung (im September 2009) findet jedoch aus Zeitgründen keinen Eingang in diese Arbeit.

1 Einführung zur Satellitenozeanographie

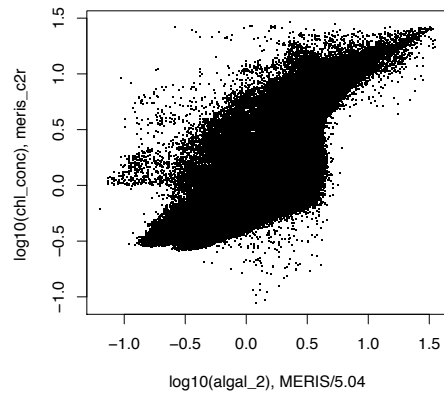
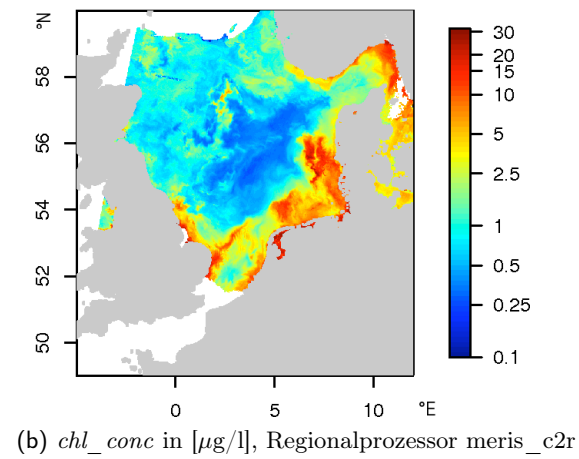
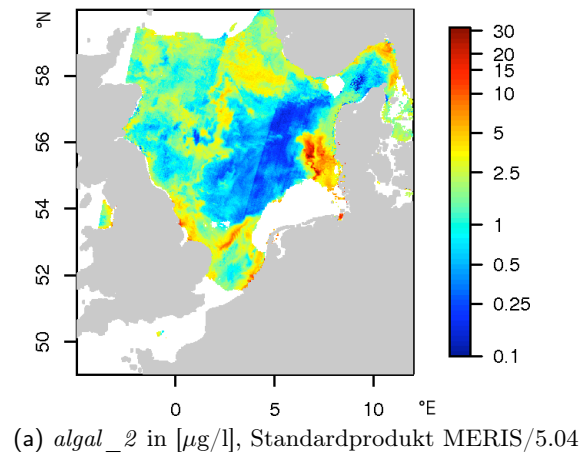


Abb. 1.2: Chlorophyllprodukt des Regionalprozessors im Vergleich zu dem des MERIS-Standardprozessors MERIS/5.04 vom 10. Mai 2006. Alle abgebildeten Punkte sind als vertrauenswürdig ausgewählt worden. Deutlich ist im Standardprodukt ein streifiger Versatz der Chlorophyllkonzentration zu erkennen, der mit Kamera 4 des Spektrometers zusammenfällt.

2 Ozeanographische Beschreibung der Nordsee

Dieses Kapitel soll und kann nur eine knappe Einführung in die Ozeanographie der Nordsee leisten. Für ausführlichere Beschreibungen ihrer physikalischen Aspekte sei auf die Zusammenstellung von Huthnance (1991) oder das einleitende Kapitel von Becker (1990) verwiesen.

2.1 Geographische Definition der Nordsee

Laut der Definition des Internationalen Hydrographischen Büros verlaufen die Grenzen der Nordsee zwischen der engsten Stelle im Englischen Kanal, von Norwegen entlang des 61°-Breitengrades bis 61° N, 0°53' W über die Shetland- und Orkneyinseln nach Schottland, ausschließlich des Skagerraks (Holligan u. a. (1989); Huthnance (1991)).

Die Nordsee besitzt eine Fläche von 575 300 km² und ein Wasservolumen von 40 300 km³.

Als Schelfmeer ist sie im Mittel etwa 70 m tief, wobei die Tiefe im Allgemeinen gen Norden zunimmt. Topographisch lässt die Nordsee sich in drei Großbereiche einteilen:

- in der *Südlichen Bucht* (zwischen 51° N und 54° N) inklusive der *Deutschen Bucht* (z.B. nach Modell ERSEM etwa 6° - 9° O und 53° - 55.6° N, Abb. 2.1b) variiert die mittlere Tiefe zwischen 20-40 m,
- in der *Zentralen Nordsee* (54° N bis 57° N) nimmt sie auf 100 m zu.
- Die *Nördliche Nordsee* (nördlich von 57° N) besitzt mittlere Wassertiefen zwischen 100 und 200 m, maximal können in der Norwegischen Rinne Tiefen bis zu 700 m auftreten.¹

Markante Formationen sind einerseits die Norwegische Rinne mit einer Breite von etwa 100 km und einer mittleren Tiefe von 300 m (maximale Tiefe im Skagerrak \simeq 700 m) und andererseits Sandbänke im südlichen Bereich, die teilweise bis auf 10 m an die Wasseroberfläche herantreten. Deren ausgedehnteste Formation ist die Doggerbank (1° - 4° O, 54° - 55° N) mit einer Wassertiefe von etwa 20 m (siehe Abb. 2.1a). Dabei sind die trockenfallenden Wattenflächen und die den Küsten vorgelagerten Inselsysteme nicht berücksichtigt.

¹Diese Unterteilung wird - in leichter Abwandlung - auch von den Bearbeitern des Continuous Plankton Recorders (CPR) verwendet, die wiederum auf die International Council for the Exploration of the Sea - kurz ICES - zurückgeht. Angaben finden sich für die Großbereiche ICES Statistical Sub-Area IVa (Northern North Sea) zwischen 57.3° N und 62° N, ICES Statistical Sub-Area IVb (Central North Sea) zwischen 53.3° N und 57.3° N und ICES Statistical Sub-Area IVc zwischen 51° N und 53.3° N.

2 Ozeanographische Beschreibung der Nordsee

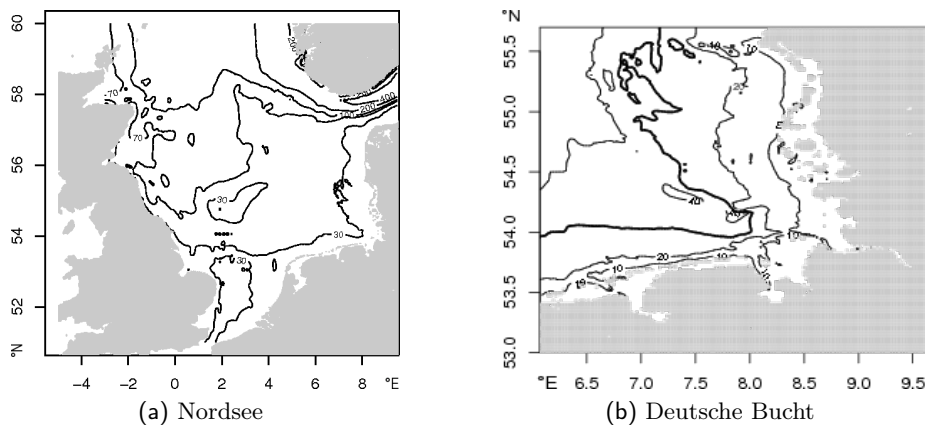


Abb. 2.1: Bathymetrie der Nordsee, wie sie im Strömungsmodell des BSH Verwendung findet, für die (a) Nordsee und für die (b) Deutsche Bucht. Gezeigt sind Isolinien der Tiefe in Meter.

Mit Rücksicht auf die Überdeckung durch Satellitendaten wird das Gesamtgebiet der Nordsee im Folgenden als die Wassermassen zwischen 49° - 60° Nord und 5° West bis 12° Ost definiert sein.

2.1.1 Untergliederung nach Wassermassen

Die Wassermassen der Nordsee lassen sich nach Lee (Lee, 1980, vgl. Abb. 8 in Huthnance, 1991) nach hydrographischen Gesichtspunkten, d.h. anhand von Salzgehalt und Temperatur, in verschiedene Regionen einteilen.

Auch die räumliche Definition von Modellregionen (z.B. nach ICES oder ERSEM) orientiert sich an dieser Gliederung und wird im Rahmen dieser Arbeit zur Einteilung der Nordsee in Regionen übernommen.

In den folgenden Kapiteln werden elf Regionen (siehe Abb. 2.2) unterschieden, wie sie für Modelluntersuchungen zur Zirkulation und zum Ökosystem der Nordsee verwendet werden (ERSEM, Lenhart u. a. (1995, 1997)). Die Benennung wird angepasst, um ähnliche Gebiete bereits namentlich kenntlich zu machen. Da hier im Gegensatz zum Modell nur die Oberflächenschicht betrachtet wird, werden die Regionen rein nach geographischen Grenzen eingeteilt, ohne eine Namensgebung für Zwei-Schicht-Systeme bei tieferen Teilen des Untersuchungsraumes notwendig zu machen.

Diese Regionen lassen sich anhand der Wassermassenbezeichnungen nach Huthnance (1991) wie folgt charakterisieren:

- area 1: Nordatlantikwasser (B) sowie Mischung aus (B) + Skagerrak (C), starke Salzgehaltsgradienten nahe der Oberfläche, Wirbelbildung (D), im Modell ERSEM werden hier Boxen für das Zwei-Schicht-System angesetzt.
- area 2: Nordatlantikwasser (B), ERSEM: Zwei-Schicht-System.

2 Ozeanographische Beschreibung der Nordsee

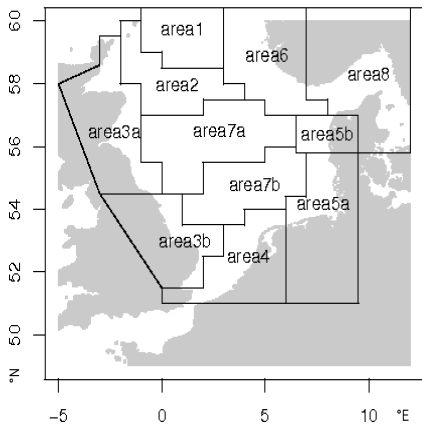


Abb. 2.2: Einteilung der Nordsee nach ERSEM-Modell basierend auf Wassermassen

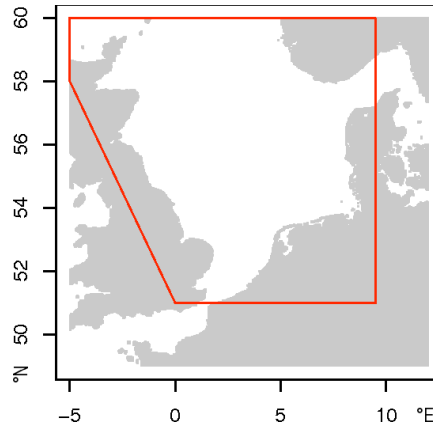


Abb. 2.3: Definition der Nordsee als einfach zusammenhängendes Gebiet

- area 3a: Schottische Küstengewässer: durch Flüsse saisonal niedriger Oberflächen-salzgehalt (A1)
- area 3b: Englische Küstengewässer: vertikal durchmischt durch Tideneinfluss (A2)
- area 4: „*Englischer Kanal*“, Kanalwasser: hohe Temperatur und Salzgehalt im Winter (A3) und Kontinentale Küstenwasser (Continental Coastal Water): geringe Salzgehalt, lokale Schichtung durch Salzgehalt, große jährliche Temperaturschwankungen (A4)
- area 5a: „*Deutsche Bucht*“, Kontinentale Küstenwasser (Continental Coastal Water, A4)
- area 5b: „*Dänische Küste*“, Mischung aus (B) + Skagerrak (C), starke Salzgehaltsgradienten nahe der Oberfläche, Wirbelbildung (D)
- area 6: „*Norwegische Rinne*“, Skagerrak (C) + Mischung (D), ERSEM: Zwei-Schicht-System.
- area 7a: „*Zentrale Nordsee*“, Nordatlantikwasser (B) + Mischung (D), ERSEM: Zwei-Schicht-System
- area 7b: „*Doggerbank*“, Mischung aus A1-A4 + B (A6) und starker ost-westlicher Salzgehaltsgradient, thermohaline Schichtung bricht im August zusammen (A5), ERSEM: Zwei-Schicht-System
- area 8: „*Skagerrak*“ (C), in ERSEM nicht vorhanden.

Regionen, die auch später Gegenstand der Untersuchung sind, erhalten neben der Nummerierung gemäß der Boxen des ERSEM-Modells auch beschreibenderen Titel.

Die Schelf-Wassermassen sind nur während der Wintermonate durch Wind und Tide vollständig durchmischt, während in der Norwegischen Rinne die thermohaline Schichtung über das ganze Jahr herrscht. In den küstennahen Bereichen gerade im Süden und Osten der Nordsee mit den sehr geringen Wassertiefen ist mit beständiger tidenbedingter vertikaler Durchmischung zu rechnen.

Durch die geringe Wassertiefe und die Nähe zu den Landmassen ist der Temperaturjahresgang in den küstennahen Gewässern am stärksten ausgeprägt. Gleichzeitig ermöglichen meteorologische Einflüsse wie solare Einstrahlung, Wind, Niederschläge, etc. eine starke interannuale Variabilität.

2.2 Definition der Untersuchungsräume

Für spätere Untersuchungen werden Räume definiert, die in erster Linie von der Einteilung nach Wassermassen ausgehen bzw. sich zum direkten Vergleich auf die in anderen Projekten getroffenen Einteilungen stützen.

2.2.1 Nordsee

Um die Wassermassen der Nordsee als einfach zusammenhängendes Gebiet untersuchen zu können, werden nur jene Datenpunkte über Wasser berücksichtigt, die in den Bereich zwischen 51° N und 60° N, sowie als westliche Grenze im Norden bis 5° W, im Süden ab 0° O und als östliche Grenze bei 9.5° O fallen (Abb. 2.3).

2.2.2 ICES-Regionen

Die Unterteilung der Nordsee gemäß ICES, die auch im ERSEM-Modell Verwendung findet, ist im Abschnitt 2.1.1 (insbesondere Abb. 2.2) dokumentiert und orientieren sich an Wassermassen und -tiefe.

2.2.3 Gebiete 1-5

Da eine Fast-Fourier-Transformation (FFT) rechteckige Untersuchungsgebiete voraussetzt, werden anhand der ICES-Einteilung der Nordsee (Kapitel 2, insbesondere Abb. 2.2) entsprechende Bereiche ausgewählt. Ebenso wird die Häufigkeit von Beobachtungen im Jahr berücksichtigt (vgl. Abb. 6.1b), sodass vermehrt Räume im südlichen Teil der Nordsee ausgesucht worden sind.

Rechteckig bezieht sich hier auf eine simple Projektion auf das Standardgitter (vgl. Abschnitt 4.3) in Länge und Breite, die Seiten der Rechtecke zeichnen sich dementsprechend durch jeweils konstante Länge oder Breite aus.

Gebiet 1, „*Doggerbank*“, fällt dabei mit dem westlichen Teil von ICES area 7b, dem nordwestlichen Teil der Doggerbank, zusammen (86×111 Pixel, $127.0 \text{ km} \times 164.7 \text{ km}$, $1.5 - 3.5^{\circ}$ O, $54 - 55.5^{\circ}$ N).

2 Ozeanographische Beschreibung der Nordsee

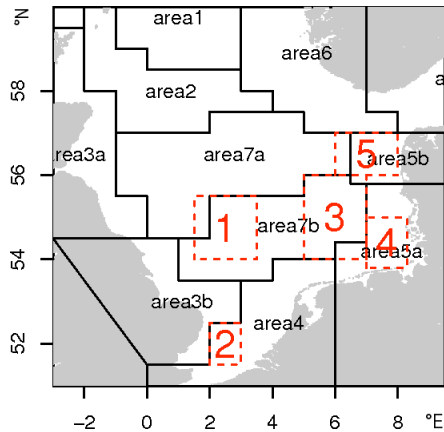


Abb. 2.4: Rechteckige Untersuchungsgebiete 1-5 angelehnt an ICES-Regionen

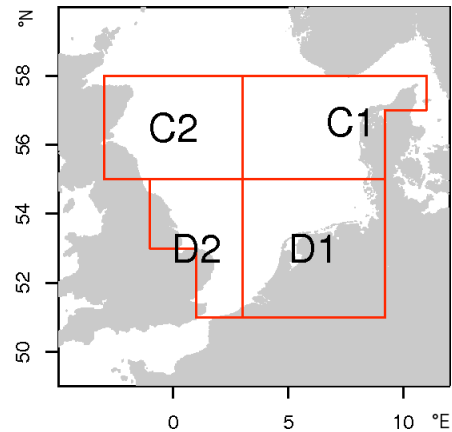


Abb. 2.5: Standardareale in der Langzeit-Auswertung des Continuous Plankton Recorders, CPR.

Gebiet 2, „*Englischer Kanal*“, liegt vollständig innerhalb von ICES area 4, relativ zentral im Einstromgebiet des Englischen Kanals und der in den Schwebstoffen sichtbaren Themsefahne bis vor die belgische Küste (43×74 Pixel, $67.0 \text{ km} \times 109.3 \text{ km}$, $2 - 3^\circ \text{ O}$, $51.5 - 52.5^\circ \text{ N}$).

Gebiet 3, „*Zentrale Nordsee*“, umfasst den östlichen Teil der Doggerbank (ICES area 7b, 86×148 Pixel, $126.3 \text{ km} \times 220.1 \text{ km}$, $5 - 7^\circ \text{ O}$, $54 - 56^\circ \text{ N}$).

Gebiet 4, „*Deutsche Bucht*“, bildet die gesamte Deutsche Bucht ab (ICES area 5a, 56×89 Pixel, $82.9 \text{ km} \times 131.7 \text{ km}$, $7 - 8.3^\circ \text{ O}$, $53.8 - 55^\circ \text{ N}$), während

Gebiet 5, „*Dänische Küste*“, vor der Dänischen Küste südlich des Skagerrak bzw. der Norwegischen Rinne zu liegen kommt (ICES area 5b, 86×75 Pixel, $121.5 \text{ km} \times 110.8 \text{ km}$, $6 - 8^\circ \text{ O}$, $56 - 57^\circ \text{ N}$).

2.2.4 Weitere Areale

Eine grobe Einteilung der Nordsee findet sich für die Datensätze des Continuous Plankton Recorders (CPR). Für diese sehr simple Einteilung der Nordsee, die die speziellen räumlichen Eigenschaften kaum berücksichtigt, können Auswertungen des Phytoplankton-Farbindex (Grünindex) direkt erhalten werden (Abb. 2.5)²:

C1 „*Östliche zentrale Nordsee*“, mit Teilen der Norwegischen Rinne und des Skagerraks

C2 „*Westliche zentrale Nordsee*“

²vgl. http://www.sahfos.ac.uk/standard_areas.htm, Sir Alister Hardy Foundation for Ocean Science, Sahfos

- D1 „Südöstliche Nordsee“, Deutsche Bucht sowie alle Küstenregionen der Niederlande und Belgiens
- D2 „Südwestliche Nordsee“, Teile der Doggerbank

2.3 Physikalisch-ozeanographische Einflüsse

Einige physikalische Einflüsse, die über den Wasserkörper auf das Phytoplankton und seine Dynamik einwirken, sollen hier zusammengestellt werden.

2.3.1 Tiden und Strömungen

Die annähernd halbtägliche (lunare) Tide (M_2 , Periode 12.42h) kommt für den Großteil der Varianz in Wasserstand und Strömung innerhalb der Nordsee auf. Die Tiden pflanzen sich zyklonisch (gegen den Uhrzeigersinn gemäß der Corioliskraft) entlang der Küsten fort, mit den größten Amplituden an der östlichen Küste von England und in der Deutschen Bucht.

Die Geschwindigkeitsamplitude der M_2 -Tide in der Deutschen Bucht beträgt etwa 0.5 m/s (vgl. Schirmer u. a. (1994)), die mittlere Strömungsgeschwindigkeit liegt im Norden und Osten bei etwa 0.2 m/s, südlich von 54° N bei 0.5 m/s (Huthnance (1991)).

Für die Durchmischung der Wassersäule ist ein Strömungsgeschwindigkeit von 1 m/s etwa so effektiv wie ein Windgeschwindigkeit von 25 m/s (Huthnance (1991)), so dass die Tidenströme insbesondere im südlichen Teil der Nordsee (südlich 54° N) einen wichtigen Anteil an der Durchmischung haben.

Die Tiden erzeugen oszillatorische Verläufe der Strömungsgeschwindigkeit (Tidenellipse³, vgl. für Deutsche Bucht in Schirmer u. a. (1994)). Das Tidensignal kann mit der zeitlichen Auflösung der Satellitenbilder von etwa 24 h nicht aufgezeichnet werden.

2.3.2 Schichtung und Fronten

Ob in einem Gebiet thermische Schichtung oder vollständige, vertikale Durchmischung auftritt, schätzen Simpson und Hunter (1974) durch das Verhältnis von Dissipation von Tidenenergie ($\propto u^3$) gegen potentielle Energie ($\propto h$) bedingt durch Erwärmung respektive Einstrahlung ab.

Der Schichtungsparameter S (engl. stratification parameter, auch Simpson-Hunter-Parameter, Pingree-Linie; Pingree (1978)) berechnet sich aus

$$S_{c.g.s.} = \log_{10} \frac{h}{C_p |u|^3}$$

³Die so genannte Tidenellipse ist eine spezielle Form der Darstellung, bei der die Strömungsvektoren - gemittelt über jeweils eine Stunde - über einen Tidenzyklus jeweils von der Spitze des vorhergehenden Vektors aufgetragen werden. Die Ellipse gibt nicht die Bahn des Wasserpartikels wieder, sondern die momentane, mittlere Geschwindigkeit an einem Ort.

2 Ozeanographische Beschreibung der Nordsee

Tab. 2.1: Übersicht der Literaturwerte für Schichtungsparameter S (Intervall nach Pingree in SI-Einheiten umgerechnet)

Autor	S (nach Gl. 2.1)	Untersuchungsgebiet	Definition therm. Front
Simpson und Hunter (1974)	1.74	Irische See	horizontaler Gradient: $\delta T = 1^\circ\text{C km}^{-1}$
Pingree (1978)	2.4 - 3.4	1. westl. Nordsee, 2. nordwestl. Nordsee, 3. zwischen Schottland und Irland, 4. Celtic Sea und westl. English Channel	1. $\Delta T = 1^\circ\text{C}$, 13°C 2. $\Delta T = 1 - 2^\circ\text{C}$, 11°C 3. 13°C 4. $\Delta T = 3^\circ\text{C}$, 15°C
Bowers und Simpson (1987)	2.7 ± 0.4	Irische See	$\Delta T = 0.5^\circ\text{C}$
Sharples (2008)	2.4 - 2.8	Keltische, Irische See	$\Delta T = 1^\circ\text{C}$
Holt und Umlauf (2008)	3.0 ± 0.8	ICES-Datensatz, Europ. Schelfmeere	$\Delta T = 0.5^\circ\text{C}$

mit Wassertiefe h [cm], Reibungskoeffizient am Boden $C_P = 0.0025 \frac{\text{s}^3}{\text{cm}^2}$, $|u|$ instantane Strömungsgeschwindigkeit [cm/s] (Amplitude des Tidenstromes)⁴. Der Querstrich steht für eine zeitliche Mittelung über einen Tidenzyklus, der hier mit der Periode der Hauptkomponente der M_2 -Tide von 12.42 h gleichgesetzt wird.⁵

Heutzutage wird eher die Definition

$$S = \log_{10} \frac{h}{U^3} \quad (2.1)$$

verwendet, mit Tiefe h [m], über einen Tidenzyklus gemittelte Geschwindigkeitsamplitude U [m/s].

Bei einem Schichtungsparameter $S = 2.9$ (im Original im c.g.s.-System $S = S_{c.g.s} = 1.5$) liegt der Übergang zwischen beständiger vollständiger Durchmischung (durch die Tidenströme) und saisonaler Schichtung. Dieser gewählte Wert fällt gut mit Beobachtungen der Frontregion von Temperaturfronten im Sommer für Wassertiefen zwischen 40-100 m (Pingree und Griffiths (1978)) zusammen. Nach Bowers und Simpson (1987) treten die meisten Tidenfronten zwischen 30 und 100 m auf, in Bereichen mit Tiefen kleiner als 30 m wird demnach keine thermische Schichtung auftreten, haline Schichtung ist aber sehr wohl möglich. Für Werte $S < 2.4$ ist das Wasser ständig tidendurchmischt, für $S > 3.4$ kommt

⁴In Pingrees numerischem Modell werden die Tidenströme als konstant mit der Tiefe angenommen. u ist die vertikal gemittelte horizontale Geschwindigkeitsamplitude.

⁵Die Umrechnung zwischen dem Parameter S nach Pingree in c.g.s.-Einheiten inklusive C_P zu den heutzutage gebräuchlicheren Parametern $\hat{S} = \log_{10} \frac{h}{U^3}$ in SI-Einheiten erfolgt so: Pingrees Frontkriterium lautet $S_{c.g.s} = 1.5$; daraus folgt (gestrichene Größen in c.g.s.):

$$S_{c.g.s} = \log_{10} \frac{h'}{C_P U'^3} = \log_{10} \frac{10^{-4}}{C_P} \frac{h}{U^3} \Rightarrow S_{c.g.s} - \log_{10} \frac{10^{-4}}{C_P} = \log_{10} \frac{h}{U^3} = S_{SI}$$

Entsprechend ist $S_{SI} (S_{c.g.s} = 1.5) = 1.5 - -1.4 = 2.9$.

es zu saisonaler Schichtung im Sommer. Erhöhte Chlorophyllkonzentrationen im Sommer fallen ebenfalls in diesen Übergangsbereich (vgl. Abschnitt 6.3.2).

Maximale Oberflächentemperaturen finden sich Ende August, wenn auch die Schichtung am stärksten ausgeprägt ist (Modell von Elliott und Clarke (1991)). Im Allgemeinen zeigt sich in der Nordsee eine stärkere Abhängigkeit der Oberflächentemperatur von der Wassertiefe als von der geographischen Breite, wobei die Effekte der Landmassen im Sommer und Winter in den sehr flachen Regionen zu starken jahreszeitlichen Temperaturschwankungen führen.

Viele *in situ* Beobachtungen insbesondere in den Sommermonaten berichten von kleinräumiger Schichtung auch in küstennahen Gewässern mit Wassertiefen unter 30 m.

2.3.3 Vertikale Durchmischung durch Wind

Der Antrieb der Durchmischung mittels Wind oder mittels der Tide spielt sich auf unterschiedlichen räumlichen Größenskalen ab: das Windfeld ist über größere Distanzen kohärent und variiert glatt über den Raum, während die Tide starke Unterschiede in Distanzen von 100 km aufweisen kann. Für die Deutsche Bucht lässt sich sogar bei einer Beschränkung auf eine Zeitdauer von einer Woche von einer konstanten Wetterlage⁶ ausgehen (Schrum (1997)). Je nach Windrichtung werden sich in der Deutschen Bucht verschiedene Schichtungsmuster ausbilden.

Die folgende Aussagen sind von Schrum aus einem Modell der Deutschen Bucht abgeleitet worden, das sich im Ausgangszustand im klimatologischen Mittel für den Monat Mai bezüglich Temperatur- und Salzgehaltsverteilung befindet:

Bei Wind aus Süd bzw. Südwest tritt die geringste Schichtung auf; es gibt keine haline Schichtung und thermische tritt nur in den tieferen Gebieten im Nordwesten auf. Um eine Thermokline bilden zu können, benötigt es eine minimale Wassertiefe von 35 m bei Tidengeschwindigkeiten von $u_0 = 0.7 \text{ m/s}$ und täglichen Temperaturschwankungen im durchmischten Oberflächenwasser von 0.2°C .

Bei Wind aus West kann auch südlich von Helgoland haline Schichtung auftreten, bedingt durch barokline Instabilität.

Trotz barokline Instabilität können solche Gebiete vollständig vertikal durchmischt sein, wenn die mittlere Tidenzirkulation groß genug ist.

Bei östlichen Winden wird die thermische Schichtung am ausgeprägtesten sein, maximale Ausdehnung findet sich bei nordöstlichen Winden. Selbst in Bereichen mit Wassertiefen unter 20 m kann sie auftreten.

Die zunehmende Windgeschwindigkeit verlagert die Thermokline in größere Tiefen und vermindert ihre Intensität.

Die thermohaline Schichtung verhindert die Zufuhr von Nährstoffen aus den unteren Wasserschichten und besitzt damit weitreichende Auswirkungen auf die ökologischen Prozesse.

Auswirkungen der Strömungsgeschwindigkeitsunterschiede zwischen Spring- und Nipp tide auf die Schichtbildung sind laut Schrum klein.

⁶constant meteorological condition, bedeutet konstant in Zeit und Raum.

2 Ozeanographische Beschreibung der Nordsee

Die Voraussage, wann Schichtung einsetzt oder sich auflöst, hängt teilweise von den räumlichen Auflösungen der verwendeten Modelle oder der Definition, wann eine Schichtung vorliegt, ab. Mit der Definition eines Temperaturunterschiedes von $T_s - T_b > 1.0^\circ\text{C}$ zwischen Oberfläche und Boden finden sich Angaben zum Einsatz der Schichtung zwischen Mitte April (Pingree) und Mitte Mai (Elliott und Clarke (1991)) und ihrer vollständigen Auflösung zwischen Ende November (Elliott) und Ende Dezember für die gesamte Nordsee mit Ausnahme des Englischen Kanals und der südlichen Nordsee, in der die starken Tidenströme die Durchmischung im ganzen Jahr aufrecht erhalten.

3 Phytoplankton in der Nordsee

In späteren Kapiteln wird das Algenprodukt von MERIS im Vordergrund stehen und dabei die mechanistischen Prozesse der Dynamik (Abschnitt 3.1) in den Hintergrund treten. Die vom Satelliten ermittelte Pigmentkonzentration (dominiert durch Chlorophyll *a*) bildet in erster Linie die oberflächennahe mittlere Chlorophyllkonzentration des Meerwassers ab, nicht jedoch die Tiefenstruktur der Algenverteilung (Abschnitt 3.2).

Satellitenfernerkundung kann Phänomene auf regionalen und globalen Skalen (Levin (1992)) beobachten und liegt damit an der Schnittstelle zwischen kleinskaligen ökologischen Studien und den großskaligen Modellen z.B. zur Identifizierung globaler Klimaänderungen. Im Allgemeinen hängt die beobachtete Variabilität des Systems von der Skala - im Sinne von räumlicher und zeitlicher Auflösung - ab (Steele (1978)).

Ökologische Prozesse finden auf

- zeitlichen Skalen von Tagen (oder Stunden) zur Beschreibung der Entwicklung von Phytoplanktonblüten bis zu Monaten beim Studium langjähriger Trends und
- räumlichen Skalen zwischen im cm-Bereich aufgelösten Wassersäulen bei eindimensionalen Ökosystemmodellen zu großen, flächigen Regionen (z.B. Abschnitt 2.2.2) zur Kalkulation von Primärproduktionsbudgets statt.

3.1 Jahresgang und Sukzession

Typischerweise wird der folgende Jahresgang des Phytoplanktons sowohl bezüglich der Abundanz als auch in der Abfolge der Arten in der Nordsee erwartet (Johns und Reid (2001)):

In den Wintermonaten reichern sich Nährstoffe (Stickstoff, Phosphor, Silikat und Ammonium) bei geringer Produktivität durch die limitierende Sonneneinstrahlung an. Sobald im Frühjahr Licht in ausreichendem Maße zur Verfügung steht, beginnen die ersten Diatomeenblüten (z.B. Mills u. a. (1994), Tett und Walne (1995), Gypens u. a. (2007), etc.). Ihre Maxima erreichen sie meist im April, wenn bereits vereinzelt thermische Schichtung auftreten kann und das Verdichten der Konzentration ermöglicht. Nachfolgend, wenn das schwindende Silikat die Produktivität der Diatomeen nicht länger begünstigt, erreichen Flagellaten und später Dinoflagellaten ihr Biomassemaximum. Dieser zweite Teil der Blüte ist hauptsächlich von den nach der Diatomeenblüte verbleibenden Nährstoffen Stickstoff und Phosphat begrenzt. Mit einsetzender Schichtung werden die mit Licht versorgten Wasser von der Nährstoffzufuhr aus größeren Tiefen abgeschnitten. In den Sommermonaten sinkt und sammelt sich das Phytoplankton in tieferen Schichten bzw. der Pyknokline, die Produktivität ist reduziert. Im Herbst mit den ersten Stürmen und

damit einhergehender vollständiger Durchmischung der Wassersäule bzw. Auflösung der Schichtung, werden den Algen erneut Nährstoffe zugänglich und eine zweite Blüte kann auftreten.

3.1.1 Phytoplanktodynamik in Ökosystemmodellen

Für die Nordsee ist eine Reihe von dreidimensionalen Ökosystemmodellen entwickelt worden, um Einblick in die jährliche Dynamik des Phytoplanktons, seine regionale Variation, die Abfolge von auftretenden Arten und die Regulation dieser Prozesse zu erhalten. Übersichten und Vergleiche verschiedener Modellstudien und ihrer Ergebnisse finden sich in Moll und Radach (2003) bzw. Radach und Moll (2006). Die eingangs formulierte Erwartungshaltung zum Jahresgang und seinen Abhängigkeiten geht dabei vielfach auf Erkenntnisse aus Modellstudien zurück.

Dabei variiert die Komplexität der Modelle abhängig von der Anzahl betrachteter Zustandsgrößen stark. Z.B. kann durch das relativ einfache Modell ECOHAM (Ecological North Sea Model, Hamburg) die Entwicklung der Frühjahrsblüte in der Nordsee trotz der Reduktion des Phytoplanktons auf nur eine Gesamtmassengröße, die durch einen Nährstoff (Phosphat oder gelösten anorganischen Stickstoff, DIN) aufgelöst wird, in einer Weise vorausgesagt werden, die großteils auch den Beobachtungen in den Satellitendaten entspricht.

Die Blüte beginnt im küstennahen Bereich hauptsächlich im Bereich des Ausstroms der Ostsee und entlang der Norwegischen Rinne (vgl. die Beispiele der frühesten Algenblüten im Jahr, z.B. Tabelle 5.2 oder Abschnitt 6.3.3) und bestimmt im April die ganze Nordsee. Mit Ausnahme einiger Küstenbereiche erreicht die Phytoplanktonabundanz durch die Nährstofflimitierung aufgrund von Schichtung im Juli ihr Minimum.

Wesentlich detailreicher gestalten sich Studien mittels ERSEM (European Regional Seas Ecosystem Model), in dem vier Nährstoffzyklen (N, P, Si, O) sowohl im Wasser als auch im Boden, Zooplanktoneinflüsse und vier Phytoplanktonrepräsentationen Picophytoplankton ($0.2 - 2 \mu\text{m}$), Flagellaten ($2 - 20 \mu\text{m}$), Diatomeen ($20 - 200 \mu\text{m}$) und Dinoflagellaten (für Zooplankton nicht essbar, $20 - 200 \mu\text{m}$) berücksichtigt werden. Die Zustandsvariable „Chlorophyll“ wird dann als Summe des Chlorophyllgehaltes dieser vier Größen bestimmt. Die theoretisch erwartete Sukzession der Gruppen lässt sich durch das Modell darstellen, jedoch ist die Übereinstimmung mit Beobachtungen zum Teil sehr gering.

3.1.2 Beginn der Frühjahrsblüte

Besonderes Augenmerk gilt der Frage, an welchen Parametern sich der Beginn der Frühjahrsblüte festmachen lässt.

Lichtbedingungen sind für die Diatomeenblüte ausschlaggebend, sodass z.B. eine hohe Konzentration von Schwebstoffen in der Wassersäule den Beginn der Blüte verzögern kann, während wolkenarme Perioden im zeitigen Frühjahr ausreichen können, um das vermehrte Wachstum auszulösen (Townsend u. a. (1994)). Die Frühjahrsblüte kann Ende Februar einsetzen, wenn sich nach einem Sturm die Lichtverhältnisse wieder zu nicht limitierend gebessert haben. Ist Schlick vorhanden, kann ein stürmisches Frühjahr Re-

suspension auslösen und die Primärproduktion durch Lichtlimitierung geringer ausfallen, sodass selbst der verstärkende Einfluss spät einsetzender Schichtung aufgehoben wird.

Der Zeitpunkt der Frühjahrsblüte hängt auch von der verbleibenden Phytoplanktonkonzentration am Ende des Winters ab.

Zusätzlich hängt der Beginn von Frühjahrsblüten zu gewissem Grad vom Spring-Nipp-Tiden-Zyklus (vgl. Sharples (2008)) ab. Hauptsächlich ist dieser Zyklus für die einsetzende Schichtung verantwortlich, wenn z.B. die schwächeren Nipptidenströmungen nicht ausreichen, um die Wassersäule weiterhin vollständig zu durchmischen. Während der Nipptiden können sich temporäre Schichtungen mit einer durchmischten Oberflächenschicht bilden, in denen die Diatomeen gute Wachstumsbedingungen vorfinden (Cloern (1991); Sharples u. a. (2006)). Dies ist letztlich eine Illustration für das Konzept der kritischen Tiefe (Sverdrup (1953)), die neben der kritischen Turbulenz (Huisman (1999)) als Modell vielfach zur Erklärung von Phytoplanktonblüten herangezogen werden (z.B. Obata u. a. (1996)).

Blüten - je nach Definition werden unterschiedliche Grenzen von Chlorophyll- oder Biomasse-Konzentrationen bemüht - können sich entwickeln, wenn die Tiefe der durchmischten Schicht geringer als die kritische Tiefe ist, oder die turbulente Durchmischung geringer als die kritische Turbulenz, wobei es sich um unabhängige Größen handelt. Letzteres bedeutet insbesondere, dass theoretisch Phytoplanktonblüten bei vollkommener vertikaler Durchmischung möglich sind. Kritische Turbulenz bzw. Tiefe sind jeweils charakteristische Funktionen des betrachteten Phytoplanktons.

In vollständig durchmischten Wassern werden die Phytoplanktonarten bevorzugt auftreten, die die größte kritische Tiefe besitzen, d.h. die in den mit der Tiefe abnehmenden Lichtverhältnissen am längsten biologisch aktiv bleiben können. Geringe vertikale Durchmischung begünstigt nicht nur die Bildung von Phytoplanktonblüten, sondern auch die Ausbildung der Thermokline, die jedoch in keinem kausalen Zusammenhang stehen (müssen).

Insbesondere der Entstehungszeitpunkt der thermalen Schichtung hat für die gesamte weitere Phytoplanktondynamik in den Sommermonaten große Auswirkung, sei es gemessen in Primärproduktion oder Sukzession (Ruardij u. a. (1997), 1D-Modell basierend auf ERSEM).

Der Wechsel dominierender Phytoplanktonarten (Sukzession) im Frühjahr vollzieht sich mit Einsetzen der thermalen Schichtung innerhalb von wenigen Wochen.

Schwankungen in der Schichtung (Intensität oder Tiefe) haben jedoch auch länger anhaltenden Einfluss auf die physikalischen und biologischen Eigenschaften der Wassersäule. Z.B. hat die Wind induzierte Durchmischung nicht nur kurzzeitige Auswirkungen auf die Tiefe der Oberflächenschicht, sondern länger anhaltende auf den Gesamtwärmeinhalt der Wassersäule (Elliott und Clarke (1991)). Die kurzfristigen Durchmischungsereignisse stellen (wahrscheinlich) die Nährstoffversorgung während des Sommers sicher.

3.1.3 Langzeitentwicklungen - Continuous Plankton Recorder

Derzeit repräsentieren Ökosystemmodelle die einzige Möglichkeit, um Langzeitentwicklungen im Jahresgang mechanistisch, d.h. bezüglich der Einflüsse von Nährstoffen, Licht,

3 Phytoplankton in der Nordsee

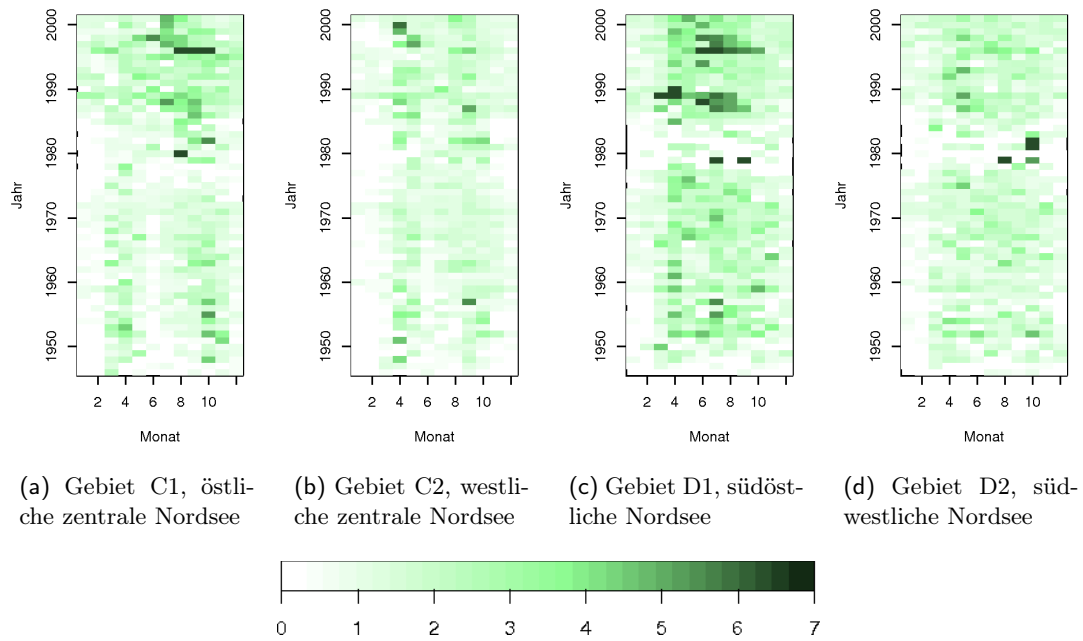


Abb. 3.1: Monatlicher Grünindex (Kategorien der Grünfärbung, ohne Einheit) des Continuous Plankton Recorder für verschiedene Regionen (vgl. Abschnitt 2.2.4)

Schichtung, aber auch Wechselwirkungen mit dem Benthos und Zooplankton, zu untersuchen. Beim Studium dieser Prozesse können Algenprodukte aus Satellitendaten (innerhalb ihrer Genauigkeit), sei es zur Assimilation in Modelle oder zur Validierung derselben, gute Beiträge liefern. Aussagen über zukünftige Trends in der interannualen Entwicklung sind natürlich allein der Modellierung zugänglich (Radach und Moll (2006)). Satellitendaten werden in zukünftigen Validierungen von Modellen eine wichtige Rolle spielen, da sie eine Überprüfung der räumlichen Voraussagen ermöglichen, und das nicht nur punktuell, sondern flächig für einen vergleichsweise kurzen Zeitraum.

Langfristige Entwicklungen lassen sich außer in Modellen am besten mit dem *Continuous Plankton Recorder (CPR)* verfolgen. Auf einem Seidenfilter ($270\ \mu\text{m}$ Maschengröße) wird die verbleibende Grünfärbung des Meeresswassers aus 10 m Tiefe ausgewertet und ein mit Farbtafeln definierter „Grünindex“ als diskrete Hilfsgröße stellvertretend für den Chlorophyllgehalt und die Menge an Phytoplankton verwendet. Zeitreihen monatlicher Mittelwerte über große Regionen der Nordsee von 1946-2001 zeigen sowohl eine deutliche Ausprägung einer Frühjahr- und einer Herbstblüte in der zentralen (C1, C2) und der südöstlichen Nordsee (D1, auch Deutsche Bucht), als auch eine Zunahme der Grünfärbung über die Jahre insgesamt (Edwards u. a. (2001), Abb. 3.1). Nur die südwestlichen Areale erscheinen uneinheitlich: weder ist die Blütenstruktur präsent, noch lässt sich eine eindeutige Zunahme im Grünindex feststellen.

3.2 Vertikalverteilung

Für die Beobachtbarkeit vom Satelliten stellt die Vertikalverteilung des Phytoplanktons eine wichtige Einflussgröße dar.

Während die Prozessoren bei der Bestimmung der Level 2-Produkte von einer homogenen Wassersäule ausgehen, findet sich in der Realität häufig ein Tiefenmaximum der Chlorophyllkonzentration. In den küstennahen Gebieten bilden sich diese Tiefenmaxima durch Absinken der Diatomeen nach der Frühjahrsblüte an der Halokline oder in der Bodenschicht, während sie in der zentralen und nördlichen Nordsee häufig mit der Pyknokline (häufig Thermokline) zusammenfallen. Entsprechende küstennahe *in situ* Beispiele sind dokumentiert in der Friesischen Front (Abb. 6 in Reid u. a. (1990)) oder entlang von Breitenkreisprofilen vor der Dänischen Küste (vgl. Bericht Gjertsen u. a. (2004)); ein Beispiel zur Konzentration des Phytoplanktons in der Thermokline beschreibt Sharples u. a. (2001).

Wird in einem Phytoplanktonmodell basierend auf kritischer Tiefe das Absinken, wie bei Diatomeen sehr verbreitet (Nielsen und Richardson (1989)), oder Prozesse wie Photoinhibition, selbstständiges Bewegungsvermögen bei einigen Arten oder Nährstofflimitierung berücksichtigt, so lässt sich das Tiefenmaximum auch theoretisch postulieren (Huisman (1999)).

Bei Tiefenprofilen von *in situ* bestimmter Chlorophyllkonzentration liegt der Oberflächenwert meist innerhalb der Standardabweichung um den Mittelwert der gesamten Wassersäule (Kommunikation Doerffer) und kann daher gut den mittleren Chlorophyllgehalt repräsentieren.

Solange das übergeschichtete Wasser geringe Absorption und Streuung aufweist, wird die Satellitenbeobachtung auch den Chlorophyllgehalt in größeren Tiefen bestimmen können. Eine genaue Studie der Abhängigkeit von Sichttiefe und abgeleitetem Algenprodukt z.B. auch unter Berücksichtigung der Tiefe der Sprungschicht, die sich mit dem Tidenstrom hebt und senkt, steht noch aus.

3.3 Artspezifische optische Eigenschaften

Allen in der Nordsee vertretenen Phytoplanktonarten ist gemein, dass sie eine oder mehrere Varianten des Pigmentes Chlorophyll enthalten. Damit wird die Bestimmung der Chlorophyllkonzentration mittels des Absorptionsspektrums zu einer praktikablen Möglichkeit, die Gesamtbiomasse des Phytoplanktons abzuschätzen. Dabei variieren die Absorptionsspektren der unterschiedlichen Chlorophylle deutlich und ihre Mischung ist letztlich in erster Linie für die Absorptionseigenschaften einer Phytoplanktonzelle verantwortlich.

Typische Absorptionsspektren (Abb. 3.2) - hier der Diatomee *Thalassiosira levandri* (grüne Linie, Absorptionsmaxima bei 438 und 674 nm) und der Haptophyceae *Phaeocystis globosa* (rote Linie, Absorptionsmaxima bei 438, 464 und 676 nm) - weisen zwei Maxima auf, die direkt auf die Kombination von Chlorophyllpigmenten (vgl. Tabelle 3.1) zurückgehen.

3 Phytoplankton in der Nordsee

Chlorophylltyp	Farbe	Absorptionsmaxima [nm]	Vorkommen
Chlorophyll <i>a</i>	blaugrün	430, 662	Cyanobakterien und alle phototrophen Eukaryoten
Chlorophyll <i>c</i>	grün	444, 576, 626	Braunalgen (Phaeophyta), Kieselalgen (Bacillariophyta), Goldalgen (Chrysophyta), Gelbgrüne Algen (Xanthophyta), Haptophyta, Dinophyta und Raphidophyceae
Chlorophyll <i>d</i>		447, 688	Rotalgen (Rhodophyta)

Tab. 3.1: Wellenlängen der Absorptionsmaxima der Chlorophylle und das Vorkommen dieser Pigmente

Weitere Pigmente sorgen für artenspezifisches Absorptionsverhalten, durch das Gruppen bei hoher spektraler Auflösung voneinander unterschieden werden können.

Einige Klassen lassen sich gesondert vom Satelliten erfassen durch ihre sehr speziellen optischen Eigenschaften. Die Kalkschale von Coccolithophoriden (wie *Emiliana huxleyi*) lässt Wasserflächen sehr hell erscheinen durch ihre starke Rückstreuung im gesamten sichtbaren Spektrum (weißer Streuer). *Phaeocystis globosa* in hohen Konzentrationen verleiht dem Wasser eine rötliche Färbung (“red tides”). Doch auch Versuche, Diatomeen anhand ihres Absorptionsverhaltens aus Satellitendaten zu erkennen, sind bereits erfolgreich durchgeführt worden (Sathyendranath u. a. (2004)). Jedoch ist selbst innerhalb einer Art ist mit Variation zu rechnen. Die Wachstumsbedingungen können Auswirkungen auf die inhärenten optischen Eigenschaften einer Art haben. Z.B. führt Lichtlimitierung bei *Thalassiosira pseudonana* zu einem Anstieg des Chlorophyllgehaltes (Stramski u. a. (2002)).

Die Artenzusammensetzung kann über die Eigenschaften der einzelnen Arten die optischen Parameter Streuung und Absorption entscheidend prägen (Ciotti (1999); Ciotti u. a. (1999); Cunningham u. a. (2003); Moore u. a. (2005); Stramski u. a. (2001)). Eine Anpassung des biooptischen Modells an die spezifischen in einer Region zu erwartenden Algen mit den ihnen eigenen inhärenten optischen Eigenschaften kann wesentlich zur Güte des Phytoplanktonproduktes beitragen.

Im Satellitenprodukt wird standardmäßig nicht nach verschiedenen Phytoplanktonklassen unterschieden, sondern lediglich die Gesamtkonzentration von Phytoplankton bzw. Chlorophyll über die Absorption der Pigmente abgeleitet. Es hat sich neben oben genannten extremen Beispielen aber bereits als möglich erwiesen, unterschiedliche Phytoplanktongruppen durch die Kombination des Chlorophyllproduktes mit abgeleiteten Streueigenschaften, d.h. durch Streukoeffizienten oder den daraus direkt abgeleiteten Anteil an suspendierten Partikeln (total suspended matter), voneinander zu trennen.

3 Phytoplankton in der Nordsee

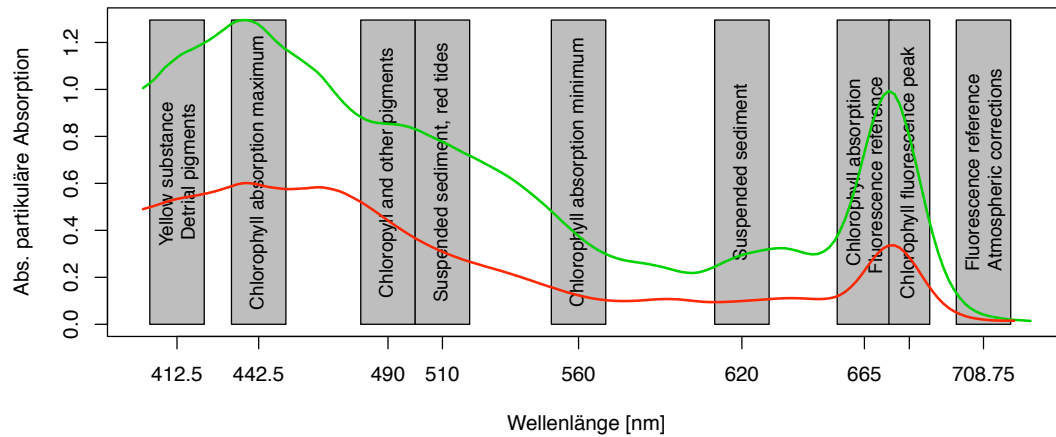


Abb. 3.2: Absorptionsspektrum der Diatomee *Thallasiosira levandri* (grüne Linie) und der Haptophyceae *Phaeocystis globosa* (rote Linie). MERIS Spektralbänder sind grau hinterlegt und ihre Funktion bei der Berechnung von Produkten vermerkt.

3.4 Das “Phytoplanktonjahr”

Frühjahrsblüten treten in der Norwegischen Rinne bereits Anfang Februar auf. Um diesen Strukturen Rechnung zu tragen, wird statt der üblichen Definition der Jahreszeiten hier im Folgenden ein „Phytoplanktonjahr“ definiert:

das *Frühjahr* erstreckt sich vom 1. Februar bis zum 14. Mai anstatt des Zeitraumes 15. Februar bis 14. Mai, als *Sommer* wird der Zeitraum 15. Mai bis 14. August, als *Herbst* der Zeitraum zwischen 15. August und 14. November bezeichnet. Durch das Vorziehen der Frühjahresgrenze erstreckt sich der *Winter* nur vom 15. November bis zum 31. Januar des Folgejahres.

Auch mit Rücksicht auf die tatsächlich vorhandenen Satellitendaten erscheint diese Einteilung sinnvoll, da vor dem 1. Februar aufgrund meist durchgängiger Bewölkung kaum mit Daten nördlich von 56° N zu rechnen ist.

Zusammenfassung

Dieser erste Teil gab einen kurzen Überblick, wie in der Satellitenozeanographie Chlorophyllprodukte abgeleitet werden. Stellvertretend standen die Daten und Prozessoren des Spektrometers MERIS im Vordergrund, da sie im weiteren Verlauf die Datenbasis aller qualitativen und quantitativen Untersuchungen bilden werden.

Für den Untersuchungsraum, die Nordsee, wurden die bekannten ozeanographischen und biologischen Prozesse dokumentiert. Dazu gehören die natürlichen Frühjahrsblüten des Phytoplankton genauso, wie das mit Einsetzen der Schichtung im Sommer entstehende Chlorophyllkonzentrationsminimum.

Durch die teilweise erheblichen Unterschiede in den Absorptionsspektren verschiedener Algengruppen ist es einerseits möglich, diese optisch zu trennen. Andererseits kann eine Anpassung der biooptischen Modelle an eine sich signifikant ändernde Artzusammensetzung notwendig werden. Dergleichen Überlegungen werden im Folgenden nicht weiter verfolgt.

Der Regionalprozessor *c2r* ist speziell auf die Gewässer und Algenvorkommen der Nordsee einschließlich ihrer Küstenregionen ausgerichtet. Er berücksichtigt in seinen modellierten inhärenten optischen Eigenschaften z.B. auch die Algengruppe der Coccolithophoriden (*b_{whit}*, weißer Streuer), die die Frühjahrsblüte in der Norwegischen Rinne dominieren. Besondere wiederkehrende Ereignisse in der Dynamik des Phytoplanktons wie die Frühjahrsblüte über der Doggerbank oder in der Norwegischen Rinne sollen im kommenden Teil anhand der verfügbaren Satellitendaten qualitativ belegt werden.

Je nach Fragestellung werden im Nachfolgenden die Daten jeweils auf die hier definierten Gebiete innerhalb der Nordsee beschränkt. Ich habe bei meinen eigenen Definitionen versucht, eine sinnvolle Kombination von möglichst großem, im Sinne der Wassermasse einheitlichen Raum mit möglichst hoher räumlicher und zeitlicher Überdeckung in Gebieten mit besonders interessanter Struktur oder Dynamik zu finden.

Teil II

Datenbearbeitung und Datenbasis

Einleitung

Auch in der Form eines Level 2-Produktes ist die weitere Aufbereitung der Satellitendaten für zeitliche oder räumliche Untersuchungen unumgänglich.

Durch die unterschiedlichen Beobachtungssituationen eines Gebietes durch einen Satelliten auf sonnensynchroner Umlaufbahn ist es notwendig, ein gemeinsames Gitter zu definieren, auf dem alle quantitativen Vergleiche und Berechnungen durchgeführt werden.

Jede Analyse dieses Datensatzes beginnt mit der eingehenden Betrachtung der verfügbaren Chlorophyllproduktkarten. Selbst wenn die Datenpunkte in ihnen nach den zuvor spezifizierten Qualitätskriterien ausgewählt werden, ist nicht auszuschließen, dass unerkannte Wolkenränder oder sehr hoch liegende Wolkenbedeckung zu fehlerhaften Chlorophyllwerten führen. Mehrere Methoden, wie diese entdeckt und aus dem Datensatz ausgeschlossen werden können, werden im Folgenden vorgestellt (Kapitel 4).

Betrachtet man die einzelnen Szenen im zeitlichen Ablauf, so fällt bei aufeinander folgenden Tagen auf, dass sich das Niveau der Chlorophyllkonzentration deutlich ändern kann, ohne dass ein dafür verantwortlicher, natürlicher Prozess benannt werden könnte. Zusammen mit den Ergebnissen der Frequenzanalyse wird den Hinweisen auf einen systematischen Fehler in den Level 2-Produkten, der mit den Winkeln des Sonnenstandes korreliert, nachgegangen und dieser Fehler ausführlich dokumentiert (Kapitel 5). Da alle Produkte aus dem Reflektanzspektrum oberhalb der Atmosphäre abgeleitet werden und nur zu etwa 10% des am Satelliten beobachteten Signals aus dem Wasser stammt, ist es unumgänglich, sich eingehender mit der faktisch möglichen Genauigkeit der berechneten Chlorophyllkonzentration auseinanderzusetzen. Ein relativ großer Fehler ist bereits nur durch diese fundamentale Abhängigkeit von der Güte der Atmosphärenkorrektur zu erwarten.

Abschließend sollen anhand einiger relativ kleinräumiger Beispiele bekannte wiederkehrende Algenblütenbildungen in der Nordsee dokumentiert werden, ohne eine Vervollständigung fehlender Datenpunkte versucht oder eine Korrektur um den systematischen Fehler vorgenommen zu haben. Neben der räumlichen und zeitlichen Dynamik, die zumindest qualitativ sehr gut durch die Satellitendaten abgebildet werden kann, werden diese Beispiele auch zur Illustration der Auswirkung des systematischen Fehlers herangezogen (Kapitel 6).

4 Datenbearbeitung

Die Datenbasis zur Untersuchung der zeitlichen und räumlichen Dynamik des Phytoplanktons in der Nordsee besteht aus den gültigen Datenpunkten der Satellitenszenen, die gemäß der Qualitätskriterien des Standardprozessors (Abschnitt 1.3.1) oder des Regionalprozessors (Abschnitt 1.3.2) ausgewählt werden. Die Szenen sind so gewählt, dass sie die Deutsche Bucht beinhalten oder streifen. Sie überdecken die Nordsee daher stärker im Südosten.

Für jeden Datenpunkt ist prinzipiell neben dem geographischen Ort auch der Zeitpunkt der Beobachtung und sämtliche Winkel zur Beschreibung der Geometrie von Sonnenstand und Satellitenposition in Relation zum Bildpunkt bekannt. Der Schwerpunkt wird auf den Level-2-Chlorophyllprodukten liegen, weitere Produkte sind aber verfügbar. Bevor geostatistische Untersuchungen folgen können, muss die Datenbasis einer Reihe von Bearbeitungsschritten unterzogen werden.

Insbesondere für den Regionalprozessor reicht die Sortierung per Qualitätskriterien nicht aus, um Wolkenränder bzw. -schatten, die sich als auffällig niedrige Chlorophyllkonzentrationen abbilden, auszuschließen. Einige allgemeine und spezielle Filter für diese Daten werden im Folgenden vorgestellt (Abschnitt 4.1 und Abschnitt 4.2).

Um Satellitenszenen verschiedener Tage quantitativ miteinander vergleichen zu können, ist es notwendig, ein gemeinsames Gitter zur Projektion der Datenpunkte festzulegen (Abschnitt 4.3).

4.1 Einfache Filter

4.1.1 Sigma-Filter

Jede normalverteilte Größe x , die durch Mittelwert m und Standardabweichung σ gut beschrieben ist, lässt sich sinnvoll von Ausreißern befreien, in dem man ihren Wertebereich auf das Intervall mit jeweils der Breite der dreifachen Standardabweichung unter- und oberhalb des Mittelwertes, $W = \{x | x \in [m - 3 \cdot \sigma, m + 3 \cdot \sigma]\}$, beschränkt.

4.1.2 Median-Filter

Räumliche Daten auf regelmäßigen Gittern können von hochfrequentem Rauschen befreit werden, indem statt des Originalbildpunktes der Median der Datenwerte in der Umgebung um den Punkt benutzt wird. Meist wird ein solcher Filter bei zugrunde liegendem Gitter durch gleitende, quadratische Fenster mit einer ungeraden Anzahl von Pixel als Kantenlänge verwendet, sodass der zentrale Punkt jeweils durch den Median aus dem

Fenster ersetzt werden kann. Ein Fenster der Größe 3×3 Pixel kann bereits zur deutlichen Glättung ausreichen.

4.2 Filter für unerkannte Wolkenränder

Bei derzeitigem Stand des Regionalprozessors ist die Wolkenerkennung noch nicht vollständig ausgereift, sodass an den Rändern von Wolken, durch Kondensstreifen oder ähnliche Bereiche mit hohen optischen Dicken der Aerosole gleichzeitig anomal niedrige Chlorophyllwerte berechnet werden. Diese Datenpunkte sind jedoch als vertrauenswürdig gekennzeichnet und lassen sich durch die einfachen Qualitätskriterien nicht aufspüren. Um artefaktfreie Komposite oder Analysen auf Einzelbildern zu erhalten, müssen diese Bildpunkte identifiziert und entfernt werden.

4.2.1 Identifizierung über optische Dicke der Aerosole

Aus der Kombination von zwei Level 2-Standardprodukten, der optischen Dicke der Aerosole bei 550 nm, τ_{550} , und der Chlorophyllkonzentration, wird jeweils ein Schwellwert für die optische Dicke abgeleitet, über dem kein Pixel mehr als vertrauenswürdig zu betrachten ist.

Die Güte dieses Filters wird nur an den Gebieten 1-5 (vgl. Abb. 2.4) getestet und kommt insbesondere vor der Autokorrelationsanalyse basierend auf mit Schätzwerten vervollständigten Datensätzen zur Anwendung (Kapitel 14).

Die Datenpunkte (*Lon, Lat, chl_conc, tau_550*) werden über τ_{550} vorgefiltert: alle Punkte, an denen die optische Dicke der Aerosole kleiner als der Mittelwert von $\log \tau_{550}$ abzüglich der dreifachen Standardabweichung σ ist, werden ausgeschlossen. Die optische Dicke der Aerosole erwies sich in den Beispieldaten eher lognormal-verteilt, sodass anhand des Logarithmus die statistischen Kenngrößen bestimmt und der Test durchgeführt wird.

Die verbleibenden τ_{550} -Werte werden in N gleich breiten Klassen zusammengefasst und ihre Anzahl festgestellt. Betrachtet wird nur der Teil der Häufigkeitsverteilung, der nach dem absoluten Maximum folgt. Im Falle nicht entdeckter Wolkenpixel wird die Häufigkeitsverteilung für hohe τ_{550} -Werte ein relatives Minimum nach dem absoluten Maximum zeigen.

Rein empirisch wird nun τ_{550} an diesem relativen Minimum oder bei 0.25 als obere Grenze festgesetzt. Ist gleichzeitig die Chlorophyllkonzentration kleiner als deren Mittelwert (der logarithmierten Konzentration), so werden diese Datenpunkte ausgeschlossen.

Dieses Verfahren liefert noch nicht optimale Ergebnisse, stellt jedoch eine große Verbesserung gegenüber den Rohdaten dar.

4.2.2 Identifizierung über die Gesamtstrahldichte

Die Summe der Gesamtstrahldichte (engl. path radiance) über die einzelnen spektralen Kanäle gibt Auskunft, wie viel Streulicht aus der Atmosphäre in jedem Pixel gesehen wird. Wasserpixel bei klarer Atmosphäre sind im Allgemeinen recht dunkel. Je höher die Gesamtstrahldichte, umso größer die Wahrscheinlichkeit, dass es sich bei einem Pixel

über Wasser um eine nicht erkannte Wolke handelt. Im nachfolgenden Beispiel werden die Pixel beibehalten, deren Gesamtstrahldichte kleiner als $0.3 \frac{1}{\text{sr}}$ ist. Allgemeingültigkeit ist für diese Schwelle nicht gegeben, sie resultiert nur aus der empirischen Beobachtung des einen nachfolgenden Beispiels (Abschnitt 4.2.3). Weitere Untersuchungen sind hier notwendig.

4.2.3 Beispiel zur Identifizierung von Wolkenrändern

Die Notwendigkeit des Filters vor der Berechnung von Schätzwerten soll an einem Beispiel illustriert werden. Die Schätzwertmethoden werden im Nachfolgenden genau vorgestellt werden (Kapitel 8 bis 12).

In der zentralen Nordsee östlich der Doggerbank (Gebiet 3) am 25. Juli 2006 enthält das Chlorophyllprodukt in der Nähe der als Wolken entdeckten Pixel (weiße Lücken) viele Datenpunkte mit so niedriger Chlorophyllkonzentration, dass diese unrealistisch im Kontext der Struktur erscheint (Abb. 4.2a). Werden alle Datenpunkte mit optischer Dicke der Aerosole $\tau_{550} > 0.25$ und Chlorophyllkonzentration $\text{chl_conc} < 1.042 \frac{\mu\text{g}}{\text{l}}$ (gemäß Abschnitt 4.2.1) aussortiert (Abb. 4.2b), so lässt sich der verbleibende Datensatz recht sinnvoll durch eine Interpolation mit dem Krigingverfahren vervollständigen (Abb. 4.1b).

In den Randbereichen von Wolken ist die Gesamtstrahldichte recht hoch (Abb. 4.2d). Schließt man alle Datenpunkte mit \sum Gesamtstrahldichte $_{\lambda} > 0.3 \frac{1}{\text{sr}}$ aus (Abb. 4.2e), so gelingt eine Rekonstruktion, die die vorhandenen Strukturen am sinnvollsten zu ergänzen scheint (Abb. 4.1c).

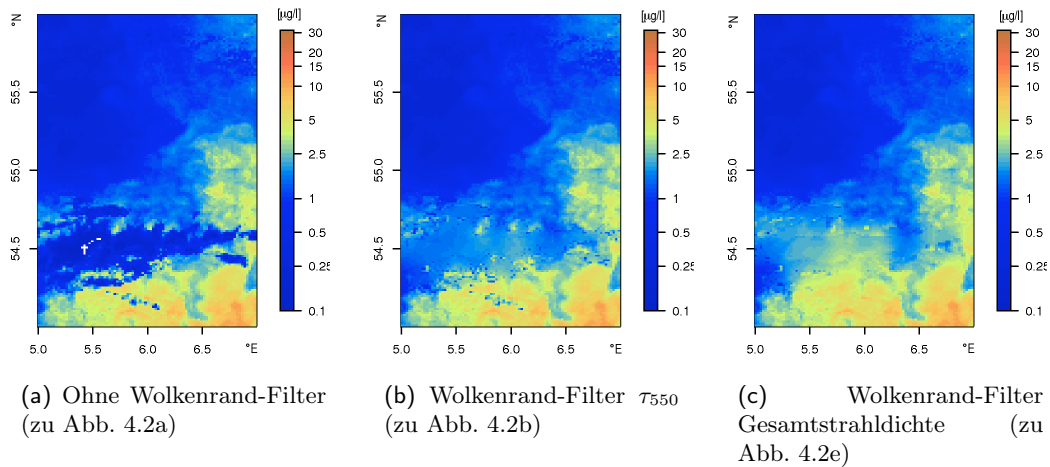


Abb. 4.1: Anwendung des Filters für Wolkenränder auf die Chlorophylldaten vom 25. Juni 2006 (vgl. Abb. 4.2) und die Auswirkung auf Ordinary Kriging Schätzwerte. Dargestellt ist jeweils der um Schätzwerte ergänzte Originaldatensatz.

4 Datenbearbeitung

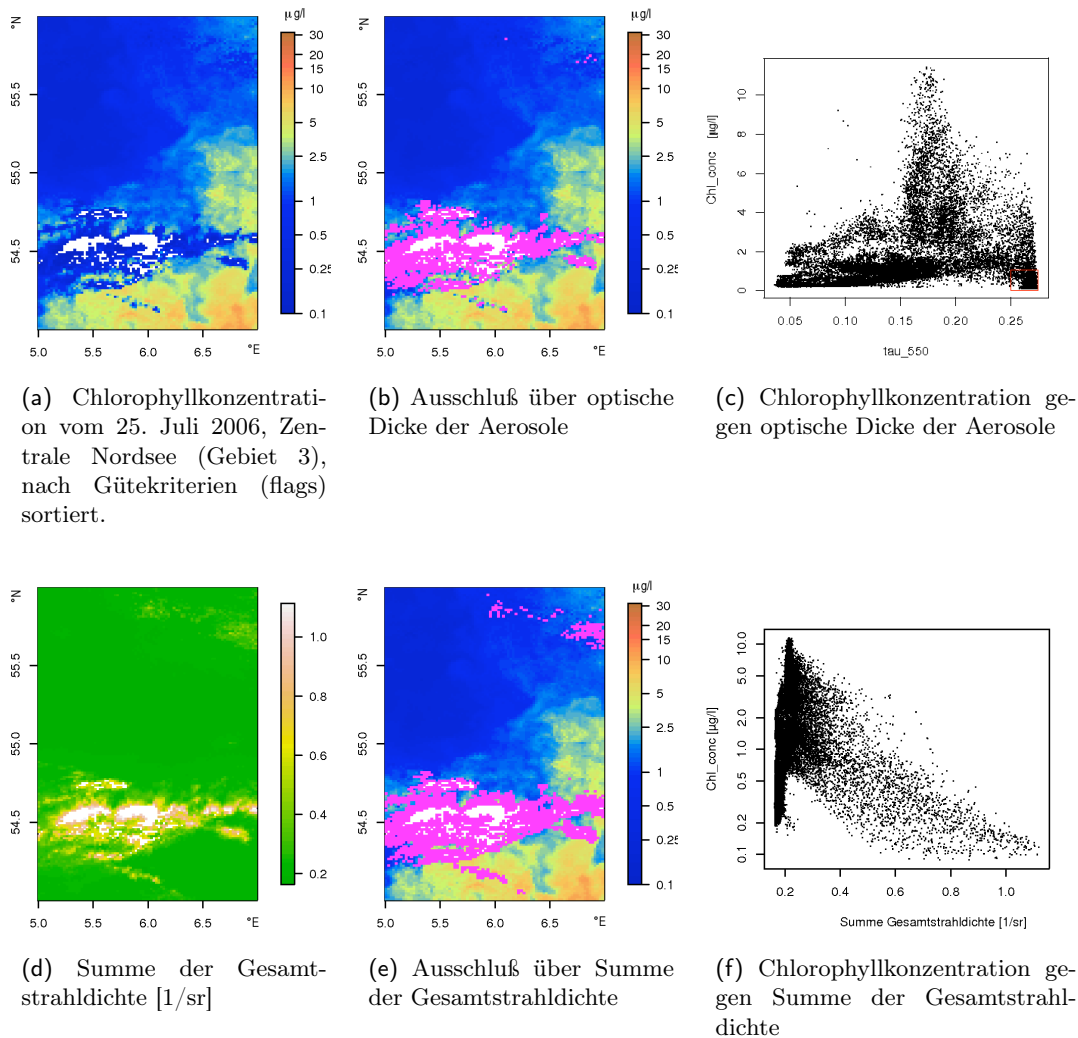


Abb. 4.2: Anwendung der Filter zur Identifizierung von Wolkenrändern für Chlorophyllwerte vom 25. Juli 2006 in der zentralen Nordsee, östl. der Doggerbank (Gebiet 3). Die durch die Filter identifizierten Pixel in Wolkenrandlagen sind jeweils in magenta markiert.

In nachfolgenden Kapiteln werden die Daten mittels der optischen Dicken der Aerosole aussortiert, da die Information der Gesamtstrahlendichte in den Textdateien mit gültigen Datenpunkten nicht mehr enthalten ist. Sollte ein anderer Filter zur Entdeckung der Wolkenränder benutzt worden sein, so wird dies vermerkt.

4.3 Standardgitter

Die Pixel zweier beliebiger MERIS-Satellitenbilder fallen im Allgemeinen nicht zusammen und erlauben damit nicht unmittelbar, die Daten für eine zeitliche Analyse zu nutzen.

Jede Methode, die die Auflösung der Daten in der Zeit berücksichtigt, muss daher ein gemeinsames Gitter definieren und die Datenpunkte auf ein solches projizieren.

Für die nachfolgenden Versuche wird auf eine besonders einfache Projektion zurückgegriffen, die lediglich ein rechteckiges Gitter in geographischer Länge und Breite heranzieht. Um eine lückenlose Projektion der Daten mit Kantenlängen der Pixel von 1200 m zu gewährleisten, wird die Anzahl der Pixel so gewählt, dass ihre mittlere Kantenlänge über das gesamte Gebiet 1.5 km beträgt. Dazu wird das Gitter zwischen 49° N bis 60° N in 815 Pixel und von 5° W bis 12° O in 730 Pixel unterteilt. Auf diesem Gitter gliedern sich die Datenpunkte in 290870 Landpixel und 304080 Wasserpixel.

Die meisten Karten in dieser Arbeit verzichten auf die Achsenbeschriftung von Länge und Breite. Die Angaben der geographischen Länge erfolgen immer bezogen auf die Ostrichtung, d.h. 5° W wird als -5° O geschrieben.

Bei allen folgenden Analysen wird das hier definierte Gitter verwendet, jedoch werden teilweise nur einzelne Unterräume berücksichtigt. In welchem dieser Unterräume (Definition in Abschnitt 2.2) die spezifische Betrachtung stattfindet, wird jeweils dokumentiert.

4.4 Zeitangaben

Tag des Jahres In vielen nachfolgenden Bildern wird der “Tag des Jahres” als zeitliche Größe verwendet. Diese Zählung beginnt jeweils am ersten Januar mit Tag 1 und endet am 31. Dezember mit Tag 365 bzw. 366. Die Schaltjahre werden dabei berücksichtigt, um eine korrekte zeitliche Analyse zu ermöglichen.

In einem früheren Stadium der Arbeit sind die Schaltjahre 2004 bzw. 2008 nicht beachtet worden, sodass der 29. Februar genauso als Tag 60 gezählt wird wie der erste März. Im Jahr 2004 liegen für den 29. Februar keine Daten vor. Im Jahr 2008 gibt es sowohl Daten für den 29. Februar als auch für den ersten März. Bezogen auf das Standardgitter überlappen sich hier 1845 Pixel, bei rein räumlichen Analysen auf kleineren Regionen (in Frage käme nur Gebiet 5) gibt es keine Überlappung der Daten, sodass keine Effekte durch den zweimal auftretenden Tag 60 zu erwarten sind. Natürlich sorgt das Auslassen dieses Tages in der Zählung bei der Analyse zeitlicher Abhängigkeiten, z.B. von der Beobachtungsgeometrie, für Unstimmigkeiten.

Uhrzeit Alle Angaben von Uhrzeiten mit den Satellitenüberflügen erfolgen in UTC (engl. Universal Time Coordinated), d.h. sie beziehen sich auf die lokale Ortszeit am Nullmeridian. In Deutschland findet die Mitteleuropäische Zeit (UTC+1) bzw. die Mitteleuropäische Sommerzeit (UTC+2) Verwendung.

5 Kritische Bewertung der Datenbasis

Bevor tiefer gehende Analysen oder Methoden zur Vervollständigung einzelner Datentage vorgestellt werden, soll der Datensatz aus gültigen Chlorophyllkonzentrationen genauer charakterisiert werden.

Die Datentage vom 1. Januar 2003 bis zum 8. Juli 2008 werden sowohl für den Standard- als auch Regionalprozessor gemäß der Qualitätskriterien des jeweiligen Prozessors auf die gültigen Datenpunkte über Wasser reduziert (vgl. Abschnitt 1.3.1 bzw. Abschnitt 1.3.2) und im Produkt des Regionalprozessors die unerkannten Wolkenränder entfernt (Abschnitt 4.2.1).

Neben allgemeinen Eigenschaften wie der statistischen Verteilung der Chlorophyllkonzentrationswerte (Abschnitt 5.1) und dem Konzept der Separabilität (Abschnitt 5.2) wird auch die räumliche und zeitliche Datenüberdeckung beschrieben (Abschnitt 5.3). Die Frequenzanalyse der (vereinfachten) Zeitreihen (Abschnitt 5.4) liefert Hinweise auf den systematischen Fehler, der im Chlorophyllprodukt zu finden ist, und der im nachfolgenden Abschnitt (Abschnitt 5.5) genauer untersucht wird.

5.1 Normalverteiltheit

Betrachtet man die Verteilung der Chlorophyllkonzentrationswerte aller Datenpunkten eines Datentages, folgt sie im Allgemeinen eher einer Log-Normalverteilung.

Viele der nachfolgenden mathematischen Verfahren basieren auf Gauss-Zufallsfunktionen oder anderweitig auf der Annahme, dass der Mittelwert und die Standardabweichung aussagekräftige Parameter zur Beschreibung der Datenmenge sind, sodass in diesen Fällen der Logarithmus der Konzentration verwendet werden wird. Des Weiteren besitzt dieser den Vorteil, die Beschränkung des Wertebereiches von $\mathbb{R}^{>0}$ auf \mathbb{R} aufzuheben.

Im weiteren Verlauf wird bei der Mittelwertberechnung explizit angegeben, ob mit logarithmierten oder tatsächlichen Konzentrationen gerechnet wird.

5.2 Separabilität

Mit Separabilität bezeichnet man das Vermögen, zeitliche und räumliche Prozesse einer Zufallsfunktion $Z(s, t)$, mit räumlichen Koordinaten $s \in \mathbb{R}^d$ und zeitlichen Koordinaten $t \in \mathbb{R}$, getrennt voneinander als eigenständige Kovarianzfunktionen zu betrachten, sodass die Kovarianz des Gesamtprozesses gerade dem Produkt der Einzelprozesse entspricht (Gneiting u. a. (2005)):

$$\text{cov} \{Z(s_1, t_1), Z(s_2, t_2)\} = \text{cov}_S(s_1, s_2) \cdot \text{cov}_T(t_1, t_2)$$

Mögliche rein räumliche Prozesse sind z.B. Einflüsse der Bathymetrie. Als ein rein zeitlicher Prozess kann die Einstrahlung gelten, wobei jedoch durch die Wolkenbedeckung räumliche Aspekte ins Spiel kommen können. Generell können räumliche bzw. zeitliche Prozesse eine Richtungsabhängigkeit aufweisen, sei es Anisotropie im räumlichen und ein Entwicklungstrend im zeitlichen Bereich. Sie unterscheiden sich fundamental in ihren Eigenschaften: der zeitliche Prozess ist immer gerichtet, da lediglich Zustände in der Vergangenheit die zukünftigen beeinflussen können.

Innerhalb des komplexen Systems, in das die Entwicklung und Beobachtung der Chlorophyllkonzentration als Maß für das Algenaufkommen eingebunden ist, werden die meisten Prozesse nur in Näherung und selten eindeutig in räumliche und zeitliche voneinander zu trennen sein. Viele Prozesse lassen sich nicht voneinander entkoppeln. So besitzt z.B. der Eintrag von Nährstoffen eine deutliche räumliche Komponente, wenn man nur den Beitrag aus Flüssen betrachtet. Dieser ist gleichzeitig von einem Jahresgang überprägt. Gleiches gilt für den Einfluss von Durchmischung und Schichtung, bei denen neben dem Wärmehaushalt durch Einstrahlung ebenso der Wind und die Tiden eine Rolle spielen.

Eine Überprüfung des Datensatzes auf Separabilität wird nicht durchgeführt, auch wenn dergleichen Methoden (Mitchell (2002); Mitchell u. a. (2005)) vorhanden sind, da bereits der Verdacht des später in diesem Kapitel diskutierten systematischen Fehlers bestanden hat. Der Versuch hätte damit von vornherein ohne belastbares Ergebnis enden müssen.

Entsprechend sind in allen späteren Verfahren, die nur räumliche oder nur zeitliche Aussagen treffen, die Effekte weiterhin vermischt und eine ursächliche Interpretation kann schon durch diesen Umstand mit Fehlern behaftet sein.

5.3 Überdeckungsstatistik

Jeder Datenpunkt einer Szene über Wasser lässt sich in eine der beiden Kategorien einteilen: es handelt sich um einen gültigen oder um einen ungültigen Datenpunkt, wobei an letzterem das Chlorophyllprodukt aus unterschiedlichsten Gründen nicht bestimmt werden konnte. Die Wolkenpixel sind eine Teilmenge der ungültigen Datenpunkte und somit nicht (unbedingt) komplementär zu den gültigen Datenpunkten.

Durch die diskrete Natur der Satellitenüberflüge kann jeder relative Orbit nur einen festen Anteil der Wasserfläche der Nordsee (als zusammenhängendes Gebiet, Abb. 2.3) sehen, der zwischen 36 und 99% liegt (Tabelle 5.1).

5.3.1 Datenüberdeckung

Der Überdeckungsgrad der Chlorophylldaten (Abb. 5.1) wird hier unabhängig von der tatsächlichen Abdeckung der Nordsee durch die jeweilige Satellitenszene immer auf die Anzahl der Wasserpixel im Nordseegebiet (vgl. Abb. 2.3) bezogen. Die Markierungen zeigen die Einteilung in die Jahreszeiten des Phytoplanktonjahres.

Unterteilt man die beobachteten Überdeckungen in Klassen von jeweils 1% Breite (von 0.5-1.5%, usw.), so folgen diese einem exponentiellen Abfall (Abb. 5.2): häufig treten

5 Kritische Bewertung der Datenbasis

<i>Orbit-Nr.</i>	Anteil [%]	<i>Orbit-Nr.</i>	Anteil [%]	<i>Orbit-Nr.</i>	Anteil [%]	<i>Orbit-Nr.</i>	Anteil [%]	<i>Orbit-Nr.</i>	Anteil [%]
8	85	108	36	194	68	309	84	409	99
37	89	123	63	223	95	323	97	423	62
51	95	137	99	237	80	352	71	452	97
80	79	151	54	266	91	366	99	466	75
94	99	180	98	280	91	380	45	495	95

Tab. 5.1: Anteil der Wasserpixel an der zusammenhängenden Nordseefläche (in %) je Orbit

Datentage mit sehr geringen Überdeckungen auf, während hohe Grade selten vorkommen. Vollständige Überdeckung ist im gesamten Datensatz nicht zu finden. In einigen wenigen Beispielen werden mehr als 80% der Wasseroberfläche gesehen.

Auf die Klasse von 0-0.5% wird in der Darstellung zugunsten ihrer Übersichtlichkeit verzichtet. Im gesamten Zeitraum entfallen allerdings 211 Datentage in diese Kategorie, davon 62 im Frühjahr, 39 im Sommer und 40 im Herbst. Der Winter besitzt fast ausschließlich Überdeckungen von weniger als 10%.

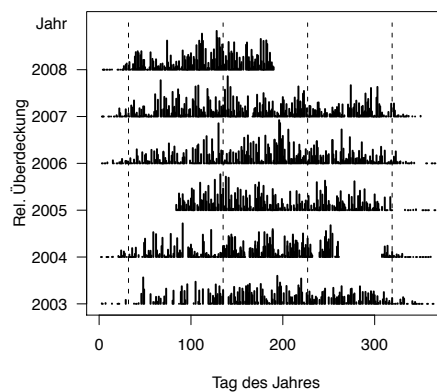


Abb. 5.1: Relative Überdeckung des Nordseegebietes mit gültigen Chlorophylldaten. Dabei entspricht die Länge eines Balken zwischen zwei aufeinander folgenden Jahren gerade 100% Überdeckung. Ohne Berücksichtigung der Schaltjahre sind die Jahreszeiten als gestrichelte Linien markiert. Jahre 2003 - 2008.

5.3.2 Wolkenüberdeckung

Der hier präsentierten Statistik der Wolkenbedeckung liegt die Aufteilung der Nordsee in Gebiete wie im Ökosystemmodell ERSEM definiert zugrunde. Diese Einteilung wird hier als ICES-Regionen bezeichnet (vgl. Abschnitt 2.1.1 und Abb. 2.2). Der Anteil der Wolkenbedeckung wird jeweils in einer solchen Region und für einen Monat über alle Jahre von 2003 bis 2007 betrachtet.

5 Kritische Bewertung der Datenbasis

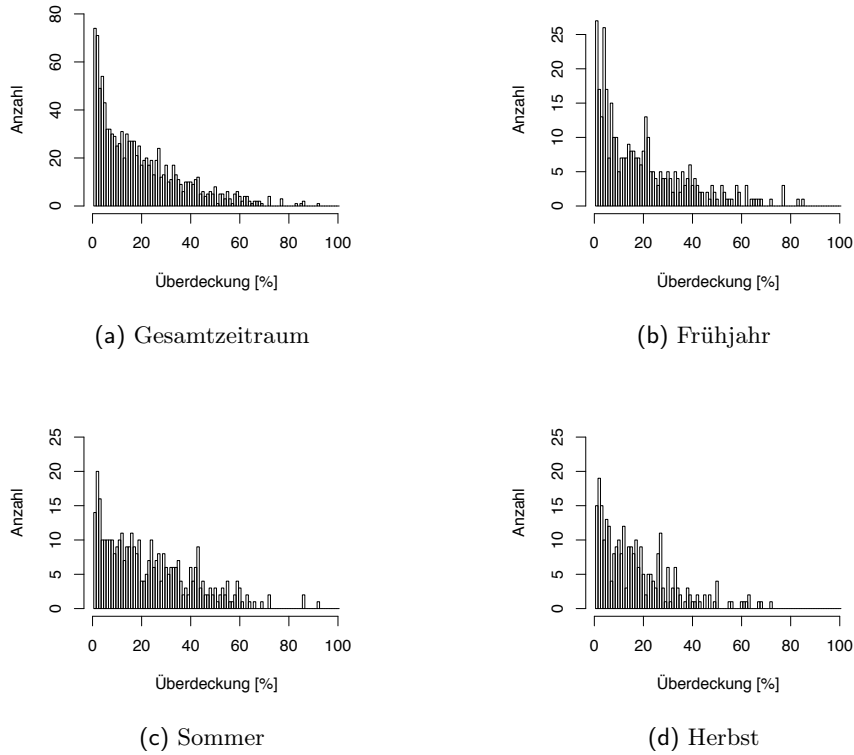


Abb. 5.2: Häufigkeit von Datentagen eines Überdeckungsgrades in Intervallen von 1%-Breite im Gesamtzeitraum von Januar 2003 bis August 2008 für das Chlorophyllprodukt des Regionalprozessors. In die nicht dargestellte Klasse von 0-0.5 % fallen im (a) Gesamtzeitraum 211, im (b) Frühjahr 62, im (c) Sommer 39, im (d) Herbst 40 Datentage.

Die Gesamtzahl von Pixel A , die innerhalb einer Szene - repräsentiert durch den Zeitpunkt (Tag) t - in einem Gebiet gesehen werden, setzt sich aus der Anzahl der gültigen Datenpixel D und den ungültigen Datenpixel zusammen. Die Anzahl gesehener Wolkenpixel sei W . Jedes Gebiet ("area") besteht bezogen auf das Standardgitter (vgl. Abschnitt 4.3) aus einer festen Anzahl von Gitterpunkten $P_{max}(\text{area})$. Da abhängig von der Lage des Orbits nicht jede Szene jedes Gebiet vollständig überdeckt, beträgt die maximal mögliche Überdeckung A_{max} zum Zeitpunkt t also gerade

$$A_{max}(t) = A(t) / P_{max}(\text{area}).$$

Je größer der Anteil gesehener Pixel ist, umso größer soll auch das Gewicht $u(t)$ für eine gewichtete Mittelung des Wolkenanteils sein. Die Gewichte werden hier definiert als

$$u(t) = \frac{A_{max}^2(t)}{\sum_{t \in \text{Monat}} A_{max}^2(t)}$$

5 Kritische Bewertung der Datenbasis

Dabei werden alle maximalen Überdeckungen, die in einem Monat unabhängig vom konkreten Jahr beobachtet wurden, zusammengefasst.

Der mittlere Wolkenbedeckungsgrad \bar{W} beträgt somit

$$\bar{W} = \sum_{t \in \text{Monat}} u(t) \cdot \frac{W(t)}{A(t)}$$

mit der mittleren Abweichung bei insgesamt N Datentagen von

$$\bar{W}_{sd} = \sqrt{\frac{1}{N} \sum_{t \in \text{Monat}} \left(u(t) \frac{W(t)}{A(t)} - \bar{W} \right)^2}$$

Der Juli und der August sind die Monate mit dem geringsten Wolkenbedeckung, doch selbst dann liegt sie noch um 20% (Abb. 5.3). Die Deutsche Bucht (area 5a) ist im Mittel außer in den Monaten April, Juli und August zu mehr als 60% unter Wolken. Außer der Tendenz von stärkerer Bewölkung im Winter als im Sommer lässt sich kein einheitlicher Jahresgang beobachten. Auch regionale Unterschiede treten nicht augenscheinlich hervor. Leicht verfälscht ist diese Statistik noch für die Wintermonate, da hier in den nördlichen Regionen fast ausschließlich eine geschlossene Wolkendecke vorliegt und diese Szenen mangels gültiger Datenpunkte im Datenbankarchiv nicht abgelegt sind und somit nicht berücksichtigt werden.

5.4 Frequenzanalyse unregelmäßiger Zeitreihen

Die Überdeckungsstatistik für den Datensatz - sei sie bezogen auf die gültigen Datenpunkte (Abschnitt 5.3.1) oder die Wolkenpixel (Abschnitt 5.3.2) - legt nahe, dass selbst bei Reduzierung der Ausdehnung des Untersuchungsgebietes keine vollständige Zeitreihe aufgestellt werden kann. Im klassischen Sinne heißt vollständig, dass eine zeitlich äquidistante Zeitserie des Chlorophyllsignals existiert. Dies erlaubte eine Frequenzanalyse per Fouriertransformation. Bereits die diskrete Natur der Überflüge in unterschiedlichen Orbits und zu verschiedenen Uhrzeiten für einen zu beobachtenden Punkt verhindert, dass der unbearbeitete Datensatz eine solche Zeitreihe ergeben kann.

Um die volle zeitliche Auflösung nutzen zu können, wird ein Lomb-Scargle-Periodogramm (Lomb (1975); Scargle (1989)), d.i. eine diskrete Fouriertransformation für unregelmäßige Zeitreihen, verwendet. Anhand von Periodogrammen, d.i. die spektrale Leistungsdichte in Abhängigkeit von der (diskreten) Frequenz, lassen sich die charakteristischen zeitlichen Signalanteile in den Zeitreihen identifizieren. In einem Frequenzintervall, das durch minimale und maximale Periodenlänge definiert wird, werden die diskreten Frequenzen gleichverteilt angesetzt, sodass im Zeitbereich die langen Perioden schlechter aufgelöst sind als die kurzen. Die (rel.) Maxima der spektralen Leistungsdichte können auf Signifikanz überprüft werden.

Die maximal ohne Informationsverlust aufzulösende Frequenz (Nyquistfrequenz) bei einer Abtastrate von f_0 beträgt $f_{Nyquist} = \frac{1}{2}f_0$. Ist die Abtastrate zu gering, so werden Signale mit Frequenzen oberhalb der Nyquistfrequenz fälschlich als solche mit geringerer

5 Kritische Bewertung der Datenbasis

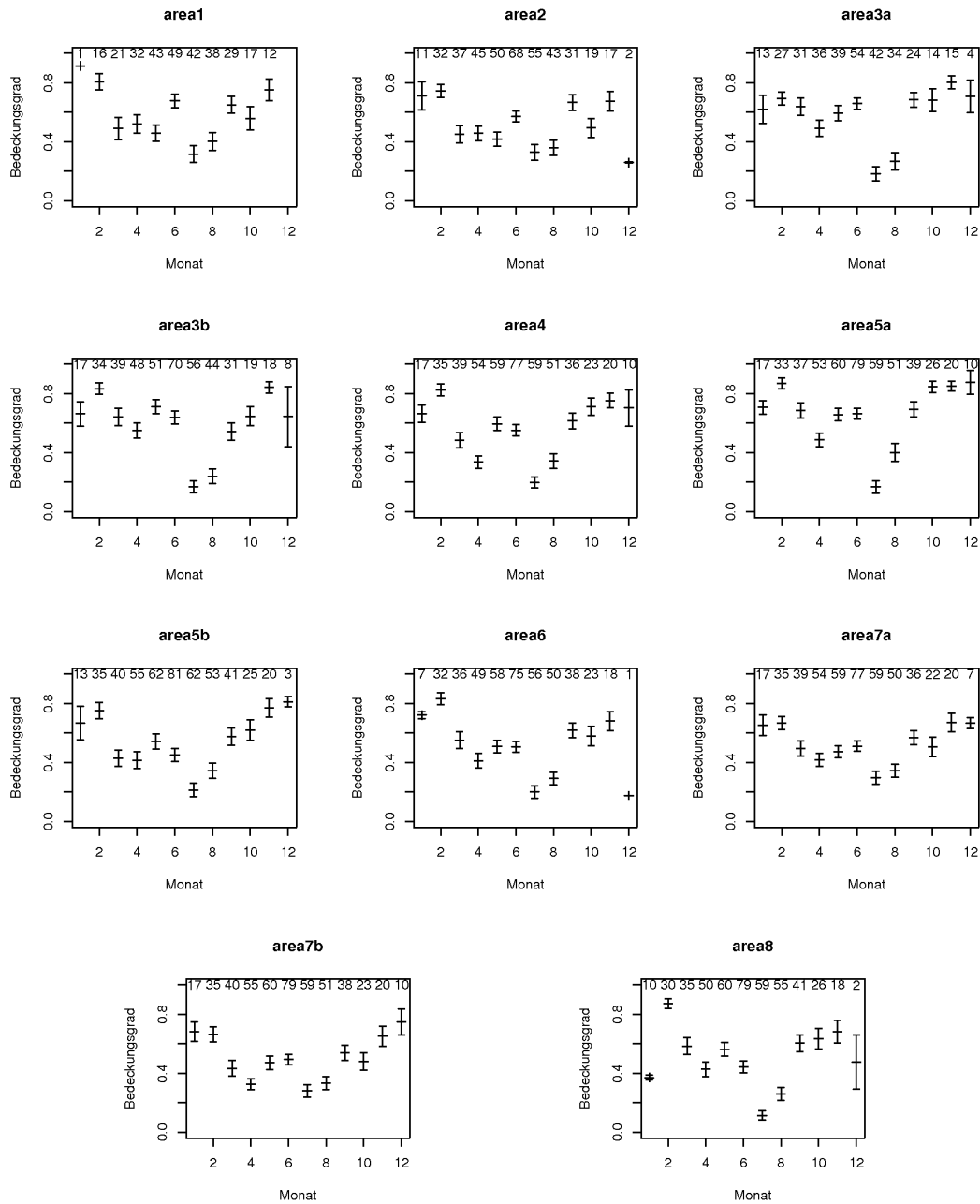


Abb. 5.3: Mittlere gewichtete Wolkenbedeckung jeweils einer ICES-Regionen, pro Monat und über den Zeitraum von 2003 bis 2007. Die Zahlen im oberen Bereich des Diagramms geben die Anzahl eingeflossener Datentage pro Monat an. Die Lage der Regionen 1-8 entnehme man Abb. 2.2.

Frequenz erkannt (Aliasing). Bei unregelmäßigen Zeitreihen ohne konstante Abtastrate kann die Nyquistfrequenz anhand der Statistik der Differenz zwischen benachbarten zeitlichen Stützstellen abgeschätzt werden. Da in der Chlorophyllzeitreihe nur wenige relativ große Zeitdifferenzen durch die Datenlücken im Winter auftreten, werden diese aus der Abschätzung ausgeschlossen. Über den Median der zeitlichen Differenzen $\Delta t_i = t_{i+1} - t_i$ von aufeinander folgenden Zeitpunkten t_i und t_{i+1} sei die Nyquistfrequenz hier definiert als:

$$f_{\text{Nyquist}} = \frac{1}{2} \cdot (\text{Median}(\Delta t_i))^{-1} \quad (5.1)$$

5.4.1 Zeitreihen punktuell zusammengefasster Chlorophyllkonzentrationen

Als Zeitreihe wird jeweils der Mittelwert der logarithmierten Chlorophyllkonzentration in einem Gebiet wie der Doggerbank (Gebiet 1), dem Englischen Kanal (Gebiet 2), der zentralen Nordsee östlich der Doggerbank (Gebiet 3), der Deutschen Bucht (Gebiet 4) und der Dänischen Küste (Gebiet 5) (vgl. Abschnitt 2.2.3) genutzt, wenn mehr als 10 gültige Pixel für den Beobachtungszeitpunkt in das Gebiet fallen. Die Punkte der Zeitreihe gehen weder auf konstante Überdeckungen noch auf eine konstante Anzahl gesehener Pixel zurück und haben somit unterschiedliche statistische Aussagekraft. Als Beobachtungszeitpunkt wird jeweils der Tag und die Uhrzeit am Mittelpunkt des Gebietes gewählt, ebenso bezieht sich die Beobachtungsgeometrie mit Satellitenposition und Sonnenstand auf diesen Punkt.

Diese fünf Gebiete sind einerseits relativ klein gewählt, sodass starke räumliche Strukturen keine Auswirkung haben und die Mittelwerte die Chlorophyllkonzentration des Gebietes repräsentieren können, andererseits groß genug, um Effekte aus Strömungstransport oder punktueller Wolkenbedeckung zu minimieren. Ist das Gebiet zu groß gewählt bzw. sein Verhalten erfahrungsgemäß räumlich sehr inhomogen, so können partielle Wolkenüberdeckungen zu massiven Unterschieden im Mittelwert führen. Für die Deutsche Bucht kann diese Überlegung in der Interpretation nicht vernachlässigt werden.

Die Chlorophyllkonzentrationen variieren teilweise sehr stark von Tag zu Tag bzw. zwischen aufeinander folgenden zeitlichen Stützstellen (Abb. 5.4). Generell treten um den Tag 100 (10. April) herum lokale oder absolute Maxima im Jahresverlauf auf. Dieses Signal, das der zu erwartenden Frühjahrsblüte zugeordnet werden kann, ist in Gebiet 2 (Englischer Kanal) besonders deutlich ausgeprägt. Dabei variiert der Zeitpunkt und -raum der Frühjahrsblüte stark: über der Doggerbank tritt sie deutlich früher auf als im Englischen Kanal oder an der Dänischen Küste. Teilweise sind auch zwei Chlorophyllmaxima im Frühjahr sichtbar, die sich mit dem Übergang von vorherrschenden Diatomeen zu nachfolgenden Flagellaten erklären lassen (z.B. Englischer Kanal im Frühjahr 2003). Die Variabilität zwischen den Jahren ist im Frühjahr und Herbst am ausgeprägtesten, während die Chlorophyllkonzentration im Sommer recht stabil und gering über die Jahre ist. Das gilt nicht für die Deutsche Bucht, in der die Phytoplanktonkonzentration über den gesamten Jahreslauf hohe Variabilität zeigt.

Modelluntersuchungen bezüglich der interannualen Variabilität bzw. der zeitlichen Dynamik, mit denen sich genauere Aussagen über die Ursachen der Chlorophyllkonzentra-

5 Kritische Bewertung der Datenbasis

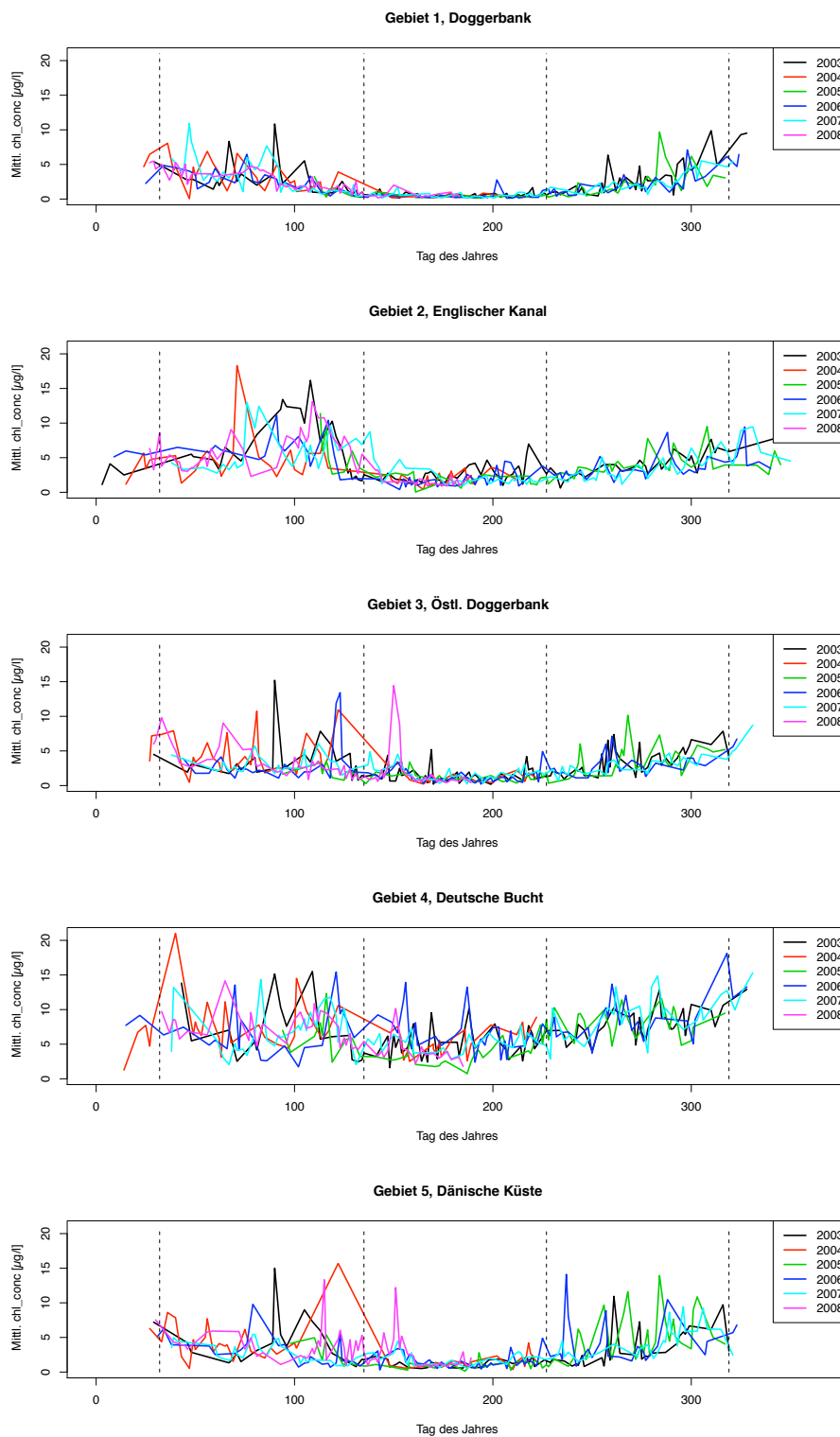


Abb. 5.4: Mittlere Chlorophyllkonzentration im Jahresgang für die Gebiete 1-5 abgeleitet aus dem Chlorophyllprodukt des Regionalprozessor in den Jahren 2003-2008. Die vertikalen Einteilungen markieren die Jahreszeiten im Phytoplanktonjahr. Die Überdeckungsstatistik zu diesen Zeitpunkten findet sich in Abb. 14.11.

5 Kritische Bewertung der Datenbasis

tionsschwankungen treffen lassen, sind jedoch kein Bestandteil der vorliegenden Arbeit.

In den Zeitreihen bildet sich eine deutliche Tendenz zu hohen Chlorophyllwerten im Winter ab, anders als Erfahrung und *in situ*-Beobachtung es erwarten lassen. Hierfür können einerseits die niedrigen Sonnenstände im Winter verantwortlich sein, die die Annahme einer planaren Struktur der Atmosphäre bei der Atmosphärenkorrektur kaum erfüllen, oder die starke Durchsetzung des Wassers mit Phaeophytin haltigem Detritus. Phaeophytin besitzt ein den Chlorophyllen sehr ähnliches Absorptionsspektrum und würde als Chlorophyll von den Prozessoralgorithmen interpretiert werden (Kommunikation Doerffer).

Andererseits lässt dieser ausgeprägte Jahresgang mit Maximum im Winter vermuten, daß es eine Beziehung zu den Sonnenstandswinkeln, die den gleichen Verlauf zeigen, gibt. Dabei geht es nicht um den Einfluss des Sonnenstandes auf die biologischen Prozesse im Wasser, sondern um den Einfluss der Geometrie auf die Berechnung des Chlorophyllproduktes. Diesem Verdacht auf einen systematischen Fehler wird der Abschnitt 5.5 gewidmet sein.

<i>Frühjahrsblüte</i>	Doggerbank	Engl. Kanal	Zentral, östl. Doggerbank	Deut. Bucht	Dän. Küste
2003	67, 90	94, 108	90, 113	90, 109	90
2004	71	71	66, 81, 106, 122	56,65,101,122,116	56, 122
2005	-	113	-	-	-
2006	76	91, 117	123	70, 121	79
2007	47, 86	76-82, 118	80, 112	83, 118, 125	80
2008	78	68, 109	64	65, 110	59, 72, 115

Tab. 5.2: Zeitpunkte (Tag des Jahres) der Frühjahrsblüte als Maxima der mittleren Chlorophyllkonzentration in den Gebieten 1-5.

<i>Herbstblüte</i>	Doggerbank	Engl. Kanal	Zentral, östl. Doggerbank	Deut. Bucht	Dän. Küste
2003	258, 293	(218) -	261	(169, 188) 261, 284	261
2005	284	278, 308	244, 268	(231) 266, 285	256, 284
2006	298	288	(225) 260	(156, 187) 266	(237) 257, 288
2007	305	-	-	(230) 262, 283	(235) 289
2008	-	-	(150)	(151)	(151)

Tab. 5.3: Zeitpunkte (Tag des Jahres) der Herbstblüte als Maxima der mittleren Chlorophyllkonzentration in den Gebieten 1-5. In Klammern sind Phytoplanktonblüten im Sommer vermerkt.

Eine weitere Auffälligkeit ist der regelmäßige Versatz der Chlorophyllkonzentration (Abb. 5.5). Über der Doggerbank (Gebiet 1) erscheint er besonders deutlich zwischen Tag 193 bis Tag 200 (12. - 19. Juli 2006, vgl. Abb. 5.5a). Für den vergleichbaren Zeitraum tritt vor der Dänischen Küste (Gebiet 5) ein dreitägiger Zyklus über sechs Tage im Chlorophyllmittelwert auf (Tag 196 - 201, 15. - 20. Juli 2006, vgl. Abb. 5.5b).

Alle äußeren, physikalischen Einflüsse, die das System derart verändern können, lassen sich ausschließen:

Gerade in diesen Sommertagen ist bei stabiler Schichtung - die im Bereich der Doggerbank (Gebiet 1) zu erwarten ist, vgl. Abschnitt 2.3.2 bzw. Abschnitt 6.3.2 - nur mit geringer Variabilität zu rechnen; von Tag zu Tag kann zwar bei einem extremen Windereignis Durchmischung hervorgerufen werden und die beobachtbare Chlorophyllkonzentration sich signifikant ändern, doch dass dergleichen regelmäßig stattfindet, ist unwahrscheinlich.

Auch der Einfluss durch einen hohen Wolkenbedeckungsgrad, der zu einem nicht repräsentativen Mittelwert der verbleibenden Datenpunkt für das Gebiet führen kann, lässt sich für die hier gezeigten Beispiele ausschließen (vgl. Pixelanzahl zu jedem Datentag in Abb. 5.5).

Zudem scheidet bei der Dimension der Gebiete (Gebiet 1: 127 km \times 164.7 km, Gebiet 5: 121.5 km \times 110.8 km) der vollständige Austausch des Wasserkörpers durch Strömungstransport aus. In Gebiet 1 sind nur geringe, über die Tage umlaufende Strömungsgeschwindigkeiten von etwa 4 cm/s dokumentiert. In Gebiet 5 strömt zwar von Tag 194 bis 196 Wasser aus dem Skagerrak mit maximaler mittlerer Geschwindigkeit von 30 cm/s ein, doch selbst diese konstante Geschwindigkeit würde nur einen Transport innerhalb von 24 h um 25.9 km nach sich ziehen und damit höchstens ein Viertel des beobachtbaren Wasserkörpers austauschen können.¹ Trotzdem können bei diesem Beispiel durch die relativ starke Verlagerung einer Front Einflüsse auf den veranschlagten Mittelwert auftreten, die eine Interpretation erschweren.

Da ozeanographische und biologische Ursachen ausgeschlossen werden können, kommt zuvorderst die Beobachtungsgeometrie in Betracht. In den Szenen zweier aufeinander folgender Tage liegt der gleiche geographische Ort aufgrund der wechselnden Orbits entweder in der östlichen oder westlichen Bildhälfte. Alle weiteren Untersuchungen an der Zeitreihe (Abschnitt 5.4.2 und Abschnitt 5.4.3) zielen darauf ab, diese Auffälligkeit zu charakterisieren. Die Fehlerbetrachtung zum Einfluss der Sonnenstandswinkel findet im Detail in Abschnitt 5.5 ihren Abschluss.

5.4.2 Frequenzanalyse unregelmäßiger Zeitreihen

Zum Vergleich des Frequenzinhaltes der Chlorophyllkonzentration wird die Zeitreihe zur Doggerbank (Gebiet 1) über den gesamten Zeitraum von 2003 bis 2008 für die unterschiedlichen Chlorophyllprodukte des Standard- und des Regionalprozessors untersucht.

Die Periodogramme werden, um eine gute Auflösung der Frequenz zu gewährleisten, mehrfach in verschiedenen Periodenintervallen berechnet: um das Verhältnis der Stärke des 3-Tages-Signals zum Jahressgang zu demonstrieren, wird ein Intervall zwischen 2 und 450 Tagen gewählt (Abb. 5.6 (a)-(c)). Da in diesem Fall die Auflösung für die langen Perioden nicht ausreicht, um die Periode des Jahressignals nicht massiv zu verfälschen, wird

¹Strömungsdaten stammen aus den Karten für mittlere Strömung der 1. Schicht (0-5 m) im Archiv des Bundesamt für Seeschifffahrt und Hydrographie BSH, unter <http://www.bsh.de/akt/dat/modell/stroemungen/Modell1.htm>

5 Kritische Bewertung der Datenbasis

eine weitere Analyse für Perioden zwischen 100 und 450 Tagen durchgeführt. Das Jahres-signal lässt sich hier mit einer Genauigkeit von einem Tag nachweisen; diese Spektren sind hier nicht dargestellt. Um die 3-Tages-Signale genauer zu charakterisieren, wird die Untersuchung für ein kleineres Periodenintervall von 2 bis 40 Tagen wiederholt (Abb. 5.6 (d)-(f)).

Die signifikanten Perioden (aus den unterschiedlichen Frequenzanalysen, siehe Tabelle 5.4) sehen wie folgt aus:

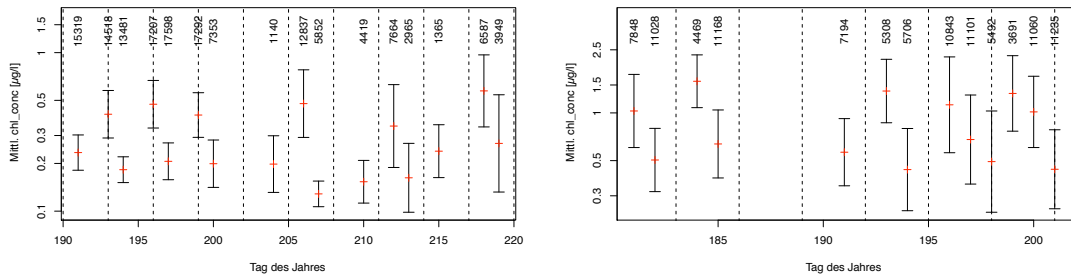
	<i>algal_1</i>	<i>algal_2</i>	<i>chl_conc</i>
Periode [Tag]	363.8 181.8 3.21 3.15 3.13	364.6 182.5 3.18 3.13	366.7 181.8 161.9 3.21 3.15 3.13
Spektrale Leistungsdichte	121.9 74.6 16.2 19.7 18.2	59.0 54.8 15.8 14.3	123.9 83.9 18.6 24.3 27.7 30.0

Tab. 5.4: Signifikante Perioden der Chlorophyllzeitreihe zwischen 2 und 450 Tagen. Die Zeitreihe stammt aus dem Gebiet der Doggerbank (Gebiet 1) und umfasst die Jahre 2003-2008. Ausgewählt aus den Maxima der spektralen Leistungsdichte für die im Text genannten Periodenintervalle.

- *algal_1*: Dieses Produkt zeigt sowohl den Jahres- als auch den Halbjahresgang. Zusätzlich erscheinen starke Signale bei einer Periode von 3.20 bzw. 3.15 Tagen.
- *algal_2*: Auch hier ist Jahres- und Halbjahresgang, sowie der 3-Tages-Zyklus mit 3.18 Tagen zu finden.
- *chl_conc*: Neben Jahres- und Halbjahresgang tritt hier erneut der 3-Tages-Zyklus auf. Drei Perioden von 3.21, 3.13 bzw. 3.15 Tagen sind signifikant.

Auch für die Zeitreihen, die als mittlere Chlorophyllkonzentration auf den Gebieten 1 bis 5 definiert sind, bleiben Jahres-, Halbjahres- und 3-Tage-Signal erhalten (Tabelle 5.5). Für letzteres Signal gibt es keine biologische oder ozeanographische Erklärung (vgl. Argumentation in Abschnitt 5.4.1). Ähnlichkeit besitzt nach drei Tagen nur die Beobachtungssituation, d.i. die Kombination von Winkel von Satellit und Sonnenstand. Dieses Signal besitzt eine höhere Frequenz als die abgeschätzte Nyquistfrequenz. Aliasing tritt hier jedoch nicht auf, denn zu erwarten wäre eine Frequenz des Aliassignals f_A bei $f_A = f_{Nyquist} - (f_{>Nyquist} - f_{Nyquist})$. Im Zeitbereich besäße eine kurze Periode von 3.18 Tagen bei einer Nyquistperiode von 6.01 Tagen ein Aliassignal mit einer Periode von 54.6 Tagen. Dergleichen tritt nicht auf, sodass die mitunter bessere zeitliche Auflösung - viele aufeinander folgende Punkte liegen nur einen Tag auseinander - ausreicht, um auch das hochfrequenteres Signal zu bestimmen.

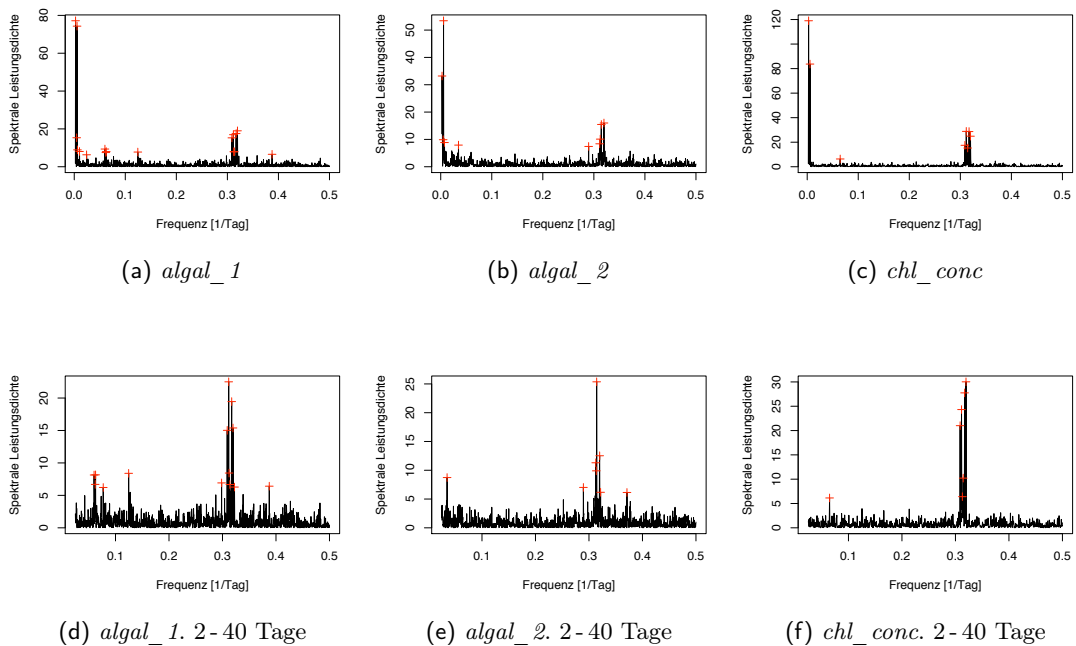
5 Kritische Bewertung der Datenbasis



(a) Doggerbank (Gebiet 1), Tage 191-220.
10. Juli-8. Aug. 2006

(b) Dänische Küste (Gebiet 5), Tage 180-205.
29. Juni-24. Juli 2006

Abb. 5.5: Ausschnitte der Zeitreihen für Gebiete 1 und 5 im Jahr 2006. Mittelwert (rote Kreuze) und Standardabweichung der logarithmierten Chlorophyllkonzentration (Regionalprozessor) in jeweils einer Region sind dargestellt. Die gestrichelten Linien markieren zur besseren Orientierung jeden dritten Tag. Die Zahlen am Oberrand geben die Anzahl von Datenpunkten an, die in die Mittelwertberechnung geflossen ist.



(a) *algal_1*

(b) *algal_2*

(c) *chl_conc*

(d) *algal_1*. 2-40 Tage

(e) *algal_2*. 2-40 Tage

(f) *chl_conc*. 2-40 Tage

Abb. 5.6: Periodizitätsuntersuchung für die Chlorophyllzeitreihe im Zeitraum 2003 bis 2008 über der Doggerbank (Gebiet 1). Die Periodendauer liegt in der ersten Analyse, (a) bis (c), zwischen 2-450 Tagen, in der zweiten, (d) bis (f), zwischen 2-40 Tagen. Dargestellt wird die spektrale Leistungsdichte gegen die Frequenz.

5 Kritische Bewertung der Datenbasis

Gebiet	Doggerbank, 1	Engl. Kanal, 2	Zentral, östl. Doggerbank, 3	Deut. Bucht, 4	Dän. Küste, 5
Periode [Tage]	366.7 181.8 161.9 3.24 3.21 3.15 3.13	368.7 184.1 150.2 3.15 3.13	367.0 182.1 3.24 3.15 3.13	369.5 183.8 3.21 3.13	364.2 183.3 3.24 3.21
Nyquist- Periode [Tage]	6.01	6.01	4.05	6.01	6.01

Tab. 5.5: Signifikante Perioden in Tagen im Chlorophylldatensatz des Regionalprozessors für alle Gebiete 1-5. Die Nyquist-Periode wird nach Gl. 5.1 bestimmt.

5.4.3 Systematischer Fehler im Jahrgang

Die Zeitreihen des Chlorophyllproduktes des Regionalprozessors für Gebiet 1-5 wird jeweils in Stücken von 121 Tagen (als gleitendes Fenster) auf ihre vorherrschende Periodizität im Bereich zwischen 2 und 40 Tagen untersucht. Dazu wird die markanteste Periode dann dargestellt, wenn die Wahrscheinlichkeit sehr gering ($p < 0.05$) ist, das es sich *nicht* um ein Maximum in der spektralen Leistungsdichte handelt, es also wahrscheinlich ein Maximum ist. Zu den weniger ausgeprägten Maxima gehören häufig solche, die mit ganzzahligen Bruchteilen der 35-Tage Wiederholungsrate der relativen Orbits korrespondieren.

Die markanteste Periode zeigt eine systematische Änderung zwischen erster und zweiter Jahreshälfte (Abb. 5.7). Die Periodizität verhält sich annähernd punktsymmetrisch zur Jahresmitte (Sommersonnenwende Tag 172/173) und beginnt in der ersten Jahreshälfte mit einem Wert von 3.15 ± 0.03 Tagen und steigt in der zweiten Jahreshälfte auf 3.21 ± 0.02 Tage an. Sie liegt damit annähernd symmetrisch um $35 \text{ Tage}/11 = 3.18$ Tage.

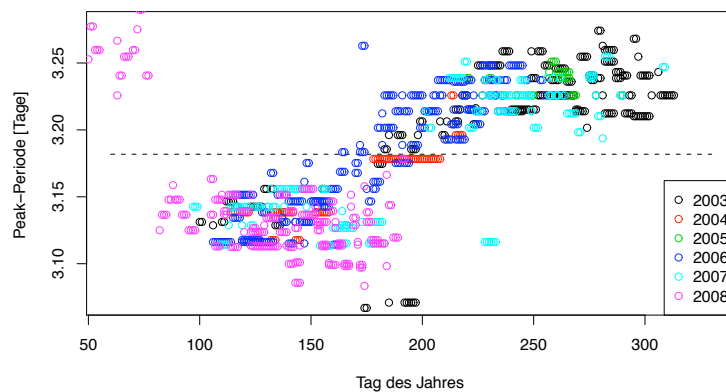


Abb. 5.7: Veränderung der Hauptperiode innerhalb eines gleitenden Fensters von 121 Tagen, jeweils aufgetragen zum Mittelpunkt des zeitlichen Fensters. Die unterschiedlichen Jahre sind farblich hervorgehoben.

Dieses Signal tritt auch bei Zeitreihen des Level 2-Produkt Sichttiefe $Z90_max$ und der Schwebstoffe auf.

Ähnliche, aber weniger stark ausgeprägte Verschiebungen der Hauptperioden finden sich auch im Standardprodukt (Abb. 5.8). Sie konzentrieren sich recht deutlich bei den Periodenlängen, die durch die Orbitfrequenz und ihre Vielfachen vorgegeben wird. Dies ist ein weiterer Hinweis darauf, dass dieses Signal ursächlich durch die Geometrie von Sonneneinstrahlung und Beobachtung beeinflusst wird. Der Jahresgang der Hauptperiode wird dagegen für die Standardprodukte nicht deutlich.

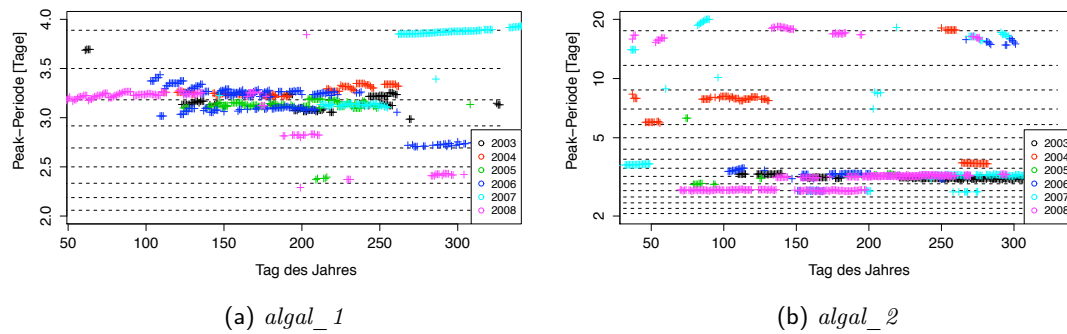


Abb. 5.8: Hauptperioden im Standardprodukt bei gleitenden Zeitreihenausschnitten von 121 Tage. Als gestrichelte Linien sind die Vielfachen der Orbitfrequenz eingetragen.

Der dreitägige Versatz ist in allen Produkten vorhanden, obwohl sie sich grundsätzlich in ihrer Atmosphärenkorrektur unterscheiden. Gleich ist ihnen jedoch, dass auf Berücksichtigung der Polarisation verzichtet wird.

5.5 Systematische Fehler im Level 2-Produkt des Regionalprozessors

Schon in Bilderserien der Chlorophyllverteilung fällt ein etwa dreitägiger systematischer Versatz der Absolutwerte der Konzentrationen auf. Als großräumiges Beispiel sei auf die Chlorophyllkonzentrationskarten von aufeinander folgenden Tagen über der Doggerbank verwiesen (siehe Abschnitt 6.3.1). In der Frequenzanalyse der Zeitreihen (Abschnitt 5.4.1, Abb. 5.4) lässt sich diese etwa dreitägige Periode in allen herangezogenen Gebieten bestätigen.

Da die Frequenzanalyse bereits eine Abhängigkeit des Produktes von der Geometrie des Sonnenstandes und der Beobachtungsposition nahe legt, stellt sich die Frage, welche der vier zur vollständigen Beschreibung der Geometrie notwendigen Winkel diesen Einfluss ausüben.

Der Beobachtungszimut ist über die linke bzw. rechte Bildhälfte einer Szene annähernd konstant und unterscheidet hauptsächlich linke von rechter Bildhälfte (Abb. 5.9). Die Beobachtungszenit variiert symmetrisch zur Mittellinie einer Szene.

Sonnenzenit und -azimut variieren jeweils glatt über ein Bild, wobei die Hauptkomponente des Gradienten im Zenit von Süd nach Nord und im Azimut von West nach Ost weist.

Die Korrelation zwischen Chlorophyllprodukt und Winkeln ist für den Sonnenazimutwinkel am stärksten ausgeprägt (Abb. 5.10). Die kaum ausgeprägte Abhängigkeit vom Beobachtungszimut spricht dagegen, dass insbesondere Polarisierungseffekte beim Blick der Satellitenkamera mit der Sonne (rechte Bildhälfte) im Gegensatz zum Blick gegen die Sonne (linke Bildhälfte) für den Versatz verantwortlich sind. Mit der linken Bildhälfte verbindet sich immer ein höherer Sonnenazimutwinkel, als auch die höheren Chlorophyllkonzentrationen.

Es sind die Winkel zur Beschreibung der Sonnenposition, die an drei aufeinander folgenden Tagen durch die unterschiedlichen Orbits einen dreistufigen Versatz durchlaufen können.

5.5.1 Instrumentenfehler und systematischer Fehler

Für eine genauere Charakterisierung des systematischen Fehlers bietet sich eine ungewöhnliche Kombination der Produkte an.

Dazu werden alle Datentage (zwischen dem 1. Januar 2003 und dem 24. Mai 2006) für die Produkte *chl_conc* (Regionalprozessor *meris_c2r*) und *algal_2* (MERIS Standardprozessor) gemäß ihrer Pixelkoordinaten (X, Y) in jeweils einem Bild zusammengefasst, ohne auf den tatsächlichen geographischen Ort Rücksicht zu nehmen (Mittelung als Mittelwert größter Wahrscheinlichkeit, vgl. Abschnitt 7.2). Bei den vom Instrument unabhängigen Strukturen sollte der Median/Mittelwert über eine Spalte oder eine Zeile des Bildes keine Abhängigkeit von der Pixelnummer zeigen. Pixelnummern in Y -Richtung der Szenen werden in Richtung des Überfluges von Norden nach Süden aufsteigend gezählt, sodass Süden in folgenden "Karten" oben und Norden unten (gespiegelt) zu liegen kommt. Die Pixelnummer nimmt in X -Richtung von links nach rechts in der Szene zu. Dabei setzt sich die Szene (in reduzierter Auflösung, von links nach rechts bzw. annähernd von West nach Ost) aus den Aufnahmen der fünf Kameras zusammen: Kamera 5 von Pixel 26 bis 272, Kamera 4 von Pixel 273 bis 468, Kamera 3 von Pixel 469 bis 650, Kamera 2 von Pixel 651 bis 844 und Kamera 1 von Pixel 845 bis 1094.

Das räumliche Mittel auf dem Pixelgitter für das Chlorophyllprodukt *algal_2* des Standardprozessors (Abb. 5.11) zeigt als Median über die Spalten des Bildes sowohl deutlich die Kameragrenzen von Kamera 4 als auch einen ansteigenden Trend mit der Pixelnummer in X -Richtung. Die sich wiederholenden Strukturen entstehen durch die relativ konstant hohe Chlorophyllkonzentration in der Deutschen Bucht und die 27 diskreten Orbits, unter denen die Nordsee nur gesehen werden kann bzw. in diesem Datensatz nur gesehen wird.

In der Variante aus dem Chlorophyllprodukt des Regionalprozessors (Abb. 5.12) bleibt der Trend mit zunehmender Pixelnummer erhalten. Statt eines starken Versatzes für Kamera 4 ist das Produkt hier jedoch eher stetig in den übrigen Verlauf, der sich stark an der spektralen Charakteristik der fünf Kameras orientiert (Fischer (2005)), eingebettet. Es ist bekannt, dass MERIS aus vier baugleichen Kameras plus dem Prototyp

5 Kritische Bewertung der Datenbasis

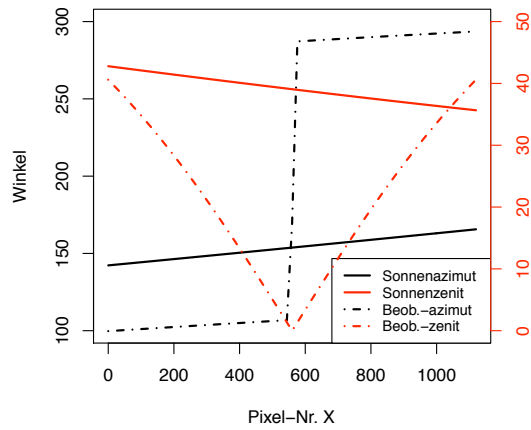


Abb. 5.9: Beobachtungs- und Sonnenstandswinkel einer Satellitenszene im Gebiet der Nordsee vom 10. Mai 2006 entlang einer Linie konstanter Pixelnummer in X-Richtung.

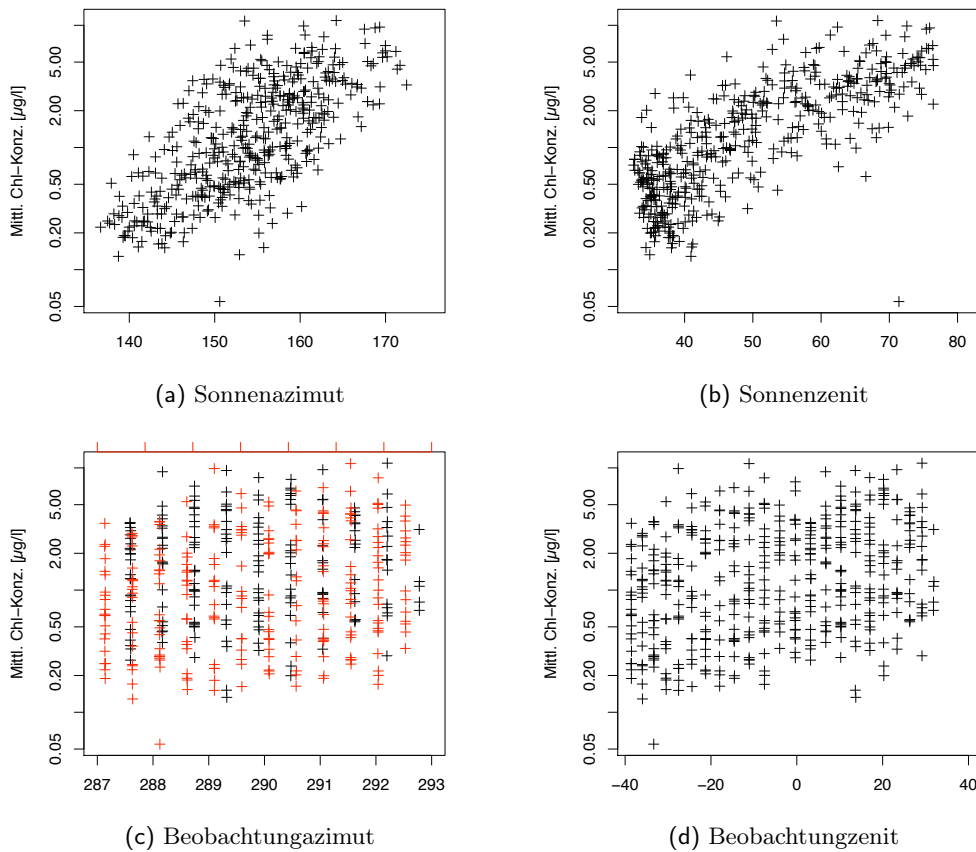


Abb. 5.10: Mittlere Chlorophyllkonzentration gegen verschiedene Beobachtungs- und Sonnenstandswinkeln im Bereich der Doggerbank (Gebiet 1) für die Jahre 2003-2008.

(Kamera 4) zusammengesetzt worden ist, und diese Instrumente sich in ihren spektralen Charakteristiken unterscheiden (Saunier u. a. (2006)). Mit dem Regionalprozessor ist dieser systematische Fehler aber weitestgehend behoben. Nichtsdestotrotz fallen alle Unstetigkeitsstellen, die selbst im Rauschen noch zu erkennen sind, mit Kameragrenzen zusammen.

Diskussion Im Vergleich unterscheiden sich die Produkte *algal_2* und *chl_conc* deutlich in der Ausprägung der Fehler. Während der Regionalprozessor den Instrumentenfehler in Kamera 4 im Vergleich zum Standardprozessor nicht mehr zeigt, wird bei ersterem der Trend ansteigender Chlorophyllkonzentration mit höherer Pixelnummer, d.h. von links nach rechts in einer Szene, wesentlich größer. Der Trend ist im Allgemeinen nur für die Level-2-Produkte zu sehen. Er ist für den Regionalprozessor etwa vier bis fünfmal so groß wie für das Standardprodukt.

Dabei gilt es zu bedenken, dass eine Summation über die Pixel im östlichen (rechten) Bildbereich bereits tendenziell einzig durch die Geographie höhere Chlorophyllkonzentrationen sieht als ein Pixel im westlichen Teil des Bildes. In diese Bereiche fallen häufiger die hohen Konzentrationen der Deutschen Bucht.

Um eventuelle Effekte der vernachlässigten Polarisation nachweisen zu können, wird die Zusammenfassung der Chlorophyllkonzentration *chl_conc* als Karte bezüglich der Pixel-Koordinaten für den Ausschnitt der Doggerbank (Gebiet 1) im Sommer wiederholt (Abb. 5.13). In diesem Zeitraum mit dem Sommerminimum ist in dieser Region hauptsächlich mit konstanter, niedriger Chlorophyllkonzentration und geringen Strukturen zu rechnen, sodass systematische Versätze im Produkt besonders gut zu erkennen sind (Abb. 5.5a). Ein eindeutiges Indiz darauf, dass hohe oder niedrige Chlorophyllwerte hauptsächlich in der rechten bzw. linken Bildhälfte auftreten, gibt es nicht, allerdings kann man von einem linken und einem rechten Plateau mit breitem Übergangsbereich sprechen.

5.5.2 Der Sonnenazimutwinkel

Von allen zur Berechnung des Produktes verwendeten Winkeln ist es der Sonnenazimutwinkel, der die stärkste Korrelation mit dem Chlorophyllprodukt zeigt (Abb. 5.10a).

Definition Sonnenazimutwinkel: Der Sonnenazimutwinkel ist innerhalb des Produktes des Regionalprozessors wie folgt definiert: vom Pixel aus schaut der Beobachter zur Sonne. Der Winkel, den diese Linie im Gegenuhrzeigersinn mit der Nordrichtung einschließt, ist der Azimutwinkel.

Der Sonnenazimut ist ein Maß der wahren Ortszeit, sodass 24 h gerade 360° entsprechen. Im Jahr verschiebt sich durch die elliptische Bahn der Erde die wahre Ortszeit gegenüber unserem 24 h-Tag um bis zu 17 min. Dieser Gang findet sich im beobachteten Sonnenazimutwinkel (vgl. Abb. 5.14, über alle Jahre, 2003-2008) wieder und lässt sich anhand der gleichen Orbits in Abhängigkeit vom Tag des Jahres (Schaltjahre sind in die Zählung eingeflossen) wiedererkennen. Nach der Korrektur mit der Zeitgleichung (= wahre

5 Kritische Bewertung der Datenbasis

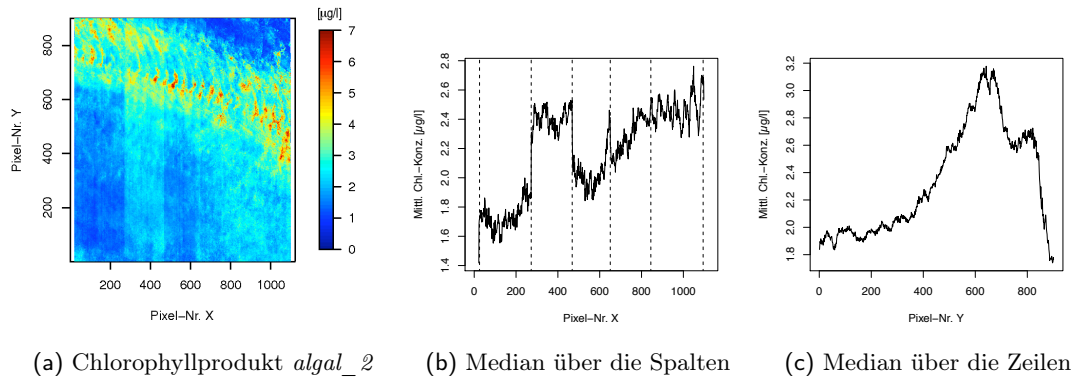


Abb. 5.11: Mittelwert des Standardprodukts *algal_2* berechnet auf einem Gitter aus Pixelkoordinaten der Einzelszenen für die Jahre 2003-2006.

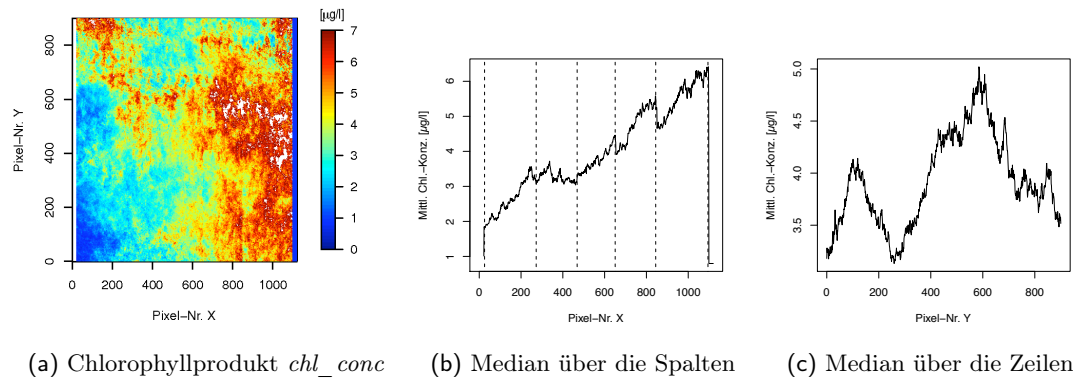


Abb. 5.12: Mittelwert des Regionalprozessorproduktes *chl_conc* auf einem Gitter aus Pixelkoordinaten der Einzelszenen für die Jahre 2003-2006.

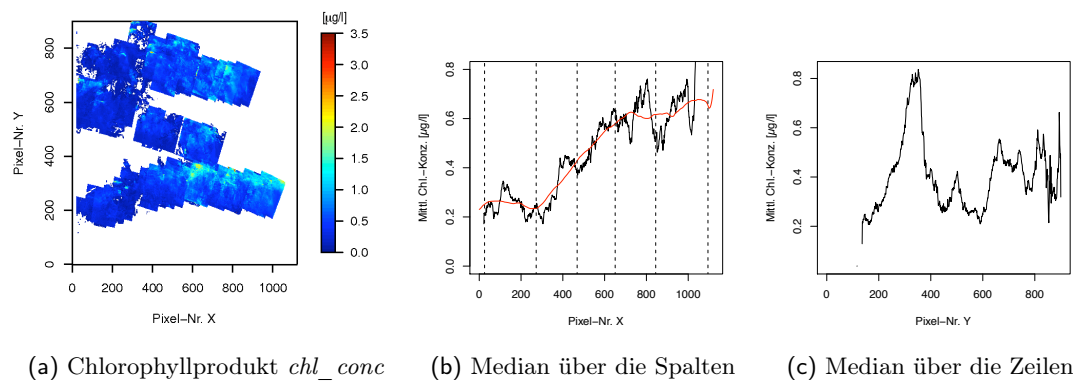


Abb. 5.13: Mittlere Chlorophyllkonzentration im Sommer über der Doggerbank (Gebiet 1) auf einem Gitter aus Pixelkoordinaten für die Jahre 2003-2006.

Ortszeit – mittlere Ortszeit) drückt der korrigierte Sonnenazimut nur noch die mittlere Ortszeit aus.

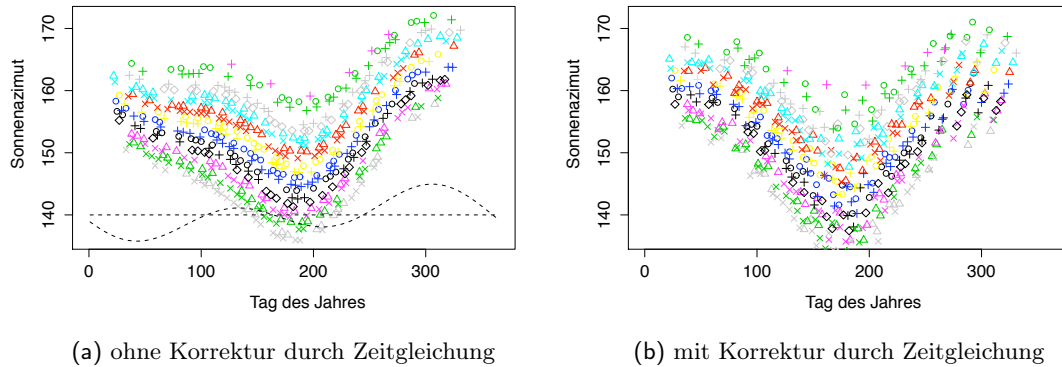


Abb. 5.14: Sonnenazimut im Jahresgang für den Punkt 54° N, 8° O. Farbe und Form der Markierungen stehen für einzelne relative Orbits. Als gestrichelte Kurve ist die Zeitgleichung eingezeichnet.

5.5.3 Theoretische Überlegung zum Sonnenazimutwinkel

Dass auch nach der Korrektur mit der Zeitgleichung ein ausgeprägter Jahresgang im Sonnenazimut verbleibt, hängt mit der lokalen Definition dieses Winkels auf der Erdoberfläche und der sich beständig im Jahresgang veränderten Lage der Nordrichtung zusammen.

Man betrachte folgendes System (Abb. 5.15): der Nullpunkt des Koordinatensystem x, y, z liegt im Zentrum der Erde (als gedachter Kugel), die x -Richtung weist zur Sonne, die y -Richtung liegt tangential zur (Kreis-)Bahn der Erde um die Sonne, die z -Richtung steht senkrecht auf der Bahnebene. Dann lässt sich ein Koordinatensystem (Erdsystem) x', y', N finden, sodass x' innerhalb der Ebene x, N liegt. N bezeichnet dabei die Nordrichtung, die gegenüber z konstant um den Winkel $\Theta = 23.44^\circ$ geneigt ist und im Verlauf eines Jahres einmal um die z -Achse rotiert. Zu einem Zeitpunkt t - gemessen in Stunden seit Beginn eines kalendarischen Jahres - ist diese Drehung um

$$\Phi(t) = \frac{2\pi}{365 \cdot 24} ((\text{doy} - 9) \cdot 24 + (\text{time} - 12))$$

fortgeschritten. Dabei steht *doy* für den Tag des Jahres (Abk. *day of year*) und *time* für die Uhrzeit in UTC. $\Phi(t) = 0$ ist gleichbedeutend mit dem Sonnenhöchststand, d.i. 12 Uhr mittags, zur Wintersonnenwende.

Die Ebene x, N definiert immer den Sonnenhöchststand: nach jedem Tag um 12 Uhr befindet sich gemäß der Überlagerung der beiden Rotationen der gleiche Punkt wieder an dieser Stelle. Für einen Punkt P, der sich gerade in dieser 12 Uhr-Position befindet, beträgt der Sonnenazimutwinkel genau 180° . Zu jedem anderen Zeitpunkt befindet sich der Punkt P um die Achse N mit dem Winkel $\alpha(t) = \frac{2\pi}{24} (\text{time} - 12)$ gedreht. Innerhalb des

5 Kritische Bewertung der Datenbasis

Erdsystems lässt sich dieser Punkt - der gleichzeitig die Normale des Beobachtungspixels darstellt - als

$$P = \begin{pmatrix} \sin \theta \cos \alpha(t) \\ \sin \theta \sin \alpha(t) \\ \cos \theta \end{pmatrix}$$

beschreiben. Durch Koordinatentransformation des Erdsystems x', y', N in das System x, y, z um $-\alpha(t)$, $-\Phi(t)$ und $-\Theta$ erhält man eine Darstellung der Richtung P im System x, y, z . Damit ist es möglich, zwei Projektionen innerhalb der durch P definierten Ebene (des Beobachtungspunktes) zu bestimmen: zum einen die Projektion der Nordrichtung in diese Ebene, zum zweiten die Projektion der konstanten Sonnenrichtung.

Der Winkel zwischen diesen beiden Projektionen ist dann der gesuchte Sonnenazimutwinkel.

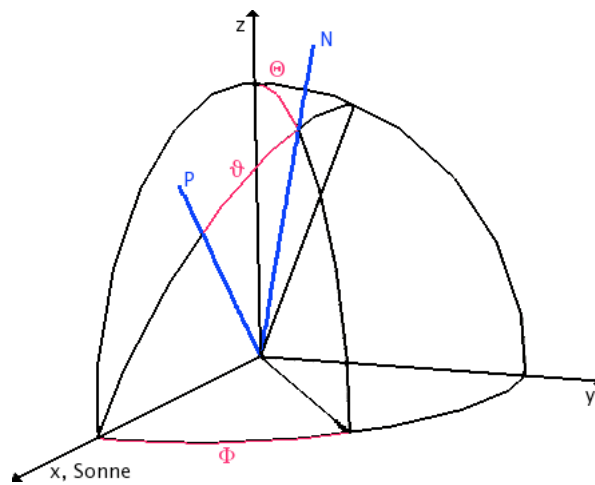


Abb. 5.15: Geometrie zur theoretischen Bestimmung des Sonnenazimuts. Eine 12 Uhr-Situation, in der Punkt P gerade in der Ebene des Sonnenhöchststandes liegt.

Für einen festen Punkt mit geographischer Breite von 54°N und Länge von 8°O (in der Deutschen Bucht) werden aus dem Datensatz der Szenen alle Uhrzeiten bestimmt, zu denen der Punkt gesehen worden ist. Da die Zeit als UTC auf 0° definiert ist und für die Definition oben die wahre Ortszeit benötigt wird, muss zu der UTC-Uhrzeit noch $\frac{8^\circ}{360^\circ} \cdot 24 \text{h}$ hinzugefügt werden. Damit lässt sich der Jahresgang des Sonnenazimuts bis auf die Korrektur der Zeitgleichung (für elliptische Umlaufbahn) simulieren (Abb. 5.16).

5 Kritische Bewertung der Datenbasis

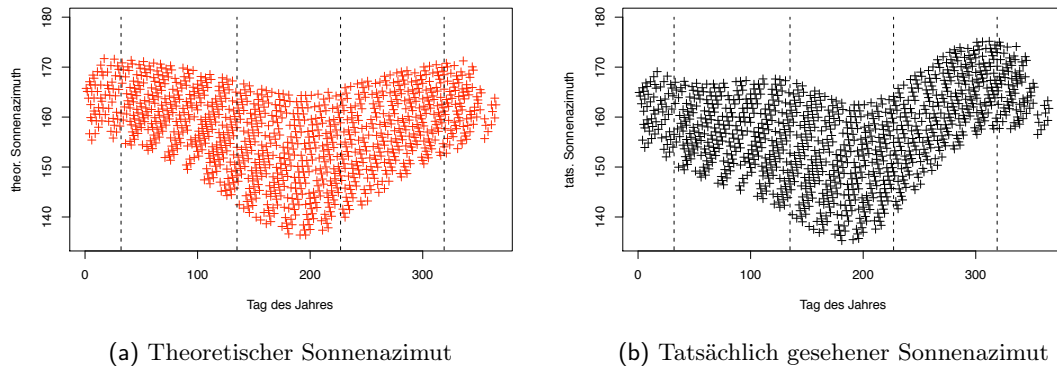


Abb. 5.16: Sonnensazimuth im Jahresgang für Position 54° N, 8° O in der Deutschen Bucht. Zur Berechnung der theoretischen Größe des Sonnensazimuths werden alle Zeitpunkte der Zeitreihe aus dem Gesamtzeitraum verwendet.

Die Frequenzanalyse der theoretischen Sonnensazimuthwinkel (Abb. 5.17) liefert den 3-Tages-Rhythmus, der hier sogar stärker als das Jahressignal ausgeprägt ist und sich hauptsächlich in der Periode von 3.18 Tagen, d.h. in der Harmonischen der Orbitwiederholungsrates, widerspiegelt. Daß der Jahresgang im Chlorophyllprodukt ausgeprägter als diese Harmonische ist (vgl. Abb. 5.6 und Tabelle 5.4), ist ein Indiz dafür, dass es neben dem Signalanteil aus den Winkeln auch einen echten Chlorophylljahresgang innerhalb der Zeitreihe gibt. Für eine genauere Untersuchung der Zeitreihe ist eine Filterung der nicht-äquidistanten Zeitreihen unumgänglich (Scargle (1989)). Dieser Bearbeitungsschritt steht weiterhin aus.

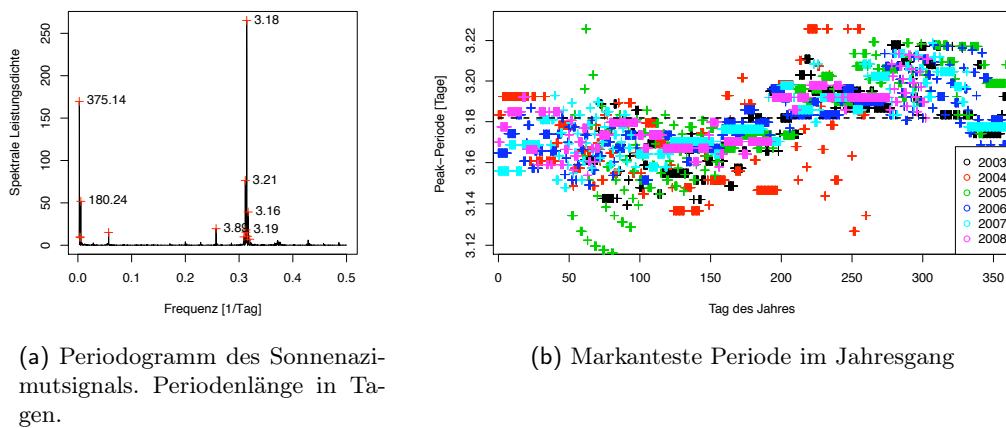


Abb. 5.17: Frequenzanalyse des (theoret.) Sonnensazimuthwinkels für Perioden zwischen 2 und 450 Tagen.

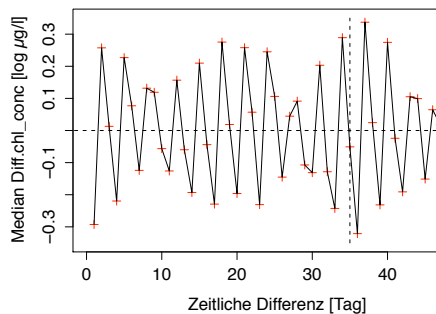
5.5.4 Eine einfache Frequenzanalyse

Ein Einfluss des Jahresgangs des Chlorophyllsignals wird in dieser Betrachtung mit folgender Überlegung weitestgehend ausgeschlossen: in der häufig beobachteten Abfolge von einem Tag hoher und anschließend niedriger Konzentration ist das Niveau des Absolutwertes, um das herum der Versatz in Erscheinung tritt, vom tatsächlichen Jahresgang des Signals abhängig. Zwei Paare von Datenpunkten können somit auf ganz unterschiedlichen Niveaus des Absolutwertes liegen, die Differenz der Konzentrationen bei Paaren mit gleichem zeitlichen Abstand ist jedoch annähernd konstant.

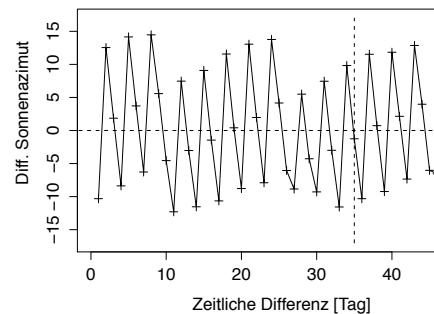
Auch bei einer zeitlichen Auflösung der Chlorophyllkonzentrationszeitreihe in Tagen - die Uhrzeit an einem festen Beobachtungspunkt schwankt maximal um etwa eine Stunde - ergibt sich kein signifikanter Unterschied zu den markanten Frequenzen bei genauer Beobachtungsurzeit. Daher wird hier als Zeitpunkt nur der Tag berücksichtigt.

Für diese Zeitreihe werden alle möglichen Datenpaare so gebildet, dass die Differenzen der Beobachtungszeitpunkte zwischen 1 und 45 Tagen liegen. Für jede Klasse der Zeitdifferenz wird jeweils die Differenz des logarithmierten Chlorophyllwertes aller Paare gemittelt (Median). Die mittlere Differenz des Chlorophyllproduktes zeigt erneut deutlich die etwa dreitägige Periode, welche überlagert wird von der 35-Tage-Periode, der Orbitwiederholungsrate (Abb. 5.18).

Wie durch den linearen Zusammenhang zwischen dem Sonnenazimutwinkel und dem Chlorophyllprodukt (jeweils der Median der Differenzen von Datenpunktpaaren für eine konstante Zeitdifferenz) deutlich wird, existiert ein enger Zusammenhang zwischen beiden (Abb. 5.19).



(a) Mittlere Differenz der Chlorophyllkonzentration (log.) zu verschiedenen zeitlichen Abständen in Tagen



(b) Mittlere Differenz des Sonnenazimuts zu verschiedenen zeitlichen Abständen in Tagen

Abb. 5.18: Mittlere Differenzen für alle Datenpunktpaare im Abstand von Tagen (Doggerbank, Gebiet 1). Die Wiederholung des Orbits nach 35 Tagen ist markiert.

5 Kritische Bewertung der Datenbasis

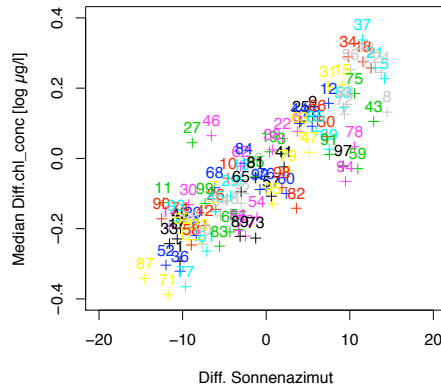


Abb. 5.19: Mittlere Differenz log. Chlorophyllkonzentration gegen Differenz Sonnenazimut (Daten über der Doggerbank, Gebiet 1). Die Zahlen geben den zeitlichen Abstand in Tagen an.

5.5.5 Mittlerer Fehler

Eine Fehlerabschätzung für das Produkt des Standard- bzw. des Regionalprozessors aus den Zeitreihen der fünf Untersuchungsgebiete soll an dieser Stelle anschließen. Wie im vorhergehenden Abschnitt wird wieder die Statistik von Punktpaaren in festem zeitlichen Abstand zueinander bemüht, um den Einfluss des tatsächlichen Jahresganges zu minimieren und hauptsächlich die Größe des Fehlers abzuschätzen, der durch den Versatz in den Datensatz eingebracht wird.

Für einen festen zeitlichen Abstand von $\Delta t = 1, 2, 3$ Tagen wird jeweils die (absolute) Differenz der Chlorophyllkonzentrationen

$$\delta_{\text{chl}}(t, \Delta t) = \|\text{chl_conc}(t) - \text{chl_conc}(t + \Delta t)\|$$

und die mittlere Konzentration

$$\overline{\text{chl_conc}}(t, \Delta t) = \frac{1}{2} (\text{chl_conc}(t) + \text{chl_conc}(t + \Delta t))$$

als Annäherung an den unbekanntem tatsächlichen Wert bestimmt. Der relative Fehler

$$\Delta_{\text{chl}}(t, \Delta t) = \frac{\delta_{\text{chl}}(t, \Delta t)}{\overline{\text{chl_conc}}(t, \Delta t)}$$

wird dann jeweils für alle Zeitpunkte t eines Monats zum mittleren Fehler zusammengefasst (Abb. 5.20).

Die relativen Fehler liegen im Allgemeinen meist über 30%. Die Abschätzung mit der geringsten Varianz im Jahresgang unabhängig vom Untersuchungsgebiet gelingt mit dem Produkt des Regionalprozessors und den Abständen der Beobachtung von drei Tagen (grüne, durchgezogene Linien in Abb. 5.20). Das konstante Niveau des Fehlers variiert jedoch mit dem Beobachtungsraum: während über der Doggerbank mit Fehlern um 40% gerechnet werden muss, sind es im Englischen Kanal 35%, in der östlichen Nordsee 50%

5 Kritische Bewertung der Datenbasis

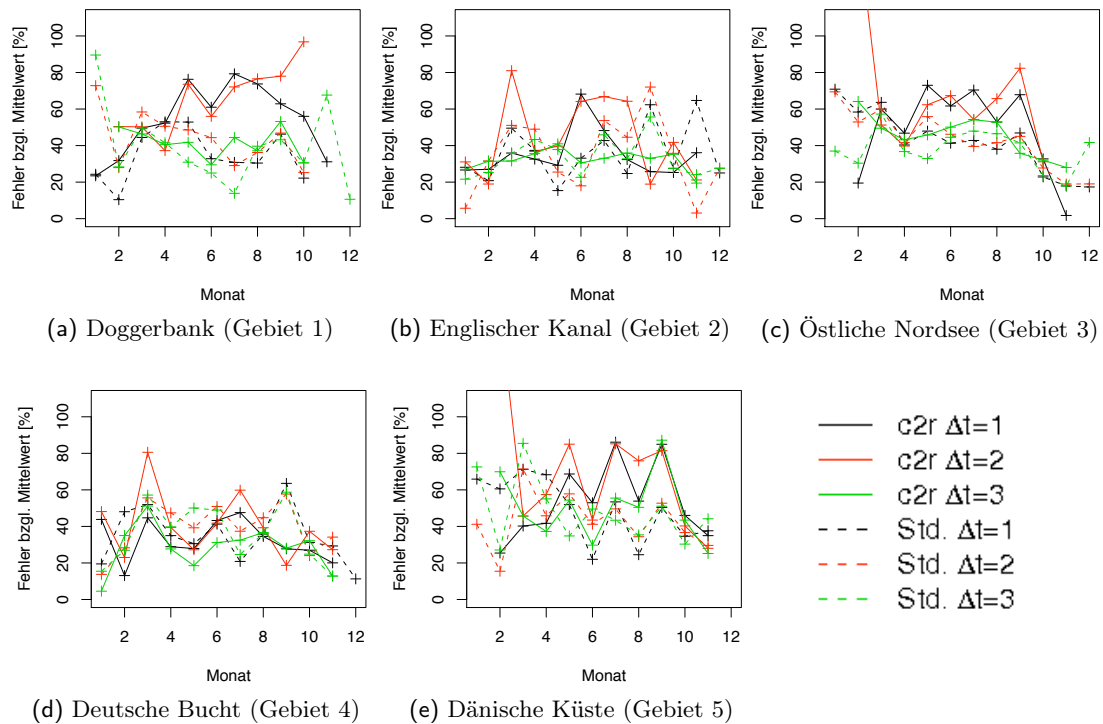


Abb. 5.20: Mittlerer relativer Fehler in den Chlorophyllzeitreihen für unterschiedliche Untersuchungsgebiete (Zeitraum: Regionalprozessor *c2r* 2003-2008, Standardprozessor *Std.* 2003-2007) aufgeschlüsselt nach Monaten und zeitlicher Distanz Δt der Beobachtungen in Tagen. Die Werte im Januar, Februar, November und Dezember beruhen meist nur auf 1-3 Punktpaaren.

und in der Deutschen Bucht 30%. Vor der Dänischen Küste kann man nicht von einem konstanten Fehler sprechen, ein mittleres Fehlerniveau liegt hier etwa bei 50%. Ein Fehler von 30% im Chlorophyllprodukt wird im allgemeinen auch für die case 1-Produkte angenommen.²

In der Deutschen Bucht ist der Fehler für $\Delta t = 3$ Tage deutlich größer als im restlichen Jahreslauf. Dies spiegelt die starke Dynamik und Anstieg der Chlorophyllkonzentration zur Frühjahrblüte wieder, die schon für einen zeitlichen Abstand von drei Tagen keine Annahme der Konstanz zulässt.

Bei den Zeitreihen zur Doggerbank, der östlichen Nordsee und der Dänischen Küste fällt auf, dass die Fehlerabschätzungen (für Punktpaare im Abstand von ein oder zwei Tagen) zum Regionalprozessor meist höhere Werte aufweisen als der Standardprozessor.

Die hohen Fehlerwerte um 70% über der Doggerbank im Sommer gehen bei annähernder Konstanz der tatsächlichen Chlorophyllkonzentration im Sommerminimum vollstän-

²Diese Größe beruht auf den Qualitätszielen für MODIS, vgl. Wayne Esaias, MODIS Ocean Validation, 1996, unter http://modis.gsfc.nasa.gov/MODIS/OCEANS/DOCS/ocean_val_chart.pdf

dig auf den Einfluss aus dem Sonnenstandswinkel zurück.

5.6 Anmerkung und Zusammenfassung

In der vorliegenden Arbeit werden die Daten bezüglich des in diesem Kapitel charakterisierten systematischen Fehlers aus folgenden Gründen *nicht* korrigiert:

Mit Hilfe der Inversen einer Fouriertransformation für unregelmäßige Zeitreihen (Scargle (1989)) ließe sich der 3-Tages-Versatz bereinigen. Eine Verzerrung der Zeitreihe durch das Jahressignal, das sich aus Anteilen zusammensetzt, die ursächlich aus dem Sonnenazimutwinkel bzw. der tatsächlichen Chlorophyllverteilung stammen, bliebe aber bestehen.

Gleichzeitig benötigen die Einzelbilder eine pixelweise Korrektur gemäß der Winkel, unter denen die Sonne gesehen wird. Es wird sinnvoller sein, einen neuen Prozessor für das Chlorophyllprodukt anstatt einer kosmetischen Korrektur des Produktes zu entwickeln.

Es ist daher zu erwarten, dass alle Verfahren mit zeitlichen Komponenten, seien es zeitliche Datenkomposite oder Bearbeitungen von Zeitreihen wie z.B. in der Bestimmung der Dekorrelationszeit (Abschnitt 14.3), Artefakte beinhalten werden.

Die hohen Chlorophyllkonzentrationen im Winter können auf folgende Weise entstehen:

- Der niedrige Sonnenstand - d.h. der Sonnenzenit ist groß - sorgt für eine Geometrie, bei der die vereinfachende Annahme einer Atmosphäre aus planaren Schichten den realen Gegebenheiten nicht mehr gerecht wird. Ein Atmosphärenmodell mit sphärischer Geometrie ist sehr aufwendig und steht nicht zur Verfügung (vgl. Anhang, Abschnitt A.3).
- Da die Oberfläche des Wasserkörpers im Winter statistisch von höheren Wellen bestimmt wird, wird sie bei niedrigem Sonnenstand so beleuchtet, dass die Streuwege der Photonen eher in einem dreidimensionalen Körper verfolgt werden müssten, um die Eindringtiefe und damit das gesehene Chlorophyll angemessen abschätzen zu können. In den Wassermodellen wird zwar eine Neigung der Wasseroberfläche angesetzt, nicht jedoch die tatsächliche kleinskalige Dimensionierung eines Wellenkörpers (Kommunikation Doerffer).
- Im Winter wird, hervorgerufen durch die höheren Windgeschwindigkeiten und dadurch induzierte vertikale Durchmischung, sehr viel Detritus in die Wassersäule resuspendiert. Der Detritus kann eine nicht unerhebliche Menge an Chlorophyllabbauprodukten wie Phaeophytin enthalten, das sich spektral ähnlich zum Chlorophyll verhält und als solches fälschlich erkannt wird (Kommunikation Doerffer).

Die bislang vernachlässigte Polarisierung scheint nach der Sachlage keine entscheidende Einflussgröße auf den systematischen Fehler zu sein. Eine Abhängigkeit von Polarisierungseffekten ließe sich am ehesten über die Berücksichtigung der Beobachtungsrichtung mit der Sonne (linke Hälfte einer Szene, $\text{view_azimuth} < 180^\circ$, hauptsächlich Rückwärtsstreuung) und gegen die Sonne (rechte Hälfte einer Szene, $\text{view_azimuth} > 180^\circ$, hauptsächlich Vorwärtsstreuung, vgl. Abb. 5.10) aus den Daten erschließen. Einzelne

5 Kritische Bewertung der Datenbasis

Szenen zeigen jedoch keinen erkennbaren Versatz der Absolutwerte zwischen linker und rechter Bildhälfte.

Ein geringfügiges, aber nicht eindeutiges Indiz für den Unterschied zwischen den Bildhälften lässt sich in den pixelweisen Verlauf der mittleren Sommerchlorophyllkonzentration über der Doggerbank hineininterpretieren (Abb. 5.5a).

Ein erster Schritt ist mit einem neuen Atmosphärenmodell zur automatischen Korrektur, welche die Polarisierung berücksichtigt, gemacht worden. Vollständige Neuberechnung und Prüfung des neuen Datensatzes (Stand: September 2009) stehen allerdings noch aus.

Ein weiterer Prozessor (Stand: August 2010) befindet sich derzeit in der Entwicklung, der die gesamte Streugeometrie in der Atmosphäre anders bewertet (vgl. Anhang, Abschnitt A.2) und gute Resultate verspricht.

6 Phytoplankton der Nordsee in Beispielen

Nach den Betrachtungen der Zeitreihen in ihrer höchsten zeitlichen Auflösung (vgl. Abschnitt 5.4) sollen in diesem Kapitel erste Eindrücke von räumlichen Strukturen der Phytoplanktonverteilung im Vordergrund stehen. Bei einem Zeitraum von nur sechs Jahren erscheint der Begriff der Klimatologie als langjährige Zustandsbeschreibung eines Raumes noch nicht angebracht. Diese Zeitspanne ist insbesondere zu kurz, um Entwicklungen über die Jahre hinweg beschreiben zu können, die globalen Umweltveränderungen zu verschulden sind.

Stattdessen werden in dieser Studie zum einen mittlere jahreszeitliche Phytoplanktonverteilungen als typische räumliche Muster vorgestellt (Abschnitt 6.1) und die Auswirkung des systematischen Fehlers (vgl. Abschnitt 5.5) auf die Mittelwertbildung von längeren Zeiträumen aufgezeigt.

Mit Rücksicht auf die Länge dieser Arbeit wird auf eine vollständige Zusammenstellung aller Karten von Chlorophyllmonatsmittel verzichtet. Es sei stattdessen auf den existierenden Algenatlas der Nordsee (Peters u. a. (2005)) verwiesen. Eine Zeitreihe mittlerer Chlorophyllkonzentrationen in monatlicher Auflösung wird in Anlehnung an den Grünindex des Kontinuierl. Planktonrekorder (CPR) dargestellt (Abschnitt 6.2).

Abschließend sollen wiederkehrende Strukturen exemplarisch in höchstmöglicher zeitlicher Auflösung anhand der Frühjahrsblüte über der Doggerbank bzw. der Norwegischen Rinne diskutiert werden (Abschnitt 6.3).

6.1 Klimatologie der Chlorophyllkonzentration nach Jahreszeiten

Da die Zeitreihe für eine klimatologische Untersuchung zu kurz ist, wird auf die Ermittlung von interannualen Variationen und Trends verzichtet und stattdessen lediglich das klimatologische Mittel bzw. das Jahreszeitenmittel präsentiert.

Klimatologisch meint dabei, dass alle Datentage für eine Jahreszeit (gemäß der Definition der Zeiträume des Phytoplanktonjahres (vgl. Abschnitt 3.4)) über alle Jahre als Median maximaler Wahrscheinlichkeit (Maximum Likelihood Median, Beschreibung des Algorithmus folgt in Abschnitt 7.2) auf dem Standardgitter (Abschnitt 4.3) pixelweise zusammengefasst werden.

Das klimatologische Mittel (Abb. 6.1a) gibt erste Hinweise auf wiederkehrende Strukturen. Im Allgemeinen folgt die Chlorophyllkonzentration in der südlichen und mittleren Nordsee stark der Bathymetrie (Abb. 2.1a): je geringer die Wassertiefe umso größer die

6 Phytoplankton der Nordsee in Beispielen

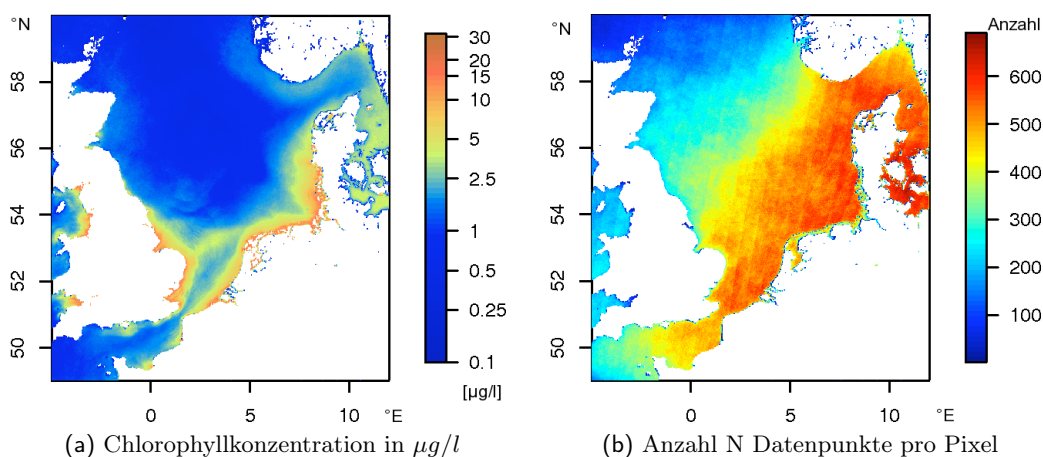


Abb. 6.1: Klimatologisches Mittel des Chlorophyllproduktes *chl_conc* gemäß des Regionalprozessors über den Gesamtzeitraum von Januar 2003 bis August 2008. Streifen in (b) gehen auf die Lage der relativen Orbits zurück.

Chlorophyllkonzentration. Die Ursache ist in der vollständigen Durchmischung dieser Flachwasserbereich durch die Tide zu finden, die für beständigen Nachschub an Nährstoffen sorgt (vgl. Abschnitt 2.3.2). Das gilt nicht für die Norwegische Rinne, in der einströmendes Atlantikwasser oder Wasser aus dem Skagerrak die Phytoplanktodynamik bestimmen. Die Frühjahrsblüte über der Doggerbank ist ebenfalls zu erahnen. Da Szenen ausgewählt werden, die den nicht zentral im Untersuchungsgebiet liegenden Punkt 8° Ost, 54° Nord enthalten, kommt es zur inhomogenen Überdeckung des Gebietes (Abb. 6.1b).

Die klimatologischen Jahreszeiten gewähren eine Übersicht, wann die vorherrschenden Strukturen in der Nordsee auftreten (Abb. 6.2) und wie ausgeprägt sie über die Jahre bzw. den Zeitraum sind (Abb. 6.3).

Das Frühjahr (Abb. 6.2b) wird im Phytoplanktonjahr bereits ab dem 1. Februar bis zum 14. Mai definiert, um die zeitige Frühjahrsblüte in der Norwegischen Rinne zu enthalten - und sie nicht im durch den systematischen Fehler stark verzerrten Winter (Abb. 6.2a) zu ignorieren. Neben dieser zeichnen sich auch die Blüte über der Doggerbank und entlang der Küsten gut ab. Letztgenannte Bereiche besitzen hohe Variabilität (Abb. 6.3b), die sowohl durch Unterschiede von Jahr zu Jahr als auch innerhalb des gewählten Zeitraumes begründet sein können.

Im Sommer stellt sich das Minimum der Chlorophyllkonzentration in der zentralen Nordsee ein, wobei in den tidendurchmischten Bereichen hohe Konzentrationen verbleiben (Abb. 6.2c). Die Darstellung legt nahe, dass das Chlorophyllprodukt als Hilfsgröße für den Schichtungszustand im Sommer benutzt werden kann (vgl. Abschnitt 6.3.2). Die Variabilität ist im Sommer sehr gering (Abb. 6.3c).

Das Abbild der Herbstsituation ist weiterhin von hohen Konzentrationen im tidendurchmischten Bereich vor den Küsten geprägt (Abb. 6.2d). Allerdings ist die Interpretation als zweite Blüte, die mit Aufbrechen der Schichtung eintritt, aufgrund der

6 Phytoplankton der Nordsee in Beispielen

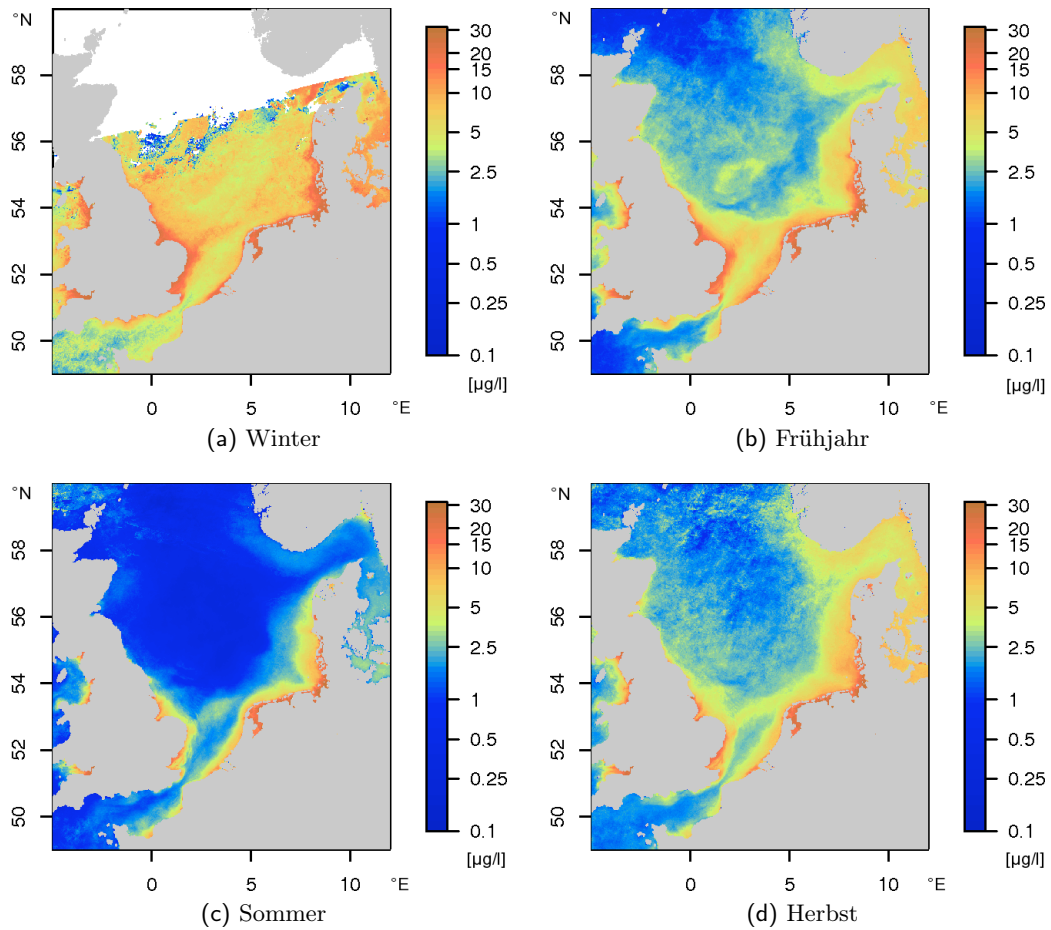


Abb. 6.2: Jahreszeiten des Phytoplanktonjahres, klimatologisches Mittel der Chlorophyllkonzentration über den Zeitraum Januar 2003 bis August 2008 in $\mu\text{g/l}$. Trotz der Beschränkung auf den Qualitätskriterien entsprechende Datenpunkte wird der Winter von unrealistisch hohen Chlorophyllkonzentrationen dominiert.

6 Phytoplankton der Nordsee in Beispielen

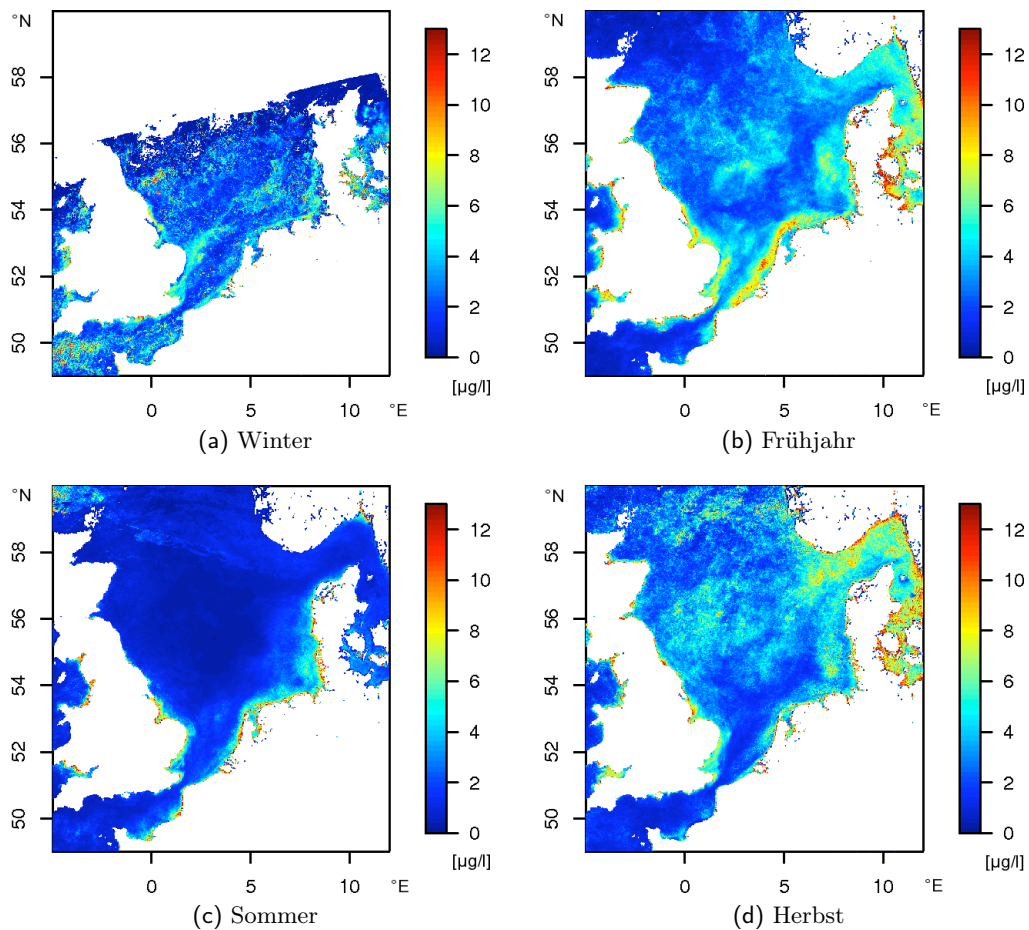


Abb. 6.3: Jahreszeiten des Phytoplanktonjahres, Standardabweichung des klimatologischen Mittels der Chlorophyllkonzentration über den Zeitraum Januar 2003 bis August 2008 in $\mu\text{g/l}$.

zunehmenden Sonnenazimutwinkel nicht gesichert.

Die Abweichung der Chlorophyllkonzentration im Winter von den Erwartungen ist eine direkte Folge des derzeitigen systematischen Fehlers (Abb. 6.2a). Der Einfluss des Sonnenazimuts auf die Chlorophyllkonzentrationen sorgt durch annähernd lineares Verhalten für extrem hohe Chlorophyllwerte im ganzen Winter. Nur ausgehend vom Sonnenazimutwinkel am Punkt 8°O , 54°N und gewichtet mit dem Überdeckungsanteil der einzelnen Szenen besitzt das Winterbild einen mittleren Sonnenazimut von 167.8° (vgl. Tabelle 6.1).

Der Effekt wird auch in den übrigen Jahreszeiten vorhanden sein, verlässliche Absolutwerte kann es erst nach einer Neuberechnung aller Satellitendaten geben. Im Moment dient der Sonnenazimut nur als Hilfsgröße zur Beschreibung der Fehlerabhängigkeit.

Bei allen Interpretationen von Mittelwerten soll nicht außer Acht gelassen werden, daß die Überdeckung pro Pixel (Abb. 6.5) sehr unterschiedlich im Gesamtgebiet ausfallen kann und somit die statistische Aussagekraft regional verschieden ist.

Jahreszeit	Frühjahr	Sommer	Herbst	Winter
Mittlerer Sonnenazimut	159.5°±5.4°	155.9°±7.0°	165.0°±6.4°	167.8°±4.9°
Minimum/Maximum	141.9°/167.6°	135.3°/166.7°	141.6°/175.2°	152.0°/175.0°

Tab. 6.1: Mittlere Sonnenazimutwinkel am Punkt 8° O, 54° N, gewichtet mit dem Anteil von gültigen Datenpunkten über der Nordsee (Großgebiet)

6.1.1 Überdeckungsstatistik

Die Datenüberdeckung bezieht sich auf die Gesamtfläche des Standardgitters, sprich alle Wasserpixel innerhalb der gesetzten geographischen Grenzen. Damit sind streng genommen auch Wasserpixel enthalten, die nicht mehr zur Nordsee gehören.

Im Winter ist die Überdeckung meist geringer als 10% pro Szene (Abb. 6.4a). Das verwendete Bilderarchiv beginnt im Frühjahr 2005 erst ab dem 1. April 2005, genauso fehlen einige Szenen aus dem Herbst 2004 zwischen dem 18. September und 3. November.

Gerade im Sommer wird ein etwa dreitägiger Rhythmus sichtbar in einer Abfolge von hoher, geringer und keiner Überdeckung (Abb. 6.4c), vermittelt durch die Lage der Orbits.

6.2 Jahresgang anhand von Monatsmitteln

Für die Standardregionen gemäß der Bearbeitung des CPR-Grünindex (vgl. Abschnitt 2.2.4) werden die mittleren, räumlichen Chlorophyllkonzentrationen aus den Monatsmitteln bestimmt und wie der Grünindex aufgetragen (vgl. Abb. 3.1). Sichere Interpretationen lassen sich unter Berücksichtigung von nicht zu niedrigem Sonnenstand nur zwischen Mitte Februar bis Anfang November erwarten. Die sich manifestierenden Konzentrationsmaxima zu Jahresende bzw. -anfang werden dementsprechend nicht betrachtet.

Im Standardprodukt lässt sich die Frühjahrsblüte im April recht gut erkennen (Abb. 6.6), mit Ausnahme des Gebietes D2 (südwestliche Nordsee), für das die zeitlichen Strukturen auch im Grünindex uneinheitlich erscheinen. Besonders augenscheinlich wird sie im Gebiet C2 (westl. zentrale Nordsee) und D1 (südöstliche Nordsee, Deutsche Bucht), wo sie nur im Jahr 2006 nicht entdeckt werden kann. Im Jahr 2005 ist das Frühjahresmaximum erst im Mai erreicht (außer in der östlichen zentralen Nordsee, C1). Im Jahr 2007 tritt das Chlorophyllmaximum in der südlichen Nordsee (D1, D2) erst im Mai auf, während es in der nördlichen (C2) im April zu beobachten ist.

Dagegen ist das Chlorophyllprodukt des Regionalprozessor stärker von zeitlichen Artefakten durchsetzt (Abb. 6.7), sodass nur im Jahr 2003 in allen Gebieten die Frühjahrsblüte im April identifiziert werden kann. Klar zu erkennen ist sie des Weiteren nur im Gebiet D1 (südöstliche Nordsee) in den Jahren 2007 und 2008, wobei in letzterem das Maximum bereits im März erreicht wird. Das Ergebnis des Regionalprozessors verhält sich hier allerdings konträr zum Standardprozessor.

Das Auftreten einer zweiten Blüte im Herbst lässt sich nur in der westlichen zentralen Nordsee (C2) im Oktober des Jahres 2006 aus den Daten des Standardprozessors vermuten. Die geringere zeitliche Auflösung wirkt sich hier eher nachteilig aus, da in

6 Phytoplankton der Nordsee in Beispielen

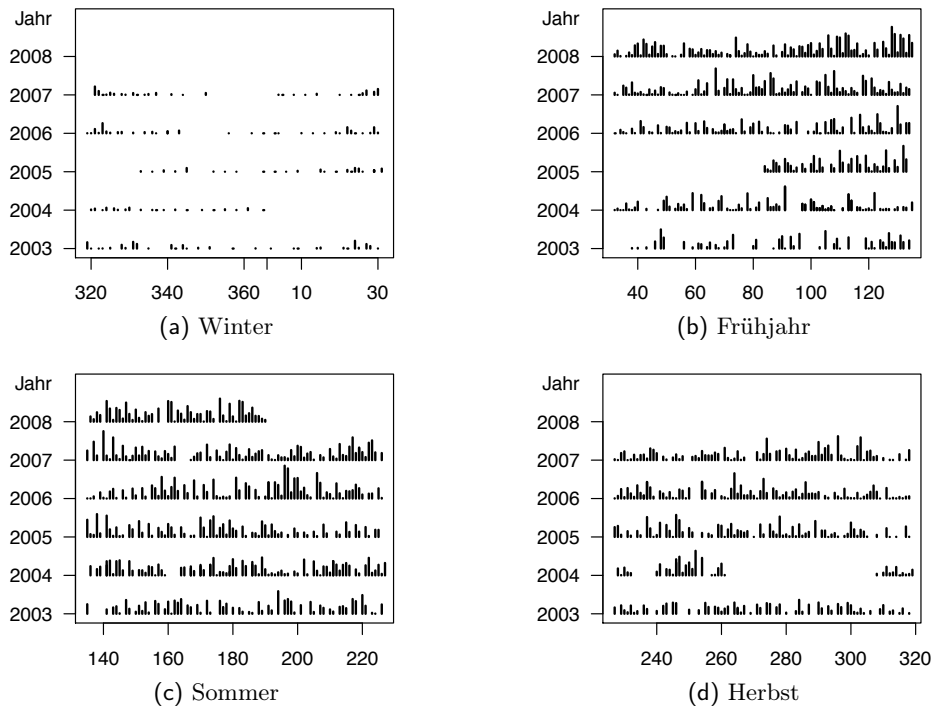


Abb. 6.4: Überdeckungsstatistik zum klimatologischen Mittel. Beiträge einzelner Datentage (als Tag des Jahres) aus den Jahren 2003 bis 2008. Die Länge der Balken zwischen zwei Jahreseinteilungen entspricht dem Anteil gültiger Datenpunkte über Wasser bezogen auf die Nordsee (Großgebiet).

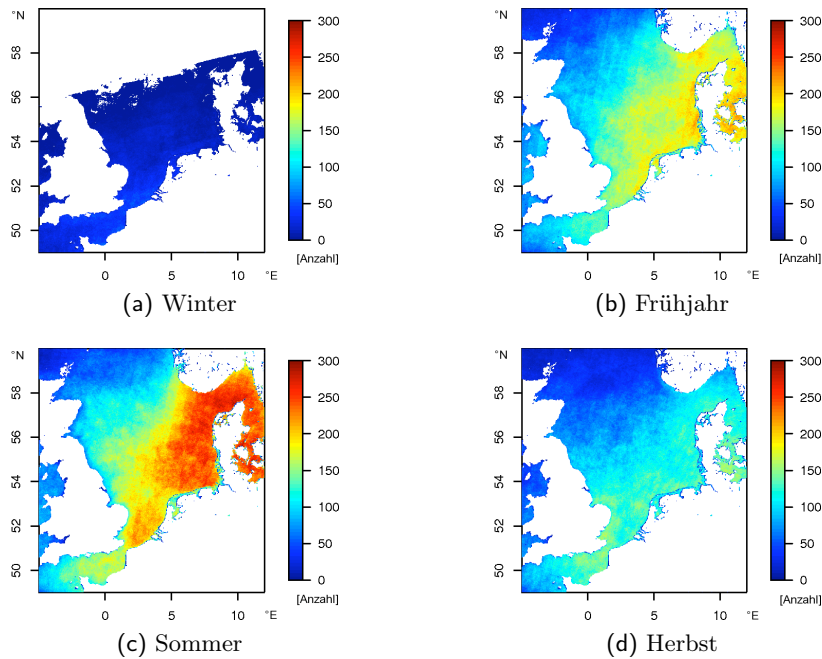


Abb. 6.5: Anzahl der Datenpunkte pro Pixel zum klimatologischen Mittel, räumliche Überdeckung. Streifigkeit entsteht durch die diskreten relativen Orbits.

6 Phytoplankton der Nordsee in Beispielen

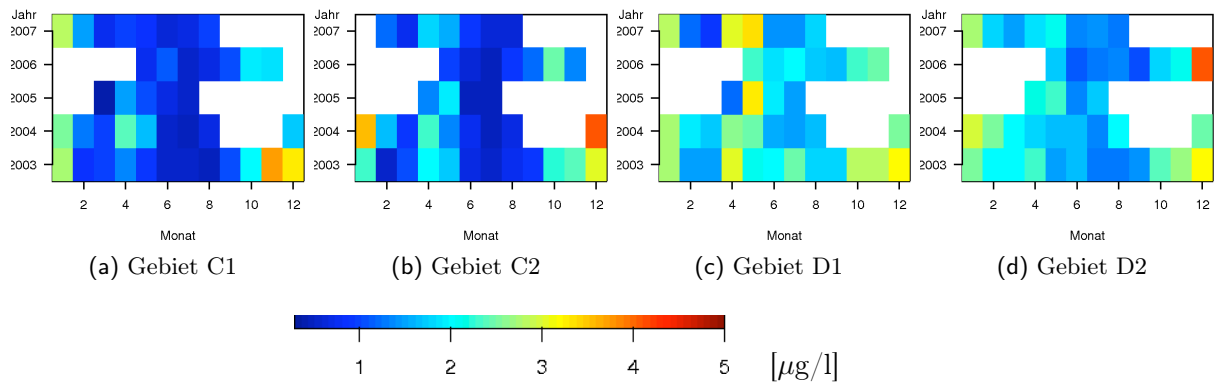


Abb. 6.6: Mittlere Chlorophyllkonzentration pro Monat und Jahr, Standardprozessor *algal_2*. Die Definition der Gebiete findet sich in Abb. 2.5.

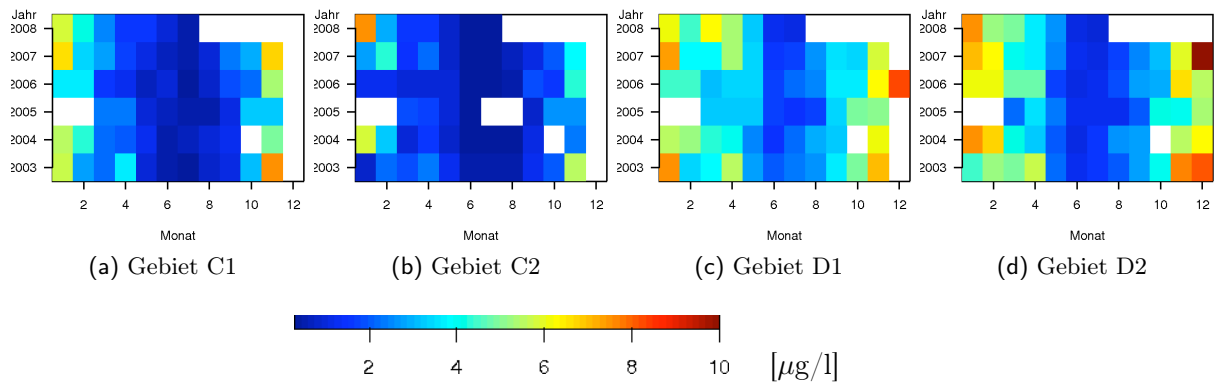


Abb. 6.7: Mittlere Chlorophyllkonzentration pro Monat und Jahr, Regionalprozessor *chl_conc*. Die Definition der Gebiete findet sich in Abb. 2.5.

den kleinräumigeren und zeitlich höher aufgelösten Zeitreihen durchaus Herbstblüten zu identifizieren sind (vgl. Tabelle 5.3).

6.3 Wiederkehrende Strukturen

Die Satellitendaten lassen eine Reihe von Entdeckungen jährlich wiederkehrender Phytoplanktonblüten zu, die sich bereits im klimatologischen Jahresmittel abzeichnen. Sie sollen hier exemplarisch für drei Regionen bildhaft dokumentiert werden.

6.3.1 Doggerbank

Im Frühjahr tritt über der Doggerbank eine Phytoplanktonblüte auf, die zwei Grundmuster aufweist. Die Blüte erscheint im März eher als zusammenhängende Fläche über dem Plateau im Flachwasserbereich (vgl. Abb. 6.8a), während sie im April hauptsächlich an

6 Phytoplankton der Nordsee in Beispielen

Datum	8./9.3.2007	30./31.3.2008	10./11.4.2005	17./18.4.2006	18./19.4.2008	21./22.4.2008
Sonnenazimut	155.9°/147.4°	156.9°/147.6°	156.8°/146.9°	159.3°/149.1°	157.3°/147.1°	159.2°/148.7°
Bildhälfte	L / L	R / L	R / L	R / L	R / L	R / L

Tab. 6.2: Sonnenazimut der Paare von unmittelbar aufeinander folgenden Datentagen jeweils am Punkt 2.5° O, 54.75° N (Mitte von Gebiet 1, Doggerbank) und die Angabe, in welche Bildhälfte der Punkt fällt (L: links, R: rechts)

den Flanken mit gleichzeitigem Minimum über der Sandbank auftritt (vgl. Abb. 6.8b). Die Abbildungen zeigen den Ausschnitt von 1-5.4° Ost und 54-56.1° Nord. Jede Zeile repräsentiert ein Jahr von 2003 bis 2008, während die Spalten Zeiträume zwischen ein und vier Tagen zusammenfassen. Die benachbarten Abbildungen von Einzeltagen innerhalb einer Zeile sind also nicht äquidistant in der Zeit, sondern so zusammengefasst, dass möglichst wenige Lücken entstehen. Einige Paare von Datentagen zeigen jedoch aufeinander folgende Tage und teilweise auch deutlich die Versätze im Chlorophyllprodukt. Diese Bildpaare werden mit einem roten Rahmen hervorgehoben.

Derartige Beispiele finden sich im März (4. Zeile: 8./9. März 2007; 30. und 31. März 2008, unterste Zeile ganz rechts im März und unterste Zeile ganz links im April), sowie im April (3. Zeile: 10./11. April 2005, 4. Zeile: 17./18. April 2006, letzte Zeile: 18./19. bzw. 21./22. April 2008). Das frühere Bild zeigt jeweils höhere Chlorophyllkonzentrationen, korrespondierend mit den höheren Sonnenazimutwinkeln (vgl. Tabelle 6.2) und der rechten Hälfte der Szene.

Die Dynamik der Muster ist von Jahr zu Jahr sehr variabel. Der Übergang von großräumigen Blüten über der Doggerbank zu hohen Konzentrationen an ihren Flanken setzt ab Ende März ein.

Die Entstehung der Frühjahrsblüten lässt sich durch folgenden Mechanismus erklären: in der durchmischten Wassersäule über der Sandbank reichern sich im Winter Nährstoffe an. Sobald die Lichtbedingungen Wachstum erlauben, beginnt die Blüte gemäß der kritischen Tiefe im Flachwasserbereich der Doggerbank. Die Wassertiefe ist hier so gering, dass die Lichtverhältnisse über die gesamte Wassersäule nicht zum limitierenden Faktor werden. Sobald die Nährstoffe im Flachwasser aufgebraucht sind, setzt sich die Blüte an den Flanken da fort, wo nährstoffreicheres Wasser in der Tiefe mit ausreichender Einstrahlung zusammentrifft. Dieser Zeitpunkt fällt auch mit dem Beginn der Schichtung zusammen (vgl. Struktur der Wassersäule im Nord-Süd-Tiefenprofil mittels der Wassertemperatur, z.B. Kröncke und Knust (1995), S.338), wenn nährstoffreiches Wasser an den Flanken aus der Tiefe von Norden oder Süden aufsteigt und für den Nachschub an Nährstoffen sorgt.

6.3.2 Schichtungsfront

Im Hochsommer sorgt die Schichtung für niedrige Chlorophyllkonzentrationen im Bereich der zentralen Nordsee, während in den (partiell) durchmischten Gebieten weiterhin hohe Chlorophyllkonzentrationen auftreten können. Das klimatologische Sommermittel und seine sehr geringe Varianz lassen Satellitenszenen erwarten, in denen sich direkt die

6 Phytoplankton der Nordsee in Beispielen

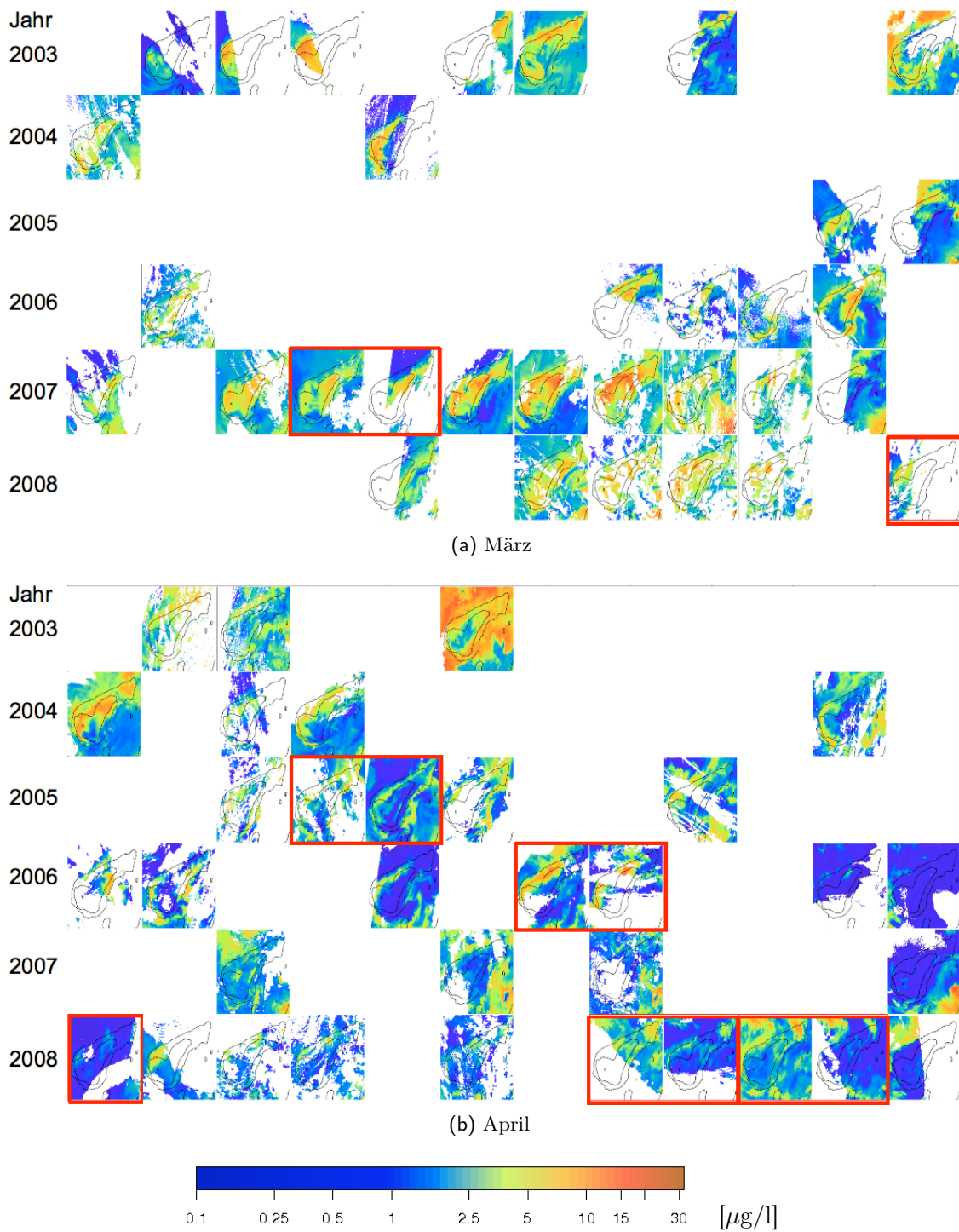


Abb. 6.8: Frühjahrsblüte des Phytoplankton über der Doggerbank anhand des Chlorophyllproduktes *chl_conc* des Regionalprozessors in $\mu\text{g/l}$. Zeilen von oben nach unten repräsentieren jeweils die Jahre 2003 bis 2008, Bildausschnitte (jeweils von 1 - 5.4° Ost und 54 - 56.1° Nord) in einer Spalte fallen in einen gemeinsamen Zeitabschnitt (Länge zwischen 1 und 4 Tagen). Paare in rotem Rahmen sind aufeinander folgende Tage.

Schichtungsfront widerspiegelt.

Die Angaben zum Sonnenazimut oder zu linker bzw. rechter Bildhälfte beziehen sich jeweils auf den Punkt 6° O, 55° N (Mittelpunkt in der zentralen Nordsee östlich der Doggerbank, Gebiet 3, vgl. Abschnitt 2.2.3).

Um die Daten vom 15. bis 18. Juli 2006 vergleichbar zu gestalten, wird der Schichtungsparameter über diesen Zeitraum gemittelt und in die Abb. 6.9 (c) - (e) eingetragen. Zwischen dem 15. und 16. Juli 2006 wird der Versatz der Chlorophyllkonzentrationen sichtbar.

Insbesondere im Jahr 2006 zeichnet sich ein starker Chlorophyllgradient ab, der teilweise recht gut mit der Linie $S = 2.9$ des Schichtungsparameters korrespondiert. Die beständig tidendurchmischten Bereiche (gestrichelte Linie zur Küste hin) fallen meist mit Strukturen erhöhter Chlorophyllkonzentration zusammen (z.B. vor der Emsmündung am 15./16. Juli 2006). Die Einflüsse von haliner Schichtung durch Flusseinträge von Süßwasser sind vermutlich für einige der Abweichungen vom vorausgesagten Schichtungsfrontverlauf verantwortlich. Dass sich die Bereiche hoher Konzentration über die Linie hinaus vor der nordfriesischen und dänischen Küste fortsetzen, wird auf den Transport durch den Reststrom zurückzuführen sein.

6.3.3 Norwegische Rinne

Bereits im Februar erfolgt in der Norwegischen Rinne eine auffällige Phytoplanktonblüte, die zumeist aus Coccolithophoriden besteht.

Als Vergleichsgröße zur Abschätzung des möglichen Fehlers in den Absolutwerten wird zu jedem Datentag der Sonnenazimutwinkel am Punkt 7° O, 56.5° N sowie dessen Lage links oder rechts in der Szene vermerkt. Die Absolutwerte der Chlorophyllkonzentrationen sind erneut mit Vorsicht zu interpretieren.

Dass es sich um eine jährlich wiederkehrende Struktur handelt, verdeutlichen die Beispiele aus verschiedenen Jahren (Abb. 6.10). Die Blüte setzt bereits Anfang Februar ein und bleibt bis in den März zu erkennen. Sie folgt den Tiefwassergebieten mit ihrer halinen Schichtung.

6 Phytoplankton der Nordsee in Beispielen

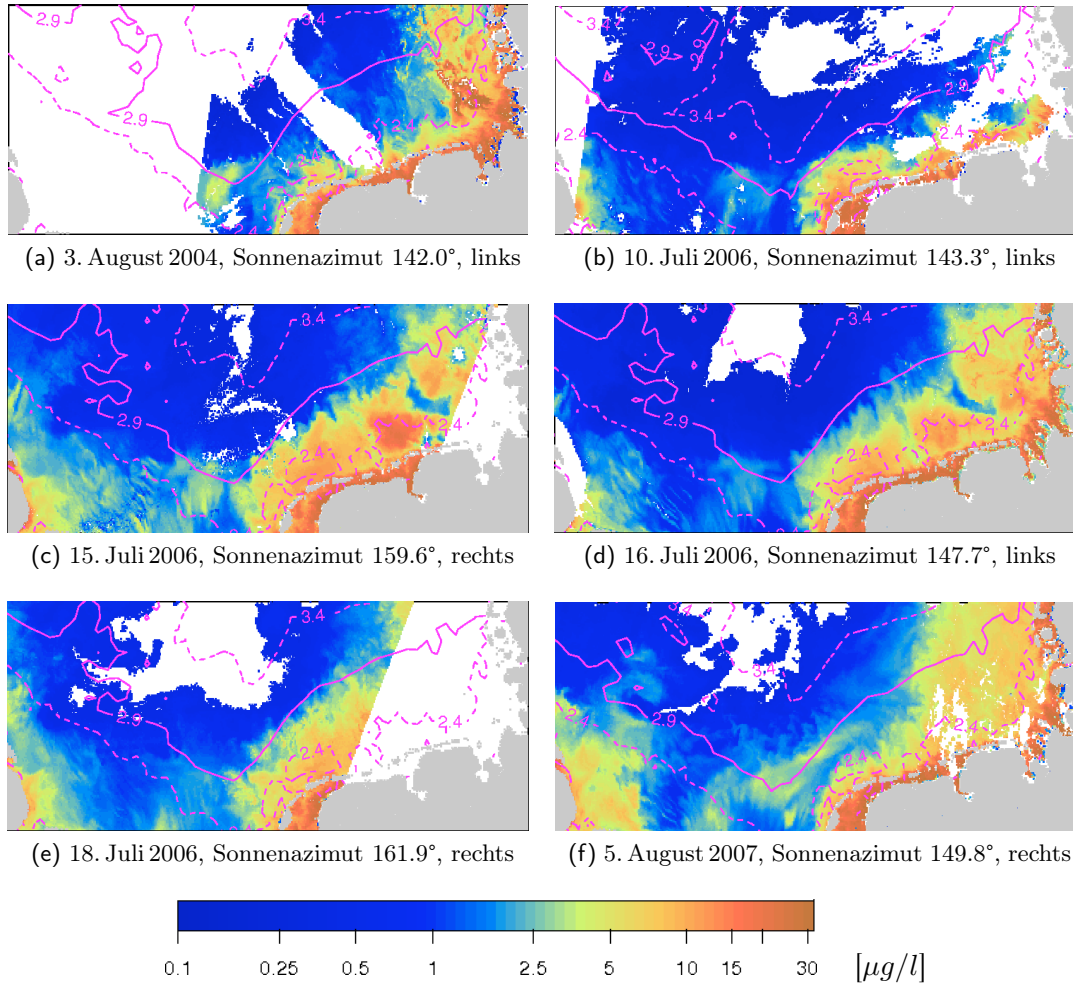


Abb. 6.9: Schichtungsfront in der südlichen Nordsee und das Chlorophyllprodukt chl_conc in $\mu\text{g}/\text{l}$. Sonnenazimut und Festlegung der Bildhälfte jeweils für den Punkt 6°O , 55°N . Die Karten zeigen den Ausschnitt von $0-9^\circ \text{Ost}$ und $53-55^\circ \text{Nord}$. Als durchgezogene Linie wurde der Schichtungsparameter $S = 2.9$ eingetragen, die gestrichelten Linien geben Bereiche partieller Schichtung wieder.

6 Phytoplankton der Nordsee in Beispielen

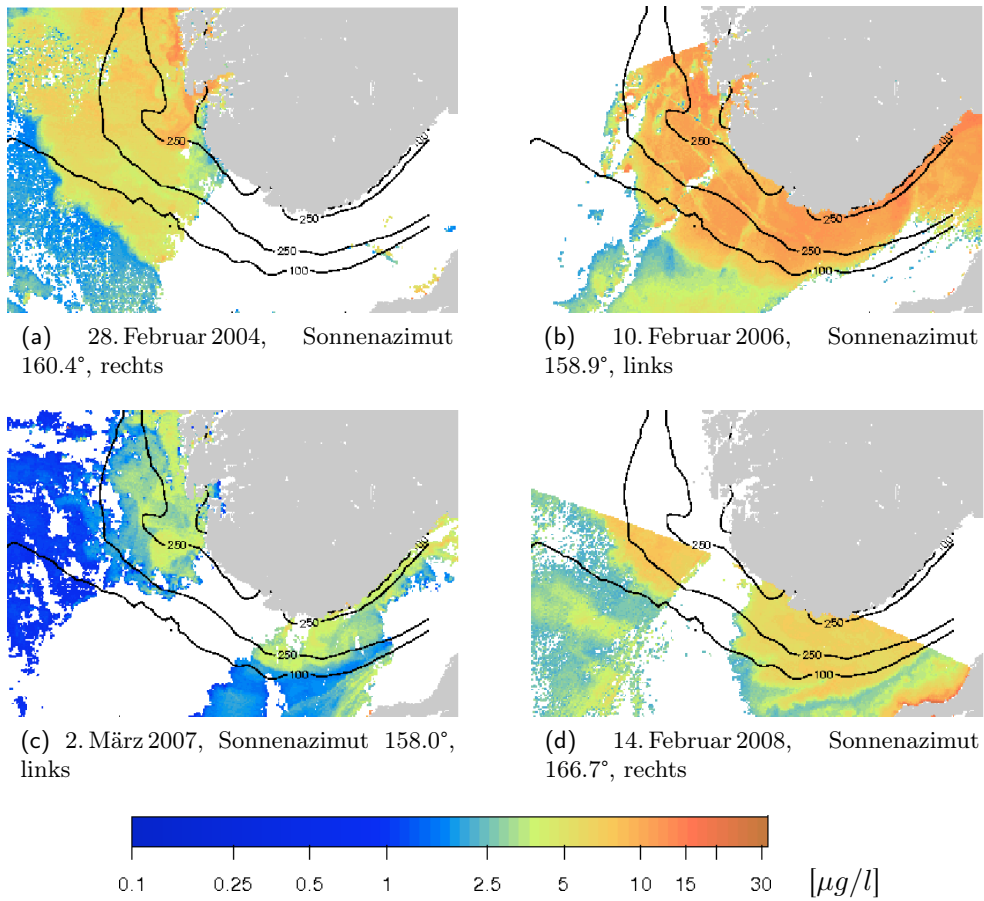


Abb. 6.10: Frühjahrsblüte Norwegische Rinne in $\mu\text{g}/\text{l}$. Die Karten zeigen den Ausschnitt von 2-10° Ost und 57-60° Nord.

Teil III

Räumliche Methoden zur Aufbereitung von Datenlücken

Übersicht der Verfahren und Methoden

Die Beispiele einzelner Szenen (Kapitel 6) vermitteln einen guten Eindruck des räumlichen Detailreichtums, der mit Satellitendaten erfasst werden kann. Durch zeitliche Mittelung werden diese Strukturen immer an Klarheit verlieren oder gänzlich verloren gehen.

Für Untersuchungen mit hoher zeitlicher und räumlicher Auslösung, wie sie für Ökosystemanalysen notwendig sein können, ist es daher wünschenswert, ein gutes Schätzwertverfahren zu finden, mit dem sich fehlende Datenpunkte unter Erhaltung der räumlichen Strukturen ergänzen lassen. Insbesondere für Karten ist die Vervollständigung der Datenlücken sinnvoll, um dem Betrachter einen geschlossenen Eindruck zu vermitteln und die Interpretation zu erleichtern.

Jede Szene in sich erscheint in ihren räumlichen Strukturen konsistent, sodass ein Schätzwertverfahren, das in erster Linie auf die räumliche Information zurückgreift, gute Ergebnisse liefern sollte.

Methoden, die zeitliche Aspekte berücksichtigen, werden unweigerlich Schätzwerte produzieren, die Artefakte durch die Unzuverlässigkeit des Absolutwertes in der Zeit enthalten.

Mit diesen Fragestellungen wird sich der zweite Teil dieser Arbeit eingehend befassen.

Den fehlenden Daten lässt sich mit unterschiedlichen Vorgehensweisen begegnen, die jeweils individuelle Vor- und Nachteile besitzen. Die Frage, wie der Datensatz sinnvoll zu bearbeiten ist, um die notwendige räumliche und zeitliche Auflösung zu erhalten bzw. zu bewahren, stellt sich für jeden zu beschreibenden Prozess neu. Belastbare Ergebnisse können die Verfahren nur dann liefern, wenn sich die Auflösung an den Skalen der zu studierenden Prozesse orientiert.

Einfache pixelweise Mittelwertbildung(en) über einen definierten Zeitraum sind häufig das Mittel der Wahl (Kapitel 7). Populär ist eine Zusammenfassung der Chlorophyllverteilung zu großflächigen Karten mit monatlicher Auflösung, da in diesem Zeitintervall annähernd vollständige Überdeckung z.B. der Nordsee zu erwarten ist.

Das althergebrachte Krigingverfahren - ursprünglich in der Lagerstättenerkundung entwickelt, um von Punktmessungen auf flächige Verteilungen zu schließen (Krige (1951); Matheron (1963)) - wird hier in seiner Anwendung und Anwendbarkeit untersucht.

Nach der theoretischen Einleitung (Kapitel 8) folgen vielfältige praxisnahe Analysen zur Bewertung der verschiedenen Krigingmethoden. Zuerst steht nur das Variogramm als Maß räumlicher Abhängigkeit der Daten im Vordergrund. Später folgen einige Kreuzvalidierungsversuche, die die Abhängigkeit der tatsächlichen Fehler von den gewählten Koordinaten, der Variogrammrealisation und der Krigingmethode verdeutlichen. Dazu wird der Datensatz auf die Deutsche Bucht beschränkt (Kapitel 9).

Um die Beziehung zwischen Wolkenstrukturen und dem tatsächlichen Fehler der Schätzwerte herzustellen, werden großräumige Kreuzvalidierungsexperimente mit künstlichen

Wolkenmasken im Bereich der gesamten Nordsee (Kapitel 10) durchgeführt.

Von weiteren Methoden zur Bestimmung von Schätzwerten wird ein Dateninterpolationsverfahren basierend auf der Hauptkomponentenanalyse vorgestellt. Die Schätzwerte - und ihre Fehler - nach Krigingmethoden und nach der Dateninterpolation werden gegenüber gestellt (Kapitel 12).

Zu der Fülle von Verfahren, die hier nicht weiter betrachtet werden, gehören unter anderem die *Optimale Interpolation*, die vielfach in der Meteorologie angewendet wird, um z.B. Windfelder aus Messungen an Stationen auf ein regelmäßiges Gitter abzuleiten (Gandin (1963)). Das Verfahren ist sehr ähnlich dem Kriging, bezieht jedoch zeitliche Information in die Analyse ein. Die *Objektive Analyse* ist ebenfalls eine Methode, die zeitliche und räumliche Information berücksichtigt, und sie wird genutzt, um Chlorophyllprodukte unterschiedlicher Satellitenmissionen miteinander zu verschneiden (Pottier u. a. (2006)). Auch spektrale Methoden zur Schätzwertbestimmung wurden vorgeschlagen (z.B. Analyse bei multipler Auflösung und Wavelet-Transformation, engl. *Multiresolution analysis and fast wavelet transform*, Pottier u. a. (2008)).

Durch den systematischen Fehler im Chlorophyllprodukt sollte derzeit auf die zeitliche Information verzichtet und nur die in sich konsistente(re) räumliche Information genutzt werden.

7 Mittelungsverfahren

Eine Reihe verschiedener Mittelungsverfahren sind vorstellbar, mit denen die Daten aus Satellitenszenen auf einem gemeinsamen Gitter vereinigt werden können. Z.B. ist die Projektion einer einzelnen Szene auf das definierte Gitter (z.B. für das in dieser Arbeit verwendete Standardgitter in Abschnitt 4.3) die einfachste räumliche Aggregation, da, sollten mehrere Originalpixel in ein Gitterpixel fallen, diese gemäß des Verfahrens zusammengefasst werden. Bereits diese einfachste aller Projektionen macht aus dem Level 2-Produkt mit seinen durch die Kamera des Satelliten definierten Pixel ein Level 3-Produkt.

Einige bekannte Verfahren und ihre Erweiterung um eine zeitliche Gewichtung sollen im Folgenden beschrieben werden.

7.1 Arithmetisches Mittel, Gewichtetes Mittel, Median

Zur Beschreibung einer normalverteilten Datenmenge kann auf den Mittelwert \bar{x} zurückgegriffen werden.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (7.1)$$

Statt der Gleichgewichtung - wie der arithmetische Mittelwert ($w_i = 1/N$) - lässt sich ein zeitlich gewichtetes Mittel konstruieren, das die zeitliche Distanz zwischen den Datenpunkten zu Zeitpunkten t_i und dem Bezugstag t_0 nutzt:

$$w(t_i) = \frac{1/(t_i - t_0)^2}{\sum_{i=1}^N 1/(t_i - t_0)^2} \Rightarrow \bar{x}_{gew} = \sum_{i=1}^N w_i \cdot x_i \quad (7.2)$$

Der Median ist bei einer Verteilung von Daten gerade der Wert, für den 50% der Stichprobe oberhalb und 50% unterhalb liegen. Bei einem normalverteilten (bzw. jedem symmetrisch verteilten) Datensatz entspricht der Median gerade dem Mittelwert. Der Vorteil der Verwendung des Medians statt des Mittelwertes ist seine Robustheit gegen einzelne Ausreißer im Datensatz.

7.2 Mittelwert und Median größter Wahrscheinlichkeit¹ - mit zeitlicher Gewichtung

Auf einem Gitter $G \in \mathbb{R}^{G_1 \times G_2}$ aus j Gitterpixel mit $j \in \{1, \dots, G_1 \cdot G_2\}$ wird der Mittelwert bzw. Median größter Wahrscheinlichkeit (abgekürzt aus dem Englischen als MLH-Mittelwert bzw. -Median) wie folgt berechnet (vgl. Dokumentation zum Level 3-Prozessor in BeamVISAT, Brockmann-Consult):

$$S_j = n_j^{w-1} \sum_i \ln(X_{i,j})$$

$$S_j^2 = n_j^{w-1} \sum_i \ln^2(X_{i,j})$$

Dabei bezeichnet $X_{i,j}$ die Werte der i Datenpunkte von einem festen Zeitpunkt t , die in das Gitterpixel j fallen. Die Aggregationssummen S und S^2 , Zwischenprodukte der Rechnung, besitzen die Dimension des Gitters. Das Gewicht n_j^{w-1} hängt bei jedem Gitterpixel j von

- n_j , der Anzahl von Datenpunkten $i = \{1, \dots, n\}$ in einem Gitterpixel j ab. Zu einem festen Zeitpunkt können je nach Definition des Gitters mehr als ein Datenpunkt in ein Gitterpixel fallen.
- w , Gewichtungskoeffizient; im Normalfall $w = \text{const.} = 0.5$.

Statt eines konstanten Gewichtungskoeffizienten wird eine zeitabhängige Funktion so gewählt, dass Datentage in zeitlicher Nähe zu einem zentralen Beobachtungstag t_0 stärker berücksichtigt werden:

$$w(t) = \frac{T}{|t - t_0| + T} \cdot 0.5 \quad (7.3)$$

Dabei ist

- T : maximale Zeitspanne zwischen den beteiligten Szenen [Tage]
- t_0 : Rekonstruktionstag [Tag des Jahres]
- t : Datentag [Tag des Jahres]

Für $t - t_0 = 0$ gilt: $w(t = t_0) = 0.5$; in allen anderen Fällen ist w kleiner als 0.5, das Gewicht $n^{w-1} = 1/n^{1-w}$ wird mit wachsender Distanz $|t - t_0|$ kleiner.

Für jeden Datentag t werden die räumlichen Aggregationssummen entsprechend angepasst:

¹engl. maximum likelihood mean/median

7 Mittelungsverfahren

$$S_j(t) = \sum_i n_j^{w(t)-1} \ln(X_{i,j})$$
$$S_j^2(t) = \sum_i n_j^{w(t)-1} \ln^2(X_{i,j})$$

In der zeitlichen Aggregation werden alle räumlichen Aggregationssummen, sowie Anzahl der Beobachtungen N pro Gitterpixel j und die Summe der Gewichte zum Zwecke der Normierung zusammengefasst:

$$\text{Sum}_{1,j} = \sum_t S_j(t); \text{Sum}_{2,j} = \sum_t S_j^2(t); \text{ebenso } N_j = \sum_t n_j(t); W_j = \sum_t n(t)_j^w.$$

Es gilt für die Größen $S_1(t), S_2(t), n(t)$ sowie die von ihnen abgeleiteten Größen $\text{Sum}_1, \text{Sum}_2, N, W$, dass sie jeweils die Dimension des Gitters besitzen.

Die weiteren Zwischenergebnisse sind

$$m_j = \frac{\text{Sum}_{1,j}}{W_j}$$
$$s_j^2 = \frac{\text{Sum}_{2,j}}{W_j} - m_j^2$$

Der MLH-Mittelwert *mean* bzw. MLH-Median *median* ist dann definiert als:

$$\text{mean}_j = \exp(m_j + 0.5 \cdot s_j^2)$$
$$\text{median}_j = \exp(m_j)$$

7.3 Übersicht der Mittelungsverfahren

Für die verschiedenen Mittelungsverfahren werden im Nachfolgenden Abkürzungen benutzt:

Mean arithmetischer Mittelwert

gew.Mean gewichteter arithmetischer Mittelwert

MLH.Mean Mittelwert größter Wahrscheinlichkeit mit konstantem Gewicht $w = 0.5$

gew.MLH.Mean Mittelwert größter Wahrscheinlichkeit mit zeitabhängigen Gewichten

MLH.Median Median größter Wahrscheinlichkeit mit konstantem Gewicht $w = 0.5$

gew.MLH.Median Median größter Wahrscheinlichkeit mit zeitabhängigen Gewichten

Die arithmetischen Mittelwerte betonen die minimal bzw. maximal vorkommenden Werte, während die Mittelwerte größter Wahrscheinlichkeit unanfälliger gegen Ausreißer sind. Durch zeitliche Gewichtung wird ein dem Vergleichstag möglichst ähnliches Ergebnis zu erreichen versucht.

8 Kriging: ein exakter, linearer Schätzer

Bevor in nachfolgenden Kapiteln die praktischen Erwägungen des Kriging im Vordergrund stehen, sollen hier die mathematischen Grundlagen zum Ordinary Kriging (Abschnitt 8.1, vgl. Olea (1999)) und den Variogrammen (Abschnitt 8.2) thematisiert werden. Ich habe mich bewußt gegen eine Übersetzung der Krigingverfahren entschieden, da mir keine moderne deutschsprachige Literatur bekannt ist, die die verschiedenen Verfahren eindeutig benennt. In alten Lehrbüchern wird lediglich von "Kriging" gesprochen, da die Abgrenzung verschiedener Verfahren gegeneinander, die erst mit dem Einsatz von Computern entstanden sind, nicht notwendig gewesen ist, da sie noch nicht existierten.

8.1 Ordinary Kriging in der Theorie

Die Mathematik des Ordinary Kriging verlangt, dass die Daten folgende *Voraussetzungen* erfüllen:

- Die Stichprobe (d.h. der Datensatz) ist eine partielle Realisierung der Zufallsfunktion $Z(x)$, x bezeichnet den Ort.
- Die Zufallsfunktion $Z(x)$ ist in 2. Ordnung stationär; das heißt, der Erwartungswert ist im gesamten Untersuchungsgebiet konstant und die Kovarianz lediglich abhängig vom Abstand h der Datenpunkte:

$$E[Z(x)] = \text{const.} = m \quad (8.1)$$

$$\begin{aligned} E[(Z(x) - m)(Z(x+h) - m)] &= E[Z(x)Z(x+h)] - m^2 \\ &= \text{Cov}[Z(x), Z(x+h)] \\ &= \text{Cov}[x, x+h] = \text{Cov}(h) \end{aligned} \quad (8.2)$$

- Alternativ ist die schwächere Bedingung ausreichend, dass die Zufallsfunktion $Z(x)$ der intrinsischen Hypothese genügt.

$$\begin{aligned} E[Z(x)] &= m \\ \text{Var}[Z(x) - Z(x+h)] &= 2\gamma(h) \end{aligned} \quad (8.3)$$

Dabei ist jede Zufallsfunktion, die in 2. Ordnung stationär ist, auch intrinsisch; umgekehrtes gilt nicht. Gl. 8.3 ist die Definition des *Variogramms* $\gamma(h)$.

8 Kriging: ein exakter, linearer Schätzer

Das Ordinary Kriging definiert einen Schätzer $\hat{Z}(x_0)$ mit der Position des Schätzwertes x_0 als eine gewichtete Summe der Zufallsfunktion $Z(x_i)$ unter der Bedingung, dass die Summe der Gewichte λ_i gleich Eins ist.

Definition des Ordinary Kriging Schätzers:

$$\hat{Z}(x_0) = \sum_{i=1}^k \lambda_i Z(x_i) \quad \text{mit} \quad \sum_{i=1}^k \lambda_i = 1 \quad (8.4)$$

Die Schätzvarianz $\sigma^2(x_0)$ wird definiert als die Varianz der Abweichung zwischen Schätzwert und Wert der Zufallsfunktion und dient als Zielfunktion des Krigingproblems (Gl. 8.5). Dabei soll $\sigma^2(x_0)$ durch einen geeigneten Satz von Gewichten minimiert werden, und wird dann als *Krigingvarianz* bezeichnet.

Definition der Krigingvarianz:

$$\sigma^2(x_0) = \text{Var}[\hat{Z}(x_0) - Z(x_0)] \quad (8.5)$$

Mithilfe des Semi-Variogrammes - im Folgenden werden die Begriffe *Semi-Variogramm* und *Variogramm* synonym verwendet - lässt sich dieses Problem unter Voraussetzung der intrinsischen Hypothese lösen:

Definition des Semi-Variogrammes:

$$2\gamma(h) = \text{Var}[Z(x) - Z(x+h)] \quad (8.6)$$

Dabei bezeichnet man γ auch als *Semi-Varianz*. Aus der Definition der Krigingvarianz (Gl. 8.5) und der Definition des Krigingschätzers (Gl. 8.4) folgt:

$$\begin{aligned} \sigma^2(x_0) &= \text{Var}[\hat{Z}(x_0) - Z(x_0)] \\ &= \text{Var}\left[\sum_{i=1}^k \lambda_i (Z(x_i) - Z(x_0))\right] \end{aligned} \quad (8.7)$$

Für jede Zufallsvariable $Z(x_i)$ und beliebige Gewichte λ_i gilt:

$$\text{Var}\left[\sum_{i=1}^k \lambda_i Z(x_i)\right] = \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z(x_i), Z(x_j)) \quad (8.8)$$

sodass folgt:

$$\sigma^2(x_0) = \sum_i \sum_j \lambda_i \lambda_j \text{Cov}(Z(x_i) - Z(x_0), Z(x_j) - Z(x_0)) \quad (8.9)$$

Für jede intrinsische Zufallsfunktion Z besteht folgender Zusammenhang zwischen Variogramm und Kovarianz:

$$\gamma(x_i - x_j) = \gamma(x_i - x) + \gamma(x_j - x) - Cov(Z(x_i) - Z(x), Z(x_j) - Z(x)), \quad (8.10)$$

sodass wiederum für die Krigingvarianz folgt:

$$\begin{aligned} \sigma^2(x_0) &= \sum_i \sum_j \lambda_i \lambda_j \{ \gamma(x_i - x_0) + \gamma(x_j - x_0) - \gamma(x_i - x_j) \} \\ &= \sum_i \lambda_i \gamma(x_i - x_0) + \sum_j \lambda_j \gamma(x_j - x_0) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_j) \quad (8.11) \\ &= 2 \sum_i \lambda_i \gamma(x_i - x_0) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_j) \end{aligned}$$

In Gl. 8.11 wurde ausgenutzt, dass die Summe der Gewichte gleich Eins ist.

Die Bestimmung der Gewichte λ_i stellt sich als Extremwertaufgabe mit Nebenbedingungen dar, die über die Methode der Lagrange-Multiplikatoren gelöst werden kann. Die Lagrange-Funktion

$$L(\lambda_1, \lambda_2, \dots, \lambda_k; \mu) = \sigma^2(x_0) + 2\mu \left(\sum_i \lambda_i - 1 \right) \quad (8.12)$$

liefert nach partiellem Ableiten nach λ_i sowie dem Lagrange-Multiplikator μ und Gleichsetzung mit Null, der notwendigen Bedingung für eine Extremstelle, ein Gleichungssystem, mit dem die Gewichte berechnet werden können.

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} L(\lambda_1, \lambda_2, \dots, \lambda_k; \mu) &= 2\gamma(x_i - x_0) - 2 \sum_j \lambda_j \gamma(x_i, x_j) + 2\mu \equiv 0 \quad (8.13) \\ \frac{\partial}{\partial \mu} L(\lambda_1, \lambda_2, \dots, \lambda_k; \mu) &= \sum_i \lambda_i - 1 \equiv 0 \end{aligned}$$

Aus Gl. 8.12 und Gl. 8.13 folgt die Formulierung zur Berechnung der Krigingvarianz für den Schätzwert an der Stelle x_0 :

aus Gl. 8.11:

$$\sigma^2(x_0) = 2 \sum_i \lambda_i \gamma(x_i - x_0) - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_j) \quad (8.14)$$

aus Gl. 8.13:

$$\begin{aligned} \sum_j \lambda_j \gamma(x_i, x_j) &= \gamma(x_i - x_0) + \mu \\ \Rightarrow \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_0) &= \sum_i \lambda_i \gamma(x_i - x_0) + \mu \quad (8.15) \end{aligned}$$

$$\sigma^2(x_0) = \sum_i \lambda_i \gamma(x_i - x_0) - \mu \quad (8.16)$$

In der Formulierung Gl. 8.16 zeigt sich deutlich, dass die Krigingvarianz hauptsächlich vom Variogramm und damit von der Datenverteilung im Bezug auf den Ort des Schätzwertes abhängt. Die Krigingvarianz ist die gewichtete Summe aus den Variogrammwerten, die Gewichte sind die Kriginggewichte, die gemeinsam mit dem Lagrangeparameter μ unter der Nebenbedingung, dass ihre Summe gleich Eins sei, berechnet wurden. In der Anwendung wird die Krigingvarianz damit aus dem theoretischen Variogramm, d.h. dem angepassten Variogramm-Modell, bestimmt unter Verwendung der gleichen Gewichte, die auch zur Bestimmung des Schätzwertes genutzt werden.

8.2 Variogramme

8.2.1 Experimentelles Variogramm

Das Variogramm $\gamma(h)$ der Zufallsfunktion $Z(x)$ ist im Allgemeinen nicht bekannt, sondern kann nur durch die Stichprobe abgeschätzt werden. Für die Lösung des Krigingproblems ist es unumgänglich, letztlich das Variogramm als Funktion des (vektoriellen oder absoluten) Abstandes h zum Bezugspunkt beschreiben zu können.

Für eine Menge von N diskreten Datenpunkten werden dazu alle Paare von Punkten betrachtet. Diese Paare werden gemäß ihrer Distanz h in diskreten Distanzklassen zusammengefasst, sodass N_h Punktpaare innerhalb dieser Klassen zu finden sind. Diese Distanzklassenverteilung N_h ist neben der Einteilung der Klassen hauptsächlich von der räumlichen Verteilung der Datenpunkte abhängig. Das experimentelle (omnidirektionale) Variogramm (Ribeiro und Diggle (2001)) berechnet sich dann als:

$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (Z(x_i + h) - Z(x_i))^2 \quad (8.17)$$

Dabei wird jegliche Richtungsabhängigkeit vernachlässigt, da einzig der Betrag der Distanz berücksichtigt wird.

Die Werte γ gezeichnet man auch als *Semi-Varianz*, die Funktion $\gamma(h)$ dagegen als Variogramm oder Semi-Variogramm.

8.2.2 Theoretisches Variogramm (Variogramm-Modell)

Die fünf Typen von Variogramm-Modellen, die im Weiteren zur Anwendung kommen, seien hier mit ihrer spezifischen Terminologie vorgestellt (vgl. Dutter (1985); Pebesma und Wesseling (1998); Pebesma (2004); Olea (1999)). In allen Fällen gilt:

Schwellwert C ist das Niveau der Varianz, das für weite Distanzen erreicht wird. Erreicht ein Variogramm für weite Distanzen ein derartiges Niveau, so ist dies ein Ausdruck der den Daten innewohnenden Stationarität (Rivoirard (1987)).

8 Kriging: ein exakter, linearer Schätzer

Reichweite a ist die Distanz, bei der der Schwellwert C (oder ein Prozentsatz davon) erreicht wird.

Distanz h ist der Abstand vom Bezugspunkt (Schätzposition) zu einem beliebigen Datenpunkt.

- Sei h die Distanz, dann ist das *sphärische Semi-Variogramm* (Matheron (1971)) definiert als

$$\gamma(h) = \begin{cases} C \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & , 0 \leq h < a \\ C & , a \leq h \end{cases} \quad (8.18)$$

mit Reichweite a und Schwellwert C .

- Sei h die Distanz, dann ist das *exponentielle Semi-Variogramm* definiert als

$$\gamma(h) = C \left(1 - \exp \left(-\frac{3h}{a} \right) \right) \quad (8.19)$$

Der Schwellwert wird asymptotisch erreicht.

- Sei h die Distanz, dann ist das *Gauß'sche Semi-Variogramm* definiert als

$$\gamma(h) = C \left(1 - \exp \left(-3 \left(\frac{h}{a} \right)^2 \right) \right) \quad (8.20)$$

Der Schwellwert wird asymptotisch erreicht.

- Sei h die Distanz, dann ist das *Bessel'sche Semi-Variogramm* definiert als

$$\gamma(h) = C \left(1 - \frac{h}{a} K_1 \left(\frac{h}{a} \right) \right) \quad (8.21)$$

mit K_1 als modifizierte Besselfunktion erster Ordnung und zweiter Art. Der Schwellwert wird asymptotisch erreicht.

- Sei h die Distanz, dann ist das *Semi-Variogramm des Nugget-Effekts* definiert als

$$\gamma(h) = C_0 (1 - H(0)) \quad (8.22)$$

C_0 ist der Nuggeteffekt, $H(0)$ steht für die Heaviside-Funktion, die den Wert 1 an der Stelle Null annimmt und ansonsten gleich Null ist. Diese Art des Modells steht für die Varianz, die zwischen extrem nah liegenden Punkten herrscht und wird im Falle der Satellitendaten durch das Bildrauschen verursacht.

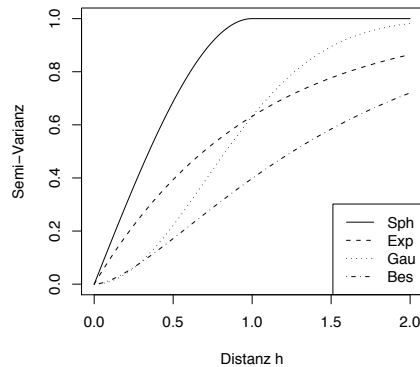


Abb. 8.1: Vier Variogramm-Modelle (sphärisch (Sph), exponentiell (Exp), Gauß'sch (Gau) und Bessel'sch (Bes)) mit Reichweite $a = 1$ und Schwellwert $C = 1$ nach Gl. 8.18 bis Gl. 8.21.

Die effektive Weite für die Modelle mit asymptotischen Verhalten ist die Distanz, bei der das Semi-Variogramm $0.95 C$ erreicht. Das exponentielle Modell erreicht diese Schwelle bei $3a$, das Gauss'sche bei $\sqrt{3}a$ und das Bessel'sche bei $4a$ (vgl. Abb. 8.1). C entspricht dabei der globalen Varianz des zugrunde liegenden Datensatzes.

8.2.3 Anisotropie

Anstelle von omnidirektionalen Variogrammen, die die Distanz als Skalar handhaben, können auch richtungsabhängige Variogramme gebildet werden, die neben dem Betrag die Richtung der Distanz berücksichtigen. In diesen Variogrammen können zwei Arten von Anisotropie auftreten:

Geometrische Anisotropie: Zeigt ein Variogramm in unterschiedlichen Richtungen die gleiche maximale Semi-Varianz, aber erreicht dieses Niveau zu verschiedenen Distanzen, so spricht man von geometrischer Anisotropie.

Zonale Anisotropie: Bei zonaler Anisotropie erreichen die Semi-Varianzen in unterschiedlichen Richtungen unterschiedliche Niveaus.

8.3 Cokriging

Im Cokriging wird zur Schätzung der Zufallsvariable $Z(x)$ eine Hilfsvariable $Y(x)$ hinzugezogen. Dabei können die Werte für exakt die gleichen Datenpunkte (engl. collocated oder isotopic) oder an unterschiedlichen Positionen (engl. heterotopic) vorliegen.

Auf eine ausführliche Darlegung der Theorie zum Cokriging wird hier verzichtet, es sei auf die Literatur (z.B. Myers (1982) oder Chiles und Delfiner (1999)) verwiesen.

Wenn in nachfolgenden Kapiteln Cokriging als Methode eingesetzt (z.B. in Abschnitt 9.5.8) und die Chlorophyllkonzentration mit der Hilfsgröße der Wassertiefe geschätzt wird, dann wird der isotope Fall hergestellt, in dem die Wassertiefe, die auf einem groberen Gitter aus dem Nordseeströmungsmodell des BSH zur Verfügung steht, zuvor auf den fehlenden Positionen des Standardgitters geschätzt wird. Auch wenn die Theorie zur Nutzung heterotoper Datensätze existiert (z.B. Rivoirard (2004)), ist Isotopie eine Voraussetzung für die Anwendbarkeit der Funktionen aus verfügbaren Software-Bibliotheken (Pebesma und Wesseling (1998); Ribeiro und Diggle (2001)).

8.4 Übersicht der Krigingverfahren

Auch die Krigingverfahren werden im Folgenden häufig abgekürzt:

OK Ordinary Kriging

ResOK Ordinary Kriging der Residuen

CoOK Ordinary Cokriging

ResCoOK Ordinary Cokriging der Residuen

Unter Residuum wird hier und im Folgenden die Differenz zwischen der Chlorophyllkonzentration des Datentages, für den die Schätzwerte zu berechnen sind, und einem Monatsmittel verstanden.

Die Abkürzungen werden daher meistens ergänzt durch die Bezeichnung der Kovariablen oder der Mittelwertmethode.

9 Kriging in der Praxis I - Deutsche Bucht

Bevor Schätzwerte für fehlende Datenpunkte ermittelt werden können, müssen sinnvolle Entscheidungen bezüglich der verschiedenen Parameter des experimentellen und theoretischen Variogramms und der Krigingmethode getroffen werden. Fragen, deren Antworten hier umrissen werden, sind z.B. die folgenden:

- Welche Datenpunkte sollen in die Berechnung des Variogramms einfließen und welchen Einfluss hat ihre Auswahl auf das Variogramm?
- Wie müssen die Art und Größe der Nachbarschaft festgelegt werden, aus der Punkte in die Berechnung des Schätzwertes einfließen sollen? Z.B. schreibt Wackernagel (2003): “In practice, because of the generally highly irregular spatial arrangement and density of the data, the definition of the size of a local neighborhood is not straightforward.”
- Wie müssen im Speziellen die Chlorophylldaten aufbereitet werden, um sie mittels Kriging weiter bearbeiten zu können?

Dieses Kapitel soll einige der Phänomene dokumentieren, die generell im Krigingverfahren auftreten können und bedient sich dabei der Illustration durch ein Beispiel für die Deutsche Bucht (Gebiet 4, vgl. Abb. 2.4) am 12. März 2007.

Kurzcharakterisierung der Daten Das Chlorophyllprodukt *chl_conc* des Regionalprozessors vom 12. März 2007 zeigt hohe Konzentrationen im küstennahen Bereich. Der nordöstliche Raum vor Sylt ist deutlich strukturiert (Abb. 9.1a).

Weder diese Konzentrationen noch ihre Logarithmen sind normalverteilt (vgl. Abb. 9.1 b, c). Durch die Korrektur mit dem gewichteten Monatsmittel (Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, ohne den Datentag, Abb. 9.1d) wird eine als normalverteilt zu bezeichnende Verteilung von Residuen geschaffen (Abb. 9.1 e, vgl. auch Gl. 9.1), die deutlich weniger räumliche Trends zeigt als die originale Chlorophyllverteilung (Abb. 9.1 f).

Ohne Berücksichtigung der Schiefe besitzen die Residuen den Mittelwert $\bar{res} = -0.0996$ (in $[\log \frac{\mu g}{l}]$), die Standardabweichung $res_{sd} = 0.1123$ (in $[\log \frac{\mu g}{l}]$) bzw. die (globale) Varianz $var_{glob} = 0.0126$ (in $[(\log \frac{\mu g}{l})^2]$). Sowohl das Monatsmittel als auch der Datentag sind auf dem Testgebiet mit $56 \times 89 = 4984$ Pixeln vollständig vorhanden.

9 Kriging in der Praxis I - Deutsche Bucht

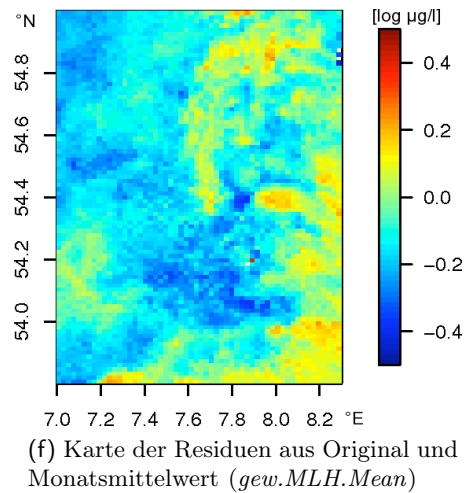
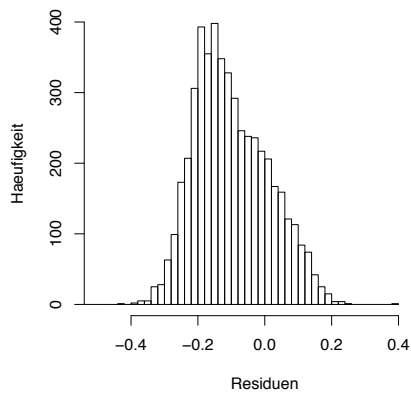
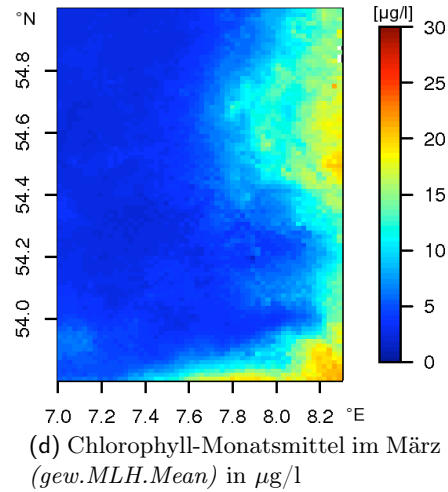
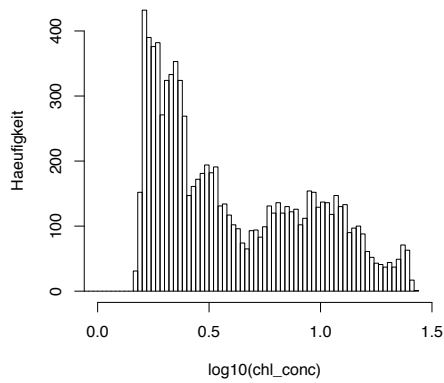
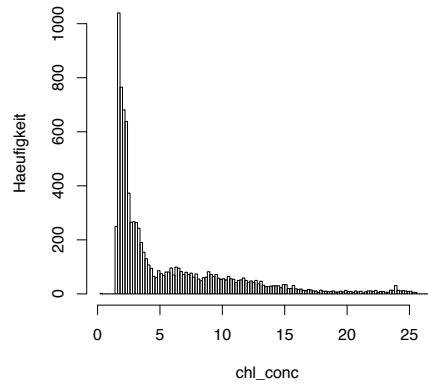
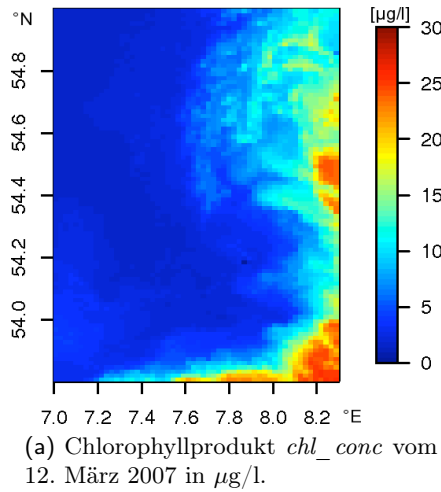


Abb. 9.1: Überblick über das Chlorophyllprodukt chl_conc des Regionalprozessors vom 12. März 2007 in der Deutschen Bucht (Gebiet 4) und dem Residuum nach der Trendkorrektur (Gl. 9.1).

9.1 Das experimentelle Variogramm

Das experimentelle Variogramm charakterisiert eine Menge von Datenpunkten dann am besten, wenn diese räumlich homogen verteilt liegen. Bei Satellitendaten ist eine homogene Verteilung nur in den wenigsten Fällen gegeben: die Daten werden durch Wolken unterbrochen oder räumlich gegliedert, weite Bereiche können ohne einen einzigen Datenpunkt sein, während an anderen Stellen die Daten dicht auf dem vorgegebenen Gitter liegen. Dabei hat ein Satellitendatensatz im Allgemeinen meist zu viele Datenpunkte, als dass alle in dieser Art der Zwei-Punkt-Statistik berücksichtigt werden könnten. Es ist daher notwendig, eine Auswahl zu treffen.

Wie stabil lässt sich das experimentelle Variogramm bei verschiedenen Probenahmefethoden bzw. von Realisation zu Realisation bestimmen?

Welchen Einfluss hat die Definition der Distanz auf die Stabilität des Variogramms?

Zwei Distanzdefinitionen bieten sich an: zum einen werden die Koordinaten in Pixelnummern mit der mittleren Kantenlänge eines Pixels als Distanz in km herangezogen (*Pixel-Distanz*), zum zweiten können die Koordinaten in Länge (Lon) und Breite (Lat) direkt als *Pseudo-Distanz* (oder *LonLat-Distanz*) verwendet werden, ohne auf die sphärische Geometrie d.h. insbesondere die Nicht-Gleichheit der Distanz in km von 1° Länge bzw. 1° Breite Rücksicht zu nehmen. Benachbarte Pixel mit mittlerer Kantenlänge von 1.5 km über die gesamte Nordsee ergeben eine Auflösung von $\Delta Lon = 0.0233^\circ$ und $\Delta Lat = 0.0135^\circ$.

Zur Vergleichbarkeit werden für alle experimentellen Variogramme fünfzehn gleichmäßige Distanzklassen verwendet, die durch die maximale Distanz einer halben Diagonallänge des rechteckigen Untersuchungsgebietes begrenzt sind.

Alle Untersuchungen zum experimentellen Variogramm nutzen als Zufallsfunktion $Z(x)$ die Residuen $res(x_i)$ am Gitterpunkt x_i gemäß

$$res(x_i) = \log chl_conc(x_i) - \log gew.MLH.Mean(x_i) \quad (9.1)$$

Die logarithmierte Chlorophyllkonzentration wird also mit dem zentrierten Monatsmittel verrechnet, um sowohl eine eher normalverteilte Zufallsgröße zu erhalten, als diese auch vom räumlichen Trend, der der Bedingung eines konstanten Erwartungswertes zuwider liefe, zu befreien.

9.1.1 Globales Variogramm

Das globale Variogramm - sprich jenes, welches aus *allen* Datenpunkten berechnet werden kann - dient im weiteren Verlauf als konstante Vergleichsgröße zu den Variogrammrealisationen aus Stichproben.

Omnidirektional und direktional Die omnidirektionalen Variogramme, die auf den zwei Distanzdefinitionen beruhen, sind sich qualitativ sehr ähnlich (Abb. 9.2).

Die direktionalen Variogramme verdeutlichen, dass zonale Anisotropie (vgl. Abschnitt 8.2.3) vorliegt, d.h. die Variabilität ist in verschiedene Raumrichtungen unterschiedlich groß.

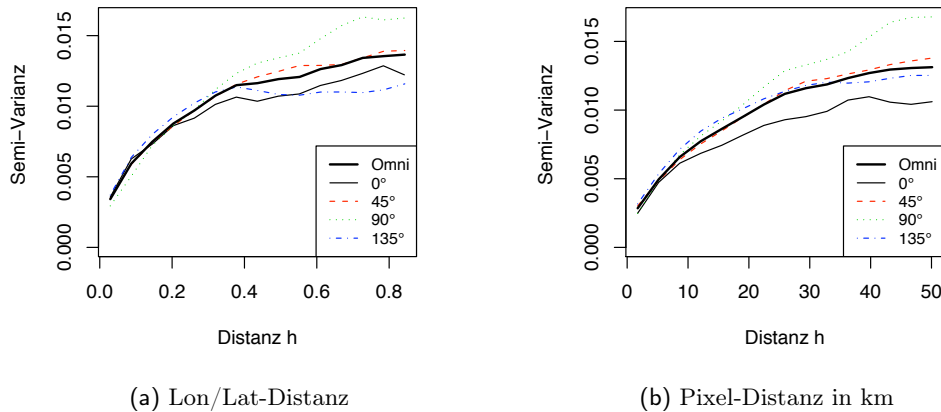


Abb. 9.2: Globales experimentelles Variogramm zur Beschreibung des vollständigen Datensatzes. Gezeigt wird das omnidirektionale und die direktionale Variogramme. Die Semi-Varianzen in den 15 Distanzklassen sind miteinander verbunden.

Besonders stark ausgeprägt ist sie für die Pixel-Distanzen im Intervall zur Nordrichtung ($\hat{=}$ 0°), bei LonLat-Distanzen sowohl für 0° , als auch für 45° .

Die beiden Koordinatendefinitionen sorgen für die Unterschiede zwischen den Variogrammen. Aus ihnen werden schließlich sowohl die Distanzen als auch die Winkel berechnet, wobei der Winkel als im Gegenuhrzeigersinn gegen Nord, sprich die y-Achse, definiert und damit den Verzerrungen im LonLat-Fall unterworfen ist. Die Winkelintervalle im LonLat-Fall sind schmäler gegen Nord als die gleichmäßigen Intervalle auf Pixelbasis.

Theoretisches Variogramm Das globale, omnidirektionale experimentelle Variogramm lässt sich am besten durch folgendes Variogramm-Modell (vgl. Abschnitt 8.2) anpassen: für die maximale Distanz der halben Diagonalen ($h_{max} = 0.8724$) wird das beste Variogramm-Modell aus $C_0 = 0.00062$, einem sphärischen Modell ($C = 0.00212$, $a = 0.0946$) und einem Exponentialmodell ($C = 0.01104$, $a = 0.2574$) zusammengesetzt (Abb. 9.3). Im Falle von Pixel-Distanzen ergeben sich die Parameter zu: $h_{max} = 76.8$ km, $C_0 = 0.00058$, sphärisches Modell ($C = 0.00195$, $a = 8.49$), Exponentialmodell ($C = 0.01194$, $a = 21.20$).

Die Güte der Anpassung durch die einzelnen Variogramm-Modelle gemäß dem Akaike Informationskriterium, AIC (vgl. Abschnitt 9.3.1), entnehme man Abb. 9.3.

9.1.2 Einfluss der Probenahme

Um Satellitenbilder der Nordsee mit etwa 300.000 Wasserpixeln bearbeiten zu können, kann nur eine Auswahl der sehr dicht liegenden Datenpunkte in das Variogramm einflie-

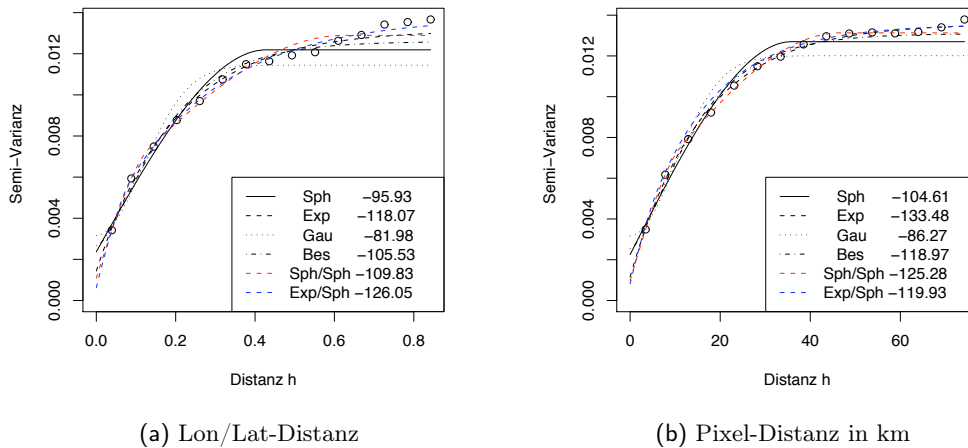


Abb. 9.3: Variogramm-Modelle (Linien) zur Anpassung des experimentellen globalen, omnidirektionalen Variogramms (Kreise) für maximale Distanz = halbe Diagonale des Untersuchungsgebietes. Zu jedem Modell wird die Güte der Anpassung nach Akaike Informationskriterium (AIC) in der Legende vermerkt. Die Abkürzungen der Modelle entsprechen denen in Abb. 8.1.

ßen. Denn aus N Wasserpixeln folgen

$$N_{Paare} = (N - 1) N/2$$

Punktepaare, was die Speichermöglichkeiten der verwendeten Software übersteigt.

Wie gut kann ein Variogramm basierend auf einer Teilmenge der Datenpunkte das globale Variogramm annähern?

Zwei Arten von Probenahmen werden gemeinsam mit den beiden Definitionen von Distanzen ausprobiert.

1. Bei der *Gitter*-Methode zerlegt ein auf dem Standardgitter (vgl. Abschnitt 4.3) basierendes, groberes Gitter das Untersuchungsgebiet in jeweils gleichgroße, nicht überlappende Unterräume. In diesen Fenstern mit der Kantenlänge von $w \times w$ Standardpixeln wird ein irgendein Pixel ausgewählt. Dementsprechend nimmt die Anzahl von verwendeten Datenpunkten für das Variogramm mit zunehmender Fenstergröße ab.
2. Die Datenpunkte können rein zufällig im gesamten Gebiet ausgewählt werden (*Zufall*-Methode). Diese Zufallsversuche nutzen jeweils die gleiche Anzahl von Punkten wie bei der *Gitter*-Methode, um den Informationsgehalt vergleichbar zu halten.

Für die einmal ausgewählten Punkte gemäß der beiden Probenahmeverfahren werden die jeweiligen Variogramme mit beiden Distanzen ermittelt. Die Probenahme wird hundertmal wiederholt, um eine statistische Aussage zur Stabilität des Variogramms je nach Größe der Stichprobenmenge treffen zu können.

Fenstergröße w	Gesamt	2	3	4	5	6	7	8
Pixelanzahl	4984	1232	522	308	187	126	96	77
Anteil an Gesamtpixel [%]	100	24.7	10.5	6.2	3.8	2.5	1.9	1.5

Tab. 9.1: Anzahl der zufällig ausgewählten Datenpunkte in Abhängigkeit von der Fenstergröße w

Grundlegend streuen die einzelnen Variogrammrealisationen sehr stark bezogen auf das globale, "wahre" Variogramm (Abb. 9.4). Je weniger Punkte in das Variogramm geflossen sind, umso größer ist die mögliche Abweichung der Semi-Varianz vom globalen Variogramm.

Der relative Fehler im Bezug auf das globale Variogramm wird für jede der hundert Variogrammrealisationen berechnet und anschließend seine Absolutwerte pro Fenstergröße und Probenahmemethode gemittelt. Unterschiede in den Fehlern in Abb. 9.5a und Abb. 9.5c entstehen daher hauptsächlich durch die Geometrie der Distanzdefinitionen.

Die Abweichungen der experimentellen Variogramme vom globalen Variogramm zeigen kein signifikant anderes oder besseres Verhalten in Abhängigkeit von der verwendeten Distanzdefinition (Abb. 9.5a, c). Wie zu erwarten wird der Fehler - und seine Streuung - mit größerem Gitter, d.h. geringerer Stichprobendichte, größer.

Der mittlere Fehler zum globalen Variogramm (je Fenstergröße und gemittelt über alle Distanzklassen) übersteigt für die zufällige Probenahme immer den für die Probenahme im Gitter (Abb. 9.5b, d). Jedoch sind alle Abweichungen nicht signifikant, da die Mittelwerte der Fehler immer in die Intervalle der Standardabweichungen der jeweils anderen Probenahmemethode fallen.

9.1.3 Konvergenz des Medians von Variogrammrealisationen

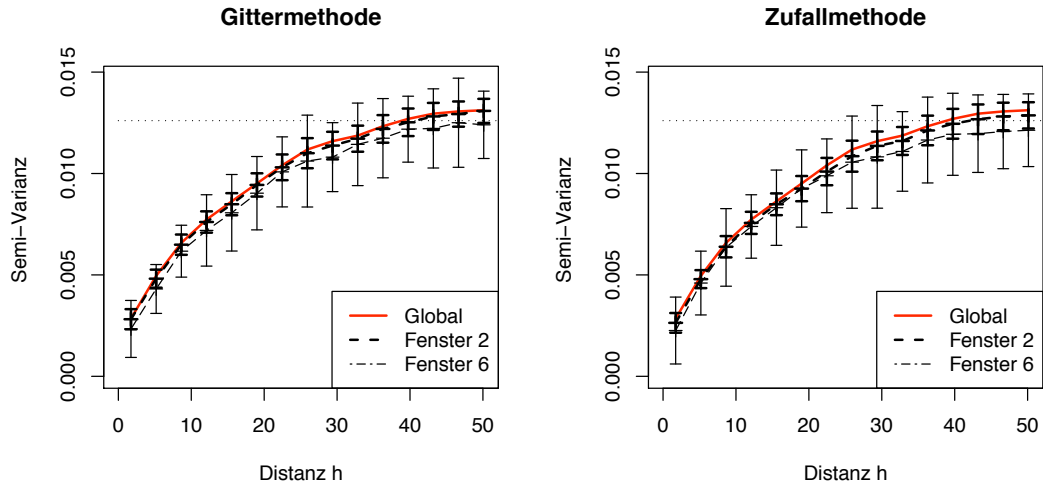
Variogrammrealisationen aus Stichproben aller Datenpunkte nähern sich dem globalen Variogramm nur bedingt gut an. Bereits die geringere Anzahl an Punktpaaren sorgt für eine veränderte Distanzklassenverteilung N_h .

In Abschnitt 9.1.2 wird daher mit einem mittleren Variogramm aus 100 Probenahmen gearbeitet, um eine mehr oder minder stabile Schätzung für das Variogramm zu bekommen. Ob 100 Probenahmen und ihre Variogramme zu einer guten Annäherung an das globale Variogramm führen bzw. wie viele Realisationen benötigt werden, um eine solche zu erhalten, wird hier überprüft.

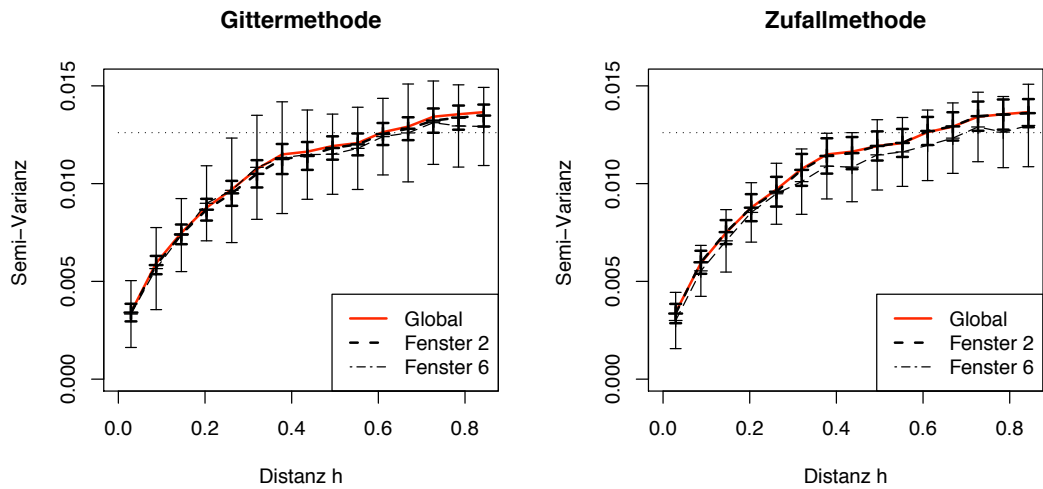
Dazu werden Variogramme für eine konstante Fenstergröße $w = 5$ bzw. eine entsprechende Anzahl zufällig gewählter Punkte und die LonLat-Distanz 10 bis 400 Realisationen bestimmt.

Unabhängig von der Anzahl N , die man als Median zusammenfasst, sind die Variogramme für diese Fenstergröße in der Konvergenz immer etwa 4% kleiner als der tatsächliche Wert (Abb. 9.6a).

Die Streuung der Variogrammabweichungen ist wie zu erwarten für die Probenahme mit Gitter im Allgemeinen geringer als für eine rein zufällige Pixelwahl (Abb. 9.6b).

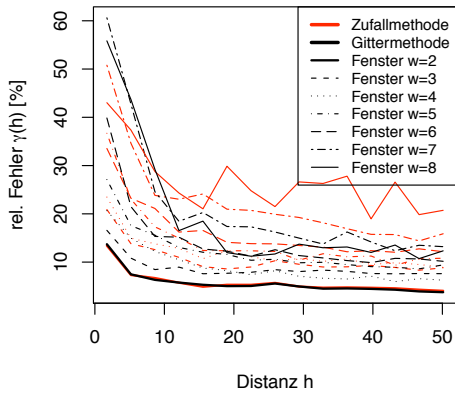


(a) Pixel-Distanz in km

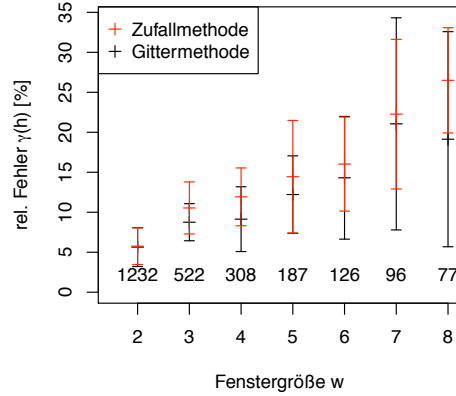


(b) Lon/Lat-Distanz in °

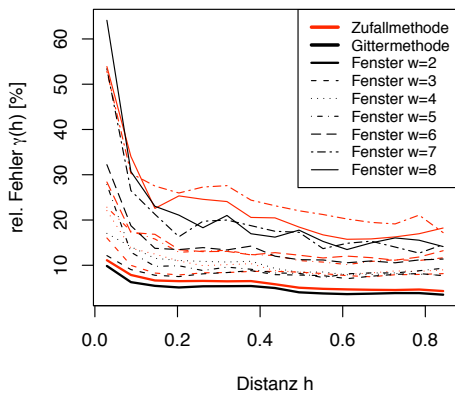
Abb. 9.4: Streuung von 100 Variogrammrealisationen abhängig von der Distanzdefinition für jeweils zwei Fenstergrößen. Die eingezeichneten Intervalle stehen für die Standardabweichung der Semi-Varianz in der jeweiligen Distanzklasse.



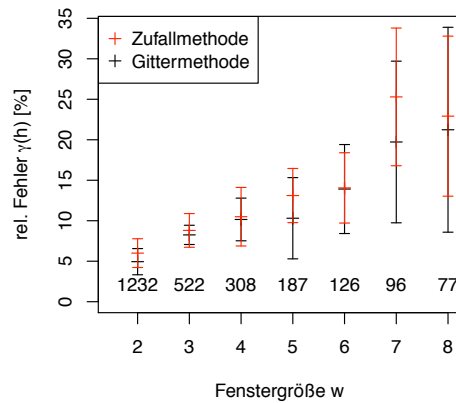
(a) Pixel-Distanz in km



(b) Pixel-Distanz

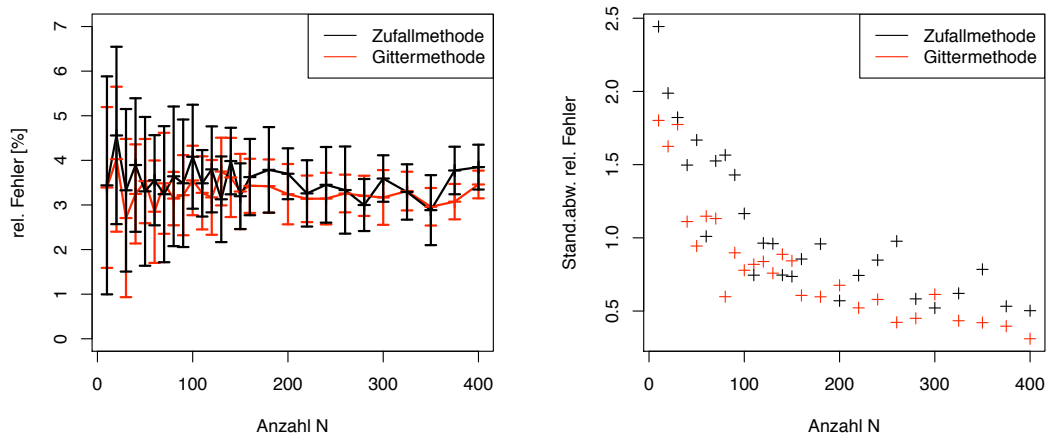


(c) LonLat-Distanz



(d) LonLat-Distanz

Abb. 9.5: Mittlerer Fehler der 100 Variogramm-Realisationen im Vergleich zum globalen Variogramm für verschiedene Arten der Probenahme in Abhängigkeit von der Distanz (a, c) sowie von der Fenstergröße (b, d). In das Diagramm geschriebene Zahlen zeigen die Anzahl ausgewählter Datenpunkte (vgl. Tabelle 9.1), aus der die Berechnung pro Variogramm-Realisation hervorgeht.



(a) Relativer Fehler des mittleren Variogramms (b) Standardabweichung der Fehler des mittleren Variogramms

Abb. 9.6: Konvergenz der mittleren Variogramme (a), als Funktion der Anzahl der Variogramrealisationen N bei einer Fenstergröße $w=5$, sowie die Standardabweichung des relativen Fehlers für die beiden Probenahmeverfahren (b).

9.1.4 Mittleres Variogramm aus unterschiedlichen Stichproben

Der Median des Variogramms aus 100 Realisationen nähert sich für die Auswahl auf dem Gitter für kleine Distanzen unabhängig von der Fenstergröße gut an das globale Variogramm an (Abb. 9.7).

Dieser Bereich ist für ein lokales Kriging, bei dem hauptsächlich Datenpunkte aus nächster Nachbarschaft um die Schätzposition berücksichtigt werden, von maßgeblicher Bedeutung.

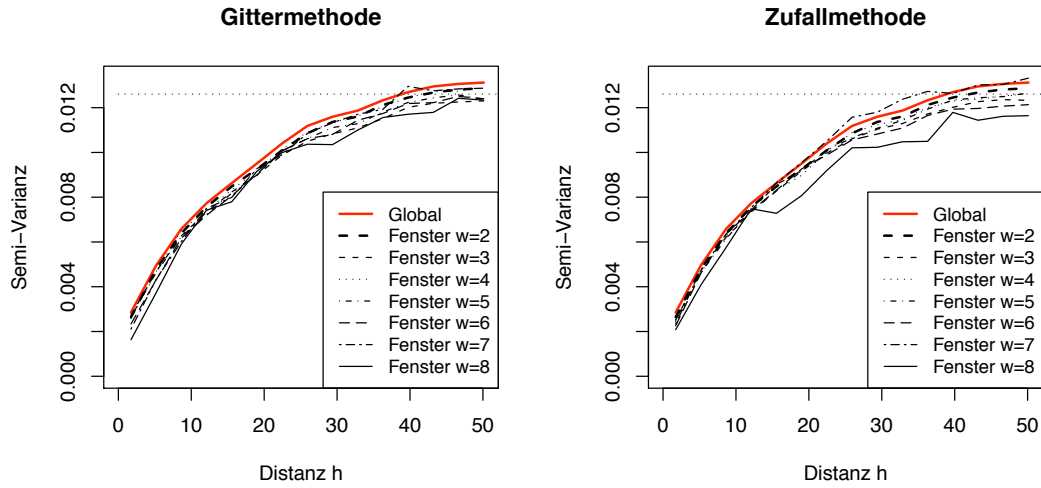
Das Ergebnis, die LonLat-Distanzen den Pixel-Distanzen vorzuziehen, folgt aus Eigenheiten dieses speziellen Datensatzes. Die Residuen zeigen deutlich mehr Variabilität in Ost/West- als in Süd/Nord-Richtung (vgl. dazu auch die direktionalen Variogramme in Abb. 9.2). Die starke Betonung der Ost-West-Komponente in den Distanzklassen sorgt dafür, dass die LonLat-Koordinaten hier als die besseren wahrgenommen werden.

9.1.5 Übertragbarkeit auf größere Datenmengen

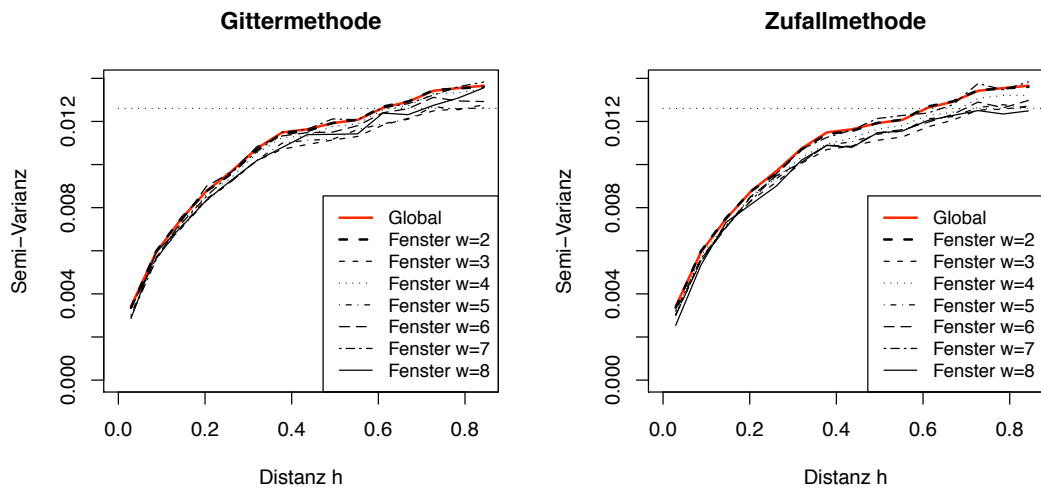
Für das Gebiet der Nordsee mit 304080 Wasserpixeln und einer angestrebten Zahl von 3000 Datenpunkten für ein Variogramm müsste die Fenstergröße des Auswahlgitters etwa $w=10$ betragen.¹ Die Festlegung auf 3000 Datenpunkte folgt aus Rechenzeitgründen.

Immer 3000 Datenpunkte zu wählen, unabhängig von der Anzahl vorhandener Daten, sorgt für ganz unterschiedliche Probenahmedichten in Abhängigkeit von Wolkenbede-

¹Abschätzung: $304080 \text{ Pixel} / 3000 \text{ Fenster} \approx 100 \text{ Pixel/Fenster}$. Bei quadratischem Fenster entspricht dies also einer Kantenlänge von $w=10$ Pixel.



(a) Pixel-Distanz in km



(b) Lon/Lat-Distanz

Abb. 9.7: Mittleres Variogramm als Funktion der Fenstergröße für verschiedene Distanzen im Vergleich zum globalen Variogramm. Dargestellt ist der Median aus 100 Realisationen.

ckung bzw. Datenverteilung. Nehmen wir für die Nordsee eine mittlere Wolkenbedeckung (über Wasser) von 40% an, so enthielte bei 3000 Fenstern mit gültigen Datenpunkten jedes Fenster etwa $0.6 \cdot 304080/3000 = 60.8$ Punkte und die Fenstergröße ließe sich zu $\sqrt{60.8} = 7.8$ Pixel abschätzen.

Eine Probenahme mit vergleichbarer Dichte bezogen auf die Gesamtpixelzahl (dies sind nur 1% der möglichen Gesamtzahl) bestünde im Beispielfall aus nur 49 Pixel, deutlich zu wenig, um die Region zu charakterisieren. Dass diese Fälle mit $w=8$ und größer starke Schwankungen zeigen, ist bei der geringen Punktauswahl nicht verwunderlich. Bei einer insgesamt größeren Datenmenge wird ein Variogramm basierend auf einer rein zufälligen Auswahl der Punkte aber auch für geringe Distanzen unkritisch sein. Solange die Auswahl zufällig erfolgt, werden auch nah beieinander liegende Punktpaare im gesamten Gebiet beprobt. Die räumliche Auflösung der experimentellen Variogramme wird sich jedoch bei gleich bleibender Distanzklassenanzahl und größerem Gebiet verringern, sodass die Genauigkeit der Beschreibung im Nahbereich im Sinne von weniger Distanzstützstellen abnehmen wird. Da mehr Punktpaare in die Statistik einfließen, kann die Genauigkeit der (experimentellen) Semi-Varianz mindestens erhalten bleiben.

9.2 Einfluss fehlender Daten auf das experimentelle Variogramm

Aus dem rechteckigen Untersuchungsgebiet wird mit einer Wolkenmaske eine Gruppe von 498 Datenpunkten ausgeschlossen, sodass 10% der Daten nicht länger zur Verfügung stehen (siehe Abb. 9.1f bzw. Abb. 9.8a).

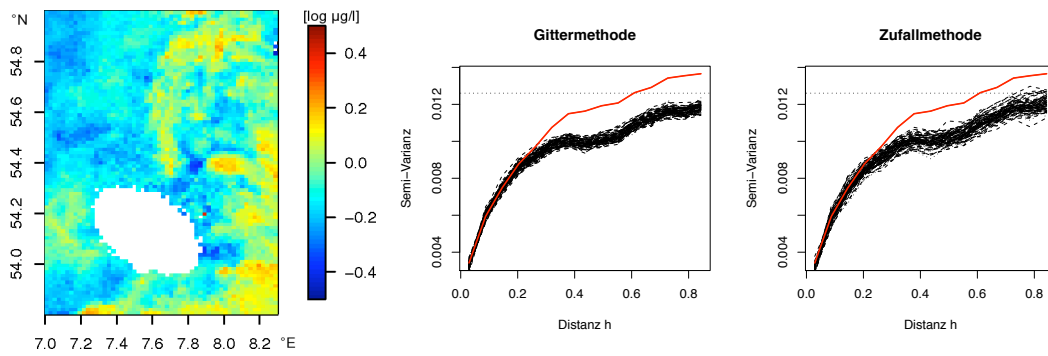
Mehrere Faktoren legen bei fehlenden Daten eine Veränderung des Variogramms nahe, die alle stark mit der Lage der Wolkenmaske zusammenhängen:

- Die räumliche Struktur der Daten hat Auswirkung auf die Distanzklassenverteilung N_h .
- Die Varianz der verbleibenden Daten var_{dat} kann größer oder kleiner als die tatsächliche globale Varianz var_{glob} sein. var_{dat} überschätzt die globale Varianz, wenn ein recht homogenes Gebiet aus den Daten ausgeschlossen wird, oder unterschätzt sie, wenn ein sehr variables Gebiet nicht länger berücksichtigt wird.

Im Folgenden werden alle Variogramme mit LonLat-Distanzen berechnet, die Auswahl der Datenpunkte erfolgt entweder mit der Gittermethode mit einer konstanten Fenstergröße von $w = 5$ oder einer entsprechenden Anzahl von zufällig gewählten Datenpunkten. Alle gezeigten Variogramme sind Mediane aus vierzig Realisationen. Die räumliche Struktur fehlender Daten wird konstant gehalten (Abb. 9.8a); das bezieht sich im ersten Versuch auf Form und Lage der Wolkenmaske, im zweiten rein auf die Form.

9.2.1 Einfluss der Probenahme bei konstant fehlenden Daten

Für eine feste Wolkenposition werden jeweils 40 Variogrammrealisationen zu einem mittleren zusammengefasst (Abb. 9.8).



(a) Residuen aus dem gewichteten Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, mit Wolkenmaske

(b) Variogramme für die Residuen aus dem gewichteten Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*.

Abb. 9.8: Einfluss konstant fehlender Daten auf die Variogramme. Zum Vergleich ist das globale Variogramm (rote Linie) eingezeichnet. Bei der Auswahl per Gittermethode beträgt die Fenstergröße $w=5$.

Die Auswahl per Gittermethode zieht eine geringere Streuung nach sich in mittleren und weiten Distanzbereichen. Gerade für die Nahbereiche nähert sich das gemittelte Variogramm sehr gut dem globalen an, unabhängig von der Probenahmemethode, während es die Semi-Varianz in größeren Entfernungen unterschätzt. Dieses Verhalten ist zu erwarten, da für die verbleibenden Datenpunkte die Varianz $\text{var}_{dat} = 0.0120$ gegenüber der globalen Varianz $\text{var}_{glob} = 0.0126$ geringer ist.

9.2.2 Einfluss der Positionierung der Wolkenmaske

Die Wolkenmaske wird im Untersuchungsraum verschoben und für 100 Positionen das Variogramm berechnet.

Ist die Varianz der verbleibenden Daten kleiner als die globale Varianz, so wird das aus 40 Realisationen gemittelte Variogramm in Abb. 9.9 grün eingezeichnet.

Es gibt keinen signifikanten Unterschied zwischen den beiden Probenahmemethoden. Beide streuen sehr weit um das globale Variogramm, wobei es die zu erwartende Tendenz gibt, dass ein Satz verbleibender Daten mit $\text{var}_{dat} < \text{var}_{glob}$ auch ein Variogramm besitzt, dass geringere Semi-Varianzen aufweist, als das globale Variogramm. Entsprechendes gilt für $\text{var}_{dat} > \text{var}_{glob}$.

9.3 Vom experimentellen Variogramm zum Variogramm-Modell

Zur Bestimmung der Kriginggewichte ist es unabdinglich, eine Funktion zur Beschreibung des experimentellen Variogramms zu finden, sodass für jede beliebige Distanz h die Semi-Varianz γ festgelegt ist.

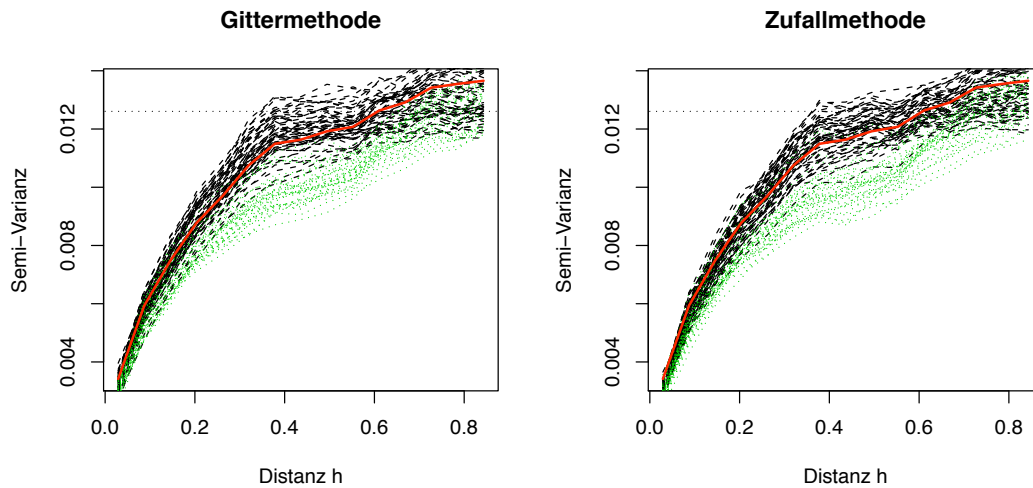


Abb. 9.9: Variogramme unter dem Einfluss von fehlenden Daten bei wechselnder Position der Wolkenmaske. Zum Vergleich das globale Variogramm (rote Linie). Ist die Varianz der verbleibenden Daten kleiner als die globale, sind die Variogramme grün eingezeichnet, $var_{dat} < var_{glob}$; im anderen Fall, $var_{dat} > var_{glob}$, werden schwarze Linien verwendet.

9.3.1 Akaike Informationskriterium

Um die Güte der Anpassung eines theoretischen Semi-Variogramms an ein experimentelles festzustellen, bietet sich das Akaike Informationskriterium (AIC , Akaike (1974) bzw. vgl. Olea (1999)) an. Es ist wie folgt definiert:

$$AIC = n \ln\left(\frac{R_m}{n}\right) + 2p \quad (9.2)$$

Dabei stehen die Variablen n, p, R_m für folgende Größen:

- n Anzahl von Distanzklassen im Variogramm, d.h. Punkte, die zur Anpassung des Modells zur Verfügung stehen.
- p Anzahl der Parameter im Modell. Die Variogramm-Modelle (vgl. Abschnitt 8.2.2) besitzen im Normalfall zwei Parameter (Schwellwert und Reichweite). Wird außerdem die Varianz direkt benachbarter Punkte angepasst (Nuggeteffekt), so erhöht sich die Anzahl auf drei Parameter. Bei zusammengesetzten Modellen wird die Anzahl entsprechend höher.

$R_m = \sum w \Delta^2$ die gewichtete Summe über die Quadrate der Differenzen zwischen der experimentellen Semi-Varianz und dem Wert des theoretischen Variogramm-Modells. Gewichtet wird mit dem Kehrwert der Varianz der Semi-Varianz in der jeweiligen Distanzklasse, $w = 1/var_{Klasse}$.

Das Modell mit dem geringsten AIC -Wert ist das beste.

9.3.2 Automatische Wahl des Variogramm-Modells

Um einen zügigen und automatischen Bearbeitungsprozess von Satellitendaten zu gewährleisten, wird eine Prozedur entworfen, die aus einem Datensatz gemäß des Akaike Informationkriteriums das beste Variogramm-Modell bestimmt.

Nach der eindeutigen Auswahl einer Stichprobe der Datenpunkte wird aus der Besetzung der Distanzklassen abgeschätzt, bis zu welcher Distanz eine Anpassung des Variogramms sinnvoll ist.

Da im weiteren Verlauf lokales Kriging (vgl. Abschnitt 9.4) verwendet wird, ist es nicht notwendig, das experimentelle Variogramm über die gesamte mögliche Distanzweite anzupassen; der Nahbereich ist ausreichend.

Die Distanzklassenverteilung N_h weist meist ein deutliches Maximum im geringen Distanzbereich auf und nimmt mit zunehmender Distanz ab. Es hat sich als praktikabel erwiesen, die zweifache Distanz des ersten, relativen Maximums in der Distanzklassenverteilung N_h oder die Hälfte der maximalen Distanz im Untersuchungsgebiet zu nutzen; die kleinere Distanz findet Verwendung.

Die Beschränkung der Distanzen kann auch Anisotropieeffekte ausschließen bzw. auf ein Maß verringern, sodass eine isotrope Betrachtung zulässig wird.

Angepasst werden Variogramm-Modelle vom Typ sphärisch, exponentiell, gauß'sch, Bessel, Nugget, und geschachtelte Modelle vom Typ sphärisch/sphärisch oder exponentiell/sphärisch (vgl. Abschnitt 8.2 insbesondere Abb. 9.3, Olea (1999)). Das Modell mit dem geringsten AIC-Wert (Abschnitt 9.3.1) wird im anschließenden Krigingverfahren genutzt.

9.4 Lokales Kriging

Zur Berechnung der Schätzwerte wird im Kriging eine gewichtete Summe der zur Schätzposition benachbarten Datenpunkte gebildet. Im *lokalen Kriging* wird diese Nachbarschaft explizit durch eine maximale Distanz um die Schätzposition oder eine feste Anzahl N_{max} nächster Nachbarn definiert.

Beide Definitionen besitzen ihre Vor- und Nachteile:

Während eine Nachbarschaft gemäß einer festen Distanz gleichzeitig eine maximale Distanz für die Bestimmung des Variogramms vorgibt, kann es bei fehlenden Datenpunkten zu sehr variablen Anzahlen von Datenpunkten innerhalb einer Nachbarschaft kommen. Dann berücksichtigt die in einen Schätzwert eingeflossene Information insbesondere die Beziehungen der Datenpunkte im Nahbereich zueinander, andererseits fallen bei zu klein gewählter Distanz eventuell kein oder nur sehr wenige Datenpunkte in die Nachbarschaft.

Eine Nachbarschaft aus einer festen Anzahl von Datenpunkten N_{max} führt einem Schätzwert immer die gleiche Menge an Information zu, jedoch kann bei fehlenden Datenpunkten die mittlere Distanz für jede Nachbarschaft stark zwischen unterschiedlichen Schätzwertpositionen variieren.

Hier ist der festen Anzahl von Punkten der Vorzug gegeben worden.

Wie die Schätzwerte von der Anzahl der verwendeten nächsten Nachbarn N_{max} bei konstantem Variogramm abhängen, soll hier ermittelt werden. Dazu werden basierend

auf allen Datenpunkten und ihrem globalen Variogramm Schätzwerte an 100 zufällig gewählten Positionen für unterschiedliche Anzahl von $N_{max} = \{10, 20, \dots, 150, 175, 200, 225, 250, 275, 300\}$ bestimmt. Der Datenwert an der jeweiligen Position wird für die Schätzung ausgeschlossen, da das Kriging sonst den exakten Funktionswert liefern wird. Diese Art von Kreuzvalidierung als direkter Vergleich von wahren Wert und Schätzwert findet sich in der englischsprachigen Literatur als "crossvalidation", sodass im Folgenden weitere ähnliche Untersuchungen mit "CV" abgekürzt werden. Die Nachbarschaft wird den Schätzpunkt in diesem Fall - außer bei Lage am Rande des Gebietes - von allen Seiten umschließen.

Als Krigingmethode dient im Nachfolgenden das Ordinary Kriging der Residuen, *Res-OK*, wobei die Residuen mit dem gewichteten Monatsmittel größter Wahrscheinlichkeit, *gew.MLH.Mean*, berechnet werden und die Distanz aus geographischer Länge und Breite abgeleitet wird.

9.4.1 Konvergenz der relativen Fehler der Schätzwerte

An jeder CV-Position konvergiert der Schätzwert mit zunehmendem N_{max} gegen einen Wert, der im Allgemeinen nicht dem tatsächlichen Wert entspricht. Es ist nicht absehbar, ob über- oder unterschätzt oder wie groß die tatsächliche Abweichung sein wird (vgl. Abb. 9.10a).

Zum Vergleich des Konvergenzverhaltens bietet es sich daher an, den relativen Fehler nicht auf den wahren Wert zu beziehen, sondern auf den Schätzwert mit einer Nachbarschaft von $N_{max} = 300$ Punkten. Für diese Anzahl an Datenpunkten ist im Allgemeinen (d.h. innerhalb dieses Beispiels mit seiner speziellen Geometrie nächster Nachbarn) Konvergenz eingetreten.

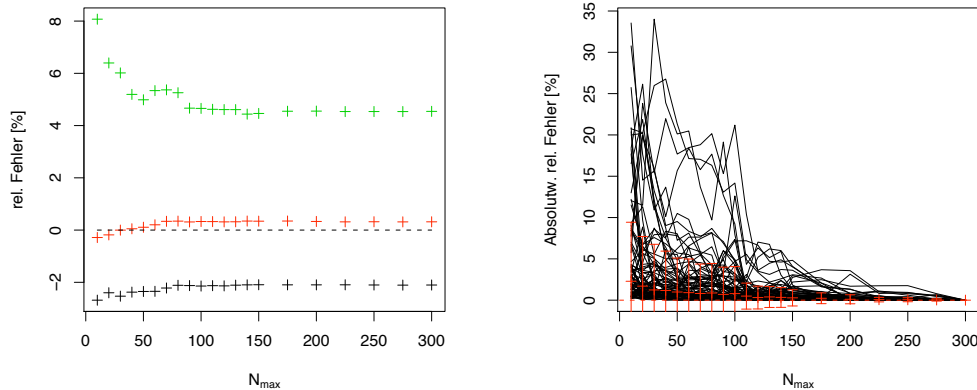
Um den so definierten relativen Fehler als reine Funktion der Größe der Nachbarschaft N_{max} betrachten zu können, wird zu jeder Klasse von N_{max} der Median aus dem Absolutwert des relativen Fehlers, bestimmt aus einer zufälligen Auswahl von 100 Kreuzvalidierungs-Punkten, berechnet. Dieser Median der relativen Fehler ist immer recht klein (vgl. Abb. 9.10, rote Markierungen, bzw. Tabelle 9.2). Ab welcher Anzahl N_{max} man Konvergenz postuliert und die entsprechende Anzahl für alle weiteren Berechnungen nutzt, bleibt dem Bearbeiter überlassen.

Da geringere Anzahl auch geringere Rechenzeit bedeutet, wird hier im Folgenden mit $N_{max} = 150$ gearbeitet, eine Schwelle, bei der in vielen Einzelverläufen das Plateau gerade erreicht wird. Die drei exemplarisch ausgewählten CV-Positionen könnten auch eine Anzahl von nur $N_{max} = 90$ rechtfertigen (Abb. 9.10a).

9.4.2 Konvergenz der Krigingvarianz

Der Versuch zur Konvergenz wird für die Krigingvarianz (Gl. 8.16) anstelle der Schätzwerte wiederholt. Hier bekommt die Lage des Schätzwertes und damit die Geometrie der nächsten Nachbarn Einfluss auf das Konvergenzverhalten.

Die Konvergenz der Krigingvarianz ist bei $N_{max} = 120$, sicherlich jedoch bei $N_{max} = 150$ erreicht (Abb. 9.11).



(a) Konvergenz der Schätzwerte für drei CV-Positionen, relativer Fehler zum jeweiligen tatsächlichen Chlorophyllwert

(b) Konvergenzverhalten der Absolutwerte des relativen Fehlers. Dieser ist für eine Position jeweils auf den Schätzwert mit $N_{max} = 300$ bezogen. Mediane und Standardabweichung in rot, ihre Zahlenwerte sind in Tabelle 9.2 dokumentiert.

Abb. 9.10: Konvergenz der Schätzwerte in Abhängigkeit von der Größe der Nachbarschaft N_{max}

Die Konvergenz kommt einzig durch die zunehmend vollständigere Abdeckung aller diskreten Raumrichtungen auf dem Gitter zustande. Betrachtet man nur eine Raumrichtung so gilt für Datenpunkte, die auf einer Geraden mit der Schätzposition liegen, dass derjenige in direkter Nähe zu dieser Position ein großes Gewicht erhält, während alle übrigen kleine Gewichte (nahe an Null) zugewiesen bekommen (Abschirmungseffekt, vgl. Wackernagel (2003)). Ab $N_{max} = 200$ erscheinen die Krigingfehler annähernd konstant, was dafür spricht, dass alle zusätzlichen Punkte nur noch Gewichte ungefähr gleich Null besitzen und daher verschwindenden Einfluss auf die Krigingfehler haben. Ab dieser Schwelle sind also alle diskreten Raumrichtungen abgedeckt.

Dieses Konvergenzverhalten hängt extrem von der Geometrie der Datenpunkte ab. In Fällen einer CV-Position benachbart zu einem Gebiet fehlender Daten kann eine wesentlich größere Anzahl nächster Nachbarn notwendig sein, um eine vollständige Winkelabdeckung zu gewährleisten.

9.4.3 Abhängigkeit der Schätzwerte vom Variogramm-Modell

An den 100 CV-Positionen werden für zwei unterschiedliche Variogramm-Modelle die Schätzwerte berechnet. Zum globalen, experimentellen Variogramm wird das am schlechtesten und das am besten angepasste Variogramm-Modell ermittelt (siehe Variogramm-Modelle in Abb. 9.3a). Die beste Anpassung gelingt mit einem zusammengesetztes Modell (Abschnitt 9.1.1), am schlechtesten eignet sich das Gauß-Modell mit den Parametern $C_0 = 0.0032$, $C = 0.0083$, $a = 0.1632$.

N_{max}	10	20	30	40	50	60	70	80	90	100	110
Median rel. Fehler [%]	2.29	1.70	1.24	1.17	0.99	0.91	0.81	0.87	0.70	0.80	0.49
Stand.abw. rel. Fehler [%]	7.14	6.01	5.52	4.76	4.11	4.07	3.56	3.57	3.30	3.27	1.57
N_{max}	120	130	140	150	175	200	225	250	275	300	
Median rel. Fehler [%]	0.32	0.39	0.34	0.28	0.23	0.11	0.065	0.055	0.031	0	
Stand.abw. rel. Fehler [%]	1.38	1.27	1.21	0.98	0.65	0.55	0.27	0.24	0.14	0	

Tab. 9.2: Konvergenzverhalten der Schätzwerte in Abhängigkeit von der Anzahl von Datenpunkten zu ihrer Berechnung, N_{max} . Der relative Fehler bezieht sich jeweils auf den Schätzwert für $N_{max} = 300$. Der Median dieses relativen Fehlers wird aus dessen Absolutwert an 100 zufällig gewählten Positionen bestimmt.

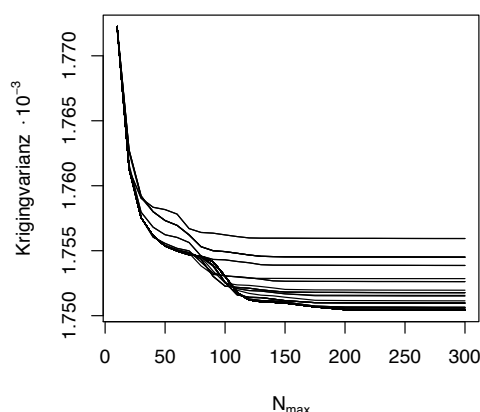


Abb. 9.11: Konvergenz der Krigingvarianz in Abhängigkeit von N_{max} für 100 Positionen.

Die absoluten Fehler sind für die gute Variogrammanpassung (im Median) geringer als für das schlechteste Variogramm-Modell. Für geringe N_{max} sind die Fehler noch vergleichbar, jedoch nehmen sie für das schlechte Variogramm mit steigender Anzahl von Punkten zu und konvergieren auf hohem Niveau. In beiden Fällen konvergieren die Fehler nicht im minimalen Wert (Abb. 9.12a). Die relativen Fehler (Median der Absolutwerte, Abb. 9.12b. Die hier dargestellten schwarzen Kreuze finden sich auch in Abb. 9.10b als rote Einträge mit Angabe der Standardabweichung) konvergieren für das schlechteste Variogramm-Modell nicht; sie sind immer deutlich größer als für den Fall des besten Modells.

9.5 Abhängigkeit der tatsächlichen Fehler von der Krigingmethode

Unter dem Begriff 'Krigingmethode' soll hier nicht ausschließlich die Unterscheidung zwischen Ordinary Kriging oder Cokriging verstanden werden, sondern auch die spezifische

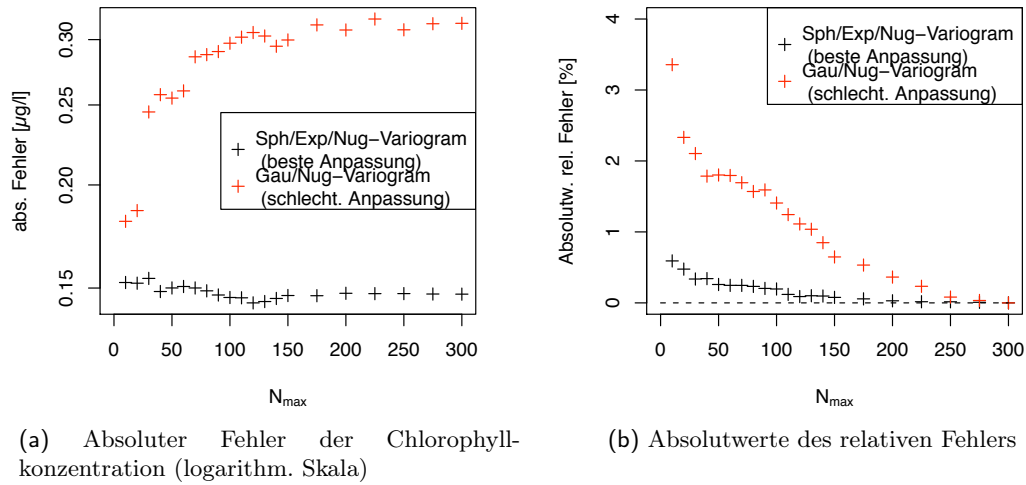


Abb. 9.12: Absoluter und relativer Fehler der geschätzten Chlorophyllkonzentrationen für das beste (schwarz) und das schlechteste (rot) Variogramm-Modell. Absolute Fehler beziehen sich auf den tatsächlichen Chlorophyllwert an der Schätzposition, relative Fehler auf den Schätzwert bei $N_{max} = 300$. Dargestellt wird der Median der Fehler von 100 willkürlich gewählten Schätzwertpositionen mit einer Nachbarschaft aus jeweils N_{max} Punkten.

Art der Trendkorrektur mit einem Monatsmittel und das Ordinary Kriging der entsprechenden Residuen.

Die Notwendigkeit zur Trendkorrektur entsteht durch den Anspruch an den Datensatz, die intrinsische Hypothese zu erfüllen, d.h. einen konstanten Erwartungswert zu besitzen. Für die Chlorophyllverteilung in der Nordsee gilt dies grundsätzlich erstmal nicht. Durch die Bildung der Residuen mit einem Monatsmittel soll dieser Zustand angenähert werden.

In den nachfolgenden Versuchen wird mit dem Ausschnitt der Deutschen Bucht vom 12. März 2007 und einer festen Wolkenposition gearbeitet (vgl. Abb. 9.8a), als Koordinaten der Datenpunkte wird Lon/Lat verwendet, die Daten für die Variogramme werden gemäß der Gittermethode mit einer Fenstergröße von $w = 5$ oder rein zufällig ausgewählt. Zum Kriging werden jeweils $N_{max} = 150$ nächste Nachbarn genutzt.

9.5.1 Lokale und globale Genauigkeit der Schätzwerte

Zur Einführung eines Maßes der lokalen und globalen Genauigkeit der Schätzwerte wird hier ein kurzer Einblick in das Zusammenspiel von Trendkorrektur und Krigingmethode gegeben.

In diesem Vorversuch werden für das Residuum aus dem gewichteten Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, das Ordinary Kriging der Residuen, *ResOK*, dem Ordinary Kriging des (logarithm.) Konzentrationen, *OK*, bzw. den Cokriging-Methoden, *CoOK* oder *ResCoOK*, gegenübergestellt.

Die Ordinary Kriging-Methoden beziehen sich auf:

9 Kriging in der Praxis I - Deutsche Bucht

Methode	ResOK	OK	CoOK.tief	CoOK.bin	CoOK.bin.tief	ResCoOK.bin
Güte der Methode (RG)	1.361	0.034	0.784	1.340	0.962	1.340
Güte der Methode (R)	1.334	0.008	0.694	1.313	0.921	1.313

Tab. 9.3: Zahlenbeispiel zur Güte der Krigingmethoden (Auswertung gemäß Gl. 9.4 zu Abb. 9.13) mit verschiedener Probenahmemethode für das Variogramm (rein zufällig, R , oder Gittermethode, RG).

- OK: $\log chl_conc \sim 1$. Variogramm und Kriging basieren auf der logarithmierten Chlorophyllkonzentration.
- ResOK: $res \sim 1$. Variogramm und Kriging basieren auf den Residuen, d.h. der Abweichung der Datenwerte zu einem um den Beobachtungszeitpunkt zentrierten Monatsmittel.

Die Cokriging-Modelle lauten:

- CoOK.tief: $\log chl_conc \sim tief$. Neben der logarithmierten Chlorophyllkonzentration wird auch ihre Korrelation mit der Tiefe zur Bestimmung der Schätzwerte genutzt.
- CoOK.bin: $\log chl_conc \sim \log \text{Monatsmittel}$. Statt die Residuen zu berechnen, wird hier die Korrelation zwischen dem Datentag und dem zentrierten Monatsmittel genutzt und die Schätzwerte auf dieser Basis berechnet.
- CoOK.bin.tief: $\log chl_conc \sim \log \text{Monatsmittel} + tief$. Die Korrelation mit dem Monatsmittel als auch mit der Wassertiefe wird zur Ermittlung von Schätzwerten verwendet.
- ResCoOK.bin: $res \sim \log \text{Monatsmittel}$. Als Hilfsvariable kommt das Monatsmittel zur Berechnung der Schätzwerte basierend auf den Residuen zum Einsatz.

Dabei steht *Monatsmittel* für den gewichteten Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, die Residuen *res* sind gemäß Gl. 9.1 definiert, bei *tief* handelt es sich um die Wassertiefe an der entsprechenden Stelle (vgl. Abschnitt 8.3).

Für diese Art von Wolkenmaske und Gebiet sowie gewählten Trend *gew.MLH.Mean* besitzt das Ordinary Kriging der log. Chlorophyllkonzentration – unabhängig von der Probenahme des Variogramms – die beste Rekonstruktion der Originalwerte (Abb. 9.13).

Als Fehler $f(x_i)$ wird die Differenz zwischen der tatsächlichen Chlorophyllkonzentration $chl_conc(x_i)$ und dem Schätzwert $estim_{Methode}(x_i)$ für einen Gitterpunkt x_i definiert:

$$f(x_i) = chl_conc(x_i) - estim_{Methode}(x_i) \tag{9.3}$$

Als beste Rekonstruktion wird diejenige verstanden, deren mittlerer absoluter Fehler $\overline{\|f\|} = \frac{1}{N} \sum_i \|f(x_i)\|$ aus allen Schätzwerten möglichst gering ist und deren Median der

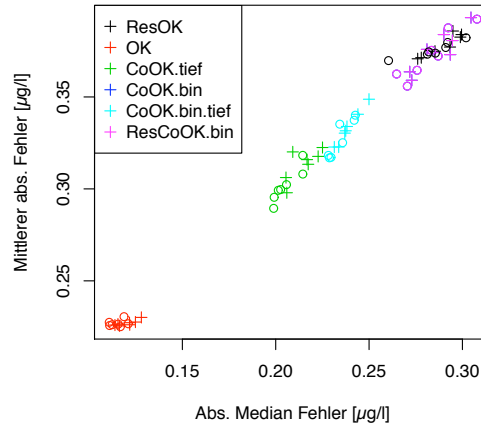


Abb. 9.13: Statistik tatsächlicher Fehler für eine Auswahl von Krigingmethoden und eine konstante Wolkenstruktur. Probenahmemethoden sind mit Kreisen für zufällige Wahl bzw. mit Kreuzen für die Gittermethode gekennzeichnet. Jeder Punkt in der Graphik repräsentiert die Fehlerstatistik einer Variogramm-Realisation, wobei für eine Probenahme Variogramme und Schätzwerte gemäß aller verwendeten Krigingmethoden berechnet werden. Das Residuum basiert auf *gew.MLH.Mean*. Die Quantifizierung der Güte der Methoden findet sich in Tabelle 9.3.

Fehler (Absolutwert) möglichst nah an Null zu liegen kommt. Anschaulich repräsentiert der Median des Fehlers die globale Eigenschaft der Rekonstruktion, die Originalwerte zu über- oder unterschätzen, während der mittlere absolute Fehler die lokale Genauigkeit repräsentiert. Die beste Rekonstruktion sollte globale und lokale Genauigkeit miteinander vereinen.

Die Güte der Methode soll über die Kombination von Fehlermedian und absolutem Fehler festgelegt werden. Das bedeutet: für jede Variogrammrealisation wird sowohl der mittlere absolute Fehler $\overline{\|f\|}$, als auch der Absolutwert des Medians der Fehler $\|\text{med}(f)\|$, beide als Funktion der Krigingmethode, auf ihren jeweiligen Wertebereich über alle Krigingmethoden normiert:

$$\overline{\|f\|} \in \left[\overline{\|f\|}_{min}; \overline{\|f\|}_{max} \right] \implies \overline{\|f\|}_{norm} \in [0; 1]$$

D.h. das Maximum von $\overline{\|f\|}$ und $\|\text{med}(f)\|$ ist danach gleich Eins, während das Minimum Null ist. Die Information, wie gut die spezifische Methode absolut dem angestrebten Fehlernullpunkt nahe kommt, geht damit verloren zugunsten eines reinen Bewertens der Methoden. Die normierten Fehlergrößen sind unabhängiger von der spezifischen Variogrammrealisation als es die absoluten Fehlergrößen wären - jedoch weit davon entfernt, tatsächlich statistisch unabhängig von der Variogrammrealisation zu sein. Insgesamt werden sie zur Güte G für jede Krigingmethode und jede Variogrammrealisation zusammengefasst:

$$G_{\text{Methode}} = \sqrt{\|f\|_{\text{norm}}^2 + \|\text{med}(f)\|_{\text{norm}}^2} \quad (9.4)$$

Die mittlere Güte aus allen Variogrammrealisationen gibt schließlich eine Rangfolge der Krigingmethoden; je kleiner dieses Gütemaß G ist, umso besser die Rekonstruktion.

Die tatsächlichen Fehlergrößen werden hier jeweils für eine konstante Auswahl von Datenpunkten zur Berechnung des Variogramms gemäß einer Probenahmemethode für alle zu vergleichenden Krigingmethoden wie beschrieben normiert. Die Güte der Krigingmethoden wird somit jeweils abhängig von der Probenahmemethode bestimmt, sodass qualitative Vergleiche nur zwischen gleichen Probenahmemethoden möglich sind. Die Cokriging-Schätzwerte werden gegenüber denen des Ordinary Kriging schlechter bewertet. Am schlechtesten schneiden die Krigingmethoden der Residuen ab (vgl. Abb. 9.13 und Tabelle 9.3).

Verhalten sich alle Residuen derart, oder sieht man hier lediglich die Tatsache widerspiegelt, dass der gewichtete Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, für diesen Datentag nicht die optimale Trendkorrektur darstellt (vgl. Abb. 9.15)? Sowohl Untersuchungen zur besten Trendkorrektur als auch zu den Krigingmethoden bei verschiedenen Kovariablen sollen das Verhalten näher beleuchten.

Die Güte ist einheitlich besser für die Schätzwerte aus den Variogrammen, die aus rein zufällig statt mittels der Gittermethode gewählten Punkten berechnet worden sind. Im Nachfolgenden wird daher auf unterschiedliche Probenahmetechniken verzichtet und die Variogrammauswahl nur noch rein zufällig getätigt.

9.5.2 Beste Trendkorrektur im Ordinary Kriging der Residuen

Welches Residuum basierend auf einem speziellen Monatsmittel liefert im Ordinary Kriging (*ResOK*) die besten Ergebnisse im Bezug auf die tatsächlichen Fehler?

Zur Auswahl stehen acht Mittelungsverfahren (vgl. Beschreibungen und Abkürzungen in Abschnitt 7.3), wobei die arithmetischen Mittel sowohl mit Chlorophyllkonzentrationen als auch deren Logarithmen berechnet werden. Letztere werden entsprechend als *Mean.log* oder *gew.Mean.log* bezeichnet.

Die Chlorophyllkonzentrationen der Datentage vom 28. Februar 2007 bis zum 27. März 2007 übersteigen im niedrigen und mittleren Konzentrationsbereich deutlich die des Vergleichstages, dem 12. März 2007 (Abb. 9.16). Das ist einerseits auf die durch die Frühjahrsblüte hohe biologisch bedingte Dynamik zurückzuführen, andererseits auf den systematischen Versatz des Konzentrationsniveaus. In diesem Beispiel wird ein Monatsmittel den Vergleichstag immer übertreffen, sodass die Residuen nicht um Null zentriert liegen werden.

Sowohl rein visuell (Abb. 9.14) als auch in der Verteilung ihrer Residuen (Abb. 9.15) lässt sich der gewichtete Mittelwert der logarithm. Konzentrationen, *gew.Mean.log*, bzw. der gewichtete Mittelwert der Konzentrationen, *gew.Mean*, als dem Original am ähnlichsten identifizieren. Bisher wird in den Krigingversuchen mit dem gewichteten Mittelwert größter Wahrscheinlichkeit, *gew.MLH.Mean*, (Abschnitt 9.1 bis Abschnitt 9.4) gerechnet, welcher das Original nicht gut annähert.

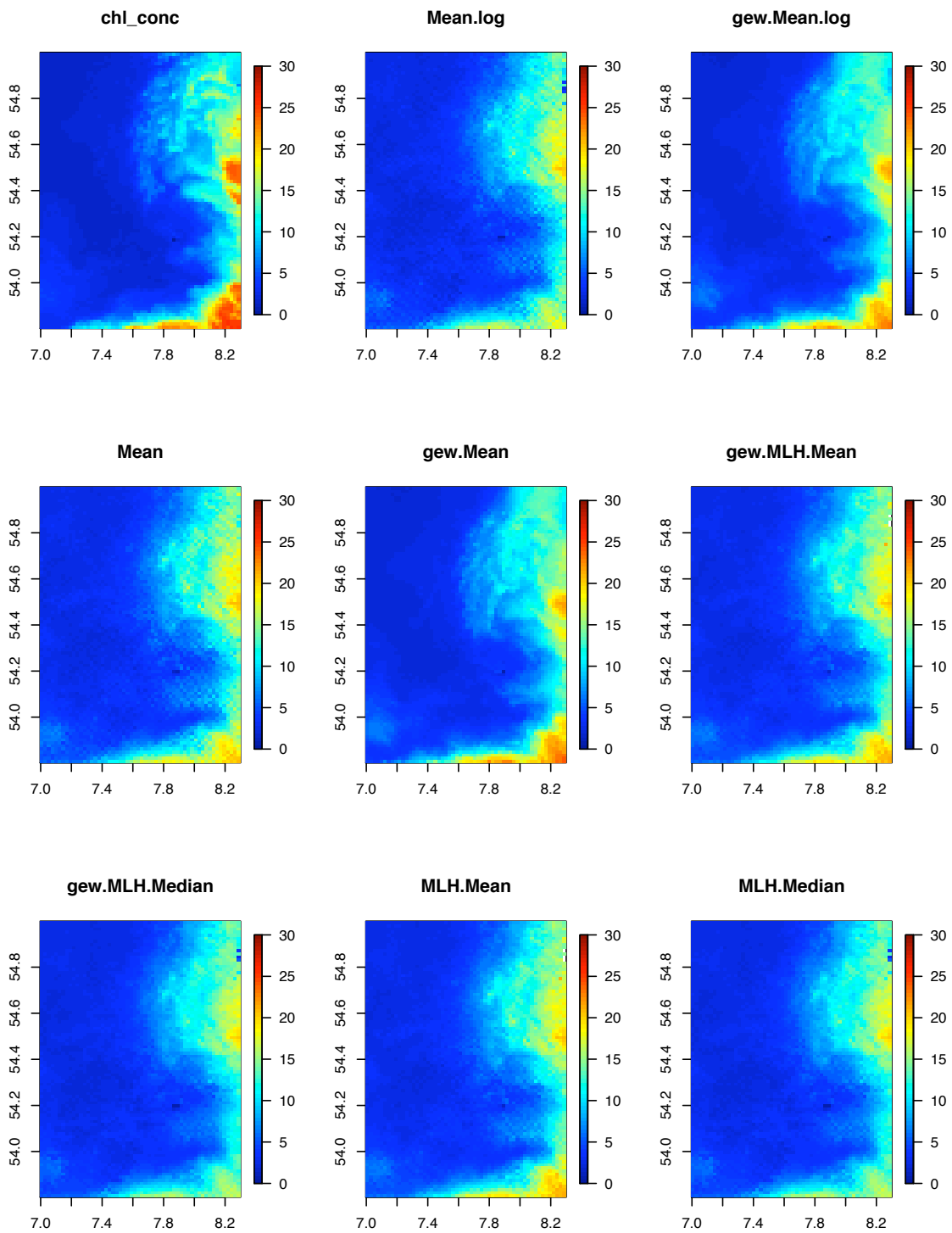


Abb. 9.14: Übersicht über die Monatsmittelwerte der Chlorophyllkonzentration in $\mu\text{g/l}$ nach verschiedenen Methoden berechnet jeweils aus Daten des Zeitraumes vom 28. Februar 2007 bis zum 27. März 2007. Die Karte links oben zeigt die originale Chlorophyllkonzentration $\mu\text{g/l}$ am 12. März 2007. Der Ausschnitt liegt in der Deutschen Bucht (Gebiet 4).

9 Kriging in der Praxis I - Deutsche Bucht

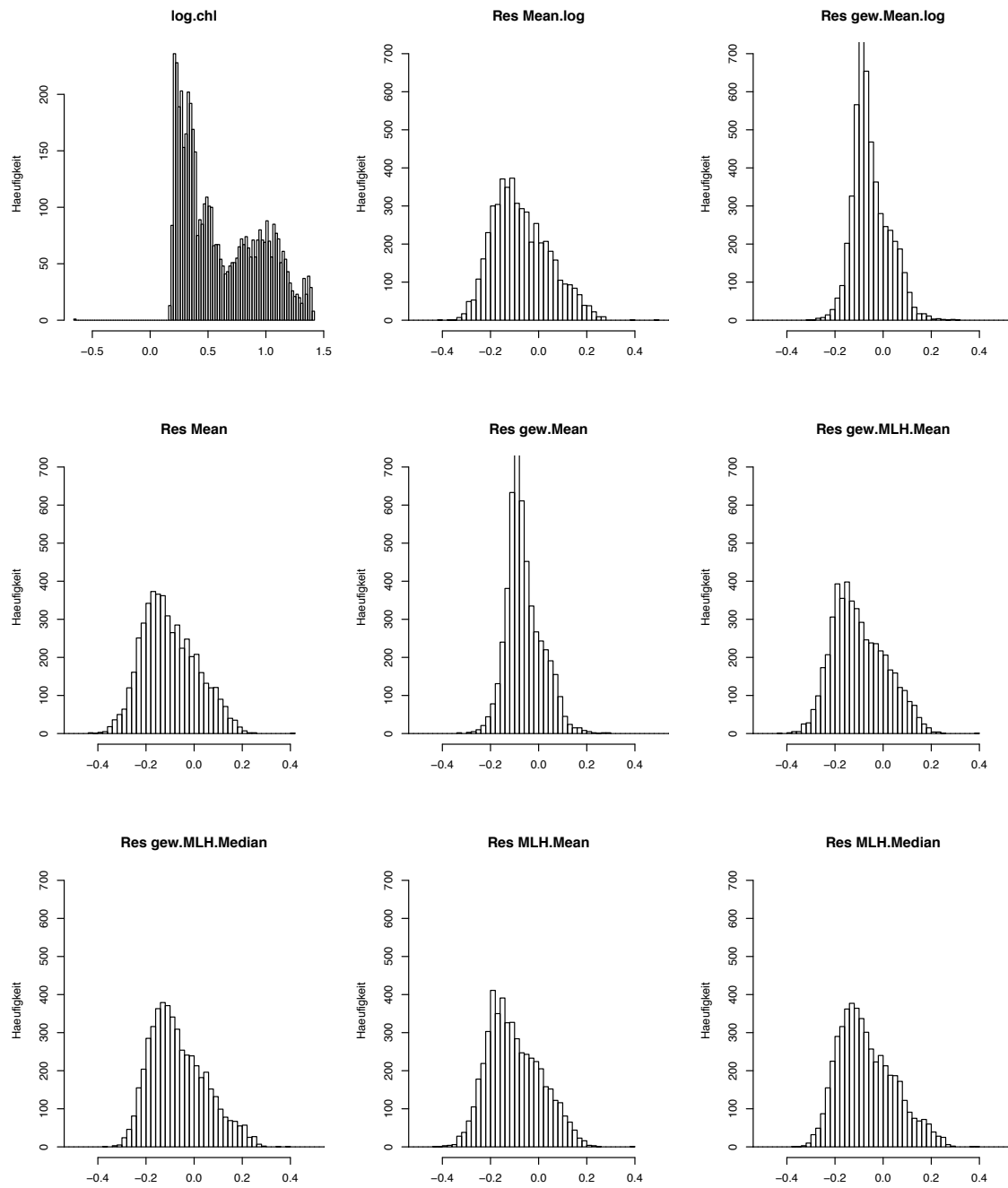


Abb. 9.15: Übersicht über die Häufigkeitsverteilung der Residuen für verschiedene Mittelungsmethoden berechnet. Zum Vergleich ist die Verteilung der logarithmierten Chlorophyllkonzentrationen gezeigt (Einheit $\log \mu\text{g}/l$). Je ähnlicher die Residuen einer Normalverteilung kommen, umso besser lassen sich die Daten durch ein Variogramm repräsentieren.

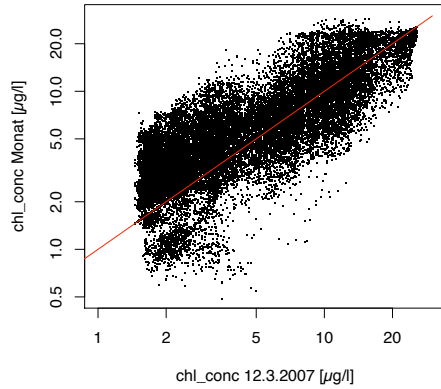


Abb. 9.16: Alle Chlorophyllwerte zwischen 28.2.2007 und 27.3.2007 jeweils pixelweise verglichen mit denen vom 12.3.2007.

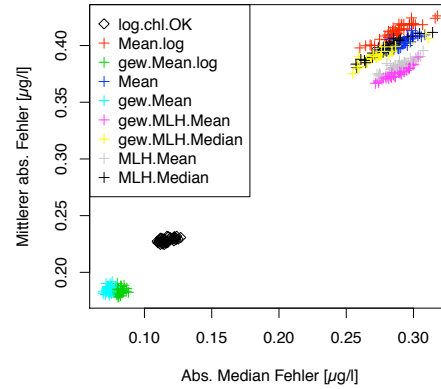


Abb. 9.17: Statistik tatsächlicher Fehler für Schätzwerte des Kriging der Residuen, *Res-OK*, die durch verschiedene Trendkorrekturen ermittelt wurden. Konstante Wolkenposition, Streuung durch Variogrammrealisationen. Die Quantifizierung der Güte findet sich in Tabelle 9.4

Trend	log.chl.OK	Mean.log		Mean		MLH.Mean		MLH.Median	
		-	gew.	-	gew.	-	gew.	-	gew.
Güte	0.275	1.395	0.032	1.374	0.007	1.309	1.284	1.329	1.302
var_{glob}	0.1169	0.0141	0.00586	0.0132	0.00584	0.0128	0.0126	0.0141	0.0140
var_{dat}	0.1180	0.0137	0.00592	0.0125	0.00590	0.0122	0.0120	0.0137	0.0137

Tab. 9.4: Güte der Trendkorrekturen anhand von Schätzwerten des Kriging der entsprechenden Residuen (vgl. Abb. 9.17) für eine konstante Wolkenposition im Datensatz vom 12. März 2007.

Bei konstanter Wolkenposition wird das Ordinary Kriging von Residuen verschiedener Mittelwerte demjenigen der Chlorophyllkonzentration gegenübergestellt. Die Fehlergrößen werden jeweils für eine konstante Probenahme zur Variogrammberechnung über alle zu vergleichenden Trendkorrekturen normiert.

Die Wahl des Trends hat deutlichen Einfluss auf die Güte der Rekonstruktion (Tabelle 9.4) gemessen am tatsächlichen Fehler (Abb. 9.17). Die beste Trendkorrektur gelingt mit dem gewichteten arithm. Mittelwert der Konzentrationen (*gew.Mean*), gefolgt von dem für logarithmierte Konzentrationen (*gew.Mean.log*), deren Schätzwerte deutlich gegenüber dem Ordinary Kriging verbessert sind. Alle übrigen Mittelwertmethoden schneiden schlechter als das OK ab. Dies wird in erster Linie durch die ausgeschlossene und verbleibende Varianz des Datensatzes bestimmt: im Falle von $\text{var}_{dat} > \text{var}_{glob}$ wird ein schwach strukturierter Bereich ausgeschlossen, der sich immer besser rekonstruieren lässt als ein stark strukturierter.

Bei der Residuenmethode werden die Abweichungen der Chlorophyllverteilung eines Tages vom monatlichen Mittel modelliert. Anschaulich bedeutet dies, dass die Methode versucht, die notwendigen Anpassungen am Monatsmittel vorzunehmen. Das Monatsmittel fungiert als Abschätzung des räumlichen Trends, der möglichst gut die vorhandenen Strukturen nachzeichnen sollte, sodass die Residuen möglichst strukturarm sind und in Näherung die Voraussetzung des konstanten räumlichen Erwartungswertes erfüllen. Dazu sind die gewichteten arithmetischen Mittel hier sehr gut in der Lage.

9.5.3 Beste Krigingmethode bei konstanter Trendkorrektur

Während nur die beiden besten Trends (gewichteter Mittelwert der (logarithm.) Konzentration, *gew.Mean* und *gew.Mean.log*, vgl. Abschnitt 9.5.2) verwendet werden, werden in den Krigingmethoden verschiedenen Kombinationen von Kovariablen durchgespielt. Zum Vergleich mit diesen Cokrigingansätzen wird immer das Ordinary Kriging hinzugezogen. Die Position der Wolkenmaske ist konstant, die Variationen entstehen durch die wiederholte Probenahme für das Variogramm.

Entsprechende Modelle sind hier:

Ordinary Kriging basierend auf den logarithmierten Konzentrationen oder auf den Residuen (*OK* bzw. *ResOK*), sowie

Cokriging basierend auf den Residuen mit den Hilfsvariablen Wassertiefe und Monatsmittel (*bin*) in verschiedenen Kombinationen (*ResCoOK.tief*, *ResCoOK.bin*, *ResCoOK.bin.tief*)

Hier zeigt sicher erneut, dass das Cokriging kein Garant für exaktere Schätzwerte ist (Abb. 9.18 und Tabelle 9.5).

Absolut konsistent lassen sich die Ergebnisse für die beiden Trends und die verschiedenen Krigingmethoden verfolgen:

- *ResOK*. Das einfache Ordinary Kriging der Residuen für Trendkorrekturen aus gewichteten Mittelwerten (*gew.Mean.log* bzw. *gew.Mean*) schafft deutlich exaktere Schätzwerte als reines Ordinary Kriging der log. Chlorophyllkonzentration. Der

Trend		gew.Mean.log				gew.Mean			
Kriging	OK.log.chl	ResOK	ResCoOK			ResOK	ResCoOK		
Hilfsvariable	-	-	tief	bin	bin, tief	-	tief	bin	bin, tief
Güte	1.159	0.280	0.097	1.300	0.705	0.223	0.072	1.355	0.791
var _{glob}	0.1169	0.00587	-0.3548	0.0109	-	0.00585	-0.3386	0.0101	-
var _{dat}	0.1180	0.00592	-0.3067	0.0104	-	0.00590	-0.2914	0.0096	-

Tab. 9.5: Güte der Krigingmethoden in Vergleich. Die zugrunde liegende Fehlerstatistik ist in Abb. 9.18 dargestellt. Für diesen Versuch mit konstanter Wolkenmaske wird die globale Varianz bzw. die Varianz der verbleibenden Datenpunkte jeweils für die zu schätzende Variable dokumentiert, für Cokrigingmethoden ist die Kovarianz angegeben.

gewichtete arithm. Mittelwert (ohne logarithmierte Konzentrationen, *gew.Mean*) sorgt für einen geringeren Median des Fehlers, d.h. höhere globale Genauigkeit, als der *gew.Mean.log*. Andererseits liegt der Mittelwert des absoluten Fehlers etwas höher für den *gew.Mean*. Die Streuung der Fehler in verschiedenen Variogrammrealisationen ist hier am geringsten.

- *ResCoOK.tief*. Den geringsten Fehler besitzen die beiden ResCoOK-Methoden, die sich der Tiefe als Kovariable bedienen. Wiederum ist für *gew.Mean.log* der Mittelwert des absoluten Fehlers am geringsten, für *gew.Mean* ist der Median der Fehler am nächsten an Null. Die Streuung der Fehlergrößen ist deutlicher als bei ResOK.
- *ResCoOK.bin*. Dies ist die schlechteste Kombination von Methoden.
- *ResCoOK.bin.tief*. Die Hinzunahme der Tiefe verbessert die Fehlerstatistik gegenüber ResCoOK.bin. Allerdings streuen die Ergebnisse sehr weit.

Nach der definierten Methodengüte (Gl. 9.4, Normierung der Fehlergrößen erfolgt jeweils bei einer konstanten Probenahme für die Variogrammberechnung über alle zu vergleichenden Krigingmethoden) wäre für dieses Beispiel das Cokriging des Residuums des gewichteten Mittelwertes mit der Hilfsvariable Tiefe, *ResCoOK.gew.Mean.tief*, die beste Methode ($G = 0.072$), gefolgt von der gleichen Cokrigingmethode jedoch mit Residuen, die aus dem gewichteten Mittelwert der log. Konzentrationen bestimmt werden, *ResCoOK.gew.Mean.log.tief*, ($G = 0.097$). Die einfachen Ordinary Kriging der Residuen, *ResOK.gew.Mean(.log)*, sind ebenfalls deutlich besser als das Ordinary Kriging der log. Chlorophyllkonzentration, *OK.log.chl*, ($G = 1.159$). Dieses Ergebnis ist keinesfalls allgemeingültig, sondern absolut auf den speziellen Datentag und die spezifische Lage dieser Wolkenmaske beschränkt.

9.5.4 Beste Trendkorrektur in Abhängigkeit von der Variation der Wolkenposition

In den bis jetzt gezeigten Beispielen beruht die Variation der Fehler nicht auf unterschiedlicher Positionierung der Wolken, sondern einzig und allein auf unterschiedlichen

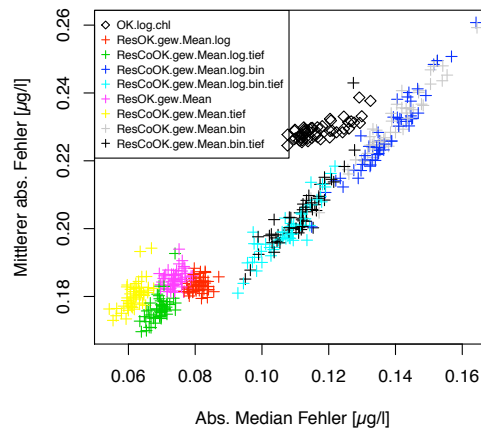


Abb. 9.18: Statistik tatsächlicher Fehler für verschiedene Ordinary Kriging und Cokriging-Methoden bei Residuen basierend auf gewichteten Monatsmitteln (*gew.Mean* bzw. *gew.Mean.log*) für eine konstante Wolkenmaske. Für eine Variogrammprobenahme werden jeweils alle Schätzwertmethoden ausgeführt. Die Quantifizierung der Güte der Methoden findet sich in Tabelle 9.5.

Variogrammrealisationen, d.h. der Auswahl von Punkten für die Berechnung des Variogramms. Als Grundlage dienen auch weiterhin sowohl die Chlorophyllkonzentrationen als auch alle Residuen unterschiedlicher Mittelungsverfahren, sodass Schätzwerte mit dem Kriging der logarithmierten Konzentrationen (*log.chl.OK*) und der verschiedenen Residuen (*ResOK*) bestimmt werden.

Variiert die Lage der Wolkenmaske und bestimmt man für jede Lage genau ein Variogramm, so ergeben sich wesentlich unschärfere Zusammenhänge zwischen der Güte der Rekonstruktion und den Krigingverfahren:

Während die Darstellung von Mittelwert des absoluten Fehlers gegen Absolutwert des Fehlermedians (Abb. 9.19a) eine große Streuung aufweist, werden die Unterschiede zwischen den Methoden durch Normierung über alle Methoden für jeweils eine Wolkenposition - was hier gleichbedeutend mit einer Variogrammrealisation ist - deutlicher erkennbar (Abb. 9.19b). Dann treten die Fehlerstatistik der Schätzwerte basierend auf den Residuen gebildet durch Monatsmittel mit den beiden Methoden *gew.Mean.log* und *gew.Mean* positiv hervor, sowie einige Realisationen basierend auf dem Kriging der log. Konzentrationen (*OK*). Dies lässt sich in der Quantifizierung ihrer Güte bestätigen (vgl. Tabelle 9.6).

Zu den Fehlergrößen einer Wolkenposition kommt jeweils noch die Streuung durch Variogrammrealisationen hinzu. Die Streuung der Fehlergrößen ausgehend von den Wolkenpositionen erlaubt keine so eindeutige Festlegung der besten Methode, wie in den vorausgehenden Fällen.

Welche Auswirkung das Verhältnis der Varianz der verbleibenden Daten, var_{dat} , zur globalen Varianz, var_{glob} , auf die Wahl der besten Trendkorrektur hat, lässt sich an diesem Versuch nicht klären: von den 50 Wolkenpositionen resultiert nur eine einzige in

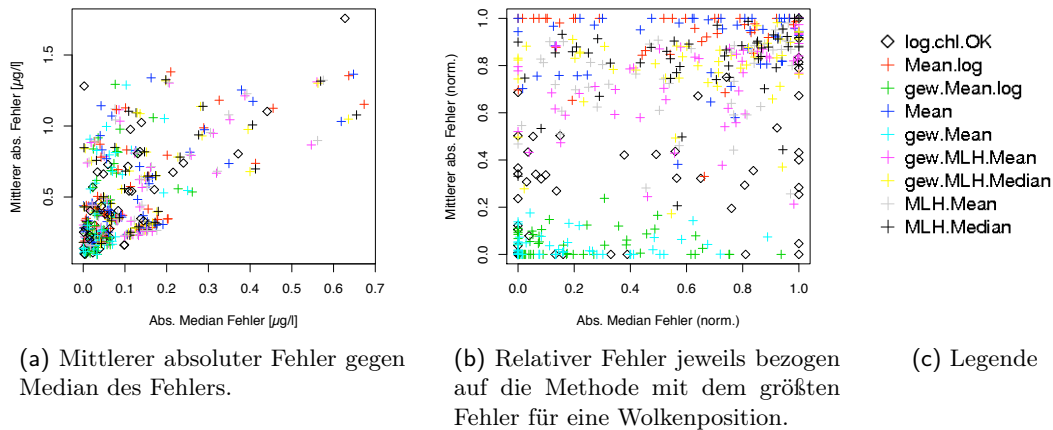


Abb. 9.19: Statistik der tatsächlichen Fehler der Schätzwerte zur Bestimmung der besten Trendkorrektur. Für 50 verschiedene Wolkenpositionen werden jeweils die Schätzwerte basierend auf unterschiedlichen Residuen (c) berechnet. Die Daten stammen vom 12. März 2007 in der Deutschen Bucht. Die Quantifizierung der Güte der Methoden findet sich in Tabelle 9.6.

Trend	Mean.log		Mean		MLH.Mean		MLH.Median		
	log.chl.OK	-	gew.	-	gew.	-	gew.	-	gew.
Güte	0.636	1.176	0.240	1.155	0.208	1.002	0.972	1.061	1.036
Std.Abw.	0.435	0.207	0.242	0.212	0.227	0.257	0.262	0.232	0.240

Tab. 9.6: Güte der Trendkorrektur basierend auf Schätzwerten von Daten vom 12. März 2007 und 50 verschiedenen Wolkenpositionen. Die lokalen und globalen Fehlergrößen sind in Abb. 9.19 illustriert.

einen Datensatz mit $var_{glob} > var_{dat}$ für $log.chl_conc$ und die verschiedenen Residuen. 30 Wolkenlagen führen zu $var_{glob} < var_{dat}$, bei 19 Wolkenpositionen sind die Verhältnisse gemischt. Eine entsprechende Aufschlüsselung und genauere Erläuterung erfolgt im weiteren Verlauf (vgl. Abschnitt 9.5.6).

9.5.5 Beste Krigingmethode in Abhängigkeit von der Variation der Wolkenposition

Für die Daten vom 12. März 2007 wird der Krigingversuch mit den zuvor bestimmten besten Trendkorrekturen, den gewichteten Monatsmitteln der (log.) Konzentrationen, $gew.Mean.log$ bzw. $gew.Mean$, wiederholt. Als Krigingmethoden werden die gleichen wie in Abschnitt 9.5.3 verwendet. Diesmal wird die Wolkenmaske (vgl. Abb. 9.8a) innerhalb des Gebietes verschoben und zu jeder Realisation der Wolkenlage genau eine Probenahme für das Variogramm durchgeführt.

Die Normierung der Fehlergrößen erfolgt für jeweils eine Wolkenposition (d.h. einer Probenahme für die Variogrammberechnung) über alle verwendeten Methoden.

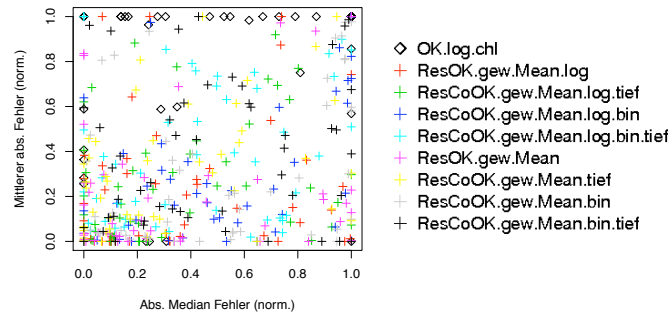


Abb. 9.20: Bestimmung der besten Krigingmethode für die Trendkorrekturen mit Monatsmitteln nach *gew.Mean.log* und *gew.Mean* bei variiertter Wolkenposition. Die Daten stammen vom 12. März 2007. Die Quantifizierung der Güte der Methoden ist in Tabelle 9.7 zu finden.

Trend		gew.Mean.log				gew.Mean			
Kriging	OK.log.chl	ResOK	ResCoOK			ResOK	ResCoOK		
Hilfsvariable	-	-	tief	bin	bin, tief	-	tief	bin	bin, tief
Güte	1.075	0.499	0.492	0.622	0.640	0.485	0.493	0.647	0.672

Tab. 9.7: Güte der Krigingmethoden für Daten vom 12. März 2007 bei variabler Wolkenposition und zwei Trendkorrekturen (vgl. Abb. 9.20).

Die Darstellung der normierten Absolutfehler und des Fehlermedians erlaubt nur das direkte Urteil, dass das Ordinary Kriging häufig als die schlechteste Methode gelten muss (Abb. 9.20). Die Quantifizierung der Güte zeichnet die Residuenkrigingmethoden als die besseren aus; dabei gibt es keinen nennenswerten Unterschied zwischen den beiden verwendeten Trendkorrekturen oder zwischen reinem Ordinary Kriging der Residuen oder ihrem Cokriging mit der Wassertiefe als Hilfsvariable (vgl. Tabelle 9.7).

9.5.6 Beste Trendkorrektur bei fester Wolkenposition und verschiedenen Datentagen

Im Gegensatz zu den vorherigen Experimenten, die jeweils nur Daten vom 12. März 2007 berücksichtigen, soll die Fehlerstatistik diesmal auf 203 unterschiedlichen Datentagen aus den Jahren 2003 bis 2008 aufbauen, für die jeweils die gleiche Wolkenmaske in fester geographischer Lage (vgl. Abb. 9.8a) zum Einsatz kommt. Dadurch wird die Geometrie des Krigingproblems konstant gehalten, während die räumliche Kovarianzstruktur der Daten variiert.

Für alle Datentage wird die Positionierung der Probenahme für das Variogramm konstant gehalten, sodass die Variogramme unterschiedlicher Methoden und Tage auf der gleichen Datenpunktgeometrie beruhen und die bereits demonstrierte Variabilität der Schätzwerte durch die Probenahme nicht zum Tragen kommt. Die Untersuchung wird

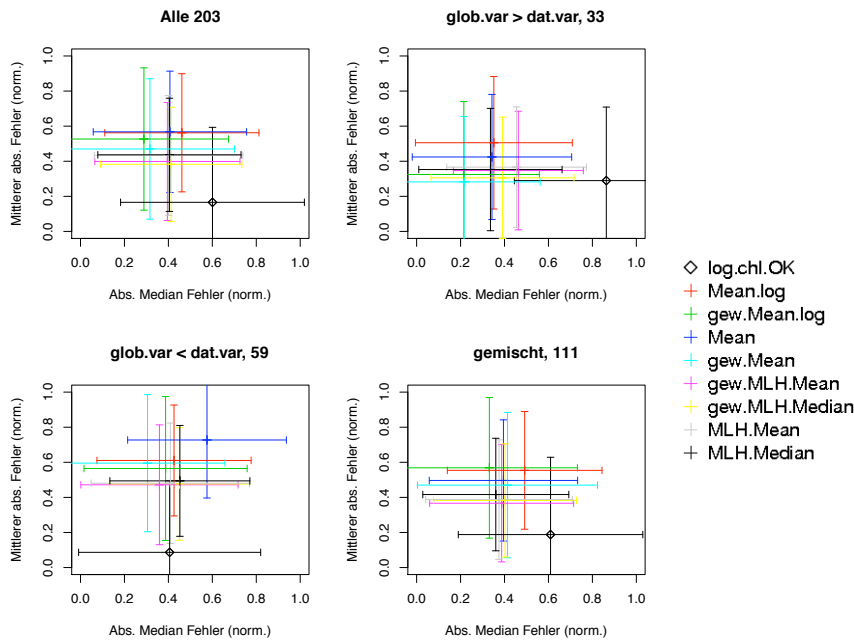


Abb. 9.21: Bestimmung der besten Trendkorrektur für eine konstante Wolkenmaske und einen Satz konstanter Positionen für die Probenahme zum Variogramm. Für jeden der 203 Datentage werden die je nach Krigingmethode verschiedenen Variogramme bestimmt. Die Quantifizierung der Güte der Methoden findet sich in Tabelle 9.8.

hier auf diese eine Variogrammrealisation beschränkt. Die Normierung der Fehlergrößen erfolgt für jeweils einen Datentag über alle zu vergleichenden Trendkorrekturen.

Was die insgesamt 203 Situationen quantitativ unterscheidet, ist das Verhältnis der Varianz der verbleibenden Daten, var_{dat} , zur globalen Varianz, var_{glob} . Dieses Verhältnis kann nicht nur für den individuellen Datentag, sondern auch für jede einzelne Trendkorrektur anders ausfallen. Es gibt 59 Fälle, in denen für alle neun Methoden $var_{dat} > var_{glob}$ gilt; entsprechende 33 Fälle mit $var_{dat} < var_{glob}$. In den restlichen 111 Fällen sind die Verhältnisse gemischt. Betrachtet man alle Fälle, so schneidet das Kriging der log. Chlorophyllkonzentrationen, $log.chl.OK$, besonders gut ab. Nur im Falle von $var_{dat} < var_{glob}$ ist das Residuenkriging basierend auf dem Monatsmittel $gew.Mean.log$ vorzuziehen (vgl. Abb. 9.21 bzw. Tabelle 9.8); der Trend $gew.Mean.log$ scheint in der Lage, wenn stark strukturierte Gebiete ausgeschlossen werden, diese am besten zu rekonstruieren.

Bei dieser Aussage gilt, dass sie nur für die Geometrie dieses Teilgebietes mit der speziellen Wolkenmaske Gültigkeit hat.

Trend	N	log.chl.OK	Mean.log		Mean		MLH.Mean		MLH.Median	
			-	gew.	-	gew.	-	gew.	-	gew.
$var_{dat} > var_{glob}$	59	0.6496	0.7959	0.8675	1.0011	0.8557	0.7662	0.7041	0.7365	0.6989
$var_{dat} < var_{glob}$	33	0.8633	0.6438	0.3778	0.64	0.5563	0.7135	0.7007	0.5692	0.6018
gemischt	111	0.8372	0.7574	0.8351	0.7061	0.8479	0.6449	0.6608	0.6582	0.6516
alle	203	0.8206	0.7798	0.7792	0.8098	0.8269	0.6887	0.6961	0.6926	0.6511

Tab. 9.8: Güte der Krigingmethoden mit Schwerpunkt auf den verschiedenen Trendkorrekturen. Für jeden der 203 Datentage werden bei konstanter Wolkenmaske und konstanter Positionierung der Probenahme je nach Krigingmethode verschiedenen Variogramme bestimmt (Illustration der Fehlerstatistik in Abb. 9.21).

9.5.7 Beste Krigingmethode bei fester Wolkenposition und verschiedenen Datentagen

Um die Kombinationen von Untersuchungen zu vervollständigen, folgt als letztes ein Versuch zur besten Krigingmethode bei fester Position der Wolkenmaske aber unterschiedlichen Datentagen. Die Auswahl der Variogrammpunkte wird pro Datentag konstant gehalten. Als Trendkorrektur wird das Monatsmittel gemäß *(gew).MLH.Median* verwendet, welches im vorherigen Versuch als bester Trend ermittelt wurde. Zusätzlich zum Ordinary Kriging wird jeweils das Cokriging mit der Tiefe berechnet.

Zu jedem Datentag wird die Normierung der Fehlergrößen über alle Methoden durchgeführt.

In diesem Fall schneidet das Ordinary Kriging basierend auf den logarithmierten Chlorophyllkonzentrationen am besten ab (*OK.log.chl*, vgl. Tabelle 9.9). Wird jedoch starke Variabilität durch die Wolkenmaske aus den Daten ausgeschlossen, d.h. $var_{dat} < var_{glob}$, so führt das Cokriging der Residuen basierend auf der Trendkorrektur *gew.MLH.Median* zur besten Rekonstruktion (*ResCoOK.tief*).

Ist der nicht gesehene Anteil von Daten vergleichsweise homogen, d.h. $var_{dat} > var_{glob}$, so kann mit dem einfachen Cokriging von Chlorophyllkonzentration und Tiefe das beste Ergebnis erzielt werden.

Signifikant ist der Unterschied der Güte in allen Fällen nicht, da die Intervalle ihrer Standardabweichung sehr breit sind und sich mit denjenigen anderer Methoden überlappen (Abb. 9.22).

9.5.8 Beispiel für Schätzwerte aus verschiedenen Krigingmethoden

Die Wolkenmaske (vgl. Abb. 9.8a) wird so gelegt, dass sie einen möglichst stark strukturierten Bereich ausschließt - homogene Bereiche werden sich immer vergleichsweise gut schätzen lassen. Jedoch ist die Entwicklung einer Methode, die auch zur Rekonstruktion komplexer Strukturen in der Lage ist, ein wichtiges Untersuchungsgebiet dieser Arbeit. Entsprechend gilt für den im Folgenden betrachteten Datensatz: $var_{dat} < var_{glob}$.

Aus den 93 Probenahmen für die Variogramme werden jeweils für eine Probenahme die Fehlergrößen normiert, sodass die Güte der Krigingmethoden unabhängig von der

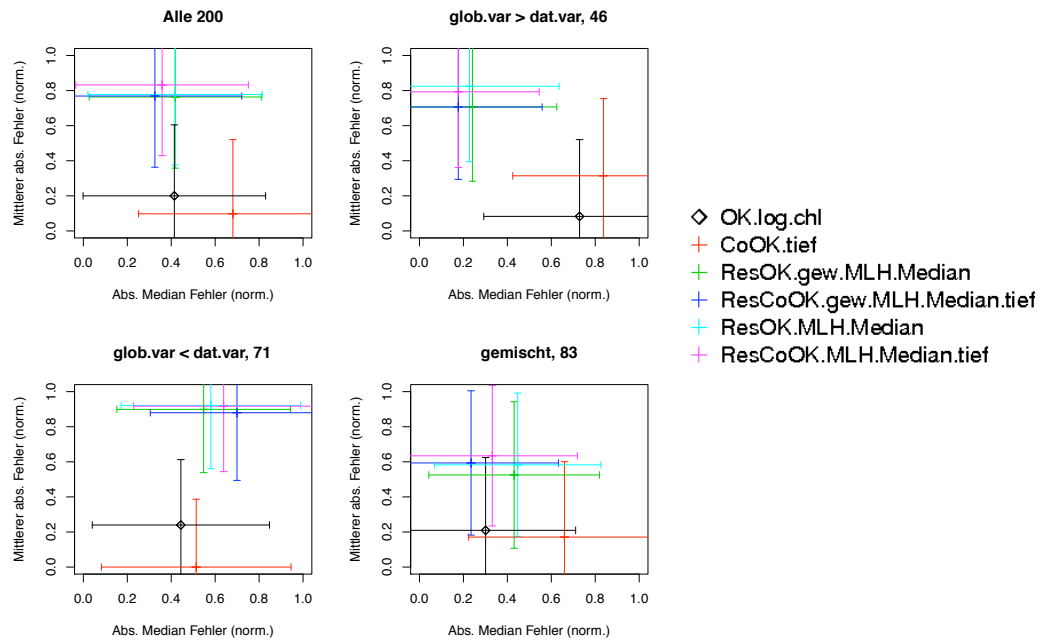


Abb. 9.22: Fehlerstatistik der Schätzwerte verschiedener Krigingmethoden bei fester Wolkenposition und verschiedenen Datentagen. Die Quantifizierung der Güte ist in Tabelle 9.9 niedergelegt.

Trend	N	MLH.gew.Median				MLH.Median	
		OK.log.chl	CoOK.tief	ResOK	ResCoOK.tief	ResOK	ResCoOK.tief
$\text{var}_{dat} > \text{var}_{glob}$	73	0.7487	0.6085	1.003	1.011	1.0002	1.0236
$\text{var}_{dat} < \text{var}_{glob}$	46	0.9144	1.000	0.897	0.7533	0.9817	0.8471
gemischt	84	0.6269	0.7926	0.8336	0.7258	0.8993	0.8249
alle	203	0.752	0.828	0.9181	0.9175	0.9884	0.9598

Tab. 9.9: Güte der Krigingmethoden bei fester Wolkenposition und verschiedenen Datentagen. Zur Größe der Fehlerintervalle vgl. Abb. 9.22.

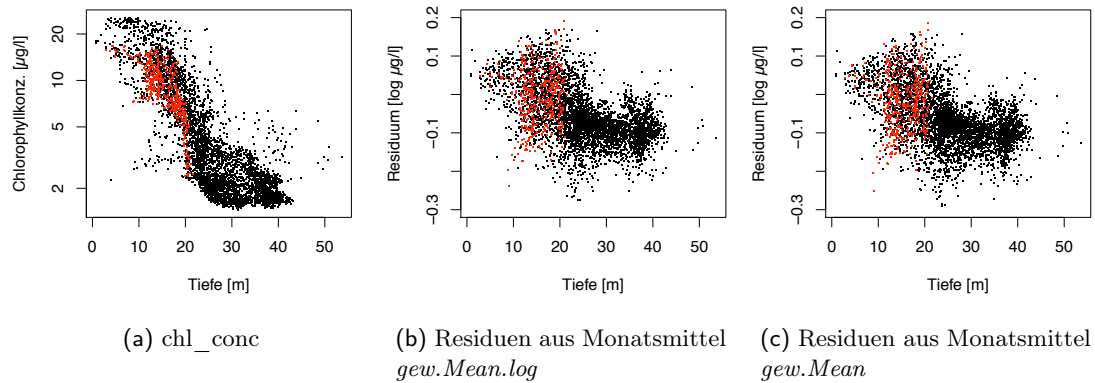


Abb. 9.23: Abhängigkeit der Chlorophyllkonzentration *chl_conc* bzw. der Residuen von der Hilfsvariable Wassertiefe

Variogrammrealisation bestimmt wird.

Die beste Trendkorrektur ist der gewichtete arithmetische Mittelwert der logarithmierten Chlorophyllkonzentration (*gew.Mean.log*, gemäß Abschnitt 9.5.6).

Die erfolgversprechendste Krigingmethode ist nach Abschnitt 9.5.7 das Cokriging der Residuen mit der Tiefe.

Zusätzlich wird auch die zweitbeste Trendkorrektur *gew.Mean* sowie das Ordinary Kriging der log. Chlorophyllkonzentrationen und dem Cokriging mit der Tiefe genutzt. Von allen wird jeweils eine Realisation berechnet und die Krigingergebnisse als Karten dargestellt (Abb. 9.24).

Tatsächlich entspricht die Rangfolge der Güte der Rekonstruktion der aus den vorherigen Versuchen erwarteten. Das Cokriging mit der Tiefe nach der Trendkorrektur mit *gew.Mean.log* liefert die beste Rekonstruktion (vgl. Tabelle 9.10).

Dass das Cokriging mit der Kovariablen 'Tiefe' für die Modelle mit reiner Chlorophyllkonzentration keine Verbesserung bringt im Gegensatz zu den Residuenmethoden, lässt sich aus dem Zusammenhang zwischen Daten und Tiefe ansehen (Abb. 9.23). Die ausgeschlossenen Datenpunkte (rot) liegen außerhalb des Schwerpunktes der verbleibenden Daten (schwarz), aus denen das Variogramm ausgewählt wird (Abb. 9.23a). In diesem Fall kann das Variogramm die ausgeschlossenen Punkte nicht gut repräsentieren und die Schätzwerte werden nicht nahe an den Originalen liegen.

9.6 Zusammenfassung und Ausblick

Nach den Untersuchungen in Kapitel 9 kann es keine eindeutige Empfehlung für *eine* Trendkorrektur oder *eine* Krigingmethode zur Bestimmung von Schätzwerten geben. Die Einzelbeispiele legen gänzlich andere Methoden je nach verbleibender und ausgeschlossener Datenstruktur nahe. Dabei ist noch unberücksichtigt, welchen Einfluss auf

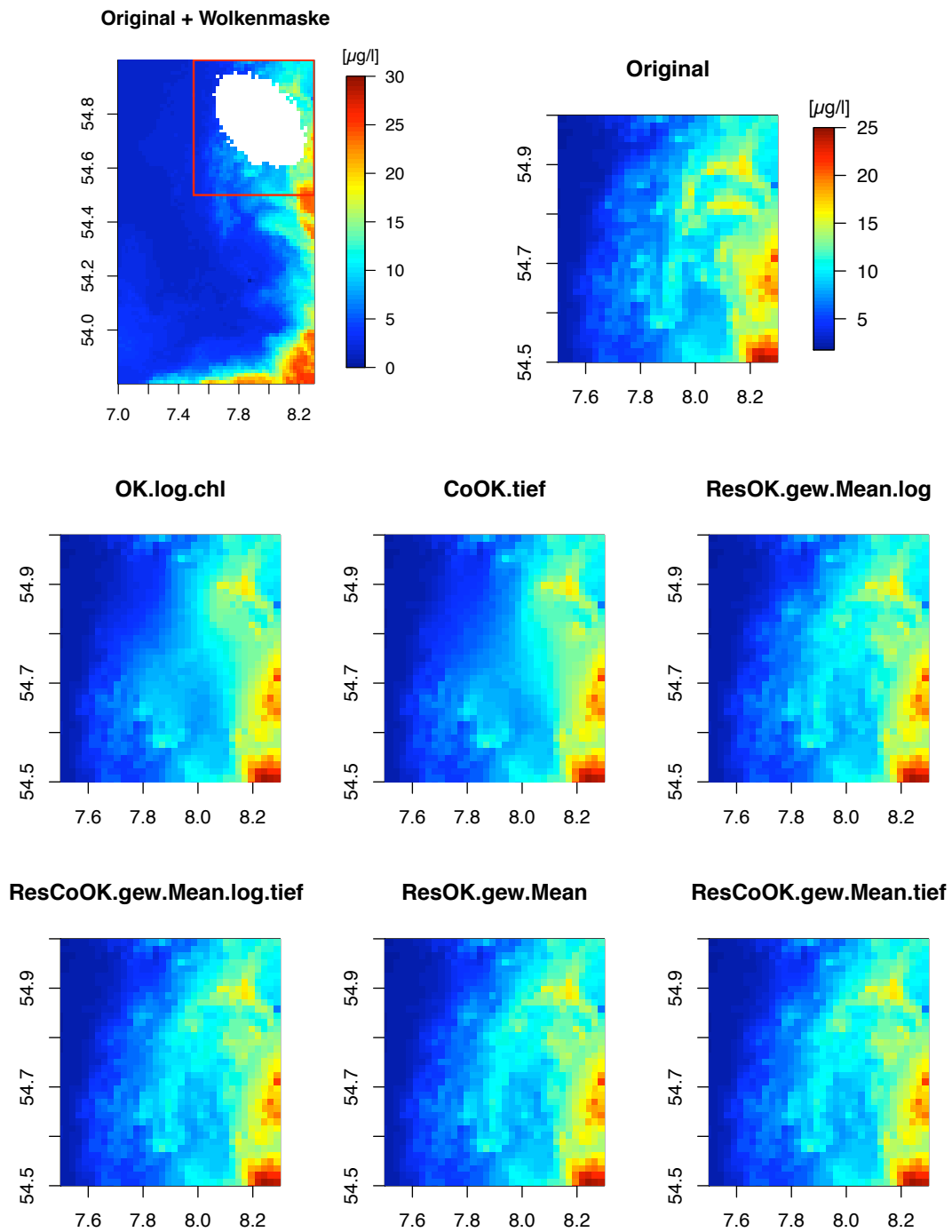


Abb. 9.24: Beispiele für Schätzwerte nach unterschiedlichen Krigingverfahren. Die Variogramme für die unterschiedlichen Methoden beruhen auf einer konstanten Probenahme. Zum Vergleich ist sowohl die originale Chlorophyllverteilung vom 12. März 2007 als Ausschnitt des Gebietes, als auch die Lage und Struktur der Wolkenmaske dargestellt. Alle Abbildungen zeigen die Chlorophyllkonzentration in $\mu\text{g/l}$, die Skala des Originalbildes gilt auch für alle übrigen. Die Güte der Methoden findet sich in Tabelle 9.10; die beste Rekonstruktion gelingt mit dem Residuen-Cokriging (ResCoOK.gew.Mean.log.tief).

Trend			gew.Mean.log		gew.Mean	
	Methode	OK.log.chl	CoOK.tief	ResOK	ResCoOK.tief	ResOK
Güte	1.017	1.389	0.104	0.078	0.130	0.097
var_{glob}	0.117	-2.535	0.0059	-0.355	0.0058	-0.339
var_{dat}	0.114	-2.400	0.0057	-0.342	0.0057	-0.326

Tab. 9.10: Güte der Krigingmethoden für Daten des Tages 12. März 2007 basierend auf 93 Variogrammrealisationen (vgl. Abb. 9.24).

die Fehler der Überdeckungsgrad oder die Struktur der Wolkenmaske nach sich zieht. Einen gewissen Einblick werden die Krigingversuche im Rahmen der Quantifizierung der Wolkenstrukturen im nachfolgenden Kapitel 10 liefern.

Die sinnvolle Wahl der Methodenkombination wird letztlich vom Bearbeiter der Daten zu treffen sein, sodass dessen persönliche Überlegungen und Vorlieben zur Uneindeutigkeit der Schätzwerte führen werden (Englund (1990)).

Dennoch können einige Ergebnisse festgehalten werden:

Variogramm

- Es ist vertretbar, statt echter Distanzen die Pseudo-Distanzen, gebildet aus Lon/Lat-Koordinaten, zu verwenden. Die Variogramme unterscheiden sich qualitativ nicht von denen auf Pixel-Distanzen basierenden. Bei der Variabilität, die durch die Probenahme der Datenpunkte zur Berechnung des Variogramms entsteht, ist der Unterschied durch die Distanzdefinition unerheblich.
- Die Wahl der Probenahmemethode ist bei der großen Anzahl von Punkten, die bei diesen Datensätzen zur Verfügung steht, ohne Relevanz im Vergleich zur Streuung, die bei jeder neuen Variogrammrealisation, d.h. mit jeder neuen Probenahme, auftritt. In einem dichten Datensatz mit zusammenhängend fehlenden Datenpunkten können die Variogramme verschiedener Realisationen etwas weniger bei einer Auswahl durch die Gittermethode als die rein zufällige Wahl streuen. Sobald die Variogramme aus Stichproben berechnet werden, ist es bezogen auf die tatsächlichen Fehler nachfolgender Krigingschätzwerte unerheblich, ob diese per Zufall- oder Gittermethode ausgewählt werden.
- Das Variogramm eines dichten, vollständigen Datensatzes, für dessen Berechnung nur ein geringer Anteil von Datenpunkten ausgewählt wird, wird im Allgemeinen im Mittel geringere Semi-Varianzen aufweisen, als das entsprechende globale Variogramm.
- Das Variogramm eines dichten Datensatzes bei fehlenden (d.h. ausmaskierten) Datenpunkten wird für den Fall $var_{glob} > var_{dat}$ die Tendenz zeigen, eine kleinere Semi-Varianz als die des globalen Variogramms zu besitzen, während im Fall $var_{glob} < var_{dat}$ die Semi-Varianz überschätzt wird.

Unter Berücksichtigung der Erkenntnis, dass die Variogrammrealisationen im Nahbereich am wenigsten streuen und nur dieser Teil für das lokale Kriging wirklich von Belang ist und gut angepasst werden muss und kann, lassen sich Krigingexperimente basierend auf derartigen Variogramm-Modellen durchaus rechtfertigen.

Nachbarschaft zum Schätzpunkt im lokalen Kriging Im lokalen Kriging gibt es eine Reihe von Möglichkeiten zur Definition der “nächsten” Nachbarn (vgl. Abschnitt 9.4). Statt eines reinen Abstandargumentes kann es zum Beispiel sinnvoll sein, jeweils Daten aus den verfügbaren Raumrichtungen (in diskreten Winkelintervallen) zu berücksichtigen. Zu der rein distanz-beschränkten Menge nächster Nachbarn kommt dann noch eine Auswahl von N_w Punkten pro noch nicht gesehenes Winkelintervall hinzu. Dieser Ansatz liefert vielversprechende Ergebnisse, ist aber noch nicht für den einfachen Gebrauch implementiert und bleibt daher unberücksichtigt.

Eine mathematischere und objektivere Beschreibung zur notwendigen Größe der Nachbarschaft findet sich z.B. in Rivoirard (1987).

Konvergenz der Schätzwerte bzw. der Krigingfehler tritt im Allgemeinen natürlich nicht bei den hier veranschlagten Anzahl $N_{max} = 150$ Punkten auf. Je nach Struktur der vorhandenen Datenpunkte (Struktur der Nachbarschaft) zu einer Schätzposition wird sich diese erst für wesentlich größere Punktmengen erreichen lassen. Jeder Datenpunkt, der in einer noch nicht gesehenen Raumrichtung hinzukommt, verändert das Verhältnis der Gewichte zueinander enorm.

Korrektur für Glättung und systematische Über- oder Unterschätzung Dass die Schätzwerte des Kriging (gerade im Ordinary Kriging) glatter, d.h. von weniger kleinkaligen Variationen als der Originaldatensatz geprägt, sind als die tatsächlichen (bei einer Kreuzvalidierung) oder die erwarteten Werte (Abb. 9.24), ist von einer Interpolationsmethode zu erwarten.

Die Fehlerstatistik des Median der Fehler zeigt, dass die Schätzwerte zu einer Variogrammrealisation die tatsächlichen Werte systematisch über- oder unterschätzen können. Zur Korrektur dieser Phänomene sind verschiedene Verfahren vorgeschlagen worden (Olea und Pawlowsky (1996); Journel u. a. (2000); Yamamoto (2005)), die sich Kreuzvalidierungsuntersuchungen oder Frequenzanalysen zu Eigen machen. Die Bewertung der Verfahren steht noch aus und wird kein Gegenstand dieser Arbeit sein.

Verlässlichkeit von Schätzwerten In den Kreuzvalidierungsexperimenten lässt sich ein Maß zur gleichzeitigen Bewertung der lokalen und globalen Genauigkeit der Schätzwerte aus der Statistik der tatsächlichen Fehler definieren. Für einen echten Anwendungsfall mit fehlenden Datenpunkten ist das Maß so nicht übertragbar. Stattdessen bietet sich z.B. die Krigingvarianz zur Abschätzung einer sinnvollen maximalen Entfernung von vorhandenen Datenpunkten an (Kapitel 11).

Auch ließe sich die Interpolationsvarianz anstelle der Krigingvarianz einführen (Yamamoto (2000)). Diese Größe ist abhängig vom Datenwert, vom Variogramm und repräsentiert die lokale Genauigkeit.

Weitere Krigingmethoden Das Ordinary Kriging setzt zu einem gewissen Maße die Stationarität der Daten voraus. In einer Chlorophyllverteilung ist der (räumliche) Erwartungswert aber selten konstant, wie es der Definition von Stationarität entspräche (vgl. Gl. 8.1). Die einfachste Abhilfe ist hier sicherlich die Verwendung der um den räumlichen Trend korrigierten Residuen im Ordinary Kriging. Zur Trendkorrektur wird hier die Verwendung von um den Untersuchungszeitpunkt zentrierten Monatsmitteln vorgeschlagen.

Als Lösung des nichtstationären Problems wird das *Universal Kriging* eingeführt, das als Hilfsvariablen den Ort berücksichtigt (Theorie z.B. in Olea (1999); Wackernagel (2003)). Ohne diese Methode hier dargestellt zu haben, sei gesagt, dass sie nicht zuverlässig bessere Ergebnisse liefert als ein Ordinary Kriging (zu Problemen beim Universal Kriging vgl. Armstrong (1984)) und zu Beginn dieser Arbeit als Methode verworfen worden ist.

Um nicht-stationären Datensätzen gerecht zu werden, werden Krigingmethoden vorgeschlagen, die z.B. Wavelet-Analyse zur Identifizierung der großskaligen Strukturen einsetzen und das Kriging nur auf den kleinskaligeren Fluktuationen des verbleibenden Residuums berechnen (Demyanov und Soltani (2001)) bzw. bei denen das Kriging im hochfrequenten Spektralbereich zum Einsatz kommt (Fuentes (2001)).

Eine neuere Entwicklung ist auch das 'Sequential Kriging' (Vargas-Guzman und Yeh (1999)) für große Datenmengen.

Während in dieser Arbeit das Krigingproblem auf ein rein räumliches reduziert wird, in dem zwar zeitliche Mittelwerte zur pixelweisen Bildung eines Residuums genutzt werden, die Operation selbst aber auf jegliche zeitliche Information verzichtet, gibt es Krigingmethoden, die explizit die räumlichen und zeitlichen Prozesse berücksichtigen. Das 'Space-Time Kriging' (z.B. Rouhani und Myers (1990); Bogaert (1996)) setzt entweder die Separabilität (Abschnitt 5.2) der Prozesse voraus und passt zeitliche und räumliche Variogramme getrennt an oder verwendet in einem Cokrigingansatz die Zeit als Hilfsvariable.

Für den derzeitigen Stand des Chlorophyllproduktes mit seinen systematischen Fehlern kann die Verwendung der zeitlichen Information nicht ohne weitreichende Korrekturen erfolgen und findet daher keine Anwendung in dieser Arbeit.

Übertragbarkeit auf einen lückenhaften Datensatz Ob das Variogramm basierend auf einer Auswahl bei fehlenden Datenpunkten das globale Variogramm unter- oder überschätzt, hängt primär davon ab, ob die Varianz des verbleibenden Datensatzes größer oder kleiner als die globale Varianz ist (siehe Abschnitt 9.2.2). Welche Trendkorrektur in Kombination mit der Krigingmethode die beste Rekonstruktion liefert, lässt sich ebenfalls an diesem Verhältnis festmachen. Im Normalfall ist die globale Varianz aber gerade *nicht* zugänglich - und was sich hier in den Kreuzvalidierungsexperimenten durch eine künstliche Wolkenmaske einfach erschließen lässt, ist bei tatsächlich lückenhaften Datensätzen als Vergleichsbasis nicht vorhanden.

Um im Anwendungsfall die globale Varianz eines Gebietes abzuschätzen, bietet es sich an, z.B. die Varianz aus klimatologischen (Monats-)Mitteln zu nutzen.

10 Kriging in der Praxis II - Nordsee

Dieses Kapitel behandelt die Übertragung der Versuche vom kleinen Kartenausschnitt der Deutschen Bucht mit einer einzigen künstlichen Wolkenmaskenform (Kapitel 9) auf großräumigere Datensätze des gesamten Nordseegebietes mit komplexen Wolkenmasken. Diese werden nach verschiedenen Vorschriften künstlich konstruiert (Abschnitt 10.1), um eine Bibliothek mathematisch definierter Strukturen zu schaffen, mit denen sich reale Wolkenformationen durch entsprechende Kombination beschreiben lassen. Dabei wird die Struktur der Wolkenbedeckung quantifiziert (Abschnitt 10.2 und Abschnitt 10.3) und mit der Güte der Rekonstruktion in Beziehung gesetzt (Abschnitt 10.4).

10.1 Künstliche Wolkenmasken

Um die räumliche Struktur der Wolkenverteilungen a priori klassifizieren zu können, werden künstliche Wolkenmasken (Abb. 10.1) hergestellt, die nach den folgenden Vorschriften aufgesetzt werden:

- “*Streifen*”: ein rechteckiger Streifen in Nord-Süd-Richtung. Auf weitere Varianten wie zwei oder drei parallele Streifen, die auch in Ost-West-, sowie NO-SW- bzw. NW-SO-Richtung verlaufen können, wird hier verzichtet.
- “*Klumpen*”: an $N \in \{50, 150\}$ zufällig ausgewählten Punkten über Wasser werden “Kreise” (in Pseudodistanzen aus Länge und Breite) als Wolken deklariert.
- “*Zufallsbewegung*”: von einem zufällig ausgewählten Wasserpixel entwickelt sich die Folge von Punkten gemäß einer homogen gleichverteilten Wahrscheinlichkeit für die Richtung und eines Potenzgesetz für die Wahrscheinlichkeit der Distanz $p(x)$ mit

$$p(x) = r_{min} (1 - x)^{-1/(\alpha-1)} \quad (10.1)$$

Der Grad der Häufung α , im Folgenden auch Ballungsstärke genannt, nimmt mit $\alpha \in \{2, 3, 4\}$ zu, da die Wahrscheinlichkeit für das Auftreten einer großen Distanz in der Abfolge der Punkte entsprechend abnimmt. In der Literatur ist diese Vorschrift auch als “Levi dust” zu finden (vgl. auch Mandelbrot (1983) oder Martinez und Saar (2002)).

Diese Vorschläge von zweidimensionalen Mustern basieren auf den 1D-Verteilungen nach Plotnick u. a. (1996) und sind entsprechend flächig erweitert worden.

Beispiele für Wolkenmasken Stellvertretend für die Wolkenmasken wird hier jeweils eine Realisation nach den oben genannten Vorschriften für 20% Bedeckung (über Wasser im Nordseegebiet, vgl. Abschnitt 2.2.1) gezeigt, die im weiteren Verlauf zum Krigingversuch herangezogen wird. Die Darstellung der *Zufallsbewegung*, $\alpha = 2$ (Abb. 10.1d), erscheint irreführend, der Bedeckungsgrad von 20% wird auch hier eingehalten.

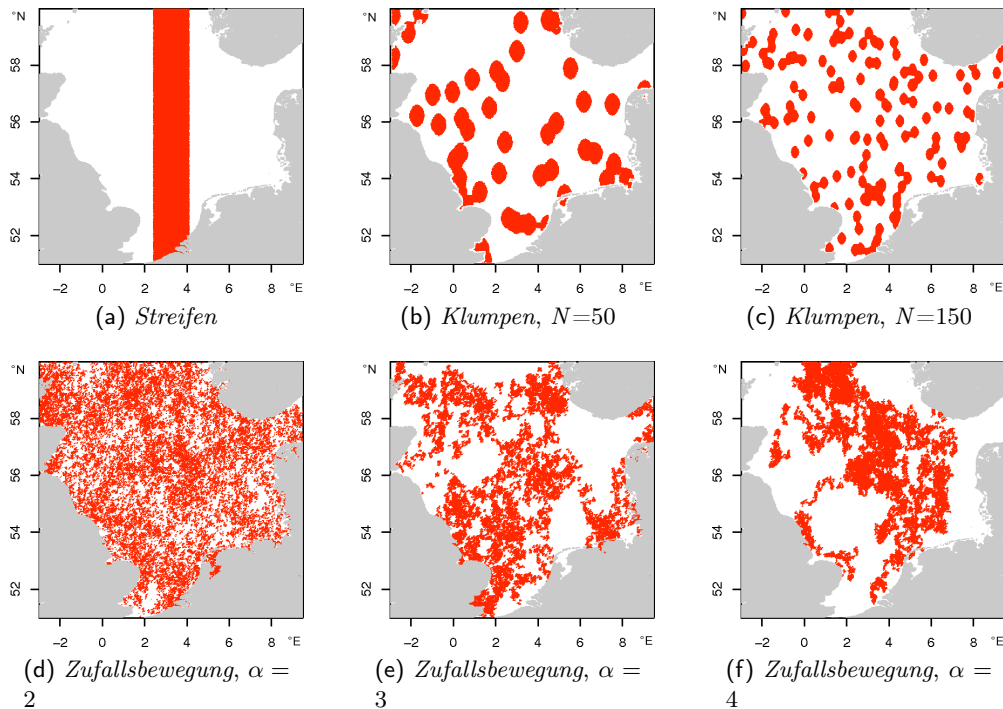


Abb. 10.1: Kartenausschnitt der Nordsee mit verschiedenen Wolkenmasken (rote Bereiche) zu jeweils 20% Bedeckung über Wasser (46034 von 230172 möglichen Wasserpixeln).

10.2 Beschreibung räumlicher Strukturen

Zur quantitativen Erfassung räumlicher Strukturen werden zwei auf binären Daten aufbauende Maße vorgeschlagen: zum einen die Distanzklassenverteilung (Abschnitt 8.2.1) und zum zweiten die Lückenhaftigkeit (engl. lacunarity).

Binär heißt hier, dass neben der räumlichen Information jedem Gitterpunkt nur die Eigenschaft gültig oder nicht-gültig zugewiesen wird. In diesem Sinne ist ein gültiges Wolkenpixel jedes Pixel über Wasser, für das kein Chlorophyllwert bestimmt werden kann. Auch wenn hier weiterhin von 'Wolkenpixeln' gesprochen wird, handelt es sich strenggenommen um die Menge aller ungültigen Datenpunkte.

10.2.1 Distanzklassenverteilung

Im Zuge der Berechnung eines experimentellen Variogramms wird die Anzahl von Punkt-paaren innerhalb einer Distanzklasse erhoben, die als einfache Zwei-Punkt-Statistik zur Beschreibung der räumlichen Verteilung der betrachteten Datenpunkte dienen kann. Zur Quantifizierung der Wolkenstruktur werden hier statt der gültigen Datenpunkte die Positionen der Wolkenpixel verwendet.

In einer zufälligen Auswahl von N Datenpunkten aus der Menge der gültigen Wolkenpixel werden alle paarweisen Kombinationen gebildet und die Anzahl von Paaren pro Distanzklasse erfasst. Anstelle der Beschränkung der Distanzklasse wie in der Variogrammberechnung werden nun Distanzen bis zur maximal möglichen Entfernung zweier Punkte im Untersuchungsgebiet zugelassen. Damit werden auch die geringer besetzten Klassen größerer Distanzen bzw. die theoretisch maximal mögliche Anzahl $N_{max} = \frac{1}{2}N \cdot (N - 1)$ von Paaren erfasst. Das gewährt die Möglichkeit zum quantitativen Vergleich verschiedener Strukturen in einem Untersuchungsgebiet bei $N = \text{const.}$ und gleicher Einteilung der Distanzklassen.

Für die Beispiele wird aus den 46034 Wolkenpixeln eine Stichprobe von $N = 4000$ Punkten zufällig ausgewählt, deren Paare bis zur maximalen Distanz von $h_{max} = 11$ (gemäß der Pseudodistanz aus Koordinaten in geographischer Länge und Breite) in 111 Distanzklassen mit der Breite $\Delta h = 0.1$ gezählt werden. Die ersten Distanzklassen enden jeweils bei 0.065, 0.15, 0.25, 0.35, etc. Die Zählungen werden auf $N_{max} = 7998000$ normiert. Neben dem omnidirektionalen Variogramm wird die Zählung auch in die Winkelintervalle um die 0° -Richtung (gen Nord zwischen -45° bis $+45^\circ$) und 90° -Richtung (gen Ost zwischen 45° bis 135°) aufgeteilt (Abb. 10.2).

Der Verlauf der Distanzklassenverteilung für die *Streifen*-Formation lässt sich wie folgt erklären (Abb. 10.2a): der Streifen erstreckt sich in Nord-Süd-Richtung und besitzt in Ost-West-Richtung eine Breite von 1.677. Das Maximum in der gerichteten Verteilung (90°) tritt gerade bei $h = 0.850$ auf, was bei der Auflösung von $\Delta h = 0.1$ sehr nah an der halben maximalen Breite liegt. Da das Maximum in der 0° -Verteilung nicht mit dem der 90° -Verteilung zusammenfällt, wird die Struktur räumlich inhomogen sein.

In den Beschreibungen der *Klumpen*-Formationen (Abb. 10.2b, Abb. 10.2c) lässt sich nur das erste relative Maximum als durch die Ausdehnung der Wolkenklumpen entstanden interpretieren. Unter Berücksichtigung der Auflösung von $\Delta h = 0.1$ fallen diese Maxima bei $h = 0.35$ für die *Klumpen*, Anzahl 50, mit mittlerem Radius von 0.37 bzw. bei $h = 0.25$ für die *Klumpen*, Anzahl 150, mit mittlerem Radius von 0.19 jeweils annähernd zusammen. Auch die nachfolgenden relativen Minima bei $h = 0.65$ bzw. $h = 0.45$ entsprechen den jeweiligen Durchmessern in dem Maße, um darin den Einfluss der grundlegenden Struktur wiederzuerkennen. Die Abweichungen ergeben sich dadurch, dass Überlappungen von zwei Klumpen auftreten können, die weitere Punktpaare im Bereich kurzer Distanzen möglich machen, letztlich aber nicht rein durch die primäre Form der Einzelstrukturen entstehen.

Der Vergleich der Distanzklassenverteilung einer rein homogen, zufälligen Verteilung (in Abb. 10.2b-f als "*Zufall*" bezeichnete, gestrichelte Linie) mit einer *Zufallsbewegung* der Ballungsstärke $\alpha = 2$ zeigt nur minimale Unterschiede, sodass letztere Struktur mit ihrer

sehr geringen Häufung als annähernd homogen, zufällig verteilt gelten kann (Abb. 10.2d).

Im direkten Vergleich der Zufallsverteilung mit einer mittleren Distanz zwischen Punktpaaren von 3.9 lässt sich der uneinheitliche Verlauf für die *Klumpen*, Anzahl 50, gut als Überlagerung zweier Effekte interpretieren: die Maxima, die durch die Häufung (relativ schmale Maxima wie bei der *Streifen*-Struktur) entstehen, überlagern den zugrunde liegenden Zufallsprozess (vgl. Abb. 10.2b). Daher sind bei der *Klumpen*-Struktur, Anzahl 150, die Häufungseffekte wesentlich weniger ausgeprägt, während der zufällige Prozess deutlich zu sehen ist (Abb. 10.2c).

Gegenüber der rein zufälligen Verteilung der *Zufallsbewegung*, $\alpha = 2$, wirkt sich bei der *Zufallsbewegung*, $\alpha = 3$, die lockere Häufung so aus, dass die Punktpaare im Mittel weiter auseinander liegen als im rein zufälligen Fall (häufigste Distanz $h = 4.3$, Abb. 10.2e). Bei der *Zufallsbewegung*, $\alpha = 4$, wird die Häufung gegenüber der zufälligen Verteilung so stark, dass die mittlere Distanz nun nur noch $h = 2.2$ beträgt (Abb. 10.2f). Mit zunehmender Häufung wird der Unterschied zu einer rein zufälligen Struktur immer größer.

Durch die Überlagerung mehrerer Effekte wird die Interpretation der Distanzklassenverteilung oftmals nicht eindeutig sein.

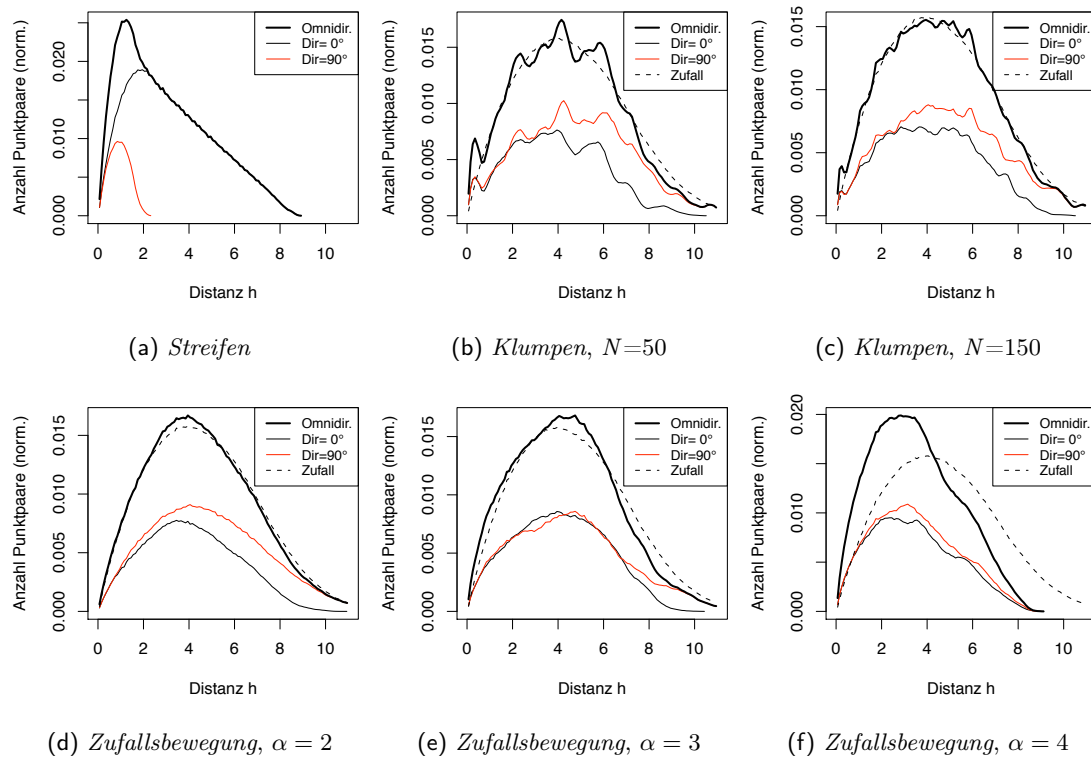


Abb. 10.2: Distanzklassenverteilung N_h für die Wolkenstrukturen in Abb. 10.1. Zum Vergleich wird für Strukturen mit Zufallsprozessen die Distanzklassenverteilung einer rein zufälligen Struktur (*Zufall*) eingezeichnet.

10.2.2 Lückenhaftigkeit

Ursprünglich wird die Lückenhaftigkeit (engl. lacunarity, von lat. lacuna, "Loch") als Maß zur Beschreibung der Eigenschaften von Fraktalen eingeführt (Mandelbrot (1983), Gefen u. a. (1983)). Als skalenabhängiges Maß für Heterogenität oder Textur eines Objektes misst sie die Abweichung eines Datensatzes, der nicht notwendigerweise fraktales Verhalten zeigen muss, von der (skalenabhängigen) Translationsinvarianz.

Anschaulich bedeutet dies: homogene Verteilungen besitzen eine geringe Lückenhaftigkeit, während stark aggregierte Verteilungen mit sehr unterschiedlicher Ausdehnung der Lücken eine hohe Lückenhaftigkeit besitzen.

Erweitert auf den zweidimensionalen Fall stellt sich der Algorithmus ("gliding box", Allain und Cloitre (1991), Plotnick u. a. (1996)) zur Berechnung der Lückenhaftigkeit wie folgt dar:

Basierend auf den Standardgitterpixeln werden quadratische Fenster definiert, deren Kantenlänge r in Pixel bemessen wird. Die Anzahl s an Wolkenpixeln im Fenster wird bestimmt und das Fenster über den gesamten Untersuchungsraum verschoben. Die Frequenzverteilung $n(s, r)$ - d.h. die Anzahl n von Fenstern der Größe r , in denen gerade s Wolkenpixel gefunden werden - wird mittels der Gesamtzahl $N(r)$ der Fenster in eine Wahrscheinlichkeitsverteilung $Q(s, r)$ umgerechnet, $Q(s, r) = \frac{n(s, r)}{N(r)}$.

Erstes und zweites Moment dieser Verteilung werden bestimmt:

$$\text{Mittelwert: } Z_1(r) = \sum sQ(s, r) = \overline{s(r)} \quad (10.2)$$

$$\text{Varianz: } Z_2(r) = \sum s^2Q(s, r) = s_s^2(r) + \overline{s^2(r)} \quad (10.3)$$

Die Lückenhaftigkeit $\Lambda(r)$ als Funktion der Fenstergröße r wird dann definiert als das Verhältnis von Varianz zu quadriertem Mittelwert

$$\Lambda(r) = \frac{Z_2}{Z_1^2} \quad (10.4)$$

oder auch

$$\Lambda(r) = \frac{s_s^2(r)}{\overline{s^2(r)}} + 1 \quad (10.5)$$

Die Lückenhaftigkeit ist eine Funktion der folgenden Größen:

- Anteil von Wolkenpixel $P = S/M$, mit Gesamtzahl von Wolkenpixel S und Gesamtzahl von Pixel M . Spärlich besetzte Verteilungen werden eine höhere Lückenhaftigkeit besitzen: es werden vornehmlich die Wahrscheinlichkeiten von kleinen Anzahlen n hoch sein, sodass der Mittelwert Z_1 klein ist. Ein kleinerer Mittelwert bedingt eine umso größere Lückenhaftigkeit.
- Größe des Fensters r . Außer für stark geclusterte Verteilungen werden große Fenster immer stärkere Translationsinvarianz zeigen als kleine, d.h. das zweite Moment

nimmt relativ zum ersten ab. Dabei ist $Q(1, 1) = P$ und $\Lambda(1) = 1/P$ nur abhängig vom Anteil an Wolkenpixel.

- Geometrie der Pixelverteilung. Bei gegebenem Anteil P und Fenstergröße r ist eine höhere Lückenhaftigkeit gleichbedeutend mit stärkerer Ballung der betrachteten Verteilung.

10.3 Berechnung der Lückenhaftigkeit in der Praxis

Einige Anpassungen werden am Gliding-Box-Algorithmus vorgenommen, um die Rechenzeit zu verkürzen.

- Da die Lückenhaftigkeit als Funktion der Fenstergröße sinnvollerweise logarithmisch dargestellt wird, wird nicht jede Fenstergröße berechnet, sondern lediglich jene von solcher Größe, sodass sie auf einer logarithmischen Skala annähernd gleichverteilt liegen.
- Zusätzlich wird die Gesamtheit aller vollständig im Untersuchungsgebiet liegenden Fenster auf einen Bruchteil (5% des Maximalwertes, je nach Größe) reduziert. Die Fenster werden zufällig ausgewählt. Sie besitzen ganzzahlige, quadratische Ausdehnungen, die von den Gitterpixeln festgelegt werden. Von der Größe des Fensters hängt die maximale Anzahl ab.

Für je eine Wolkenmaske W (Methode, Bedeckungsgrad) werden jeweils zehn Sätze (5% aller möglichen Fenster einer Größe) zufällig gewählter Fenster ausgewertet. Die Standardabweichung von der mittleren Lückenhaftigkeit ist gering (Abb. 10.3), in Bereichen kleiner Skalenlängen fällt sie höher aus als auf den größeren Skalenlängen. Insgesamt ist die Variabilität aber so gering, dass die reduzierte Fensterzahl ohne weiteres akzeptiert werden kann.

Wie kommt die spezifische Abhängigkeit der Lückenhaftigkeit mit der Fenstergröße zustande?

Die einfachste Einsicht gewährt das Beispiel der *Zufallsbewegung*, $\alpha = 2$ (Abb. 10.3d). Die Punktverteilung unterscheidet sich kaum von einer homogen zufällig verteilten Wolkenmaske (vgl. Abb. 10.4a und Abb. 10.4d) und besitzt annähernd konstante Punktdichte im gesamten Gebiet. Je größer die Fenster werden, umso wahrscheinlicher repräsentieren die betrachteten Punkte gerade die mittlere Punktdichte. Bei 50% Bedeckung wird die gesehene Punktdichte bald unabhängig von der Fenstergröße. Es gibt nur noch wenige Klassen s , die überhaupt besetzt sind, vornehmlich die eine, die gerade dem Überdeckungsgrad entspricht. Damit wird $Q(s, r) = \text{const.}$ und die Lückenhaftigkeit strebt gegen Eins; $\Lambda(r) \rightarrow 1$.

Stark gehäufte Verteilungen (Abb. 10.3a, Abb. 10.3f) ermöglichen Fenster, in denen sich nur Lücke befindet ($s = 0$); diese werden für die Statistik nicht gezählt. Für den breiten *Streifen* gibt es drei mögliche Szenarien für ein Fenster: es befindet sich im Randbereich des Streifens und sieht daher Lücke und Streifen, es liegt vollständig im Streifen, oder es ist so groß, dass es auf beiden Seiten eine Lücke sieht. Mit zunehmender Breite

des Streifens bei steigender Überdeckung bleibt die Wahrscheinlichkeit für Fenster über einem Streifenrand gleich, jedoch liegen sie bis zu größeren Fenstergrößen vollständig in den Streifen. Deshalb setzt der Abfall zu geringerer Lückenhaftigkeit bei steigendem Bedeckungsgrad zunehmend später ein.

Die übrigen Verteilungen stellen Mischungen und Übergangszustände dar. Die *Zufallsbewegungen* mit $\alpha = 3$ oder $\alpha = 4$ besitzen gegenüber der zufälligen Verteilung der *Zufallsbewegung*, $\alpha = 2$, jeweils die höhere Lückenhaftigkeit. Wie optisch und nach Vorschrift zu erwarten, nimmt mit der Ballung die Lückenhaftigkeit zu. Bei den *Klumpen*, Anzahl 150 (Abb. 10.3c), setzt der Abfall zu geringerer Lückenhaftigkeit früher ein als im Fall der *Klumpen*, Anzahl 50 (Abb. 10.3b). Dies ist in sofern anschaulich zu verstehen, als die *Klumpen*-Formation, Anzahl 150, mit vielen kleinen Wolkenklumpen ab einer geringeren Skalenlänge einer rein zufälligen Verteilung gleicht.

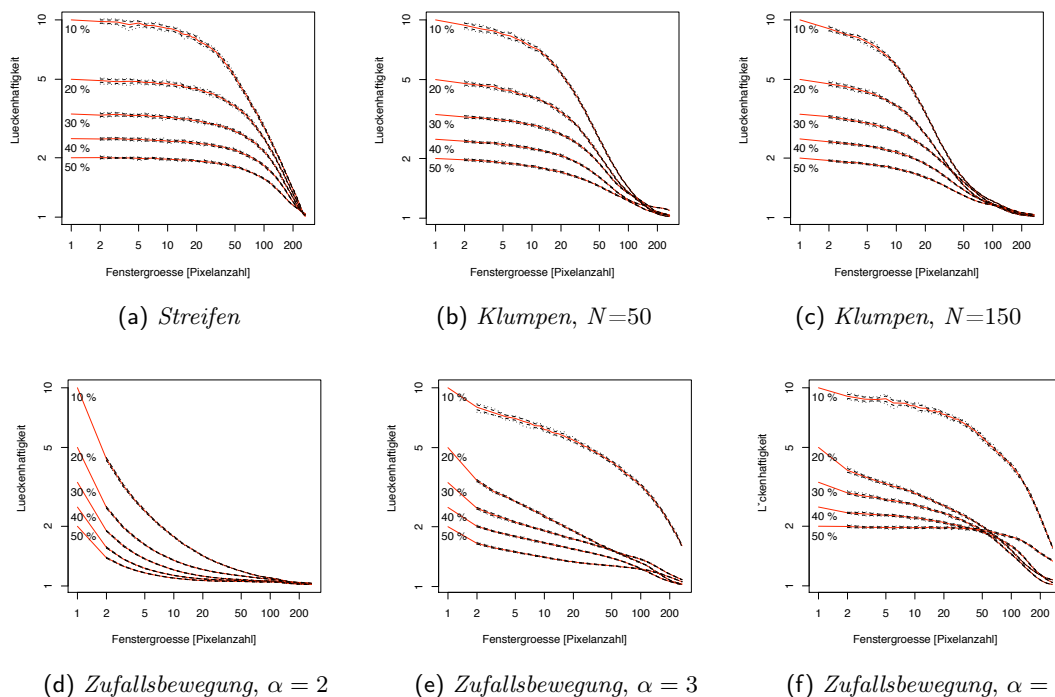


Abb. 10.3: Variabilität der Lückenhaftigkeit durch zufällige Auswahl der Untersuchungsfenster bei konstanter Wolkenmaske. Aus 10 Sätzen von Fenstern wird die mittlere Lückenhaftigkeit (schwarze Linie) und das Intervall der Standardabweichung (rote Linien) bestimmt.

Die Lückenhaftigkeit der *Zufallsbewegungen* $\alpha = 3$ und $\alpha = 4$ mit einem Bedeckungsgrad von 10% verhalten sich gegenüber den restlichen Mustern dieser Gruppe anders (Abb. 10.3e, Abb. 10.3f). Es kann sich hier um Randeffekte bei der Anwendung des Gliding-Box-Algorithmus handeln (Feagin u. a. (2007)), die bei geringem Bedeckungsgrad kombiniert mit starker Heterogenität auftreten können.

Nachdem die weitestgehende Konstanz der Lückenhaftigkeit trotz zufälliger Auswahl der Fenster gezeigt ist, bleibt die Frage, ob die Muster mit zufälligen Anteilen sich auch in unterschiedlichen Realisationen eindeutig durch die Lückenhaftigkeit unterscheiden lassen. Dazu werden gemäß der Vorschriften jeweils 10 Wolkenrealisationen zu 20% Bedeckung aufgesetzt und ihre jeweilige Lückenhaftigkeit bestimmt.

Selbst bei Strukturen, die auf Zufallsprozessen beruhen, variiert sie teilweise nur geringfügig zwischen verschiedenen Musterrealisierungen (Abb. 10.4). Die großen Streuungen in der *Zufallsbewegung*, $\alpha = 4$, können wie zuvor durch Randlagen der Struktur erklärt werden.

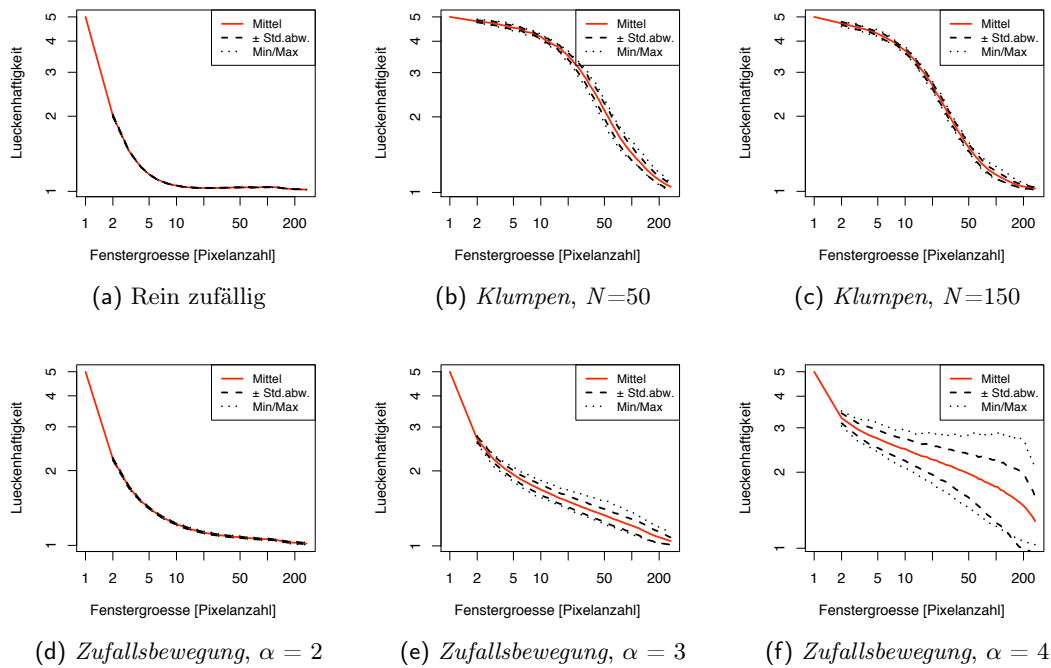


Abb. 10.4: Variabilität der Lückenhaftigkeit bei Wolkenmustern mit zufälligen Anteilen, *Klumpen* und *Zufallsbewegung*, je 10 Realisationen zu 20% Bedeckungsgrad.

Bei geringer Ballungsstärke ($\alpha = 2$) verhält sich die *Zufallsbewegung*-Verteilung ähnlich einer homogen zufällig verteilten, so dass sich verschiedenen Realisierungen kaum in ihrer Struktur unterscheiden. Je grober die Strukturelemente eines zufälligen Musters, umso stärker streut die Lückenhaftigkeit zwischen den Realisationen. Die Formation der *Klumpen*, Anzahl 150 (zufällige Verteilung 150 gleichgroßer Klumpen) und der *Zufallsbewegung*, $\alpha = 2$ (homogen zufällig verteilt), streuen deutlich weniger als die *Klumpen*-Struktur, Anzahl 50, wo mit weniger aber größeren Klumpen die gleiche Bedeckung erreicht wird. Mit zunehmender Ballungsstärke α steigt nicht nur die Lückenhaftigkeit: durch eine eventuelle Randlage der geclusterten Punkte streut sie auch wesentlich stärker. Dennoch lassen diese zehn Realisierungen pro Mustervorschrift den Schluss zu, dass sich die Strukturen auf bestimmten Skalenbereichen immer deutlich unterscheiden und

die Möglichkeit einer eindeutiger Zuordnung besteht (Abb. 10.5).

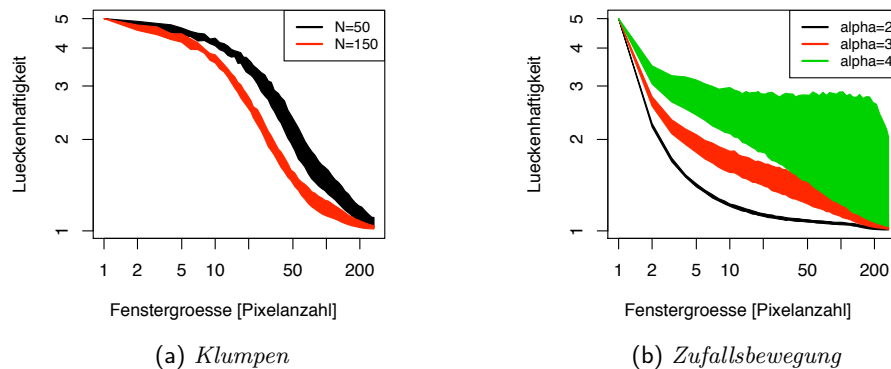


Abb. 10.5: Bereiche der Lückenhaftigkeit für Wolkenmasken mit 20% Bedeckungsgrad. Die Flächen zeigen den Bereich zwischen minimaler und maximaler Lückenhaftigkeit aus 10 Realisationen der Wolkenstrukturen. Bei den *Klumpen* wird zwischen Strukturen aus 50 oder 150 Einzelformen unterschieden, bei der *Zufallsbewegung* variiert die Ballungsstärke zwischen $\alpha = 2, 3, 4$.

10.4 Krigingexperiment mit künstlichen Wolkenmasken

Die kommenden Versuche orientieren sich an der Bewertung der besten Trendkorrektur für die Erstellung der Residuen oder der besten Krigingmethode anhand der Fehlerstatistik, wie sie zuvor am Beispiel der Deutschen Bucht durchgeführt worden sind (Abschnitt 9.5). Dabei werden die tatsächlichen Fehler der Schätzwerte im Folgenden speziell im Kontext der Wolkenmaskenstruktur gesehen. Um ein (annähernd) einfach zusammenhängendes Gebiet untersuchen zu können, wird das Standardgitter auf den expliziten Teilbereich der Nordsee beschränkt (Abb. 2.3).

10.4.1 Aufbereitung der Daten

Die Untersuchung basiert auf dem Chlorophyllprodukt des Regionalprozessors vom 10. Mai 2006, sowie auf den vorhandenen Datentagen im Zeitraum von ± 15 Tagen. Für dieses Beispiel wird der Ansatz gewählt, die Gültigkeit der Datenpunkte mittels der Summe der Gesamtstrahldichte (Abschnitt 4.2.2) festzulegen. Für jedes Bild wird eine spezielle Grenze gewählt, unterhalb derer die Daten als nicht durch unerkannte Wolken verfälscht angenommen werden. Bei dieser Festlegung ist das subjektive Augenmaß ausschlaggebend gewesen.

Nach einer ersten Übersicht über die Datentage steht fest, dass der Versatz der Chlorophyllkonzentration (Abschnitt 5.5) sich signifikant auf das Ergebnis auswirkt. Somit wird hier eine grobe Korrektur der Daten vor der Berechnung der Mittelwerte versucht.

Dazu wird ein Ausschnitt in der zentralen Nordsee gewählt (zwischen $3-4^\circ \text{O}$, $56-56.5^\circ \text{N}$), von dem gute Datenüberdeckung und relativ geringe Strukturiertheit zu er-

warten ist. Die mittlere Chlorophyllkonzentration zeigt Abhängigkeit vom Sonnenazimut (Abb. 10.6). Gemäß des angedeuteten linearen Zusammenhangs werden die Konzentrationen um die Differenz zwischen Chlorophyllwerten auf der Gerade vom jeweiligen Datentag und dem Bezugstag, dem 10. Mai 2006, korrigiert. Dies ist jedoch nur ein grobes Näherungsverfahren, da der räumliche Chlorophyllmittelwert natürlich auch von biologischen Prozessen abhängig ist und der Sonnen(azimut)winkel zusätzlich von West nach Ost ansteigt (vgl. Profile zur Winkelabhängigkeit, Abb. 5.9, bzw. systematischer Trend über die Pixel in x-Richtung, Abb. 5.11 und Abb. 5.12). Für längere Zeiträume ist dieser Ansatz daher nicht geeignet.

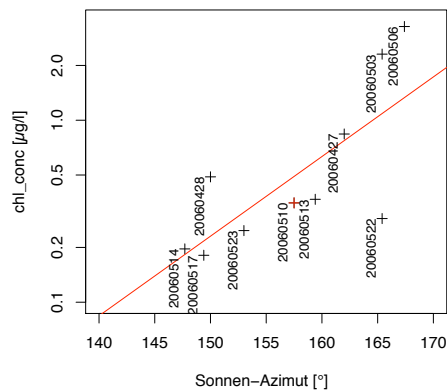


Abb. 10.6: Median der Chlorophyllkonzentration (log. Skala) in Abhängigkeit vom Sonnenazimut für die Datentage zwischen dem 27. April und 23. Mai 2006. Der Bezugstag für die Korrektur ist der 10. Mai. Die Datenpunkte sind mit dem Datum in der Form JJJJMMTT beschriftet.

Diese Satellitenszene (gültiger Datenanteil im Gesamtgebiet: 204378 von 230172 Wasserpixeln, das entspricht 89%) wird mithilfe des Cokriging unter Einbeziehung der Wassertiefe vervollständigt. Ebenso werden alle acht Monatsmittel¹ durch das Cokriging unter Berücksichtigung der Wassertiefe (CoOK.tief) ergänzt.

Die Monatsmittel aus versatzkorrigierten Daten erscheinen etwas glatter gegenüber dem unkorrigierten Fall. Insbesondere sind die Grenzen der einzelnen Bilder im unkorrigierten Fall deutlich im zentralen Nordseebereich, sowie einige Artefakte sehr hoher oder sehr niedriger Chlorophyllkonzentration in unmittelbarer Nachbarschaft zueinander zu erkennen (Abb. 10.7).

¹Folgende Tage tragen zum Gesamtbild bei: 27., 28. und 30. April sowie 1.-7., 13., 14., 16., 17., 19., 20., 22. und 23. Mai 2006.

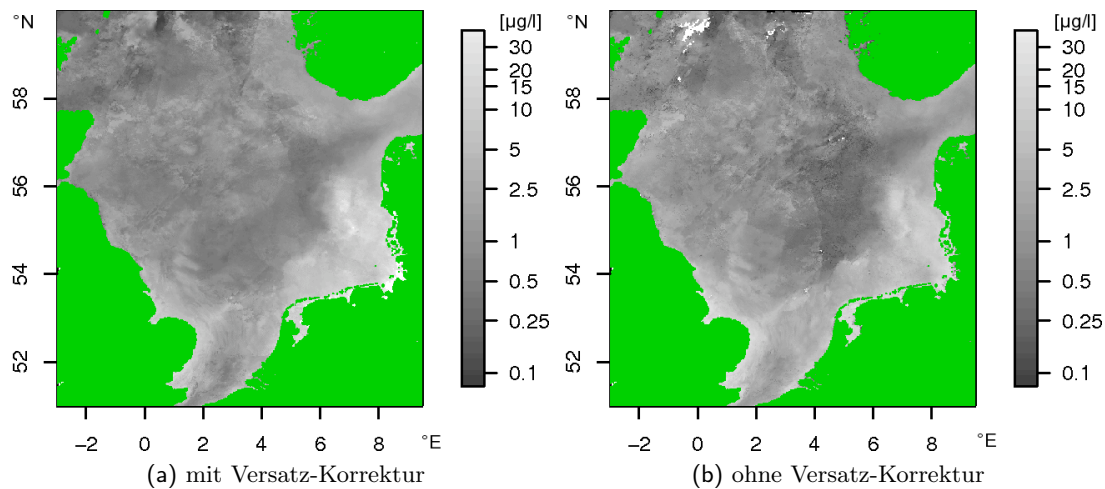


Abb. 10.7: Median größter Wahrscheinlichkeit (*MLH.Median*) für um den Versatz korrigierte und unkorrigierte Datensätze als Monatsmittel der Chlorophyllkonzentration in $\mu\text{g/l}$ zentriert um den 10. Mai 2006.

10.4.2 Beste Trendkorrektur

Wie im kleinräumigen Beispiel (Kapitel 9) wird hier erneut getestet, welche Trendkorrektur zum besten Ergebnis bezüglich der tatsächlichen Fehler der Schätzwerte führt. Die Wolkenmasken als Funktion des Musters und des Bedeckungsgrades (*Wolkenmaske* (Muster, %), Definitionen in Abschnitt 10.1, Beispiele für u.a. verwendete Realisationen in Abb. 10.1) werden wie zuvor beschrieben als sechs Mustervorschriften mit Wolkenbedeckungsgraden von 10, 20 und 30% den Originaldaten überlagert. Anschließend wird das Ordinary Kriging der Residuen für die verschiedenen Trendkorrekturen durchgeführt.

Die Fehlermaße mittlerer absoluter Fehler und absoluter Median des Fehlers werden für eine Wolkenmaske, d.h. mit konstantem Bedeckungsgrad und konstantem Muster, normiert, sodass die Rangfolge der Trendmethoden ablesbar wird. Da jede Wolkenmaske andere Datenpunkte ausschließt, werden die Punkte für die Variogramme jeweils pro Wolkenmaske neu gewählt. Die Vergleiche enthalten daher auch die volle Variabilität durch die Probenahme für das Variogramm.

Mehrere Aussagen lassen sich aus diesem Versuch ableiten:

1. Berechnet man die mittlere Güte aus allen Mustern und allen Bedeckungsgraden, sodass nur noch ihre Abhängigkeit von der Trendkorrektur betrachtet wird, so zeigt sich, dass das Ordinary Kriging gegenüber den Methoden des Kriging der Residuen genauere Ergebnisse liefert (Tabelle 10.1).
2. Die mittlere Güte, jeweils in Abhängigkeit von einer Klasse von Mustern und dem Überdeckungsgrad, gibt einen detaillierten Einblick (Tabelle 10.2): Mit zunehmendem Bedeckungsgrad lassen sich die Wolkenstrukturen anhand der Güte trennen (Abb. 10.8). Dabei kann die Formation eines *Streifens* immer deutlich schlechter

gegenüber den übrigen rekonstruiert werden. Die *Klumpen*-Struktur führt zu einer starken Streuung in der Über- oder Unterschätzung der Werte, während die Fehlermediane der *Zufallsbewegung*-Formationen (ab 20% Bedeckungsgrad) im Vergleich nur wenig um den Wert Null streuen. Dabei lassen sich Regelmäßigkeiten bezüglich der Trendkorrekturen finden, wenn diese Methoden nach Wolkenstrukturen aufgeschlüsselt betrachtet werden (Abb. 10.9).

3. Für eine globale Bewertung des Einflusses der Wolkenmuster auf die Rekonstruktionsgüte wird die Normierung der Fehlergrößen jeglicher Trendkorrekturen noch einmal über alle Wolkenmasken jeglicher Bedeckungsgrade wiederholt. Die mittlere Güte je nach Wolkenmuster, aber unabhängig von Trendkorrektur und Bedeckungsgrad, zeigt deutlich (Tabelle 10.3): mit der Stärke der Häufung sinkt die Güte der Rekonstruktion.

Im Allgemeinen gilt, dass die beste Rekonstruktion bei der Wahl zwischen einem beliebigen Ordinary Kriging der Residuen oder reinem Ordinary Kriging immer mit letzterem gelingt, sei es unabhängig von der Wolkenstruktur oder in jedem Fall innerhalb einer Gruppe von verschiedenen Trendkorrekturen bei konstanter Wolkenstruktur und Bedeckungsgrad. In dieser speziellen Realisation gibt es nur eine Ausnahme für *Klumpen*, 20%, Anzahl 50. Für diese Struktur schneidet das Ordinary Kriging der logarithmierten Chlorophyllkonzentrationen am schlechtesten ab.

Methode	OK.log.chl	Mean.log		Mean		MLH.Mean		MLH.Median	
		-	gew.	-	gew.	-	gew.	-	gew.
Güte	0.3082	0.6633	0.6714	0.565	0.5795	0.5907	0.5971	0.5551	0.5616
Std.abw.	0.3041	0.3189	0.3211	0.3245	0.3198	0.3061	0.2569	0.2807	0.2632

Tab. 10.1: Beste Trendkorrektur über alle Wolkenmasken unabhängig vom Bedeckungsgrad. Mittelung der Fehlerstatistik jeweils für eine Trendkorrekturmethode. Die Fehlerstatistik der Schätzwerte ist in Abb. 10.9 illustriert.

Wolkenbedeckung	10%	20%	30%
<i>Streifen</i>	0.6548 ± 0.1652	1.1425 ± 0.1588	1.1551 ± 0.1241
<i>Klumpen</i>	0.6971 ± 0.246	0.7265 ± 0.1584	0.6602 ± 0.1174
<i>Zufallsbewegung</i>	0.7396 ± 0.31	0.3234 ± 0.1262	0.3978 ± 0.1433

Tab. 10.2: Güte der Rekonstruktion je nach Wolkenmuster und Bedeckungsgrad unabhängig von der Trendkorrektur. Dies ist die Quantifizierung zur Fehlerstatistik in Abb. 10.8

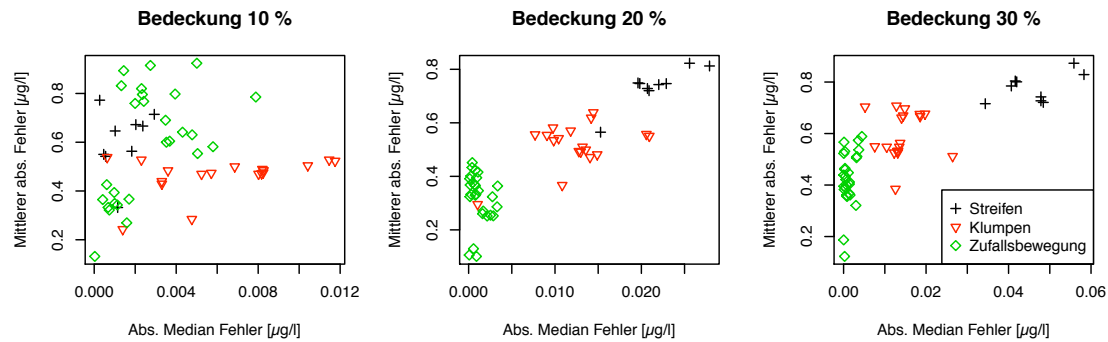


Abb. 10.8: Fehlerstatistik der Schätzwerte verschiedener Verfahren zu Chlorophylldaten vom 10. Mai 2006 für verschiedene künstliche Wolkenmasken, die sich in Bedeckungsgrad und Struktur unterscheiden. Quantifizierung der Güte bezogen auf Wolkenmuster findet sich in Tabelle 10.2.

	<i>Streifen</i>	<i>Klumpen</i>		<i>Zufallsbewegung</i>		
Wolkenmuster	1	Anz = 50	Anz = 150	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$
Güte	0.8562	0.5722	0.5368	0.3158	0.3911	0.5224
Std.Abw.	0.2539	0.1376	0.109	0.1062	0.217	0.2786

Tab. 10.3: Güte der Rekonstruktion allein abhängig vom Wolkenmuster, ohne Berücksichtigung der Bedeckung oder der Trendkorrektur. Basiert auf der Fehlerstatistik in Abb. 10.9.

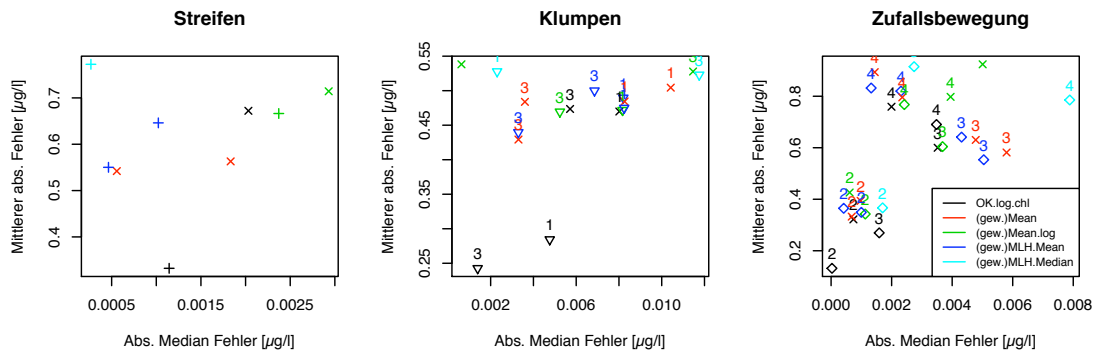
10.4.3 Beste Krigingmethode

Mit der zuvor bestimmten besten Trendkorrektur, das sind die Monatsmittelwerte gemäß dem Mittelwert bzw. Median größter Wahrscheinlichkeit (*MLH.Mean* und *MLH.Median*), wird erneut die Güte der Schätzwerte aus unterschiedlichen Krigingmethoden überprüft. Dabei wird jeweils die Fehlerstatistik zum Kriging der Residuen, *ResOK*, bzw. zum Cokriging der Residuen mit der Wassertiefe, *ResCoOK.tief*, mit der des Kriging der logarithmierten Chlorophyllkonzentration, *OK.log.chl*, bzw. dem Cokriging dieser Variable unter Zuhilfenahme der Tiefe, *CoOK.tief*, verglichen. Als Wolkenmasken dienen nur diejenigen für 20% Wolkenbedeckung (Abb. 10.1). Die Daten stammen vom 10. Mai 2006.

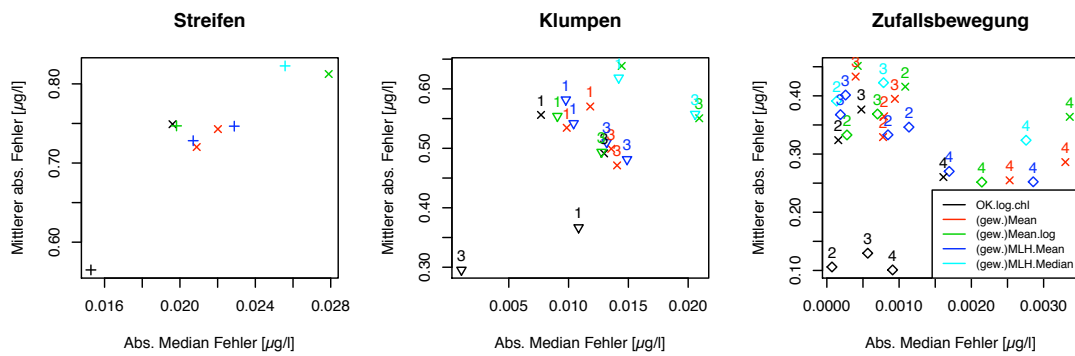
Normiert werden die Fehlergrößen hier jeweils für eine konstante Wolkenmaske über alle zu vergleichenden Krigingmethoden. Unter Berücksichtigung aller Wolkenmasken gelingt die Rekonstruktion am besten mit *CoOK.tief*, gefolgt von *OK.log.chl* (Tabelle 10.4).

Der breite Wolkenstreifen bleibt weiterhin die Struktur, die am schlechtesten zu rekonstruieren ist (siehe Abb. 10.10a). Je größer die zusammenhängenden Gebiete fehlender Daten sind, umso schlechter ist auch die Rekonstruktion bei *Klumpen*-Strukturen. Für die Wolkenmasken nach der *Zufallsbewegung* lassen sich die besten Rekonstruktionen erwarten.

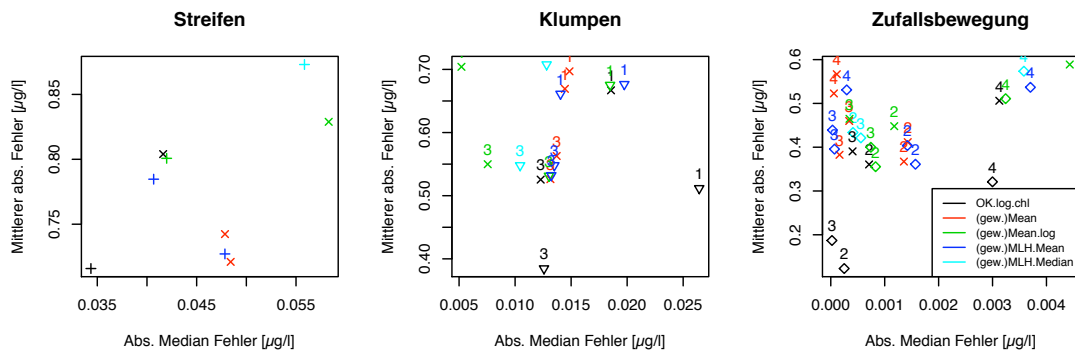
Die Aussage zur besten Krigingmethode lässt sich wiederum nur unter Vorbehalt treffen: sie basiert auf den gewählten Repräsentanten für die Wolkenstrukturen und auf der



(a) Bedeckungsgrad 10%, aufgeschlüsselt nach Trendkorrektur und Wolkenstruktur



(b) Bedeckungsgrad 20%, aufgeschlüsselt nach Trendkorrektur und Wolkenstruktur



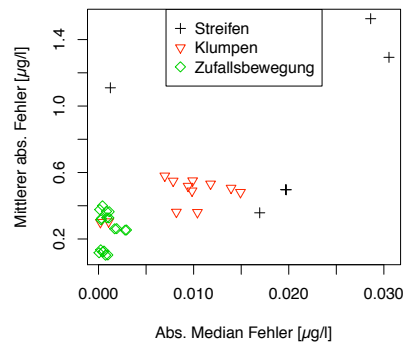
(c) Bedeckungsgrad 30%, aufgeschlüsselt nach Trendkorrektur und Wolkenstruktur

Abb. 10.9: Beste Trendkorrektur anhand der Bewertung tatsächlicher Fehler der ResOK-Schätzwerte. Verschiedene Wolkenstrukturen und Bedeckungsgrade werden hier im Vergleich zu Abb. 10.8 getrennt aufgezeigt. Trendkorrekturen werden in (a) - (c) nach Farben unterschieden, die Markierungen × stehen für die gewichtete Variante. Die Quantifizierung der besten Trendkorrektur findet sich in Tabelle 10.1. Die Güte der Rekonstruktion in Abhängigkeit von der Wolkenstruktur ohne Berücksichtigung des Bedeckungsgrades findet sich in Tabelle 10.3.

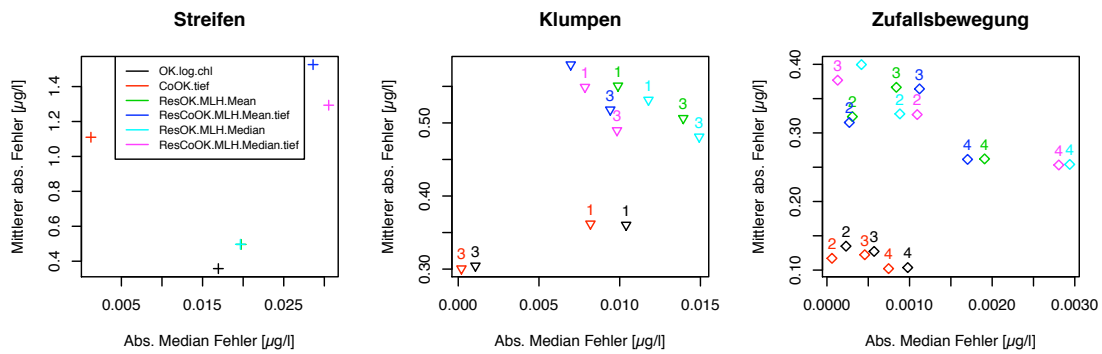
räumlichen Statistik des speziellen Datentages.

Innerhalb dieses Rahmens gilt (Abb. 10.10b):

- *Streifen*: Hier werden für *OK.log.chl* und *CoOK.tief* mehrere Variogrammrealisationen berücksichtigt, die erhebliche Streuung der Schätzwerte in Abhängigkeit vom Variogramm nahe legen. Besonders auffällig ist der große Unterschied zwischen den Realisationen von *CoOK.tief*. Beste Methode ist hier recht deutlich das *OK.log.chl*, gefolgt von *ResOK.MLH.Mean* bzw. *ResOK.MLH.Median*.
- *Klumpen*: Für die stärker geballte Verteilung (mit "1" gekennzeichnet) liefert *ResCoOK.tief* die besten Ergebnisse unabhängig von der Wahl des Mittelwertes als *MLH.Mean* oder *MLH.Median*. Für die näher an einer vollständig zufälligen Verteilung liegenden Variante (mit "3" gekennzeichnet) ergibt sich die Abfolge von *CoOK.tief*, *OK*, *ResOK.MLH.Mean*, *ResOK.MLH.Median*, *ResCoOK.MLH.Median.tief* und *ResCoOK.MLH.Mean.tief*.
- *Zufallsbewegung*: Drei Stufen zunehmender Ballungsstärke $\alpha \in \{2, 3, 4\}$ werden unterschieden. Zu jedem dieser Fälle bietet *CoOK.tief* die beste Rekonstruktion vor dem Ordinary Kriging, *OK*. Die übrigen Methoden folgen in ihrer Güte keiner Systematik. Im Allgemeinen sind auch hier die stärker gehäufte Wolkenkonstrukte ungenauer zu rekonstruieren.



(a) Direkter Vergleich der Fehlerstatistik in Abhängigkeit von den Wolkenstrukturen ohne Beachtung der Krigingmethode. *Streifen*: schwarze Kreuze, *Klumpen*: rote Dreiecke, *Zufallsbewegung*: grüne Rauten.



(b) Fehlerstatistik der Schätzwerte verschiedener Krigingmethoden getrennt nach Wolkenstrukturen.

Abb. 10.10: Bestimmung der besten Krigingmethode anhand der Fehlerstatistik der Schätzwerte in Abhängigkeit von der Wolkenstruktur. Alle Wolkenmasken besitzen eine Überdeckung von 20% über Wasser (Abb. 10.1). In (b) wird die Fehlerstatistik farblich nach den verschiedenen Krigingverfahren aufgeschlüsselt: Ordinary Kriging (schwarz), Cokriging (rot), (Co-)Kriging der beiden Residuen (grün, blau, cyan, magenta). Die Zahlen in der Graphik der *Klumpen*-Struktur stehen für deren Anzahl ($1 \hat{=} N = 50$, $3 \hat{=} N = 150$), bzw. bei der *Zufallsbewegung* für die Ballungsstärke α . Quantifizierung der Güte der Krigingmethoden findet sich in Tabelle 10.4.

Methode	OK.log.chl	CoOK.tief	MLH.Mean		MLH.Median	
			ResOK	ResCoOK.tief	ResOK	ResCoOK.tief
Güte	0.3154	0.1268	1.0956	1.1332	1.2732	1.1848

Tab. 10.4: Beste Krigingmethode bei konstantem Bedeckungsgrad von 20% und unterschiedlichen Wolkenmustern. Die zugrundeliegende Fehlerstatistik ist in Abb. 10.10 dargestellt.

10.5 Diskussion: Beste Rekonstruktion mittels Kriging

Zwei Ziele sollen mit den in diesem Kapitel vorgestellten Untersuchungen erreicht werden:

1. Die Struktur einer Wolkenverteilung wird quantifiziert und es wird eine Beziehung zur zu erwartenden Güte der Rekonstruktion hergestellt.
2. Die beste Krigingmethode soll als Kombination von Wahl der Zielvariable, einer Trendkorrektur und eines Krigingverfahrens (mit Hilfsvariable) bestimmt werden.

Sowohl die Distanzklassenverteilung als Bestandteil der Zwei-Punkt-Statistik als auch die Lückenhaftigkeit als flächenhafte Statistik zu unterschiedlichen Skalenlängen kann die Flächenstruktur der fehlenden Datenpunkte ausreichend quantifizieren. Es können stark gehäufte von eher zufällig verteilten Mustern unterschieden werden. Aus den Maxima der Distanzklassenverteilung lassen sich mittlere Ausdehnungen der Strukturobjekte ablesen. Diese Strukturparameter können mit einfacheren Maßzahlen wie Bedeckungsgrad und Kompaktheit (siehe Peura und Iivarinen (1997) für eine Übersicht weiterer einfacher Zahlen zur Beschreibung von zweidimensionalen Objekten) ergänzt werden.

Die gewählten Beispiele zur Genauigkeit der Rekonstruktion können nicht belegen, dass ein Kriging der Residuen zu den besten Ergebnissen führt. Unter den Residuen werden dabei die um den zeitlich gewichteten Mittelwert korrigierten Datenpunkte des Bezugstages verstanden. Durch den systematischen Versatz der Chlorophyllkonzentration verschiedener Datentage stellt der zeitlich gewichtete Mittelwert nicht notwendigerweise die beste Annäherung an den Bezugstag dar. Die Trendkorrektur ist damit nicht so effektiv, wie sie theoretisch sein könnte. Der Vorzug eines zeitlich gewichteten Mittelwertes ist die Berücksichtigung der räumlichen Dynamik des Untersuchungsraumes. Dazu ist es notwendig, dass pro Pixel mehrere Datenpunkte an verschiedenen zeitlichen Stützstellen vorhanden sind. Durch die Diskontinuität im Niveau der Chlorophyllkonzentration kann ein Komposit, das an vielen Stellen nur aus einer zeitlichen Stützstelle gebildet wird, pixelweise sowohl deutlich zu hohe als auch zu niedrige Konzentrationen veranschlagen. Wird eine Trendkorrektur mit einem (gewichteten) Monatsmittels durchgeführt, wird in den Datensatz eines einzelnen Tages die zeitliche (systematische) Variabilität eingebracht.

Deshalb kann von Schätzwerten aus dem Ordinary Kriging der logarithmierten Chlorophyllkonzentration ohne Trendkorrektur im Moment angenommen werden, dass sie unabhängig von der Wolkenstruktur und dem Bedeckungsgrad zur genauesten Rekonstruktion führen.

Die Synthese aus dem besten Schätzwertverfahren und der Klassifizierung der Strukturen fehlender Daten ist noch nicht abgeschlossen. Durch die systematische Ungenauigkeit ist es derzeit nicht möglich, verlässliche Tests zur Genauigkeit der Rekonstruktion aus dem Kriging der Residuen durchzuführen.

11 Wann ist Kriging einsetzbar?

In der einschlägigen Literatur (z.B. Wackernagel (2003)) heißt es, dass Kriging nur interpolierend, das heißt an Punkten, die möglichst in allen Raumrichtungen von Datenpunkten umschlossen sind, angewendet werden sollte. Durch die Eigenheit der Satellitendaten, eine sehr dicht liegende, meist stark räumlich konzentrierte Menge von Datenpunkten zu sein, und die Verwendung des lokalen Kriging mit einer festen Anzahl von berücksichtigten Datenpunkten in der nächsten Nachbarschaft tritt häufig der Fall auf, dass die Schätzwerte nur auf Daten aus einem sehr beschränkten Raumbereich beruhen. Insbesondere wenn die zu rekonstruierenden Werte an Positionen liegen, die durch breite Wolkenbänder bzw. stark gehäufte Wolken mit relativ hohem Bedeckungsgrad zustande kommen oder tatsächlich am Rande der aufgezeichneten Daten liegen, kann man vielfach nicht mehr von Interpolation sprechen.

Doch bei welcher Distanz - und räumlichen Verteilung - zwischen Schätzposition und den Daten zieht man die Grenze zwischen Inter- und Extrapolation?

Wie lassen sich die Positionen identifizieren, an denen vermittelt durch die vorhandenen Datenpunkte und ihre räumliche Struktur verlässliche Schätzwerte erwartet werden können?

Wie groß darf der Bedeckungsgrad durch Wolken sein, damit alle Positionen der Datenlücken geschätzt werden können?

Mein Vorschlag zu Bestimmung der Verlässlichkeit von Schätzwerten beruht auf der stark von der Datenstruktur abhängigen Krigingvarianz. Diese auch "Krigingfehler" genannte Größe ist allerdings nicht zu verwechseln mit dem tatsächlichen Fehler. Ihre räumlichen Strukturen sind im Allgemeinen unabhängig voneinander (vgl. Abb. 11.2).

Der Wertebereich der Krigingvarianz ermittelt aus der Kreuzvalidierung kann als Maß dienen, ob eine Position einen glaubwürdigen Schätzwert erhalten kann oder nicht.

Um den Gedanken zu illustrieren, werde ich auf Vorheriges erneut eingehen.

Die Chlorophylldaten des Regionalprozessors vom 11. Juni 2006 werden mit der Wolkenmaske vom 2. Mai 2005 versehen, sodass ein Datentag mit etwa 87% Wolkenbedeckung - durchaus nicht untypisch für die Nordseeregion - simuliert wird.

Unter der Annahme, die Daten genügen der Forderung nach einer Zufallsfunktion mit Stationarität zweiter Ordnung (oder erfüllen die intrinsische Hypothese), lässt sich das Semi-Variogramm bestimmen. Die Kovarianz zwischen Datenpunktpaaren hängt dann nur von ihrer (vektoriellen) Distanz ab (Gl. 8.2) und lässt sich bei ausreichend isotropen Verhalten in einer omnidirektionalen Repräsentation zusammenfassen.

Typischerweise - und idealerweise - zeigt die Semi-Varianz monoton wachsendes Verhalten mit der Distanz (Abb. 11.1).

Die Kriging-(oder Fehler-)Varianz (Gl. 8.5) lässt sich als Summe der Semi-Varianz $\gamma(h_i)$ mit den Kriginggewichten λ_i und dem Lagrange-Multiplikator μ ausdrücken (vgl.

11 Wann ist Kriging einsetzbar?

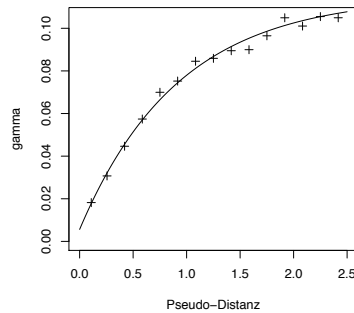


Abb. 11.1: Variogramm für die Chlorophylldaten des Regionalprozessors vom 11. Juni 2006 unter Ausschluß der Wolkenpunkte vom 2. Mai 2005.

Gl. 8.16). Unter der Nebenbedingung normierter Gewichte werden diese berechnet, indem die Fehlervarianz minimiert wird. Im Allgemeinen sind die Gewichte zur Bemessung des Einflusses eines Datenpunktes umso größer, je näher der jeweilige Datenpunkt an der Schätzwertposition liegt. Sie nehmen mit der Entfernung stark ab, wenn die Datenpunkte auf einer Linie definiert durch die Schätzposition und den nächstliegenden Datenpunkt liegen (Abschirmungseffekt, vgl. Beispiel in Wackernagel (2003)). Datenpunkte, die äquidistant auf einem Kreis mit der Schätzposition als Mittelpunkt liegen, bekämen alle das gleiche Gewicht zugewiesen.

Die Krigingvarianz spiegelt in erster Linie die räumliche Verteilung der Daten wieder. Abgesehen vom Lagrange-Multiplikator hängt sie von der Semi-Varianz und den Gewichten ab (Abb. 11.2). Im Beispiel werden im lokalen Kriging jeweils die 100 nächsten Nachbarn einbezogen. Die auffälligen Strukturen von geringeren Varianzen trotz weiterer Entfernung zu den Daten entstehen durch die Datenverteilung: wenn z.B. bei benachbarten Schätzwertpositionen ein Datenpunkt in einer zuvor unberücksichtigten Raumrichtung hinzukommt, ändert sich die Zusammensetzung der Gewichte sehr stark, da dieser einzelne Punkt starke Gewichtung erhalten wird, die Gewichte insgesamt aber normiert sind (Abb. 11.2a). Der tatsächliche Fehler spiegelt dagegen nicht die Datenverteilung sondern Bereiche hoher räumlicher Varianz wider, deren Rauigkeit das Kriging nicht erfassen kann (Abb. 11.2b).

Die Kreuzvalidierung beinhaltet folgendes Vorgehen: an der Position eines existierenden Datenpunktes wird der Schätzwert unter Ausschluss besagten Punktes gemäß der gewählten Methode berechnet. Der ermittelte Schätzwert wird mit dem tatsächlichen Wert verglichen.

Zwei extreme Möglichkeiten für die Wahl eines Kreuzvalidierungspunktes zeigt Abb. 11.3. Die mittlere Distanz ist im ersten Fall (Abb. 11.3a) sehr gering, die Datenpunkte liegen nah zusammen und hintereinander aufgereiht, sodass der Abschirmungseffekt für wenige große Gewichte in unmittelbarer Nachbarschaft der Schätzposition, in der Mehrzahl aber für kleine Gewichte nah an Null für die größeren Distanzen sorgen wird. In die Krigingvarianz fließen hier hauptsächlich kleine Semi-Varianzen ein. Die Krigingvarianz abgeleitet

11 Wann ist Kriging einsetzbar?

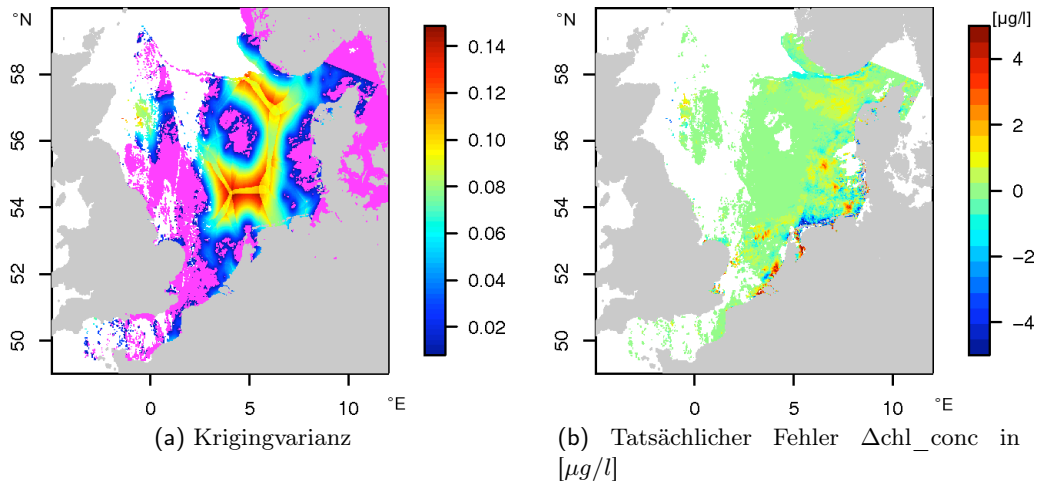


Abb. 11.2: Die Fehlergrößen Krigingvarianz und tatsächlicher Fehler der Schätzwerte für den Datensatz vom 11. Juni 2006 unter Ausschluß der Datenpunkte unter der Wolkenformation vom 2. Mai 2005. In (a) sind die verwendeten Datenpunkte in magenta markiert.

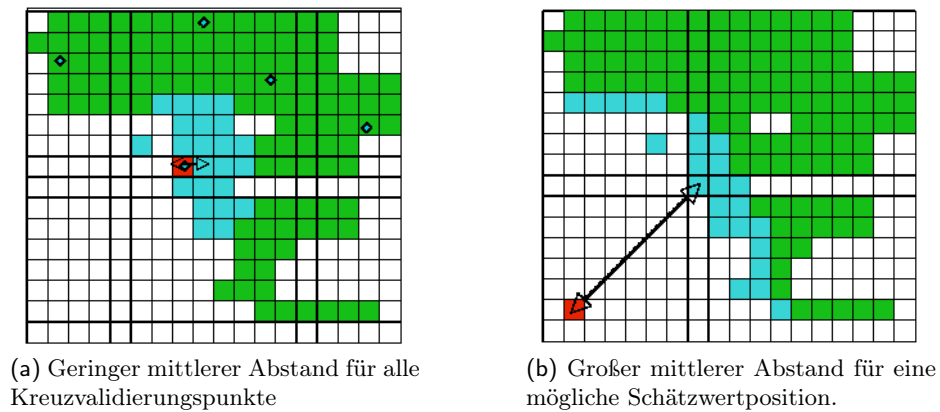


Abb. 11.3: Zwei Umgebungen nächster Nachbarn (blau) zu einer Schätzwertposition (rot) mit geringem Abstand bei der Kreuzvalidierung oder großem mittleren Abstand bei einer Datenlücke. Die Datenpunkte, an denen die Kreuzvalidierung vorgenommen wird, sind mit blauen Rauten markiert (a). Grüne Pixel repräsentieren vorhandene Datenpunkte, weiße stehen für Datenlücken.

11 Wann ist Kriging einsetzbar?

aus der Kreuzvalidierung ist somit minimal, da in diesen Fällen die geringsten Distanzen zwischen Datenpunkten und Schätzwertpunkten auftreten. Der Wertebereich der Krigingvarianz aus der Kreuzvalidierung beschreibt den Bereich minimaler Fehlervarianz und kann genutzt werden, um den Bereich der Interpolation zu charakterisieren.

Sollte die Schätzposition nicht in der Nähe der Daten liegen (Abb. 11.3b) wird die mittlere Distanz recht groß, die Daten sind nicht mehr (so stark) aufgereiht, sodass mehr oder minder gleich große Gewichte die Folge sind (kein Abschirmungseffekt). Die Krigingvarianz wird vergleichbar mit der Semi-Varianz der mittleren Distanz sein.

In dem alle Voraussagen mit Krigingvarianzen außerhalb ihres Wertebereiches in der Kreuzvalidierung ausgeschlossen werden, lassen sich diejenigen Positionen finden, die von Daten umgeben oder nah an Datenpunkten liegen, sodass die Voraussagemethode einer Interpolation nahe kommt.

Die Krigingvarianz, wie sie hier als Auswahlkriterium genutzt wird, hängt im hier praktizierten lokalen Kriging von einer ganzen Reihe von Faktoren und ihrem Zusammenspiel ab. Die hier zusammengestellten Überlegungen stellen lediglich die Basis dar, um daraus ein belastbares Kriterium zu entwerfen.

Es bleibt zu untersuchen, ob sehr große Datenlücken mit Hilfe der hier angestellten Überlegung sukzessive geschätzt werden können, in dem die Schätzwerte aus dem Intervall der in der Kreuzvalidierung bestimmten Krigingvarianz in nachfolgenden Schritten mit zu den Datenpunkten hinzukommen. Dieser Vorschlag entspräche einem groberen "Sequential Kriging", bei dem jeder einzelne neue Schätzwert sofort für die nächste Schätzposition als Datenpunkt mit zur Verfügung steht.

12 Iteratives

Dateninterpolationsverfahren - DINEOF

Nach der eingehenden Betrachtung und Bewertung der Krigingverfahren soll eine weitere Methode zur Schätzwertbestimmung zum Abschluß dieses Teiles vorgestellt werden.

Mithilfe der Hauptkomponenten-Analyse findet das (zeitliche) Dateninterpolationsverfahren DINEOF (*Data Interpolating Empirical Orthogonal Functions*) auf iterativem Wege Schätzwerte zu fehlenden Datenpunkten (Alvera-Azcarate u. a. (2005); Beckers und Rixen (2003); Beckers u. a. (2006)). Aufgrund seiner relativen Schlichtheit und Robustheit hat es sich zurzeit als Standardmethode durchgesetzt.

Nach einer kurzen methodischen Einführung werden zwei Anwendungsbeispiele im direkten Vergleich mit Schätzwerten aus Krigingverfahren vorgeführt: zum einen für Daten vom 12. März 2007 in der Deutschen Bucht (Gebiet 4, vgl. Abschnitt 9.5.8) und zum zweiten für die Chlorophyllverteilung vom 10. Mai 2005 auf dem Gesamtgebiet der Nordsee (vgl. Abschnitt 10.4)

12.1 Einführung zur Methode DINEOF

Aus der Menge aller Datentage können diejenigen in der Dateninterpolation berücksichtigt werden, die bezogen auf das gewählte Gitter und auf die darin mögliche Anzahl von Wasserpixel mehr als 5% Datenüberdeckung besitzen; dieses Kriterium sei zu Zeitpunkten $t = \{1, \dots, T\}$ erfüllt. Eine gültige Gitterposition ist definiert als diejenige, die mindestens zu drei Zeitpunkten mit gültigen Datenpunkten besetzt ist; ihre Anzahl sei M .

Daraus wird die Matrix $X \in \mathbb{R}^{M \times T}$ zusammengestellt, mit M Zeilen gültiger Gitterpixel und T Spalten von Beobachtungszeitpunkten. Es gilt im Allgemeinen, dass die räumliche Dimension wesentlich größer ist als die zeitliche, $M \gg T$. Dabei sind die Zeitpunkte weder notwendig äquidistant, noch berücksichtigt die Abfolge der Gitterpunkte ihre tatsächliche räumliche Distanz und damit Beziehung zueinander. Jede Zeile repräsentiert lediglich einen festen Gitterpunkt, jede Spalte einen festen Zeitpunkt. Die Werte der Matrix X werden um den globalen Mittelwert - also sowohl über die räumliche als auch über die zeitliche Dimension gemittelt - reduziert. Für Chlorophyllkonzentrationen bezieht sich der Mittelwert sinnvollerweise auf deren Logarithmen, aus denen auch die Einträge der Matrix X bestehen.

Alle fehlenden Daten werden auf den Startwert $x_{start} = 0$ gesetzt und entsprechen damit dem globalen Mittelwert.

Um die optimale Anzahl N an Hauptkomponenten für die Rekonstruktion des speziellen Datensatzes zu bestimmen, wird eine Kreuzvalidierung, d.h. der Ausschluß vorhandener Datenpunkte und Vergleich der Schätzwerte zum Originalwert, abgekürzt als CV , durchgeführt. Dazu werden aus den gültigen Datenpunkten mit mehr als 10 zeitlichen Beobachtungen eine Anzahl $n_{CV} = \text{Minimum}(0.01 \cdot M \cdot T + 40, 0.03 \cdot M \cdot T)$ ausgewählt und ihre Funktionswerte an den Stellen x_{CV} vorübergehend auf Null gesetzt, $X[x_{CV}] = 0 \rightarrow X_{CV}$.

Nun beginnt der iterative Prozess: Die Hauptkomponenten X_{PCA} der Matrix X_{CV} werden bestimmt. An Stellen fehlender Werte $X_{CV}[x]$, seien sie durch die Kreuzvalidierung bedingt oder tatsächlich fehlend, wird der Wert der ersten Hauptkomponente eingefügt, $X_{CV}[x] = X_{PCA,1}[x]$. Der Fehler (definiert als $\text{error}_i = \sqrt{\sum_x (X_{CV}[x, i] - X_{CV}[x, i - 1])^2}$ für Iteration i) zwischen ursprünglichem fehlenden Wert und Schätzwert wird notiert. Basierend auf der Matrix X_{CV} , die nun auch erste Schätzwerte enthält, wird die Hauptkomponentenanalyse erneut ausgeführt und die alten Schätzwerte durch die neuen ersetzt. Mit jeder Iteration werden sich die Schätzwerte zum vorhergehenden Durchgang weniger verändern. Ist die Änderung kleiner als δ , $\text{error}_i - \text{error}_{i-1} < \delta$, so wird die Iteration mit einer Hauptkomponente als konvergent betrachtet und die Berechnung abgebrochen.

Nach der nächsten Berechnung der Hauptkomponentenanalyse werden nun zwei Hauptkomponenten zur Rekonstruktion verwendet. Der gesamte Prozess wird bis zur Konvergenz der Fehleränderung bei k_{max} Hauptkomponenten fortgesetzt.

Die ausgewählten CV-Punkte ermöglichen jeweils bei Konvergenz die Bestimmung des tatsächlichen Fehlers in Abhängigkeit von der Komponentenanzahl. Dieser zeigt (im Falle gutes Verhaltens des Datensatzes) ein Minimum bei N Komponenten (Abb. 12.4).

Für die schlussendliche Rekonstruktion werden die CV-Daten wieder hinzugefügt. Die bereits aus dem Vorversuch geschätzten Werte an den Stellen fehlender Datenpunkte verbleiben als verbesserte Startwerte in X . Wiederum wird die Hauptkomponentenanalyse durchgeführt, die Schätzwerte werden jeweils aus den optimalen N Hauptkomponenten zusammengesetzt. Bei Konvergenz der Schätzwertfolge wird die Berechnung abgebrochen.

Auf diese Weise erhält man eine gleichzeitige Rekonstruktion aller fehlender Daten in Raum und Zeit, ohne jedoch auf räumliche oder zeitliche Abhängigkeiten der Daten wirklich Rücksicht zu nehmen.

Anmerkungen zur Umsetzung

- Die Autoren Beckers u. a. (2006) bemerken, dass erst ab einer Datenmenge, deren zeitliche Überdeckung an jedem Gitterpunkt größer als 30 Tage ist, eine gute Rekonstruktion erfolgen kann. Um einen Methodenvergleich mit dem Kriging zu verwirklichen, werden für den DINEOF-Datensatz statt der Daten eines Monats wesentlich längere Zeiträume herangezogen, um diese Bedingung annähernd zu erfüllen.
- Bei einem sehr lückenhaften Datensatz mit vielen Punkten von geringerer zeitlicher Überdeckung als der empfohlenen stellt nur das Kriterium, die CV-Punkte aus denjenigen auszuwählen, die am häufigsten in der Zeit gesehen werden, sicher, dass

der tatsächliche Fehler in Abhängigkeit von der Komponentenzahl ein ausgeprägtes Minimum besitzt. Im Beispiel der Nordsee werden z.B. nur Gitterpunkte zugelassen, die mehr als 10 zeitliche Datenpunkte besitzen.

- Die Abhängigkeit der tatsächlichen Fehler vom Konvergenzkriterium δ ist unvorhersagbar.
- Die Methode liefert generell Schätzwerte aus einem unbeschränkten Wertebereich, was im Extremfall zu Rekonstruktionen führen kann, die die gesamte globale Varianz in einzelnen wenigen Punkten konzentrieren, sie aber kaum über die gesamte Fläche zu verteilen versuchen. Es ist daher durchaus sinnvoll, den Wertebereich der Schätzwerte zu beschränken und z.B. nur Schätzwerte innerhalb des Wertebereiches des Datensatzes X zuzulassen. Schätzwerte, die größer/kleiner sind, werden dementsprechend auf die obere/untere Schranke gesetzt.
- Die eingehenden Zeitpunkte müssen nicht äquidistant sein. Die räumliche wie zeitliche Struktur - wie im Sinne des Grundgedanken jeder Interpolation, dass benachbarte Punkte einander ähnlich sein sollten - wird nicht verwendet.

12.2 Beispiel DINEOF in der Deutschen Bucht

Im direkten Vergleich zum Versuchsansatz mit künstlicher Wolkenmaske in der Deutschen Bucht (Gebiet 4, vgl. Abschnitt 9.5.8) wird für die DINEOF-Methode ein Datensatz aus der ersten Hälfte des Jahres 2007 zusammengestellt, der die räumliche Dimension $M = 4984$ und die zeitliche Dimension $T = 66$ besitzt. Alle Gitterpixel haben minimal eine zeitliche Überdeckung von 25 Tagen, im Mittel 34.5 Tage, sodass sie die Empfehlung von mindestens 30 zeitlichen Stützstellen pro räumliches Pixel recht gut erfüllen. Wie im Test der Krigingmethoden wird aus den Daten vom 12. März 2007 die Wolkenmaske ausgeschlossen. Die 3329 CV-Punkte werden danach willkürlich aus der Gesamtmenge gültiger Datenpunkte aus X ausgewählt.

Die Rekonstruktion erfolgt immer unter Beschränkung des Wertebereiches auf denjenigen der Matrix X . Als Konvergenzkriterium werden $\delta = 0.5$, $\delta = 0.1$, $\delta = 0.01$ und $\delta = 0.001$ gesetzt, wobei für $\delta < 0.1$ die Festlegung der Hauptkomponentenanzahl weiterhin mit $\delta = 0.1$ erfolgt und nur die abschließende Berechnung der Schätzwerte tatsächlich die kleinere Schranke nutzt. Zusätzlich wird in einem Versuch mit $\delta = 0.01$ die Anzahl von CV-Punkten auf $N_{CV} = 5000$ erhöht, um das Verfahren zumindest im Ansatz auf eine Abhängigkeit von der Größe der CV-Stichprobe zu testen.

Die Prozedur wird für jeden Satz (δ, N_{CV}) mehrfach wiederholt, sodass Einflüsse aus der zufälligen Auswahl der CV-Punkte sichtbar werden können.

Zum Vergleich der Rekonstruktionen wird ihre Güte berechnet (vgl. Gl. 9.4 in Abschnitt 9.5.1); normiert werden die Fehlergrößen bezogen auf alle Realisationen, für alle δ und alle N_{CV} .

Legt man die beste Anzahl von Hauptkomponenten über die Güte der Rekonstruktion fest, so hängt sie hauptsächlich von der Auswahl der CV-Punkte ab (vgl. Abb. 12.1) und

nur in zweiter Linie von der Wahl des Konvergenzkriteriums δ . Je größer die Schranke gewählt wird, je weniger Iterationen also durchgeführt werden, umso weiter scheint die mögliche Anzahl von Hauptkomponenten streuen zu können. Die besten Rekonstruktionen werden für 12 (bzw. 13) Komponenten erreicht, relativ unabhängig von der Größe des Konvergenzkriteriums.

Eine eindeutige Aussage zum Einfluss des Konvergenzkriteriums auf die Güte der Rekonstruktion lässt sich so nicht treffen. Der Versuch müsste für jeweils eine feste Auswahl von CV-Punkten mit allen δ wiederholt werden, sodass die Fehlergrößen bezogen auf eine feste Konstellation von CV-Punkten normiert werden könnten.

Im Vergleich zu einem Satz von Krigingschätzwerten (d.h. verschiedene Krigingmethoden bei Variogrammen basierend auf der gleichen Datenpunktauswahl) besitzen die Rekonstruktionen nach DINEOF den kleinsten absoluten Fehlermedian, während bei den Krigingmethoden (zumindest die ResOK, vgl. auch Abb. 9.24) der mittlere absolute Fehler deutlich geringer ist (Abb. 12.2). Dieses Ergebnis spiegelt die methodischen Unterschiede gut wieder: während im Kriging die Varianz lokal minimiert wird (minimale mittlere absolute Fehler), geschieht dies beim DINEOF bezogen auf die globale Varianz (minimale Über- bzw. Unterschätzung).

Die Anzahl der DINEOF-Versuche ist nicht ausreichend, um zu entscheiden, ob sich tatsächlich die Tendenz abzeichnet, dass mit kleinerem δ der mittlere absolute Fehler sinkt, die Über- bzw. Unterschätzung aber gleichzeitig zunimmt (Abb. 12.2b).

Die starke lokale Strukturiertheit des Originals kann DINEOF hier jedenfalls schlechter rekonstruieren als ein Krigingverfahren (Abb. 12.3).

12.3 Beispiel DINEOF im Gebiet der Nordsee

Für den Datensatz vom 10. Mai 2006 und die Wolkenmaske *Klumpen* (20% Bedeckung, Anzahl=150, Abb. 10.1c) wird der Versuch der Rekonstruktion aus insgesamt $T = 29$ Datentagen (innerhalb eines Zeitraumes zwischen dem 10. April und 10. Juni 2006) bei $M = 229962$ gültigen Datenpunkten im Nordseegebiet (vgl. Abb. 2.3) unternommen. In diesem Zeitraum treten (vgl. Abschnitt 10.4.1) deutlich sichtbare Versätze in der Chlorophyllkonzentration auf, die durch den langen Zeitraum von zwei Monaten aber nicht vereinfacht korrigiert werden können. Sowohl der biologische als auch der Transportanteil des Signals müsste hier einbezogen werden. Statt dergleichen zu versuchen, sei nur auf diese Fehlerquelle verwiesen und die Daten in ihrem derzeitigen Zustand verwendet.

Die mittlere zeitliche Überdeckung pro Pixel beträgt 10 Datentage (minimal 3, maximal 24), sodass die Voraussetzung von einer Überdeckung > 30 Tage nicht im Ansatz erfüllt ist. Die CV-Auswahl wird daher nicht willkürlich im Datensatz durchgeführt, sondern beginnt bei den Pixel mit höchster zeitlicher Überdeckung, bei denen jeweils ein Punkt in der Zeit ausgewählt wird, und setzt sich mit Pixel abnehmender zeitlicher Überdeckung fort, bis N_{CV} erreicht ist.

Die Kreuzvalidierung nutzt $N_{CV} = 65619$ Datenpunkte und bestimmt die Anzahl notwendiger Hauptkomponenten mit $N = 5$ (Abb. 12.4).

Für die *Klumpen*-Formation (Anzahl 150, 20% Überdeckung, vgl. Abb. 10.1c) wird mit

12 Iteratives Dateninterpolationsverfahren - DINEOF

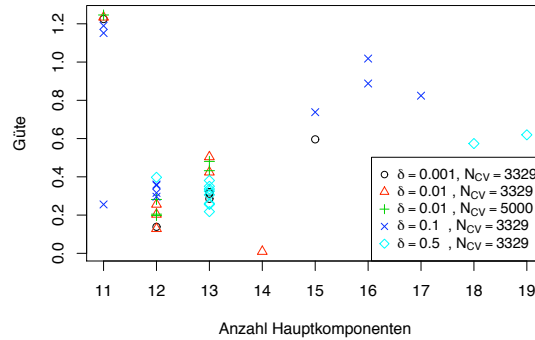
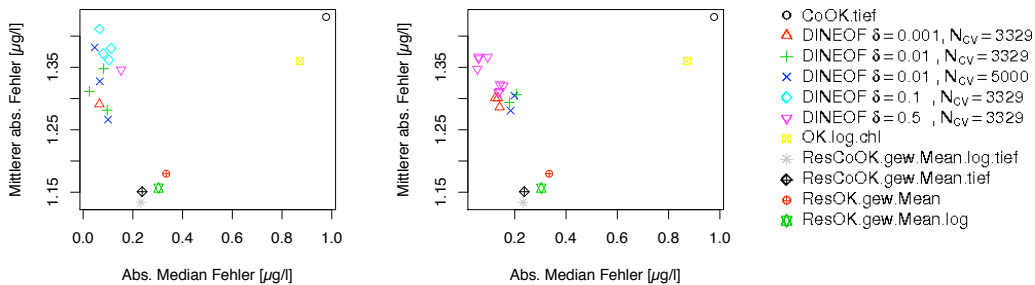


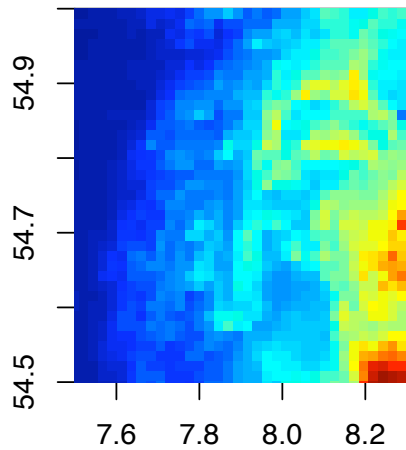
Abb. 12.1: Variabilität der Anzahl der für die Dateninterpolation zu benutzenden Hauptkomponenten für verschiedene Konvergenzkriterien δ und mehrere Realisationen zur Wahl der CV-Punkte. Die Güte ist definiert nach Gl. 9.4, die Normierung der globalen und lokalen Fehler bezieht sich jeweils auf alle dargestellten Versuche. Als Datensatz dient die Chlorophyllkonzentration vom 12. März 2007 in der Deutschen Bucht, die Schätzwerte werden an den Positionen einer Wolkenmaske (vgl. Abb. 9.24) erhoben.



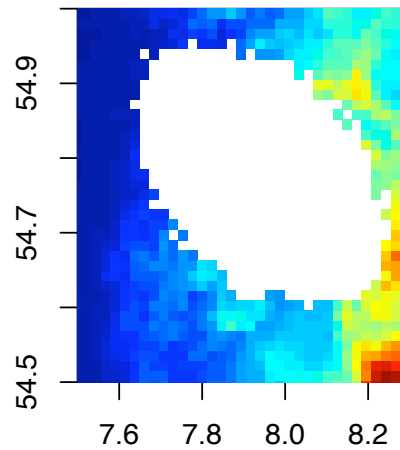
(a) DINEOF aus 12 Hauptkomponenten

(b) DINEOF aus 13 Hauptkomponenten

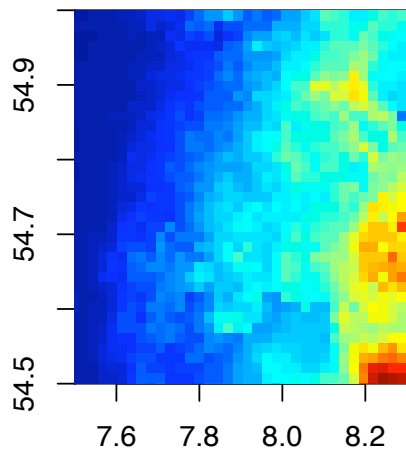
Abb. 12.2: Fehlerstatistik der Schätzwerte aus den Dateninterpolationsversuchen für verschiedene Konvergenzkriterien δ und Größen der Kreuzvalidierungsstichprobe im Vergleich zu der Fehlerstatistik der Rekonstruktionen unterschiedlicher Krigingmethoden. Als Datensatz dient die Chlorophyllkonzentration vom 12. März 2007 in der Deutschen Bucht, die Schätzwerte werden an den Positionen einer Wolkenmaske (vgl. Abb. 9.24) erhoben.



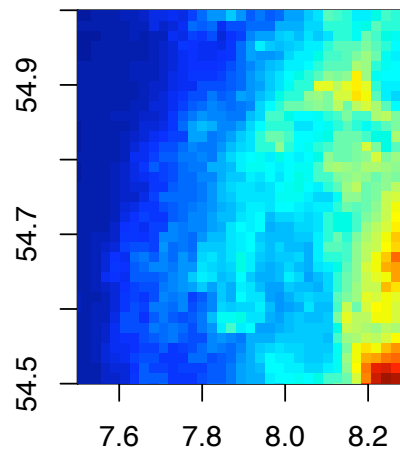
(a) Chlorophyllverteilung vom 12. März 2007



(b) Chlorophyllverteilung mit Wolkenmaske



(c) DINEOF mit Konvergenzkriterium $\delta = 0.001$ bei einer benutzten Anzahl von 12 Hauptkomponenten



(d) Cokriging der Residuen mit der Hilfsvariable Wassertiefe nach einer Trendkorrektur mit dem gewichteten Monatsmittel logarithmierter Konzentrationen (*ResCoOK gew.Mean.log tief*)

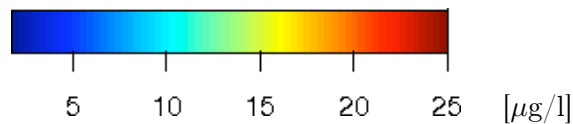


Abb. 12.3: Beispiel der Rekonstruktionsergebnisse für die Chlorophyllkonzentration in $[\mu\text{g}/\text{l}]$ gemäß der Dateninterpolation, DINEOF, und dem Cokriging der Residuen. Die Fehlerstatistik für diese beiden Verfahren ist in 12.2a dokumentiert. Gezeigt ist ein Ausschnitt aus dem als "Deutsche Bucht" bezeichneten Gebietes 4 (vgl. Abb. 9.24), die als Ganzes Grundlage für die Rekonstruktionen ist.

drei Kombinationen von Konvergenzparametern δ und Beschränkung des Wertebereiches gearbeitet:

- Im CV-Prozess wird als Konvergenzkriterium $\delta = 0.5$ genutzt.
- Im abschließenden Schätzprozess wird
 - $\delta = 10^{-4}$ mit uneingeschränktem Wertebereich oder
 - $\delta = 0.5$ mit eingeschränktem Wertebereich, sodass derjenige der Schätzwerte dem des Datensatzes entspricht, oder
 - $\delta = 0.01$ mit eingeschränktem Wertebereich verwendet.

Diese Schwellwerte stehen im starken Kontrast zu denen in der Software-Variante von Alvera-Azcarate u. a. (2005), bei denen Konvergenz erst mit $\delta = 10^{-9}$ erreicht wird. Die Implementierung ihres Algorithmus in R lässt eine exakte Reproduktion ihrer Methode nicht zu. Gleichzeitig bietet ihre momentan verfügbare Software nicht die Option, für die CV-Auswahl nur eine bestimmte Menge an Punkten zu definieren, sodass sich das Fehlerverhalten in Abhängigkeit der Komponentenanzahl bei dem stark lückenhaften und in der Zeit unterbestimmten System nicht als glatter Verlauf mit Minimum darstellt. Die Bestimmung der optimalen Anzahl von Hauptkomponenten muss daher scheitern - und daran auch die Rekonstruktion. Somit kommt hier nur die Reproduktion des Algorithmus als R-Skript zur Anwendung.

Verglichen werden die DINEOF-Rekonstruktionen mit Krigingmethoden, bei denen nur eine einzige Untermenge von Datenpunkten zur Berechnung der methodenabhängigen Variogramme ausgewählt wird.

Die geringsten Fehler (lokal und global) lassen sich durch Cokriging der logarithmierten Chlorophyllkonzentration mit der Hilfsvariable Wassertiefe, *CoOK.tief*, und dem normalen Kriging, *OK.log.chl*, erzielen (Abb. 12.5). Bezüglich der lokalen Genauigkeit (repräsentiert durch den mittleren absoluten Fehler) schneiden die DINEOF-Rekonstruktionen jeweils wie zu erwarten am schlechtesten ab. Eine Systematik im Verhalten der globalen Fehler als Funktion des Konvergenzkriteriums δ ist nicht zu erkennen.

Die Schätzwerte für die *Klumpen*-Wolkenmaske unterscheiden sich auch optisch stark für die (hier beste) Krigingmethode *CoOK.tief* und das DINEOF-Verfahren (Abb. 12.7). Während das Krigingverfahren rein visuell als glatt wahrzunehmende räumliche Verteilungen produziert, ist in der DINEOF-Rekonstruktion deutlich die Wolkenmaske als Bereich zur erkennen, der nicht glatt genug erscheint. Von allgemeiner Aussagekraft ist dieses Beispiel aber nicht, da die zeitliche Überdeckung insgesamt zu niedrig ist und nicht die empfohlene Mindestzahl von Zeitpunkten trifft. Zusätzlich wird der systematische Fehler mit den Versätzen von Tag zu Tag für eine deutliche Überschätzung der globalen Varianz im Datensatz führen.

Gibt es eine Abhängigkeit der Rekonstruktion im DINEOF von der Struktur der Wolkenmaske? Dazu wird ein weiteres Experiment für diesen Datensatz mit der *Streifen*-Formation mit 20% Bedeckung durchgeführt (Abb. 10.1a). Die Cokrigingmethoden fallen mit extrem großen lokalen Fehlern (mittlerer abs. Fehler = 6-10) aus der Untersuchung heraus. Während auch die übrigen Krigingmethoden jeweils geringere lokale Genauigkeit

12 Iteratives Dateninterpolationsverfahren - DINEOF

beim Übergang zur *Streifen*-Struktur erhalten, ist diese für das DINEOF vergleichsweise stabil (Abb. 12.6).

12 Iteratives Dateninterpolationsverfahren - DINEOF

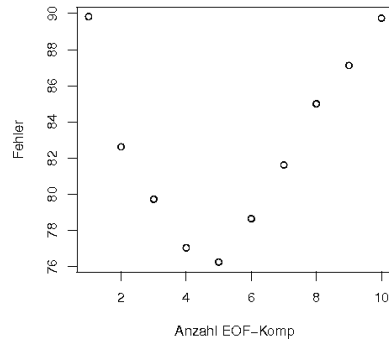


Abb. 12.4: Beste Anzahl an Hauptkomponenten bestimmt mittels Kreuzvalidierung. Mit “Fehler” wird hier die Summe der absoluten Abweichung zwischen Rekonstruktion und tatsächlichem Wert bezeichnet. Die Schätzwerte werden an den Positionen der Wolkenmaske *Klumpen*, $N = 150$, mit einem Bedeckungsgrad von 20% (vgl. Abb. 10.1c) berechnet. Als Datensatz der Rekonstruktion dient das Chlorophyllprodukt vom 10. Mai 2006.

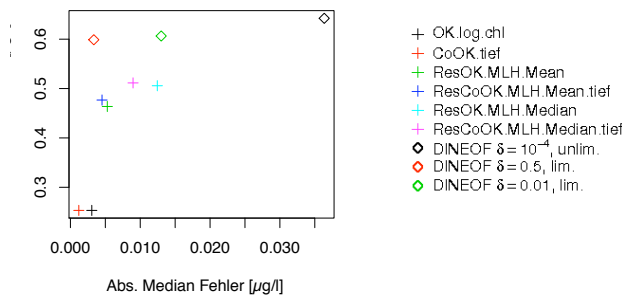
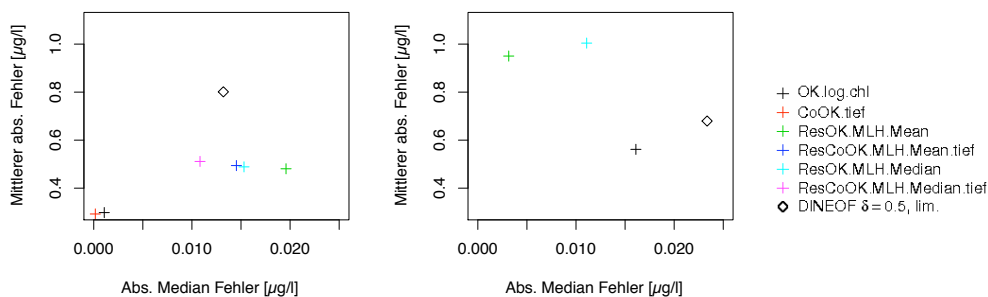


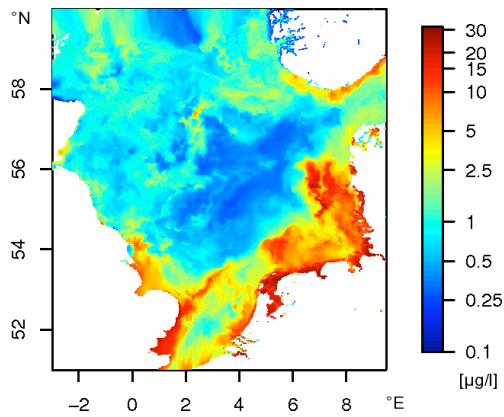
Abb. 12.5: Abhängigkeit der Fehlerstatistik der Dateninterpolation DINEOF vom Konvergenzkriterium δ in der Rekonstruktion für den Chlorophylldatensatz der Nordsee vom 10. Mai 2006. Als Wolkenformation dienen die *Klumpen*, $N = 150$, mit 20% Bedeckung (vgl. Abb. 10.1c).



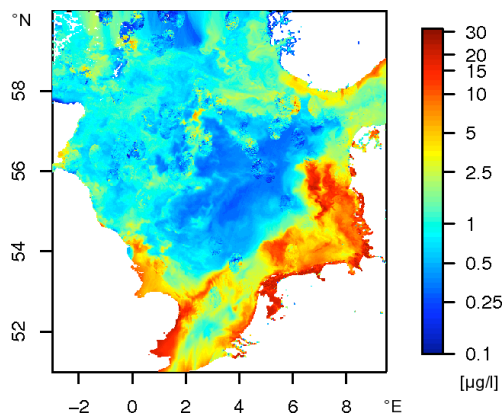
(a) Wolkenmaske *Klumpen*, $N = 150$, 20% (vgl. Abb. 10.1c)

(b) Wolkenmaske *Streifen*, 20% (vgl. Abb. 10.1a)

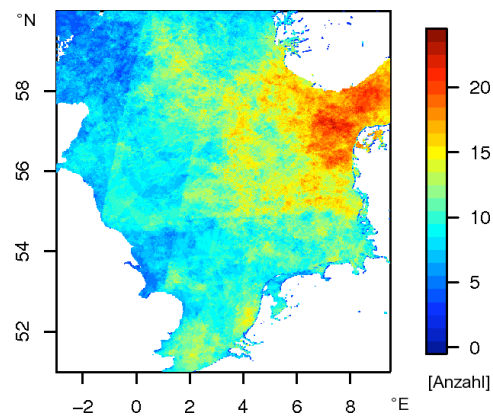
Abb. 12.6: Fehlerstatistik der Rekonstruktion für verschiedene Krigingverfahren und die Dateninterpolation in Abhängigkeit von der Struktur der Wolkenmaske für den Chlorophylldatensatz vom 10. Mai 2006.



(a) Cokriging der log. Chlorophyllkonzentration mit der Hilfsvariable Tiefe



(b) Dateninterpolation DINEOF mit Konvergenzkriterium $\delta = 0.5$ und beschränkte Wertebereich der Rekonstruktion



(c) Zeitliche Überdeckung des Datensatzes zur Dateninterpolation DINEOF in Tagen

Abb. 12.7: Karten der Chlorophyllkonzentration zusammengesetzt aus Schätzwerten und dem Originalbild vom 10. Mai 2006 in $[\mu\text{g/l}]$ gemäß dem Cokrigingverfahren (a) bzw. der Dateninterpolation (b). Als Wolkenmaske dient die *Klumpen*-Formation, $N = 150$, mit 20% Bedeckung (vgl. Abb. 10.1c). Die zu erwartende lokale Genauigkeit im DINEOF wird direkt mit der zeitlichen Überdeckung (c) korrelieren.

Zusammenfassung

Zwei grundsätzlich verschiedene Schätzwertverfahren sind in den vorherigen Kapiteln vorgestellt worden: die Krigingmethode in zahlreichen Varianten und das Dateninterpolationsverfahren basierend auf der Hauptkomponentenanalyse. Da die Dateninterpolation auch auf die Qualität der zeitlichen Information angewiesen ist, kann sie derzeit nicht mit der Genauigkeit der Krigingverfahren konkurrieren.

Bei allen methodischen Unterschieden ist den Schätzwerten gemeinsam, daß sie letztlich keine neue Information kreieren, sondern aus geschickt gewichteten vorhandenen Datenpunkten auf die zu schätzenden schließen.

Die Ergebnisse - wie zuvor in Kapitel 9 und Kapitel 10- zeigen, dass *eine* beste Schätzwertmethode nicht eindeutig empfohlen werden kann. Die Wahl der besten Schätzwertmethode ist vielmehr komplex von den Faktoren der Daten- und Wolkenstruktur, sowie der zu erwartenden Variabilität an der Schätzposition abhängig.

Es wird vorgeschlagen, die Nordsee bzw. jedes Pixel danach zu charakterisieren, ob es eher im Mittel in einem hochstrukturierten oder homogenen Bereich liegt. Entsprechend dieser Erfahrungswerte können jeweils regional oder sogar punktweise ein angemessenes Krigingverfahren ausgesucht werden.

Dabei sollten nicht nur räumliche sondern auch zeitliche Vorkenntnisse einfließen. Während viele Räume im Sommer sehr homogen sind, können sie zu anderen Jahreszeiten starke Strukturen zeigen. Ein Ansatz zur quantitativen Beschreibung der Strukturiertheit der Nordseeregion und einzelner Gebiete wird im Folgenden mit Autokorrelationsmethoden vorgestellt. Einen ersten Überblick gibt auch bereits die Varianz der Jahreszeitenmittel (vgl. Abb. 6.3), die bisher allerdings nicht systematisch in Überlegungen zu einer regional-besten Schätzwertmethode einbezogen worden ist. Diese statistische Größe ließe sich als Varianz des klimatologischen Monatsmittels noch deutlich verfeinern. In welcher Reihenfolge die Verfahren sinnvollerweise anzuwenden sind bzw. ob bereits berechnete Schätzwerte eines Verfahrens in nachfolgenden als vorhandene Datenpunkte gehandhabt werden sollten und welchen Einfluß eine solches Vorgehen - aus dem 'Sequential Kriging' (Vargas-Guzman und Yeh (1999)) nicht unbekannt - auf die Güte der Rekonstruktion insgesamt hätte, bleibt für weitere Untersuchungen offen.

Teil IV

Räumliche und zeitliche Autokorrelation

Einleitung

Bis jetzt haben sich die Untersuchungen im Rahmen dieser Arbeit auf die Möglichkeiten der Vervollständigung der Satellitendatensätze (Kapitel 9 bis Kapitel 12), die Charakterisierung des systematischen Fehlers im Chlorophyllprodukt (Kapitel 5) und einige Versuche qualitativer Interpretation räumlicher Strukturen (Kapitel 6) konzentriert.

In diesem abschließenden Teil soll die quantitative Analyse der Chlorophyllverteilungen im Vordergrund stehen.

Etablierte Methoden verwenden die Autokorrelation, um räumliche Zusammenhänge (Shevyrnogov und Vysotskaya (2006)) oder die zeitliche Dynamik (Denman und Abbott (1988, 1994)) der Strukturen zu beschreiben. Aufgrund des systematischen Fehlers wird auf das Maß der skalenabhängigen Korrelationszeit bzw. der Kohärenz verzichtet, da mit starken Artefakten durch die zeitliche Komponente der Analyse zu rechnen ist.

Nach einer methodischen Einführung in die räumliche (Abschnitt 13.1) bzw. zeitliche Autokorrelation (Abschnitt 13.2) wird der Chlorophylldatensatz mit diesen Verfahren analysiert. Der Einfluss von Schätzwerten auf die räumliche Analyse (Abschnitt 14.1) wird kurz charakterisiert, bevor diese durchgeführt wird.

In der zeitlichen Untersuchung werden zwei Datensätze unterschiedlicher Auflösung verwendet. Dabei ist die Analyse der Zeitreihe der Monatsmittel (Abschnitt 14.2) eher auf die Entdeckung jährlich wiederkehrender Zyklen ausgerichtet, während die Untersuchung an der Zeitreihe mit höchster, d.h. annähernd täglicher, Auflösung (Abschnitt 14.3) den Schwerpunkt auf die individuelle Dynamik der Chlorophyllverteilung einzelner Regionen und Jahreszeiten legt.

13 Theoretische Betrachtungen zur Autokorrelation

13.1 Räumliche Autokorrelation

Die räumliche Autokorrelation $R(\vec{r})$ des Feldes $T(\vec{r})$ ist definiert als dessen Faltung

$$R(\vec{r}) = T(\vec{r}) \star T(\vec{r}) = \int T(\vec{r}') T(\vec{r}' - \vec{r}) d\vec{r}' \quad (13.1)$$

Gemäß der Fouriertransformation lässt sich die Transformierte des Faltungsintegrals als eine Multiplikation des Spektrums im Fourierraum behandeln:

$$R_{xx}(\vec{r}) = \mathcal{F}\{T(\vec{r}) \star T(\vec{r})\} = \mathcal{F}\{T(\vec{r})\} \cdot \mathcal{F}\{T(\vec{r})\} \quad (13.2)$$

$R_{xx}(\vec{r})$ heißt auch Autospektrum des Feldes T .

$$R(\vec{r}) = \mathcal{F}^{-1}\left\{\left(\mathcal{F}\{T(\vec{r})\}\right)^2\right\} = \mathcal{F}^{-1}\{R_{xx}\} \quad (13.3)$$

Im diskreten Fall werden die räumlichen Daten $T(r_{i,j}) \in \mathbb{R}^{N \times M}$ auf einem regelmäßigen Gitter mit annähernd quadratischen Pixel $\Delta x \approx \Delta y$ in einem rechteckigen Gebiet mit $i = \{0, \dots, N-1\}$ und $j = \{0, \dots, M-1\}$ betrachtet. Bei der Berechnung des Autospektrums werden alle Einträge in der Matrix der Fourierkoeffizienten zu T quadriert.

Um die Auflösung im Wellenzahlbereich zu erhöhen (Denman und Abbott (1988, 1994)), werden die Daten $T(r_{i,j})$ in eine größere Umgebung $U \in \mathbb{R}^{L \times L}$ mittig eingeschrieben, deren Dimension $L \times L$, mit $L > N, M$, den Ansprüchen der FFT nachkommt und L einer Potenz von 2 entspricht. Die Daten $T(r_{i,j})$ werden zu $\hat{T}(r_{i,j})$ zentriert:

$$\bar{T} = \frac{1}{MN} \sum_{i=1}^N \sum_{j=1}^M T(r_{i,j}) \quad (13.4)$$

$$\hat{T}(r_{i,j}) = T(r_{i,j}) - \bar{T} \quad (13.5)$$

Im Falle der Verteilung der Chlorophyllkonzentration setze man T sinnvollerweise als logarithmierte Konzentration ein, um eine annähernde Normalverteilung und damit die Bedeutungshaftigkeit des arithmetischen Mittelwertes zu gewährleisten. Die Werte innerhalb der Umgebung werden alle auf Null gesetzt, entsprechend dem Mittelwert der zentrierten Daten. Die Ränder von \hat{T} werden mit Cosinusflanken (Tukey-Fenster, Harris (1978)) geglättet, um ein starkes Überschwingen durch die Unstetigkeiten zwischen Umgebung und Feld zu verringern.

$$U = \begin{pmatrix} 0 & & \dots & 0 \\ & \ddots & & \\ \vdots & & \hat{T}(r_{i,j}) & \vdots \\ 0 & & \dots & 0 \end{pmatrix}, \bar{U} = 0 \quad (13.6)$$

Die Fouriertransformation $\mathcal{F}\{U\}$ besitzt eine spektrale Auflösung von $\Delta k = \frac{1}{L\Delta x}$ gegenüber $\mathcal{F}\{T\}$ mit $\Delta k_x = \frac{1}{N\Delta x}$ bzw. $\Delta k_y = \frac{1}{M\Delta x}$.

Die kleinste, sinnvolle Wellenzahl k_{min} ergibt sich aus der größten Dimension von \hat{T} (nach Abtasttheorem, Shannon (1949)) zu $k_{min} = \frac{1}{2N\Delta x}$, die maximale, sinnvolle Wellenzahl $k_{max} = \frac{1}{\Delta x}$.

Die Autokorrelation von U - und damit mit entsprechender Auswahl innerhalb der Matrix auch von \hat{T} - ist der Realteil der inversen Fouriertransformation des quadrierten Autospektrums:

$$R(i \cdot \Delta x, j \cdot \Delta x) = \Re \left(\mathcal{F}^{-1} \left\{ \mathcal{F}\{U\}^2 \right\} \right) \quad (13.7)$$

Dabei stehen die Einträge gemäß der Indizes i, j für die jeweilige Verschiebung um $\Delta \vec{x} = (i \cdot \Delta x, j \cdot \Delta x)$ des Feldes T gegen sich selbst. Entsprechend wird die Autokorrelation R - falls notwendig - skaliert, sodass bei einer Verschiebung von Null $R(0, 0) = 1$ gilt.

Als *Korrelationslänge* λ_{Korr} wird der Betrag der Verschiebung definiert, bei dem

$$R(\Delta \vec{x}; \|\Delta \vec{x}\| = \lambda_{Korr}) = \exp(-1) \quad (13.8)$$

gilt. Da dieser Wert in Abhängigkeit von der Richtung der Verschiebung bei unterschiedlichen Längen auftreten kann und nur eine Übersicht über die Korrelationslänge unter der Annahme von Isotropie gewünscht wird, bietet es sich an, Autokorrelationen im gleichen Distanzintervall zu mitteln und damit - ähnlich wie das omnidirektionale Variogramm - auf eine richtungsunabhängige Distanz zu reduzieren (Shevyrnogov und Vysotskaya (2006)).

13.2 Zeitliche Autokorrelation

Das Feld $T(r_{i,j}, t_k)$ liegt für die diskreten Positionen $r_{i,j}$ auf einem gleichmäßigen Gitter vor ($i = \{0, \dots, N-1\}$, $j = \{0, \dots, M-1\}$) und wird für diskrete, äquidistante Zeitpunkte t_k , $k = \{0, \dots, K\}$ beobachtet. Sowohl in der Zeit als auch im Raum können Beobachtungen fehlen. Das klimatologische Mittel \hat{T} von T wird pixelweise als zeitlicher Erwartungswert berechnet:

$$\hat{T}(r_{i,j}) = \langle T(r_{i,j}, t) \rangle_t \quad (13.9)$$

Für jede zeitliche Realisation des Feldes wird die Anomalie $A(r_{i,j}, t_k)$ zum klimatologischen Mittel gebildet:

13 Theoretische Betrachtungen zur Autokorrelation

$$A(r_{i,j}, t_k) = T(r_{i,j}, t_k) - \widehat{T}(r_{i,j}) \quad (13.10)$$

Anomalien können sowohl aus den Chlorophyllkonzentrationen oder den Logarithmen der Konzentrationen berechnet werden.

Bei jedem Pixel $r_{i,j}$ wird nach allen existierenden Paaren von Anomaliewerten $A(t)$ und $A(t + \tau)$ mit jeweils festem zeitlichen Abstand $\tau = \{1, \dots, W\}; W < K$ gesucht. Gibt es für das Pixel $r_{i,j}$ und die zeitliche Distanz τ eine Anzahl $N(r_{i,j}, \tau)$ von Paaren, so berechnet sich die zeitliche Autokorrelation $R(r_{i,j}, \tau)$ als

$$R(r_{i,j}, \tau) = \frac{1}{N(r_{i,j}, \tau)} \sum_{t=1}^{K-\tau} A(r_{i,j}, t) \cdot A(r_{i,j}, t + \tau) \quad (13.11)$$

Die Korrelationszeit τ_{korr} als Maß dafür, bei welcher zeitlichen Distanz τ noch positive Korrelation vorliegt, wird nach vorheriger Skalierung auf $R(r_{i,j}, \tau = 0) = 1$ durch

$$R(r_{i,j}, \tau_{korr}) = \exp(-1) \quad (13.12)$$

definiert (Denman und Abbott (1988, 1994)).

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

Wie stark variieren die räumlichen Strukturen und ab welcher Distanz sind die Chlorophyllkonzentrationen unkorreliert?

Wie verändert sich die Korrelationslänge (Def. 13.8) im Jahresgang, bzw. gibt es eine jahreszeitliche Abhängigkeit der Korrelationslänge? Das heißt, gibt es Zeiträume mit eher kleinräumigen oder großräumigen Chlorophyllstrukturen und lassen sich diese unterscheiden?

Nach wie vielen Tagen haben sich auftretende Strukturen soweit verändert, dass sie nicht mehr korreliert sind?

Welche Zyklen im Sinne von wiederkehrenden Strukturen durchläuft die Nordsee auf einer zeitlichen Skala von Monaten?

Diese Fragen sollen in nachfolgenden Experimenten mit dem Chlorophylldatensatz anhand von Korrelationslänge oder Korrelationszeit beantwortet werden. Die Tabelle 14.1 gibt Auskunft über die benutzten Datensätze, ihre Zeiträume und Untersuchungsgebiete in nachfolgenden Abschnitten. In jedem Fall werden die Anomalien bzw. die zentrierten Felder explizit definiert, da zum Vergleich auch Chlorophylldaten des Standardprozessors herangezogen werden.

Einleitend wird der Einfluss von Schätzwerten auf die Berechnung der Autokorrelation getestet.

Autokorrelation Methode	Auflösung Δx bzw. Δt Analysegebiet	Datensatz	Ziel
räumlich (vgl. Abs.13.1 und Abs. 14.1)	Pixel $\Delta x = 1.5$ km, Gebiete 1-5	Datentage >95%, Jan. 2003 - Juli 2008	Korrelationslänge (Gl. 13.8) im Jahresgang
zeitlich (vgl. Abs. 13.2 und Abs. 14.2)	$\Delta t = 1$ Monat, Nordsee	<i>MLH.Median</i> Monatsmittel Jan. 2003 - Juli 2008	Zyklen der Nordsee
zeitlich (vgl. Abs. 13.2 und Abs. 14.3)	$\Delta t = 1$ Tag, Gebiete 1-5	alle Datentage, Jan. 2003 - Juli 2008	Korrelationszeit (Gl. 13.12) für Jahreszeiten

Tab. 14.1: Übersicht von räumlichen und zeitlichen Autokorrelationsuntersuchungen. In allen Fällen wird das Chlorophyllprodukt des Regionalprozessors analysiert, bei der Zeitreihe von Monatsmitteln zusätzlich das Standardprodukt *algal_2*.

14.1 Räumliche Autokorrelation

Die Berechnung der räumlichen Autokorrelation über die Fouriertransformation (gemäß Abschnitt 13.1) verlangt vollständige, rechteckige Untersuchungsräume, ein Kriterium, dass in den Gebieten 1-5 innerhalb des gesamten Zeitraumes nur 13 mal erfüllt wird. Mit dem Ziel, eine Aussage über die zeitliche Entwicklung der Korrelationslänge in den Gebieten treffen zu können, wird es daher unumgänglich sein, unvollständige Datentage mittels eines Schätzwertverfahrens zu vervollständigen. Welchen Einfluss der Anteil und die Positionierung fehlender Datenpunkte auf die Autokorrelation hat, soll im Folgenden kurz an einem Beispiel erläutert werden.

Die Chlorophylldaten des Regionalprozessors werden wie folgt bearbeitet:

- nicht erkannte Wolkenränder werden über die optische Dichte der Aerosole identifiziert und aussortiert (vgl. Abschnitt 4.2.1).
- Datentage mit mehr als 95% Datenüberdeckung im jeweiligen Gebiet werden für die Untersuchung ausgewählt und fehlende Datenpunkte mittels Ordinary Kriging geschätzt.
- Für die Fouriertransformation werden die Logarithmen der Chlorophyllkonzentrationen zentriert und in eine größere, quadratische Umgebung eingeschrieben. Der Übergang zwischen Umgebung und zentrierten Daten wird zur Minderung der Unstetigkeit mit Cosinusflanken (Tukey-Fenster, Harris (1978)) geglättet.
- Nachdem die zweidimensionale Autokorrelation $R(i \cdot \Delta x, j \cdot \Delta x)$ bezüglich $R(0, 0) = 1$ skaliert worden ist, wird sie zu einer richtungsunabhängigen Autokorrelation zusammengefasst, d.h. über konstante Distanzintervalle gemittelt. Anhand dieser Verläufe wird die Korrelationslänge bestimmt.

14.1.1 Einfluss von Schätzwerten auf die räumliche Autokorrelation

Bis jetzt ist quantitativ untersucht worden, wie gut Schätzwerte im Mittel an die Originalwerte heranreichen (z.B. für das Kriging-Verfahren in Kapitel 9 und Kapitel 10 oder die Dateninterpolationsmethode DINEOF in Kapitel 12), nicht jedoch, wie gut ihre räumliche Struktur derjenigen des Originals entspricht. Es ist zu erwarten, dass durch die glättenden Eigenschaften des Kriging sich die räumliche Struktur insbesondere in kleinen Distanzbereichen verändert.

Wie groß ist aber die Wirkung von Schätzwerten in Abhängigkeit vom Grad und der Struktur fehlender Datenpunkte durch Wolkenbedeckung auf die räumliche Autokorrelation?

Als Beispiel dienen Daten vom 8. August 2003 in der zentralen Nordsee östlich der Doggerbank (Gebiet 3), die das Gebiet vollständig überdecken (Abb. 14.1).

Mit Wolkenmasken, deren Bedeckungsgrade zwischen 5-50% liegen sowie stark konzentrierte oder eher gleichverteilte Strukturen besitzen (Abb. 14.2; vgl. Abschnitt 10.1, "Zufallsbewegung"), werden Datenpunkte entfernt. An ihrer Stelle werden Ordinary Kriging Schätzwerte eingefügt.

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

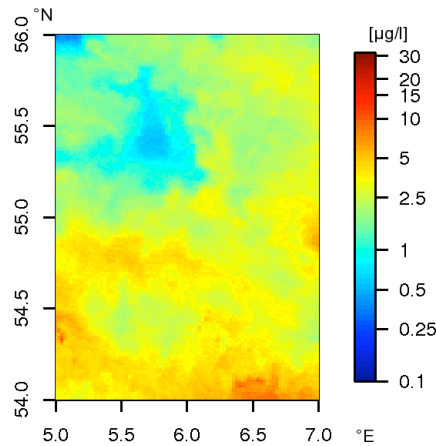


Abb. 14.1: Die Chlorophyllkonzentration in $\mu\text{g/l}$ vom 8. August 2003 dient als Vergleichsbasis für die Veränderung der räumlichen Autokorrelation beim Ersetzen von Datenpunkten durch Schätzwerte.

Die Differenz zwischen der räumlichen Autokorrelation der Originaldaten und des durch Schätzwerte aufgefüllten Datensatzes ist gemäß der Erwartung besonders groß bei kleinen Distanzen für die stark konzentrierten Wolkenformen (Abb. 14.3). Je stärker die Häufung fehlender Datenpunkte, umso weniger bleibt in den Schätzwerten von der ursprünglichen Struktur erhalten. Während die annähernd gleichverteilte Wolkenmaske für größere Differenzen bei höherem Bedeckungsgrad sorgt, ist die Interpretation für die gehäuften Verteilungen nicht so einfach. Hier muss zusätzlich berücksichtigt werden, wie die Daten beschaffen sind, die durch die Wolkenmaske ausgeschlossen werden - eine Information, die bei echten Wolken a priori nicht zugänglich ist. Von diesem zugrunde liegenden Grad von Strukturiertheit hängt ab, wie gut die Schätzwerte tatsächlich die Originale repräsentieren können: werden stark variable Gebiete durch die Maske ausgeschlossen, so können die Schätzwerte die Struktur nur ungenügend wiedergeben und die rekonstruierte räumliche Autokorrelation unterscheidet sich erheblich vom Original. Werden dagegen relativ homogene Gebiete ausgeschlossen, so sind die Schätzwerte eine gute Repräsentation der Originalwerte und die Differenz der räumlichen Autokorrelation fällt geringer aus. Diese Überlegung kann erklären, warum die größten Differenzen im Falle konzentrierter Strukturen nicht direkt vom Bedeckungsgrad abhängen (vgl. Abb. 14.3b).

Insgesamt sind die Abweichungen der Korrelation durch die Einführung von Schätzwerten aber so gering (Abb. 14.4a), dass sie selbst bei einem Anteil von 50% Schätzwerten unabhängig von der Häufung kaum von der Autokorrelation des vollständigen Originals zu unterscheiden ist. In diesem Beispiel beträgt die Abweichung maximal 1.5%.

In weiteren Versuchen (Gebiet 1: 13. Juli 2003, 10. Mai 2006, 16. Juli 2006, 21. September 2006, 7. Mai 2008; Gebiet 3: 15. Juli 2003, 8. August 2003, 8. August 2004, 26. Juli 2006, 21. September 2006, 22. April 2008, 7. Mai 2008, 10. Mai 2008) beträgt sie maximal 5% im stark konzentrierten Fall (Abb. 14.4c). Für einen Bedeckungsgrad von nur 5% beträgt die maximale prozentuale Abweichung nur 0.2% - unabhängig von der Häufung - und ist daher gut mit dem

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

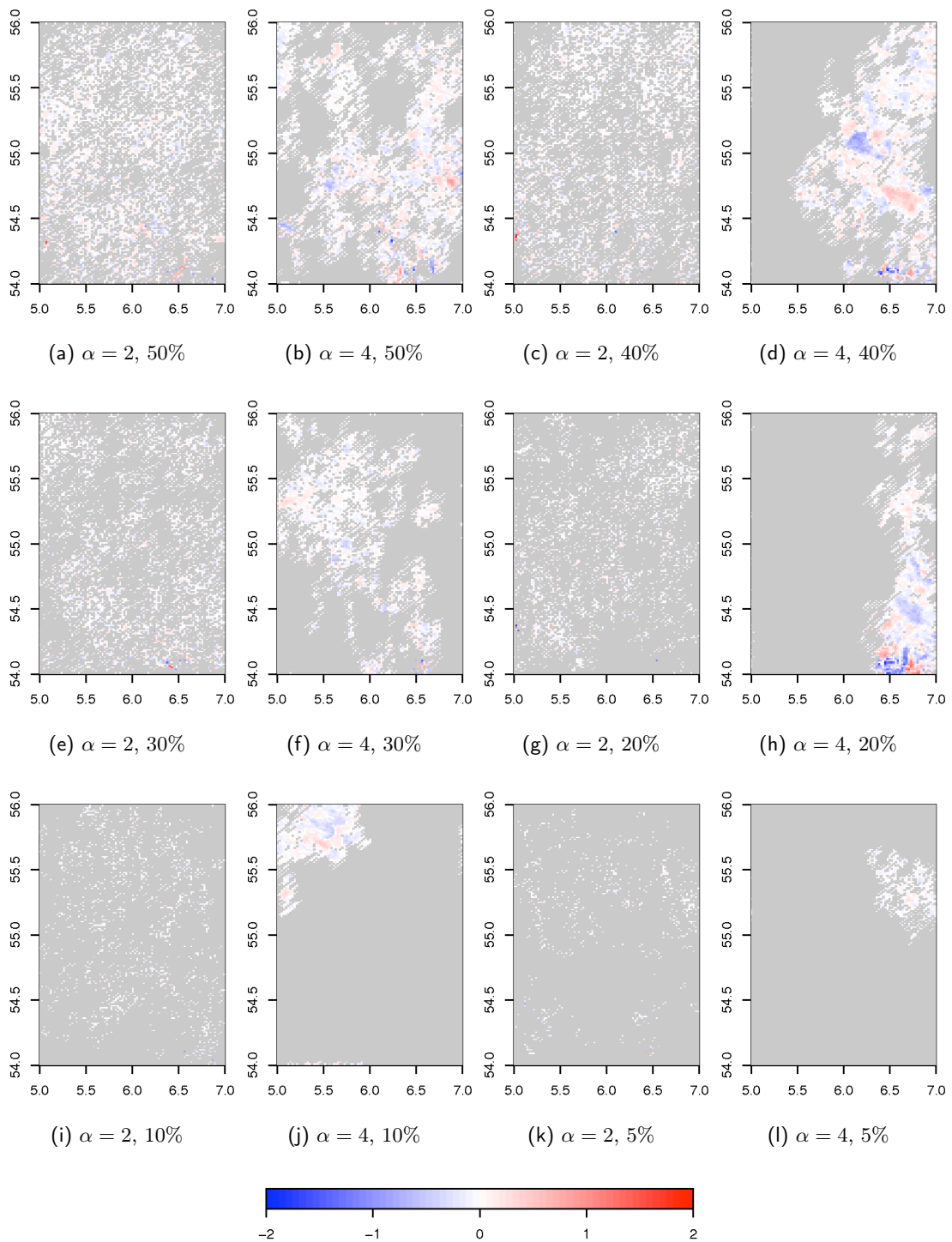


Abb. 14.2: Tatsächlicher Fehler in $\mu\text{g}/\text{l}$ zwischen Rekonstruktion der Chlorophyllkonzentrationen mittels Ordinary Kriging (mit Wolkenmasken *Zufallsbewegung* der Ballungsstärken $\alpha = 2, 4$ und Bedeckungsgraden zwischen 5-50%) und den Originalwerten (Abb. 14.1).

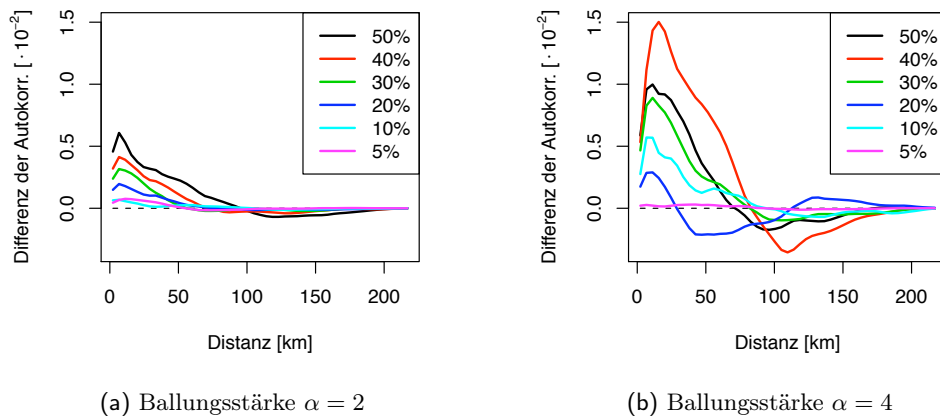


Abb. 14.3: Differenz der räumlichen (richtungsunabhängigen) Autokorrelation der vervollständigten Datensätze zu der des Originals für verschiedene Anteile der Schätzwerte bezogen auf die Gesamtzahl von Datenpunkten über Wasser. Die Rekonstruktionen werden an Stelle von *Zufallsbewegungs*-Strukturen mit einer Ballungsstärke $\alpha = 2, 4$ und Bedeckungsgraden zwischen 5-50% eingefügt (vgl. Abb. 14.2).

Original gleichzusetzen.

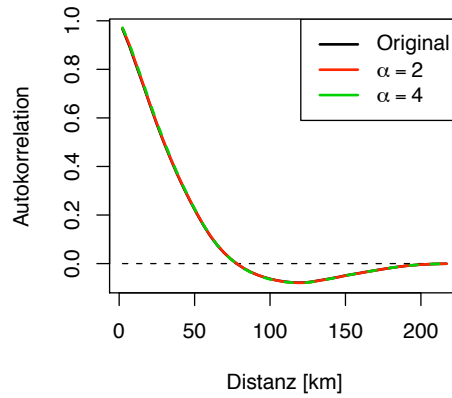
14.1.2 Bestimmung von Korrelationslängen

Aus den Datentagen des Regionalprozessors zwischen dem 3. Januar 2003 und dem 8. Juli 2008 werden diejenigen ausgewählt, die in den fünf Gebieten jeweils eine Überdeckung größer als 95% aufweisen (vgl. Tabelle 14.2).

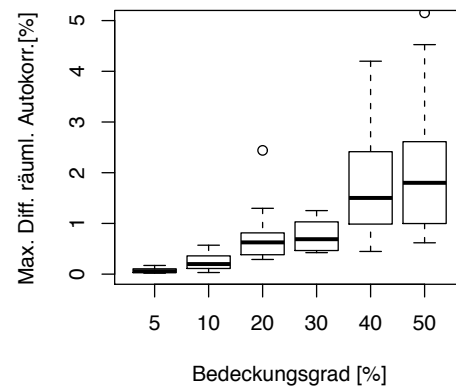
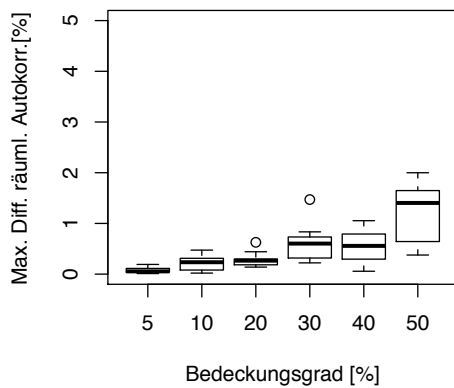
Gebiet	Anzahl gesamt	<i>Frühjahr</i> 1. Februar - 14. Mai	<i>Sommer</i> 15. Mai - 14. August	<i>Herbst</i> 15. August - 14. November
Doggerbank (1)	29	10 (1 0 2 2 1 4)	16 (4 0 3 3 4 2)	3 (0 0 2 1 0 0)
Engl. Kanal (2)	61	26 (4 0 2 5 7 8)	30 (6 2 7 9 3 3)	5 (1 0 2 1 1 0)
Zentr. Nordsee (3)	48	22 (4 1 3 1 5 8)	17 (3 3 3 4 2 2)	9 (1 1 3 4 0 0)
Deutsche Bucht (4)	83	31 (5 0 4 3 8 11)	38 (5 8 4 9 5 7)	14 (2 1 4 5 2 0)
Dänische Küste (5)	69	24 (1 2 5 3 7 6)	40 (5 7 5 15 4 4)	5 (0 0 2 3 0 0)

Tab. 14.2: Anzahl der Datentage für die Gebiete 1-5 für die Jahreszeiten Frühjahr, Sommer und Herbst mit mehr als 95% Überdeckung. In Klammern findet sich die Anzahl von Beispielen nach Jahren aufgeschlüsselt, jeweils von 2003 bis 2008.

Gemäß der Untersuchung des Einflusses von Schätzwerten (vgl. Abschnitt 14.1.1) ist hier immer mit gegenüber dem Original vernachlässigbaren Abweichungen der Autokor-



(a) Räumliche Autokorrelation für 50% Bedeckung.



(b) Maximale Differenz, Ballungsstärke $\alpha = 2$

(c) Maximale Differenz, Ballungsstärke $\alpha = 4$

Abb. 14.4: Autokorrelation der Chlorophyllverteilung vom 8. August 2003 bzw. der mit Schätzwerten vervollständigten Datensätze. Die Rekonstruktionen werden anstelle von Datenpunkten in der Struktur der *Zufallsbewegung* mit einer Ballungsstärke $\alpha = 2, 4$ und verschiedenen Bedeckungsgraden eingefügt (vgl. Abb. 14.2).

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

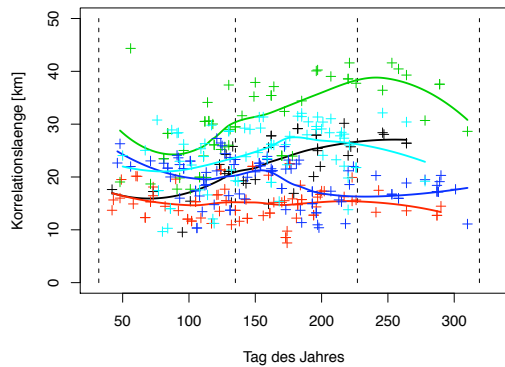


Abb. 14.5: Korrelationslängen im Jahresgang für die Gebiete 1-5 (farbcodiert: 1: schwarz, 2: rot, 3: grün, 4: blau, 5: cyan) ohne Berücksichtigung des Beobachtungsjahres.

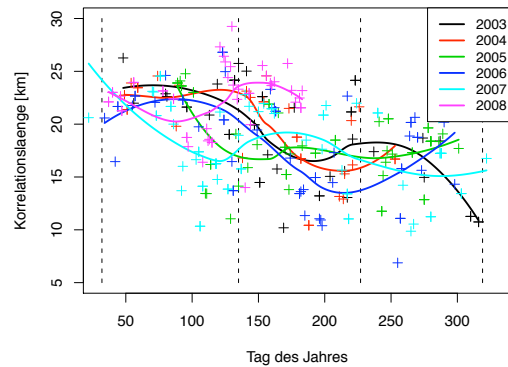


Abb. 14.6: Jahresgang der Korrelationslänge jeweils für die Jahre 2003 bis 2008 in der Deutschen Bucht (Gebiet 4). Datensatz aus >60% Überdeckung.

relation zu rechnen.

Nachdem die fehlenden Datenpunkte mittels Ordinary Kriging gefüllt und die zweidimensionale räumliche Autokorrelation auf eine richtungsunabhängige eindimensionale reduziert wurde, kann die (mittlere) Korrelationslänge bestimmt werden.

Um ihre zeitliche Entwicklung zu verdeutlichen, werden die Korrelationslängen als Funktion der Zeit (in Tagen) mittels einer stark glättenden lokalen Regression (Cleveland und Devlin (1988)) angepasst.

14.1.3 Diskussion

Für die einzelnen Gebiete werden für jeden Datentag, der ausreichende Überdeckung besitzt, die Korrelationslängen bestimmt. Aus den einzelnen zeitlichen Ergebnissen lässt sich der Verlauf der Korrelationslänge als Funktion der Zeit mit einer zeitlichen Auflösung von Tagen ableiten (Abb. 14.5). Hier bleibt unberücksichtigt, in welchem Jahr die Beobachtung gemacht worden ist. Einzig der Anzahl von Tagen nach Beginn des Jahres wird Rechnung getragen. Auf eine Berücksichtigung der interannualen Variabilität wird in diesem Fall verzichtet.

Bis auf die Region des Englischen Kanals (Gebiet 2) zeigen die Gebiete einen deutlich strukturierten Jahresgang: die Doggerbank, die zentrale Nordsee und die Bereiche vor der Dänischen Küste (Gebiete 1, 3, 5) zeichnen sich durch geringere Korrelationslängen im Frühjahr und einem Anstieg zum Sommer aus.

- Im Bereich der Doggerbank (Gebiet 1) steigt die Korrelationslänge von 17 km im Frühjahr kontinuierlich bis auf 26 km zum Ende des Sommers an. Da für den Herbst nicht ausreichend Daten zur Verfügung stehen, lässt sich ein erneutes Abfallen der Korrelationslänge nicht belegen. Dieses Ergebnis deckt sich mit der Dynamik der

Region, in der kleinräumig stark strukturierte Verteilungen von Chlorophyllkonzentrationen im Frühjahr (Frühjahrsblüten) von wesentlich strukturärmeren in den Sommermonaten abgelöst werden.

- Die Region im Englischen Kanal (Gebiet 2) zeichnet sich durch eine annähernd gleich bleibende Korrelationslänge von 15 km aus. Hier prägt sich kein saisonales Signal aus, die auftretenden kleinräumigen Strukturen sind im gesamten Jahresverlauf zu finden.
- In der zentralen Nordsee östlich der Doggerbank (Gebiet 3) herrscht im Frühjahr noch Korrelation bei einer Länge von 25 km (der Punkt bei Tag 56, dem 25. Februar 2003, ist ein Ausreißer), steigt etwa ab Tag 100 (10. April) den gesamten Sommer hindurch an, bis sie am Anfang der Herbstes (Tag 245, 2. September) ihr Maximum bei 38 km findet. Danach nimmt sie deutlich wieder ab. Dies ist insgesamt die Region mit den größten Korrelationslängen, was für überwiegend große Strukturen bzw. Strukturarmut im kleinskaligen Bereich spricht. Für diese Region in der zentralen Nordsee mit sommerlicher Schichtung und Chlorophyllminimum erscheint diese Interpretation durchaus sinnvoll. Gleichzeitig ist das Gebiet 3 schon per Definition das größte der Untersuchungsgebiete und erlaubt die größten Verschiebungen.
- Die Deutsche Bucht (Gebiet 4) präsentiert einen uneinheitlichen Gang der Korrelationslänge: das Frühjahr beginnt mit einem leichten Abfall und geht in den Sommer mit einem leichten Anstieg zum (lokalen) Maximum von 20 km über, bevor die Werte noch im Sommer auf etwa 17 km fallen und dort verbleiben. Die geringe Strukturiertheit des mittleren Verlaufes liegt in der interannualen Variabilität begründet. Um die Unterschiede in den Jahresgängen zu verdeutlichen, wird der Datensatz für dieses Gebiet auf alle Datentage mit mehr als 60% Überdeckung erweitert (Abb. 14.6). Nicht alle Jahresgänge folgen einem ähnlichen Muster. Ob damit direkt auf starke interannuale Variabilität geschlossen werden darf, ist durch die ungleichmäßige zeitliche Überdeckung nicht gesichert.
- Die Korrelationslänge der Chlorophyllverteilungen, die vor der Dänischen Küste (Gebiet 5) beobachtet wird, steigt von ihrem Minimum im Frühjahr von etwa 20 km (Tag 90, 31. März) zu einem Maximum im Sommer von 28 km (Tag 190, 9. Juli) an. Auch hier ist im Sommer mit geringer Strukturierung respektive größerer Einheitlichkeit der Chlorophyllverteilung zu rechnen.

14.2 Zeitliche Autokorrelation der Chlorophyllmonatsmittel

Für den Zeitraum von Januar 2003 bis Juli 2008 werden jeweils die Monatsmittel (als Mediane größter Wahrscheinlichkeit, MLH-Median, vgl. Abschnitt 7.2) der Chlorophyllkonzentration im Bereich der Nordsee ermittelt. Als Chlorophyllprodukte werden sowohl die des Regionalprozessors auch auch des Standardprozessors jeweils auf dem Standardgitter zusammengefasst. Zum Zeitpunkt der Berechnung der Monatsmittelwerte existieren beide Wolkenrandfilter (vgl. Abschnitt 4.2.1 bzw. Abschnitt 4.2.2) noch nicht, sodass

Wolkenränder nur über einen ungenaueren, statistischen Filter aufgefunden werden. Aus Zeitgründen, und da die einzelnen Monatsmittel hinreichend Artefakt frei erscheinen, wird die Berechnung nicht wiederholt. Über den gesamten Zeitraum wird jeweils das klimatologische Mittel (ebenfalls als MLH-Median) berechnet.

Um hochfrequentes Bildrauschen (z.B. durch nicht entdeckte Wolkenränder) zu minimieren, wird jedes Monatsmittel mit einem 3×3 -Pixel Median-Filter (vgl. Abschnitt 4.1.2) im logarithmischen geglättet.

Drei Definitionen für die Anomalien (Gl. 13.10) bilden die Basis für die folgende Untersuchung:

- $A_{\text{chl_conc}} = \text{chl_conc} - \overline{\text{chl_conc}}$: Differenz der Chlorophyllkonzentration eines Monatsmittels und des klimatologischen Mittels über den Gesamtzeitraum für das Produkt des Regionalprozessors. Diese Bearbeitung des Datensatzes berücksichtigt insbesondere die hohen Chlorophyllkonzentrationswerte.
- $A_{\log(\text{chl_conc})} = \log(\text{chl_conc}) - \log(\overline{\text{chl_conc}})$: Differenz der Logarithmen der Chlorophyllkonzentration eines Monatsmittels und des klimatologischen Mittels über den Gesamtzeitraum für das Produkt des Regionalprozessors. Bei dieser Konstellation werden geringe Konzentrationen stärker betont.
- $A_{\text{algal_2}} = \text{algal_2} - \overline{\text{algal_2}}$: Differenz der Chlorophyllkonzentration eines Monatsmittels und des klimatologischen Mittels über den Gesamtzeitraum für das Produkt des Standardprozessors.

Aus diesen Anomalien wird die zeitliche Autokorrelation nach Gl. 13.11 ermittelt.

Nachdem ein Sigma-Filter (vgl. Abschnitt 4.1.1) auf die Autokorrelation $R(\tau)$ mit $\tau \in \{0, 1, \dots, 24\}$ Monaten angewendet worden ist, werden diese pixelweisen zeitlichen Serien durch $R(r_{i,j}, \tau = 0)$ skaliert.

Diese Karten $R(r_{i,j}, \tau)$ sind Grundlage für die individuellen Auswertungen.

14.2.1 Nordsee

Die Korrelation $R(\tau)$ unterscheidet sich qualitativ und quantitativ für die Anomalien der verschiedenen Datensätze (Abb. 14.7¹):

- Basierend auf chl_conc folgt das Mittel von R über alle Pixel zu einer festen Zeitdifferenz τ hauptsächlich dem Jahresgang. Die Korrelation ist für $\tau = 12, 24$ Monate maximal und positiv, während sie für weite Zeitdifferenzbereiche fast konstant Null beträgt. Die Standardabweichung ist im Vergleich zu den beiden anderen Datensätzen ($\log(\text{chl_conc})$ und algal_2) am geringsten.
- Für $\log(\text{chl_conc})$ deutet sich eine andere Abhängigkeit an: weiterhin ist der Jahresgang mit Maxima der Korrelation bei $\tau = 12, 24$ Monaten dominierend (und am stärksten ausgeprägt für die verschiedenen Chlorophyllmaße), jedoch sind die Zeitbereiche zwischen den Maxima anti- bzw. unkorreliert. Gleichzeitig gibt es erste

¹Es fehlt jeweils der Punkt $R(\tau = 0) = 1$.

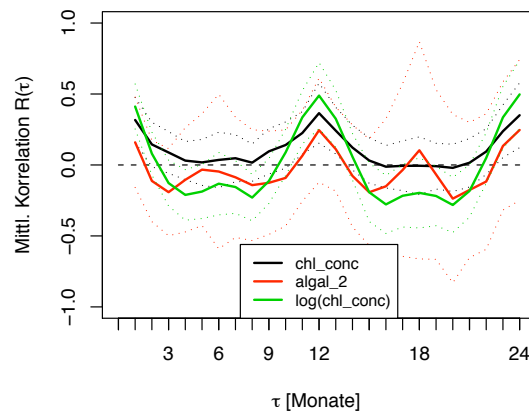


Abb. 14.7: Mittel (durchgezogene Linie) und Standardabweichung (gestrichelte Linie als Intervall um das Mittel) der zeitlichen Autokorrelation über die ganze Nordsee als Funktion der Zeitdifferenz τ für die Chlorophyllkonzentration gemäß Regional- bzw. Standardprozessor.

Anzeichen für eine halbjährliche Struktur. Das Intervall der Standardabweichung ist relativ eng, insbesondere um die Maxima.

- Für *algal_2* besitzt die Korrelation einen deutlichen Halbjahresgang. Allerdings ist hier auch die Standardabweichung am größten und gerade in den Bereichen des Maximums bei $\tau = 6, 18$ Monaten wird das Intervall möglicher Korrelationswerte sehr groß.

Da für jedes Pixel innerhalb der Nordsee eine solche Abhängigkeit der Korrelation zugänglich ist, werden Periodizitätsuntersuchungen durchgeführt und die vorherrschende Periodendauer als Karte der Nordsee zusammengetragen. Dazu wird die Auflösung der Karte reduziert, 5×5 Pixel des Standardgitters ergeben nun ein neues Pixel.

Die Nordsee lässt sich grob in Gebiete mit jährlichem bzw. halbjährlichem Zyklus unterteilen (Abb. 14.8). Dabei ist “Zyklus” nicht direkt gleichbedeutend mit “Chlorophyllkonzentrationsgang”. Schließlich wird ein in zeitlichem Abstand wiederkehrender niedriger Chlorophyllwerte genauso zu einer positiven Korrelation führen wie ein hoher.

Wiederum gibt es große Unterschiede zwischen den drei Chlorophylldatensätzen:

- Für *algal_2* zeigt die Karte der vorherrschenden Periodendauer (Abb. 14.8a) einige stark ausgeprägte Bereiche: aus den weiten Bereichen eines halbjährlichen Zyklus sticht der Raum über der Doggerbank mit jährlichem Zyklus (11-13+ Monate) hervor. Er fällt (bis auf den Ostteil) erstaunlich gut mit der ICES-Region 7b zusammen.
- Die Struktur der vorherrschenden Periodendauer ist für die logarithmierte Konzentration des Chlorophyllproduktes des Regionalprozessors, *log(chl_conc)* in wesentlichen Teilen sehr verschieden (Abb. 14.8b).

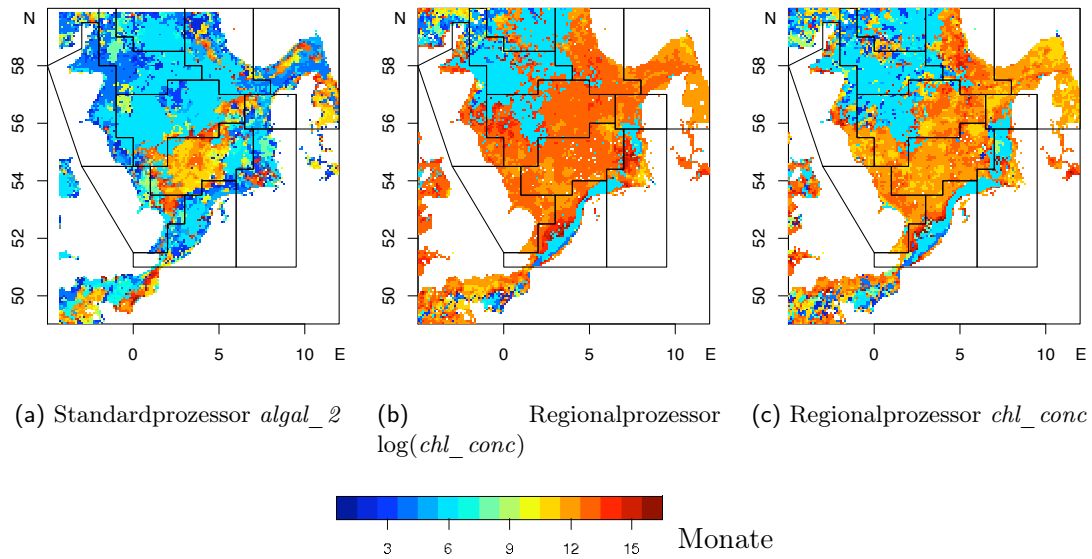


Abb. 14.8: Karte mit vorherrschender Periodendauer (in Monaten) der pixelweisen zeitlichen Autokorrelation

- Der wichtigste Unterschied zwischen den maximal korrelierten Perioden aus *chl_conc* oder $\log(chl_conc)$ für den vom Jahresgang geprägten Teil liegt in der südlichen und östlichen Nordsee. Während die logarithmierte Konzentration eine Periode von 13 Monaten nahe legt, sind es im unlogarithmierten Fall 12 Monate (Abb. 14.8c).

14.2.2 ICES-Regionen in der Nordsee

Innerhalb der ICES-Regionen (vgl. Abb. 2.2) wird die Korrelation $R(\tau)$ für jede zeitliche Verschiebung τ gemittelt und auf die vorherrschende Periodendauer zeitlicher Korrelation untersucht.

Wie es die Karten (Abb. 14.8) nahe legen, unterscheidet sich die Korrelation als Funktion der zeitlichen Verschiebung sowohl nach Gebiet als auch nach Chlorophyllprodukt.

In der nördlichen Nordsee (Region 1) herrschen für alle Chlorophyllprodukte sechsmo-
natige Zyklen vor (Abb. 14.9a). Qualitativ besteht jedoch trotz der gleichen (mittleren) Hauptperiode ein großer Unterschied zwischen den Produkten: während für *chl_conc* (schwarze Linie) der Verlauf zwischen relativen Maxima bei $\tau = 6, 12, 18, 24$ Monaten mit positiven Korrelationswerten und unkorrelierten Bereichen ($R \simeq 0$) abwechselt, tritt positive Korrelation für $\log(chl_conc)$ (grüne Linie) nur als Jahressignal auf. Zu halb-jährlichen Distanzen ist das Signal dagegen unkorreliert ($R(\tau = 6, 18) \simeq 0$), während das Signal sonst eher antikorreliert erscheint. Für *algal_2* wechselt das Signal zwischen anti- und positiv korreliert.

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

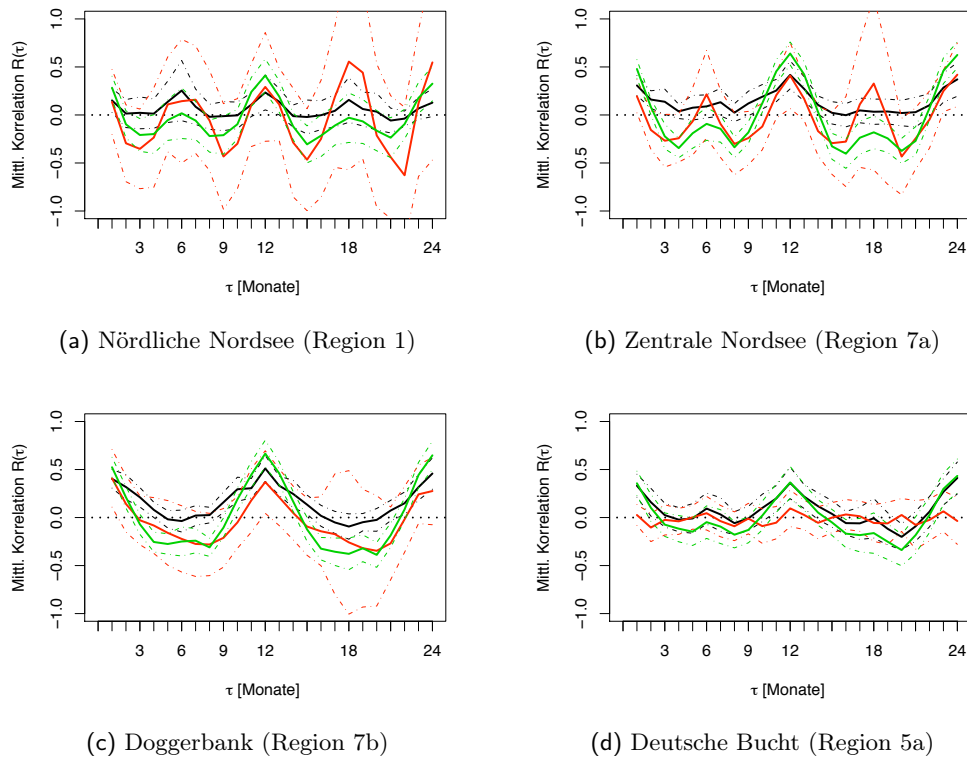


Abb. 14.9: Mittlere Autokorrelation $R(\tau)$ für jeweils eine Modellregion (ICES, ERSEM, vgl. Abb. 2.2). Verschiedene Anomalien sind farblich gekennzeichnet: basierend auf chl_conc : schwarz, $algal_2$: rot, $log\ chl_conc$: grün (vgl. Abb. 14.7). Gestrichelte Linien stehen jeweils für das Intervall der Standardabweichung um den Mittelwert.

In der zentralen Nordsee (Region 7a) verhalten sich die Signale sehr ähnlich zu Region 1 (Abb. 14.9b). Der Verlauf von chl_conc ist jedoch wesentlich näher an einem ganzjährlichen Zyklus, ebenso die Korrelation von $log(chl_conc)$, bei der die Amplituden des Ganzjahressignals stärker ausfallen als zuvor.

Ganz im Zeichen von jährlichen Zyklen steht die Doggerbank-Region (7b, Abb. 14.9c). Während das Signal von chl_conc aber zwischen den Maxima positiver Korrelation und Unkorreliertheit pendelt, zeigen $log(chl_conc)$ und $algal_2$ antikorreliertes Verhalten.

Die stark variable Deutsche Bucht (Region 5a, Abb. 14.9d) besitzt gegenüber den anderen Gebieten eine nur schwach ausgeprägte zeitliche Korrelation. Das Jahressignal ist nur in den Produkten $log(chl_conc)$ und chl_conc sichtbar, im $algal_2$ erscheint das Signal durchgängig als unkorreliert. Das Jahressignal in den Produkten des Regionalprozessors kann hier lediglich aus dem systematischen Fehler stammen, wie er in den Jahresgängen der Chlorophyllkonzentration (Abb. 5.4) zutage tritt.

14.2.3 Kleinere Gebiete in der Nordsee

Die mittlere Korrelation soll in den kleineren Gebieten (vgl. ihre Definition und Lage in Abb. 2.4) zur Vervollständigung der Untersuchung des Nordseegebietes betrachtet werden.

Die Signale im Bereich der Doggerbank (Gebiet 1) ähneln denjenigen aus der größeren Region 7b (Abb. 14.10a). Ein auffälliger Unterschied zeigt sich im *algal_2*-Signal, das jeweils zu $\tau = 7 - 8, 20$ Monaten, d.h. etwa ein bis zwei Monate nach dem halben Jahr das Minimum in der Korrelation erreicht und dann vergleichsweise stark antikorreliertes Verhalten nahe legt.

In der Region des Englischen Kanals (Gebiet 2) zeigen die Produkte des Regionalprozessors einen erkennbaren Jahresgang (Abb. 14.10b, vgl. zuvor Region 5a, Abb. 14.9d), für das Standardprodukt *algal_2* liegt erst nach zwei Jahren wieder deutbare positive Korrelation vor.

In der zentralen Nordsee (Gebiet 3) dominiert das Jahressignal die Produkte des Regionalprozessors, während das *algal_2*-Signal hochfrequenter Anteile besitzt (Abb. 14.10c).

Wie zuvor beschrieben (vgl. Region 5a) zeigt das *algal_2*-Signal hauptsächlich unkorreliertes Verhalten in der Deutschen Bucht (Gebiet 4) (Abb. 14.10d), während in den *chl_conc*-Produkten ein Jahressignal erkennbar bleibt. Auffällig ist - ähnlich wie im Bereich der Doggerbank (Gebiet 1) - ein Minimum der Korrelation nach $\tau = 8, 20$ Monaten.

Das Areal vor der dänischen Küste (Gebiet 5) wird geprägt von mehr oder minder stark ausgeprägten Jahresgängen (Abb. 14.10e), wobei insbesondere das Signal von *algal_2* uneinheitlicher erscheint - dies dürfte wiederum ein Einfluss des systematischen Fehlers im Regionalprozessor sein.

14.3 Zeitliche Autokorrelation der Chlorophylldatentage

Für 1324 Datentage zwischen dem 3. Januar 2003 und dem 8. Juli 2008 (siehe Überdeckungsstatistik Abb. 14.11) liegen Chlorophyllwerte des Regionalprozessors vor. Für die fünf zuvor definierten Gebiete (Abschnitt 2.2.3) werden jeweils alle Daten zusammengestellt, wenn sie mehr als 5% Überdeckung der Region gewährleisten, um so die Aussortierung von nicht erkannten Wolkenpixel zu ermöglichen (vgl. Abschnitt 4.2.1).

Die beobachtbare Dynamik in den Karten der Chlorophyllkonzentration legt nahe, dass es bei einer zeitlichen Auflösung von Tagen im Gegensatz zu den Monatsmitteln nicht sinnvoll ist, eine Bestimmung der Autokorrelation mit Verschiebung τ über den gesamten Zeitraum vorzunehmen. Innerhalb jedes Gebietes wird die Untersuchung daher auf jeweils eine Jahreszeit (vgl. Definition des Phytoplanktonjahres, Abschnitt 3.4) über alle beteiligten Jahre beschränkt. Aus Tabelle 14.3 entnehme man die Anzahl einfließender Datentage. Der Winter wird aufgrund mangelnder Überdeckung nicht betrachtet. Im Gegensatz zur räumlichen Autokorrelation wird hier ausschließlich mit Originalpixel gearbeitet und keinerlei Schätzwerttechnik angewendet.

Die pixelweise zu analysierenden Anomalien $A(r_{i,j}, t)$ sind hier durch die logarithmierte Chlorophyllkonzentration $T(r_{i,j}, t)$ mit $t \in \{\text{Frühjahr, Sommer, Herbst}\}$ und das jeweilige

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

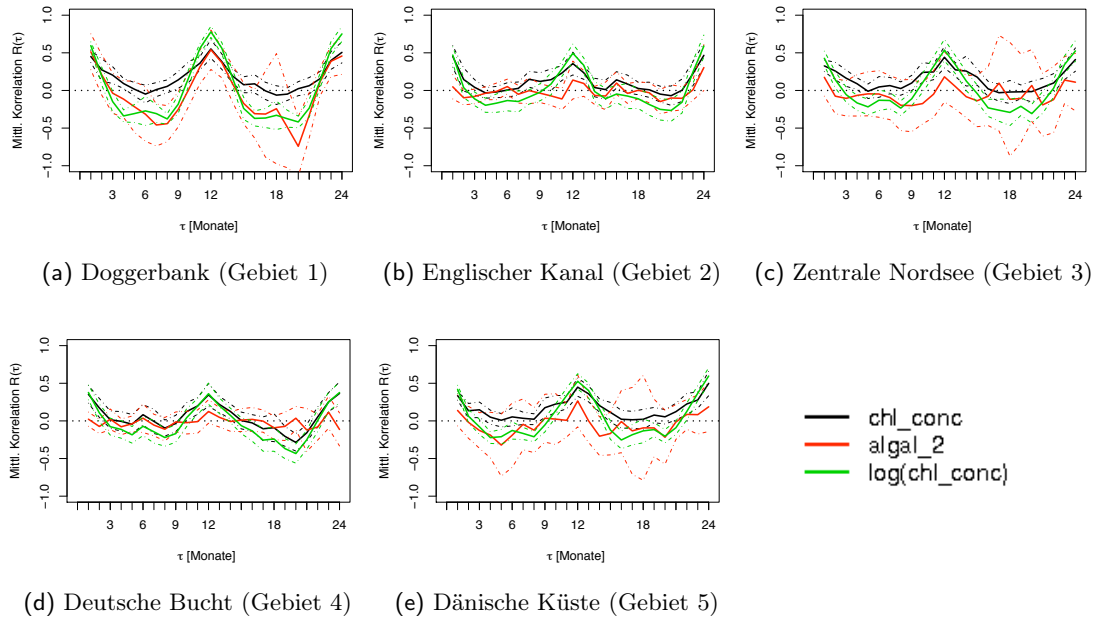


Abb. 14.10: Mittlere Autokorrelation $R(\tau)$ für jeweils ein Gebiet. Verschiedene Anomalien. Gestrichelte Linien stehen jeweils für das Intervall der Standardabweichung um den Mittelwert.

klimatologische Mittel als Logarithmus des MLH-Medians der Chlorophyllkonzentration $\hat{T}_{\text{Jahreszeit}}(r_{i,j})$ gegeben.

Um die Korrelation $R(r_{i,j}, \tau)$ skalieren zu können, wird die maximale Korrelation für $\tau = 0$ ermittelt und diese im Weiteren auf $R(r_{i,j}, \tau = 0) = 1$ gesetzt. Alle übrigen Autokorrelationswerte werden gemäß Gl. 13.11 für $\tau \in \{1, \dots, 45\}$ [Tage] berechnet und anschließend skaliert.

Eine Mittelung über alle Pixel in der jeweiligen Region ergibt somit typische Verläufe der Korrelation in Abhängigkeit von der zeitlichen Distanz τ in Tagen pro Gebiet und Jahreszeit.

Zur Bestimmung der Korrelationszeit wird jeweils der Verlauf $R(\tau)$ mit lokaler Regression angepasst (Abb. 14.12, rote Linie). Inwiefern die so bestimmte Zeit zu einer sinnvollen Interpretation bei dem derzeitigen Datensatz führen kann, bleibt durch die starken Artefakte aus dem systematischen Fehler fraglich (vgl. Abschnitt 5.5).

Während im Frühjahr über der Doggerbank (Gebiet 1) ausgeprägte Strukturen zu finden sind, die dafür verantwortlich sind, dass das 3-Tage-Signal nicht stark hervortritt, wird im Sommer mit seinen annähernd zeitlich konstanten, homogenen Chlorophyllverteilungen dieses Signal deutlich sichtbar (Sommerminimum bei Schichtung). Hier löst die Autokorrelation fast ausschließlich den systematischen Versatz des Chlorophyllproduktes auf, der sich in der extremen Streuung in $R(\tau)$ niederschlägt. Für $\tau = 3$ Tage wird die Korrelation wie durch die Frequenzanalyse zu erwarten maximal (vgl. Abb. 5.18), da dies eine zeitliche Differenz ist, bei der sich die Beobachtungsgeometrie aus Satelli-

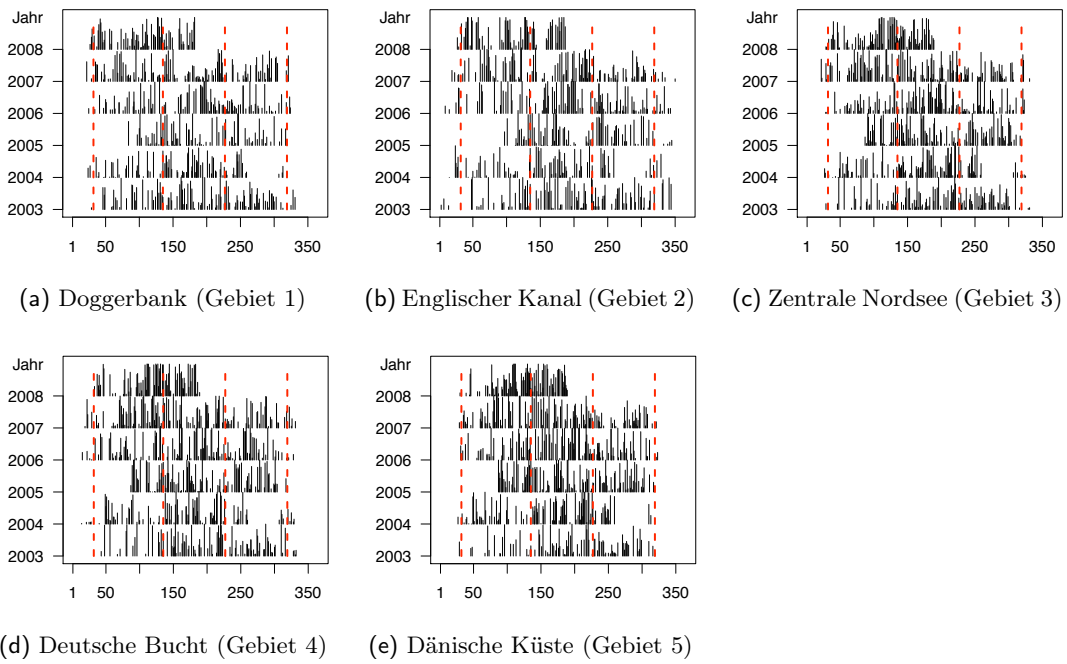


Abb. 14.11: Überdeckungsstatistik als Überdeckungsgrad pro Tag für die Gebiete aufgeschlüsselt nach Jahren 2003 bis 2008 ausgehend von Daten des Regionalprozessors nach der Entfernung von Wolkenrändern (tau_550-Filter, Abschnitt 4.2.1). Die Balkenlänge zwischen zwei aufeinander folgenden Jahren entspricht gerade 100% Datenüberdeckung.

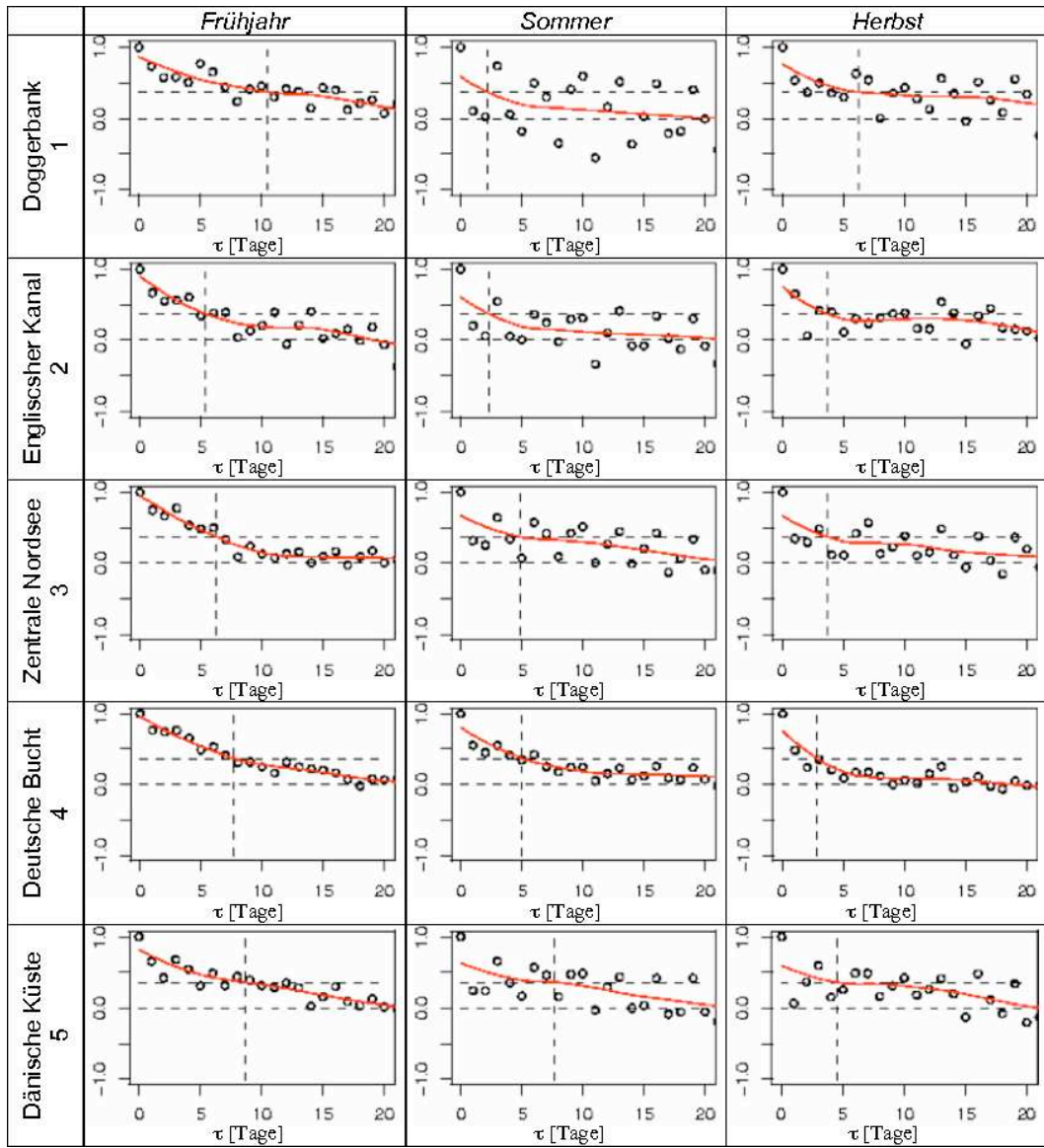


Abb. 14.12: Korrelation R in Abhängigkeit von zeitlicher Distanz τ für Chlorophyllkonzentrationen des Regionalprozessors in verschiedenen Gebieten und Jahreszeiten mit zeitlicher Auflösung von $\Delta t = 1$ Tag.

14 Räumliche oder zeitliche Autokorrelation von Chlorophylldaten

Gebiet	<i>Frühjahr</i> 1. Februar - 14. Mai	<i>Sommer</i> 15. Mai - 14. August	<i>Herbst</i> 15. August - 14. November
Doggerbank (1)	178 (27,21,14,31,37,48)	195 (34,42,26,33,40,20)	141 (35,17,26,34,29,0)
Engl. Kanal (2)	142 (24,19,12,19,29,39)	173 (28,32,27,32,35,19)	138 (31,13,27,36,31,0)
Zentr. Nordsee (3)	197 (23,25,21,35,40,53)	254 (38,54,41,44,44,33)	164 (39,20,37,34,34,0)
Deutsche Bucht (4)	185 (17,26,18,34,45,45)	225 (37,41,38,40,37,32)	151 (36,15,37,33,30,0)
Dänische Küste (5)	192 (16,30,25,35,39,47)	240 (34,47,41,39,43,36)	156 (33,16,39,35,33,0)

Tab. 14.3: Anzahl der Datentage für die Gebiete 1-5 in den Jahreszeiten Frühjahr, Sommer und Herbst mit mehr als 5% Datenüberdeckung. In Klammern findet sich die Anzahl von Datentagen nach Jahren aufgeschlüsselt, jeweils von 2003 nach 2008.

tenposition und Sonnenstand am meisten ähnelt und bei angenommener Konstanz der tatsächlichen Chlorophyllkonzentration auch das abgeleitete Chlorophyllprodukt ähnliche Absolutwerte aufweist. Die Korrelationen der übrigen Gebiete (mit Ausnahme der Deutschen Bucht, Gebiet 4) streuen im Sommer am stärksten - mit hohen Korrelationswerten für $\tau = 3, 6, 10, 13, 16, 19$ Tage. Das sind die zeitlichen Differenzen, bei denen im Mittel die geringste Differenz zwischen Beobachtungswinkeln auftritt.

Das etwa dreitägige Signal ist als Artefakt ist auch zu den übrigen Jahreszeiten zu erkennen, im Frühjahr ist es am geringsten ausgeprägt. Zu dieser Jahreszeit sind die auftretenden Chlorophyllstrukturen und ihre Dynamik stärker als der systematische Versatz. Diese Aussage gilt auch für die Deutsche Bucht, deren zeitlich stark variable Chlorophyllverteilungen im Jahr durchgehend für ein stärkeres Signal als der systematische dreitägige Versatz sorgen.

Die Korrelationszeit ist für jedes Gebiet im Frühjahr maximal, variiert aber stark zwischen den Gebieten (Abb. 14.12). Während die Chlorophyllkonzentrationen über der westlichen Doggerbank (Gebiet 1) 10.5 Tage korreliert sind, verkürzt sich diese Zeit im Englischen Kanal (Gebiet 2) auf 5.4 Tage.

Für die östliche zentrale Nordsee, die Deutsche Bucht und den Bereich vor der dänischen Küste (Gebiete 3, 4, 5) nimmt die Korrelationszeit mit der Jahreszeit ab. Da sich in der Zentralen Nordsee und vor Dänemark (Gebiet 3 und 5) im Sommer und Herbst aber deutliche Einflüsse des Artefaktes durch Beobachtungswinkel zeigen, sollte die Interpretation auf die Deutsche Bucht (Gebiet 4) beschränkt bleiben. Dort scheint die Stabilität der Strukturen mit der Jahreszeit abzunehmen.

Zusammenfassung

Die zeitliche Autokorrelation von monatlichen Chlorophyllkonzentrationskarten bot sich an, um die Nordseeregion (pixelweise) auf ganz- oder halbjährlichen Zyklen zu untersuchen. Ganzjährige Zyklen treten z.B. im Bereich der Doggerbank auf, wo die wiederkehrende Frühjahrsblüte dominiert. Weite Teile der zentralen und nördlichen Nordsee sowie der Küstenregionen folgen einem halbjährlichen Gang. Eine Ausnahme bildet die Deutsche Bucht, wo ganzjährige Zyklen ausgeprägter sind, aber auch kürzere Abfolgen dominieren können. Durch den Einfluss des systematischen Fehlers sind die jährlichen Zyklen mit Vorsicht zu interpretieren, da sie lediglich aus den erhöhten Werten im Winter entstanden sein können.

Anhand von kleinräumigen Beispielen konnte gezeigt werden, dass die rein räumliche Autokorrelation die Strukturen quantifizieren kann. Da sie nur aus den Daten einer Szene berechnet wird, ist sie weitestgehend vom systematischen Fehler unbeeinflusst. Die Korrelationslänge als Funktion der Zeit kann ein sinnvolles Maß für die Dynamik der Strukturen in den gewählten Untersuchungsgebieten darstellen. Sie kommt als ein quantitatives Maß für die Fleckenhaftigkeit (engl. patchiness) in Frage, die in ökologischen Untersuchungen vielfach diskutiert wird (vgl.). Letztlich kann die Korrelationslänge jedoch nur ein erster Schritt sein: um ökologische Fragestellungen nach der Ursache der auftretenden Strukturen zu beantworten, ist es unumgänglich, die Daten auf verschiedenen Skalen zu betrachten und zu interpretieren. Maßgebliche Einflüsse können sowohl aus kleinskaligen Prozessen stammen als auch von großskaligen Beschränkungen überlagert werden (Levin (1992)). Gleichzeitig können die beobachteten Strukturen nicht primär auf biologische Aktivität sondern auf rein physikalische (Transport-) Prozesse zurückzuführen sein. Bei der Beobachtung von ozeanischen Strömungen wird daher in Gebieten mit geringen Gradienten in der Oberflächentemperatur auf die Chlorophyllkonzentration zurückgegriffen (Gower u. a. (1980)).

Im Bereich der Doggerbank ist die Chlorophyllverteilung besonders heterogen zu Beginn der Frühjahrsblüte, während zum Sommer die Heterogenität ab- und die Korrelationslänge zunimmt. Der Jahresgang der spezifischen Strukturen läuft in allen Jahren gleich ab. Dies gilt auch für die Zentrale Nordsee. In den von thermaler Schichtung im Sommer beeinflussten Gebieten folgen die Strukturen mit geringer interannualer Varianz aufeinander.

Das Gebiet vor der Dänischen Küste befindet sich in der Randregion zur Norwegischen Rinne und wird immer wieder von deren Wassermasse beeinflusst. Dadurch können in kurzen zeitlichen Intervallen starke Fronten auftreten oder verschwinden. Ohne Berücksichtigung der Strömungsverhältnisse ist die Korrelationslänge für dieses Gebiet nicht zu interpretieren.

Für die stark heterogene Chlorophyllverteilung der Deutschen Bucht reicht dieses ein-

fache Verfahren ebenfalls nicht aus, um die Dynamik der Strukturen zu analysieren. Ein multivariater Ansatz zur Interpretation steht noch aus (basierend auf Seuront und Lagadeuc (1998)).

Die Methode gibt keine Auskunft darüber, auf welchem Konzentrationsniveau die homogenen Verteilungen zu finden sind, d.h. ob die Homogenität z.B. durch das Sommerminimum oder ein räumlich ausgedehnte Planktonblüte zustande kommt.

Die Dekorrelationszeit wird sich erst nach Bereinigung des systematischen Fehlers interpretieren lassen.

Teil V

Zusammenfassung und Ausblick

15 Zusammenfassung und Ausblick

15.1 Zusammenfassung

Anhand von Satellitendaten habe ich innerhalb dieser Arbeit den räumlich-zeitliche Verlauf der Phytoplanktonverteilung in der Nordsee analysiert. Es war meine Zielsetzung, quantitative Maße zur Beschreibung der Strukturen und Dynamik zu verwenden. Dazu war es notwendig, schrittweise die methodische Grundlage für eine solche Analyse zu schaffen und die Daten entsprechend zu bearbeiten:

1. Die Daten sind bezüglich zufallsverteilter und systematischer Fehler analysiert worden.
2. Ich habe verschiedene geostatistische Verfahren getestet, um, z.B. aufgrund von Wolken, fehlende Datenpunkte möglichst exakt abschätzen zu können. Viele Analyseverfahren benötigen vollständige Datensätze.
3. Die Daten wurden hinsichtlich der zeitlich-räumlichen Entwicklung analysiert und ausgewertet bzw. die Anwendbarkeit der Methoden trotz des systematischen Fehlers oder der eingefügten Schätzwerte überprüft.

Diese Untersuchungen basieren auf Karten der Chlorophyllkonzentration des abbildenden Spektrometers MERIS (Umweltsatellit ENVISAT). Gegenüber *in situ* Beprobungen zeichnet sich dieser Datensatz durch seine (relativ) hohe räumliche Auflösung bei annähernder Gleichzeitigkeit der Messung für ein Gebiet mit der Ausdehnung der Nordsee aus. Der Strukturreichtum dieses dynamischen Lebensraumes kann in detailreichen Momentaufnahmen sehr gut erfasst werden. Erst mit Hilfe von Satellitendaten lassen sich Strukturanalysen durchführen, da die auf schiffgestützten Einzelmessungen beruhenden Verteilungskarten große Lücken aufweisen und die Daten räumlich und zeitlich nicht zusammenpassen.

Der Datensatz erstreckt sich hier über einen Zeitraum von sechs Jahren (Januar 2003 bis August 2008) und beinhaltet sowohl das Chlorophyllprodukt des Standardprozessors als auch eines Regionalprozessors, der speziell für die Gewässer der Nordsee entwickelt worden ist. Die betrachteten Auswertungsverfahren und Produkte sind für den Einsatz in case 2-Wasser gedacht.

Es wurde mit sehr unterschiedlichen zeitlichen Auflösungen experimentiert, um das Problem der mangelnden räumlichen Überdeckung zu umgehen und zu äquidistanten Zeitserien zu gelangen. Durch eine Methode zur Frequenzanalyse unregelmäßiger Zeitreihen konnte dieser Ansatz mit seiner reduzierten zeitlichen Auflösung umgangen und die höchstmögliche Auflösung verwendet werden. Diese Analyse bot mit dem Signal der Periode von 3.18 Tagen den quantitativen Hinweis auf einen systematischen Fehler in den

Chlorophyllprodukten, der bereits in der qualitativen, visuellen Auswertung erkennbar war.

15.1.1 Fehler innerhalb des Chlorophyllproduktes

Die Charakterisierung und Abschätzung der Fehler des Chlorophyllproduktes bildete den ersten Schwerpunkt.

Es gibt drei Auffälligkeiten im Datensatz, die über die Sonnenstandswinkel in Beziehung zueinander stehen (vgl. Anhang, Kapitel A und Müller u. a. (2010)):

1. In den Zeitreihen fällt ein starker Trend zu hohen Chlorophyllkonzentrationen im Winter auf, der in erster Linie mit dem Sonnenzenitwinkel korreliert ist.
2. Bei aufeinander folgenden Beobachtungstagen zeigt sich im gleichen Gebiet von Szene zu Szene ein Versatz in der Chlorophyllkonzentration.
3. Es lässt sich ein Trend ansteigender Konzentration von (quasi) West nach Ost entlang einer Scanzeile feststellen.

Die Suche nach ihrer Ursache (bzw. Ursachen) wurde durch diese Arbeit angeregt und erste Ergebnisse gefunden. Die letzte Korrektur ist bis heute nicht vollständig abgeschlossen.

Bereits die Strahldichten über Wasser (Level 2, water leaving radiance reflectance, vgl. Abb. A.5) zeigen die Drift entlang einer Scanzeile. Da dieses Zwischenprodukt beim Regionalprozessor vor der Ableitung der Wasserinhaltsstoffe entsteht, wird die Quelle mit großer Wahrscheinlichkeit in der Atmosphärenkorrektur zu finden sein. Diese ist für Gewässer mit sehr hohen Konzentrationen von Substanzen, die im blau-grünen Spektralbereich die Sonnenstrahlung absorbieren (Chlorophyll und andere organische Substanz), besonders schwierig und stößt dort an ihre Grenzen. Die bestehende Korrelation zwischen Chlorophyllkonzentration (aller Produkte des Standardprozessors wie des Regionalprozessors (Müller u. a. (2010))) legt erst einmal eine einfache Ursache nahe: im Winter mit den niedrigen Sonnenständen beschreibt die Näherung einer planaren Atmosphäre die reale sphärische Geometrie nicht mehr gut. Ab einem Sonnenzenitwinkel von etwa 59° Grad beträgt die Abweichung zwischen den Weglängen in den unterschiedlichen Modellen bei einer Standardatmosphäre von 50 km Höhe etwa 1 % (bei 75° sind es 5 %). Bei dem geringen aus dem Wasser stammenden Signalanteil (<10%) sind diese Abweichungen in der Weglänge innerhalb der Atmosphäre nicht zu vernachlässigen und schlagen sich damit in den Produkten nieder.

Der Versatz zwischen aufeinander folgenden Tagen lässt sich auf diese Art nicht erklären. Gerade bei der Analyse der Daten aus den Sommermonaten mit niedriger und annähernd konstanter Chlorophyllkonzentration wird der Versatz besonders augenscheinlich. Es gibt weitere Ideen, wie die Definition der Geometrie des Streuprozesses ein derartiges Artefakt verursachen könnte. Dieser Ansatz befindet sich allerdings noch in der Erprobung.

Der tägliche Versatz innerhalb einer Region und der Trend über eine Scanzeile können vermutlich auf die gleiche Ursache zurückgeführt werden. Allein die Drift entlang

der Scanzeile reicht aus, um eine Zeitserie mit überlagertem Signal von 3.18 Tagen zu produzieren.

Die Genauigkeit des Datensatzes ist hauptsächlich durch diesen systematischen Fehler beschränkt.

In der Abfolge von Chlorophyllkarten treten Konzentrationsunterschiede von ca. 30% für einen festen geographischen Punkt auf. Diese können nicht durch Vorgänge in der Nordsee, wie den Gezeitenzyklus oder den Strömungstransport, nicht erklärt werden.

Als Hilfsvariable zur Demonstration der Fehlerabhängigkeit wurde in dieser Arbeit hauptsächlich auf den Sonnenazimutwinkel zurückgegriffen, der durch seine lokale Definition sowohl den Jahresgang als auch starke Variabilität im Sommer zeigt. Er bildet damit alle drei oben benannten Phänomene ab - auch wenn es keinen kausalen Zusammenhang geben muss.

Zusätzlich sorgen im Standardprodukt die unterschiedlichen Kameras, aus denen das Spektrometer zusammengesetzt ist, für den größten Anteil des Fehlers. Die zweite Kamera von Westen (Kamera 4) liefert Radianzspektren, die zu deutlich erhöhten Chlorophyllkonzentrationen führen. Bei Zeitreihen, die auf Daten mit enger räumlicher Begrenzung beruhen, kann dies zur Diskontinuität des Verlaufes führen.

Der Chlorophylldatensatz wurde ohne Bereinigung des Fehlers verwendet. Die Prozesse der Fehlerentstehung sind bis dato nicht vollständig geklärt. Da die Beobachtungs- und Sonnenstandsgeometrie für jeden Datenpunkt variiert, wird eine vollständige Korrektur aller Pixel notwendig sein. Statt mit aufwendigen empirischen Korrekturen zu arbeiten, wird eine Neuprozessierung unter Berücksichtigung der teilweise noch zu identifizierenden Fehlerquellen die sinnvollere Alternative sein.

Bei allen Untersuchungen, die zeitliche Komponenten enthalten, ist daher abhängig von der zeitlichen Auflösung mit spezifischen Artefakten zu rechnen:

- Bei zeitlicher Auflösung von einem Tag wird das Chlorophyllsignal an einem festen geographischen Ort von einem (annähernden) dreitägigen Rhythmus dominiert, der durch die diskreten Orbits und die relativ zur Szene wechselnde Lage des Ortes in Kombination mit der Drift zustande kommt.
- In Monats- und Jahreszeitenmitteln des Chlorophyllproduktes lassen sich die täglichen Schwankungen gut glätten. Der Jahresgang des Sonnenstandes wirkt sich aber vollständig auf das zusammengefügte Produkt aus und sorgt entgegen der Erwartung für sehr hohe Konzentrationen im Winter. Daher müssen alle Szenen mit einem Sonnenzenitwinkel $> 70^\circ$ von der Zeitreihenanalyse vorerst ausgeschlossen werden.

Beschränkt man sich auf einen isolierten Zeitpunkt, so können die Chlorophyllkarten detaillierte Auskunft über räumliche Strukturen geben, die mit punktuellen Messungen nicht aufzulösen wären.

Unter der Annahme, der tatsächliche mittlere Chlorophyllgehalt der betrachteten Wassersäule liegt im Intervall der durch den Sonnenstand beeinflussten Absolutwerte des Satellitenchlorophyllproduktes, ist bei diesen mit einem Fehler von mindestens 30% zu rechnen. Dies entspricht dem mittleren Fehler, der für case 1-Wasser Chlorophyllprodukte

angegeben wird. Gleichzeitig ist es eine Abschätzung der minimalen Größe des Fehlerintervalls für jeden einzelnen Datenpunkt des Chlorophyllproduktes. Die Abschätzung basiert auf Daten des Sommers, wenn eine annähernd konstant niedrige Chlorophyllkonzentration angenommen werden kann. Entsprechend treten bei Auswerteverfahren, die das Chlorophyllprodukt in Regionen mit Phytoplanktonminimum im Sommer betrachten, zu dieser Jahreszeit die größten Artefakte auf, da bei niedrigen Absolutwerten der Konzentration die größten relativen Fehler durch den Versatz entstehen.

15.1.2 Geostatistische Schätzwertverfahren

Viele klassische Methoden der Datenanalyse benötigen einen in Bezug auf Raum und/oder Zeit vollständigen Datensatz. Daher wurde der Umgang mit fehlenden Datenpunkten, die in der Satellitenozeanographie in erster Linie durch Wolkenüberdeckung entstehen, zum zweiten Schwerpunkt meiner Arbeit.

Hauptsächlich habe ich verschiedene Krigingverfahren und ihr Zusammenspiel mit der räumlichen Verteilung von Datenlücken, der Aufbereitung der Zielvariable mittels (räumlichen) Trendkorrekturen und dem Einfluss der Variogrammrealisation auf die Güte der Schätzwerte betrachtet.

Gute Ergebnisse liefert bereits die einfache lokale Ordinary Kriging Methode, wenn nur Datenpunkte aus der nächsten Nachbarschaft um die Schätzposition in deren Schätzwert einfließen. Neben der verkürzten Rechenzeit zeigt sich im Nahbereich keine räumliche Anisotropie, sodass das richtungsabhängige Variogramm gut auf ein omnidirektionales reduziert werden kann.

Für jede lokale Nachbarschaft bzw. Punktpaare mit geringem Abstand gilt des Weiteren die Annahme des konstanten Erwartungswertes bereits für die (logarithmierte) Chlorophyllkonzentration, sodass die mathematische Voraussetzung für die Krigingmethoden erfüllt ist.

Die Untersuchung zeigt, dass komplexere Methoden, die mehr Informationen berücksichtigen, nicht zwangsläufig zu genaueren Schätzwerten führen. So schneidet z.B. das Co-Kriging bei Verwendung der Wassertiefe nicht wesentlich besser ab als das Ordinary Kriging der logarithmierten Konzentrationen, obwohl die Tiefe mit der Chlorophyllkonzentration in den Küstenbereichen der Nordsee stark korreliert ist.

Auch eine räumliche Trendkorrektur ist nur im Falle von sehr großen Datenlücken erforderlich.

Für eine Trendkorrektur schlage ich verschiedene Monatsmittel vor. Auf diese Weise ist es möglich, sowohl die zeitliche Information in gewissen Umfang zu nutzen, als auch die teilweise stark ausgeprägten Muster in der Chlorophyllverteilung zu berücksichtigen. Der Rekonstruktionsprozess modelliert dann nur die Abweichungen vom Monatsmittel.

Erfahrungsgemäß reichen die Daten aus einem Monat (abgesehen von den Wintermonaten mit häufiger Wolkenbedeckung) aus, um das Gebiet der Nordsee vollständig abzudecken.

Je nach Schwerpunkt der Fragestellung können unterschiedliche Mittelwerte zum Einsatz kommen: der arithmetische Mittelwert ist sensitiv für die zeitlichen Maxima und Minima und kann bei Untersuchungen zu Blütenstrukturen sinnvoll sein. Dagegen be-

sitzt der Median die größtmögliche Unempfindlichkeit gegen Ausreißer und kann gut die mittlere Verteilung über einen längeren Zeitraum repräsentieren. Diese Maße werden für die logarithmierte Chlorophyllkonzentration berechnet, da diese eher normalverteilt ist. Unempfindlich ist auch der Mittelwert größter Wahrscheinlichkeit (Maximum Likelihood Mean).

Theoretisch kann durch zeitliche Gewichtung der Daten die größtmögliche Ähnlichkeit zu einem Bezugstag hergestellt werden. Durch den systematischen, zeitlichen Fehler gilt diese Annahme für diesen Datensatz allerdings im allgemeinen nicht und konnte nicht experimentell belegt bzw. widerlegt werden.

Bei großen Datensätzen, wie einem Satellitenbild, kann das Variogramm nur aus einer beschränkten Auswahl an Datenpunkten berechnet werden.

Die Schätzwerte und damit die Güte der Rekonstruktion variieren deutlich mit den unterschiedlichen Variogrammrealisationen. Die Güte wird charakterisiert durch den mittleren absoluten Fehler als lokale Fehlergröße und den Absolutwert des Medians des Fehlers als globale Größe. Die Schätzwerte hängen selbst bei einer einzigen Krigingmethode von der Auswahl der Datenpunkte ab, die für die Charakterisierung des räumlichen Zusammenhanges gewählt worden sind. Gleichzeitig bestimmt die räumliche Struktur des zu schätzenden Bereiches bzw. der zur Verfügung stehenden Daten die Güte der Schätzwerte. Große, zusammenhängende Datenlücken können immer am ungenauesten rekonstruiert werden.

Die beste Kombination von Trendkorrektur und Schätzwertverfahren hängt von der Datenauswahl zur Variogrammrealisation und der räumlichen Struktur der Zielvariable im zu schätzenden Bereich ab. Je nach räumlicher Verteilung der fehlenden Datenpunkte und der Stärke der dort zu erwartenden Variabilität können ganz unterschiedliche Verfahrenskombinationen sich als die jeweils besten erweisen. Eine eindeutige Empfehlung lässt an dieser Stelle nicht aussprechen.

Als zwei mögliche Maße zur Quantifizierung von Wolkenstrukturen habe ich die Distanzklassenverteilung, d.h. die Statistik der Distanz zwischen Punktpaaren, und die Lückenhaftigkeit (hier "lacunarity") verwendet. Die Kombination beider Maße kann eine räumliche Struktur recht eindeutig beschreiben und könnte als Grundlage dienen, eine Beziehung zur erwarteten Güte der Rekonstruktion zu formulieren. Beim derzeitigen Untersuchungsstand ist es allerdings fraglich, ob eine solche Beziehung tatsächlich allgemeingültiger Natur wäre oder nicht vielmehr einzig für den hier betrachteten Untersuchungsraum mit seinen sehr speziellen Verteilungsmustern Gültigkeit hätte. Da die Güte der Rekonstruktion auch von der räumlichen Variabilität in der Umgebung der Schätzposition abhängt, die sich durch die Varianz des zeitlichen Mittelwertes an dieser Stelle abschätzen ließe, ist die lokale Gültigkeit wesentlich wahrscheinlicher. Ohne einen optimalen Datensatz, d.h. einen ohne systematische Fehler, ist die Formulierung der Abhängigkeit von Schätzposition, Wolkenstruktur und zum besten Ergebnis führenden Kombination von Krigingmethode und Trendkorrektur nicht abschließend möglich.

Im Gegensatz zum Ordinary Kriging, bei dem die Fehlervarianz lokal minimiert wird, beruht die auf der Hauptkomponentenanalyse basierende iterative Dateninterpolationsmethode DINEOF darauf, die globale Varianz in Raum *und* Zeit zu minimieren. Dabei nimmt sie jedoch keine Rücksicht auf räumliche Abhängigkeiten in den Daten. Durch

den systematischen Fehler führt diese Methode zu ungenaueren Schätzwerten. Sie hat allerdings den Vorteil, sehr schnell alle fehlenden Datenpunkte im ganzen betrachteten Datensatz rekonstruieren zu können.

15.1.3 Ergebnisse der Analyseverfahren

Der jahreszeitliche Verlauf der Phytoplanktonverteilung in der Nordsee, wie er aus *in situ* Beobachtungen und Simulationen in der Literatur beschrieben wird, lässt sich in den Satellitendaten wiederfinden.

Wie zu erwarten gibt es große regionale Unterschiede, wann Frühjahrs- oder Herbstblüten einsetzen bzw. ob sich Konzentrationsminima im Sommer ergeben. Die auffälligsten Blütenstrukturen entstehen ab Mitte Februar in der Norwegischen Rinne und von März bis April über der Doggerbank und lassen sich in ihrer Entwicklung gut verfolgen. Mit den Analysemethoden wurden nur die Strukturen der Verteilung untersucht, nicht jedoch ihre Ursachen (Licht, Schichtung, Nährstoffangebot, Fraß durch Zooplankton).

Im Bereich der Doggerbank lässt sich durch die geringe interannuale Variabilität aus der Information aller Jahre eine typische Dynamik ableiten: nach dem Beginn der Blüte in den zusammenhängenden Flachwasserbereichen im März, zieht sie sich in die Flankenbereiche der Doggerbank hinaus, um dort zum Sommerminimum (ab Mitte Juni) zu vergehen. Diese qualitative Analyse lässt sich mithilfe der Korrelationslänge gut quantifizieren.

Die Korrelationslänge ist ein Maß für die mittlere Ausdehnung räumlicher Strukturen und gibt den Distanzbereich an, über den die Chlorophyllkonzentration ähnlich bleibt. Damit kommt sie als Maßzahl für den von Ökologen oftmals verwendeten Begriff der Lückenhaftigkeit oder "Fleckenhaftigkeit" (hier ist "patchiness" gemeint) in Frage.

Als Funktion der Zeit kann sie die Dynamik der Strukturen in den gewählten Untersuchungsgebieten veranschaulichen. Auch eine Abschätzung der natürlichen interannualen Variabilität ist auf diese Weise möglich: während die Strukturentwicklungen in der zentralen südlichen Nordsee und im Bereich der Doggerbank sich kaum von Jahr zu Jahr unterscheiden, lassen sich in der Deutschen Bucht starke interannuale Schwankungen feststellen. In den zuvor genannten Gebieten mit ihrem Sommerminimum in der Chlorophyllkonzentration entwickelt sich die Korrelationslänge immer von kleinteiligeren, heterogenen Konzentrationsmustern zu eher homogenen Verteilungen im Sommer.

Die Methode gibt keine Auskunft darüber, auf welchem Konzentrationsniveau die homogenen Verteilungen zu finden sind, d.h. ob die Homogenität z.B. durch das Sommerminimum oder ein räumlich ausgedehnte Planktonblüte zustande kommt.

Da die Korrelationslänge jeweils nur aus den Daten einer Szene berechnet wird, ist sie kaum störungsempfindlich gegenüber dem systematischen Fehler.

Die Interpretation der beiden zeitlichen Autokorrelationsexperimente muss dagegen diesen Einfluss berücksichtigen.

Bei der zeitlichen Autokorrelation von Monatsmittel-Karten wurde jeweils pixelweise die Periode des vorherrschenden Signals bestimmt. Es konnte eine Karte geschaffen werden, in der die Längen der vorherrschenden Zyklen ausgewiesen werden. Hauptsächlich treten ganz- und halbjährliche Zyklen auf.

Ganzjährliche Zyklen finden sich z.B. im Bereich der Doggerbank. Weite Teile der zentralen und nördlichen Nordsee sowie der Küstenregionen folgen eher einem halbjährlichen Gang. Eine Ausnahme bildet die Deutsche Bucht, wo ganzjährige Zyklen ausgeprägter sind. Die Interpretation kann in dieser Stelle nicht viel weiter als dieses phänomenologische Urteil gehen, da aus den Zeitreihen bekannt ist, dass ein starkes Jahressignal mit hohen Konzentrationen im Winter das tatsächliche Jahressignal von wiederkehrenden Frühjahrsblüten überlagert. Das Maximum der Korrelation bei einer Verschiebung von einem Jahr kann also hauptsächlich ein Artefakt des systematischen Fehlers sein.

Wiederholt man dieses Experiment für kleinere Gebiete und mit Daten von täglicher Auflösung, so lässt sich die Dekorrelationszeit ableiten. Diese ist ein Maß dafür, wie lange räumliche Strukturen in einem Gebiet erhalten bleiben. Trotz des starken Einflusses des Versatzes, der als Artefakt für eine regelmäßig hohe Korrelation nach 3 Tagen sorgt, können mit dieser Methode Gebiete wie die Deutsche Bucht analysiert werden. Durch ihre stete Heterogenität macht sich das Artefakt nur wenig bemerkbar. Das beschriebene Zeitmaß nimmt hier vom Frühjahr zum Herbst ab.

15.2 Ausblick

Die Fragestellungen zur Quantifizierung zeitlicher Prozesse werden erst dann belastbar zu beantworten sein, wenn der Datensatz von systematischen Fehlern befreit wurde. Für ökologische Fragestellungen, die auf Zeitreihenanalysen hinauslaufen, ist der Datensatz damit nur unter Vorbehalt geeignet.

Mit dem bereinigten Datensatz wird sicherlich auch eine vergleichende Studien zum Einsatz der Frühjahrsblüte möglich werden. In Analogie zu den Modellstudien an Bojendaten wird sich die Untersuchung der Abhängigkeit vom Spring- und Nipptidenzyklus auf flächige Dimensionen erweitern lassen (vgl. Sharples (2008)). Auch die Assimilation in ein Ökosystemmodell wird Einblick in die kausalen Zusammenhänge der beobachteten Dynamik geben können. Zur Zeit werden Satellitendaten bereits zur Validierung von Ökosystemmodellen genutzt (z.B. Tiedje u. a. (2010)).

Die Vergleichbarkeit mit *in situ* Verfahren wie dem Durchflussfluorometer in der Ferrybox (Petersen u. a. (2008)) ist nur in begrenztem Maße gegeben. Während der qualitative Verlauf sich nach Berücksichtigung des Strömungstransportes gut in Einklang bringen lässt, können große quantitative Unterschiede auftreten. Ob diese durch die Wahl der optischen Methode (Fluoreszenz- versus Absorptionsmessung), den Unterschied im betrachteten Wasserkörper (Probennahme in konstant 4 m Tiefe oder variable Eindringtiefe des Lichtes in Abhängigkeit von Inhaltsstoffen) oder den systematischen Fehler im Chlorophyllprodukt zustande kommen, kann nicht auf einfache Weise entschieden werden. Im letzten Fall ist eine Abweichung bei Profilen in Ost-West-Richtung bzw. bei Profilen in Nord-Süd-Richtung im Bereich höherer Breiten mit ihren niedrigen Sonnenstandswinkeln zu erwarten. Ein Beispiel mit letzterem Fall zeigte deutlich die mit nördlicher Breite (zwischen 65-72°N) ansteigende Diskrepanz zwischen erhöhter Chlorophyllkonzentration gemäß MERIS und wesentlich konstanterer Fluorometermessung (Folkestad und Sørensen 2010).

Eine eingehendere Untersuchung des Einflusses der Vertikalverteilung des Phytoplanktons auf die vom Satelliten gesehene Konzentration steht ebenfalls noch aus. Mit den Fluoreszenzprofilen, die ab 2009 mit dem so genannten "Scanfish" im Rahmen des Projektes COSYNA der GKSS aufgezeichnet wurden, werden sich entsprechende Studien verwirklichen lassen.

Sicherlich ist es sinnvoll, bei Analysen in Raum und Zeit nicht auf Datentage zu verzichten, die mit wenigen geostatistischen Schätzwerten vervollständigt werden können. Wie viel Prozent noch als wenig aufgefasst werden können, hängt dabei stark von der persönlichen Interpretation des Bearbeiters ab. Aber auch die Lage der Schätzpositionen kann ausschlaggebend sein, da von der zu erwartenden Variabilität der Konzentration an der Position auch die Genauigkeit unter Verwendung verschiedener Schätzwertmethoden abhängen wird.

Die regional zu erwartende räumliche (und zeitliche) Variabilität ist z.B. durch die Varianz der Jahreszeiten- oder Monatsmittel erfassbar. Sie könnte als wertvolle Entscheidungshilfe dienen, in welchen Bereichen (oder auch pixelweise) bzw. zu welchen (Jahres-) Zeiten welche Methodenkombination zur Schätzwertbildung am sinnvollsten ist. Jedoch ist bis jetzt weder dieser Zusammenhang noch die daran anknüpfende Hierarchisierung der verschiedenen Verfahren systematisch untersucht worden. Unter letzterem verstehe ich ein vereinfachtes sequentielles Kriging, bei dem Gruppen von Schätzpositionen mit verschiedene Methoden bearbeitet werden, anstatt eines tatsächlichem punktweise sequentiellen Vorgehens (Vargas-Guzman und Yeh (1999)). Die Schätzwerte könnten z.B. erst in direkter Nachbarschaft zu Datenpunkten bzw. in sehr homogenen Gebieten berechnet werden und danach unter Verwendung aller Punkte in die heterogeneren Gebiete ausgedehnt werden.

Bei der Wahl des Verfahrens zur Einführung von Schätzwerten muss derzeit die Kenntnis über den systematischen Fehler berücksichtigt werden: Verfahren, die auf der Analyse der zeitlichen Information beruhen, können nur im Rahmen des systematischen Fehlers genau schätzen. Sie werden von den beschriebenen Artefakten durch den Versatz des Chlorophyllkonzentrationsniveaus beeinflusst sein.

Moderne Schätzwertverfahren, die mit Hilfe der Wavelet- oder Fouriertransformation (z.B. Demyanov und Soltani (2001); Fuentes (2001)) ihre Rekonstruktion räumlicher Zusammenhänge skalenabhängig durchführen, oder solche, die räumliche *und* zeitliche Abhängigkeit gleichermaßen beachten, können erst nach der Neuberechnung des Datensatzes sinnvoll getestet werden. Die Erfahrung mit komplexeren Verfahren wie dem Cokriging oder dem Universalen Kriging lässt jedoch keine entscheidende Verbesserung erwarten.

Auch die Frage, ob die Residuen aus einem zeitlich gewichteten Monatsmittel eine exaktere Rekonstruktion des Bezugstages ermöglichen, wird erst an dem bereinigten Datensatz zu überprüfen sein.

Im Rahmen der quantitativen Auswertung der Chlorophyllkonzentrationsverteilung wird die Autokorrelationsanalyse zum Vergleich mit einem rein physikalischen Parameter wie der Sichttiefe (hier angenähert als Satellitenprodukt z_{90}) wiederholt werden. Dadurch ließen sich die rein biologischen von den überlagerten physikalischen Prozessen trennen. Es wäre ersichtlich, wie stark z.B. die Korrelationslänge tatsächlich die biolo-

gische Dynamik der Phytoplanktonverteilung widerspiegelt, oder ob sie doch eher von Mustern aus Transportprozessen dominiert wird.

Auch weitere Maße zur quantitativen Beschreibung der Strukturen stehen noch zur Diskussion.

So kann die Definition der Lückenhaftigkeit z.B. auf einen kontinuierlichen Wertebereich erweitert werden und dann als skalenabhängiges Strukturmaß dienen. Dergleichen findet in der Analyse terrestrischer Datensätze zur Vegetationsklassifizierung oder in der Stadtentwicklung bereits Anwendung (Robertson (1994); Alves und Filho (2001); Greenhill u. a. (2002)) und sollte auf Chlorophyllkarten übertragbar sein. Ähnlich wie bei der Korrelationslänge sollte diese Analyse auch bei unkorrigiertem Datensatz möglich sein.

Für die weiterführenden Größen in der Autokorrelationsanalyse, wie die skalenabhängige Dekorrelationszeit oder Kohärenzlänge (vgl. Denman und Abbott (1994); Abbott und Letelier (1998)), ist die Korrektur unumgänglich.

Diese Arbeit konnte an vielen Stellen lediglich das Potential des Chlorophylldatensatzes aufzeigen und zu vorläufigen Ergebnissen gelangen. Doch allein der Anstoß und die Aussicht auf einen überarbeiteten Datensatz rechtfertigt in meinen Augen die hier zusammengetragene Dokumentation von Methoden, Versuchen und Ideen. Die Überarbeitung wird im Rahmen des kommenden Klimaprojektes der ESA (Climate Change Initiative) große Relevanz besitzen, da nur so die angestrebten strengen Gütekriterien für eine Klimavariablen Ozeanfarbe bzw. Chlorophyllkonzentration erfüllt werden können.

Anhang

Programme und Hilfsmittel

Diese Arbeit ist fast vollständig auf einem Macintosh iBook G4 (1.2 GHz) entstanden.

Es wurde fast ausschließlich mit frei verfügbarer Software gearbeitet. Neben den beiden unten aufgeführten Programmen wurden noch verwendet: *Eclipse* - eine Entwicklungsumgebung für Java-Programme zum Auslesen und Schreiben der Satellitendaten in ein für die Statistiksoftware lesbares Format.

Dieser Text wurde mit Hilfe von *LyX*, einer graphischen Oberfläche für LaTeX, gesetzt.

BeamVisat

Die Software *BeamVisat* (Brockmann-Consult) ist ein Werkzeug zur Bearbeitung der Satellitendaten der Instrumente auf Envisat. Neben der Verarbeitung der Daten innerhalb der Umgebung mit graphischer Oberfläche, ist es auch möglich, auf die Funktionalität der zugrunde liegenden Java-Bibliotheken direkt zurückzugreifen und z.B. das Ausschreiben und Aussortieren der Satellitenszenen zu automatisieren. Die Datenbasis wurde auf diese Weise erstellt und liegt, um die weitere Bearbeitung zu vereinfachen, als Sammlung von Textdateien für jeden einzelnen Datentag vor.

R

R (R-Development-Core-Team (2007)) ist die OpenSource-Variante der Statistiksoftware und Programmiersprache *S*. Insbesondere unter Mac OS X besitzt *R* enorme Benutzerfreundlichkeit (Skripteditor mit farblicher Hervorhebung der Syntax, Anzeige der Parameter einer Funktion, Anwahl von Funktionen innerhalb eines Skriptes direkt über ein Menü, etc.) und höhere Stabilität in längeren Rechnungen mit größeren Datenmengen.

Die stetig wachsende Funktionalität von *R* liegt in den Bibliotheken begründet. Für geostatistische Bearbeitungen lässt sich auf die Funktionen aus *geoR* (Ribeiro und Diggle (2001)) oder *gstat* (Pebesma und Wesseling (1998); Pebesma (2004)) zurückgreifen. Die meisten Karten wurden mit *fields* (Nychka (2007)) erstellt.

A Weiterführende Untersuchungen zum systematischen Fehler

Die Genauigkeit des Datensatzes hat sich hauptsächlich als durch einen systematischen Fehler beschränkt erwiesen. Die Analyse der Zeitserie der mit dem c2r-Algorithmus abgeleiteten Chlorophyllkonzentrationen für einen festen Ort in der Nordsee zeigt ein regelmäßiges zeitliches Muster mit einer Periode von 3.18 Tagen. Hierbei treten Konzentrationsunterschiede von ca. 30% auf, die durch Vorgänge in der Nordsee, wie den Gezeitenzyklus, nicht erklärt werden können. Bei der Suche nach den Ursachen ergibt sich z.B. eine signifikante Korrelation mit dem Azimutwinkel des Sonnenstandes.

Die drei auffälligen Ausprägungen des systematischen Fehlers (d.h. hohe Chlorophyllkonzentration im Winter, Drift entlang einer Scanzeile und Signal mit Periode von 3.18 Tagen) können mit dem Sonnenstand in Beziehung gesetzt werden (Müller u. a. (2010)). Diese Übersicht identifiziert potentielle Quellen in der Atmosphärenkorrektur, ohne auf die physikalischen Prozesse und ihre Modellierung einzugehen.

Verschiedene Ansätze, die im Rahmen der Betrachtung zur Geometriekorrektur ausprobiert wurden, sind hier kurz dokumentiert.

A.1 Korrektur der Beobachtungsgeometrie - 1

Diese Versuche beruhen auf der Annahme, daß die im MERIS-Produkt angegebenen Winkel die Erdkrümmung nicht berücksichtigen. Entsprechende Korrekturen werden hier am Beobachtungszenithwinkel durchgeführt.

Weitere Versuche direkt auf Ebene der Prozessierung der Level 1-Daten haben am Ende eine Abhängigkeit vom Beobachtungszenitwinkel ergeben. Der Azimut-Beobachtungswinkel (relativ zur Sonne, d.h. die Differenz zwischen Beobachtungs- und Sonnenazimut) $\Delta\phi = \phi_{view} - \phi_{sun}$ und Zenit-Beobachtungswinkel θ wird auf Kartesische Koordinaten umgerechnet, um Artefakte nahe des Nadirs zu vermeiden:

$$\begin{aligned}x &= -\sin\theta \cdot \cos\Delta\phi \\y &= \sin\theta \cdot \sin\Delta\phi \\z &= \cos\theta\end{aligned}$$

Diese kartesischen Koordinaten sind vom neuronalen Netz (d.h. vom Prozessor) benötigte Parameter zur Atmosphärenkorrektur. Die x-Koordinate hat dabei über eine Bildzeile einen leicht asymmetrischen sigmoidalen Verlauf (vgl. Abb. A.3). Ihre Amplitude an beiden Enden hängt naturgemäß vom Beobachtungsnadirwinkel ab. Da sich der Sonnenazimutwinkel über eine Scanzeile ebenfalls, wenn auch nur im Bereich von etwa

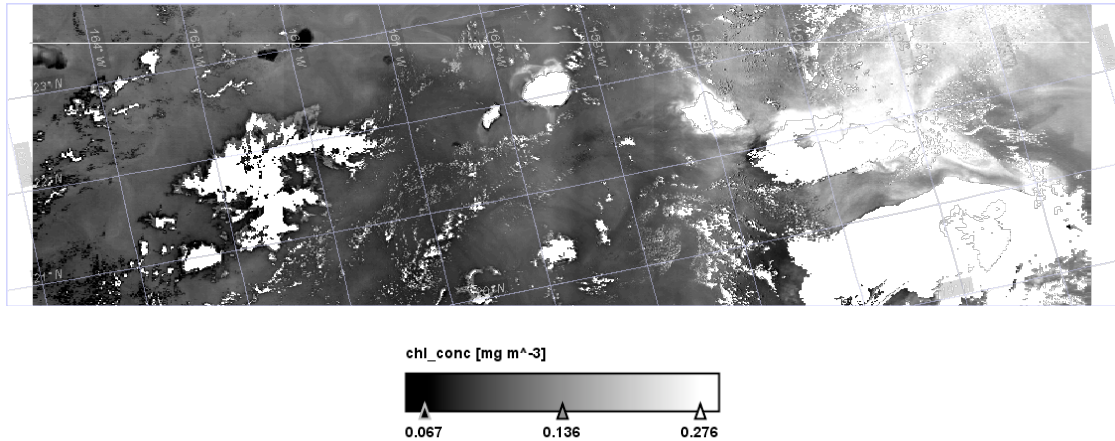


Abb. A.1: Chlorophyllkonzentration im pazifischen Ozean vom 6. April 2004. Die Transekte, an der der Einfluß des Beobachtungszenitwinkels getestet wird, ist als weiße Linie markiert.

15 Grad ändert, ergibt sich eine hohe Korrelation zwischen dem Fehler und dem Sonnenazimutwinkel.

Die Analysen basieren derzeit auf einer Szene aus dem pazifischen Ozean, wo von einer homogenen Chlorophyllkonzentration über eine Bildzeile ausgegangen werden kann (Abb. A.1). Die hohe Korrelation zwischen Reflexion im blauen Spektralbereich bzw. der Chlorophyllkonzentration und der x-Koordinate (bzw. dem Azimutwinkel der Sonne) findet sich hier leicht bestätigt (vgl. in Abb. A.4 jeweils die Verläufe ohne Korrektur).

Wird der Beobachtungszenitwinkel mit einem von ihm abhängigen Faktor geändert und damit die Amplitude der x-Koordinate, kann eine konstante Reflexion und Chlorophyllkonzentration entlang der Scanzeile (Abb. A.4b) erreicht werden. Die Änderung des Beobachtungszenitwinkels, die notwendig ist, um eine konstante Verteilung entlang einer Scanzeile zu erreichen, entspricht dem Unterschied des Beobachtungswinkels am Satelliten und dem am Boden nach Einrechnung der Erdkrümmung. Allerdings sollen die in den MERIS-Daten angegebenen Winkel die Erdkrümmung bereits berücksichtigen.

Die gezeigten Korrekturen definieren sich wie folgt:

- Korrektur 1: $\theta_{view,1} = 1.1364 \cdot \theta_{view}$
- Korrektur 2: $\theta_{view,2} = \theta_{view} + ||\text{Pixel_X} - 560|| \cdot 0.0093247 - 0.004793$
- Korrektur 3: $\theta_{view,3} = 1.18 \cdot \theta_{view}$
- Korrektur 4: $\theta_{view,4} = 1.12 \cdot \theta_{view}$

Die beste Korrektur gelingt ausgehend von der Chlorophyllkonzentration mit Korrektur 2.

Wird jedoch die das Wasser verlassende Strahldichte als Zielgröße betrachtet, dann wird es bei dieser wellenlängenabhängigen Größe sinnvoll, die Güte der Korrekturen insgesamt mit der mittleren Strahldichte pro Band zu gewichten. Das Untersuchungsgebiet

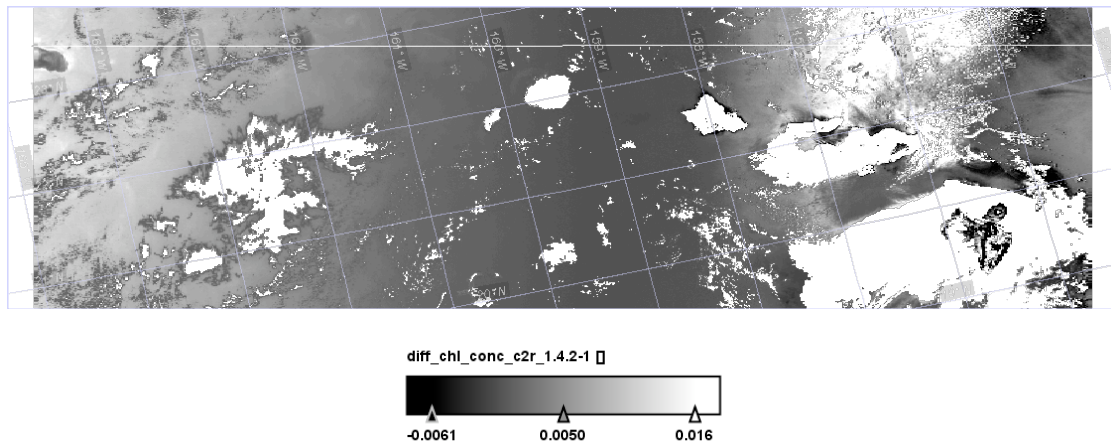
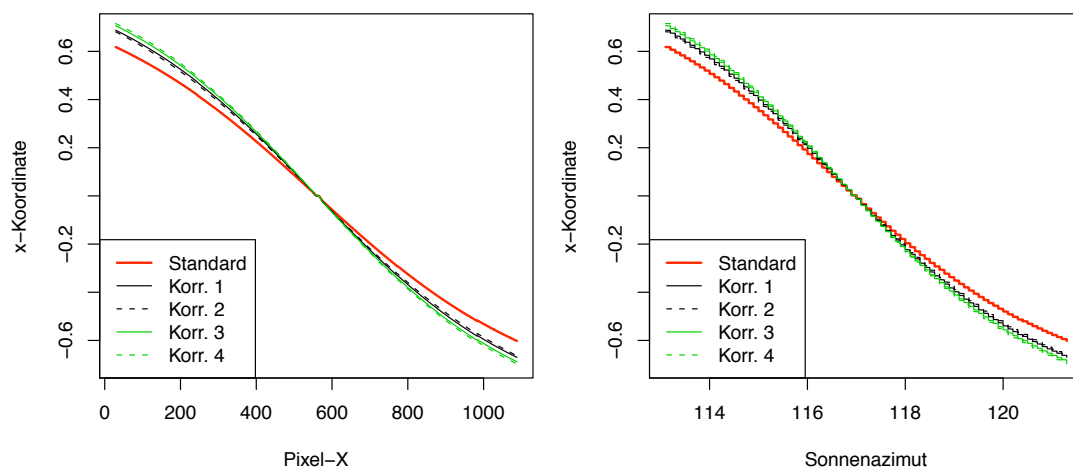


Abb. A.2: Differenz Chlorophyllkonzentrationen des Regionalprozessor ohne und mit Korrektur der Geometrie.



(a) als Funktion der Pixel-X-Nummer

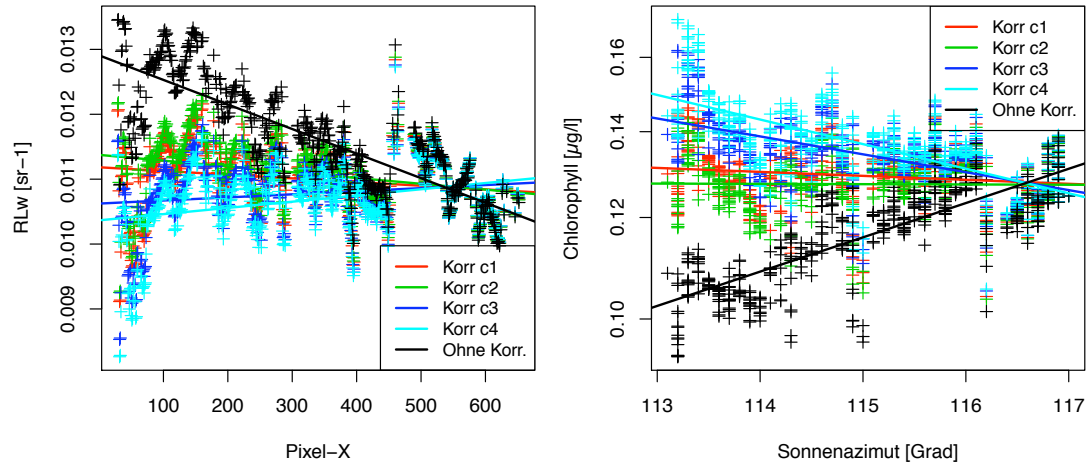
(b) als Funktion des Sonnenazimuts

Abb. A.3: x-Koordinate entlang einer Bildzeile für eine Szene vom 6. April 2004 bei Hawaii als Funktion der Pixelkoordinate X bzw. des Sonnenazimutwinkels. Verschiedene Korrekturen (1-4) des Beobachtungszenitwinkel führen zu den dargestellten Änderungen in der x-Koordinate.

A Weiterführende Untersuchungen zum systematischen Fehler

Korrektur	Transektenlänge N=510	Transektenlänge N=1010 (jeweils von Osten)
1	$-1.13e-06$	$-4.22e-05$
2	$-3.30e-05$	$-7.49e-05$
3	$1.12e-04$	$5.49e-05$
4	$1.65e-04$	$9.97e-05$

Tab. A.1: Vergleich der gewichteten mittleren Steigung der Ausgleichsgeraden bei der das Wasser verlassenden Strahldichte in Abhängigkeit von den Korrekturen.



(a) Strahldichte über Wasser bei $\lambda = 412.3$ nm und mit verschiedenen Korrekturen für den Beobachtungszenitwinkel.

(b) Chlorophyllkonzentration über eine Bildzeile mit verschiedenen Korrekturen des Beobachtungszenitwinkels.

Abb. A.4: Verlauf der Strahldichte über Wasser bei 412.3 nm bzw. der Chlorophyllkonzentration über eine Bildzeile für eine Szene vom 6. April 2004 bei Hawaii abhängig von der Korrektur der Beobachtungsgeometrie.

wird auf den östlichen Teil der Szene beschränkt, wo keine direkte Reflektion zu erwarten ist, sowie letztlich auf die 429 Datenpunkte, die innerhalb eines Intervalls niedriger optischer Dicke der Aerosole (zwischen $\tau_{550} = \{0.05, 0.185\}$) und Chlorophyllkonzentration zwischen $\{0.09, 0.14\} \mu\text{g/l}$ liegen. Nur die Kameragrenze und der Versatz zwischen Kamera 4 und 3 (bei Pixel-X = 468/9) läßt den Datensatz dann noch von einem gut durch eine lineare Beziehung zu beschreibenden Prozess abweichen. Korrektur 1 liefert basierend auf den Strahldichten die beste Veränderung zu einem konstanten Profil.

Da sich der Sonnenazimutwinkel über eine Bildzeile ebenfalls, wenn auch nur im Bereich von etwa 8 Grad ändert, ergibt sich eine hohe Korrelation zwischen der Chlorophyllkonzentration und dem Sonnenazimutwinkel. Dieser Trend tritt bereits in der Reflektanz des Wassers auf und zwar zunehmend mit abnehmender Wellenlänge im blaugrünen Spektralbereich, und wirkt sich damit auf die Ableitung der Chlorophyllkonzentration aus (Abb. A.5).

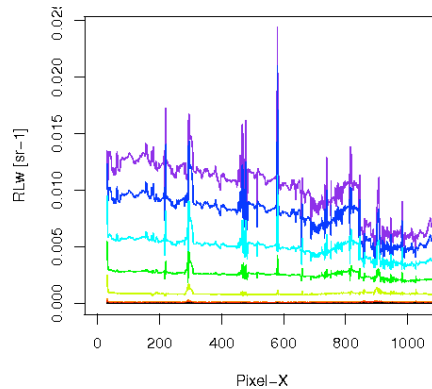


Abb. A.5: Strahldichte über Wasser in Abhängigkeit von der Wellenlänge

Eine Reihe von Experimenten werden sich an dieser Stelle anschließen.

Derzeit beruht die Bestimmung der Korrektur nur auf einer einzigen Szene, weitere Beispiele sind hier notwendig.

Die Anwendung des Prozessors mit Korrektur auf den Nordsee-Datensatz steht noch aus. Von besonderem Interesse wären die Neuberechnung der kurzen Zeitreihen, bei denen viele aufeinanderfolgende Datentage für das gleiche Gebiet gefunden wurden. Sollte sich hier eine deutliche Verminderung des dreitägigen Versatzes abzeichnen, so wäre ein weiteres Indiz für die Rechtfertigung dieser Korrektur gefunden.

Eine weitere ungeprüfte Möglichkeit eröffnet sich durch die Neuberechnung der Level 1-Produkte, die gerade abgeschlossen wird. Auf ihrer Basis ließen sich die Produkte des Regionalprozessors mit und ohne Korrektur vergleichen.

A.2 Korrektur der Beobachtungsgeometrie - 2

Dieser Ansatz ist im Moment der vielversprechendste und befindet sich derzeit (Stand: August 2010) in der Erprobung.

Statt die lokalen, an der Erdoberfläche definierten Winkel für die Atmosphärenkorrektur zu verwenden, werden alle Richtungen auf die Position des Satelliten bezogen (Abb. A.6), d.h. das Koordinatensystem aus Zenitposition und Flugrichtung. Für die Streuung in der Atmosphäre ist es unerheblich, wie sich die Winkel an der gekrümmten Oberfläche verhalten. Für eine Scanzeile - unter der Annahme, die Bewegung des Satelliten während der Zeit ist zu vernachlässigen - ist die Sonnenrichtung (Zenit und Azimut) im Bezug auf den Satelliten konstant. Zu jedem Pixel ist der Beobachtungszenitwinkel am Satelliten eindeutig festgelegt, der Beobachtungsazimut wird am sinnvollsten auf die Flugrichtung bezogen und ist jeweils für linke und rechte Bildhälfte in einer Scanzeile konstant. Da sich die Flugrichtung von Zeile zu Zeile ändert, sind auch die Azimutwinkel weiterhin von der Position des Satelliten abhängig.

Die lokalen Winkel, die im Satellitenprodukt aufgezeichnet sind, sind für die Streugeometrie im Wasser aber von entscheidender Bedeutung.

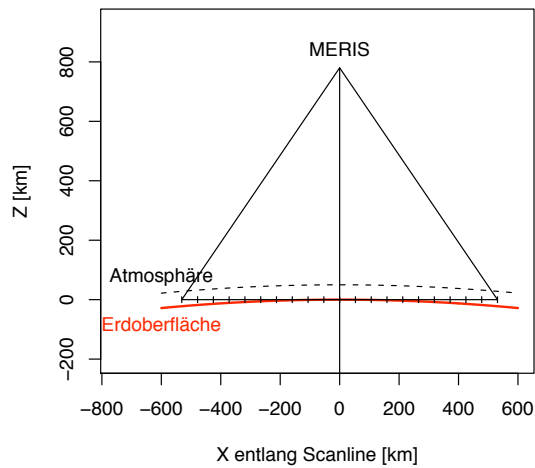


Abb. A.6: Streugeometrie bezogen auf Satellitenposition unabhängig von der gekrümmten Erdoberfläche (θ)

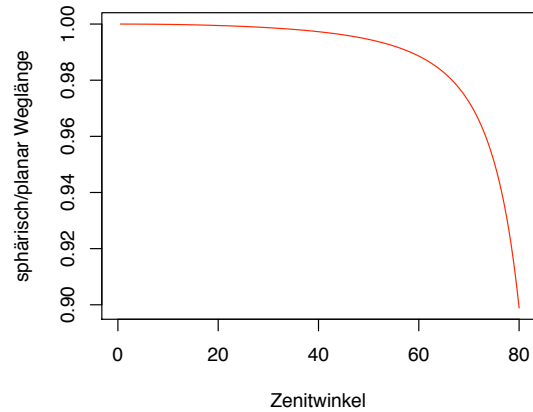


Abb. A.7: Abweichung der Weglängen zwischen sphärischem und planarem Modell einer Atmosphäre mit 50 km Höhe in Abhängigkeit vom (lokalen) Sonnenzenitwinkel

A.3 Abweichung der Weglänge in planarem oder sphärischem Atmosphärenmodell

Für den hohen Sonnenzenithwinkel im Winter führt die Annahme einer planaren Atmosphäre statt der tatsächlichen sphärischen Geometrie zu einer Abweichung in der Weglänge führen.

Dabei beträgt bei einem Sonnenstand von $\text{view}_{\text{zenith}} \simeq 59^\circ$ die Abweichung der Weglängen in einer 50 km dicken Atmosphäre (Höhe der Standardatmosphäre im Modell) etwa 1%, bei 75° sind es 5% (Abb. A.7).

Literaturverzeichnis

- [Abbott und Letelier 1998] ABBOTT, Mark R. ; LETELIER, Ricardo M.: Decorrelation scales of chlorophyll as observed from bio-optical drifters in the California Current. In: *Deep-Sea Research II* 45 (1998), S. 1639–1667
- [Akaike 1974] AKAIKE, Hirotugu: A new look at the statistical model identification. In: *IEEE Transactions on Automatic Control* 19 (1974), Nr. 6, S. 716 – 723
- [Allain und Cloitre 1991] ALLAIN, C. ; CLOITRE, M.: Characterizing the lacunarity of random and deterministic fractal sets. In: *Physical Review A* 44 (1991), Nr. 6, S. 3552–3558
- [Alvera-Azcarate u. a. 2005] ALVERA-AZCARATE, A. ; BARTH, A. ; RIXEN, M. ; BECKERS, J.M.: Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. In: *Ocean Modelling* 9 (2005), S. 325–346
- [Alves und Filho 2001] ALVES, Sinesio ; FILHO, Mauro B.: Multiscale measurements of fragmented cities: Enhancing Urban Analysis through lacunarity based measures. (2001)
- [Armstrong 1984] ARMSTRONG, M.: Problems with Universal Kriging. In: *Mathematical Geology* 16 (1984), Nr. 1
- [Becker 1990] BECKER, G. A.: *Warnsignale aus der Nordsee*. Jose L. Lozan and Walter Lenz and Eike Rachor and Burkhard Watermann und Hein von Westernhagen, 1990
- [Beckers u. a. 2006] BECKERS, J.-M. ; BARTH, A. ; ALVERA-AZCARATE, A.: DINEOF reconstruction of clouded images including error maps – application to the Sea-Surface Temperature around Corsican Island. In: *Ocean Science* 2 (2006), S. 183–199
- [Beckers und Rixen 2003] BECKERS, J. M. ; RIXEN, M.: EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. In: *Journal of Atmospheric and Oceanic Technology* 20 (2003), S. 1839–1856
- [Bissett u. a. 2001] BISSETT, W. P. ; SCHOFIELD, Oscar ; GLENN, Scott ; CULLEN, John J. ; MILLER, William L. ; PLUEDDEMAN, Albert J. ; MOBLEY, Curtis D.: Resolving the Impacts and Feedback of Ocean Optics on Upper Ocean Ecology. In: *Oceanography* 14 (2001), Nr. 3, S. 30–53
- [Bogaert 1996] BOGAERT, Patrick: Comparison of Kriging Techniques in a Space-Time Context. In: *Mathematical Geology* 28 (1996), Nr. 1

Literaturverzeichnis

- [Bowers und Simpson 1987] BOWERS, D. G. ; SIMPSON, J. H.: Mean position of tidal fronts in European-shelf seas. In: *Continental Shelf Research* 7 (1987), Nr. 1, S. 35–44
- [Brockmann-Consult] BROCKMANN-CONSULT: *Beam: Earth Observation Toolbox and Development Platform*. – URL www.brockmann-consult.de/cms/web/beam. – Zugriffsdatum: 2009
- [Chiles und Delfiner 1999] CHILES, Jean-Paul ; DELFINER, Pierre: *Geostatistics: Modeling Spatial Uncertainty*. John Wiley and Sons, Inc., 1999 (Wiley series in probability and statistics. Applied probability and statistics section)
- [Ciotti 1999] CIOTTI, Áurea: Influence of Phytoplankton Communities on Relationships between Optical Properties of Coastal Surface Waters. (1999)
- [Ciotti u. a. 1999] CIOTTI, Áurea M. ; CULLEN, John J. ; LEWIS, Marlon R.: A semi-analytical model of the influence of phytoplankton community structure on the relationship between light attenuation and ocean color. In: *Journal of Geophysical Research* 104 (1999), Nr. C1, S. 1559–1578
- [Cleveland und Devlin 1988] CLEVELAND, William S. ; DEVLIN, Susan J.: Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. In: *Journal of the American Statistical Association* 83 (1988), Nr. 403, S. 596–610
- [Cloern 1991] CLOERN, James E.: Tidal stirring and phytoplankton bloom dynamics in an estuary. In: *Journal of Marine Research* 49 (1991), S. 203–221
- [Cunningham u. a. 2003] CUNNINGHAM, Alex ; MCKEE, David ; CRAIG, Susanne ; TARRAN, Glen ; WIDDICOMBE, Claire: Fine-scale variability in phytoplankton community structure and inherent optical properties measured from an autonomous underwater vehicle. In: *Journal of Marine Systems* 43 (2003), S. 51–59
- [Demyanov und Soltani 2001] DEMYANOV, V. ; SOLTANI, S.: Wavelet analysis residual kriging vs. neural network residual kriging. In: *Stochastic Environmental Research and Risk Assessment* 15 (2001)
- [Denman und Abbott 1988] DENMAN, Kenneth R. ; ABBOTT, Mark R.: Time Evolution of Surface Chlorophyll Patterns From Cross-Spectrum Analysis of Satellite Color Images. In: *Journal of Geophysical Research* 93 (1988), 6798, Nr. C6, S. 6789
- [Denman und Abbott 1994] DENMAN, Kenneth R. ; ABBOTT, Mark R.: Time scales of pattern evolution from cross-spectrum analysis of Advanced Very High Resolution Radiometer and Coastal Zone Color Scanner imagery. In: *Journal of Geophysical Research* 99 (1994), Nr. C4, S. 7433–7442
- [Doerffer und Peters 2006] DOERFFER, Roland ; PETERS, Marco: MERIS Regional Case 2 Water BEAM Extension Flags ATBD / GKSS Research Centre Geesthacht, Institute for Coastal Research. 2006. – Forschungsbericht

Literaturverzeichnis

- [Doerffer und Schiller 1997] DOERFFER, Roland ; SCHILLER, Helmut: ATBD: Pigment index, sediment and gelbstoff retrieval from directional water leaving radiance reflectances using inverse modelling technique / GKSS Research Centre, Geesthacht. 1997 (2.12). – Algorithm Theoretical Basis Document
- [Doerffer und Schiller 2007] DOERFFER, Roland ; SCHILLER, Helmut: The MERIS Case 2 water algorithm. In: *International Journal of Remote Sensing* 28 (2007), Nr. 3-4, S. 517–535
- [Dutter 1985] DUTTER, Rudolf: *Geostatistik*. Teubner, Stuttgart, 1985 (Mathematische Methoden in der Technik)
- [Edwards u. a. 2001] EDWARDS, Martin ; REID, Philip ; PLANQUE, Benjamin: Long-term and regional variability of phytoplankton biomass in the Northeast Atlantic (1960–1995). In: *ICES Journal of Marine Science* 58 (2001), S. 39–49
- [Elliott und Clarke 1991] ELLIOTT, A. J. ; CLARKE, T.: Seasonal stratification in the northwest European shelf seas. In: *Continental Shelf Research* 11 (1991), Nr. 5, S. 467–492
- [Englund 1990] ENGLUND, Evan J.: A Variance of Geostatisticians. In: *Mathematical Geology* 22 (1990), Nr. 4, S. 417–455
- [ESA 2000] ESA: MERIS Product Handbook / European Space Agency. 2000. – Forschungsbericht
- [Feagin u. a. 2007] FEAGIN, R.A. ; WU, X.B. ; FEAGIN, T.: Edge effects in lacunarity analysis. In: *Ecological Modelling* 201 (2007), S. 262–268
- [Fischer 2005] FISCHER, Adrian: *Validierung des globalen MERIS- METEO- Wasserdampfprodukts mit Mikrowellenradiometer-Messungen*, Freie Universität Berlin, Diplomarbeit, 2005
- [Folkestad und Sørensen 2010] FOLKESTAD, Are ; SØRENSEN, Kai: Monitoring of Phytoplankton and Water Quality in the Arctic. In: *Presentation ESA Living Planet Symposium 2010* Norwegian Institute for Water Research (NORWAY) (Veranst.), 2010
- [Fuentes 2001] FUENTES, Montserrat: A high frequency kriging approach for non-stationary environmental process. In: *Environmetrics* 12 (2001), S. 469–483
- [Gandin 1963] GANDIN, L. S.: *Objective Analysis of Meteorological Fields*. Gidrometeor. Izdat., 1963
- [Gefen u. a. 1983] GEFEN, Yuval ; MEIR, Yigal ; MANDELBROT, Benoit B. ; AHARONY, Amnon: Geometric Implementation of Hypercubic Lattices with Noninteger Dimensionality by Use of Low Lacunarity Fractal Lattices. In: *Physical Review Letters* 50 (1983), Nr. 3, S. 145–148

Literaturverzeichnis

- [Gjertsen u. a. 2004] GJERTSEN, Karen ; RONG, Vigdis ; BROKER, Karen ; ENERSEN, Svein E.: Report on oceanographic cruises and data stations, 2004 / Havforskningsinstituttet, Institute of Marine Research. 2004 (4). – Forschungsbericht
- [Gneiting u. a. 2005] GNEITING, Tilmann ; GENTON, Marc G. ; GUTTORP, Peter: Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry / Department of Statistics, University of Washington. 2005 (475). – Forschungsbericht
- [Gower u. a. 1980] GOWER, J. F. R. ; DENMAN, K. L. ; HOLYER, R. J.: Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. In: *Nature* 288 (1980), S. 157–159
- [Greenhill u. a. 2002] GREENHILL, Darrel ; RIPKE, Lennart T. ; HITCHMAN, Adrian P. ; JONES, Graeme A. ; WILKINSON, Graeme G.: Use of Lacunarity index to characterize sub-urban areas for land use planning using IKONOS-2 Multispectral imagery. (2002)
- [Gypens u. a. 2007] GYPENS, Nathalie ; LACROIX, Geneviève ; LANCELOT, Christiane: Causes of variability in diatom and Phaeocystis blooms in Belgian coastal waters between 1989 and 2003: A model study. In: *Journal of Sea Research* 57 (2007), S. 19–35
- [Harris 1978] HARRIS, Fredric J.: On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. In: *Proceedings of the IEEE* Bd. 66, 1978, S. 51–83
- [Holligan u. a. 1989] HOLLIGAN, P. M. ; AARUP, T. ; GROOMS, S. B.: The North Sea: Satellite Colour Atlas. In: *Continental Shelf Research* 9 (1989), S. 667–765
- [Holt und Umlauf 2008] HOLT, Jason ; UMLAUF, Lars: Modelling the tidal mixing fronts and seasonal stratification of the Northwest European Continental shelf. In: *Continental Shelf Research* 28 (2008), S. 887–903
- [Huisman 1999] HUISMAN, Jef: Critical depth and critical turbulence: Two different mechanisms for the development of phytoplankton blooms. In: *Limnol. Oceanogr.* 44 (1999), S. 1781–1787
- [Huthnance 1991] HUTHNANCE, J. M.: Physical Oceanography of the North Sea. In: *Ocean and Shoreline Management* 16 (1991), S. 199–231
- [IOCCG 2007] IOCCG: Ocean-Colour Data Merging. In: *Reports of the International Ocean-Colour Coordinating Group* 6 (2007)
- [Johns und Reid 2001] JOHNS, D. G. ; REID, P. C.: An Overview of Plankton Ecology in the North Sea / SAHFOS. 2001. – Forschungsbericht
- [Journal u. a. 2000] JOURNAL, Andre G. ; KYRIAKIDIS, Phaedon C. ; MAO, Shuguang: Correcting the Smoothing Effect of Estimators: A Spectral Postprocessor. In: *Mathematical Geology* 32 (2000), Nr. 7

Literaturverzeichnis

- [Krige 1951] KRIGE, D. G.: A statistical approach to some basic mine valuation problems on the Witwatersrand. In: *Journal of the Chemical, Metallurgical and Mining Society of South Africa* (1951), S. 119–139
- [Kröncke und Knust 1995] KRÖNCKE, I. ; KNUST, R.: The Dogger Bank: a special ecological region in the central North Sea. In: *Helgoländer Meeresunters.* 49 (1995), S. 335–353
- [Lee 1980] LEE, A. J.: North Sea: physical oceanography. In: BANNER, F. T. (Hrsg.) ; COLLINS, M. B. (Hrsg.) ; MASSIE, K. S. (Hrsg.): *The North-West European Shelf Seas: The Sea Bed and the Sea in Motion* Bd. 2. Amsterdam : Elsevier, 1980, S. 467–493
- [Lenhart u. a. 1995] LENHART, Hermann J. ; RADACH, Günther ; BACKHAUS, Jan O. ; POHLMANN, Thomas: Simulations of the North Sea circulation, its variability, and its implementation as hydrodynamical forcing in ERSEM. In: *Netherlands Journal of Sea Research* 33 (1995), Nr. 3, S. 271–299
- [Lenhart u. a. 1997] LENHART, Hermann J. ; RADACH, Günther ; RUARDIJ, Piet: The effects of river input on the ecosystem dynamics in the continental coastal zone of the North Sea using ERSEM. In: *Journal of Sea Research* 38 (1997), S. 249–274
- [Levin 1992] LEVIN, Simon A.: The Problem of Pattern and Scale in Ecology. In: *Ecology* 73 (1992), Nr. 6, S. 1943–1967
- [Lomb 1975] LOMB, N. R.: Least-squares frequency analysis of unequally spaced data. In: *Astrophysics and Space Science* 39 (1975), S. 447–462
- [Mandelbrot 1983] MANDELBROT, B.B.: *The Fractal Geometry of Nature*. W. H. Freeman, New York, 1983
- [Martinez und Saar 2002] MARTINEZ, Vicent J. ; SAAR, Enn: *Statistics Of The Galaxy Distribution*. CRC Press, 2002
- [Matheron 1963] MATHERON, G.: Principles of geostatistics. In: *Economic Geology* 58 (1963), S. 1246–1266
- [Matheron 1971] MATHERON, G.: *The Theory of Regionalized Variables and its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, 1971
- [MERIS-SPPA-Team 2008] MERIS-SPPA-TEAM: Evolution of the MERIS Instrument Processing Facility. 2008. – Forschungsbericht
- [Mills u. a. 1994] MILLS, David K. ; TETT, Paul ; NOVARINO, Gianfranco: The spring bloom in the South Western North Sea in 1989. In: *Netherlands Journal of Sea Research* 33 (1994), Nr. 1, S. 65–80

Literaturverzeichnis

- [Mitchell u. a. 2005] MITCHELL, Matthew W. ; GENTON, Marc G. ; GUMPERTZ, Marcia L.: Testing for separability of space–time covariances. In: *Environmetrics* 16 (2005), S. 819–831
- [Mitchell 2002] MITCHELL, Matthew W.: *Testing Separability of Covariances for Space-Time Processes*, NC STATE UNIVERSITY, Dissertation, 2002
- [Moll und Radach 2003] MOLL, Andreas ; RADACH, Günther: Review of three-dimensional ecological modelling related to the North Sea shelf system Part 1: models and their results. In: *Progress in Oceanography* 57 (2003), S. 175–217
- [Moore u. a. 2005] MOORE, C. M. ; LUCAS, Mike I. ; SANDERS, Richard ; DAVIDSON, Russell: Basin-scale variability of phytoplankton bio-optical characteristics in relation to bloom state and community structure in the Northeast Atlantic. In: *Deep-Sea Research I* 52 (2005), S. 401–419
- [Morel 1974] MOREL, A.: *Optical Aspects of Oceanography*. Kap. Optical properties of pure water and pure seawater, S. 1–24, Academic, San Diego, Calif., 1974
- [Morel und Prieur 1977] MOREL, Andre ; PRIEUR, Louis: Analysis of variations in ocean color. In: *Limnology and Oceanography* 22 (1977), Nr. 4, S. 709–722
- [Müller u. a. 2010] MÜLLER, Dagmar ; KRASEMANN, Hajo ; DOERFFER, Roland: Analysis of MERIS chlorophyll time series (2003–2008) or systematic errors in MERIS Level 2 Products. In: *ESA Living Planet Symposium* Bd. SP-686, 2010
- [Myers 1982] MYERS, Donald E.: Matrix Formulation of Co-Kriging. In: *Mathematical Geology* 14 (1982), Nr. 3
- [Nielsen und Richardson 1989] NIELSEN, Torkel G. ; RICHARDSON, Katherine: Food chain structure of the North Sea plankton communities: seasonal variations of the role of the microbial loop. In: *Mar Ecol Prog Ser* 56 (1989), S. 75–87
- [Nychka 2007] NYCHKA, Doug: *fields: Tools for spatial data*. 2007. – URL <http://www.image.ucar.edu/GSP/Software/Fields>
- [Obata u. a. 1996] OBATA, Atsushi ; ISHIZAKA, Joji ; ENDOH, Masahiro: Global verification of critical depth theory for phytoplankton bloom with climatological in situ temperature and satellite ocean color data. In: *Journal of Geophysical Research* 101 (1996), Nr. C9, S. 20657–20667
- [Olea 1999] OLEA, Ricardo A.: *Geostatistics for engineers and earth scientists*. Kluwer Academic Publishers, 1999
- [Olea und Pawlowsky 1996] OLEA, Ricardo A. ; PAWLOWSKY, Vera: Compensating for Estimation Smoothing in Kriging. In: *Mathematical Geology* 28 (1996), Nr. 4
- [Pebesma 2004] PEBESMA, Edzer J.: Multivariable geostatistics in S: the gstat package. In: *Computers and Geosciences* 30 (2004), S. 683–691

Literaturverzeichnis

- [Pebesma und Wesseling 1998] PEBESMA, Edzer J. ; WESSELING, Cees G.: gstat: a program for geostatistical modelling, prediction and simulation. In: *Computers and Geosciences* 24 (1998), Nr. 1
- [Peters u. a. 2005] PETERS, S.W.M. ; ELEVELD, M. ; R.PASTERKAMP ; WOERD, H. van der ; DEVOLDER, M. ; JANS, S. ; PARK, Y. ; RUDDICK, K. ; BLOCK, T. ; BROCKMANN, C. ; DOERFFER, R. ; KRASEMANN, H. ; RÖTTGERS, R. ; SCHÖNFELD, W. ; JØRGENSEN, P.V. ; TILSTONE, G. ; MARTINEZ-VICENTE, V. ; MOORE, G. ; SØRENSEN, K. ; HØKEDAL, J. ; JOHNSEN, T.M. ; LØMSLAND, E.R. ; AAS, E.: *Atlas of Chlorophyll-a concentration for the North Sea based on MERIS imagery of 2003*. IVM report, VU- IVM, Amsterdam, January 2005
- [Petersen u. a. 2008] PETERSEN, W. ; WEHDE, H. ; KRASEMANN, H. ; COLIJN, F. ; SCHROEDER, F.: FerryBox and MERIS - Assessment of coastal and shelf sea ecosystems by combining in situ and remotely sensed data. In: *Estuarine, Coastal and Shelf Science* 77 (2008), S. 296–307
- [Peura und Iivarinen 1997] PEURA, M. ; IIVARINEN, J.: Efficiency of Simple Shape Descriptors. In: *In Aspects of Visual Form*, World Scientific, 1997, S. 443–451
- [Pingree 1978] PINGREE, R. D.: Mixing and Stabilization of Phytoplankton Distributions on the Northwest European Continental Shelf. In: STEELE, John H. (Hrsg.): *Spatial Pattern in Plankton Communities*. Plenum Press, New York, 1978, S. 181–220
- [Pingree und Griffiths 1978] PINGREE, R. D. ; GRIFFITHS, D. K.: Tidal Fronts on the Shelf Seas Around the British Isles. In: *Journal of Geophysical Research* 83 (1978), Nr. C9, S. 4615–4622
- [Plotnick u. a. 1996] PLOTNICK, Roy E. ; GARDNER, Robert H. ; HARGROVE, William W. ; PRESTEGAARD, Karen ; PERLMUTTER, Martin: Lacunarity analysis: A general technique for the analysis of spatial patterns. In: *Physical Review E* 53 (1996), Nr. 5, S. 5461–5467
- [Pottier u. a. 2006] POTTIER, Claire ; GARCON, Veronique ; LAMICOL, Gilles ; SUDRE, Joel ; SCHAEFFER, Philippe ; TRAON, Pierre-Yves L.: Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic Using Weighted Averaging and Objective Analysis. In: *IEEE Transactions on Geoscience and Remote Sensing* 44 (2006), Nr. 2
- [Pottier u. a. 2008] POTTIER, Claire ; TURIEL, Antonio ; GARÇON, Véronique: Inferring missing data in satellite chlorophyll maps using turbulent cascading. In: *Remote Sensing of Environment* 112 (2008), S. 4242–4260
- [R-Development-Core-Team 2007] R-DEVELOPMENT-CORE-TEAM: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (Veranst.), 2007. – URL <http://www.R-project.org>

- [Radach und Moll 2006] RADACH, Günther ; MOLL, Andreas: Review of three-dimensional ecological modelling related to the North Sea shelf system. Part II: Model validation and data needs. In: GIBSON, R. N. (Hrsg.) ; ATKINSON, R. J. A. (Hrsg.) ; GORDON, J. D. M. (Hrsg.): *Oceanography and Marine Biology: An Annual Review* Bd. 44. Taylor and Francis Group, 2006, S. 1–60
- [Reid u. a. 1990] REID, E. C. ; LANCELOT, C. ; GIESKE, W. W. C. ; HAGMEIER, E. ; WEICHART, G.: Phytoplankton of the North Sea and its dynamics: a review. In: *Netherlands Journal of Sea Research* 26 (1990), Nr. 2, S. 295–331
- [Ribeiro und Diggle 2001] RIBEIRO, Paulo J. ; DIGGLE, Peter J.: geoR: a package for geostatistical analysis. In: *R-NEWS* 1 (2001), Nr. 2, S. 15–18. – URL <http://CRAN.R-project.org/doc/Rnews/>
- [Rivoirard 1987] RIVOIRARD, J.: Two Key Parameters When Choosing the Kriging Neighborhood. In: *Mathematical Geology* 19 (1987), Nr. 8
- [Rivoirard 2004] RIVOIRARD, Jacques: On Some Simplifications of Cokriging Neighborhood. In: *Mathematical Geology* 36 (2004), Nr. 8
- [Robertson 1994] ROBERTSON, A.: Directionality, fractals and chaos in wind-shaped forests. In: *Agricultural and Forest Meteorology* 72 (1994), S. 133–166
- [Rouhani und Myers 1990] ROUHANI, Shahrokh ; MYERS, Donald E.: Problems in Space-Time Kriging of Geohydrological Data. In: *Mathematical Geology* 22 (1990), Nr. 5
- [Ruurdij u. a. 1997] RUARDIJ, Piet ; HAREN, Hans V. ; RIDDERINKHOF, Herman: The impact of thermal stratification on phytoplankton and nutrient dynamics in shelf seas: a model study. In: *Journal of Sea Research* 38 (1997), S. 311–331
- [Sathyendranath u. a. 2004] SATHYENDRANATH, Shubha ; WATTS, Louisa ; DEVRED, Emmanuel ; PLATT, Trevor ; CAVERHILL, Carla ; MAASS, Heidi: Discrimination of diatoms from other phytoplankton using ocean-colour data. In: *MEPS* 272 (2004), S. 59–68
- [Saunier u. a. 2006] SAUNIER, Sébastien ; GORYL, Philippe ; DELWART, Steven ; LUDOVIC, Bourg: MERIS Full Resolution Products, Geometry Aspects. In: DANESY, D. (Hrsg.): *Second Working Meeting on MERIS and AATSR Calibration and Geophysical Validation* European Space Agency (Veranst.), 2006 (ESA SP-615)
- [Scargle 1989] SCARGLE, Jeffrey D.: Studies in Astronomical Time Series Analysis III. Fourier Transforms, Autocorrelation function, and Cross-Correlation functions of unevenly spaced Data. In: *The Astrophysical Journal* 343 (1989), S. 874–887
- [Schirmer u. a. 1994] SCHIRMER, F. ; ESSEN, H.-H. ; GURGEL, K.-W. ; SCHLICK, T. ; HESSNER, K.: Local Variability of Surface Currents Based on HF-Radar Measurements. In: SÜNDERMANN, J. (Hrsg.): *Circulation and Contaminant Fluxes in the North Sea*. Springer-Verlag Berlin Heidelberg, 1994, S. 270 – 289

- [Schrum 1997] SCHRUM, Corinna: Thermohaline stratification and instabilities at tidal mixing fronts: results of an eddy resolving model for the German Bight. In: *Continental Shelf Research* 17 (1997), Nr. 6, S. 689–716
- [Seuront und Lagadeuc 1998] SEURONT, Laurent ; LAGADEUC, Yvan: Spatio-temporal structure of tidally mixed coastal waters: variability and heterogeneity. In: *Journal of Plankton Research* 20 (1998), Nr. 7, S. 1387–1401
- [Shannon 1949] SHANNON, Claude E.: Communication in the Presence of Noise. In: *Proceedings of the IRE* Bd. 37, 1949, S. 10–21
- [Sharples 2008] SHARPLES, Jonathan: Potential impacts of the spring-neap tidal cycle on shelf sea primary production. In: *Journal of Plankton Research* 30 (2008), Nr. 2, S. 183–197
- [Sharples u. a. 2001] SHARPLES, Jonathan ; MOORE, C. M. ; RIPPETH, Tom P. ; HOLLIGAN, Patrick M. ; HYDES, David J. ; FISHER, Neil R. ; SIMPSON, John H.: Phytoplankton distribution and survival in the thermocline. In: *Limnol. Oceanogr.* 46 (2001), Nr. 3, S. 486–496
- [Sharples u. a. 2006] SHARPLES, Jonathan ; ROSS, Oliver N. ; SCOTT, Beth E. ; GREENSTREET, Simon P. ; FRASER, Helen: Inter-annual variability in the timing of stratification and the spring bloom in the North-western North Sea. In: *Continental Shelf Research* 26 (2006), S. 733–751
- [Shevyrnogov und Vysotskaya 2006] SHEVYRNOGOV, A. ; VYSOTSKAYA, G.: Spatial distribution of chlorophyll concentration seasonal dynamics types in the ocean based on the autocorrelation analysis of SeaWiFS data. In: *Advances in Space Research* 38 (2006), S. 2176–2181
- [Siegismund und Schrum 2001] SIEGISMUND, Frank ; SCHRUM, Corinna: Decadal changes in the wind forcing over the North Sea. In: *Clim Res* 18 (2001), S. 39–45
- [Simpson und Hunter 1974] SIMPSON, J. H. ; HUNTER, J. R.: Fronts in the Irish Sea. In: *Nature* 250 (1974), S. 404–406
- [Steele 1978] STEELE, J. H.: Some comments on plankton patches. In: STEELE, J. H. (Hrsg.): *Spatial Pattern in Plankton Communities*. Plenum Press, New York, 1978, S. 1–10
- [Stramski u. a. 2001] STRAMSKI, Dariusz ; BRICAUD, Annick ; MOREL, André: Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. In: *Applied Optics* 40 (2001), Nr. 18, S. 2929–2945
- [Stramski u. a. 2002] STRAMSKI, Dariusz ; SCIANDRA, Antoine ; CLAUSTRE, Hervé: Effects of temperature, nitrogen, and light limitation on the optical properties of the marine diatom *Thalassiosira pseudonana*. In: *Limnol. Oceanogr.* 47 (2002), S. 392–403

- [Sverdrup 1953] SVERDRUP, H.U.: On conditions for the vernal blooming of phytoplankton. In: *J. Cons. Int. Explor. Mer* 18 (1953), S. 287–295
- [Tett und Walne 1995] TETT, P. ; WALNE, A.: Observations and simulations of hydrography, nutrients and plankton in the southern North Sea. In: *Ophelia* 42 (1995), S. 371–416
- [Tiedje u. a. 2010] TIEDJE, Bente ; MOLL, Andreas ; KALESCHKE, Lars: Comparison of temporal and spatial structures of chlorophyll derived from MODIS satellite data and ECOHAM3 model data in the North Sea. In: *Journal of Sea Research* 64 (2010), S. 250–259
- [Townsend u. a. 1994] TOWNSEND, David W. ; CAMMEN, Leon M. ; HOLLIGAN, Patrick M. ; CAMPBELL, Daniel E. ; PETTIGREW, Neal R.: Causes and consequences of variability in the timing of spring phytoplankton blooms. In: *Deep-Sea Research I* 41 (1994), Nr. 5, S. 747–765
- [Vargas-Guzman und Yeh 1999] VARGAS-GUZMAN, J. A. ; YEH, T.-C. J.: Sequential kriging and cokriging: Two powerful geostatistical approaches. In: *Stochastic Environmental Research and Risk Assessment* 13 (1999), S. 416–435
- [Wackernagel 2003] WACKERNAGEL, Hans: *Multivariate Geostatistics: An Introduction with Applications*. 3. Springer, 2003
- [Yamamoto 2000] YAMAMOTO, Jorge K.: An Alternative Measure of the Reliability of Ordinary Kriging Estimates. In: *Mathematical Geology* 32 (2000), Nr. 4
- [Yamamoto 2005] YAMAMOTO, Jorge K.: Correcting the Smoothing Effect of Ordinary Kriging Estimates. In: *Mathematical Geology* 37 (2005), Nr. 1

Danksagung

Ich danke Herrn Prof. Hartmut Graßl für die ermutigenden Gespräche, die mich beizeiten vor kleineren Irrwegen bewahrt und mir Inspiration und Bestätigung gebracht haben;

ich danke Herrn Dr. Roland Doerffer, der mir nicht nur mit zahlreichen Ideen und unschätzbarem Fachwissen zur Seite stand, sondern auch meine Arbeit beharrlich hinterfragte und mich so dazu antrieb, meine Ergebnisse immer weiter zu präzisieren;

Dr. Hajo Krasemann danke ich für unsere Diskussionen, die mich immer wieder dazu gebracht haben, auf ein Niveau verständlicher Sprache zurückzukehren und meine Gedanken entsprechend sortieren zu müssen.

Steffen Gehnke danke ich für stets offene Ohren bei ernst und nicht so ernst gemeinten Fragen von fachlicher oder nicht so fachlicher Natur - und den entsprechenden Antworten.

Meinen übrigen Kollegen aus der Arbeitsgruppe KOF der GKSS gilt ebenfalls mein Dank; insbesondere Wolfgang Schönfeld für einige hilfreiche Hinweise zum Programmieren, Ulrike Kleeberg für viele Gespräche auch abseits der fachlichen Problematik und Kerstin Heymann für das vermittelte Gefühl, tatsächlich einen wertvollen Beitrag zu leisten.

Meinen Freunden Heike und Christoph, Marcel, Bert, Karsten, meiner Familie und einer großen Anzahl Ungenannter danke ich für sehr viel Geduld und Mitleidensfähigkeit. Ohne euch, die ihr mein Jammern und Klagen erduldet und mit neuen Perspektiven hilfreich zur Seite gestanden habt, hätte ich niemals die notwendige Hartnäckigkeit aufbringen und letztlich entdecken können, dass ich wirklich von Herzen gerne Forscher bin.