

UMWELTFORSCHUNGSPLAN DES
BUNDESMINISTERIUMS FÜR UMWELT,
NATURSCHUTZ UND REAKTORSICHERHEIT

Forschungsbericht 298 25 235
UBA-FB 000285



**Entwicklung statistischer Verfahren für
Ringversuche, Methodenvalidierung und
Auswertung von marinen Daten
Teilvorhaben II:
Zusammenstellung und Bewertung von
statistischen Verfahren zur Auswertung
von biologischen und chemischen Daten
in der MUDAB**

von

PD Dr. habil. Steffen Uhlig

unter Mitarbeit von

Kirsten Simon

Norbert Schick

Susan Köhler

Daniel Rothmaler

quo data

Gesellschaft für Qualitätsmanagement und Statistik mbH

Im Auftrag des Umweltbundesamtes

Diese TEXTE-Veröffentlichung kann bezogen werden bei

Vorauszahlung von 10,00 €

durch Post- bzw. Banküberweisung,
Verrechnungsscheck oder Zahlkarte auf das

Konto Nummer 4327 65 - 104 bei der
Postbank Berlin (BLZ 10010010)
Fa. Werbung und Vertrieb,
Ahornstraße 1-2,
10787 Berlin

Parallel zur Überweisung richten Sie bitte
eine schriftliche Bestellung mit Nennung
der **Texte-Nummer** sowie des **Namens**
und der **Anschrift des Bestellers** an die
Firma Werbung und Vertrieb.

Der Herausgeber übernimmt keine Gewähr
für die Richtigkeit, die Genauigkeit und
Vollständigkeit der Angaben sowie für
die Beachtung privater Rechte Dritter.
Die in dem Bericht geäußerten Ansichten
und Meinungen müssen nicht mit denen des
Herausgebers übereinstimmen.

Herausgeber: Umweltbundesamt
Postfach 33 00 22
14191 Berlin
Tel.: 030/8903-0
Telex: 183 756
Telefax: 030/8903 2285
Internet: <http://www.umweltbundesamt.de>

Redaktion: Fachgebiet II 3.3
Bettina Hengst

Berlin, August 2002

Inhaltsverzeichnis

2.1	EINLEITUNG.....	123
2.2	STATISTISCHE METHODEN ZUR AUSWERTUNG MARINER DATEN	124
2.2.1	STATISTISCHE METHODEN IM BLMP	124
2.2.1.1	Übersicht.....	124
2.2.1.1.1	Berichtshefte des BLMP	124
2.2.1.1.2	Trendanalysen in den Jahresberichten von MURSYS.....	126
2.2.1.2	Bewertung.....	126
2.2.2	TRENDANALYSEN AUF DER BASIS DES VERALLGEMEINERTEN BERLINER VERFAHRENS	128
2.2.2.1	Übersicht.....	128
2.2.2.2	Bewertung.....	130
2.2.3	TRENDANALYSEN IN INTERNATIONALEN ARBEITSGRUPPEN.....	131
2.2.3.1	Übersicht.....	131
2.2.3.1.1	Der LOESS-Smoother	131
2.2.3.1.2	Methode von Theil.....	132
2.2.3.1.3	Der Test von Mann-Kendall	132
2.2.3.1.4	Auf dem LOESS-Smoother basierende Trend-Tests.....	133
2.2.3.2	Bewertung.....	134
2.2.4	SCHLUSSFOLGERUNGEN UND EMPFEHLUNGEN	134
2.3	DIE FUNKTIONEN EINER TRENDANALYSE	136
2.3.1	VISUELLE DARSTELLUNG DER DATEN DURCH EINE TRENDLINIE	137
2.3.2	STATISTISCHER TREND-TEST	139
2.3.3	BESTIMMUNG DER LINEAREN TRENDKOMPONENTE.....	141
2.3.4	VERGLEICH DES GEGENWÄRTIGEN NIVEAUS MIT EINEM REFERENZNIVEAU ODER DEM NIVEAU ZU EINEM REFERENZZEITPUNKT	142
2.3.5	EMPFEHLUNGEN	142
2.4	ANALYSE VON JAHRESDATEN.....	144
2.4.1	SMOOTHER-MATRIX	144
2.4.2	TESTS BASIEREND AUF DEM LOESS SMOOTHER	147

2.4.2.1	Kontrast-Test	147
2.4.2.2	Varianzanalyse der Quadratsumme	148
2.4.2.3	Test auf linearen Trend.....	149
2.4.3	TEST AUF NICHT-LINEAREN TREND	150
2.4.3.1	Approximation erster Ordnung.....	151
2.4.3.2	Approximation zweiter Ordnung.....	151
2.4.4	VERGLEICH MIT EINEM SOLLWERT.....	154
2.4.4.1	Konfidenzintervall für das lineare Regressionsmodell.....	155
2.4.4.2	Ein auf dem LOESS-Smoother beruhendes Konfidenzintervall für den linearen Trend.....	156
2.4.4.3	Konfidenzintervall für den auf dem LOESS-Smoother basierenden systematischen Trend	157
2.4.4.4	Vorhersageintervall auf der Basis des LOESS-Smoother	158
2.4.5	ZUR DIFFERENZ ZWISCHEN ANFANG-ENDE-KONTRAST UND LINEAREM REGRESSIONSKOEFFIZIENTEN.....	160
2.4.5.1	Jährliche Veränderungsrate und der Anfang-Ende-Kontrast.....	160
2.4.5.2	Jahresveränderungsrate und der lineare Regressionskoeffizient	161
2.4.5.3	Beispiel	161
2.4.5.4	Verhalten in langen Zeitreihen	162
2.4.6	SCHLUSSFOLGERUNGEN	164
2.5	DER LOESS-SMOOTHER UNTER EINBEZIEHUNG VON MONATSDATEN.....	166
2.5.1	EINFÜHRUNG	166
2.5.2	DEFINITION	166
2.5.3	TESTS	167
2.5.4	BEISPIEL.....	168
2.5.5	SCHLUSSFOLGERUNGEN	173
2.6	TRENDANALYSEN MIT KOVARIABLEN: QUECKSILBERGEHALTE VON SCHOLLEN AUS DER DEUTSCHEN BUCHT 1985 – 1999.....	174
2.6.1	DESKRIPTIVE ANALYSE	174
2.6.2	ADJUSTIERUNG BEZÜGLICH DER INDIVIDUELLEN LÄNGE	180
2.6.3	ADJUSTIERUNG BEZÜGLICH DER MITTLEREN LÄNGE DER POPULATION.....	181

2.6.4	SCHLUSSFOLGERUNGEN FÜR DIE STICHPROBENPLANUNG	186
2.7	TRENDANALYSEN FÜR ABUNDANZ UND BIOMASSE.....	187
2.7.1	DATENBASIS	187
2.7.2	AUSWERTUNG.....	188
2.7.3	ÜBERSICHT DER ERGEBNISSE	196
2.8	VARIANZKOMPONENTEN FÜR EROD.....	204
2.8.1	VARIANZ-KOMPONENTEN FÜR LOG-EROD IM MÄRZ.....	204
2.8.2	VARIANZ-KOMPONENTEN FÜR LOG-EROD IM HERBST.....	206
2.9	EMPFEHLUNGEN	212
2.9.1	UNIVARIATE DATEN	212
2.9.2	TRENDANALYSEN	213
2.9.3	EVALUATION VON MONITORINGPROGRAMMEN.....	214
2.9.4	METHODENENTWICKLUNG	214
2.9.4.1	Weitere Verbesserungen der auf dem LOESS-Smoother basierenden Trendanalyse.....	214
2.9.4.1.1	Reduzierung künstlicher Schwingungen in der Trendlinie.....	214
2.9.4.1.2	Robuste Glättungsverfahren	215
2.9.4.1.3	Power und nachweisbarer Trend	215
2.9.4.1.4	Erarbeitung von Protokollen zur Charakterisierung der Zeitreihen.....	215
2.9.4.1.5	Trendanalyse auf der Basis Gamma-verteilter Daten	216
2.9.4.2	Raum-zeitliche Analysen.....	216
2.9.4.3	Poweranalysen.....	216
2.10	LITERATUR	219
2.11	ABBILDUNGSVERZEICHNIS.....	222
2.12	TABELLENVERZEICHNIS	224
2.13	VERZEICHNIS DER VERWENDETEN GLEICHUNGSSYMBOLS	225
2.14	ABKÜRZUNGSVERZEICHNIS	227
	DANKSAGUNG	228

2.1 Einleitung

Ziel des Vorhabens war die Zusammenstellung bislang verwendeter statistischer Methoden zur Auswertung mariner Daten im Rahmen des BLMP, eine Bewertung hinsichtlich ihrer Eignung sowie die Erarbeitung von Empfehlungen für die Anwendung bestimmter statistischer Verfahren für marines Monitoring. Um diese Arbeiten zu realisieren, wurde eine enge Zusammenarbeit sowohl auf nationaler als auch internationaler Ebene angestrebt. Hervorzuheben ist dabei insbesondere die Kooperation im Rahmen von ICES-WGSAEM, sowie die Zusammenarbeit mit Herrn Wasmund, Institut für Ostseeforschung, Warnemünde und Herrn Harms, Bundesanstalt für Fischerei. Die Ergebnisse dieser Zusammenarbeit sind in dem vorliegenden Abschlussbericht zusammengefasst.

Im Einzelnen erfolgt in Kapitel 2 eine Übersichtsdarstellung sowie Bewertung bislang verwendeter Methoden, wobei neben univariaten Kennwerten Trendanalysen im Vordergrund stehen. Da für Trendanalysen eine Vielzahl unterschiedlicher Methoden vorgeschlagen werden, wurde in Kapitel 3 eine Darstellung unterschiedlicher Trendfunktionen aufgenommen. Diese Darstellung beinhaltet zugleich eine Konzeption für die Erarbeitung von Trendanalyseprotokollen.

In den weiteren Kapiteln werden verschiedene Fragestellungen in Übereinstimmung mit den Arbeitsschwerpunkten internationaler Arbeitsgruppen behandelt: Kapitel 4 gibt eine Übersicht über die Methodik zur Trendanalyse auf Basis von Jahresdaten mittels des LOESS-Smoother, während in Kapitel 5 die Entwicklung von Verfahren zur Trendermittlung auf Basis von Monats- und Jahresdaten dargestellt ist. Kapitel 6 beinhaltet die Darstellung einer Methodik zur Trendanalyse unter Berücksichtigung von störenden Kofaktoren, wobei exemplarisch die Ergebnisse einer statistischen Analysen der Quecksilbergehalte bei Schollen in der Deutschen Bucht vorgestellt werden. Die Ergebnisse der Trendanalysen zu Biomasse und Abundanz ausgewählter Artengruppen des HELCOM-Programmes werden in Kapitel 7 präsentiert, während sich Kapitel 8 mit Methoden zur Überprüfung der Eignung von Monitoringprogrammen hinsichtlich der Sensitivität zur Erfassung eines Trends befasst. Die Methodik wird exemplarisch für EROD-Daten vorgestellt. Empfehlungen für einzelne Methoden ebenso wie für Monitoringprogramme werden in Kapitel 9 diskutiert.

2.1 Statistische Methoden zur Auswertung mariner Daten

Um anhand mariner Daten aus der Meeresumweltdatenbank (MUDAB) und der Umweltprobenbank beispielsweise zeitliche Trends erfassen zu können, gibt es eine ganze Reihe unterschiedlicher statistischer Verfahrensmöglichkeiten. Eine wichtige Literaturquelle für die statistischen Vorgehensweisen sind die Berichtshefte des BLMP, die jedenfalls bis 1993 neben dem in Tabellenform aufbereiteten Zahlenmaterial jeweils spezielle statistische Auswertungen beinhalten. Weiterhin finden sich Auswertungen zu Daten des BLMP in den vom Bundesamt für Seeschifffahrt und Hydrographie herausgegebenen Berichten zum Meeresumwelt-Reportsystem MURSYS. Überlegungen und Analysen zur Zerlegung von BLMP-Zeitreihen in Trend und Saison werden in einer Arbeit von Frank (1997) angestellt. Auf internationaler Ebene werden Methoden zur Auswertung von BLMP-Daten (sowie der Monitoringdaten anderer Staaten) insbesondere im ICES (International Council for the Exploration of the Sea) diskutiert, namentlich in den Berichten von WGSAAEM (Working Group on the Statistical Aspects of Environmental Monitoring) und des ACME (Advisory Committee of the Marine Environment). Die vorliegende Untersuchung erhebt nicht den Anspruch auf Vollständigkeit, was aufgrund der Fülle durchgeführter Trendanalysen auch kaum möglich wäre. Vielmehr geben die folgenden Abschnitte eine Übersicht sowie eine Bewertung häufig verwendeter Methoden.

2.1.1 Statistische Methoden im BLMP

2.1.1.1 Übersicht

2.1.1.1.1 Berichtshefte des BLMP

In den Berichtsheften zum BLMP werden ausschließlich univariate statistische Kennwerte zur Beschreibung der Lage und der Streuung von univariaten Stichproben $x_1, \dots, x_N$ eingesetzt.

Dabei handelt es sich um folgende Parameter:

(1) Arithmetisches Mittel:

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

(2) Median:

$$x_{0.5} = \begin{cases} (x_{(N/2)} + x_{(N/2+1)})/2 & \text{falls } N \text{ gerade} \\ x_{((N+1)/2)} & \text{sonst} \end{cases}, \quad (\text{Gleichung 2})$$

wobei $x_{(1)}, \dots, x_{(N)}$ die der Größe nach geordnete Stichprobe bezeichnet.

(3) Geometrisches Mittel

$$\bar{x}_g = \sqrt[N]{x_1 x_2 \cdots x_N} \quad (\text{Gleichung 3})$$

(4) Standardabweichung

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (\text{Gleichung 4})$$

(5) Spannweite

$$r = x_{(N)} - x_{(1)} \quad (\text{Gleichung 5})$$

(6) Konfidenzintervall für den Erwartungswert (95%)

$$\text{mit der Untergrenze } x_u = \bar{x} - t_{N-1, 1-\alpha/2} s \quad (\text{Gleichung 6})$$

$$\text{und der Obergrenze } x_o = \bar{x} + t_{N-1, 1-\alpha/2} s, \quad (\text{Gleichung 7})$$

wobei $t_{N-1, 1-\alpha/2}$ = das $(1 - \alpha/2)$ -Quantil der t-Verteilung mit $(N-1)$

Freiheitsgraden

$\alpha = 0.05$ = das vorgegebene Signifikanzniveau

Die genannten Kennwerte werden eingesetzt, um

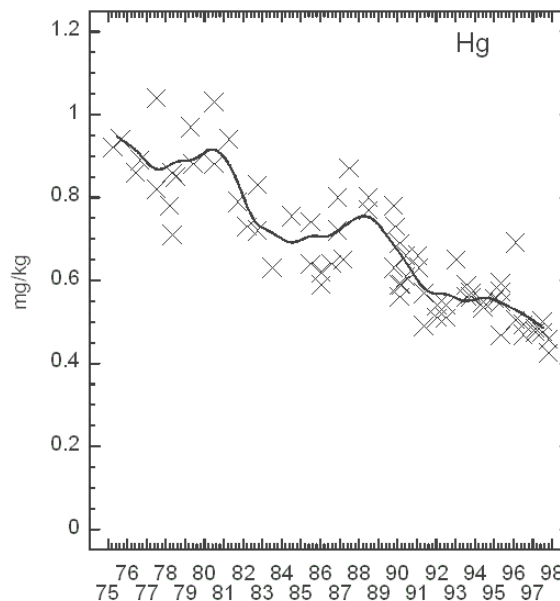
- Informationen zum Niveau und zur Streuung der einzelnen Messwerte zu erhalten
- Informationen zum Vertrauensbereich des ermittelten Niveaus zu erhalten (mittels Konfidenzintervall)
- grafische Vergleiche zwischen verschiedenen Messstationen bezüglich Niveau und Streuung durchführen zu können
- zeitliche Trends an einer Messstation grafisch erfassen zu können

Zu betonen ist dabei, dass es sich um univariate Kennwerte handelt, die jeweils nur für einen Zeitpunkt und eine Messstation ermittelt werden. Eine weitergehende Verrechnung erfolgt nicht. Gleichwohl werden die Ergebnisse in grafische Form so dargestellt, dass es möglich wird, räumliche oder zeitliche Trends zu erkennen. So wird beispielsweise das arithmetische Mittel mit dem Konfidenzintervall für den Erwartungswert kombiniert und im zeitlichen Vergleich

wiedergegeben (z.B. die Kontamination von Flussschwabeneier 1987, beschrieben im Berichtsheft 90/91).

2.1.1.1.2 Trendanalysen in den Jahresberichten von MURSYS

Neben univariaten Parametern, wie sie vorigen Abschnitt beschrieben wurden, findet man im MURSYS-Jahresbericht 1998 in einer Reihe von grafischen Darstellungen den Einsatz einer geglätteten Trendfunktion, deren Konstruktion allerdings nicht ausgeführt wird:



Die Glättungsfunktion ist eine nichtlineare und nichtmonotone Funktion, wie sie mit den klassischen Regressionsverfahren nicht ermittelt werden kann. Es ist daher davon auszugehen, dass es sich um ein rein mathematisches Glättungsverfahren handelt. Statistische Vertrauensbereiche werden nicht angegeben. Vielmehr wird auf der Basis kombinierter grafischer Darstellungen von Einzelwerten und Trendlinie mit der Evidenz der dargestellten Trends argumentiert.

2.1.1.2 Bewertung

Die in den Berichtsheften des BLMP verwendeten Kennwerte sind univariater Natur, beziehen sich also nur auf jeweils einen Orts- und Zeitpunkt. Eine weitergehende Verrechnung erfolgt nicht. Gleichwohl werden die Ergebnisse in grafische Form so dargestellt, dass es möglich wird, räumliche oder zeitliche Trends zu erkennen. So wird beispielsweise das arithmetische Mittel

mit dem Konfidenzintervall für den Erwartungswert kombiniert und im zeitlichen Vergleich wiedergegeben. Dies ermöglicht einen guten Eindruck der etwaigen zugrundeliegenden Trends, allerdings noch keine quantitative Beurteilung des Trends auf einer gesicherten statistischen Grundlage. So ist mit den verwendeten Parametern keine statistisch adäquate Berechnung der prozentualen Reduktion der Konzentrationen möglich. Erforderlich wäre deshalb der Einsatz von statistischen Trend-Tests zum Beispiel auf Basis des LOESS-Trends in Verbindung mit entsprechenden Konfidenzintervallen. Diese Methoden werden in den weiteren Kapiteln diskutiert.

Bei allen verwendeten Parametern handelt es sich mit einer Ausnahme um deskriptive Methoden, die nur dazu dienen, Niveau (Lokation) und Streuung (Dispersion) eines Datensatzes zu beschreiben, die allerdings nicht geeignet sind, weitergehende statistisch abgesicherte Schlussfolgerungen ziehen zu können. Einzig das Konfidenzintervall, welches dem Bereich der Inferenzstatistik zuzurechnen ist, ermöglicht u.a. einen statistisch abgesicherten Vergleich zweier Stichproben. Dabei muss allerdings einschränkend vermerkt werden, dass die berechneten Konfidenzintervalle auf der Basis von teilweise nur 10 Messwerten ermittelt wurden, und somit wegen der hier vorausgesetzten Normalverteilung als problematisch anzusehen sind. Eine Überprüfung auf Normalverteilung erfolgt nicht. Dies erweist sich als erheblicher Mangel, wenn zum Beispiel die Datensätze bis auf einige wenige extreme Ausreißer sehr homogen verteilt sind. Ausreißer können die Standardabweichung in extremer Weise verzerren und verursachen dann eine wesentliche Vergrößerung der zugehörigen Konfidenzintervalle. Bei Verwendung robuster Methoden würden diese Ausreißer einen sehr begrenzten Einfluss auf die Ergebnisse haben. Umgekehrt kann die Annahme der Normalverteilung auch eine falsche Datensicherheit vorspiegeln, wenn im jeweiligen Datensatz zwar keine extremen Abweichungen vorhanden sind, wohl aber die theoretische Verteilung von der Normalverteilung deutlich abweicht. In diesem Falle wird die Länge des Konfidenzintervalls von Jahr zu Jahr oder von Station zu Station starken Schwankungen unterliegen, was ebenfalls falsche Schlussfolgerungen verursachen kann.

Auffällig ist der Umstand, dass das Konfidenzintervall ausschließlich auf Basis der Originaldaten und damit auf Grundlage des arithmetischen Mittel ermittelt wird. Man kann das Konfidenzintervall auch auf Basis des geometrischen Mittels berechnen, indem zunächst alle Berechnungen mit den gewöhnlichen Formeln auf Basis der logarithmierten Werte durchgeführt werden, bevor eine Rücktransformation der ermittelten Konfidenzgrenzen mittels der

Exponentialfunktion erfolgt. Dies dürfte speziell bei rechtsschiefen Verteilungen, wie sie bei Konzentrationsmesswerten sehr häufig sind, eine deutlich bessere Konfidenzschätzung ermöglichen.

Die im MURSYS-Jahresbericht 1998 verwendete Trendfunktion gestattet eine optische Verdichtung der Daten auf einen zugrundeliegenden Trend. Statistische Vertrauensbereiche werden in den Darstellungen des Jahresberichts nicht dargestellt. Vielmehr wird auf der Basis kombinierter grafischer Darstellungen von Einzelwerten und Trendlinie mit der Evidenz der dargestellten Trends argumentiert. Soll jedoch eine genauere Quantifizierung des Trends erfolgen, oder ist es zweifelhaft, ob der vermutete Trend möglicherweise durch die zufälligen Schwankungen von Jahr zu Jahr verursacht wurde, sind weitergehende Trendanalysen erforderlich. Außerdem ist es aus statistischer Sicht fraglich, welchen Informationsgewinn der Betrachter erhält, wenn ein Diagramm der Einzelmesswerte zusätzlich eine statistisch nicht abgesicherte Trendlinie beinhaltet. Es ist nicht auszuschließen, dass die Trendlinie dem Betrachter als evident erscheint, obwohl kein statistischer Nachweis eines Trends möglich ist. Es besteht die Gefahr, dass je nach Spezifizierung des Trendanalyseverfahrens sehr unterschiedliche Resultate erzielt werden können. So wäre es im angegebenen Beispiel (Teilabschlußbericht II.1) aus statistischer Sicht durchaus vertretbar, anstelle eines nichtlinearen Trendverlaufs eine lineare Regressionsfunktion anzupassen.

2.1.2 Trendanalysen auf der Basis des Verallgemeinerten Berliner Verfahrens

2.1.2.1 Übersicht

In Frank (1997) wird die Anwendung des Verallgemeinerten Berliner Verfahrens zur Analyse von Schadstoffdaten im Rahmen des BLMP vorgeschlagen. Ziel dieses Verfahrens ist die Zerlegung von Zeitreihen $Y(t)$ in Trend und Saison:

$$Y(t) = u(t) + s(t) + e(t), \quad (\text{Gleichung 8})$$

wobei t den Zeitpunkt, $u(t)$ die Trendfigur, $s(t)$ die Saisonfigur und $e(t)$ den zufälligen Fehler bezeichnet. Die Ermittlung der Parameter für dieses Modell basiert auf der Minimierung eines Glattheitsmasses, welches über einen geeignet festgelegten Trend-Saison-Differentialoperator

definiert ist. Bezüglich der Details sei auf die Arbeit von Hebbel (1997) verwiesen. In der Arbeit von Frank wird ein Spezialfall betrachtet, der im folgenden skizziert wird:

Für jeden Zeitpunkt t wird eine ungewichtete lokale Regression durchgeführt, wobei folgendes Regressionsmodell verwendet wird:

$$Y(t) = \mu + \beta t + \gamma_1 \sin \frac{2\pi t}{P} + \gamma_2 \cos \frac{2\pi t}{P} + \gamma_3 \sin \frac{4\pi t}{P} + \gamma_4 \cos \frac{4\pi t}{P} + e(t) \quad \text{(Gleichung 9)}$$

Dabei sind μ , β sowie γ_1 , γ_2 , γ_3 und γ_4 die zu schätzenden Parameter, deren Werte allerdings für jeden Zeitpunkt t unterschiedlich sind (deshalb *lokale* Regression). P bezeichnet die Dauer einer Periode, d.h. in diesem Falle die Zeitdauer eines Jahres. Zur Berechnung der Parameter wird jeweils nur ein Teil der Daten herangezogen, der sogenannten Stützbereich, der in Abhängigkeit vom jeweils betrachteten Zeitpunkt t festgelegt wird. Dieser Stützbereich wurde in der genannten Arbeit mit einer Länge von 3 Jahren festgelegt, wobei jeweils jene Daten einbezogen wurden, die dem Zeitpunkt t am nächsten liegen. Dies bedeutet, dass am Anfang und am Ende der Zeitreihe jeweils alle Daten der ersten bzw. letzten drei Jahre verwendet werden, während bei allen Zeitpunkten ‚innerhalb‘ der Zeitreihe jeweils alle Daten verwendet werden, die in einem zeitlichen Abstand von nicht mehr als 18 Monaten erhoben wurden.

Zur Bestimmung der Trendfigur im Zeitpunkt t wird

$$u(t) = \mu + \hat{\beta}t \quad \text{(Gleichung 10)}$$

gesetzt, wobei $\hat{\beta}$ die lokale Regressionsschätzung des Parameters β bezeichnet. Die Saisonfigur wird ermittelt gemäss

$$s(t) = \hat{\gamma}_1 \sin \frac{2\pi t}{P} + \hat{\gamma}_2 \cos \frac{2\pi t}{P} + \hat{\gamma}_3 \sin \frac{4\pi t}{P} + \hat{\gamma}_4 \cos \frac{4\pi t}{P}, \quad \text{(Gleichung 11)}$$

wobei die Parameter $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4$ die lokalen Regressionsschätzungen für $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ bezeichnen.

Diese Berechnung wird für jeden Zeitpunkt auf der Basis des jeweils neu festzulegenden Stützbereichs wiederholt, so dass sich im Ergebnis eine in der Regel gute Anpassung an den tatsächlichen Verlauf von Trend und Saison ergibt. Das Verfahren kann durch flexible Modellierung und Festlegung der Stützbereiche an verschiedene stochastische Bedingungen angepasst werden kann. Dies wird durch die entsprechende Festlegung der Länge des

Stützbereiches, durch die optionale Einbeziehung einer quadratischen oder kubischen Trendkomponente sowie durch Berücksichtigung höherer Oberwellen erreicht. Allerdings gibt es keine Kriterien, mit der die genannten Flexibilitätsparameter festzulegen sind, d.h. es gibt kein Protokoll, welches sicherstellt, dass jeder Anwender des Verfahrens zum gleichen Ergebnis kommt.

2.1.2.2 Bewertung

Das Verallgemeinerte Berliner Verfahren ermöglicht eine flexible Trendmodellierung auch bei nichtlinearen und nichtmonotonen Trendverläufe und kann durch flexible Modellierung und Festlegung der Stützbereiche an verschiedene stochastische Bedingungen angepasst werden. Im Gegensatz zu konventionellen Regressionsansätzen ermöglicht das Verfahren auch dann noch eine gute Anpassung, wenn es sich um sehr lange Zeitreihen handelt, bei denen unter Umständen stark wechselnden stochastischen Bedingungen vorliegen können.

Da keine allgemeinen Kriterien zur Spezifizierung des Trendanalyseverfahrens verfügbar sind, besteht aufgrund der großen Flexibilität des Verfahrens allerdings die Gefahr, dass jeder Anwender zu einer anderen Trendanpassung kommt. Wenn also nicht nur eine grafisch ansprechende Trendkurve benötigt wird, müssten zur Festlegung von Trend und Saisonfigur zusätzliche Kriterien festgelegt werden.

Dies wird dadurch erschwert, dass zur Zeit keine statistischen Tests verfügbar sind, anhand derer die implizite Parametrisierung des Modells überprüft werden können, d.h. mit denen die genannten Kriterien festgelegt werden könnten. Weiterhin ist festzustellen, dass das beschriebene Verfahren explorativer Natur ist, d.h. auch eine Bestimmung von Vertrauensbändern ist nicht vorgesehen. Somit kann das Verallgemeinerte Berliner Verfahren sehr gut genutzt werden, um etwaige Saisoneffekte sichtbar zu machen und zugleich einen Trend zu ermitteln, doch aufgrund des Fehlens weiterer inferenzstatistischer Verfahren ist die Interpretierbarkeit der Ergebnisse eingeschränkt.

2.1.3 Trendanalysen in internationalen Arbeitsgruppen

2.1.3.1 Übersicht

Gemäss der OSPAR-Konvention zum Schutz des Nordost-Atlantiks sind “regelmässig gemeinsame Beurteilungen der Qualität der Meeresumwelt und deren Entwicklung für das Meeresgebiet oder Regionen oder Subregionen desselben durchzuführen”, und “in diese Beurteilung eine Bewertung der Wirksamkeit der zum Schutz der Meeresumwelt getroffenen und geplanten Maßnahmen sowie die Festlegung von Handlungsprioritäten einzubeziehen” (Anon. 1994). Zur Erfüllung dieser Anforderungen erfolgen Monitoring-Untersuchungen in den Vertragsstaaten der OSPAR-Kommission seit 1996 entsprechend den Anforderungen des Joint Assessment and Monitoring Programme (JAMP), wobei in Deutschland die Messungen im Rahmen des Bund/Länder-Messprogramms durchgeführt werden.

Ein wichtiges Instrument zur Auswertung dieser Messungen sind statistische Trendanalysen. Zu nennen ist hierbei insbesondere das auf dem LOESS-Smoother basierende Glättungsverfahren, welches u.a. von der MON-Arbeitsgruppe unter OSPAR diskutiert wird, sowie das nichtparametrische Verfahren von Theil, welches eine robuste Regressionsgerade liefert. Neben den eigentlichen Trendberechnungsverfahren werden jeweils auch Trend-Tests eingesetzt, so der Test von Mann-Kendall als auch verschiedene auf dem LOESS-Smoother basierende Trend-Tests.

Neben den eigentlichen Trendberechnungsverfahren werden jeweils auch Trend-Tests eingesetzt, so der Test von Mann-Kendall als auch verschiedene auf dem LOESS-Smoother basierende Trend-Tests. Die genannten Verfahren sollen im folgenden kurz vorgestellt werden. In allen Fällen wird eine Stichprobe $x_1, \dots, x_N$ von Jahreswerten vorausgesetzt, wobei es zunächst nicht von Belang ist, ob es sich bei diesen Werten um Einzelmessungen oder um aggregierte Werte handelt, also z.B. um den Median oder das arithmetische Mittel mehrerer Einzelmessungen.

2.1.3.1.1 Der LOESS-Smoother

Zur visuellen Darstellung von Trendlinien eignen sich Glättungsverfahren, beispielsweise (kubische) Splines oder der sogenannte LOESS-Smoother, während eine einfache polynomiale Regression vielfach zu einer schlechten Modellanpassung führt. Der Auswahl eines Verfahrens

sind letztlich kaum Grenzen gesetzt, denn es gibt kein statistisches Kriterium, welches in eindeutiger Weise ein bestimmtes Verfahren bevorzugt. Für die Erfassung von Trends in Biota wird vom ICES der LOESS-Smoother empfohlen, basierend auf einem Datenfenster von 7 Jahren, d.h. jeder Schätzwert basiert auf den Daten von jeweils sieben Jahren (des zugehörigen Jahres sowie der drei vorangegangenen und der drei nachfolgenden Jahre, wobei am Anfang und am Ende der Zeitreihe eine Modifikation des Datenfensters erforderlich ist). Dies ermöglicht für typische Monitoringzeitreihen in der Regel eine günstige Anpassung an den allgemeinen Trendverlauf. Die Bestimmung der Schätzwerte auf Basis der jeweils 7 Jahreswerte erfolgt mittels einer gewichteten linearen Regression, wobei jene Jahreswerte das größte Gewicht haben, die dem zu schätzenden Wert zeitlich am nächsten liegen. (Die durch den zeitlichen Abstand bestimmte Gewichtsfunktion hat kubische Gestalt.)

2.1.3.1.2 Methode von Theil

Neben dem allgemeinen (nichtlinearen) Trendverlauf besteht vielfach auch Interesse an dem linearen Anteil des Trends in einer vorgegebenen Zeitspanne. Ausgedrückt werden kann dieser lineare Anteil durch die (positive oder negative) Steigung der linearen Trendlinie oder als prozentualer Anstieg oder Rückgang. Die Bestimmung des Steigungsparameters kann entweder mittels einer linearen Regression erfolgen oder mittels der nichtparametrischen Methode von Theil, bei der zunächst für alle Jahre i und j mit $i > j$ die Differenzen $(y_i - y_j)/(i - j)$ ermittelt werden und hiervon der Median gebildet wird. Diese Methode bietet sich als Ersatz für die lineare Regressionsmethode insbesondere dann an, wenn mit Ausreißern aufgrund von Messfehlern zu rechnen ist. Man sollte sich dessen bewusst sein, dass speziell bei stark nichtlinearen Trendverläufen die verschiedenen Methoden zu sehr unterschiedlichen Ergebnissen führen können.

2.1.3.1.3 Der Test von Mann-Kendall

Solange die Zeitreihe durch zufällige Schwankungen von Jahr zu Jahr dominiert wird und kein deutlicher Trend zu erkennen ist, ist die Ermittlung einer Trendlinie und daraus abgeleiteter Größen von zweifelhaftem Nutzen, wenn nicht nachgewiesen wird, dass sich im der Trendverlauf reale Veränderungen zeigen. Die Überprüfung dieser Frage erfolgt mit Hilfe eines statistischen Trend-Tests. Ein solcher Test basiert auf der Anfangsvermutung (Nullhypothese), dass kein systematischer Trend vorliegt, und dass alle Schwankungen nur durch zufällige Veränderungen der Einzelwerte verursacht werden. Erst wenn durch einen geeigneten Trend-Test zu einer vorgegebenen Irrtumswahrscheinlichkeit α (meist wird α mit 5% festgelegt) diese

Anfangsvermutung widerlegt werden kann, kann der Trend als signifikant angesehen werden, d.h. als nachgewiesen gelten.

Der Trend-Test von Mann-Kendall ist ein in vielen Fällen sehr trennscharfes Testverfahren, welches ohne weitere Verteilungsvoraussetzungen wie z.B. die Normalverteilung auskommt. Der Test basiert auf einer Prüfgröße, die sich aus der Differenz der beiden folgenden Zahlen ergibt:

- der Anzahl jener Zahlenwerte in der Stichprobe, bei denen der frühere Wert kleiner ist als der spätere Wert,
- der Anzahl jener Zahlenwerte in der Stichprobe, bei denen der frühere Wert größer ist als der spätere Wert.

Formal lässt sich diese Prüfgröße darstellen als

$$S = \sum_{i < j} \text{sgn}(x_j - x_i), \quad (\text{Gleichung 12})$$

wobei

$$\text{sgn}(y) = \begin{cases} 1 & \text{falls } y > 0 \\ 0 & \text{falls } y = 0 \\ -1 & \text{falls } y < 0 \end{cases}. \quad (\text{Gleichung 13})$$

Wenn kein Trend vorliegt, ist davon auszugehen, dass S um den Wert Null pendelt. Nur bei deutlichem monotonen Trend wird S stark positiv (bei positivem Trend) oder stark negativ (bei negativem Trend) ausfallen. Somit ist die Nullhypothese genau dann abzulehnen, wenn der Betrag von S einen kritischen Wert übersteigt. Dieser kritische Wert der Prüfgröße ist vom Signifikanzniveau und der Anzahl der Beobachtungen abhängig.

2.1.3.1.4 Auf dem LOESS-Smoother basierende Trend-Tests

Während der Test von Mann-Kendall zum Nachweis monotoner Trends verwendet werden kann, sind auf dem LOESS-Smoother basierende Tests auch zum Nachweis nichtmonotoner Trends geeignet. Diese Tests sind in verschiedenen Varianten verfügbar, die zu jeweils unterschiedlichen Zwecken dienen. Ihnen allen gemeinsam ist jedoch das von der Varianzanalyse abgeleitete Konstruktionsprinzip des F-Tests, bei dem zwei von verschiedenen Residualquadratsummen abgeleitete Varianzterme den Zähler und den Nenner der Prüfgröße bilden.

2.1.3.2 Bewertung

Zur Trendanalyse werden eine ganze Reihe von Verfahren eingesetzt und diskutiert, wobei je nach Anwendungsfall unterschiedliche, möglicherweise mehrere Verfahren eingesetzt werden können. Die verschiedenen zur Trendanalyse von Jahresdaten zur Verfügung stehenden Verfahren haben in der Arbeitsgruppe WGSÄEM einige Diskussionen zur Bewertung und Einschätzung der Verfahren, insbesondere im Hinblick auf die Einsatzmöglichkeiten ausgelöst. Ergebnis dieser Diskussionen war insbesondere eine konzeptionelle Grundlegung statistischer Trendanalysen für Monitoringzwecke, die im folgenden Kapitel vorgestellt wird. Zu beachten ist in jedem Falle, dass die Leistungsfähigkeit der diskutierten Verfahren hinsichtlich der Möglichkeit, zu statistisch abgesicherten Testergebnissen zu kommen, sehr unterschiedlich ausfällt. Vielfach ermöglichen gerade jene Verfahren, welche besonders flexibel bei der Erfassung von Trends sind, keine Durchführung eines statistischen Trend-Tests. Somit hat jede der bislang zur Auswertung von BLMP-Daten verwendeten Trendanalysemethoden spezifische Vor- und Nachteile, so dass empfohlen wird, flexible Protokolle für die Trendanalyse zu erarbeiten, die den jeweiligen Anforderungen möglichst nahe kommen.

2.1.4 Schlussfolgerungen und Empfehlungen

Die im Rahmen des BLMP eingesetzten univariaten Verfahren sind für die meisten Anwendungen praktikabel, jedoch unter gewissen Voraussetzungen, speziell bei stark ausreißerbelasteten oder linksschiefen Verteilungen, möglicherweise in manchen Fällen irreführend, so dass sich eine zusätzliche Überprüfung der Daten anhand weiterer Kriterien empfiehlt. Hierfür geeignet sind z.B. Boxplots, welche die Verteilung der Daten grafisch anhand der Extremwerte und der Quartile verdeutlichen, oder nichtparametrische Parameter für Schiefe und Kurtosis, welche die Abweichungen von der Normalverteilung messen. In den marktgängigen Softwarepaketen sind verschiedene Parameter verfügbar, so dass hier auf eine Festlegung verzichtet werden soll.

Jede der bislang zur Auswertung von BLMP-Daten verwendeten Trendanalysemethoden hat spezifische Vor- und Nachteile, so dass empfohlen wird, flexible Protokolle für die Trendanalyse zu erarbeiten, die den jeweiligen Anforderungen möglichst nahe kommen. Grundlage dieser Protokolle sollte die in Kapitel 3 dargestellte Konzeption einer Trendanalyse sein.

Es ist festzuhalten, dass das in der Statistik vielfach diskutierte Kriterium der Robustheit nicht in allen Fällen gleichermaßen zu bewerten sind. So ist es allein von der Fragestellung abhängig, ob ein Ausreißer möglichst geringen Einfluss haben soll und gegebenenfalls eliminiert bzw. ein ausreißer-robustes Verfahren eingesetzt werden soll, oder ob für das Monitoringprogramm gerade die Erfassung von Ausreißern (z.B. durch Unfälle verursacht) höchste Priorität haben. Bei dieser Betrachtung ist auch zu prüfen, ob es sich um Einzelausreißerwerte innerhalb eines Jahres handelt (bei einzelnen Messungen bzw. Individuen), oder ob die Gesamtheit aller Messwerte eines Jahres bezüglich ihres Mittelwertes als Ausreißer anzusehen ist. So sind für manche Fragestellungen Ausreißer auf der Messwerte-Ebene möglichst zu eliminieren, während Ausreißer auf der aggregierten Ebene von großem Interesse sind.

Schließlich ist darauf hinzuweisen, dass Verfahren, die für normalverteilte Daten konzipiert wurden, vielfach auch für nichtnormalverteilte Daten einsetzbar sind. So ist aufgrund des Zentralen Grenzwertsatzes davon auszugehen, dass selbst bei stark nicht-normalverteilten Daten der zugehörige Mittelwert näherungsweise normalverteilt ist, mit der Folge, dass Verfahren auf der Basis der Normalverteilung problemlos eingesetzt werden können. Selbst wenn dies nicht gewährleistet ist, lassen sich Verfahren zur Trendanalyse zumindest zur Bestimmung der Trendlinie in der Regel verwenden. In diesem Falle ist nur zu beachten, dass die Testergebnisse mit Vorsicht zu betrachten sind. Gleichwohl hat es sich eingebürgert, dass selbst bei gewissen Verletzungen der Modellannahmen diese Testergebnisse – in Ermangelung anderer Verfahren – akzeptiert werden, zumindest dann, wenn ein entsprechendes Protokoll zur Vorgehensweise vorliegt.

2.2 Die Funktionen einer Trendanalyse

Vergleicht man die unterschiedlichen in der Literatur vorgeschlagenen Verfahren zur Trendanalyse, so wird schnell deutlich, dass jedes dieser Verfahren seine spezifischen Vor- und Nachteile hat. Die Eignung eines Verfahrens ist nicht nur von der statistischen Verteilung der Daten abhängig zu machen, sondern auch vom angestrebten Zweck der Trendanalyse. Diese Zusammenhänge sind Gegenstand von Kapitel 3. Es handelt sich dabei um eine Überarbeitung eines Textes von ICES-WGSAEM 1999. Dieser Text wurde in den Abschlussbericht aufgenommen, da neben der Darstellung der verschiedenen Funktionen einer Trendanalyse auch eine allgemeine Konzeption für die Erarbeitung eines Trendanalyseprotokolls vorgestellt wird.

Folgende Funktionen einer Trendanalyse von Monitoringzeitreihen sind von besonderem Interesse:

A) Visuelle Darstellung der Daten durch eine Trendlinie

Eine visuelle Darstellung beispielsweise mittels einer Trendlinie ermöglicht einerseits eine Zusammenfassung der Einzelwerte und andererseits die Aufdeckung von Anomalien, z.B. von Ausreißern.

B) Statistischer Trend-Test

Ein statistischer Trend-Test ermöglicht die Feststellung, ob von einem statistisch gesicherten Trend ausgegangen werden kann. Andernfalls spricht nichts gegen die Anfangsvermutung, dass es sich bei den beobachteten Schwankungen nur um zufällige Veränderungen handelt, die nicht durch einen systematischen Trend verursacht sind. Nur wenn ein statistisch gesicherter Trend vorliegt, erscheint aus statistischer Sicht eine weitergehende statistische Auswertung der Zeitreihe als zweckmäßig.

C) Bestimmung der linearen Trendkomponente für einen Anstieg oder Rückgang

Eine Trendanalyse sollte die Ermittlung des zugrundeliegenden linearen Trends in einem vorgegebenen Zeitraum ermöglichen.

D) Vergleich des gegenwärtigen Niveaus mit einem Referenzniveau oder dem Niveau eines früheren Zeitpunkts

Sofern ein Referenzwert oder ein Referenzzeitpunkt gegeben ist, stellt sich im Rahmen der Trendanalyse auch die Frage, ob das aktuelle Niveau höher oder niedriger als der Referenzwert liegt oder ob sich seit dem Referenzzeitpunkt eine Erhöhung oder Reduzierung ergeben hat.

Alle hier genannten Funktionen lassen sich mit dem LOESS-Smoother, dem Verfahren von Theil, dem Trend-Test auf der Basis des LOESS-Smoother sowie dem Test von Mann-Kendall realisieren. Jedoch lassen sich nicht alle Funktionen durch ein und dasselbe Analyseverfahren gleichzeitig in völlig befriedigender Weise erfüllen. So ist zur Erfassung eines linearen Trends ein Schätzverfahren angemessen, welches eine lineare Trendfunktion unterstellt. Ein solches Schätzverfahren ist jedoch nicht zweckmässig, um den generellen Trendverlauf abzubilden. Somit erscheint es sinnvoller, jede Funktion mit dem jeweils am besten geeigneten Analyseverfahren zu realisieren.

In den folgenden Abschnitten werden die verschiedenen Funktionen ausführlich diskutiert.

2.2.1 Visuelle Darstellung der Daten durch eine Trendlinie

Eine visuelle Trendlinie ermöglicht nicht nur eine Zusammenfassung der Daten, sondern erleichtert auch den visuellen Vergleich verschiedener Regionen oder Kontaminanten, wenn die zugehörigen Trendlinien innerhalb einer Grafik dargestellt werden. Anhand geeignet definierter Trendlinien lässt sich visuell entscheiden, ob es einen generellen Trend für verschiedene Regionen oder verschiedene Kontaminanten gibt, oder inwieweit spezifische Unterschiede vorliegen. Zusätzliche Informationen lassen sich durch die Einbeziehung von Konfidenz- und Vorhersagebereichen gewinnen.

Zur visuellen Darstellung eignen sich Glättungsverfahren, also insbesondere der LOESS-Smoother, während eine einfache polynomiale Regression vielfach zu einer schlechten Modellanpassung führt. Der Auswahl eines Verfahrens sind letztlich kaum Grenzen gesetzt, denn es gibt kein statistisches Kriterium, welches in eindeutiger Weise ein bestimmtes Verfahren bevorzugt.

Wenn mit einzelnen sehr stark abweichenden Datenwerten (Jahreswerten) zu rechnen ist, stellt sich die Frage, ob diese mit ihrem vollen Gewicht in die Bestimmung der Trendlinie einbezogen

werden sollen, oder ob ihr Einfluss mit einem geeigneten Robustifizierungsverfahren reduziert werden soll. Abbildung 1 zeigt für eine hypothetische Zeitreihe, die in Jahr $x=5$ einen stark abweichenden Einzelwert aufweist, eine mit einem LOESS-Smoother ermittelte Trendlinie. Offenbar 'zieht' der stark abweichende Einzelwert in $x=5$ die Trendlinie nach 'oben', so dass für die Jahre $x=3, 4, 6$ und 7 die Trendlinie oberhalb der jeweiligen Einzelwerte verläuft. Wenn ein Grund zu der Annahme besteht, dass der stark abweichende Einzelwert in Jahr $x=5$ nicht die realen Verhältnisse repräsentiert, sondern durch Messfehler verursacht wurde, liegt es nahe, durch eine geeignete Robustifizierung des Glättungsverfahrens den Einfluss dieses potentiellen Ausreißers zu reduzieren. Dies kann durch eine iterative Anwendung eines gewichteten Glättungsverfahrens erfolgen, bei dem die Gewichte von der Größe der jeweiligen Residuen bestimmt sind. Die Wirkung eines solchen robustifizierten Glättungsverfahrens ist in Abbildung 2 verdeutlicht. Offenbar hat der Wert im Jahr $x=5$ nahezu keinen Einfluss mehr, so dass sich nun eine konvexe Form der Trendfunktion ergibt, die den übrigen Einzelwerten sehr gut angepasst ist.

Vor- und Nachteile einer Robustifizierung der Trendermittlung sind für jeden Anwendungsfall abzuwiegen. Die zugrundeliegende Verteilung der Messfehler ist ebenso zu berücksichtigen wie die Frage nach dem jeweiligen genauen Zweck der Trendberechnungen. Bislang werden hauptsächlich nicht-robustifizierte Glättungsverfahren eingesetzt, was allerdings auch dadurch begründet ist, dass robustifizierte Glättungsverfahren in statistisch-methodischer Hinsicht nicht mehr so einfach zu handhaben sind und die Berechnung von Konfidenz- oder Vorhersageintervallen nicht geklärt ist.

Abb. 1: LOESS-Trend bei einer Zeitreihe mit einem stark abweichenden Wert

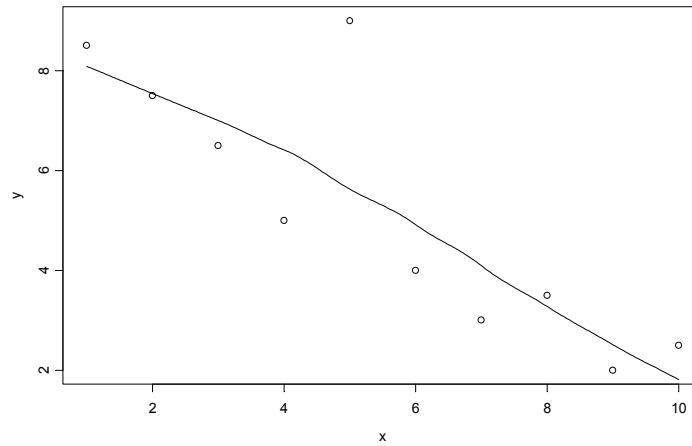
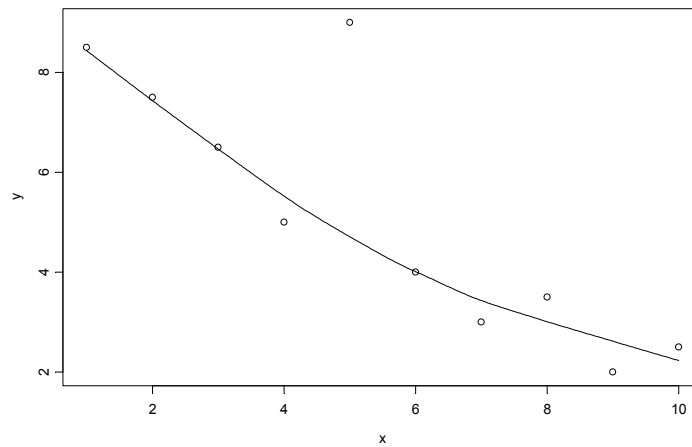


Abb. 2: Robustifizierter LOESS-Trend bei einer Zeitreihe mit einem stark abweichenden Wert



2.2.2 Statistischer Trend-Test

Solange die Zeitreihe durch zufällige Schwankungen von Jahr zu Jahr dominiert wird und kein deutlicher Trend zu erkennen ist, ist die Ermittlung einer Trendlinie und daraus abgeleiteter Größen von zweifelhaftem Nutzen, wenn nicht nachgewiesen wird, dass sich im der Trendverlauf reale Veränderungen zeigen. Die Überprüfung dieser Frage erfolgt mit Hilfe eines statistischen Trend-Tests. Ein solcher Test basiert auf der Anfangsvermutung (Nullhypothese), dass kein systematischer Trend vorliegt, und dass alle Schwankungen nur durch zufällige

Veränderungen der Einzelwerte verursacht werden. Erst wenn durch einen geeigneten Trend-Test zu einer vorgegebenen Irrtumswahrscheinlichkeit α (meist wird α mit 5 % festgelegt) diese Anfangsvermutung widerlegt werden kann, kann der Trend als signifikant angesehen werden, d.h. als nachgewiesen gelten.

Die Auswahl eines geeigneten Trend-Tests ist von einer Reihe von Aspekten abhängig zu machen. Dies betrifft zum einen die Frage, welche Arten eines systematischen Trends von Interesse sind, also die Frage, ob nur lineare oder auch nichtlineare Trends zu berücksichtigen sind, und ob neben monotonen Trends auch nichtmonotone Trends in die Betrachtungen einzubeziehen sind. Weiterhin ist zu berücksichtigen, ob die statistischen Verteilungsannahmen für den jeweiligen Trend-Tests erfüllt sind (z.B. Normalverteilung) und inwieweit der Test als ausreißerrobust angesehen werden kann. Nicht zuletzt sollte die jeweilige Power oder Trennschärfe eines Tests berücksichtigt werden, also die Wahrscheinlichkeit, mit der ein tatsächlich vorhandener Trend auch erkannt wird.

Ausgehend von den Tests von Mann-Kendall und dem auf dem LOESS-Smoother basierenden Trend-Test, gibt die folgende Tabelle einen groben Überblick über die Power der folgenden Testverfahren für verschiedene Trendverläufe.

1. Trend-Test von Mann-Kendall: ein klassischer nichtparametrischer Trend-Test
2. Compound Mann-Kendall Test: ein auf dem Test-Prinzip des Mann-Kendall basierender Trend-Test
3. Smoother Test: ein auf dem LOESS-Smoother basierender Test
4. Robustifizierter Smoother Test: ein auf dem robustifizierten LOESS-Smoother basierender Test.

Details zu den genannten Verfahren findet man in der unten genannten Literatur.

Tab. 1: Eignung verschiedener Trendtests

	Power bei Vorliegen der Normalverteilung. Anmerkung: Im Falle von Abweichungen von der Normalverteilung kann die Power sowohl höher als auch niedriger ausfallen. Die Angaben sind daher als Richtwert zu betrachten, da die Normalverteilung nicht als Voraussetzung der statistischen Methodik betrachtet werden kann			Robust
	Linearer Trend □□□□□□	Monotoner Trend □□□□□□□□	Nicht-monotoner Trend □□□□□□	
Mann-Kendall □□□□□□□□	Sehr gut □□	Befriedigend □□□□□□□□	Schlecht □□□□□□□□	Ja □□
Compound □□□□□□□□□□	Gut	Befriedigend	Befriedigend	Ja
Mann-Kendall □□□□□□□□	□□	□□□□□□	□□□□□□	□□□□
Smoother □□□□□□□□□□	Gut □□	Sehr gut □	Sehr gut □	Nein □□
Robustifizierter □□□□□□ Smoother	Gut	?	?	Ja

Für den auf dem robustifizierten Smoother basierenden Test konnte noch keine abschliessende Einstufung vorgenommen werden, da die Eigenschaften robuster Verfahren noch weiterer detaillierter Untersuchungen bedürfen.

2.2.3 Bestimmung der linearen Trendkomponente

Neben dem allgemeinen (nichtlinearen) Trendverlauf besteht vielfach auch Interesse an dem linearen Anteil des Trends in einer vorgegebenen Zeitspanne. Ausgedrückt werden kann dieser lineare Anteil durch die (positive oder negative) Steigung der linearen Trendlinie oder als prozentualen Anstieg oder Rückgang. Zusätzlich ist die Angabe eines Konfidenzintervalls für den Steigungsparameter hilfreich.

Die Bestimmung des Steigungsparameters kann entweder mittels einer linearen Regression erfolgen oder mittels der nichtparametrischen Methode von Theil. Eine Robustifizierung der linearen Regressionsmethode bietet sich an, wenn mit Ausreißern aufgrund von Messfehlern zu rechnen ist. Man sollte sich dessen bewusst sein, dass speziell bei stark nichtlinearen Trendverläufen die verschiedenen Methoden zu sehr unterschiedlichen Ergebnissen führen können.

2.2.4 Vergleich des gegenwärtigen Niveaus mit einem Referenzniveau oder dem Niveau zu einem Referenzzeitpunkt

Wenn ein Referenzwert oder ein Referenzzeitpunkt vorliegt, kann ein Vergleich mit dem gegenwärtigen Niveau vorgenommen werden. Hierzu bildet man die Differenz zwischen dem auf der Basis der Trendlinie geschätzten aktuellen Niveau und dem Referenzwert bzw. dem geschätzten Niveau zum Referenzzeitpunkt. Um eine statistische Bewertung dieser Differenz vornehmen zu können, wird zusätzlich eine Abschätzung der Standardabweichung dieser Differenz oder eine Konfidenzschätzung benötigt. Zusätzlich könnte ein statistischer Test vorgenommen werden, um zu überprüfen, ob die ermittelte Differenz signifikant von Null verschieden ist. Allerdings zeigen Simulationsrechnungen, dass die Trennschärfe eines solchen Tests vielfach unbefriedigend ist. Dies verweist auf den Umstand, dass die verwendeten Schätzmethoden vielfach hohe Schätzfehler verursachen. Verbesserungen der statistischen Methodik sind deshalb wünschenswert.

2.2.5 Empfehlungen

Wenn in erster Linie monotone Trends zu untersuchen sind, wird die Verwendung des LOESS-Smoother zur visuellen Darstellung des zugrundeliegenden Trends empfohlen, weiterhin der statistische Trend-Test von Mann-Kendall, sowie zur Ermittlung des linearen Trends die Methode von Theil. Zum Vergleich des aktuellen Niveaus mit dem Referenzniveau wird der Einsatz des LOESS-Smoother empfohlen.

Wenn hingegen auch nichtmonotone Trends zu betrachten sind, wird anstelle des Trend-Tests von Mann-Kendall die Anwendung des auf dem LOESS-Smoother basierenden Testverfahrens empfohlen, während die Bestimmung des linearen Trends nicht mittels des Verfahrens von Theil, sondern mittels des gewöhnlichen Regressionskoeffizienten für eine lineare Regression erfolgen sollte.

Beide Empfehlungen sind nicht als abgeschlossen oder endgültig anzusehen. Je nach der weiteren Entwicklung des weiteren methodischen Instrumentariums ist damit zu rechnen, dass in der Zukunft weitere statistische Verfahren zum Einsatz kommen können.

Dies betrifft nicht nur den Einsatz robuster Verfahren, sondern auch den Einsatz von Verfahren, welche nicht auf Jahreswerten basieren, sondern welche unmittelbar auf Einzelwerten basieren.

2.3 Analyse von Jahresdaten

In diesem Abschnitt wird der LOESS Smoother zur Analyse von Jahreswerten beschrieben. Hinsichtlich des theoretischen Hintergrundes sei auf die Arbeit von Fryer und Nicholson (1999) sowie auf die Monographie von Hastie und Tibshirani (1990) verwiesen. Implementiert wurde der LOESS-Smoother in dem Statistikpaket S-PLUS, sowie in dem sogenannten „Trend-y-Tector“ der Niederlande, der zur Analyse mariner Jahresdaten entwickelt wurde. Das letztgenannte Programm ist eine Internet-Anwendung, die unter www.waterland.net (unter Monitoring – OSPAR-Workshop) aufgerufen werden kann. In anderen Statistik-Paketen sind zum Teil ebenfalls verschiedene Glättungsverfahren verfügbar, doch werden vielfach andere Glättungsmethoden eingesetzt.

Grundsätzlich ist der LOESS-Smoother ein flexibles Instrument zur Trendanalyse, welches zur Erfassung von Trends in marinen Daten eingesetzt werden kann. Einschränkend ist allerdings darauf zu verweisen, dass die Trendanalyse nach einem jeweils zu spezifizierenden Protokoll erfolgen sollte, welches die Funktionen der Trendanalyse ebenso wie die Natur der jeweiligen Daten berücksichtigt.

Zu erwähnen ist an dieser Stelle, dass in WGSaEM 2001 verschiedene Erweiterungen des LOESS-Smoother zur Einbeziehung der unterschiedlichen Genauigkeit der einzelnen Werte vorgeschlagen werden. Die vorgeschlagenen Methoden sind allerdings noch nicht abschließend diskutiert, insbesondere da die Festlegung des jeweiligen „Genauigkeitsgrades“ erhebliche methodische Probleme nach sich zieht. Es wird daher darauf verzichtet, die Vorschläge im einzelnen vorzustellen.

2.3.1 Smoother-Matrix

Der LOESS-Smoother ist ein linearer Smoother, d.h. es gibt eine quadratische Smoother-Matrix S , welche durch Multiplikation mit dem Beobachtungsvektor

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{(Gleichung 14)}$$

Die Breite von Δ_t steuert die Glättung, wobei von Nicholson und Fryer (1999) vorgeschlagen wird, Δ_t so festzulegen, dass die Zeitspanne der jeweils einbezogenen Daten gleich bleibt, d.h.

$$\Delta_t = \max \left\{ \frac{\text{span} + 1}{2}, \text{span} - \min\{t - 1, n - t\} \right\}. \quad (\text{Gleichung 16})$$

Mit der Designmatrix X des einfachen Regressionsmodells

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad (\text{Gleichung 17})$$

und der Gewichtungsmatrix

$$W_t = \begin{pmatrix} w_{t1} & 0 & \cdots & 0 \\ 0 & w_{t2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_{tn} \end{pmatrix} \quad (\text{Gleichung 18})$$

ergibt sich der LOESS-Trend zum Zeitpunkt x_t durch $S'_t Y$, wobei S'_t die i -te Zeile von $X(X'W_tX)^{-1}X'W_t$ ist. Die sich daraus ergebende Smoother-Matrix kann dargestellt werden als

$$S = \begin{pmatrix} S'_1 \\ \vdots \\ S'_n \end{pmatrix}. \quad (\text{Gleichung 19})$$

Es ist anzumerken, dass die Güte der Trendschätzung bei fehlenden Werten sehr stark in Mitleidenschaft gezogen werden kann. Dies gilt insbesondere dann, wenn in einem bestimmten Zeitabschnitt sehr viele Werte fehlen. Die Trendlinie kann in diesem Fall sehr stark von einzelnen Werten beeinflusst sein, so dass sich leicht ein scheinbarer Trend ergeben kann. Das Verfahren sollte deshalb – in der vorliegenden Form – nur dann eingesetzt werden, wenn höchstens 1 bis 2 Werte in der Reihe fehlen.

2.3.2 Tests basierend auf dem LOESS Smoother

2.3.2.1 Kontrast-Test

Der auf dem Kontrast zwischen Anfang und Ende der Trendlinie basierende sogenannte Kontrast-Test ist in kommerziell verfügbaren Programmen nicht verfügbar und muss deshalb selbst programmiert werden. Die zugrundeliegenden Größen sind jedoch in einigen Programmen (SPLUS) verfügbar, so dass die Programmierung mit vertretbarem Aufwand möglich ist. Implementiert wurde der Kontrast-Test in dem frei verfügbaren Softwareprogramm Trend-y-tector.

Die Differenz $z(n)-z(1)$ zwischen dem Trendwert zu Beginn und am Ende der Zeitreihe wird auf Basis des Kontrast-Tests überprüft mittels der Teststatistik

$$t = \frac{z(n) - z(1)}{s \sqrt{\sum_{i=1}^n (S_{1i} - S_{ni})^2}} . \quad (\text{Gleichung 20})$$

Dabei bezeichnet S_{ki} das Element (k,i) der Smoother-Matrix S und s die Residualstandardabweichung für den Smoother:

$$s = \sqrt{\frac{1}{df_{error}} \sum_{t=1}^n (y(t) - z(t))^2} \quad (\text{Gleichung 21})$$

mit

$$df_{error} = n - tr(2S - SS') = 3.7224 + 0.7450 \cdot (n - 6) \quad (\text{Gleichung 22})$$

Freiheitsgraden. Die letztgenannte Gleichung gilt unter der Annahme, dass keine fehlenden Werte vorliegen. Die Nullhypothese wird abgelehnt falls:

$$H_1: t < -t_{1-\alpha} \quad (\text{Rückgang})$$

$$H_2: t < -t_{1-\alpha/2} \text{ oder } t > t_{1-\alpha/2} \quad (\text{Rückgang oder Anstieg})$$

$$H_3: t > t_{1-\alpha} \quad (\text{Anstieg})$$

wobei t_p das p-Quantil der t-Verteilung mit df_{error} Freiheitsgraden bezeichnet.

2.3.2.2 Varianzanalyse der Quadratsumme

Tests für den linearen und den nicht-linearen Anteil des Trendverlaufs können abgeleitet werden von der von Fryer und Nicholson (1999) beschriebenen ANOVA-Tabelle. Hierzu betrachtet man die Quadratsumme

$$SS_{total} = \sum_{t=1}^n (y(x_t) - \bar{y})^2 \quad (\text{Gleichung 23})$$

$$SS_{linerror} = \sum_{t=1}^n (y(x_t) - (a + bx_t))^2 \quad (\text{Gleichung 24})$$

$$SS_{error} = \sum_{t=1}^n (y(x_t) - z(x_t))^2 \quad (\text{Gleichung 25})$$

mit den Freiheitsgraden

$$df_{total} = n - 1 \quad (\text{Gleichung 26})$$

$$df_{linerror} = n - 2 \quad (\text{Gleichung 27})$$

$$df_{error} = n - tr(2S - SS') \quad (\text{Gleichung 28})$$

und den Regressionskoeffizienten a und b . df_{error} ist bestimmt durch den Erwartungswert von SS_{error} unter der Annahme unabhängiger und normalverteilter Zufallsabweichungen mit der Varianz σ^2 :

$$\begin{aligned} E[SS_{error} / \sigma^2] &= E[(y'(I - S')(I - S)y) / \sigma^2] = tr[(I - S)(I - S')] \\ &= tr(I) - 2tr(S) + tr(SS'), \end{aligned} \quad (\text{Gleichung 29})$$

Dabei bezeichnet y den Vektor der beobachteten Daten und I die Einheitsmatrix. Es ist festzuhalten, dass SS_{error} / σ^2 nicht exakt Chi-Quadrat-verteilt ist. Jedoch kann die Verteilung durch eine Chi-Quadrat-Verteilung mit df_{error} Freiheitsgraden approximiert werden.

Die totale Quadratsumme SS_{total} lässt sich in drei Komponenten zerlegen

$$SS_{total} = SS_{lin} + SS_{nonlin} + SS_{error} \quad (\text{Gleichung 30})$$

wobei

$$SS_{lin} = SS_{total} - SS_{linerror} \quad (\text{Gleichung 31})$$

mit

$$df_{total} - df_{nonlin} = 1 \quad (\text{Gleichung 32})$$

Freiheitsgrad und

$$SS_{nonlin} = SS_{linerror} - SS_{error} \quad (\text{Gleichung 33})$$

mit

$$df_{linerror} - df_{error} = tr(2S - SS') - 2 \quad (\text{Gleichung 34})$$

Freiheitsgrade.

Diese Quadratsumme kann nun verwendet werden, um zu überprüfen, ob ein linearer oder ein nicht-linearer Trend vorliegt. Die genannte Zerlegung ist in kommerziellen Programmen leider nicht verfügbar und muss selbst programmiert werden. Im Programm SPLUS werden jedoch eine Reihe der zugrundeliegenden Größen bereitgestellt, so dass eine Programmierung relativ leicht realisiert werden kann.

2.3.2.3 Test auf linearen Trend

Ob ein linearer Trend vorliegt, kann mittels der Teststatistik

$$F_{lin} = \frac{SS_{lin}}{SS_{error}} \frac{df_{error}}{1} \quad (\text{Gleichung 35})$$

überprüft werden. Auch diese Teststatistik ist nicht in kommerziellen Statistikpaketen verfügbar. Die Teststatistik ist – unter der Annahme von Normalverteilung und stochastischer Unabhängigkeit F-verteilt mit (1,n-2) Freiheitsgraden, sofern kein Trend vorliegt. Die Verwendung dieser Teststatistik ist gleichwertig mit einer Verwendung der t Statistik

$$t = \frac{b}{s} \sqrt{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (\text{Gleichung 36})$$

mit df_{error} Freiheitsgraden und dem Regressionskoeffizienten b

$$b = \frac{\sum_{t=1}^n x_t y_t - \frac{1}{n} \left(\sum_{t=1}^n x_t \right) \left(\sum_{t=1}^n y_t \right)}{\sum_{t=1}^n x_t^2 - \frac{1}{n} \left(\sum_{t=1}^n x_t \right)^2} \quad (\text{Gleichung 37})$$

für den Anstieg. Die Nullhypothese wird abgelehnt falls:

$$H_1: t < -t_{1-\alpha} \quad (\text{Rückgang})$$

$$H_2: t < -t_{1-\alpha/2} \text{ oder } t > t_{1-\alpha/2} \quad (\text{Rückgang oder Anstieg})$$

$$H_3: t > t_{1-\alpha} \quad (\text{Anstieg})$$

wobei t_p das p-Quantil der t-Verteilung mit df_{error} Freiheitsgrade bezeichnet.

Dieser Trendtest ist dem oben beschriebenen Kontrasttest sehr ähnlich.

2.3.3 Test auf nicht-linearen Trend

Ob ein nicht-linearer Trend vorliegt, kann mittels der Teststatistik

$$F_{nonlin} = \frac{SS_{nonlin}}{SS_{error}} \frac{df_{error}}{df_{linerror} - df_{error}} = \left(\frac{SS_{linerror}}{SS_{error}} - 1 \right) \frac{df_{error}}{df_{linerror} - df_{error}} \quad (\text{Gleichung 38})$$

überprüft werden. Diese Teststatistik lässt sich interpretieren als ein Mass für den Anstieg der Residualvarianz, sofern anstelle eines Smoothers ein lineares Regressionsmodell verwendet wird. Falls die Modellanpassung mit einem Smoother nur geringfügig besser als mit einer linearen Regression ist, liegt F_{nonlin} nahe bei 0. Der Test auf einen nichtlinearen Trend ist, da es sich um ein methodisch neues Verfahren handelt, ebenfalls nicht in kommerziellen Programmen verfügbar und aufwendiger zu programmieren, als die weiter oben angegebenen Tests.

2.3.3.1 Approximation erster Ordnung

Unter der Annahme von Linearität (und Unabhängigkeit und Normalverteilung) kann die Verteilung von SS_{nonlin} / σ^2 angenähert werden durch eine Chi-Quadrat Verteilung mit $df_{linerror} - df_{error}$ Freiheitsgraden. Daher kann die Verteilung der Prüfgröße F_{nonlin} durch the F-Verteilung mit $(df_{linerror} - df_{error}, df_{error})$ Freiheitsgraden approximiert werden. Daraus folgt, dass die Nullhypothese (dass kein nicht-linearer Trend vorliegt) abzulehnen ist, sofern

$$F_{nonlin} > F(df_{linerror} - df_{error}, df_{error}, 1 - \alpha) \quad (\text{Gleichung 39})$$

gilt. Dabei bezeichnet $F(df_1, df_2, 1 - \alpha)$ das $(1 - \alpha)$ -Quantil der F-Verteilung mit (df_1, df_2) Freiheitsgraden. Dieses Kriterium wird im folgenden als Kriterium erster Ordnung bezeichnet. Es ist festzuhalten, dass beträchtliche Differenzen zwischen diesem Kriterium erster Ordnung und dem “wahren” kritischen Wert bestehen können. In Tabelle 2 sind diese kritischen Werte gegenübergestellt. Der wahre kritische Wert wurde in einer Simulationsstudie ermittelt. Fryer und Nicholson (1999) schlagen eine verbesserte Approximation vor, doch auch damit können noch erhebliche Abweichungen auftreten.

Tab. 2: Kritische Werte für den Test auf Nichtlinearität

n	$df_{linerror} - df_{error}$	df_{error}	Kritischer Wert erster Ordnung	Wahrer kritischer Wert	Kritischer Wert erster Ordnung	Wahrer kritischer Wert
			$\alpha=0.05$		$\alpha=0.01$	
10	1.30	6.70	5.347	4.456	11.454	8.438
20	3.85	14.15	3.132	2.934	5.076	4.496

2.3.3.2 Approximation zweiter Ordnung

Eine Approximation zweiter Ordnung ergibt sich wie folgt. Die Verteilung der Quadratsumme SS_{error} kann angenähert werden durch

$$\frac{\text{tr}(K_{error})}{\sigma^2 \text{tr}(K'_{error} K_{error})} SS_{error} \sim \chi \left(\frac{\text{tr}(K_{error})^2}{\text{tr}(K'_{error} K_{error})} \right) \quad (\text{Gleichung 40})$$

wobei K_{error} die Quadratmatrix ist, welche durch die quadratische Form

$$SS_{error} = y'(I - S)'(I - S)y =: y'K_{error}y \quad (\text{Gleichung 41})$$

definiert ist. $\chi(df)$ bezeichnet dabei the Chi-Quadrat-Verteilung mit df Freiheitsgraden. Die Verteilung der Quadratsumme SS_{nonlin} kann angenähert werden durch

$$\frac{\text{tr}(K_{nonlin})}{\sigma^2 \text{tr}(K'_{nonlin} K_{nonlin})} SS_{nonlin} \sim \chi(\text{tr}(K_{nonlin})^2 / \text{tr}(K'_{nonlin} K_{nonlin})) \quad (\text{Gleichung 42})$$

wobei K_{nonlin} jene Quadratmatrix bezeichnet, welche durch die quadratische Form

$$SS_{nonlin} = y'(I - L)'(I - L)y - y'(I - S)'(I - S)y = y'(2S - S'S - L)y =: y'K_{nonlin}y \quad (\text{Gleichung 43})$$

mit der Regressions-Matrix

$$L = X(X'X)^{-1}X'. \quad (\text{Gleichung 44})$$

bestimmt ist. Dabei bezeichnet X die Design-Matrix

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \quad (\text{Gleichung 45})$$

Mit dieser Approximation zweiter Ordnung ergibt sich die Verteilung der Prüfgröße

$$F_{nonlin}^* = \frac{SS_{nonlin} / df_1}{SS_{error} / df_2} = \left(\frac{SS_{linerror}}{SS_{error}} - 1 \right) \frac{df_2}{df_1}, \quad (\text{Gleichung 46})$$

mit

$$df_1 = \frac{\text{tr}(K_{nonlin})^2}{\text{tr}(K'_{nonlin} K_{nonlin})}, \quad df_2 = \frac{\text{tr}(K_{error})^2}{\text{tr}(K'_{error} K_{error})}, \quad (\text{Gleichung 47})$$

approximativ durch

$$\frac{\frac{\text{tr}(K_{error})}{\text{tr}(K'_{error} K_{error})}}{\frac{\text{tr}(K_{nonlin})}{\text{tr}(K'_{nonlin} K_{nonlin})}} F_{nonlin}^* \sim F(df_1, df_2). \quad (\text{Gleichung 48})$$

$F(df_1, df_2)$ bezeichnet dabei die F-Verteilung mit (df_1, df_2) Freiheitsgraden. Die Nullhypothese (dass kein nicht-linearer Trend vorliegt) wird abgelehnt, falls

$$F_{nonlin}^* > \frac{\frac{\text{tr}(K_{nonlin})}{\text{tr}(K'_{nonlin} K_{nonlin})}}{\frac{\text{tr}(K_{error})}{\text{tr}(K'_{error} K_{error})}} F(df_1, df_2, 1-\alpha), \quad (\text{Gleichung 49})$$

wobei $F(df_1, df_2, 1-\alpha)$ das $(1-\alpha)$ -Quantil der F-Verteilung mit (df_1, df_2) Freiheitsgraden bezeichnet.

Eine Gegenüberstellung des kritischen Wertes zweiter Ordnung zum wahren kritischen Wert erfolgt in Tabelle 3. Es zeigt sich, dass die Abweichungen erheblich geringer ausfallen als beim Kriterium erster Ordnung.

Tab. 3: Kritische Werte für den Test auf Nichtlinearität, Approximation zweiter Ordnung

n	df_1	df_2	$\frac{\text{tr}(K_{nonlin})}{\text{tr}(K'_{nonlin} K_{nonlin})}$	$\frac{\text{tr}(K_{error})}{\text{tr}(K'_{error} K_{error})}$	Kritische Werte $\alpha=0.05$		Kritische Werte $\alpha=0.01$	
					Kritischer Wert erster Ordnung	Wahrer kritischer Wert	Kritischer Wert erster Ordnung	Wahrer kritischer Wert
10	1.484	6.869	0.875	0.976	4.593	4.456	9.61	8.438
20	4.067	14.36	0.946	0.985	2.956	2.934	4.756	4.496

Die grundsätzliche Vorgehensweise zur Überprüfung auf Nichtlinearität gestaltet sich damit wie folgt:

1. Berechne

$$df_1 = \frac{\text{tr}(K_{\text{nonlin}})^2}{\text{tr}(K'_{\text{nonlin}} K_{\text{nonlin}})} \quad (\text{Gleichung 50})$$

und

$$df_2 = \frac{\text{tr}(K_{\text{error}})^2}{\text{tr}(K'_{\text{error}} K_{\text{error}})} \quad (\text{Gleichung 51})$$

2. Berechne den kritischen Wert

$$\frac{\frac{\text{tr}(K_{\text{nonlin}})}{\text{tr}(K'_{\text{nonlin}} K_{\text{nonlin}})}}{\frac{\text{tr}(K_{\text{error}})}{\text{tr}(K'_{\text{error}} K_{\text{error}})}} F(df_1, df_2, 1 - \alpha) \quad (\text{Gleichung 52})$$

3. Berechne die Prüfgröße

$$F^*_{\text{nonlin}} = \frac{SS_{\text{nonlin}} / df_1}{SS_{\text{error}} / df_2} = \left(\frac{SS_{\text{linerror}}}{SS_{\text{error}}} - 1 \right) \frac{df_2}{df_1} \quad (\text{Gleichung 53})$$

Falls die Prüfgröße den kritischen Wert übersteigt, ist die Nullhypothese (wonach kein nicht-linearer Trend vorliegt) abzulehnen.

Diese Überprüfung müsste selbst programmiert werden, da die Berechnungen nicht in kommerziellen Programmen erhältlich sind.

2.3.4 Vergleich mit einem Sollwert

Um den aktuellen Trendwert mit einem vorgegebenen Sollwert θ zu vergleichen, liegt es nahe, für den Trendwert ein Konfidenzintervall zu ermitteln. Hierfür gibt es mehrere Berechnungsmöglichkeiten, die in diesem Abschnitt beschrieben werden. Die Berechnungen basieren auf dem Konzept des allgemeinen additiven Modells von Hastie und Tibshirani (1990).

2.3.4.1 Konfidenzintervall für das lineare Regressionsmodell

Das klassische $(1-2\alpha)$ Konfidenzintervall für den linearen Trendwert im Jahr m kann geschrieben werden als

$$[a + bm - t(n-2, 1-\alpha) s_{linerror} \delta(m) \quad , \quad a + bm + t(n-2, 1-\alpha) s_{linerror} \delta(m)]$$

(Gleichung 54)

wobei $a+bm$ die Regressionsgerade mit den Regressionskoeffizienten bezeichnet,

$$a = \frac{1}{n} \sum_{t=1}^n y_t - \frac{b}{n} \sum_{t=1}^n x_t \quad ,$$

(Gleichung 55)

$$b = \frac{\sum_{t=1}^n x_t y_t - \frac{1}{n} \left(\sum_{t=1}^n x_t \right) \left(\sum_{t=1}^n y_t \right)}{\sum_{t=1}^n x_t^2 - \frac{1}{n} \left(\sum_{t=1}^n x_t \right)^2} \quad ,$$

(Gleichung 56)

$s_{linerror}$ die Residualstandardabweichung im linearen Regressionsmodell,

$$s_{linerror} = \sqrt{\frac{1}{n-2} SS_{linerror}} \quad ,$$

(Gleichung 57)

und $\delta(m)$ den Stichprobenfaktor

$$\delta(m) = \sqrt{\frac{1}{n} + \frac{\left(m - \frac{1}{n} \sum_{t=1}^n x_t \right)^2}{\sum_{t=1}^n x_t^2 - \frac{1}{n} \left(\sum_{t=1}^n x_t \right)^2}} \quad .$$

(Gleichung 58)

Dieses Konfidenzintervall kann zur Überprüfung der folgenden Hypothesen eingesetzt werden:

(i) Aufdecken kritischer Situationen

$$H_{00}: a+bm \leq \theta \text{ versus } H_{01}: a+bm > \theta$$

(ii) Nachweis, dass die Situation unkritisch ist

$$H_{10}: a+bm \geq \theta \text{ versus } H_{11}: a+bm < \theta$$

Eine kritische Situation (H_{01}) ist beim Signifikanzniveau α zu erkennen, falls die obere Begrenzung des Konfidenzintervalls niedriger als θ ist, d.h. falls

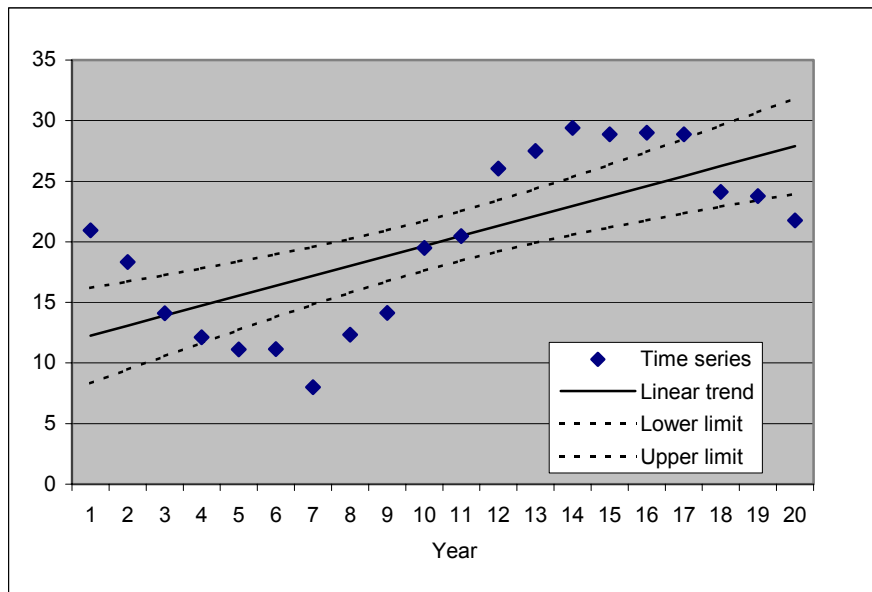
$$a + bm + t(n - 2, 1 - \alpha) s_{\text{linerror}} \delta(m) < \theta. \quad (\text{Gleichung 59})$$

Der Nachweis, dass die Situation unkritisch ist (H_{11}), ist beim Signifikanzniveau α zu erkennen, falls die untere Grenze des Konfidenzintervall höher als θ ist, d.h. falls

$$a + bm - t(n - 2, 1 - \alpha) s_{\text{linerror}} \delta(m) > \theta. \quad (\text{Gleichung 60})$$

Dieses Konfidenzintervall und die entsprechenden Tests sollten im Falle starker Nichtlinearität nicht eingesetzt werden. In Abbildung 3 wird ein Beispiel für solch eine Situation dargestellt.

Abb. 3: Linearer Trend mit 90%-Konfidenzintervall für eine nichtlineare Zeitreihe



Es ist festzuhalten, dass das hier beschriebene Konfidenzintervall auf der Annahme beruht, dass die Abweichungen vom linearen Trend stochastisch unabhängig und normalverteilt sind.

2.3.4.2 Ein auf dem LOESS-Smoother beruhendes Konfidenzintervall für den linearen Trend

Das approximative $(1-2\alpha)$ Konfidenzintervall für die lineare Trendkomponente im Jahr m kann geschrieben werden als

$$[a + bm - t(df_{error}, 1 - \alpha)s\delta(m) \quad , \quad a + bm + t(df_{error}, 1 - \alpha)s\delta(m)] . \quad (\text{Gleichung 61})$$

wobei s die Residualstandardabweichung des LOESS-Smoother bezeichnet.

Die Hypothesen wurden bereits im vorhergehenden Abschnitt beschrieben. Jedoch unterscheidet sich die Interpretation der Ergebnisse der Tests, da sich die Aussage nun nur auf die lineare Trendkomponente bezieht, und nichtlineare Trendkomponenten nicht berücksichtigt werden.

Eine kritische Situation ist bei Signifikanzniveau α nachgewiesen, falls die obere Grenze des Konfidenzintervall geringer als θ ist, d.h. falls

$$a + bm + t(df_{error}, 1 - \alpha)s\delta(m) < \theta . \quad (\text{Gleichung 62})$$

Der gute Zustand kann als nachgewiesen gelten, wenn die untere Grenze des Konfidenzintervalls größer ist als θ , d.h. wenn

$$a + bm - t(df_{error}, 1 - \alpha)s\delta(m) > \theta . \quad (\text{Gleichung 63})$$

Dieses Konfidenzintervall und die entsprechenden Tests beziehen sich auf die lineare Trendkomponente, welche einen Teil des systematischen Trends bildet. Es ist darauf hinzuweisen, dass das beschriebene Konfidenzintervall nicht in kommerziellen Programmen implementiert ist.

2.3.4.3 Konfidenzintervall für den auf dem LOESS-Smoother basierenden systematischen Trend

Das approximative $(1-2\alpha)$ Konfidenzintervall für den systematischen Trend im Jahr m kann geschrieben werden als

$$[z(m) - t(df_{error}, 1 - \alpha)s\tau(m) \quad , \quad z(m) + t(df_{error}, 1 - \alpha)s\tau(m)] . \quad (\text{Gleichung 64})$$

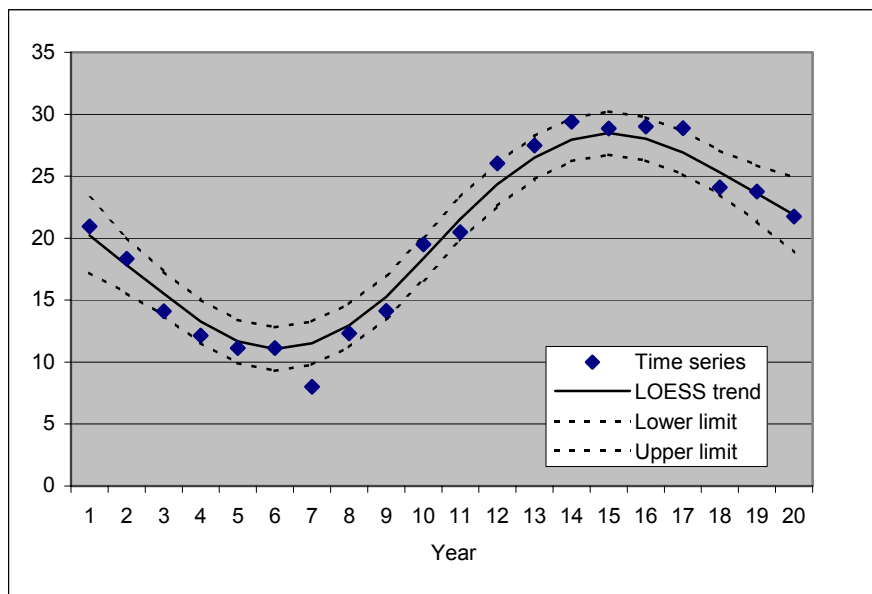
wobei

$$\tau(m) = \sqrt{(SS')_{mm}} . \quad (\text{Gleichung 65})$$

$(SS')_{mm}$ bezeichnet den m -ten Diagonalwert der Produktmatrix SS' . Dieses Konfidenzintervall wird von Fryer und Nicholson (1999) eingesetzt.

Ebenso wie bei den oben dargestellten Konfidenzintervallen kann sowohl der Nachweis des guten Zustands als auch die Aufdeckung kritischer Situationen von Interesse sein. Die Testentscheidungen erfolgen analog.

Abb. 4: Loess-Trend einer nicht-linearen Zeitreihe mit einem 90%-Konfidenzintervall



Wie beim linearen Trend kann auch der allgemeine LOESS-Trend täuschen. Daher ist zu empfehlen, die Ergebnisse zu vergleichen und zu überprüfen, ob sich die Resultate für die verschiedenen Konfidenzintervalle unterscheiden. Die Berechnung des Konfidenzintervalls kann mit dem Programmpaket SPLUS realisiert werden.

2.3.4.4 Vorhersageintervall auf der Basis des LOESS-Smoother

Das $(1-2\alpha)$ Vorhersageintervall im Jahr m kann geschrieben werden als

$$\left[z(m) - t(df_{error}, 1 - \alpha) s \gamma(m) \quad , \quad z(m) + t(df_{error}, 1 - \alpha) s \gamma(m) \right]. \quad (\text{Gleichung 66})$$

wobei

$$\gamma(m) = \sqrt{1 + (SS')_{mm}} \quad (\text{Gleichung 67})$$

und $(SS')_{mm}$ bezeichnet den m -ten Diagonalwert der Produktmatrix SS' .

Die Begrenzung des Intervalls kann folgendermassen interpretiert werden:

- (i) Aufdecken einer kritischen Situation:

$$z(m) - t(df_{error}, 1 - \alpha) s \gamma(m) > \theta \quad (\text{Gleichung 68})$$

Wenn die untere Grenze des Vorhersageintervalls den Sollwert θ übersteigt, übersteigen mehr als $(1 - \alpha) \times 100\%$ der Beobachtungen im statistischen Modell den Sollwert.

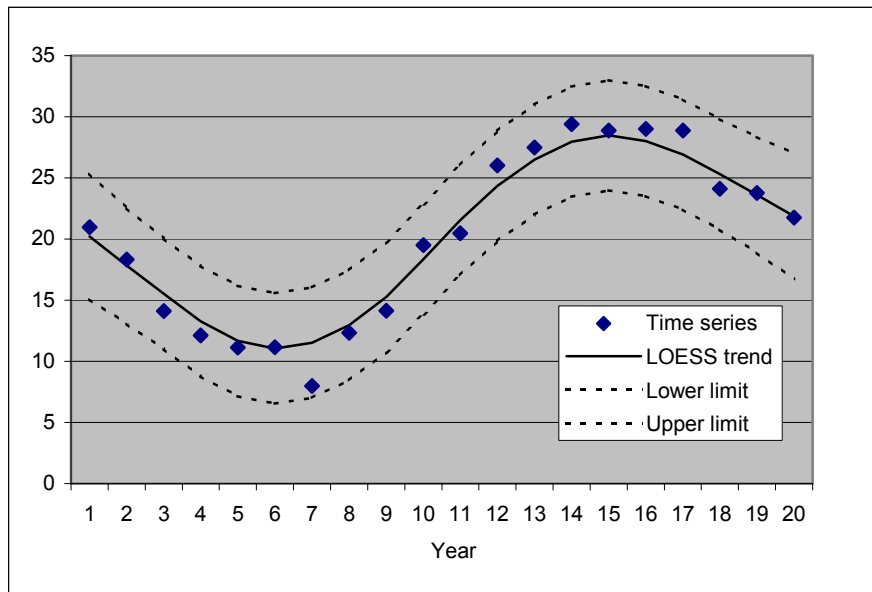
- (ii) Nachweis, das die Situation unkritisch ist:

$$z(m) - t(df_{error}, 1 - \alpha) s \gamma(m) < \theta \quad (\text{Gleichung 69})$$

Wenn die obere Grenze des Vorhersageintervalls unter dem Sollwert θ liegt, überschreiten weniger als $\alpha \times 100\%$ der Beobachtungen im statistischen Modell diesen Wert.

Abbildung 5 zeigt ein Beispiel für ein Vorhersageintervall.

Abb. 5: 90% Vorhersageintervall für eine nichtlineare Zeitreihe



Die Berechnung des Vorhersageintervalls in kommerziellen Programmen ist nicht unmittelbar möglich. Gleichwohl stellt das Programm SPLUS jene Größen bereit, mit der eine Berechnung des Vorhersageintervalls vergleichsweise schnell erfolgen kann.

2.3.5 Zur Differenz zwischen Anfang-Ende-Kontrast und linearem Regressionskoeffizienten

Die Entscheidung zwischen dem linearen Regressionskoeffizienten und der Differenz zwischen Anfang und Ende der geschätzten Trendlinie ist vom Zweck der Analyse sowie von der Länge der Zeitreihe abhängig zu machen. Beide Parameter können für die Trendanalyse nützlich sein und stehen in direktem Zusammenhang zur jährlichen Veränderungsrate des Trendniveaus.

2.3.5.1 Jährliche Veränderungsrate und der Anfang-Ende-Kontrast

Mit dem "wahren" Niveau $\mu_1, \mu_2, \dots, \mu_n$ in den Jahren $1, 2, \dots, n$ und den zugehörigen Trendschätzungen $z(1), \dots, z(n)$ erhält man eine Schätzung der jährlichen Veränderungsrate

$$\frac{\mu_n - \mu_1}{n - 1} \quad (\text{Gleichung 70})$$

durch

$$\frac{z(n) - z(1)}{n - 1}. \quad (\text{Gleichung 71})$$

2.3.5.2 Jahresveränderungsrate und der lineare Regressionskoeffizient

Der lineare Regressionskoeffizient kann interpretiert werden als eine Schätzung eines gewichteten Durchschnitts der jährlichen Veränderungsrate $\frac{\mu_{n-i+1} - \mu_i}{n - 2i + 1}$ für die gesamte Zeitreihe ($i=1$) und für folgende innere Abschnitte der Zeitreihe: von $t=2$ zu $t=n-1$, von $t=3$ zu $t=n-2$, von $t=4$ zu $t=n-3$ etc:

$$\left((n-1)^2 \frac{\mu_n - \mu_1}{n-1} + (n-3)^2 \frac{\mu_{n-1} - \mu_2}{n-3} + \dots + (n-2k+1)^2 \frac{\mu_{n-k+1} - \mu_k}{n-2k+1} \right) / m, \quad (\text{Gleichung 72})$$

mit

$$m = (n-1)^2 + (n-3)^2 + \dots + (n-2k+1)^2 \quad (\text{Gleichung 73})$$

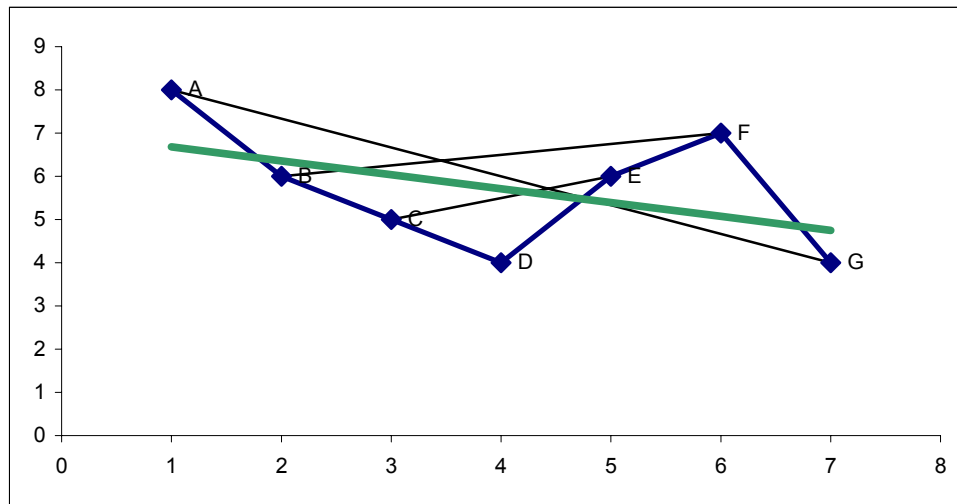
und

$$k = \begin{cases} \frac{n-1}{2} & k \text{ ungerade} \\ \frac{n}{2} & k \text{ gerade} \end{cases}. \quad (\text{Gleichung 74})$$

2.3.5.3 Beispiel

Abbildung 6 verdeutlicht für $n=7$ den gewichteten Durchschnitt der Veränderungsrate zwischen Anfang und Ende (von A bis G), zwischen dem 2. und 6. Jahr (von B bis F) und zwischen dem 3. und 5. Jahr (von C bis E).

Abb. 6: Darstellung des linearen Trends als gewichtetes Mittel der Veränderungsraten



Die entsprechenden Veränderungsraten und Gewichtungen sind in der folgenden Tabelle dargestellt. Der gewichtete Durchschnitt beträgt -0.321 . Dies entspricht dem Anstieg der grauschattierten Linie der linearen Regression in Abbildung 9.

Tab. 4: Berechnung des Regressionskoeffizienten als gewichtete Summe

	Kontrast	Anstieg	Gewicht	Anstieg x Gewicht
Von A to G:	-4	$-2/3$	$6^2/56=64\%$	-0.429
Von B to F:	1	0.25	$4^2/56=29\%$	0.071
Von C to E:	1	0.5	$2^2/56=7\%$	0.036
Summe:				-0.321

Scheinbar wird der lineare Regressionskoeffizient durch den Anfang-Ende-Kontrast dominiert, da bei $n=7$ dessen Gewichtsanteil fast $2/3$ beträgt.

Das bedeutet, dass für kurze Zeitreihen systematische Differenzen zwischen dem Anfang-Ende-Kontrast und der Schätzung des linearen Trends relativ gering sind.

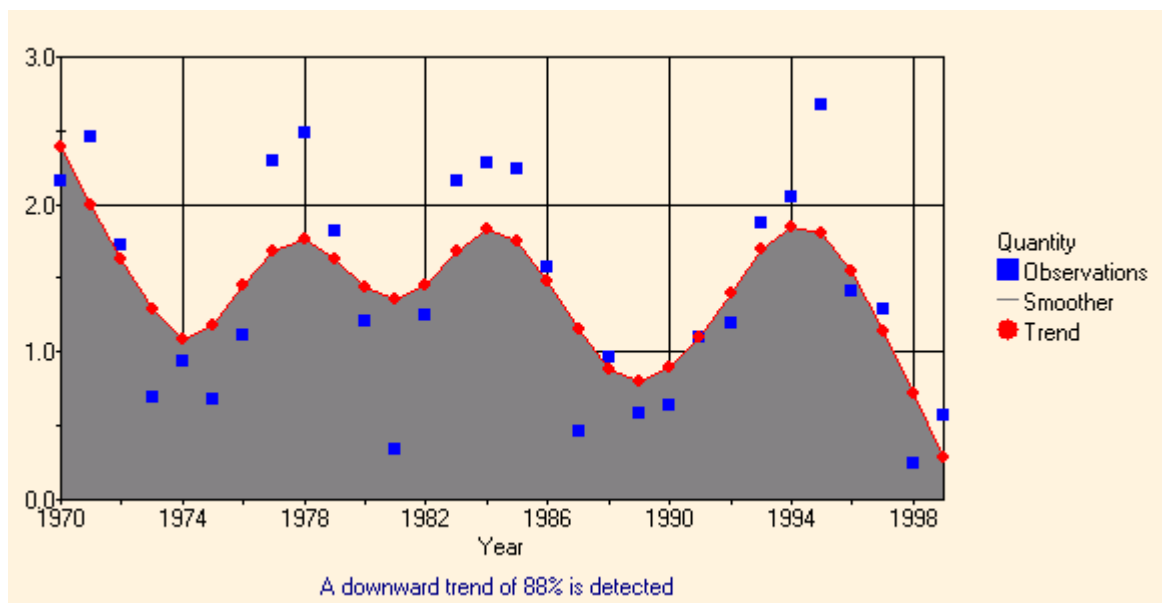
2.3.5.4 Verhalten in langen Zeitreihen

In langen Zeitreihen können erhebliche Differenzen zwischen dem Regressionskoeffizienten sowie dem Anfang-Ende-Kontrast auftreten.

Nicht-monotone Trendmuster

Wenn in langen Zeitreihen klimatische Prozesse einen nicht-monotonen Trendverlauf verursachen, kann die Berechnung der Veränderungsrate mit dem Anfang-Ende-Kontrast täuschen, wie in Abbildung 7 exemplarisch dargestellt. Die Veränderungsrate von -88 % gilt nur für den Zeitraum von 1970-1999, und spiegelt ansonsten nicht den realen längerfristigen Trend wider. Eine Bewertung auf Basis eines linearen Trends wäre hier angemessener.

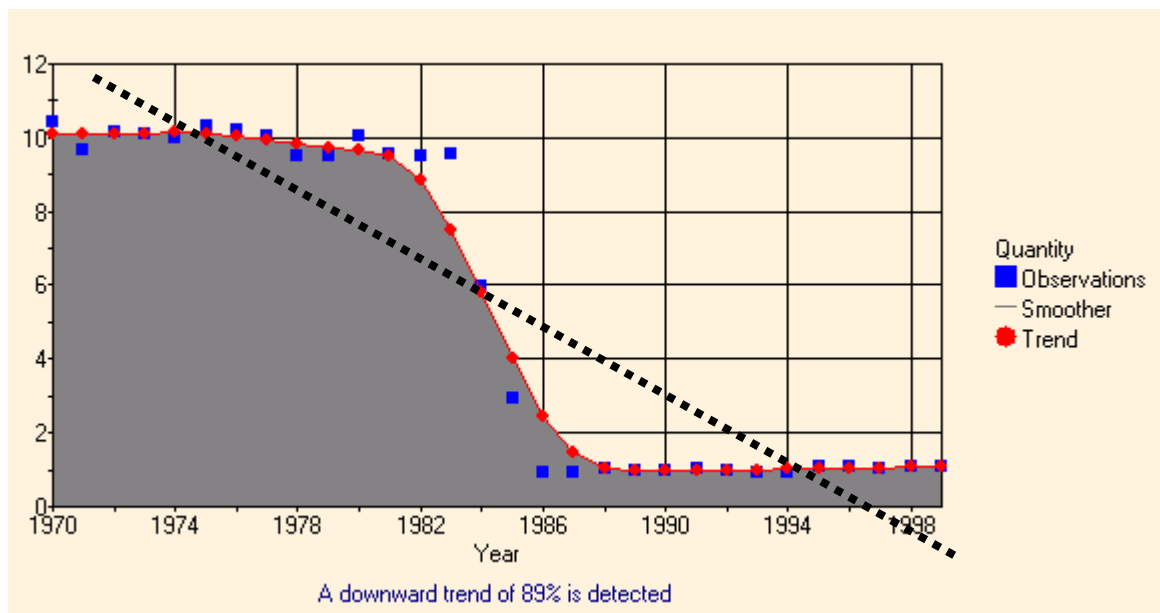
Abb. 7: LOESS-Trend bei über mehrere Jahre wirksamen, klimatisch bedingten Schwankungen



Monotone nicht-lineare Trendmuster

Wenn in langen Zeitreihen ein monotoner, nichtlinearer Trend vorliegt, ist die Verwendung des linearen Regressionskoeffizienten irreführend. Wie in Abbildung 8 dargestellt, kann die berechnete relative Reduktion 100 % (ohne logarithmische Transformation) übersteigen. Bei Verwendung einer logarithmischen Transformation ergibt sich eine relative Reduktion von immer noch 96 %, wohingegen die wahre relative Reduktion bei 89 % liegt. Die Verwendung des Anfang-Ende-Kontrast wäre in diesem Fall angemessener.

Abb. 8: LOESS-Trend und linearer Trend bei abruptem Rückgang der Konzentrationswerte



2.3.6 Schlussfolgerungen

Eine formale Trendüberprüfung sollte sowohl den linearen als auch den nicht-linearen Trendanteil überprüfen. Die Testergebnisse vervollständigen einander und geben einen ersten Eindruck von der Trendstruktur.

Tab. 5: Interpretation der Testergebnisse

			Test für die Linearität der Trendkomponente, Abschnitt 1.1.4	
			H_0 : linearer Regressionskoeffizient (slope) = 0	
			H_0 wird akzeptiert: slope = 0	linearer Trend erfasst (H_0 abgelehnt)
Test für Nichtlinearität, Abschnitt 2 (H_0 : Trend ist linear)	H_0 wird akzeptiert: No	Nichtlinearität	<i>Nur Zufallsvariationen; keine systematischen Veränderungen</i>	<i>linearer Trend</i>
	Nichtlinearität erfasst (H_0 abgelehnt)		<i>Nicht-monotone Trendmuster ohne allgemeine Tendenz</i>	<i>Umfassende systematische Veränderungen</i>

Wenn Nichtlinearität in der Zeitreihe vorliegt, dann sollte das Konfidenzintervall für die Sollwertüberprüfung basierend auf der linearen Regression nicht angewendet werden. Alle andere Methoden, die auf dem linearen Regressionskoeffizient basieren, können gleichwohl eingesetzt werden.

Der Test auf linearen Trend kann eingesetzt werden, um zu überprüfen, ob eine Tendenz des Anstiegs bzw. Abnahme vorhanden ist. Der Anfang-Ende-Kontrast-Test wird eingesetzt, um zu überprüfen, ob sich die geschätzten Niveaus im ersten und im letzten Jahr signifikant unterscheiden.

Verfahren, die auf dem linearen Regressionskoeffizienten basieren, sind vorzuziehen, wenn klimatische Prozesse ein nicht-monotones Trendmuster verursachen, wie in Abbildung 7 dargestellt. Ansonsten liegt es nahe, für Tests und Konfidenzintervalle den Anfang-Ende-Kontrast zu verwenden.

2.4 Der LOESS-Smoother unter Einbeziehung von Monatsdaten

2.4.1 Einführung

Während die Analyse von Jahresdaten naturgemäß keine Berücksichtigung von Saisonkomponenten erforderlich macht, ist letzteres bei der Analyse von Monatsdaten unumgänglich. Dieser Abschnitt behandelt daher die Einbeziehung von Saisonkomponenten in den LOESS-Smoother. Die Methodik ist eine vollständige Neuentwicklung, so dass die Berechnungen nicht in kommerziell verfügbaren Programmen realisiert werden können, sondern selbst programmiert werden müssen.

2.4.2 Definition

Im folgenden wird vorausgesetzt, dass die Proben monatlich oder zweiwöchentlich genommen werden und dass y_i ($i=1, \dots, n$) die Beobachtung zum Zeitpunkt x_i kennzeichnet. Weiterhin wird vorausgesetzt, dass die Messzeitpunkte x_i in Einheiten eines Jahres ausgedrückt werden, d.h. die Differenz zwischen zwei aufeinanderfolgenden Tagen beträgt $1/365,25$. In Analogie zur Konstruktion der Smothermatrix für Jahresdaten erfolgt die Konstruktion auf der Basis unterjähriger Daten durch eine Serie von gewichteten Regressionsanalysen separat für jeden einzelnen Zeitpunkt x_t unter Verwendung aller Daten, die innerhalb des Zeitintervalls $x_t \pm \Delta_t$ liegen. Das Gewicht der Beobachtung y_i in der lokalen Regression für den Zeitpunkt x_t wird bestimmt durch die Entfernung zwischen den Zeitpunkten x_i und x_t ,

$$w_{ti} = \begin{cases} \left(1 - \left(\frac{|x_t - x_i|}{\Delta_t} \right)^3 \right)^3 & \text{for } 0 \leq \frac{|x_t - x_i|}{\Delta_t} \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{Gleichung 75})$$

Die Breite von Δ_t steuert die Glättung und beruht auf dem Vorschlag von Nicholson und Fryer (1999) für den LOESS-Smoother ohne Saisonalität. Δ_t wird so festgelegt, dass die Zeitspanne der jeweils einbezogenen Daten gleich bleibt, d.h.

$$\Delta_t = \max \left\{ \frac{\text{span} + 1}{2}, \text{span} - \min\{t - 1, n - t\} \right\}. \quad (\text{Gleichung 76})$$

Mit der Designmatrix X des Regressionsmodells mit einer saisonalen Komponente zweiter Ordnung

$$X = \begin{pmatrix} 1 & x_1 & \sin 2\pi x_1 & \cos 2\pi x_1 & \sin 4\pi x_1 & \cos 4\pi x_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \sin 2\pi x_n & \cos 2\pi x_n & \sin 4\pi x_n & \cos 4\pi x_n \end{pmatrix}, \quad (\text{Gleichung 77})$$

dem Beobachtungsvektor

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad (\text{Gleichung 78})$$

und der Gewichtungsmatrix

$$W_t = \begin{pmatrix} w_{t1} & 0 & \dots & 0 \\ 0 & w_{t2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & w_{tn} \end{pmatrix} \quad (\text{Gleichung 79})$$

ergibt sich der LOESS-Trend zum Zeitpunkt x_t durch $S'_t Y$, wobei S'_t die i -te Zeile von $X(X'W_tX)^{-1}X'W_t$ ist. Die sich daraus ergebende Smoother-Matrix kann dargestellt werden als

$$S = \begin{pmatrix} S'_1 \\ \vdots \\ S'_n \end{pmatrix}. \quad (\text{Gleichung 80})$$

2.4.3 Tests

Tests für den linearen Trend, für den nichtlinearen Trend, für die Differenz zwischen dem Niveau am Anfang und Ende der Zeitreihe und für Konfidenzintervalle ergeben sich analog zu den Berechnungen auf der Basis von Jahresdaten. Ein Test für Saisonalität kann folgendermaßen durchgeführt werden: Sei K_{error} die quadratische Matrix, die durch die quadratische Form

$$SS_{error} = Y'(I - S)(I - S)Y = Y'K_{error}Y \quad (\text{Gleichung 81})$$

definiert ist, sei K_{season} die quadratische Matrix, die durch

$$SS_{season} = Y'(I - S_{noseason})(I - S_{noseason})Y - Y'(I - S)(I - S)Y$$

$$=: Y' K_{season} Y, \quad (\text{Gleichung 82})$$

definiert ist, und sei $S_{noseason}$ die entsprechende Smoother-Matrix für den LOESS-Smoother ohne Saisonalität. Sei weiterhin die Prüfgröße

$$F_{season} = \frac{SS_{season} / df_1}{SS_{error} / df_2}, \quad (\text{Gleichung 83})$$

wobei

$$df_1 = \frac{\text{tr}(K_{season})^2}{\text{tr}(K'_{season} K_{season})}, \quad df_2 = \frac{\text{tr}(K_{error})^2}{\text{tr}(K'_{error} K_{error})}. \quad (\text{Gleichung 84})$$

Unter der Nullhypothese (dass keine Saisonalität vorliegt) kann die Verteilung von

$$\frac{\frac{\text{tr}(K_{error})}{\text{tr}(K'_{error} K_{error})}}{\frac{\text{tr}(K_{season})}{\text{tr}(K'_{season} K_{season})}} F_{season} \quad (\text{Gleichung 85})$$

durch eine F-Verteilung mit (df_1, df_2) Freiheitsgrade angenähert werden. Daher kann die Nullhypothese, dass keine saisonalen Effekte vorhanden sind, abgelehnt werden, falls

$$F_{season} > \frac{\frac{\text{tr}(K_{season})}{\text{tr}(K'_{season} K_{season})}}{\frac{\text{tr}(K_{error})}{\text{tr}(K'_{error} K_{error})}} F(df_1, df_2, 1 - \alpha). \quad (\text{Gleichung 86})$$

2.4.4 Beispiel

Die Zeitreihe der monatlich gemessenen ortho-P Konzentration [mg/l] bei Herbrum/Ems 1984-1997 ist in Abbildung 9 dargestellt. Abbildung 12 zufolge scheint es so, dass in den ersten 15 Monaten das mittlere Niveau deutlich über den Werten in den Monaten 16-30 liegt. Dies wird zwar auf Basis der über die Jahre gemittelten Konzentration („Mean“) deutlich, nicht jedoch mit dem auf den Jahresdaten basierenden Smoother („Predicted“) mit einer Spanne von sieben Jahren (Abb. 10). Auch mit einer reduzierten Spanne von fünf Jahren (Abb. 11) ergibt sich ein kontinuierlicher, monoton fallender LOESS-Trend. Nachteilig ist auch der Umstand, dass die Untergrenze des Konfidenzintervalls am Ende der Zeitreihe negativ wird. Die hierfür ursächliche große Breite des Konfidenzintervalls erklärt sich aus der ungleichen zeitlichen Verteilung der Zeitreihe.

Abb. 9: Ortho-Phosphat bei Herbrum/Ems [mg/l] 1984-1997

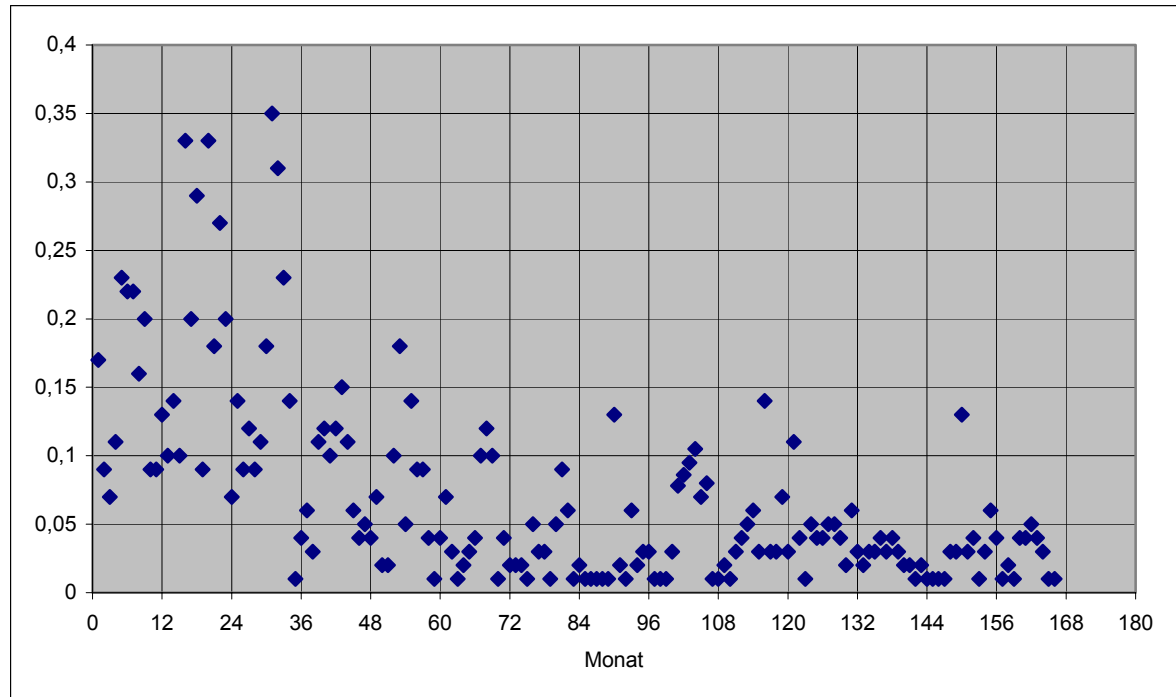


Abb. 10: LOESS-Trend basierend auf den mittleren Jahreskonzentrationen, berechnet mit einer Fensterbreite von 7 Jahren

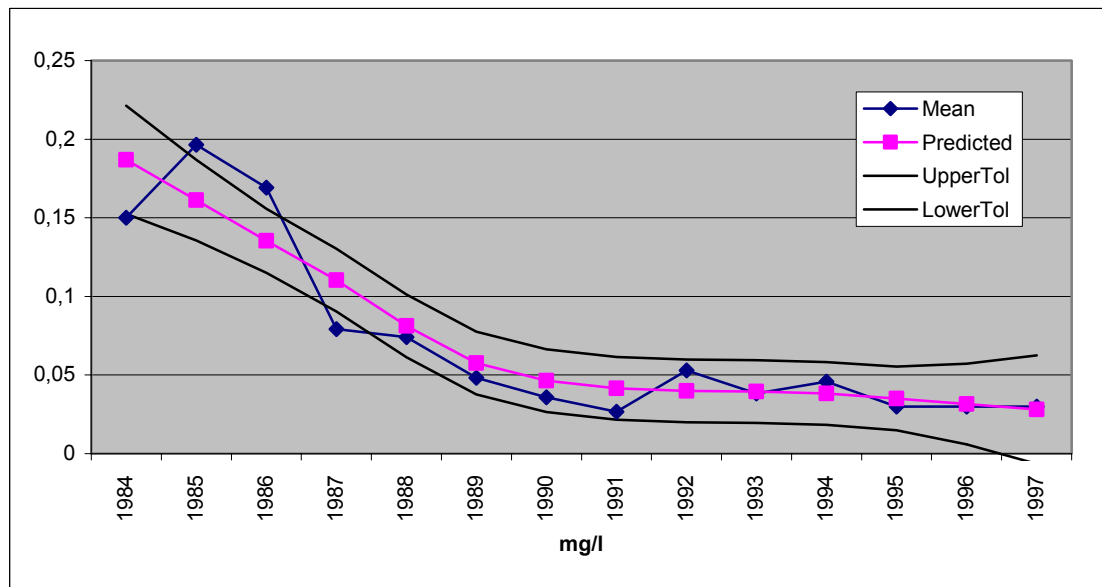
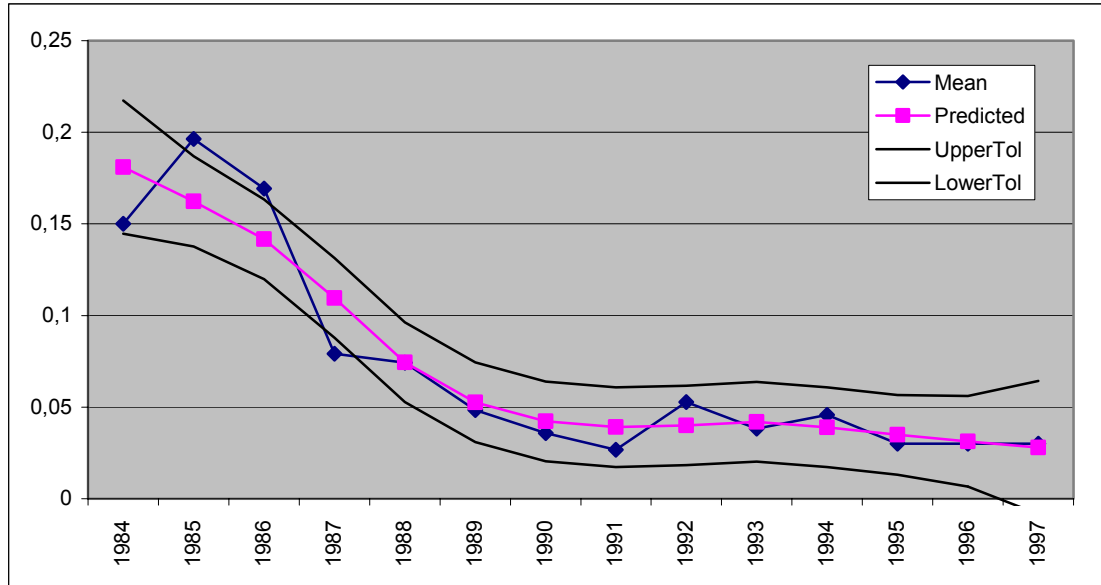
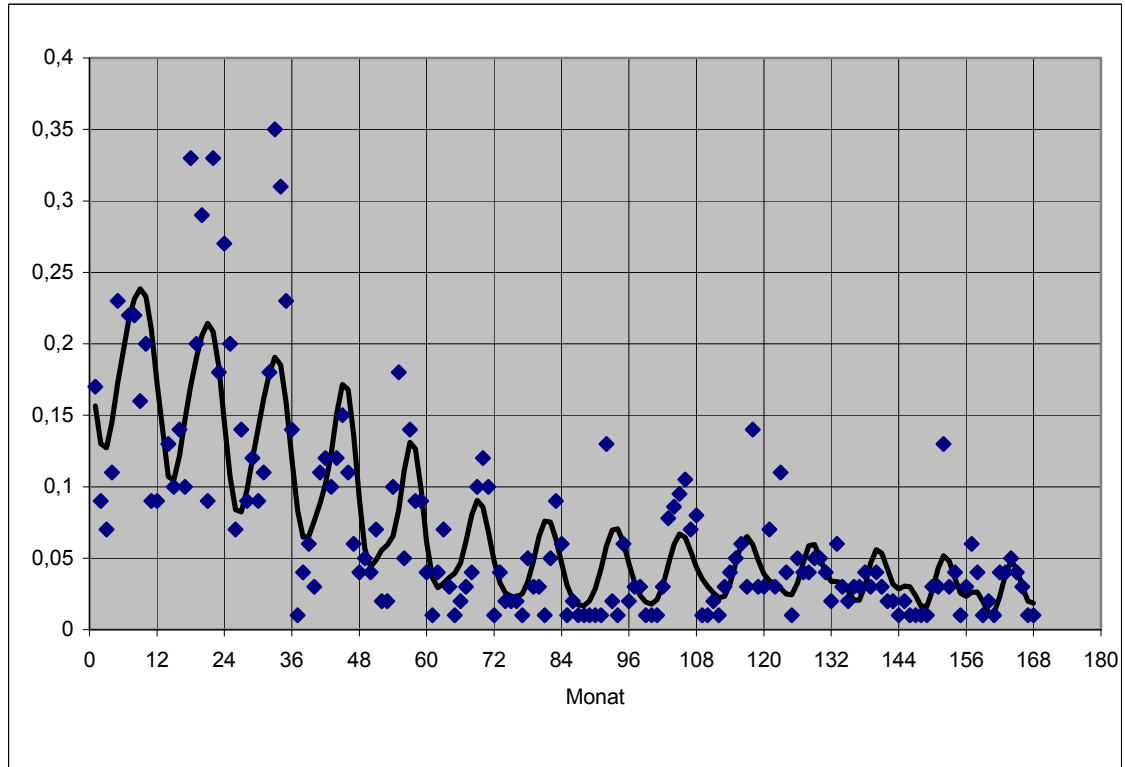


Abb. 11: LOESS-Trend basierend auf den mittleren Jahreskonzentrationen, berechnet mit einer Fensterbreite von 5 Jahren



Verbesserungen sind möglich, wenn man einen LOESS-Smoother verwendet, der auf monatlichen Daten beruht. Die Anpassung der Glättung mit einer Spanne von 85 Monaten wird in Abbildung 12 dargestellt.

Abb. 12: LOESS-Trend basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 85 Monaten



Von den geschätzten Mittelwertkonzentrationen und den entsprechenden Konfidenzintervallen können saisonbereinigte Trends abgeleitet werden. Das Ergebnis ist in Abbildung 13 dargestellt. Aufgrund des höheren Stichprobenumfangs ist das Konfidenzintervall kleiner als für die Glättung, die auf den jährlichen Daten beruht. Bessere Ergebnisse im Hinblick auf die Flexibilität der Trendkurve können mit einer Spanne von 49 Monaten erreicht werden (siehe Abbildung 14).

Abb. 13: Jahresmittelwert des LOESS-Trends, basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 85 Monaten

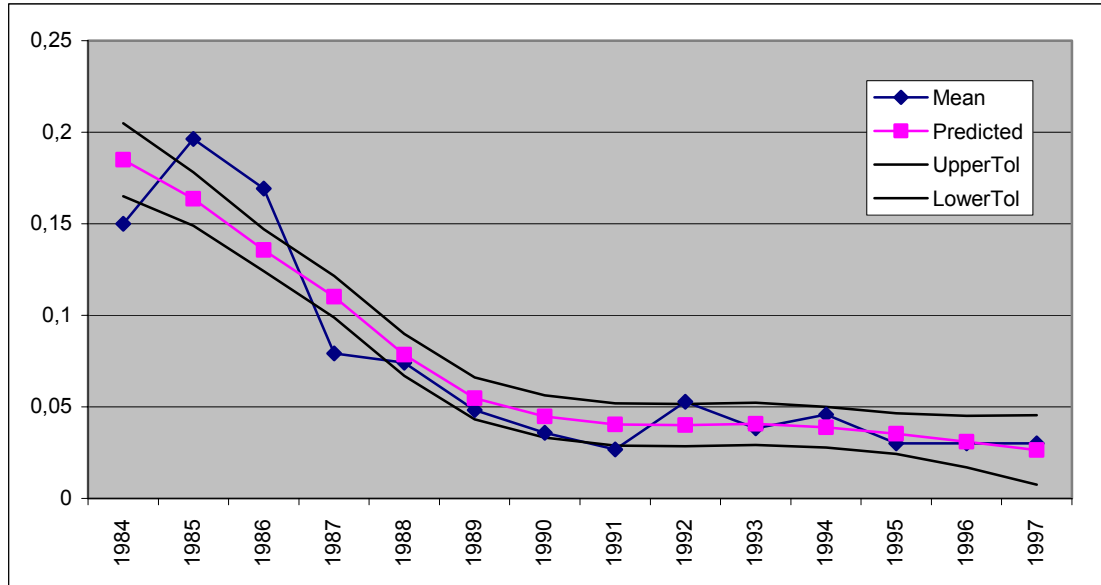
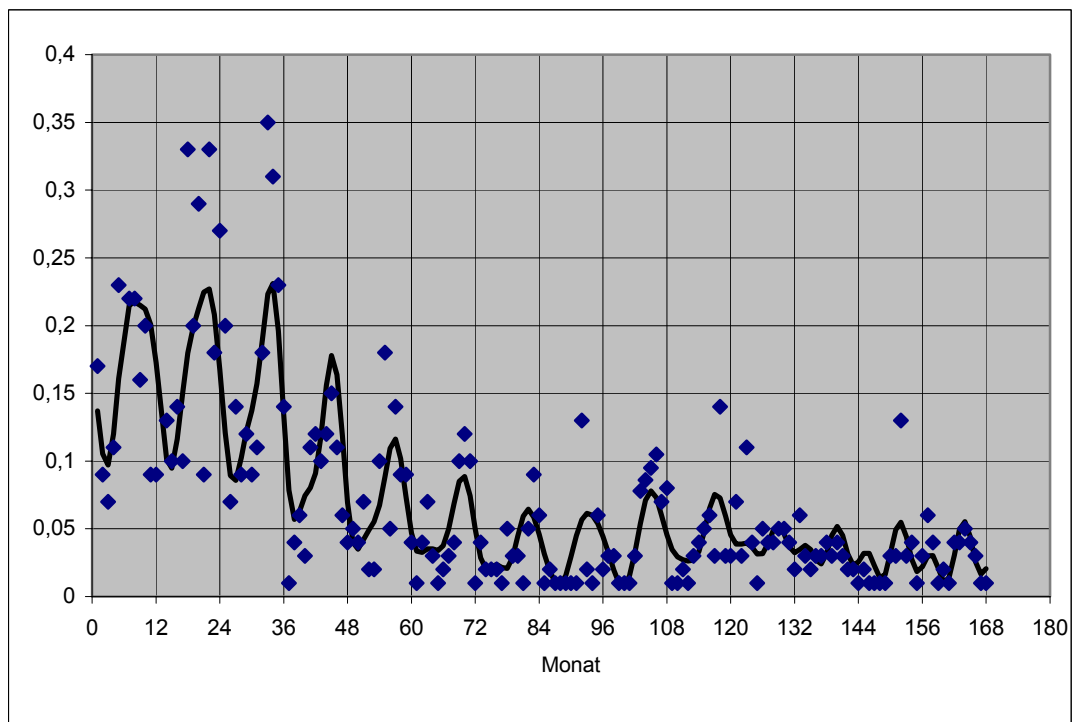


Abb. 14: LOESS-Trend basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 49 Monaten



2.4.5 Schlussfolgerungen

Die Analyse von Monatsdaten mit dem LOESS-Smoother bietet den Vorteil, dass das Konfidenzintervall in der Regel schmaler ist, als wenn die Berechnung auf der Basis von Jahresdaten erfolgt. Weiterhin bietet die Analyse von Monatsdaten die Möglichkeit die dem LOESS Smoother zugrundeliegende Spanne zu reduzieren. Das ist besonders dann von Bedeutung, wenn angenommen wird, dass anthropogene Veränderungen innerhalb einer kurzen Zeitspanne wirksam werden. Wenn daher eine befriedigende Anpassung der zugrundeliegenden Monatsdaten gelingt, ist eine Auswertung auf der Basis von Monatsdaten einer Auswertung auf Basis von Jahresdaten vorzuziehen. Vorteile ergeben sich insbesondere bei Zeitreihen aus einem Zeitraum von nicht mehr als 10 Jahren.

Es ist darauf hinzuweisen, dass die Anpassung auf Basis von Monatsdaten aufgrund der Datenlage vielfach nur bei Flussdaten möglich ist, während für die Messungen im Meer in der Regel eine gröbere zeitliche Auflösung verwendet werden muss. Grundsätzlich eignet sich das beschriebene Verfahren jedoch auch zur Anwendung auf Quartalswerte oder auf im Winter gemessene Nährstoffkonzentrationen, wenn man – aufgrund einer starken Saisonalität – die Einbeziehung der Sommerwerte vermeiden möchte.

2.5 Trendanalysen mit Kovariablen: Quecksilbergehalte von Schollen aus der Deutschen Bucht 1985 – 1999

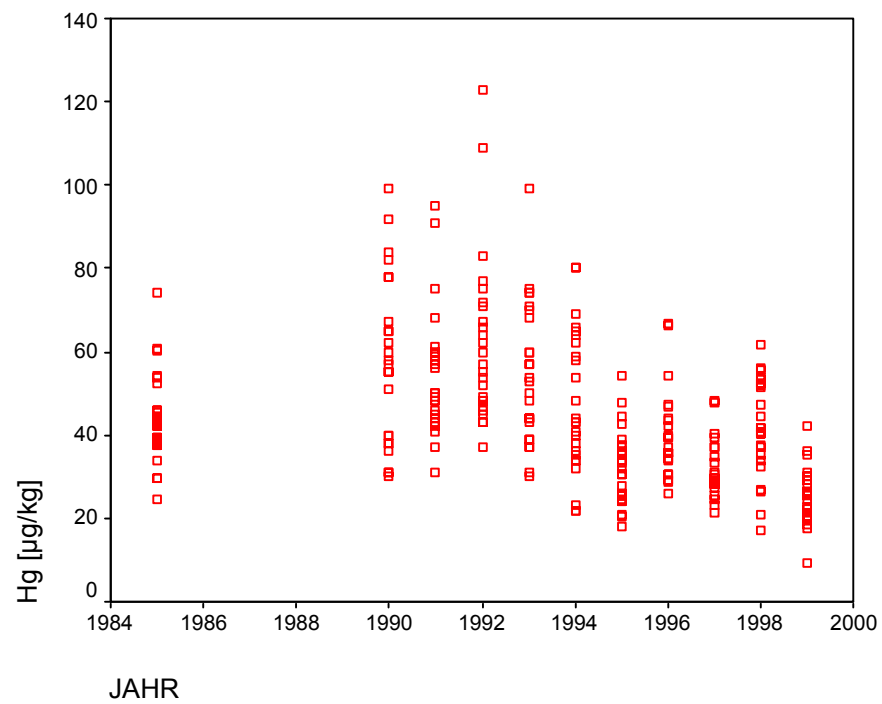
In diesem Abschnitt werden die Ergebnisse einer Trendanalyse vorgestellt, die anhand der Quecksilbergehalte von Schollen aus der Deutschen Bucht durchgeführt wurde. Die Daten wurden freundlicherweise von Herrn Dr. Harms, Bundesforschungsanstalt für Fischerei, zur Verfügung gestellt. Alle Berechnungen wurden mit dem Programm S-Plus durchgeführt.

Gegenstand der Trendanalyse ist einerseits die Abhängigkeit der Konzentration von Länge und Gewicht, und andererseits die Frage, ob für die Trendanalyse eine Betrachtung der individuellen Messergebnisse notwendig ist, oder ob eine Betrachtung der Jahresmittelwerte ausreicht, um den Trendverlauf hinreichend genau bestimmen zu können.

2.5.1 Deskriptive Analyse

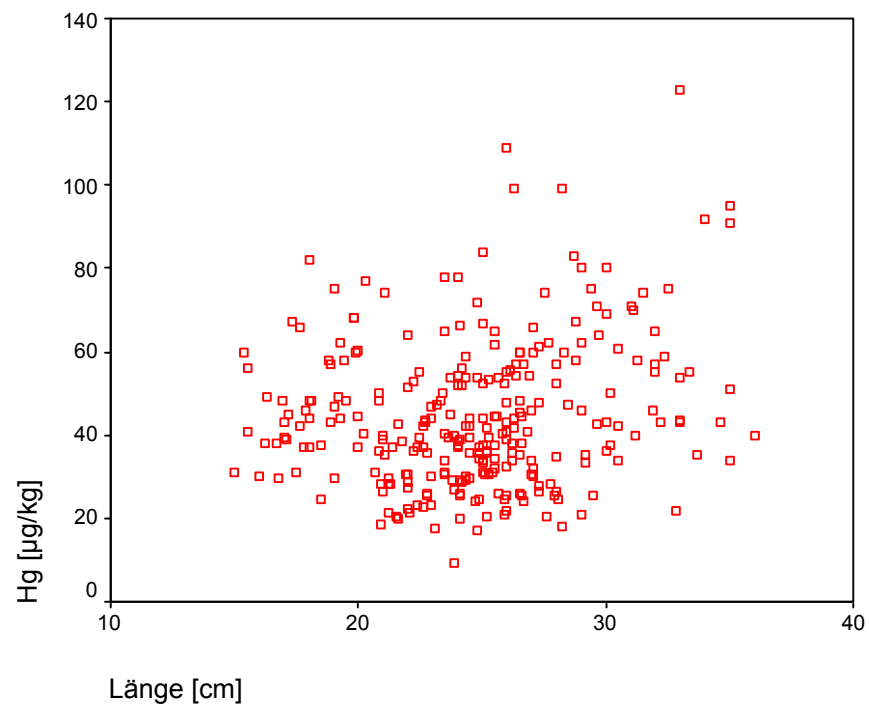
Die grafische Darstellung der gemessenen Konzentrationen zeigt eine deutliche Abhängigkeit der Konzentrationen vom jeweiligen Jahr (Abb. 15).

Abb. 15: Hg-Konzentrationen (Einzelwerte) im Zeitablauf



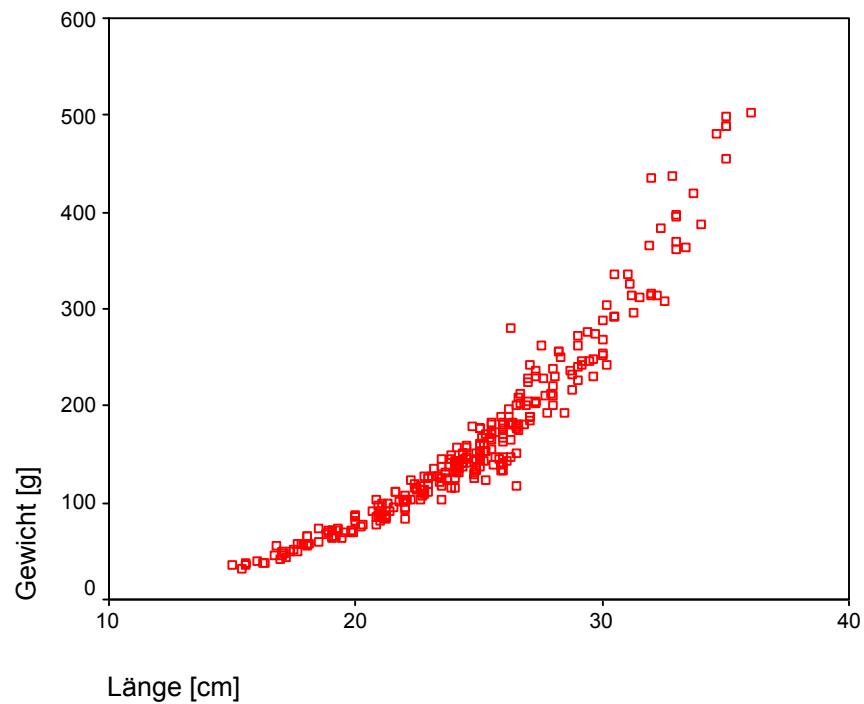
Weiterhin zeigt sich ein leichter Zusammenhang zwischen Länge und Konzentration (Abb. 16).

Abb. 16: Hg-Konzentrationen (Einzelwerte) in Abhängigkeit von der Länge



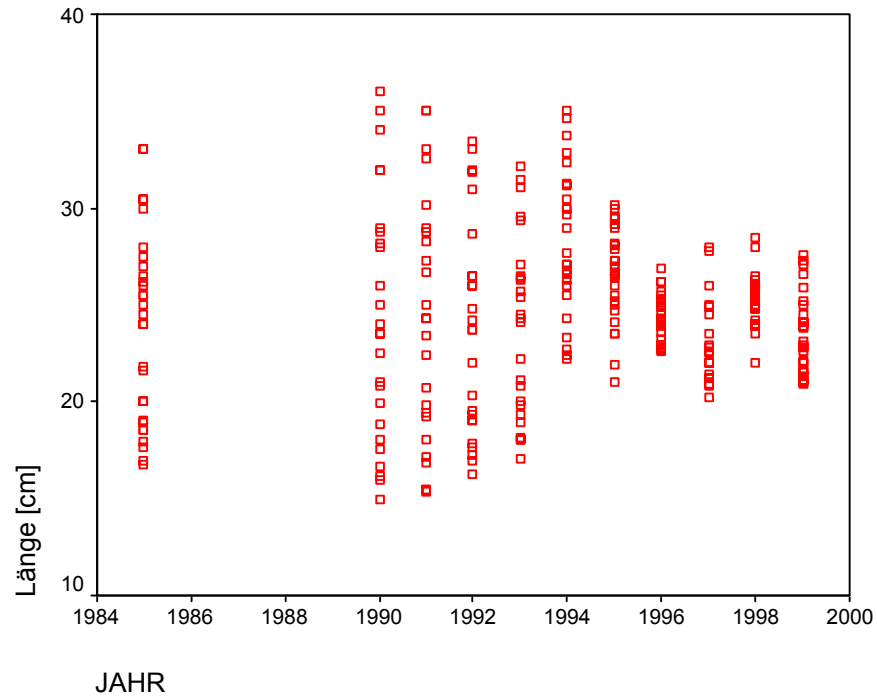
Ein sehr guter Zusammenhang besteht zwischen Länge und Gewicht (Abb. 17), d.h. es ist keine separate Adjustierung bezüglich beider Parameter sinnvoll.

Abb. 17: Beziehung zwischen Gewicht und Länge der Schollen



Die folgende Grafik zeigt eine auffällige Abnahme der Streuung der Größe der Individuen im Zeitablauf. Nach Rücksprache mit Herrn Dr. Harms von der Bundesforschungsanstalt für Fischerei ist dafür jedoch nicht die Fangtechnik verantwortlich zu machen. Möglicherweise hat sich aufgrund veränderter Umweltbedingungen die Population im Fanggebiet verändert.

Abb. 18: Länge der Schollen im Zeitablauf



Bevor eine Zeitreihenanalyse durchgeführt werden kann, ist zunächst der Effekt unterschiedlicher Längen auf die Konzentrationswerte zu überprüfen und zu quantifizieren. Hierzu wird für jedes Jahr separat eine Regressionsanalyse durchgeführt, deren Ergebnisse im folgenden dargestellt werden:

Tab. 6: Analysen der Regressionsanalysen für die Einzeljahre

1985

Koeffizienten

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	22,829	8,842		2,582	,016	4,686	40,972
	Länge [cm]	,868	,362	,419	2,398	,024	,125	1,611

a Abhängige Variable: Hg [µg/kg]

1990

Koeffizienten

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	27,185	14,729		1,846	,078	-3,284	57,655
	Länge [cm]	1,263	,587	,409	2,152	,042	,049	2,478

a Abhängige Variable: Hg [µg/kg]

1991

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	25,102	10,725		2,341	,028	2,916	47,288
	Länge [cm]	1,235	,431	,513	2,865	,009	,343	2,127

1992

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	34,321	17,827		1,925	,067	-2,557	71,198
	Länge [cm]	1,165	,729	,316	1,599	,123	-,342	2,672

1993

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	18,787	16,494		1,139	,266	-15,334	52,909
	Länge [cm]	1,487	,678	,416	2,193	,039	,085	2,889

1994

		Nichtstandardisier te Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	39,183	26,591		1,474	,154	-15,824	94,191
	Länge [cm]	,300	,931	,067	,322	,750	-1,625	2,225

1995

		Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	30,683	18,838		1,629	,115	-8,039	69,404
	Länge [cm]	3,845E-02	,704	,011	,055	,957	-1,409	1,486

1996

		Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	72,801	43,266		1,683	,106	-16,701	162,304
	Länge [cm]	-1,352	1,766	-,158	-,765	,452	-5,004	2,301

1997

		Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	31,174	17,330		1,799	,088	-5,099	67,447
	Länge [cm]	4,168E-02	,750	,013	,056	,956	-1,527	1,610

1998

		Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	73,226	44,933		1,630	,117	-19,725	166,177
	Länge [cm]	-1,229	1,773	-,143	-,693	,495	-4,897	2,438

1999

		Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten	t	P-Wert	95% Konfidenzintervall Für B	
Modell		B		Beta			Untere Schranke	Obere Schranke
1	(Konstante)	12,842	16,153		,795	,435	-20,574	46,257
	Länge [cm]	,536	,680	,162	,790	,438	-,869	1,942

Sehr auffällig ist der Längeneinfluss bis 1993, d.h. für diese Jahre ist deutlich erkennbar, dass bei größeren bzw. älteren Fischen eine höhere Hg-Konzentration vorliegt. Seit 1994 ist allerdings kein Längeneinfluss mehr nachweisbar, da die zugehörigen P-Werte das Signifikanzniveau von 5 % für alle Jahre ab 1994 deutlich überschreiten. Dies erklärt sich einerseits aus dem seitdem deutlich reduzierten mittleren Konzentrationsniveau, und zum anderen aus der wesentlich geringeren Variabilität der Größe der Schollen.

2.5.2 Adjustierung bezüglich der individuellen Länge

Wenn von Jahr zu Jahr wechselnde Größenverhältnisse vorliegen, ist fraglich, inwieweit eine Adjustierung auf eine einheitliche Länge sinnvoll ist. Zum Vergleich werden im folgenden

Jahresmittelwert und adjustierter Jahresmittelwert, bezogen auf eine durchschnittliche Länge von 24,7 cm, gegenübergestellt:

Tab. 7: Vergleich von arithmetischen Mittel und längenadjustiertem Mittelwert

Const	Längeneffekt	Mittlere Länge [cm]	Adjustierter Jahresmittelwert Hg [µg/kg]	Arithmetisches Mittel Hg [µg/kg]
22,83	0,868	23,9414	44,2696	43,61
27,185	1,263	24,2960	58,3811	57,88
25,102	1,235	24,1120	55,6065	54,88
34,321	1,165	23,8600	63,0965	62,12
18,787	1,487	23,8960	55,5159	54,32
39,183	0,3	28,3280	46,593	47,68
30,683	0,038	26,6464	31,6216	31,71
72,8	-1,352	24,4760	39,4056	39,72
31,174	0,0417	23,0190	32,20399	32,13
73,226	-1,229	25,3080	42,8697	42,11
12,842	0,536	23,6840	26,0812	25,55
		24,7223		

Der Vergleich zeigt, dass eine Adjustierung der individuellen Konzentrationsmesswerte bezüglich der Länge unnötig erscheint. Dies erklärt sich im wesentlichen aus der nur sehr geringen Schwankungsbreite der Größe in den Jahren bis 1993. Erst im Jahr 1994 ist ein deutlicher Anstieg der mittleren Länge feststellbar, der jedoch unwirksam bleibt, da zugleich der Längeneffekt nicht mehr statistisch signifikant ist, wie aus den Ergebnissen der Tabelle 6 hervorgeht. Gänzlich unwirksam ist eine individuelle Adjustierung bezüglich der Länge jedoch nicht, denn dieses Verfahren gestattet die leichtere Identifizierung einzelner Ausreißer.

2.5.3 Adjustierung bezüglich der mittleren Länge der Population

Eine nähere Betrachtung der Jahresmittelwerte legt den Verdacht nahe, dass hinsichtlich der Populationsparameter doch ein Zusammenhang zwischen Länge und Konzentration besteht. Daher wurde ein GAM-Modell angepasst, in dem in Erweiterung des im Trend-y-tector implementierten Verfahrens zusätzlich ein Längeneffekt einbezogen ist. Bei Anwendung des Programms SPLUS ist dieses Modell wie folgt zu programmieren:

```
Call: gam(formula = Concentration ~ lo(Year) + Length, family = gaussian,
  data = SP1, na.action = na.omit, control = list(epsilon = 0.001,
  bf.epsilon = 0.001, maxit = 50, bf.maxit = 10, trace = F))
```

Deviance Residuals:

```
      2      3      4      5      6      7      8
0.4556095 -3.108917 4.089274 0.6910638 0.703858 -7.154133 4.577062
      9     10     11
-3.99108 7.347766 -3.610504
```

(Dispersion Parameter Für Gaussian family taken to be 43.47691)

Null Deviance: 1407.081 on 9 Freiheitsgrade

Residual Deviance: 182.6533 on 4.201157 Freiheitsgrade

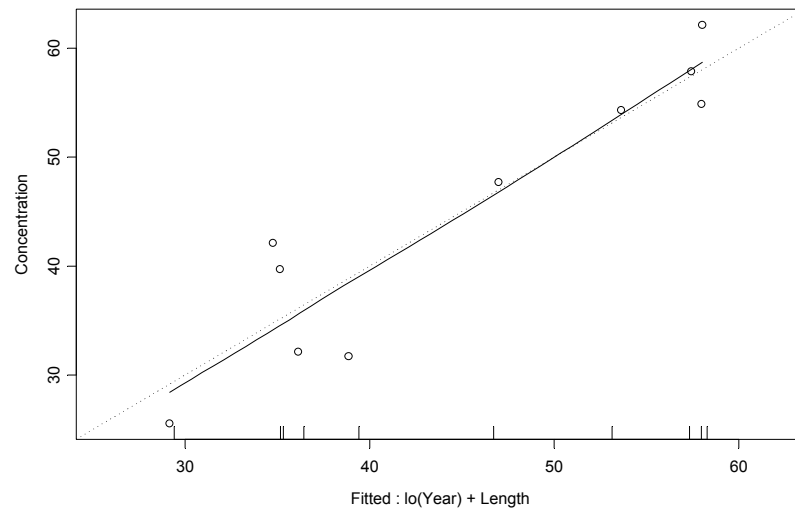
Number von Local Scoring Iterations: 1

DF Für Terms und F-values Für Nonparametric Effects

	Df	Npar	Df	Npar	F	Pr(F)
(Intercept)	1					
lo(Year)	1	2.8	1.350338	0.3695327		
Length	1					

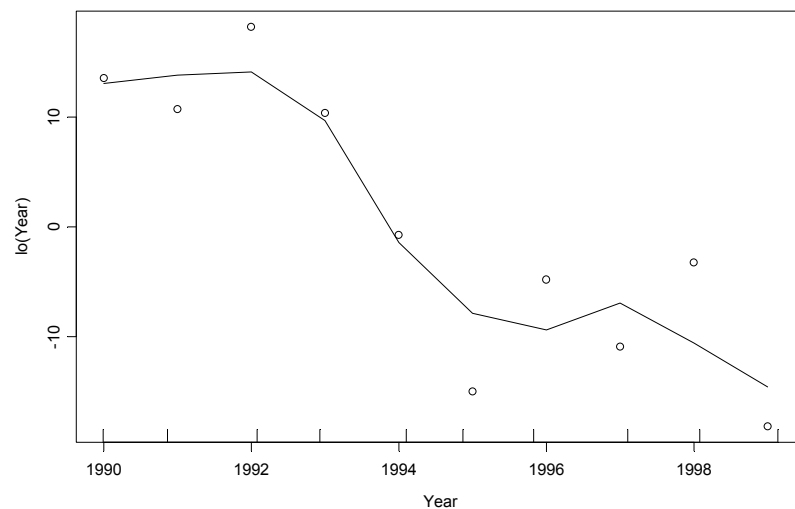
Ein signifikanter Nachweis der Effekte ist mit diesem Modell zwar nicht möglich, jedoch fällt bei Betrachtung der gegen die jeweiligen Trendwerte aufgetragenen Residuen (Abb. 19) ins Auge, dass die Zufallsschwankungen der Residuen erheblich geringer sind als die Schwankungen der Jahresmittelwerte:

Abb. 19: Residualplot des GAM-Modells auf Basis der Jahresdaten



Auffällig ist weiterhin der deutlichere Rückgang der geschätzten mittleren Konzentration bereits im Jahr 1994, der sich aus dem Längeneffekt erklärt:

Abb. 20: Trend des GAM-Modells auf Basis der Jahresdaten



Diese Modellanpassung ist statistisch nicht signifikant, wie der P-Wert von 0,3695 für den Trend. Es ist daher naheliegend, ein äquivalentes Modell auf Basis der Individualdaten anzupassen:

```
Call: gam(formula = Hg ~ lo(Jahr) + Laenge, family = gaussian, data = BLMPHG,
  subset = Jahr > 1989, na.action = na.omit, control = list(
  epsilon = 0.001, bf.epsilon = 0.001, maxit = 50, bf.maxit = 10,
  trace = F))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-29.52209	-9.885626	-2.057176	8.012095	55.65474

(Dispersion Parameter for Gaussian family taken to be 200.5082)

Null Deviance: 83570.71 on 248 degrees of freedom

Residual Deviance: 48765.13 on 243.2077 degrees of freedom

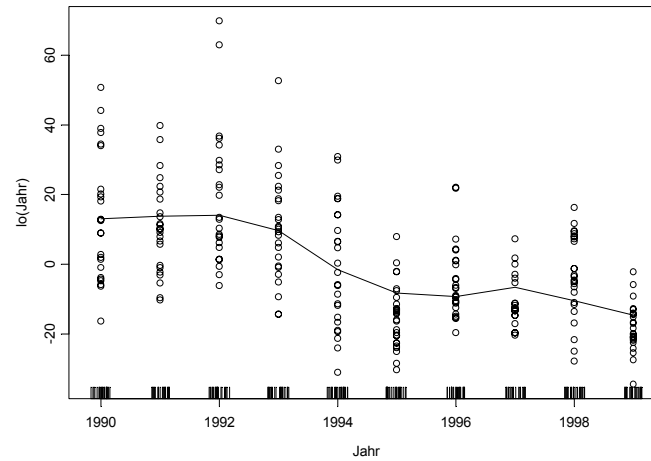
Number of Local Scoring Iterations: 1

DF for Terms and F-values for Nonparametric Effects

	Df	Npar	Df	Npar	F	Pr(F)
(Intercept)	1					
lo(Jahr)	1	2.8	10.3148	3.916543e-006		
Laenge	1					

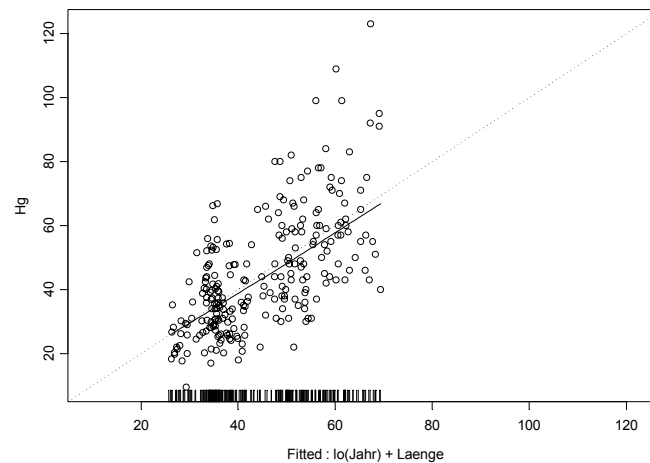
Bei Betrachtung der Einzelwerte ist ein signifikanter Nachweis der Jahrestrends offenbar möglich. Der Trendverlauf (Abb. 21) ist nahezu identisch zu dem auf Basis der Jahresmittelwerte berechneten Trends:

Abb. 21: Trend des GAM-Modells auf Basis der Einzelwerte



Wiederum zeigt sich, dass ein erheblicher Anteil der Variabilität der Einzelwerte durch das Modell erklärt werden kann:

Abb. 22: Residualplot des GAM-Modells auf Basis der Einzelwerte



Das verwendete Modell erweist sich aufgrund der Signifikanz des Jahrestrends zunächst dem auf Jahresmittelwerten basierenden Modell als überlegen. Gleichwohl ist bei Verwendung des auf den Einzelwerten basierenden Modells zu beachten, dass weitere statistische Effekte, z.B. zusätzliche zufällige Jahreseffekte die Gültigkeit der Aussagen einschränken können. Auch im vorliegenden Fall liegen solche Effekte möglicherweise in geringem Umfang vor. Sofern ein

zufälliger Jahreseffekt nicht mehr vernachlässigbar ist, sollte die Auswertung auf Basis von Jahresdaten durchgeführt werden.

2.5.4 Schlussfolgerungen für die Stichprobenplanung

Eine wichtige Schlussfolgerung aus der statistischen Analyse besteht darin, dass die Schadstoffkonzentrationen nicht nur von einer Größenklasse erfasst werden sollten, sondern möglichst von mehreren Größenklassen, um die Variation zwischen den Jahreswerten zu reduzieren und um potentielle Gefahren besser identifizieren zu können (z.B. die hohen Hg-Gehalte bei alten Schollen bis 1994). Die Berücksichtigung von Schollen nur der Länge von ca. 25 cm lässt keinerlei Schluss auf die Belastung anderer Schollen zu. Auch die Standardabweichung des Steigungsparameters B ist um so größer, je kleiner die Spanne der Längen ist, so dass eine etwa notwendig werdende Adjustierung im Falle restriktiver Fangvoraussetzungen nicht mehr möglich ist. Am günstigsten ist jedenfalls eine stratifizierte Stichprobenplanung, dass heißt die gleichmäßige Berücksichtigung verschiedener Größenklassen.

2.6 Trendanalysen für Abundanz und Biomasse

Für Abundanz und Biomasse ausgewählter Algengruppen wurde auf Basis der HELCOM-Daten umfangreiche Trendanalysen durchgeführt, die in diesem Bericht aus Platzgründen nicht wiedergegeben werden können, sondern auf CD-ROM abgespeichert wurden. Die Daten wurden freundlicherweise durch Herrn Dr. Wasmund vom Institut für Ostseeforschung, Warnemünde zur Verfügung gestellt. Alle Berechnungen wurden unter DELPHI durchgeführt, da mit Standardstatistikpaketen keine hinreichende Flexibilität bei der Auswertung gewährleistet werden konnte.

2.6.1 Datenbasis

Folgenden Algengruppen wurden einer Analyse unterzogen:

- NOSTOCOP (Nostocophyceae)
- DIATOMOP (Diatomophyceae)
- DINPHYCE (Dinophyceae)

Von artspezifischen Untersuchungen wurde aufgrund der Fülle des vorliegenden Materials zunächst abgesehen, zumal sich zeigte, dass die Einschränkung auf einzelne Arten zu höheren Anteilen fehlender Werte führt, welches wiederum methodische Probleme nach sich zieht.

Die Auswertungen erfolgten stationsweise jeweils unter Einbeziehung aller Daten der jeweiligen „Jahreszeit“. Diese sind nach vorherigen HELCOM-Einschätzungen definiert:

Tab. 8: Festlegung der „Jahreszeiten“

	Ostsee (d.h. von der Arkonasee bis in die nördl. eigentl. Ostsee)	Kattegat/Beltsee-Gebiet
Frühjahr	März-Mai	Februar-April
Sommer	Juni-September	Mai-August
Herbst	Oktober-Dezember	September-November
Winter	Januar-Februar	Dezember-Januar

Für folgende Stationen wurden die Trendanalysen durchgeführt:

- Kattegat: BMP R1 bis BMP R7
- Großer Belt: BMP P1, BMP P2
- Öresund: BMP Q1, BMP Q2
- Kieler Bucht: BMP N1 bis BMP N4, GAUSS7
- Mecklenburger Bucht: BMP M1, BMP M2, BMP M2A, 5, 22, 23

- Arkonasee: BMP K4 bis BMP K8; küstennah im Rügen: 9, 11.
- Pommersche Bucht: BMP K57, BMP K58, 162, 131, 160, 133
- Bornholmsee: BMP K2, 200
- südl. Teil der östl. Gotlandsee: BMP K1
- Danziger Bucht: BMP L1
- östliche Gotlandsee: BMP J1, (BMP J2 ist küstennah), 286
- westliche Gotlandsee: BMP I1
- nördliche eigentliche Ostsee: BMP H1 bis BMP H3, 280, 282, 284, 285, LL22A, LL22B

Da bei den dänischen Biomasse-Daten teilweise mit anderen Umrechnungsfaktoren gearbeitet worden war, wurden, um Verzerrungen zu vermeiden, bei den Auswertungen für die Biomasse nur jene Stationen berücksichtigt, bei denen keine dänische Daten enthalten waren.

2.6.2 Auswertung

Die Auswertung der gruppenspezifischen Biomassen und Abundanzen wurde nach folgendem Schema durchgeführt:

1. Bildung des arithmetischen Mittelwertes über alle Ergebnisse innerhalb einer Saison.
2. Berechnung des Logarithmus (Log10) der arithmetischen Mittelwerte
3. Berechnung des LOESS-Smoother mit den zugehörigen Tests.
4. Statistischer Test auf Trendbruch

Auf die Bestimmung des Medians wurde verzichtet, da dieser die zu erwartenden Maximalwerte nicht angemessen berücksichtigt. Das arithmetische Mittel ist zwar sehr empfindlich gegenüber

Ausreißern, jedoch bewirkt die nachfolgende Trendanalyse auf Basis der logarithmierten Mittelwerte, dass besonders stark abweichende Einzelwerte nicht unangemessen großen Einfluss auf das Ergebnis haben können.

Alle Ergebnisse wurden in Datenblättern zusammengefasst. Exemplarisch werden auf den folgenden Seiten die Datenblätter nur für die Abundanzen der Kieselalgen (DIATOMOP_A) bei Station BMP K4, die Abundanzen der Dinoflagellaten (Dinphyce_A) bei Station BMP K2 (Frühjahr) und der Biomasse von Blaualgen (NOSTOCOP_B) bei Station BMP K2 dargestellt.

Erläuterungen zu den vorstehenden Datenblättern

Die Ergebnisse von drei Trend-Tests werden dargestellt:

- Test von Mann-Kendall (MK)
- Test auf linearen Trend, basierend auf dem LOESS-Smoother
- Test auf systematischen Trend, basierend auf dem LOESS-Smoother

Bei den beiden zuerst genannten Tests erfolgt sowohl eine Prüfung auf einen Anstieg (upward) als auch auf einen Rückgang (downward). Ergebnis ist jeweils der P-Wert. Bei einem Wert unter 0,05 ist auf dem Signifikanzniveau von 5 % von einem signifikanten Trend auszugehen.

Alle Tests werden auf die logarithmierten jährlichen Saison-Mittelwerte der Abundanzen bzw. Biomassen angewandt. Während auf der ersten Seite des Datenblatts die ursprünglichen Messergebnisse dargestellt sind, enthält die zweite Seite die logarithmierten Mittelwerte (links oben) sowie den zugehörigen systematischen Trendverlauf mit Konfidenzintervall (rechts oben). Die Abbildung links unten beinhaltet die Modellanpassung mittels linearer und quadratischer Regression sowie mittels des Theil-Slopes.

2.6.3 Übersicht der Ergebnisse

Die Datenblätter zu den einzelnen Artengruppen und Stationen sind aus Platzgründen nur auf einer CD-ROM verfügbar, die dem Umweltbundesamt vorliegt. Eine Übersicht über die Resultate liefern die folgenden Tabellen. Darin bezeichnet D einen signifikanten Rückgang und U einen signifikanten Anstieg. X bezeichnet ein Feld, für welches keine Daten vorhanden sind.

Tab. 9: Ergebnisse der Trendanalysen für die Abundanz von Kieselalgen

Kieselalgen: Abundanz					
Station		1. Quartal	2. Quartal	3. Quartal	4. Quartal
I62	MK				X
	Linear				X
BMP H1	MK				X
	Linear				X
BMP H2	MK				X
	Linear				X
BMP H3	MK				X
	Linear				X
BMP I1	MK	D		U	X
	Linear	D		U	X
BMP J1	MK				X
	Linear	D			X
BMP K1	MK				X
	Linear				X
BMP K2	MK				X
	Linear			D	X
BMP K4	MK		D		X
	Linear		D		X
BMP K5	MK				X
	Linear				X
BMP K6	MK				X
	Linear				X
BMP K7	MK				X
	Linear				X
BMP K8	MK				X
	Linear				X
BMP L1	MK		U		X
	Linear		U		X
BMP M1	MK				X
	Linear				X
BMP M2	MK				X
	Linear				X
BMP N1	MK	D			X
	Linear				X
BMP N3	MK			U	X
	Linear			U	X
BMP N4	MK				X
	Linear				X
BMP P1	MK				X
	Linear		D		X
BMP Q2	MK				X
	Linear				X
BMP R1	MK				X
	Linear		D		X
BMP R2	MK	D			X
	Linear	D			X
BMP R3	MK				X
	Linear				X
BMP R4	MK	D	D		X
	Linear		D		X
BMP R6	MK		U		X
	Linear		U		X
BMP R7	MK		U	U	X
	Linear			U	X
GAUSS7	MK				X
	Linear				X

Tab. 10: Ergebnisse der Trendanalysen für die Biomasse von Kieselalgen

Kieselalgen: Biomasse					
Station		1. Quartal	2. Quartal	3. Quartal	4. Quartal
BMP H1	MK		D		X
	Linear	D	D		X
BMP H2	MK	D			X
	Linear	D			X
BMP I1	MK			U	X
	Linear			U	X
BMP J1	MK				X
	Linear				X
BMP K1	MK				X
	Linear				X
BMP K2	MK				X
	Linear				X
BMP K4	MK				X
	Linear				X
BMP K5	MK				X
	Linear				X
BMP K6	MK				X
	Linear				X
BMP K7	MK	D	D	D	X
	Linear	D	D		X
BMP K8	MK				X
	Linear				X
BMP M1	MK	D		U	X
	Linear				X
BMP M2	MK				X
	Linear				X
BMP N3	MK			U	X
	Linear			U	X
BMP R6	MK				X
	Linear				X

Tab. 11: Ergebnisse der Trendanalysen für die Abundanz von Dinoflagellaten

Dinoflagellaten: Abundanz				
Station	1. Quartal	2. Quartal	3. Quartal	4. Quartal
162 MK				X
	Linear			X
BMP H1	MK			X
	Linear			X
BMP H2	MK			X
	Linear			X
BMP H3	MK			X
	Linear			X
BMP I1	MK	U		X
	Linear	U		X
BMP J1	MK	U		X
	Linear	U		X
BMP K1	MK	U		X
	Linear	U		X
BMP K2	MK	U		X
	Linear	U		X
BMP K4	MK	U		X
	Linear	U		X
BMP K5	MK	U		X
	Linear			X
BMP K6	MK			X
	Linear			X
BMP K7	MK			X
	Linear			X
BMP K8	MK		U	X
	Linear			X
BMP L1	MK			X
	Linear			X
BMP M1	MK			X
	Linear			X
BMP M2	MK	U		X
	Linear			X
BMP N1	MK			X
	Linear		U	X
BMP N3	MK	U	U	X
	Linear	U	U	X
BMP N4	MK		U	X
	Linear		U	X
BMP P1	MK			X
	Linear			X
BMP Q2	MK			X
	Linear		D	X
BMP R1	MK		D	X
	Linear		D	X
BMP R2	MK		D	X
	Linear		D	X
BMP R3	MK		D	X
	Linear			X
BMP R4	MK	D	D	X
	Linear	D	D	X
BMP R6	MK			X
	Linear			X
BMP R7	MK			X
	Linear			X
GAUSS7	MK	U		X
	Linear		U	X
LL22A	MK	D		X
	Linear		D	X

Tab. 12: Ergebnisse der Trendanalysen für die Biomasse von Dinoflagellaten

Dinoflagellaten: Biomasse					
Station		1. Quartal	2. Quartal	3. Quartal	4. Quartal
BMP H1	MK			D	X
	Linear			D	X
BMP H2	MK		U	D	X
	Linear		U		X
BMP I1	MK	U			X
	Linear	U			X
BMP J1	MK		D		X
	Linear				X
BMP K1	MK	U			X
	Linear	U			X
BMP K2	MK	U			X
	Linear	U			X
BMP K4	MK	U	D		X
	Linear	U	D		X
BMP K5	MK	U			X
	Linear				X
BMP K6	MK				X
	Linear				X
BMP K7	MK	D		D	X
	Linear	D		D	X
BMP K8	MK	U			X
	Linear	U			X
BMP M1	MK				X
	Linear				X
BMP M2	MK		U		X
	Linear		U		X
BMP N3	MK				X
	Linear				X
BMP R3	MK	D	D		X
	Linear	D	D		X
BMP R6	MK				X
	Linear				X

Tab. 13: Ergebnisse der Trendanalysen für die Abundanz von Blaualgen

Blaualgen: Abundanz					
Station		1. Quartal	2. Quartal	3. Quartal	4. Quartal
BMP H1	MK			U	X
	Linear				X
BMP H2	MK				X
	Linear				X
BMP H3	MK				X
	Linear				X
BMP I1	MK	D			X
	Linear				X
BMP J1	MK				X
	Linear				X
BMP K1	MK				X
	Linear				X
BMP K2	MK				X
	Linear				X
BMP K4	MK				X
	Linear				X
BMP K5	MK				X
	Linear				X
BMP K6	MK	U			X
	Linear	U			X
BMP K7	MK		D	D	X
	Linear			D	X
BMP K8	MK				X
	Linear				X
BMP L1	MK		U		X
	Linear				X
BMP M1	MK				X
	Linear				X
BMP M2	MK				X
	Linear				X
BMP N1	MK				X
	Linear				X
BMP N3	MK				X
	Linear				X
BMP N4	MK				X
	Linear				X
BMP P1	MK		D		X
	Linear		D		X
BMP Q2	MK		D	D	X
	Linear		D		X
BMP R1	MK		D	D	X
	Linear		D	D	X
BMP R2	MK				X
	Linear				X
BMP R3	MK				X
	Linear				X
BMP R4	MK		D		X
	Linear		D		X
BMP R6	MK				X
	Linear				X
BMP R7	MK				X
	Linear				X
GAUSS7	MK				X
	Linear		U		X

Tab. 14: Ergebnisse der Trendanalysen für die Biomasse von Blaualgen

Blaualgen: Biomasse					
Station		1. Quartal	2. Quartal	3. Quartal	4. Quartal
BMP H1	MK	U			X
	Linear				X
BMP H2	MK				X
	Linear				X
BMP I1	MK				X
	Linear	D			X
BMP J1	MK		D		X
	Linear				X
BMP K1	MK			U	X
	Linear			U	X
BMP K2	MK		D		X
	Linear		D		X
BMP K4	MK				X
	Linear				X
BMP K5	MK				X
	Linear				X
BMP K6	MK				X
	Linear				X
BMP K7	MK	D	D	D	X
	Linear	D	D	D	X
BMP K8	MK				X
	Linear				X
BMP M1	MK				X
	Linear				X
BMP M2	MK				X
	Linear				X
BMP N3	MK		D		X
	Linear		D		X
BMP R3	MK		D		X
	Linear		D		X

Für das vierte Quartal lagen keine bzw. unzureichende Daten vor. Für die anderen Felder lagen Daten vor, doch bedeutet ein leeres Feld nicht, dass nicht möglicherweise doch ein Trend gegeben ist. So ergibt sich aufgrund der sehr unterschiedlichen Datenlage, aufgrund sehr unterschiedlicher Stichprobenumfänge und unterschiedlicher Variabilität auch eine sehr unterschiedliche Trendsensitivität des Testverfahrens. Ein leeres Feld besagt somit, dass keine Aussage hinsichtlich der Existenz eines Trends möglich ist.

Aus den Tabellen lässt sich ablesen, dass für Kieselalgen im ersten Quartal fallende Trends vorherrschen, während im dritten Quartal ansteigende Trends dominieren. Im zweiten Quartal fällt auf, dass sich für die Biomasse bis auf eine Ausnahme keine ansteigende Trends zeigen, während für die Abundanzen sowohl ansteigende als auch fallende Trends festgestellt werden können. Für Dinoflagellaten sind im ersten Quartal sowohl bei den Abundanzen als auch bei den Biomassen überwiegend steigende Trends festzustellen, während im zweiten Quartal sowohl ansteigende als auch fallende Trends auftreten. Für Blaualgen fällt die relativ hohe Anzahl fallender Trends im zweiten Quartal auf.

die Biomasse der Dinoflagellaten fällt die große Anzahl fallender Trends im zweiten Quartal auf.

Auf eingehendere Erklärungen zu den beobachteten Trends soll hier verzichtet werden, da diese von Biologie und Hydrographie abhängig sind und eine Vielzahl von Faktoren als Ursache für statistisch signifikante Trends in Frage kommen. Eine genaue Beurteilung der Ergebnisse erfordert daher eine sehr sorgfältige Bewertung. In diesem Zusammenhang sei beispielhaft auf die in einigen wenigen Fällen gegensätzliche Trendentwicklung in Frühjahr und Sommer hingewiesen, zum Beispiel bei der Biomasse der Dinoflagellaten bei Station BMP K4, die verschiedenste Ursachen haben kann.

2.7 Varianzkomponenten für EROD

Um zu überprüfen, inwieweit die EROD-Aktivität ein Biomarker ist, welcher für ein zeitliches Monitoring geeignet ist, werden in diesem Kapitel die Varianzkomponenten berechnet, welche die Streuung der EROD-Aktivität zwischen verschiedenen Individuen bzw. zwischen Jahren charakterisiert. Um Störungen durch Kofaktoren so weit als möglich auszuschließen, erfolgt die Schätzung getrennt in zwei Datenmengen, die separat in zwei Perioden, März und Herbst, erhoben wurden.

2.7.1 Varianz-Komponenten für log-EROD im März

Die Märzdaten der EROD-Aktivität wurden für eine eingehende Analyse ausgewählt, da hierfür ein umfangreicher Datensatz mit Daten aus einem Zeitraum von sechs Jahren vorliegt und ein nur geringer Anteil von Messungen unterhalb der Nachweisgrenze zu verzeichnen ist.

Die Schätzung der Varianz-Komponenten für die logarithmierte EROD-Aktivität erfolgte für Kliesche in der Größenklasse 20–25 cm, nach Geschlechtern getrennt. Diese Größenklasse entspricht den JAMP-Richtlinien für zeitliches Trendmonitoring

Die Varianzkomponenten wurden auch für die Größenklassen < 10 cm und 10–15 cm geschätzt. Für jedes Geschlecht und jede Größenklasse beinhaltet Tabelle 15 folgende Informationen:

- Die Zahl der Jahre T mit wenigstens einer Beobachtung;
- Den kleinsten und größten Wert von n_t , die Zahl der Beobachtungen in jedem Jahr (ausgenommen die Jahre, wo es überhaupt keine Daten gab);
- Die Schätzungen der Varianzkomponenten zwischen den Individuen, σ_w^2 , und zwischen den Jahren, σ_y^2 , von log-EROD, mit näherungsweisen Standardfehlern;
- Die Schätzungen der zugehörigen relativen Standardabweichungen $100\sigma_w$ und $100\sigma_y$; diese geben (näherungsweise) die prozentuale Variabilität der EROD-Aktivität an.

Tab. 15: Varianzkomponenten der EROD-Aktivität im März

Geschlecht	Größe (cm)	T	n_t		σ_w^2		$100\sigma_w$		σ_y^2		$100\sigma_y$	
			min	max	est	Standard- fehler			est	Standard- fehler		
Männlich	< 10	4	11	41	0.36	0.04	60		0.10	0.09	32	
	10–15	4	25	73	0.59	0.06	77		0.07	0.07	26	
	15–20	5	1	154	0.56	0.05	75		0.69	0.54	83	
	20–25	4	2	31	0.66	0.15	81		0.98	0.94	99	
Weiblich	< 10	4	15	77	0.52	0.08	72		0.00	0.01	0	
	10–15	5	3	50	0.69	0.09	83		0.00	0.01	0	
	15–20	7	4	74	1.43	0.15	120		0.50	0.36	71	
	20–25	5	5	41	1.57	0.24	125		0.95	0.77	97	

Die EROD-Aktivität bei älteren Weibchen (> 15 cm) unterliegt offenbar einer stärkeren Variabilität als bei jüngeren Weibchen (< 15 cm) oder bei Männchen. Das kann durch die Geschlechts- und Größendifferenzen verursacht sein, die sich in der Laichperiode besonders auswirken.

Die Schätzung der Variabilität der Population von Jahr zu Jahr unterliegt einem sehr großen Schätzfehler. Dies spiegelt die Schwierigkeit der Schätzung dieser Varianzkomponente für relativ kurze Zeitreihen wider.

Die genannten Varianzkomponenten werden für die Schätzung der statistischen Power (d. h. Wahrscheinlichkeit, daß ein vorhandener Trend auch tatsächlich erkannt wird) eines zeitlichen Monitoringprogramms der EROD-Aktivität für Kliesen verwendet. Die zugrundeliegende Theorie findet man bei Nicholson et al. (1997). Die untenstehende Tabelle beinhaltet die statistische Power zum Nachweis einer jährlichen Erhöhung bzw. Abnahme der EROD-Aktivität um 10 % bzw. 20 % über einen Zeitraum von 10 Jahren auf dem 5 % Signifikanzniveau mit einem Stichprobenumfang von jeweils 10 oder 20 Individuen. Zum Beispiel läßt sich bei männlichen Kliesen der Größenklasse 10-15 cm eine exponentielle Abnahme bzw. Zunahme der Erod-Aktivität um jährlich 10 % nach einem Zeitraum von 10 Jahren mit einer Power/Trennschärfe von 60 % bei einer Stichprobengröße von 10 Individuen nachweisen. Mit einer Stichprobengröße von jährlich 20 Individuen steigt die Power auf 99 %.

Tab. 16: Power für den Nachweis eines Trends der EROD-Aktivität bei Kliesen

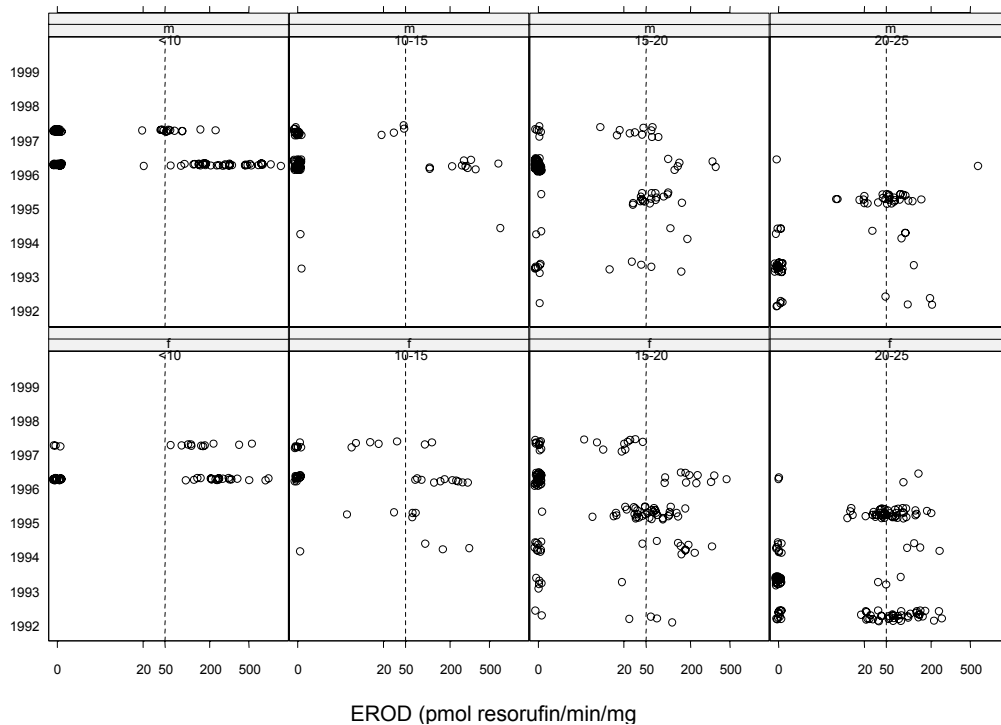
		Jährliche Veränderung: 10 %		Jährliche Veränderung: 20 %	
		10 Individuen	20 Individuen	10 Individuen	20 Individuen
männlich	< 10 cm	0.58	0.64	0.99	0.99
	10–15 cm	0.60	0.99	0.99	0.99
	15–20 cm	0.15	0.16	0.46	0.47
	20–25 cm	0.12	0.13	0.35	0.36
Weiblich	< 10 cm	0.93	0.99	0.99	0.99
	10–15 cm	0.86	0.99	0.99	0.99
	15–20 cm	0.17	0.19	0.51	0.56
	20–25 cm	0.12	0.12	0.33	0.35

Diese Ergebnisse sollten wegen des hohen Standardfehlers bei der Bestimmung von σ_y^2 nicht überinterpretiert werden. Zu beachten ist, dass der Stichprobenumfang (10 oder 20 Individuen) nur einen begrenzten Einfluss auf die Power hat und augenscheinlich nur bei männlichen Kliesen der Größenklasse 10-15 cm eine deutliche Wirkung hat. Problematisch ist ferner, dass die Interpretierbarkeit der Ergebnisse von EROD-Messungen in der Laichperiode eingeschränkt ist.

2.7.2 Varianz-Komponenten für log-EROD im Herbst

Um zu prüfen, ob die EROD-Aktivität im Herbst möglicherweise besser für ein zeitliches Monitoring geeignet ist, erfolgte eine weitere Analyse der EROD-Aktivitäten der Monate September, Oktober und November, die zusammengefasst betrachtet wurden, um ein ausreichend großes Datenkollektiv zu erhalten. Die Messwerte sind in Abbildung 29 nach Geschlecht, Jahr und Größenklassen getrennt dargestellt, wobei eine Darstellung in einer Skala der vierten Wurzel verwendet wurde, um die Schiefe der Verteilung zu reduzieren. Aufgrund eines hohen Anteils von Messungen unter der Nachweisgrenze lassen sich die Methoden aus Abschnitt 2.7.1 nicht einsetzen.

Abb. 29: Plot der EROD-Niveaus im Herbst getrennt nach Geschlecht, Jahr und Größenklasse
(Die Plot-Symbole sind mit einer zusätzlichen Zufallskomponente versehen, um den Umfang der Messungen unter der Bestimmungsgrenze zu verdeutlichen. Die senkrechte gestrichelte Linie in jeder Einzelgrafik entspricht der Schwelle von 50 pmol resorufin/min/mg protein.)



Die statistische Modellierung von EROD-Daten mit einem nicht unerheblichen Anteil von Messungen unter der Bestimmungsgrenze wirft eine Reihe methodischer Probleme auf. Ein vergleichsweise einfacher Ansatz besteht darin, nicht die EROD-Werte selbst zu betrachten, sondern nur den Anteil jener Werte, die oberhalb eines geeigneten Schwellenwertes liegen, der als signifikant erhöhte EROD-Aktivität aufgefasst werden könnte. Vorausgesetzt, dass der Schwellenwert über der Nachweisgrenze liegt, kann eine individuelle EROD-Messung immer als über oder unter dem Schwellenwert liegend klassifiziert werden. Für den prozentualen Anteil der Individuen mit erhöhter EROD-Aktivität kann nun ein Trendmodell angepasst werden und die entsprechende Power berechnet werden, mit der ein Anstieg des Anteils der Individuen mit erhöhter EROD-Aktivität nachgewiesen werden kann.

Um dies zu illustrieren, werden die Daten für Kliesen zwischen 15 und 20 cm betrachtet und der Schwellenwert mit 50 pmol resorufin/min/mg protein festgelegt. In Tabelle 17 bezeichnet n_t die Anzahl der Individuen zwischen 15 und 20 cm der Probe des Jahres t , und y_t ist die Anzahl der Individuen mit erhöhter EROD-Aktivität (über dem Schwellenwert). Wenn p_t die Wahrscheinlichkeit bezeichnet, dass ein Fisch im Jahr t eine erhöhte EROD-Aktivität aufweist, kann für y_t eine binomiale Verteilung angenommen werden gemäß

$$y_t \sim \text{Binomial}(n_t, p_t). \quad (\text{Gleichung 87})$$

Die untenstehende Tabelle gibt die Werte von n_t , y_t , und die Schätzungen von p_t , getrennt nach Geschlecht und Jahr wieder.

Tab. 17: EROD-Aktivität von Kliesen. Werte von n_t , y_t , und die Schätzungen von p_t , getrennt nach Geschlecht und Jahr

Jahr	Weibchen			Männchen		
	n_t	y_t	p_t	n_t	y_t	p_t
1992	6	3	0.50	1	0	0.00
1993	6	0	0.00	12	2	0.17
1994	19	9	0.47	4	2	0.50
1995	38	21	0.55	19	11	0.58
1996	39	11	0.28	51	6	0.12
1997	19	0	0.00	17	4	0.24

Es ist kein Hinweis von systematischen Trends in den Wahrscheinlichkeiten p_t zu erkennen, aber es sind deutliche Schwankungen festzustellen, insbesondere bei den Weibchen. Analog zu den log-EROD-Aktivitäten im März kann vermutet werden, dass es einen zufälligen Jahreseffekt gibt, der diese Variabilität erklärt.

Ein verallgemeinertes lineares gemischtes Modell liefert den statistischen Rahmen für die Modellierung dieser Daten. Vorausgesetzt es liegt kein systematischer Trend vor, kann das folgende Modell separat für beide Geschlechter angepasst werden. Wie oben wird für die Beobachtungen y_t angenommen, dass sie binomial verteilt sind, $y_t \sim \text{Binomial}(n_t, p_t)$. Weiterhin wird angenommen, dass die zugrundeliegenden Wahrscheinlichkeiten stochastisch variieren mit

$$\text{s logit}(p_t) = \log\left(\frac{p_t}{1-p_t}\right) \sim N(\mu, \sigma_y^2). \quad (\text{Gleichung 88})$$

Die Logit-Transformation wird hierbei verwendet, um die Wahrscheinlichkeiten auf den gesamten Definitionsbereich der Normalverteilung umzurechnen. μ bezeichnet den im Logit-Modell ausgedrückten Mittelwert, und σ_y^2 ist die Varianzkomponente für den zufälligen Jahreseffekt. Die untenstehende Tabelle 18 gibt die Schätzungen von μ und σ_y^2 nach Geschlecht mit dem approximativen Standardfehler wieder.

Tab. 18: Schätzergebnisse für das Logit-Modell auf Basis der EROD-Aktivität im Herbst

	μ		σ_y^2	
	Schätzung	Standardfehler	Schätzung	Standardfehler
Weiblich	-1.08	0.66	2.03	1.62
Männlich	-1.01	0.47	0.76	0.80

Die Mittelwerte μ beziehen sich auf die logistische Skala. Nach Rücktransformation zeigt sich, dass im Durchschnitt bei ungefähr 25 % der weiblichen Tiere und 27 % der männlichen Tiere eine erhöhte EROD-Aktivität festzustellen ist.

Um die Power eines Monitoringprogramms zum Nachweis eines zeitlichen Trends zu ermitteln, wird angenommen, dass in jedem Jahr $t = 1 \dots T$ bei je n männlichen Tieren (weiblichen Tieren) in der Größenklasse von 15 bis 20 cm die EROD-Aktivität ermittelt wird. Dann sind y_t und p_t wie im vorigen definiert, aber nun wird folgendes vorausgesetzt:

$$\text{logit}(p_t) = \log\left(\frac{p_t}{1-p_t}\right) \sim N(\mu_t, \sigma_y^2), \quad (\text{Gleichung 89})$$

mit

$$\mu_t = \mu_1 + \beta(t-1). \quad (\text{Gleichung 90})$$

μ_t bezeichnet nun die in logistischer Skalierung ausgedrückte Wahrscheinlichkeit, mit der der Schwellenwert überschritten wird, wobei angenommen wird, dass sich μ_t über die Zeit linear verändert, wobei β die Stärke des Trends darstellt. Wird das Hauptaugenmerk darauf gelegt, einen etwaigen Anstieg der EROD-Aktivität zu entdecken, könnte die Nullhypothese und die Alternativhypothese folgendermaßen aussehen:

$$H_0: \beta \leq 0$$

$$H_1: \beta > 0.$$

Diese Hypothesen können durch die Schätzung von β unter Verwendung eines verallgemeinerten gemischten linearen Modells getestet werden, wobei $\hat{\beta}/\text{s.e.}(\hat{\beta})$ mit dem $100(1 - \alpha)$ %-Quantil der Standardnormalverteilung verglichen wird. Dabei bezeichnet $\hat{\beta}$ ist die Schätzung von β und $\text{s.e.}(\hat{\beta})$ ist der Standardfehler von $\hat{\beta}$. α bezeichnet das vorgegebene Signifikanzniveau.

Die untenstehende Tabelle 19 gibt die Power eines 10-Jahres-Programms zum Nachweis einer ansteigenden EROD-Aktivität auf dem Signifikanzniveau von 5 % wieder. Dabei wird ein Stichprobenumfang von jährlich 10 bzw. 20 Individuen unterstellt. Die Power ist von β und μ_1 abhängig. Für μ_1 wird mit dem Wert von -1 der aktuelle status quo unterstellt (entsprechend einem Anteil von 27 %). Um die Interpretation zu erleichtern, werden auch die zurücktransformierten Werte von μ_t , das Verhältnis von Individuen oberhalb der Schwellenwerte in den letzten Jahren des Monitorings, dargestellt. Die Power wird durch Simulationen berechnet werden, wobei aufgrund des hohen Rechenaufwandes jeder Wert in der Tabelle auf nur 100 unabhängigen Realisierungen basiert. Dies erklärt auch die deutlichen Schwankungen der Ergebnisse.

Tab. 19 Power zum Nachweis eines Trends der EROD-Aktivität im Herbst auf Basis eines 10-Jahresprogramms

β	$\text{logit}^{-1}(\mu_t)$ %	Weiblich		Männlich	
		$n = 10$	$n = 20$	$n = 10$	$n = 20$
0.0	27	0.11	0.08	0.01	0.03
0.1	47	0.18	0.12	0.25	0.26
0.2	69	0.25	0.39	0.53	0.51
0.3	85	0.45	0.54	0.77	0.79
0.4	93	0.70	0.68	0.88	0.97
0.5	97	0.83	0.90	0.96	0.99

Für $\beta = 0.3$ würde der Anteil der Individuen oberhalb des Schwellenwerts über die 10-Jahres-Monitoringperiode von 27 % auf 85 % ansteigen. Mit jährlich 10 gemessenen Weibchen ergibt sich eine approximative Power von 45 %. Vergrößert man die Stichprobe auf 20 weibliche Fische, steigt die Power auf ungefähr 54 %. Für männliche Tiere betragen die entsprechenden Werte 77 % und 79 %. Die erhöhte Power für männliche Individuen ist dadurch bedingt, dass hier die zufälligen Abweichungen von Jahr zu Jahr geringer sind. Anzumerken ist zunächst,

dass einer Verdoppelung des Stichprobenumfangs eine vergleichsweise geringe Verbesserung der Power mit sich bringt. Auffällig ist auch der Umstand, dass nur ein sehr drastischer Anstieg der EROD-Aktivität in der gesamten Population eine Power von mindestens 90 % ermöglicht.

2.8 Empfehlungen

2.8.1 Univariate Daten

Die im Rahmen des BLMP eingesetzten univariaten Verfahren sind für die meisten Anwendungen praktikabel, jedoch unter gewissen Voraussetzungen, speziell bei stark ausreißerbelasteten oder linksschiefen Verteilungen, möglicherweise in manchen Fällen irreführend, so dass sich eine zusätzliche Überprüfung der Daten anhand weiterer Kriterien empfiehlt. Hierfür geeignet sind z.B. Boxplots, welche die Verteilung der Daten grafisch anhand der Extremwerte und der Quartile verdeutlichen, oder nichtparametrische Parameter für Schiefe und Kurtosis, welche die Abweichungen von der Normalverteilung messen. In den marktgängigen Softwarepaketen sind verschiedene Parameter verfügbar, so dass hier auf eine Festlegung verzichtet werden soll. Es ist zu betonen, dass für eine beschreibende Darstellung von univariaten Daten keinerlei Verteilungsanforderungen notwendig sind. Gleichwohl ist es immer notwendig, sich den Verteilungstyp der betrachteten Daten vor Augen zu halten, um Fehlinterpretationen der Resultate zu vermeiden. Es kann daher auch sinnvoll sein, Tests auf Normalverteilung durchzuführen, obwohl eine „blinde“ Fixierung auf die Verteilungsannahmen nicht immer hilfreich ist. Zugespitzt könnte man formulieren, dass die Beantwortung der Frage, ob Daten als normalverteilt angesehen werden können, letztlich alleine vom Stichprobenumfang abhängig ist. Die Normalverteilung ist ein theoretisches Konstrukt, und in der Realität kommt diese Verteilung nur näherungsweise vor. Es ist deshalb tatsächlich nur eine Frage des Stichprobenumfangs, ob diese größeren oder kleineren Abweichungen von der Normalverteilung tatsächlich entdeckt werden oder nicht. Wichtig aus statistischer Sicht erscheint deshalb, dass die in der Statistik verfügbaren Kennwerte kritisch verwendet werden, wobei sowohl „robuste“ als auch weniger robuste Methoden eingesetzt werden sollten. Der Median hat seine Berechtigung ebenso wie das arithmetische Mittel, und es kommt nicht nur darauf an, ob ein statistischer Kennwert grundsätzlich einsetzbar oder nicht, sondern ob dieser Kennwert liefert, was man von ihm erwartet. Es wird daher empfohlen, in der Diskussion dieser Kennwerte immer die reale Fragestellung zu berücksichtigen. Es könnte auch hilfreich sein, für konkrete, wiederkehrende Fragestellungen in den BLMP-Berichtsheften beispielsweise im Rahmen eines Workshops entsprechende Richtlinien zur Anwendung und Interpretation statistischer Kennwerte zu erarbeiten.

2.8.2 Trendanalysen

Jede der bislang zur Auswertung von marinen Daten im nationalen oder internationalen Rahmen verwendeten Trendanalysemethoden hat spezifische Vor- und Nachteile, so dass empfohlen wird, flexible Protokolle für die Trendanalyse zu erarbeiten, die den jeweiligen Anforderungen möglichst nahe kommen. Grundlage dieser Protokolle sollte die in Kapitel 3 dargestellte Konzeption einer Trendanalyse sein. Die Erarbeitung eines solchen Protokolls sollte in mehreren Schritten erfolgen: Zunächst kann anhand vorliegender Daten eine exemplarische Analyse erfolgen, die dann beispielsweise im Rahmen von Workshops diskutiert und überarbeitet wird, mit dem Ziel, zu verbindlichen Festlegungen zu kommen. Die internationale Einbindung solcher Arbeiten sollte gewährleistet sein.

Eine vielfach diskutierte und grundlegende Frage bei der Festlegung eines Protokolls ist das Kriterium der Robustheit. Dieses ist nicht in allen Fällen gleichermaßen zu bewerten. So ist es allein von der Fragestellung abhängig, ob ein Ausreißer möglichst geringen Einfluss haben soll und gegebenenfalls eliminiert bzw. ein ausreißer-robustes Verfahren eingesetzt werden soll, oder ob für das Monitoringprogramm gerade die Erfassung von Ausreißern (z.B. durch Unfälle verursacht) höchste Priorität haben. Bei dieser Betrachtung ist auch zu prüfen, ob es sich um Einzelausreißerwerte innerhalb eines Jahres handelt (bei einzelnen Messungen bzw. Individuen), oder ob die Gesamtheit aller Messwerte eines Jahres bezüglich ihres Mittelwertes als Ausreißer anzusehen ist. So sind für manche Fragestellungen Ausreißer auf der Messwerte-Ebene möglichst zu eliminieren, während Ausreißer auf der aggregierten Ebene von großem Interesse sind. Schließlich ist darauf hinzuweisen, dass Verfahren, die für normalverteilte Daten konzipiert wurden, vielfach auch für nichtnormalverteilte Daten einsetzbar sind. So ist aufgrund des Zentralen Grenzwertsatzes davon auszugehen, dass selbst bei stark nicht-normalverteilten Daten der zugehörige Mittelwert näherungsweise normalverteilt ist, mit der Folge, dass Verfahren auf der Basis der Normalverteilung problemlos eingesetzt werden können. Selbst wenn dies nicht gewährleistet ist, lassen sich Verfahren zur Trendanalyse zumindest zur Bestimmung der Trendlinie in der Regel verwenden. In diesem Falle ist nur zu beachten, dass die Testergebnisse mit Vorsicht zu betrachten sind. Gleichwohl hat es sich eingebürgert, dass selbst bei gewissen Verletzungen der Modellannahmen diese Testergebnisse – in Ermangelung anderer Verfahren – akzeptiert werden, zumindest dann, wenn ein entsprechendes Protokoll zur Vorgehensweise vorliegt.

2.8.3 Evaluation von Monitoringprogrammen

Zur Beurteilung eines Monitoringprogramms hat sich die Durchführung von Poweranalysen bewährt. Es wird daher empfohlen, derartige Poweranalysen als Qualitätssicherungsinstrument für Monitoringprogramme generell durchzuführen, um Monitoringprogramme besser beurteilen zu können. Dies kann im Einzelfall auch dazu führen, dass ein Messprogramm korrigiert wird. Beispielhaft werden in dem vorliegenden Bericht zwei Poweranalysen für EROD-Daten vorgestellt. Die Ergebnisse belegen, dass die Power für ein EROD-Monitoring nicht befriedigend ist.

Poweranalysen sind in besonderem Maße zweckmäßig für Monitoringprogramme, die darauf ausgelegt sind, einen systematischen Trend nachweisen zu können. Dabei ist allerdings festzuhalten, dass die Genauigkeit der Poweranalyse sehr stark von der Genauigkeit der zugrundeliegenden Varianzkomponentenschätzungen abhängig ist. In der Vergangenheit hat sich die Statistische Arbeitsgruppe des ICES (WGSAEM) mehrfach mit der Ermittlung dieser Varianzkomponenten befasst, so zum Beispiel bei der Untersuchung der Variabilität von Schadstoffgehalten in Vogeleiern (WGSAEM 1999) und im Rahmen des VIC-Programms (WGSAEM 1999+2000). Die genannten Untersuchungen haben allerdings eine sehr begrenzte Aussagekraft, da wichtige Varianzkomponenten entweder ganz fehlen oder der Schätzfehler sehr groß ist. Daraus ist schließlich die Empfehlung abzuleiten, dass die zur Beurteilung der Power benötigten Daten der Statistischen Arbeitsgruppe des ICES oder entsprechenden Fachleuten für Statistik zur Verfügung gestellt werden sollten, um die jeweilige Power zu ermitteln.

2.8.4 Methodenentwicklung

2.8.4.1 Weitere Verbesserungen der auf dem LOESS-Smoother basierenden Trendanalyse

Die große Bedeutung des LOESS-Smothers, bekannte Probleme dieses Verfahrens und die große Bandbreite der Daten machen eine ganze Reihe von Weiterentwicklungen erforderlich.

2.8.4.1.1 Reduzierung künstlicher Schwingungen in der Trendlinie

Durch geeignete Modifikation der LOESS-Matrix ließen sich künstliche Schwingungen am Anfang und Ende der Zeitreihe vermeiden: Die aktuell implementierte Glättung hat eine

konstante Bandbreite von sieben Jahren, was bedeutet, dass die vorhergesehenen Niveaus im entscheidenden Jahr durch die letzten sechs Jahre beeinflusst ist, während in der Mitte des Zeitraums die maximale Entfernung nur drei Jahre beträgt. Das bedeutet unter widrigen Bedingungen, dass künstliche Trends am Anfang und Ende der Zeitreihe erscheinen. Methoden mit variabler Bandbreite erscheinen daher zweckmäßig, um ein Überspringen zu vermeiden. Es sollte weiterhin geprüft werden, ob der Einsatz von Methoden mit stetig veränderlicher Bandbreite bessere Ergebnisse ermöglicht. In diesem Zusammenhang ist insbesondere der Spline-Smoother zu benennen, der in einem gewissen Sinne optimal ist.

2.8.4.1.2 Robuste Glättungsverfahren

Der LOESS-Smoother ist sehr stark ausreißerabhängig, und speziell bei Ausreißern am Anfang und Ende der Zeitreihe können grobe Verzerrungen auftreten. Es erscheint daher zweckmäßig, eine geeignete Robustifizierung des Glättungsverfahrens zu entwickeln. In diesem Zusammenhang könnte auch die Entwicklung von Verfahren von Interesse sein, bei denen die Bandbreite und der Einfluss einzelner Daten von einem Uncertainty-Wert abhängig gemacht wird, der entweder aus den Daten selbst ermittelt wird, oder der auf der Basis des jeweiligen QS-Standards bestimmt wird. Verwiesen sei hierbei auch auf die in der Gruppe MON unter OSPAR diskutierten Vorschläge (Nicholson, Fryer 2001). Vielversprechend erscheint auch die Entwicklung eines robusten Verfahrens mit dynamisch bestimmtem Glättungsparameter.

2.8.4.1.3 Power und nachweisbarer Trend

Um die Eignung eines Monitoringprogramms für einen bestimmten Zweck schnell überprüfen zu können, sollten explizite Formeln für den nachweisbaren Trend sowie die Power unter unterschiedlichen Voraussetzungen erarbeitet werden.

2.8.4.1.4 Erarbeitung von Protokollen zur Charakterisierung der Zeitreihen

Generell erforderlich erscheint die Erarbeitung von Protokollen, in denen die im marinen Monitoring vorkommenden Zeitreihen hinsichtlich typischer Eigenschaften klassifizierbar gemacht werden können. Dies betrifft z.B. ihre Charakterisierung hinsichtlich der dominierenden Elemente der Variabilität, also z.B. Saison, Schwankungen von Jahr zu Jahr, sowie die Häufigkeit des Auftretens von Ausreißern. Hilfreich wäre ein solches Protokoll, um dann in einem zweiten Schritt möglichst schnell zu einer geeigneten Verfahrensweise bei der Trendermittlung kommen zu können. Diese Arbeit ist auch im Zusammenhang mit der

Durchführung von Poweranalysen zu sehen, die zur Bestimmung der Trendsensitivität dienen sollen.

2.8.4.1.5 Trendanalyse auf der Basis Gamma-verteilter Daten

Da der LOESS-Smoother implizit symmetrisch verteilte Daten mit konstanter Varianz unterstellt, ist die Anwendung des Verfahrens naturgemäß eingeschränkt. Dies wurde u.a. bei der Auswertung der HELCOM-Daten deutlich, für die eine Logarithmierung der Daten unumgänglich wurde, um sinnvolle Ergebnisse zu erhalten. Da durch eine Logarithmierung jedoch zugleich Abundanzspitzen „heruntergewichtet“ werden, erscheint die Entwicklung spezifisch angepasster Methoden zwingend erforderlich. Hierzu bietet sich an, das LOESS-Verfahren für Gamma-Verteilungen zu erweitern. Die Gamma-Verteilung ist nichtnegativ und rechtsschief, und entspricht in ihrem Charakter häufig den tatsächlich beobachteten Verteilungen. Beispiel einer Gamma-Verteilung ist die bekannte Chi-Quadrat-Verteilung: Die Gamma-Verteilung kann als Verallgemeinerung dieser Verteilung angesehen werden.

2.8.4.2 Raum-zeitliche Analysen

Nur geringe Erfahrungen bestehen bislang mit räumlichen oder gar raum-zeitlichen Analysen, bei denen bislang nur erste Vorarbeiten auf der Basis des GAM-Modells realisiert werden konnten. Statistische Arbeiten zur Erfassung raum-zeitlicher Trends sind bislang noch recht selten, so dass hier noch keine abschließende Bewertung möglich ist. Das methodische Instrumentarium ist sehr anspruchsvoll und die Ergebnisse manchmal nicht einfach zu interpretieren. Gleichwohl wird der Einsatz raum-zeitlicher Analysemethoden schon deshalb als wesentlich angesehen, weil eine simultane Verwertung von Daten unterschiedlicher Stationen eine deutlich höhere Power bzw. Trennschärfe bezüglich einer Erfassung von zeitlichen Trends ermöglicht, als wenn jeweils nur isolierte Analysen für die einzelnen Stationen erfolgen. Um die Entwicklung dieser Methoden voranzutreiben, wird empfohlen, eine stationsübergreifende Analyse der HELCOM-Daten zu initiieren, wobei ein erweitertes GAM-Modell entwickelt werden sollte.

2.8.4.3 Poweranalysen

Die Erfassung der einzelnen Varianzkomponenten, welche das stochastische Verhalten der Zeitreihen bestimmen, ist ein wesentliches Hilfsmittel, um adäquate Stichprobenplanungen und

Monitoringkonzepte entwickeln zu können. Dabei ist allerdings zu beachten, dass die bislang verwendeten Verfahren sehr ausreißerempfindlich sind und zudem wegen der meist geringen Stichprobenumfänge die Ergebnisse nicht sehr zuverlässig sind. Es stellt sich daher die Aufgabe, das Konzept der Poweranalysen unter Verwendung robuster Methoden zu verbessern. Es wird empfohlen, diese Arbeiten im Zusammenhang mit der Realisierung konkreter Powerstudien zur Beurteilung von Monitoringprogrammen durchzuführen.

2.9 Zusammenfassung

Gegenstand dieses Teilvorhabens ist eine Übersichtsdarstellung sowie Bewertung bislang im BLMP verwendeter Methoden, wobei neben univariaten Kennwerten Trendanalysen im Vordergrund stehen. Weiterhin beinhaltet der Bericht eine Übersicht über die Methodik zur Trendanalyse auf Basis von Jahresdaten mittels des LOESS-Smoothers. Auf Basis dieses Verfahrens wurden Trendanalysen zu Biomasse und Abundanz ausgewählter Artengruppen des HELCOM-Programmes durchgeführt, die ebenfalls vorgestellt werden. Parallel wurden Trendanalysen auf Basis des Tests von Mann-Kendall durchgeführt. Erweiterungsmöglichkeiten zur Berücksichtigung von störenden Kofaktoren und für monatliche Daten werden entwickelt und diskutiert, wobei exemplarisch die Ergebnisse einer statistischen Analysen der Quecksilbergehalte bei Schollen in der Deutschen Bucht präsentiert werden. Einen weiteren Schwerpunkt der Arbeit bilden Methoden zur Überprüfung der Eignung von Monitoringprogrammen für die jeweils vorgesehenen Zwecke. Diese Methodik wird exemplarisch für die EROD-Aktivität von Klieschen vorgestellt.

2.10 Literatur

BLMP Berichtsheft (1993):

Gemeinsames Bund/Länder-Meßprogramm für die Nordsee-Gewässergütemessungen im Küstenbereich der Bundesrepublik Deutschland, Gütedaten 1990/91; Hannover

BLMP Berichtsheft (1994):

Gemeinsames Bund/Länder-Meßprogramm für die Nordsee-Gewässergütemessungen im Küstenbereich der Bundesrepublik Deutschland, Gütedaten 1992/93; Hannover

Cleveland, W.S. (1979):

Robust locally-weighted regression and smoothing scatterplots. Journal von the American Statistical Association 74 829-836.

Frank, B. (1997):

Statistische Verfahren zur Beobachtung und Bewertung von Schadstoffeinträgen in die Hohe See im Rahmen des Bund-Länder-Messprogramms zur Überwachung von Nord- und Ostsee

Fryer R.J.; Nicholson, M.D. (1999):

Using smoothers for comprehensive assessments of contaminant time series in marine biota. Submitted to ICES J. Mar. Sci.

Fryer R.J.; Nicholson, M.D. (1999):

Assessing Power of Trend Tests for Monotonic Trends. In ICES-WGSAEM Report 1999, Annex 4.

Hastie T.J.; Tibshirani R.J. (1990):

Generalized Additive Models. Chapman und Hall, London.

Hebbel, H. (1997):

Verallgemeinertes Berliner Verfahren VBV. In K. Edel, K.-A. Schäffer und W. Stier: Analyse saisonaler Zeitreihen. Physica-Verlag, 83-93.

ICES-WGSAEM Report 1998

ICES-WGSAEM Report 1999

ICES-WGSAEM Report 2000

ICES-WGSAEM Report 2001

Kendall, M.G. (1975):

Rank Correlation Methods, 4th ed. Charles Griffin, London.

Maxwell, D. (1999):

The Compound Mann-Kendall Test. In ICES-WGSAEM Report 1999, Annex 6.

MURSYS (1998):

Jahresbericht 1998: Meeresumwelt-Reportsystem. Herausgegeben vom BSH.

Nicholson, M.D.; Fryer R.J. (2001):

Weighting Procedures for Assessing Trend Data of Variable Quality. In ICES-WGSAEM Report 2001, Annex 5.

Nicholson, M.D.; Fryer R.J.; Maxwell D.L. (1997):

Report von the ICES/OSPAR Workshop on the identification von statistical methods for temporal trends. Annex 9. A study of the power von various methods for detecting trends. ICESCM 1997/Env:11.

Nicholson, M.D.; Fryer, R.J.; Maxwell, D. (1997):

A study of the power of various methods for detecting trends. Report of the ICES/OSPAR, Workshop on the Identification von Statistical Methods Für Trend Detection. ICES CM 1997/Env:11.

Nicholson, M.D.; Fryer, R.J.; Maxwell, D. (1998):

The influence von individual outlying observations on four methods von detecting trends. In ICES-WGSAEM Report 1998. ICES CM 1998/E:8.

Uhlig, S. (1996):

Spatio-temporal analysis von contaminant-concentrations in water und sediment. In ICES-WGSAEM Report 1996.

Uhlig, S (1999):

Assessing Accuracy von Estimators for the Percentage of Reduction of the Contaminant Level. In ICES-WGSAEM Report 1999, Annex 5.

Uhlig, S (2001):

The LOESS smoother: incorporation of uncertainty data and the behavior with missing values. In ICES-WGSAEM Report 2001, Annex 6.

2.11 Abbildungsverzeichnis

Abb. 1:	LOESS-Trend bei einer Zeitreihe mit einem stark abweichenden Wert	139
Abb. 2:	Robustifizierter LOESS-Trend bei einer Zeitreihe mit einem stark abweichenden Wert	139
Abb. 3:	Linearer Trend mit 90%-Konfidenzintervall für eine nichtlineare Zeitreihe	156
Abb. 4:	Loess-Trend einer nicht-linearen Zeitreihe mit einem 90%-Konfidenzintervall	158
Abb. 5:	90% Vorhersageintervall für eine nichtlineare Zeitreihe	160
Abb. 6:	Darstellung des linearen Trends als gewichtetes Mittel der Veränderungsraten	162
Abb. 7:	LOESS-Trend bei über mehrere Jahre wirksamen, klimatisch bedingten Schwankungen	163
Abb. 8:	LOESS-Trend und linearer Trend bei abruptem Rückgang der Konzentrationswerte	164
Abb. 9:	Ortho-Phosphat bei Herbrum/Ems [mg/l] 1984-1997	169
Abb. 10:	LOESS-Trend basierend auf den mittleren Jahreskonzentrationen, berechnet mit einer Fensterbreite von 7 Jahren	169
Abb. 11:	LOESS-Trend basierend auf den mittleren Jahreskonzentrationen, berechnet mit einer Fensterbreite von 5 Jahren	170
Abb. 12:	LOESS-Trend basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 85 Monaten	171
Abb. 13:	Jahresmittelwert des LOESS-Trends, basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 85 Monaten	172
Abb. 14:	LOESS-Trend basierend auf den Monatswerten, berechnet mit einer Fensterbreite von 49 Monaten	172
Abb. 15:	Hg-Konzentrationen (Einzelwerte) im Zeitablauf	175
Abb. 16:	Hg-Konzentrationen (Einzelwerte) in Abhängigkeit von der Länge	176
Abb. 17:	Beziehung zwischen Gewicht und Länge der Schollen	177
Abb. 18:	Länge der Schollen im Zeitablauf	178
Abb. 19:	Residualplot des GAM-Modells auf Basis der Jahresdaten	183
Abb. 20:	Trend des GAM-Modells auf Basis der Jahresdaten	183

Abb. 21:	Trend des GAM-Modells auf Basis der Einzelwerte	185
Abb. 22:	Residualplot des GAM-Modells auf Basis der Einzelwerte	185
Abb. 23:	Datenblatt Trendanalyse der Abundanz von Kieselalgen, Station BMP K4, Teil A	190
Abb. 24:	Datenblatt Trendanalyse der Abundanz von Kieselalgen, Station BMP K4, Teil B	191
Abb. 25:	Datenblatt Trendanalyse der Abundanz von Dinoflagellaten, Station BMP K2, Teil A	192
Abb. 26:	Datenblatt Trendanalyse der Abundanz von Dinoflagellaten, Station BMP K2, Teil B	193
Abb. 27:	Datenblatt Trendanalyse der Biomasse von Blaualgen, Station BMP K2, Teil A	194
Abb. 28:	Datenblatt Trendanalyse der Biomasse von Blaualgen, Station BMP K2, Teil B	195
Abb. 29:	Plot der EROD-Niveaus im Herbst getrennt nach Geschlecht, Jahr und Größenklasse	207

2.12 Tabellenverzeichnis

Tab. 1:	Eignung verschiedener Trendtests	141
Tab. 2:	Kritische Werte für den Test auf Nichtlinearität	151
Tab. 3:	Kritische Werte für den Test auf Nichtlinearität, Approximation zweiter Ordnung	153
Tab. 4:	Berechnung des Regressionskoeffizienten als gewichtete Summe	162
Tab. 5:	Interpretation der Testergebnisse	164
Tab. 6:	Analysen der Regressionsanalysen für die Einzeljahre	179
Tab. 7:	Vergleich von arithmetischen Mittel und längenadjustiertem Mittelwert	181
Tab. 8:	Festlegung der „Jahreszeiten“	187
Tab. 9:	Ergebnisse der Trendanalysen für die Abundanz von Kieselalgen	197
Tab. 10:	Ergebnisse der Trendanalysen für die Biomasse von Kieselalgen	198
Tab. 11:	Ergebnisse der Trendanalysen für die Abundanz von Dinoflagellaten	199
Tab. 12:	Ergebnisse der Trendanalysen für die Biomasse von Dinoflagellaten	200
Tab. 13:	Ergebnisse der Trendanalysen für die Abundanz von Blaualgen	201
Tab. 14:	Ergebnisse der Trendanalysen für die Biomasse von Blaualgen	202
Tab. 15:	Varianzkomponenten der EROD-Aktivität im März	205
Tab. 16:	Power für den Nachweis eines Trends der EROD-Aktivität bei Klieschen	206
Tab. 17:	EROD-Aktivität von Klieschen. Werte von n_t , y_t , und die Schätzungen von p_t , getrennt nach Geschlecht und Jahr	208
Tab. 18:	Schätzergebnisse für das Logit-Modell auf Basis der EROD-Aktivität im Herbst	209
Tab. 19:	Power zum Nachweis eines Trends der EROD-Aktivität im Herbst auf Basis eines 10-Jahresprogramms	210

2.13 Verzeichnis der verwendeten Gleichungssymbole

$x_1, \dots, x_N$	Stichprobe
$x_{(1)}, \dots, x_{(N)}$	Nach der Größe sortierte Stichprobe
$\bar{x}$	Arithmetisches Mittel
$x_{0,5}$	Median
$\bar{x}_g$	Geometrisches Mittel
s	Standardabweichung
R	Spannweite
x_w, x_o	Unter- und Obergrenze eines Konfidenzintervalls
$t_{N-1, 1-\alpha/2}$	$(1 - \alpha / 2)$ -Quantil der t-Verteilung mit $(N-1)$ Freiheitsgraden
α	Signifikanzniveau
$Y(t)$	Zeitreihe (stochastischer Prozess)
$u(t)$	Trend
$s(t)$	Saison
$e(t)$	Zufällige Abweichung
μ	Parameter für den Mittelwert
β	Parameter für einen zeitlichen Trend
$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	Parameter für die Saisonfigur
t	Zeitpunkt
y_i	Beobachteter Wert
i, j	Indices
S_{MK}	Prüfgröße des Mann-Kendal-Tests
sgn	Vorzeichenoperator
x	Platzhalter
Y	Beobachtungsvektor
X	Designmatrix
W	Gewichtungsmatrix
w	Gewicht
Δ	Bandbreite
$span$	Zeitspanne der zu berücksichtigenden Daten

n	Stichprobenumfang
t	Teststatistik des t-Tests
df	Freiheitsgrad
t_p	p-Quantil der t-Verteilung
SS	Quadratsumme
a, b	Regressionskoeffizienten
y	Vektor der beobachteten Daten
I	Einheitsmatrix
σ^2	Varianz
F	F-verteilte Teststatistik
$F(df_1, df_2, 1 - \alpha)$	$(1 - \alpha)$ -Quantil der F-Verteilung mit (df_1, df_2) Freiheitsgraden
tr	Spur-Operator
$\chi(df)$	Chi-Quadrat-Verteilung mit df Freiheitsgraden
K, L	Matrizen
$\delta(m)$	Stichprobenfaktor
m	Index
$z(m)$	Trendschätzung im Jahr m
k	Index
x_i, x_t	Messzeitpunkt
B	Steigungsparameter
σ_y^2	Varianzkomponente für den Jahreseffekt
σ_w^2	Varianzkomponente für den Individueneffekt
T	Anzahl der untersuchten Jahre
$Logit$	Logit-Operator
$N(\mu, \sigma^2)$	Normalverteilung

2.14 Abkürzungsverzeichnis

BLMP	Bund/Länder-Messprogramm für Ost- und Nordsee
ICES	International Council for the Exploration of the Sea
WGSAM	Working Group on the Statistical Aspects of Environmental Monitoring
LOESS	Glättungsverfahren auf Basis einer lokalen gewichteten Regression
HELCOM	Helsinki-Kommission
MUDAB	Meeresumweltdatenbank
MURSYS	Meeresumwelt-Reportsystem
ACME	Advisory Committee of the Marine Environment des ICES
OSPAR	Oslo-Paris-Kommission
Min	Minimum
Max	Maximum
NPar	Nichtparametrischer Wert
MK	Mann-Kendall
VIC	Voluntary International Contaminant-Monitoring for Temporal Trends

Danksagung

An dieser Stelle sei Frau Dr. Carstens (UBA) für ihr großes Engagement für das Forschungsprojekt gedankt. Danken möchten wir auch Herrn Dr. Wasmund (Institut für Ostseeforschung, Warnemünde) sowie Herrn Dr. Harms von der Bundesforschungsanstalt für Fischerei für die Bereitstellung von Daten, wertvolle Diskussionen und Anregungen. Zahlreiche Diskussionen mit weiteren ausgewiesenen Experten trugen zu den Ergebnissen des Projektes bei. Zu nennen sind hier insbesondere Herr Dr. Heinrich (Bundesamt für Seeschifffahrt und Hydrographie), Herr Dr. Duin (Den Haag), Herr Dr. Fryer (Aberdeen), Herr Dr. Nicholson (London) sowie Herr Dr. Bjerkeng (Oslo).