

# PIK Report

---

No. 75

ANGEWANDTE STATISTIK

PIK-WEITERBILDUNGSSEMINAR  
2000/2001

Friedrich-Wilhelm Gerstengarbe  
(Hrsg.)



---

POTSDAM INSTITUTE  
FOR  
CLIMATE IMPACT RESEARCH (PIK)

---

Autor:  
Dr. F.-W. Gerstengarbe (Hrsg.)  
Potsdam-Institut für Klimafolgenforschung e.V.  
PF 60 12 03, D-14412 Potsdam  
Tel.: +49-331-288-2586  
Fax: +49-331-288-2695  
E-mail: gerstengarbe@pik-potsdam.de

Herausgeber:  
Dr. F.-W. Gerstengarbe

Technische Ausführung:  
U. Werner

---

POTSDAM-INSTITUT  
FÜR KLIMAFOLGENFORSCHUNG  
Telegrafenberg  
Postfach 60 12 03, 14412 Potsdam  
GERMANY  
Tel.: +49 (331) 288-2500  
Fax: +49 (331) 288-2600  
E-mail-Adresse:pik@pik-potsdam.de

---

POTSDAM, MÄRZ 2002

## **Abstract**

Training courses are sporadically given at PIK to enable all scientists to become acquainted with present problems from other scientific disciplines and approaches on how to solve them. A training course on mathematical statistics was given in the winter term 2000-2001 within this framework. The lectures focussed on the statistical analysis of time series. They dealt with general problems on how to apply statistical methods, data analysis, methods of statistical analysis, persistence, the robustness of statistical methods, the use of multivariate methods, and the verification of model results. In addition, an overview of the uncertainties of climate change scenarios was given in a special lecture.

Except for the lecture on 'data analysis' and the 'use of multivariate methods' (the latter has already been published in: v. Storch, H., Zwiers, F.W. (1999): *Statistical Analysis in Climate Research*", Cambridge University Press, ISBN 0 521 45071 3, 494 pages), the present PIK report includes all lectures of this training course.

We would like to take the opportunity to thank all lecturers here, especially for providing their manuscripts.

## **Zusammenfassung**

Am PIK werden in loser Reihenfolge Weiterbildungsveranstaltungen durchgeführt, um allen Wissenschaftlern die Möglichkeit zu geben, aktuelle Probleme und deren Lösungsansätze aus anderen Fachbereichen kennenzulernen. In diesem Rahmen wurde im "Wintersemester" 2000-2001 eine Weiterbildung in mathematischer Statistik durchgeführt. Der Schwerpunkt lag bei den Vorträgen auf der statistischen Analyse von Zeitreihen. Dabei wurden Themen behandelt, die von allgemeinen Problemen in der Anwendung statistischer Verfahren, über die Datenanalyse, die Methoden der statistischen Analyse, die Persistenz, die Robustheit statistischer Verfahren und die Nutzung multivariater Verfahren bis hin zur Verifikation von Modellergebnissen reichten. In einem Sondervortrag wurde außerdem eine Übersicht über die Unsicherheit von Klimaänderungsszenarien gegeben.

Bis auf den Beitrag zur Datenanalyse und den zum "Einsatz multivariater Verfahren" (Letzterer ist bereits veröffentlicht in: v. Storch, H., Zwiers, F.W. (1999): *Statistical Analysis in Climate Research*", Cambridge university Press, ISBN 0 521 45071 3, 494 Seiten) enthält der vorliegende PIK Report die Vorträge der an dieser Veranstaltung beteiligten Referenten.

Allen Beteiligten sei an dieser Stelle nochmals gedankt, insbesondere für die Bereitstellung der Manuskripte.



## Inhaltsverzeichnis:

<b>Abstract</b> .....	<b>3</b>
<b>Zusammenfassung</b> .....	<b>3</b>
<b>1 Probleme in der Anwendung Statistischer Verfahren</b>	
<i>Friedrich-Wilhelm Gerstengarbe</i> .....	9
1.1 Einleitende Bemerkungen .....	9
1.2 Das Problem der Stationarität .....	9
1.3 Das Stichprobenproblem .....	10
1.4 Das Problem der Wahl der Verteilungsfunktion .....	12
1.5 Das Problem des Stichprobenumfangs .....	15
1.6 Das Problem des Stichprobenfehlers .....	16
1.7 Das Problem des zugrundeliegenden Prozesses .....	20
1.8 Abschließende Bemerkungen .....	21
1.9 Literatur .....	21
<b>2 Statistische Analyse von Zusammenhängen</b>	
<b>Methodik, Probleme, und einige Beispiele aus dem Bereich der Klimatologie</b>	
<i>Christian-D. Schönwiese</i> .....	23
2.1 Einführung und Motivation .....	23
2.2 Zweidimensionale Korrelationsrechnung .....	24
2.3 Zweidimensionale Regressionsrechnung .....	33
2.4 Mehrdimensionale Korrelations- und Regressionsrechnung .....	36
2.5 Ausblick und Wertung .....	38
2.6 Literatur .....	40
<b>3 Zur Berücksichtigung der Persistenz in meteorologischen Zeitreihen</b>	
<i>Peter C. Werner</i> .....	43
3.1 Einleitung .....	43
3.2 Methoden zur Schätzung der Persistenz .....	44
3.3 Berücksichtigung der Persistenz in statistischen Tests .....	49
3.4 Zeitliche Änderungen der Persistenz .....	50
3.5 Zusammenfassung .....	53
3.6 Literatur .....	53

<b>4</b>	<b>Robuste statistische Verfahren zur Auswertung mangelbehafteter Daten</b>	
	<i>Jens Taubenheim</i> .....	55
4.1	Mangelbehaftete Daten .....	55
4.2	Robuste statistische Auswertungsverfahren .....	56
4.3	Robuste Verfahren zur Korrelations- und Regressionsanalyse .....	57
4.4	Neue Entwicklungen durch das Paradigma "Datentiefe" .....	60
4.5	Die Regressionstiefe nach ROUSSEEUW und HUBERT (1999) .....	61
4.6	Die CAT-Line .....	62
4.7	Literatur .....	64
<b>5</b>	<b>Validation von Modellen</b>	
	<i>Andreas Hense</i> .....	67
5.1	Einleitung .....	67
5.2	Theorie .....	67
	5.2.1 Univariate reelle Zufallsvariable .....	68
	5.2.2 Multivariate reelle ZVA's .....	68
	5.2.3 Modell-Beobachtungsvergleich .....	69
	5.2.4 Schatzen .....	70
	5.2.5 Darstellung von Ergebnissen .....	70
	5.2.6 Modell-Daten Vergleiche .....	72
	5.2.7 Felder und korrelierte Zeitreihen .....	75
5.3	Validation of model predictions and analysis of measurements - a unified Bayesian view .....	75
	5.3.1 Introduction .....	75
	5.3.2 Univariate example .....	77
	5.3.3 Multivariate method .....	79
	5.3.4 Numerical analysis of observations .....	82
5.4	References .....	83
<b>6</b>	<b>How uncertain are regional climate change scenarios? Examples for Europe and the Alps</b>	
	<i>Dimitrios Gyalistras</i> .....	85
6.1	Introduction .....	85
6.2	Scenarios of atmospheric circulation changes for the North Atlantic and Europe .....	85
6.3	Scenarios of temperature and precipitation changes for the European Alps ..	87
6.4	Which scenarios are most trustworthy? .....	89
6.5	Explosion of uncertainty in impact studies? .....	90

6.6 Should the main goal be to reduce the uncertainties? . . . . .	91
6.7 Conclusions. . . . .	91
6.8 References . . . . .	92
<b>Liste der Autoren . . . . .</b>	<b>95</b>



# 1 Probleme in der Anwendung Statistischer Verfahren

*Friedrich-Wilhelm Gerstengarbe*

*Zum unverzichtbaren Rüstzeug jedes Fälschers gehört eine vertiefte Kenntnis der statistischen Tricks, die einen in die Lage versetzen, nach Belieben über Berechnungen mit dem richtigen Resultat zu verfügen um mit mathematischer Strenge jedwede Ausgeburt der Phantasie belegen zu können.*

*FEDERICO DI TROCCHIO, "Der große Schwindel"*

## 1.1 Einleitende Bemerkungen

Es ist bekannt, dass es in der Entwicklung des Klimas beachtliche natürliche Schwankungen gibt, die schon in ihrer Abfolge die Menschheit vor schwer zu lösende Probleme stellen können. Die diese Schwankungen überlagernden anthropogen verursachten Klimaänderungen werden die Probleme mit großer Wahrscheinlichkeit noch verstärken. Es gibt also eine Fülle von Problemen die zu lösen sind, um die zu erwartenden negativen Auswirkungen zu erfassen, die Wahrscheinlichkeit ihres Eintretens zu bestimmen und entsprechende Handlungsstrategien zu deren Abschwächung zu entwickeln. Da diese Probleme in der Regel komplexer Natur sind werden auch komplexe Ansätze zu deren Lösung verlangt. Dies bedeutet generell eine Abkehr von der bis in die heutige Zeit hinein praktizierten linearen Betrachtungsweise einzelner Größen und den Übergang zur Untersuchung und Beschreibung der Wechselwirkung möglichst vieler Einflußgrößen. Damit werden hohe Anforderungen an die Qualität und den richtigen Einsatz der zur Verfügung stehenden Methoden gestellt. Was in diesem Zusammenhang im Rahmen statistischer Untersuchungen zu beachten ist, soll im folgenden behandelt werden.

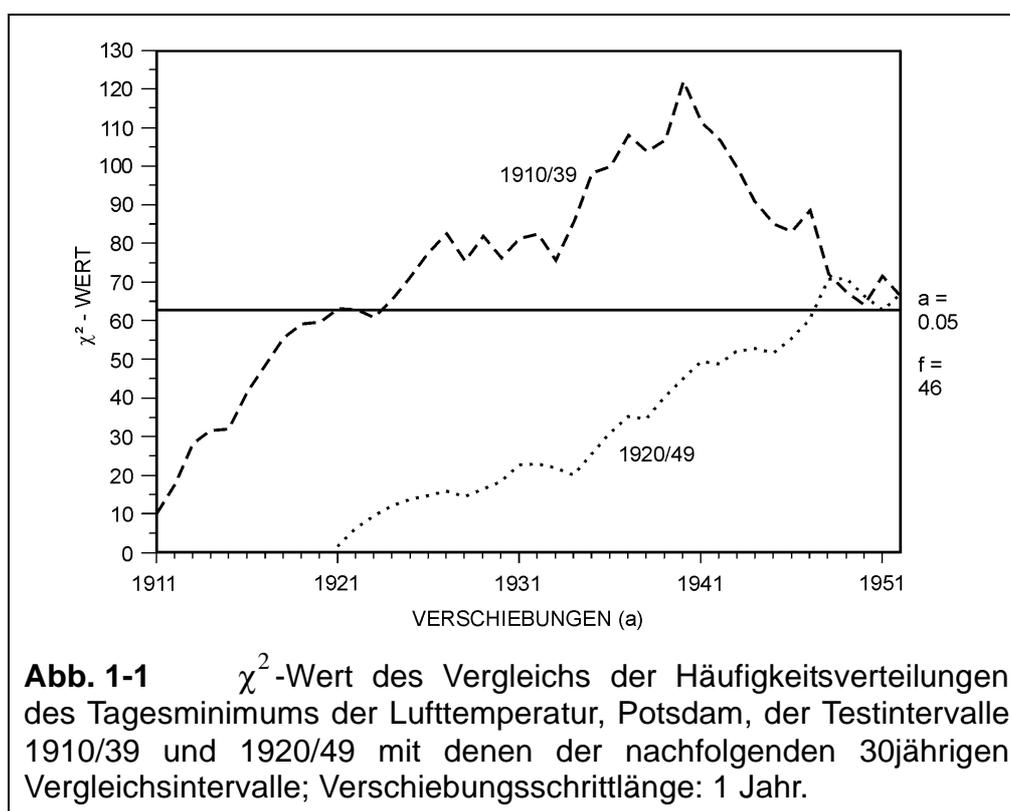
## 1.2 Das Problem der Stationarität

*Regel: Die gesamte zur Zeit vorhandene mathematische Statistik beruht auf der Grundlage stationärer stochastischer Prozesse.*

Das heißt, dass man zwar mit diesen statistischen Verfahren Instationaritäten aufdecken kann, aber nicht in der Lage ist, instationäre Prozesse zu beschreiben. Sind Prozesse instationär, wie es zum Beispiel bei Klimaänderungen der Fall ist, muß diese Problematik bei der Anwendung statistischer Verfahren beachtet werden. Ein Beispiel soll dies demonstrieren:

Die Tagesminima der Lufttemperatur Potsdams von 1910 bis 1981 werden zu 30jährigen Häufigkeitsverteilungen, die schrittweise um jeweils 1 Jahr verschoben werden, zusammengefaßt (also 1910/39, 1911/40, ... 1952/81). Mit Hilfe des  $\chi^2$ -Tests wird geprüft, ob die Stichprobe des Zeitraums 1910/39 mit den jeweils nachfolgenden Zeiträumen identisch ist, also beide aus ein und derselben Grundgesamtheit stammen. Die gleiche Untersuchung wird auch für den Zeitraum 1920/49 im Vergleich mit den nachfolgenden Zeiträumen durchgeführt. Die Ergebnisse sind in Abb. 1-1 dargestellt. Für den Bezugszeitraum 1910/29 (gestrichelte Kurve) gilt, dass ab 1924/53 die Nullhypothese (beide Verteilungen stammen aus ein und derselben Grundgesamt-

heit) verworfen werden muß, da ab diesem Zeitpunkt das Signifikanzniveau für  $\beta = 0.05$  überschritten wird. Ganz anders verhält sich die Stichprobe für den Zeitraum 1920/49 (gepunktete Kurve), für die erst ab der Stichprobe 1948/77 die Nullhypothese verworfen werden muß. Ursache für das frühzeitige Überschreiten des Signifikanzniveaus im 1. Vergleich ist die Instationarität der Reihe, die mit einem Trend behaftet ist. Der 2. Vergleich zeigt aber, dass es innerhalb dieser instationären Reihe Perioden gibt, die man als quasi-stationär bezeichnet und auf die die Palette der statistischen Verfahren problemlos angewandt werden kann.



### 1.3 Das Stichprobenproblem

*Regel: Jede Stichprobe muß aus unabhängig und identisch verteilten Werten bestehen.*

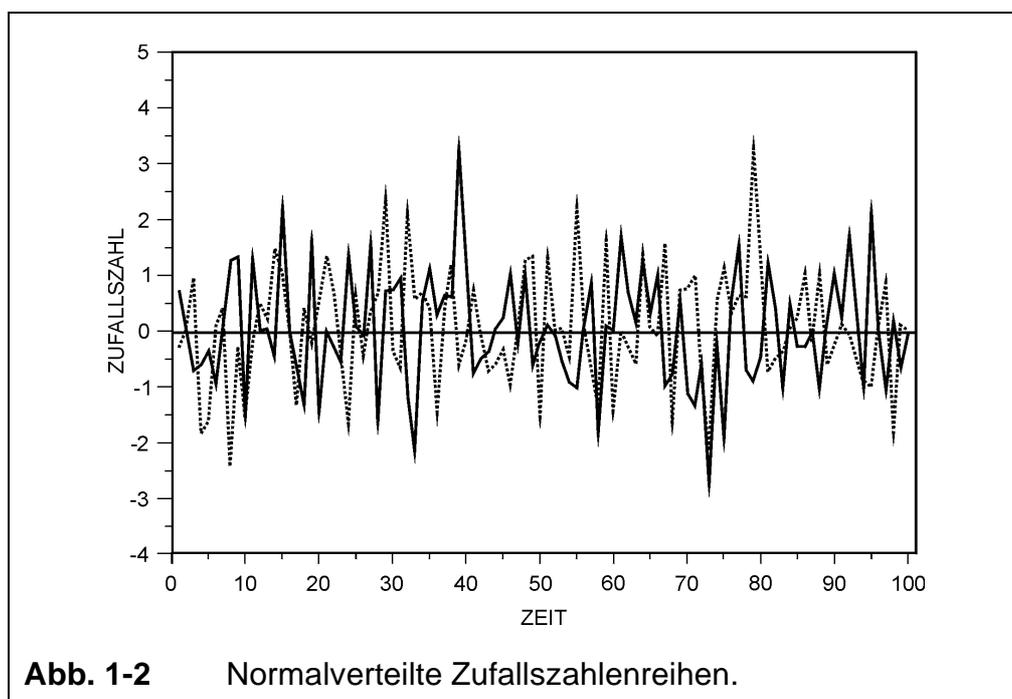
Die Unabhängigkeit der Werte ist dann gegeben, wenn eine Erhaltungsneigung statistisch nicht nachweisbar ist (d.h., die Autokorrelation geht gegen 0!) bzw. wenn man bei einer vorhandenen Erhaltungsneigung von der tatsächlichen Werteanzahl  $n$  der Stichprobe zur effektiven Werteanzahl  $n_e$  übergeht. In diesem Fall wird automatisch die Anzahl der Freiheitsgrade vermindert (BARTHELIS, 1935; OLBERG, 1972; TAUBENHEIM, 1969, 1974).

Unter identisch verteilten Zufallsgrößen versteht man eine Folge, für deren Verteilungsfunktion  $F$  die Beziehung gilt:

$$F(x_1) = F(x_n) \quad (n = 2, 3, \dots) \quad (1-1)$$

Das heißt, dass jeder mögliche, mit einer bestimmten Wahrscheinlichkeit auftretende Wert dem zugrundeliegenden Prozeß nicht widersprechen darf, also der jeweilige stochastische Prozeß bekannt sein muß, um das richtige statistische Modell zur Beschreibung dieses Prozesses wählen zu können. Instationäre Prozesse sind ein typisches Beispiel für nichtidentische Stichprobenverteilungen.

Erzeugt man mittels eines Zufallszahlengenerators zwei beliebige normalverteilte Zufallszahlenreihen (s. Abb. 1-2) so sind diese unabhängig und identisch verteilt.



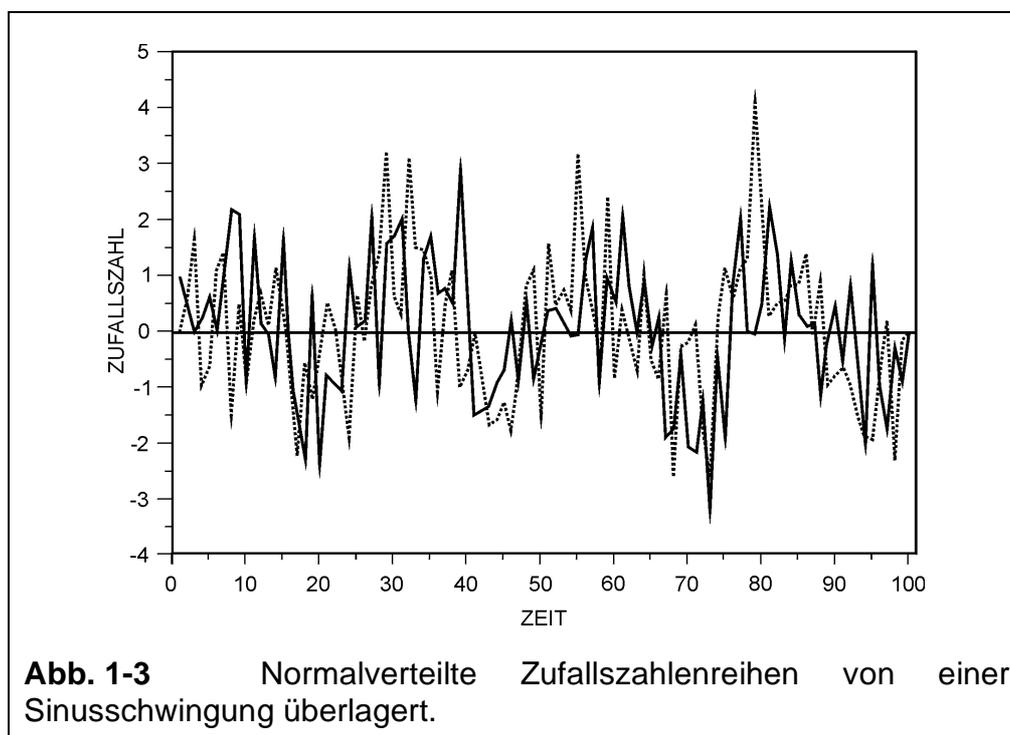
Dies kann man prüfen, indem man den Korrelationskoeffizienten zwischen den beiden Reihen berechnet und dessen Signifikanz prüft. In Tab. 1-1 sind die Ergebnisse dafür angegeben.

	ZZR	ZZRS
Werteanzahl $n$	100	100
Effektive Werteanzahl $n_{eff}$	95	55
Korrelationskoeffizient $r$	0.105	0.265
Signifikanzschränke $r_n$	0.205	0.205
Signifikanzschränke $r_{n_{eff}}$	0.211	0.2689

**Tab. 1-1** Vergleich zweier Zufallszahlenreihen, a) ohne aufgeprägte Schwingung (ZZR), b) mit aufgeprägter Schwingung (ZZRS).

Der Korrelationskoeffizient für ZZR beträgt 0.105. Er liegt sowohl unter der Signifikanzschränke für  $r_n$  als auch  $r_{n_{eff}}$ , das heißt, dass es keine statistisch gesicherte von

Null verschiedene Korrelation zwischen den beiden Reihen gibt. Dies war auch zu erwarten. Prägt man beiden Zufallszahlenreihen eine einfache Sinusschwingung auf (s. Abb. 1-3), sieht das Ergebnis anders aus.



Für die tatsächliche Werteanzahl  $n$  überschreitet jetzt der Korrelationskoeffizient mit  $r = 0.265$  die Signifikanzschränke  $r_n = 0.205$ . Das heißt, dass die Nullhypothese, beide Zufallszahlenreihen stammen aus ein und derselben Grundgesamtheit nicht verworfen werden kann! Reduziert man die tatsächliche Werteanzahl  $n$  auf die effektive  $n_{eff}$  liegt der Korrelationskoeffizient wieder unterhalb der Signifikanzschränke, das heißt, dass es keine statistisch gesicherte von Null verschiedene Korrelation zwischen den beiden Reihen gibt. Damit ist nachgewiesen, dass es unbedingt erforderlich ist bei jeder statistischen Analyse von Zeitreihen die Erhaltungsneigung oder auch Autokorrelation zu berücksichtigen.

#### 1.4 Das Problem der Wahl der Verteilungsfunktion

*Regel: Eine theoretische Verteilungsfunktion, deren Parameter aus einer empirischen Verteilung geschätzt werden, ist über den Bereich der aus der empirischen Verteilung vorhandenen Stützstellen hinaus nicht gültig (MÜLLER, 1988).*

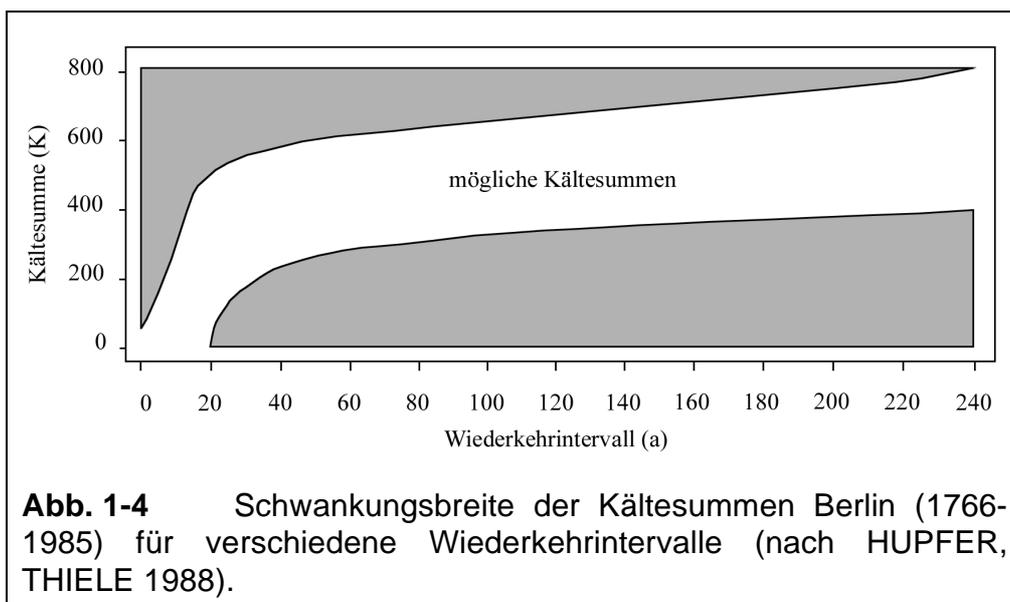
Diese Regel gilt grundsätzlich, solange keine Modellvorstellung über den betrachteten Prozeß vorliegt. Liegt eine Modellvorstellung vor, läßt ein Anpassungstest die Weiterarbeit mit einer theoretischen Verteilungsfunktion zu. In diesem Fall ist es möglich, Konfidenzintervalle für diese theoretische Verteilungsfunktion aufzustellen, die auch über den Stützstellenbereich hinaus Gültigkeit haben. Allerdings würde sich das Konfidenzband in den Extrapolationsbereichen derart weitens, dass kaum noch praktische Schlüsse zu erwarten sind (MÜLLER, 1988).

Am Beispiel der bekannten FISHER-TIPPETT-Verteilungen, dies sind die derzeit gängigen Extremwertverteilungen, soll demonstriert werden, welche Probleme auftreten, wenn die Regel nicht beachtet wird. Die drei Verteilungstypen sind wie folgt definiert:

$$\begin{aligned}
 \text{Typ I} \quad \Lambda(x) &= \exp(-e^{-x}) & -\infty < x < \infty \\
 \text{Typ II} \quad \phi_{\alpha}(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} & (\alpha > 0) \\
 \text{Typ III} \quad \Psi_{\alpha}(x) &= \begin{cases} \exp(-(-x)^{\alpha}) & x \leq 0 \\ 1 & x > 0 \end{cases} & (\alpha > 0)
 \end{aligned} \tag{1-2}$$

Der Typ I ist auch als Gumbel-Verteilung bekannt, Typ II ist die konkave, Typ III die konvexe Variante.

In einer Arbeit von THIELE (1987) werden die Kältesummen von Berlin (1766-1982) einer statistischen Extremwertanalyse unterzogen (Kältesumme = Absolutwert der Summe aller Temperaturen  $< 0^{\circ}\text{C}$  über einen definierten Zeitraum). Dabei wurden mit Hilfe der Gumbel-Verteilung Wiederkehrintervalle von 1-240 Jahren berechnet, das heißt also, welcher Wert kommt im Mittel jedes Jahr einmal vor, oder nur einmal in zwei Jahren usw. bis hin zum Auftreten eines Wertes nur einmal innerhalb von 240 Jahren. Die dabei auftretenden Schwankungsbreiten, die auf der Basis verschiedener Bezugsintervalle erzeugt wurden, sind in Abb. 1-4 dargestellt. Betrachtet man zum Beispiel in der Abbildung den Kältesummenwert 400 K, so kann man feststellen, dass die Wahrscheinlichkeit des Auftretens dieses Wertes für 20- bis über 200jährige Wiederkehrintervalle gleich groß ist. Dies ist eine für praktische Belange sinnlose Aussage, deren Ursache in der schon erwähnten Instationarität des betrachteten Prozesses liegt.



Weitere Fehler können bei der Berechnung von Wiederkehrintervallen auftreten, wenn das Verteilungsmodell der Grundgesamtheit nicht oder nur ungenügend

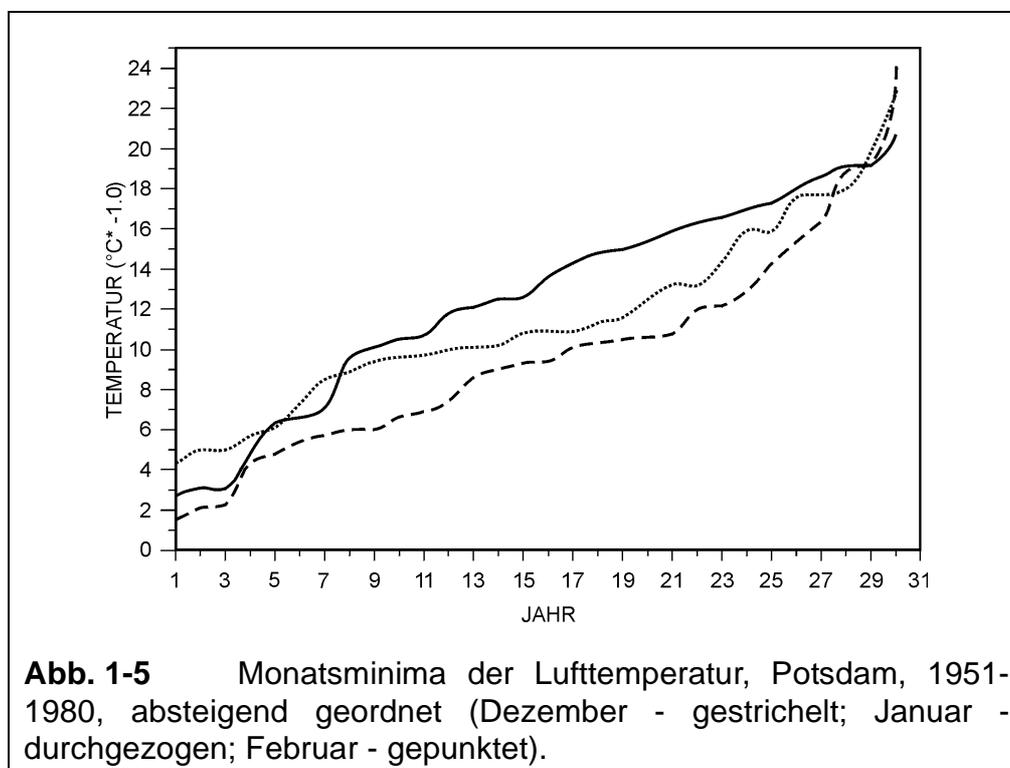
bekannt ist. So spielt zum Beispiel die Relation zwischen "mittleren" extremen Werten und den Ausreißern für die Anpassung der theoretischen Verteilung an die Stichprobenverteilung eine wichtige Rolle:

In der Tabelle (Tab. 1-2) sind für verschiedene Eintrittswahrscheinlichkeiten (Wiederkehrintervalle) die zu erwartenden Monatsminima der Lufttemperatur (berechnet nach FISHER-TIPPETT, 1928) für die Station Potsdam (1951-1980) angegeben.

Monat	Wiederkehrintervall (a)							
	10	15	20	25	30	40	50	100
Dezember	-17.8	-19.4	-20.4	-21.2	-21.8	-22.7	-23.4	-25.5
Januar	-19.3	-19.5	-19.7	-19.8	-19.9	-20.0	-20.1	-20.3
Februar	-18.2	-19.5	-20.3	-20.9	-21.5	-22.3	-22.9	-24.8

**Tab. 1-2** Wiederkehrintervalle für die Monatsminima der Lufttemperatur berechnet aus der FISHER-TIPPETT-Verteilung Typ I, Potsdam 1951-1980.

Klimatologisch betrachtet, ist der Dezember als der wärmste, der Januar als der kälteste Wintermonat anzusehen. Dazu stehen die Temperaturwerte mit einer Eintrittswahrscheinlichkeit von 1% (100jähriges Wiederkehrintervall) im Widerspruch. Selbst wenn man davon ausginge, die drei Wintermonate als klimatologisch einheitlich zu betrachten, wäre der Temperaturunterschied von mehr als 5 K zwischen Dezember und Januar falsch.



**Abb. 1-5** Monatsminima der Lufttemperatur, Potsdam, 1951-1980, absteigend geordnet (Dezember - gestrichelt; Januar - durchgezogen; Februar - gepunktet).

Ursache dafür ist zum einen die konkave Anpassung der theoretischen Verteilungsfunktion an die realen Werte des Monats Dezember mit einem zu steilen Anstieg im oberen Kurvenbereich und zum anderen die konvexe Anpassung für den Januar mit einem zu flachen Anstieg im oberen Kurvenbereich (s. Abb. 1-5).

## 1.5 Das Problem des Stichprobenumfangs

*Regel: Praktisch verwertbare Aussagen sind nur dann zu erwarten, wenn erhebliche Stichprobenumfänge vorliegen.*

Was ist unter einem erheblichen Stichprobenumfang zu verstehen? An zwei Beispielen soll diese Problematik erläutert werden.

Wie bekannt, ist die Breite jedes Konfidenzintervalls dem Stichprobenumfang direkt proportional. In der folgenden Tabelle (Tab. 1-3) sind die Konfidenzgrenzen nach dem KOLMOGOROV-Test für  $\beta = 5\%$  und verschiedene  $n$  angegeben.

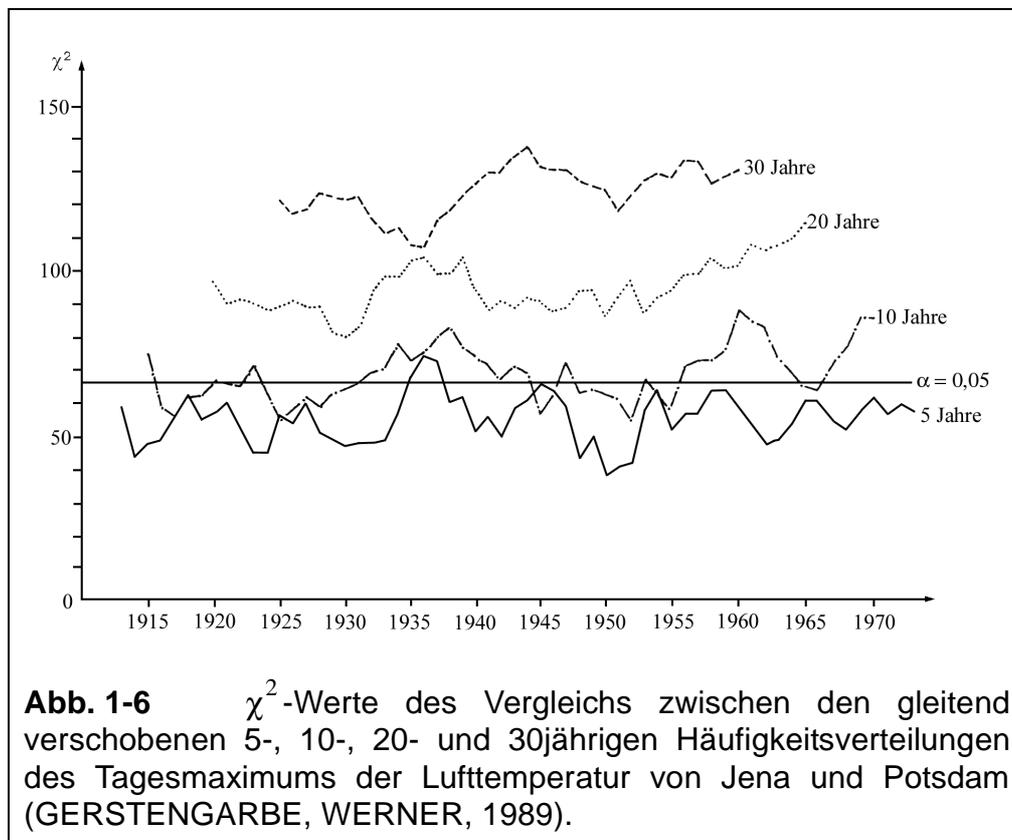
$N$	10	20	30	50	100	500	1000	30000
Konfidenzgrenze (%)	41.3	29.5	24.2	18.9	13.4	6.0	4.3	0.8

**Tab. 1-3** Konfidenzgrenzen für den KOLMOGOROV-Test.

Aus der Tabelle (Tab. 1-3) ist sofort ersichtlich, dass die Konfidenzgrenzen für kleine  $n$  derart breit sind, dass eine praktische Verwertbarkeit der aus solchen Stichprobenumfängen gewonnenen Ergebnisse im Prinzip nicht möglich ist. Zu beachten ist auch, dass bei großen  $n$  trotz relativ kleiner Konfidenzgrenzen die Variationsbreite möglicher Ereignisse ansteigt. So liegt zum Beispiel für  $n = 30000$  und einer Konfidenzgrenze von  $\pm 0.8\%$  die Variationsbreite für ein Ereignis  $x$  mit der Häufigkeit  $h$  zwischen  $h - 240 < h < h + 240$ . Für  $h = 240$  ist die Untergrenze = 0! Bei der Betrachtung zum Beispiel von Extremen sind Werteanzahlen von 240 schon als sehr groß anzusehen. Nimmt man an, es gäbe innerhalb einer Stichprobe von 30000 Werten 240 Extremwerte, so lautet die statistische Aussage, dass ein solcher Extremwert bei gleicher statistischer Sicherheit keinmal oder auch 240 mal auftreten kann. Dies ist für die Praxis eine irrelevante Aussage.

Im Rahmen einer Untersuchung (GERSTENGARBE, WERNER, 1989) wurde überprüft, ob sich die Häufigkeiten des Tagesmaximums der Lufttemperatur von Jena und Potsdam im Zeitverlauf unterscheiden. Dazu wurden 5-, 10-, 20-, und 30-jährige Häufigkeitsverteilungen gebildet, die jeweils gleitend entlang der Zeitachse verschoben wurden (Verschiebungsintervall 1 Jahr). Diese Häufigkeitsverteilungen der Daten von Jena und Potsdam wurden für jedes Intervall mittels des  $\chi^2$ -Tests miteinander verglichen. Die zeitliche Abfolge der  $\chi^2$ -Werte ist in Abb. 1-6 dargestellt (Zeitraum 1911/75). Es ist sofort zu erkennen, dass die Größe der  $\chi^2$ -Werte von der Werteanzahl in den einzelnen Häufigkeitsverteilungen abhängt, sieht man einmal von den zeitlichen Variationen ab. Damit ist eine Interpretation im Sinne einer Annahme der Nullhypothese - die Verteilungen von Jena und Potsdam stammen aus ein und derselben Grundgesamtheit - (wenigstens für Teilabschnitte) als falsch anzusehen. Abschlie-

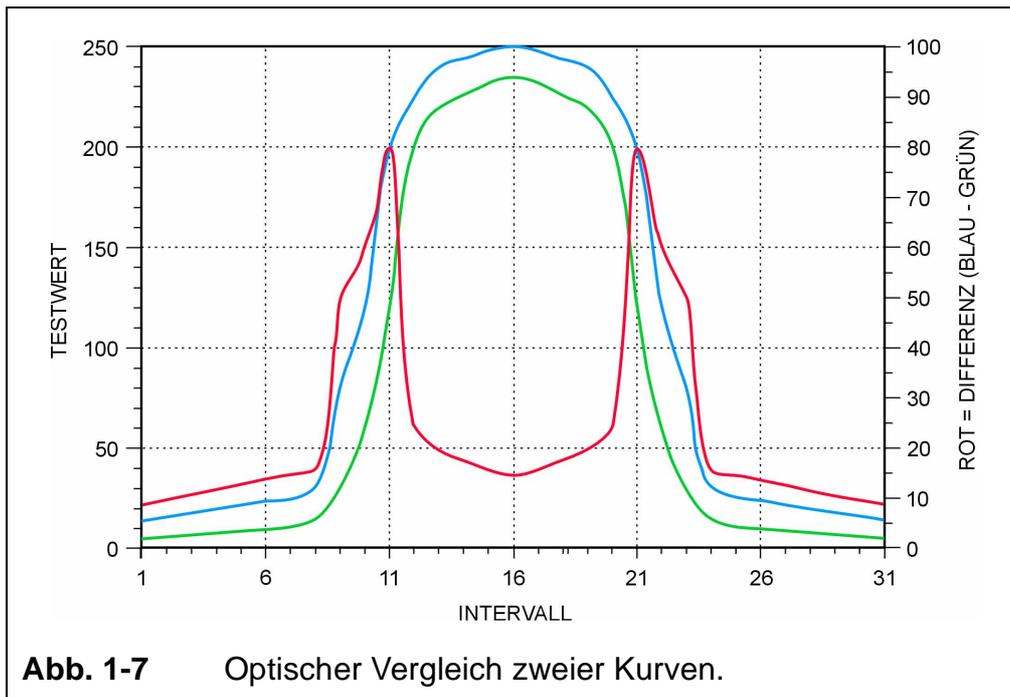
ßend sei noch bemerkt, dass die gleichen Probleme bei der oft praktizierten Testung der Güte der Anpassung einer theoretischen an eine empirische Verteilung auftreten, wo ja gerade die Nullhypothese bestätigt werden soll.



## 1.6 Das Problem des Stichprobenfehlers

*Regel: Jede statistische Aussage ist ohne Angabe ihres Fehlers unbrauchbar.*

Aus allen bisher angeführten Beispielen wird klar, dass eine Angabe des statistischen Fehlers in jedem Fall unbedingt erforderlich ist. Häufig versucht man diese Anforderung zu umgehen, indem ein einfacher graphischer Vergleich durchgeführt wird. Zum Beispiel werden zwei zu vergleichende Kurven (Zeitreihen, Werte der Wahrscheinlichkeitsdichtefunktion u.ä.) in ein gemeinsames Koordinatensystem eingetragen und optisch miteinander verglichen. Bei einer subjektiv guten Übereinstimmung wird dann geschlußfolgert, dass die Werte der beiden Stichproben aus ein und derselben Grundgesamtheit stammen. Ein einfaches Beispiel soll zeigen, wie unsinnig ein derartiges Vorgehen ist. In Abb. 1-7 ist der Verlauf zweier Kurven (grün und blau) dargestellt, die offensichtlich die gleiche Struktur aufweisen und hoch miteinander korreliert sind. Das diese beiden Kurven sich trotzdem deutlich unterscheiden, erkennt man erst, wenn man deren Differenzenreihe erstellt. Sie ist in Abb. 1-7 dargestellt. Aus ihr geht hervor, dass die Differenzen zum Teil in der Größenordnung der Originalwerte der beiden Kurven liegen. Damit verbietet sich eindeutig die aufgrund des optischen Eindrucks gemachte Feststellung, dass beide Kurven gut übereinstimmen. Hier ist eine entsprechende Fehlerbetrachtung notwendig.



Noch problematischer wird es, wenn man versucht, zweidimensionale Strukturen (z.B. Druck- oder Temperaturfelder) miteinander auf diese Art und Weise zu vergleichen. Ein Verfahren, das sich auch heute noch einer breiten Beliebtheit erfreut. Ein Beispiel soll die Problematik erläutern:

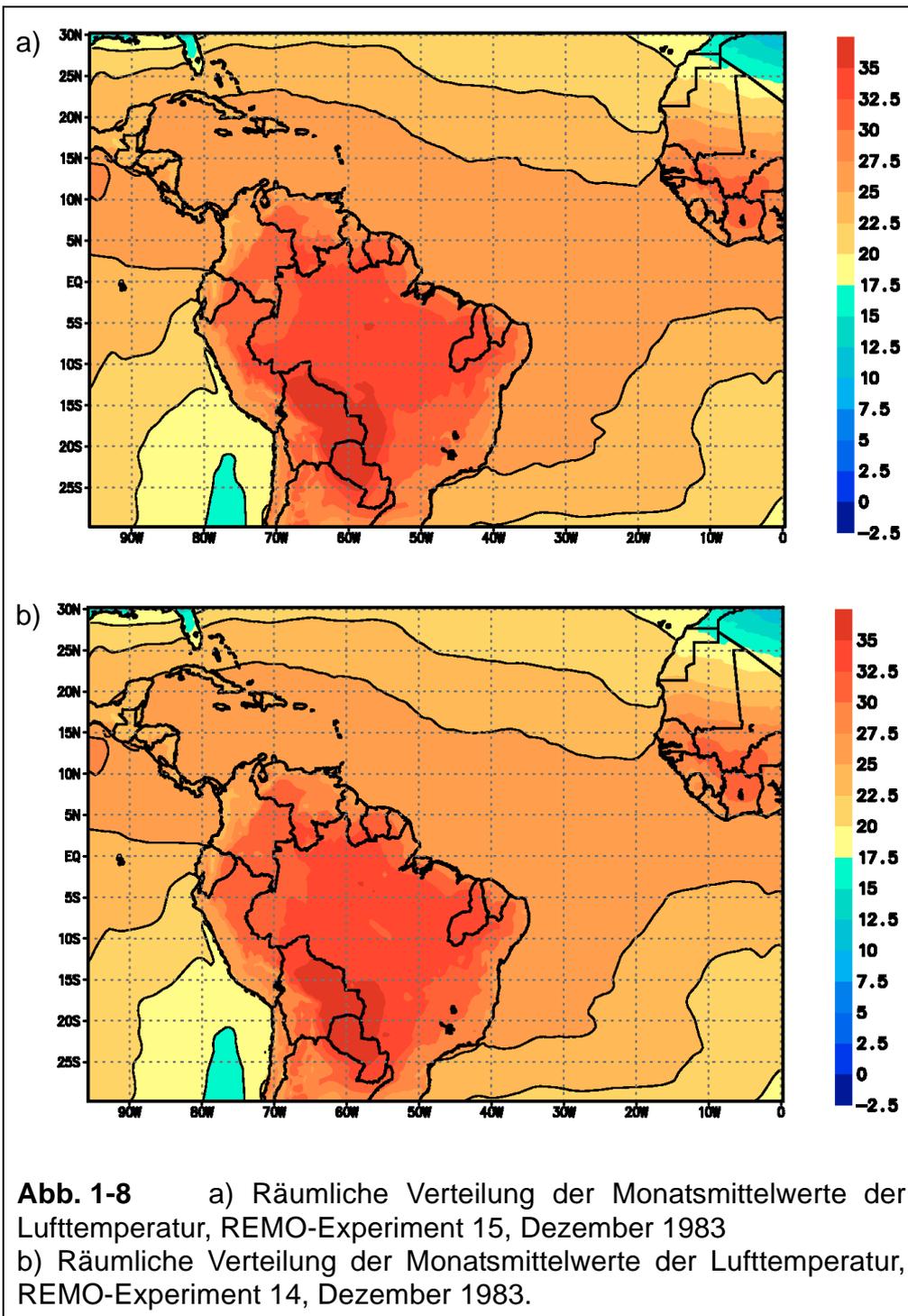
In Abb. 1-8a und Abb. 1-8b sind die Ergebnisse für zwei Experimente mit dem regionalen Klimamodell REMO dargestellt. Es handelt sich dabei um die räumliche Verteilung der Monatsmittelwerte der Lufttemperatur in Bodennähe für den Dezember 1983. Ein erster Blick über die räumliche Verteilung der Temperatur signalisiert eine gute Übereinstimmung zwischen beiden Experimenten.

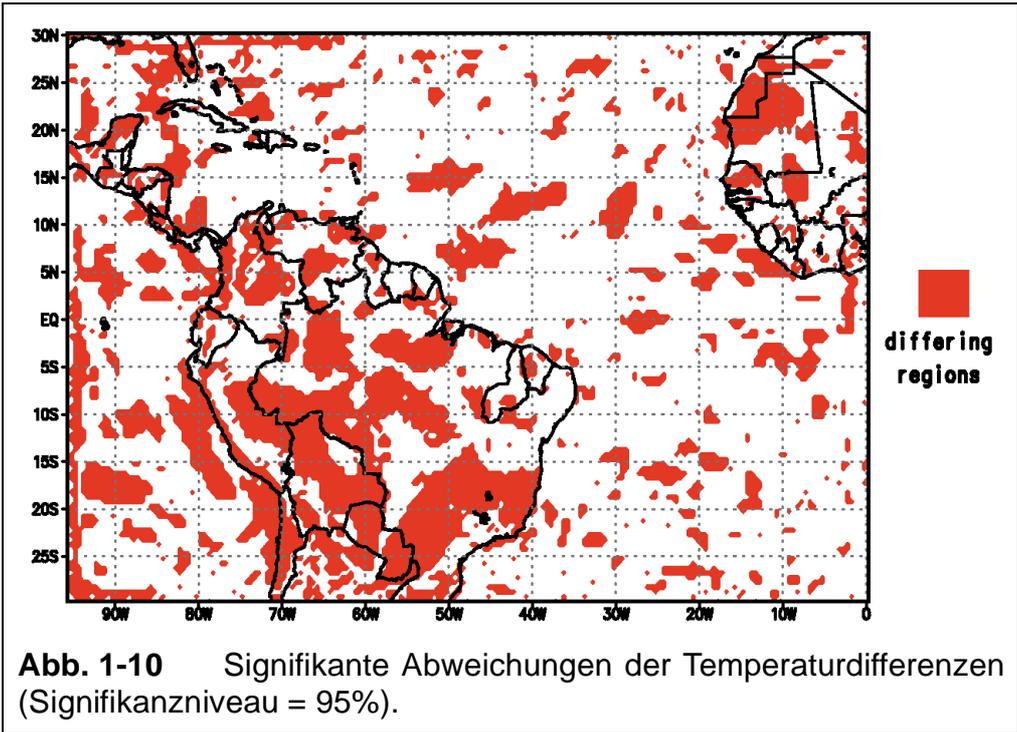
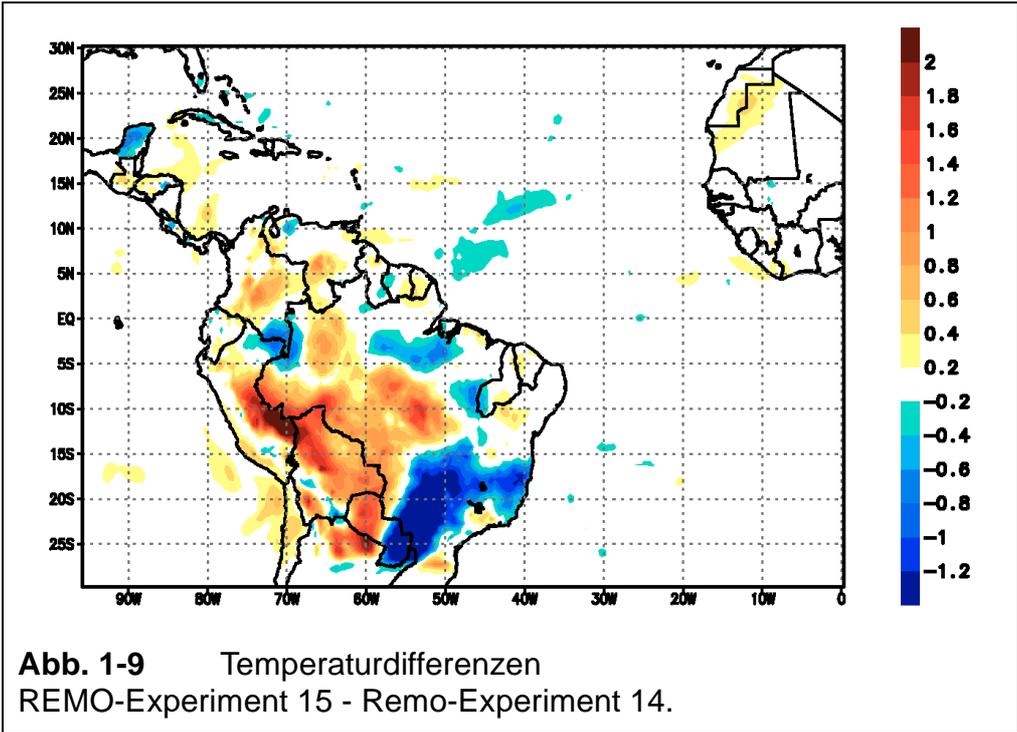
Auch der Vergleich der Gebietsmittelwerte sowohl für das Gesamtgebiet als auch die Unterteilung in Land und Seeregionen führt zu dem gleichen Ergebnis:

	Gesamt	Land	See
Experiment 14	25.6 °C	29.4 °C	23.9 °C
Experiment 15	25.5 °C	29.3 °C	23.9 °C

Dieses positive Bild verändert sich schlagartig, wenn man die Temperaturdifferenzen zwischen den beiden Experimenten an jedem Gitterpunkt bildet. Sie sind in Abb. 1-9 dargestellt. Jetzt reichen die Abweichungen von +2 °C bis zu -1.2 °C. dass diese Abweichungen zum großen Teil signifikant sind, zeigen die dazu berechneten Testergebnisse in Abb. 1-10.

Damit ist nachgewiesen, dass die Bewertung eines Ergebnisses ohne Fehlerangabe eine wissenschaftlich nicht korrekte Vorgehensweise darstellt.



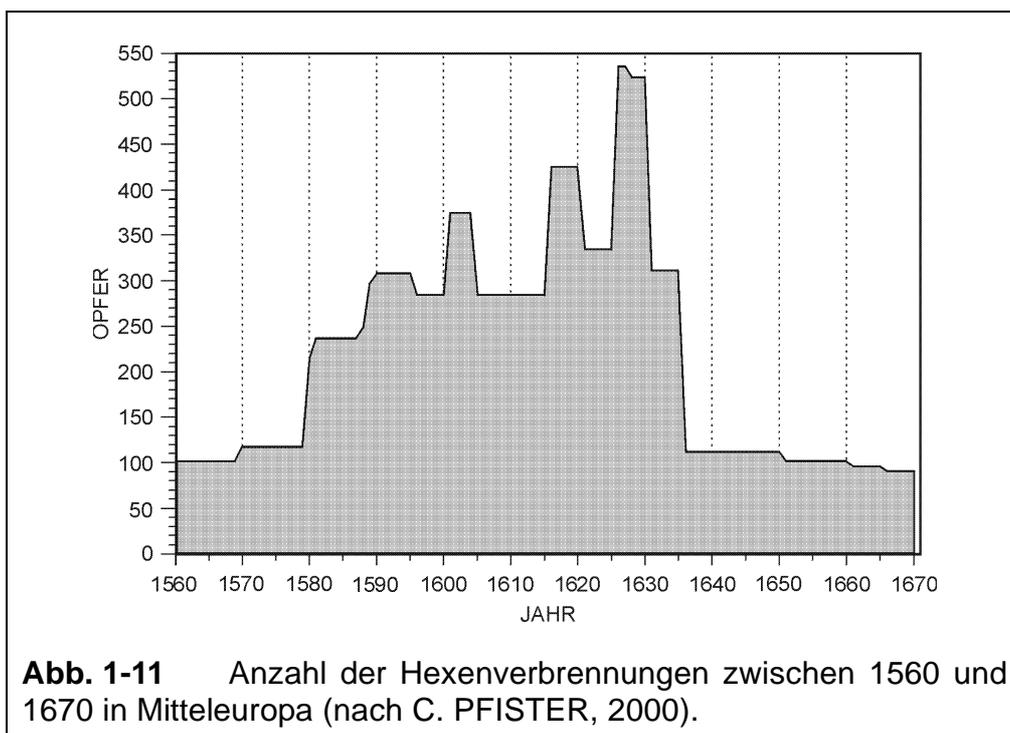


## 1.7 Das Problem des zugrundeliegenden Prozesses

*Regel: Ein statistisch gesicherter Zusammenhang darf nur dann interpretiert werden, wenn der zugrundeliegende Prozeß bekannt ist.*

Findet man einen statistisch gesicherten Zusammenhang zwischen zwei Vorgängen (z.B. zwischen zwei Zeitreihen) muß noch geklärt werden, ob diese Vorgänge ursächlich zusammenhängen, oder nur zufällig miteinander korreliert sind. An einem drastischen Beispiel soll gezeigt werden, wie die Schlußfolgerungen bei Unkenntnis des zugrundeliegenden Prozesses aussehen können:

Stellen wir uns vor, wir würden im Jahr 1670 leben und eine statistische Auswertung des in Abb. 1-11 dargestellten Geschehens vornehmen. Es handelt sich um die Anzahl der Hexenverbrennungen zwischen 1560 und 1670 in Mitteleuropa. Die Frage lautet: Gibt es einen ursächlichen Zusammenhang zwischen der Anzahl der Hexenverbrennungen und den in diesem Zeitraum aufgetretenen Klimaextremen? Aus den Klimaaufzeichnungen ist bekannt, dass sich zwischen 1580 und 1635 die klimatische Situation in Mitteleuropa deutlich verschlechtert hat mit eingelagerten Perioden eines gehäuften Auftretens von Extremereignissen (1590, 1600, 1615, 1625).



Das heißt, dass der Anstieg der Hexenverbrennungen in diesen Perioden tatsächlich stark gekoppelt ist mit der sich verschlechternden klimatischen Situation begleitet von Mißernten, Hungersnöten etc.. Mit dem Kenntnisstand von 1670 würden wir die Wirkungskette wie folgt aufbauen:

- Klimaextreme treten auf
- es kommt zu Mißernten und Hungersnöten
- die sind Werke des Bösen (Teufels)
- Vertreter des Bösen sind Hexen

- die Hexen sind zu vernichten
- das Unheil wird beseitigt.

Da nach einer gewissen Zeit verstärkter Hexenverbrennungen die Klimaextreme tatsächlich zurückgehen, kommt der Zeitgenosse zwangsläufig zu dem Schluß, dass die Maßnahme vermehrt Hexen zu verbrennen, tatsächlich zum Erfolg geführt hat.

Es ist leicht, von unserem heutigen Wissensstand aus, über eine für uns derartig abartige Logik zu lächeln. Nur, wir befinden uns in exakt der gleichen Situation, wenn wir zwei Prozesse miteinander vergleichen und keine oder auch nur ungenaue Kenntnisse über die tatsächlichen Wechselwirkungen zwischen den Prozessen haben!

## 1.8 Abschließende Bemerkungen

Mit den wenigen oben angeführten Beispielen, sollte gezeigt werden, mit welchen Problemen eine korrekte Anwendung statistischer Verfahren im allgemeinen verbunden ist. Die Probleme verschärfen sich noch, wenn multivariate statistische Verfahren zur Beschreibung komplexer Zusammenhänge zum Einsatz kommen. Um hier so wenig wie möglich Fehler zu machen ist es ein unbedingtes Muß eines jeden "Statistikers" jeden Schritt in der Kette der notwendigen Bearbeitungsschritte von der Datenaufbereitung und -prüfung über die Methodenauswahl bis hin zur Interpretation der Ergebnisse auf der Basis der vorhandenen Prozeßkenntnisse sorgfältig zu prüfen und ebenso abzuarbeiten.

## 1.9 Literatur

- Barthels, J. 1935: Zur Morphologie geophysikalischer Zeitfunktionen. *S.B. Preuß. Akad. Wiss.*, **30**, 504-522.
- Fisher, R. A., Tippett, L. H. C. 1928: Limiting forms of the frequency distribution of the largest or smallest members of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180-190.
- Gerstengarbe, F.-W., Werner, P. C. 1989: Die zeitliche Variabilität von Klimacharakteristiken in ihrer Bedeutung für den mesoskaligen Bereich. *Abh. Meteorol. Dienst der DDR*, **141**.
- Müller, P. H. 1888: Persönliche Mitteilung.
- Olberg, M. 1972: Bemerkungen zur Verwendung der effektiven Anzahl unabhängiger Stichprobenwerte bei statistischen Testverfahren in der Meteorologie und Geophysik. *Gerlands Beitr. Geophysik*, **81**, 57-64.
- Thiele, R. 1987: Zur Klassifikation und zeitlichen Variabilität der Winter im Berliner Raum. Diplomarbeit, Humboldt-Universität, Berlin.
- Taubenheim, J. 1969: Statistische Auswertung geophysikalischer und meteorologischer Daten. Akad. Verlagsgesellschaft Geest & Portig, Leipzig.
- Taubenheim, J. 1974: Zur Berücksichtigung der Autokorrelation bei der statistischen Signifikanzprüfung von Korrelationen zwischen zwei Zeitreihen. *Gerlands Beitr. Geophysik*, **83**, 413-416.

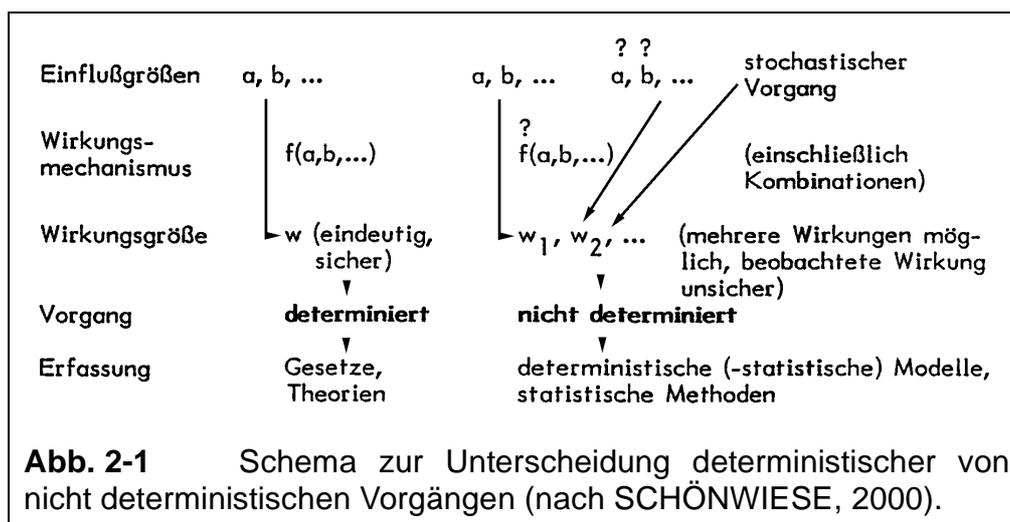


## 2 Statistische Analyse von Zusammenhängen Methodik, Probleme, und einige Beispiele aus dem Bereich der Klimatologie Christian-D. Schönwiese

### 2.1 Einführung und Motivation

„Statistik ist eine Zusammenfassung von Methoden, die uns erlauben, vernünftige optimale Entscheidungen im Fall von Ungewißheiten zu treffen“, so lautet eine der klassischen Definitionen (nach WALD, zitiert in SACHS, 1999). Die „Ungewißheit“, mit der die Wissenschaft quer durch alle Fachdisziplinen nur allzu oft konfrontiert ist, bereitet vielen Schwierigkeiten, insbesondere denen, die „exakte“ Wissenschaft betreiben wollen und alles andere als suspekt ansehen, was tatsächlich oder scheinbar davon abweicht. Daher rührt eine nicht selten anzutreffende Abneigung, meist verbunden mit einem gewissen Unverständnis, gegenüber der Statistik.

Es ist aber schlicht Realität, dass sich nicht alle Vorgänge exakt, d.h. vollständig und sicher und beliebig genau berechnen oder gar prognostizieren lassen wie eine Sonnenfinsternis, deren Eintreten ein Beispiel für einen solchen exakt berechenbaren Vorgang ist, der als *determiniert* bezeichnet wird. Die häufigste Situation, die im Gegensatz dazu einen *nicht determinierten Vorgang* charakterisiert, vgl. Abb. 2-1, ist, dass wir zwar wissen, welche Größen ihn prinzipiell beeinflussen sollten, uns aber die quantitativen Ursache-Wirkung-Zusammenhänge unbekannt sind. Wir beobachten bzw. messen dann typischerweise, trotz gleicher Ausgangssituation (gleicher Randbedingungen) unterschiedliche Wirkungen, genauer gesagt, die Wirkungsgröße/n tritt/treten in unterschiedlichen Zahlenwerten in Erscheinung, die in der Statistik als *Häufigkeitsverteilungen* festgehalten und Grundlage für weitere Analysen und Interpretationen sind.



Die Möglichkeit, dass auch die Einflußgrößen nicht bekannt sind oder dass es sich um einen (reinen) Zufallsprozeß handelt, wollen wir hier ausklammern und statt dessen unterschiedliche Ausgangssituationen (Randbedingungen) zulassen. Dann lautet die entsprechende statistische Frage: Welche *Zusammenhänge* bestehen zwischen den uns im Prinzip bekannten Einflußgrößen und den von uns beobachteten Wir-

kungsgrößen? Weiterhin nehmen wir den günstigen Fall an, dass alle diese Größen quantitativ beobachtbar/meßbar sind. Dann können wir das Problem *numerisch* behandeln, d.h. die Zahlenwerte dieser Größen statistisch analysieren und interpretieren.

Statistik bedeutet in diesem speziellen Fall, dass nicht determinierte Vorgänge, bei quantitativ bekannten Einfluß- und Wirkungsgrößen, hinsichtlich ihrer Zusammenhänge untersucht werden sollen. Die alternative Methodik dazu ist die determinierte Modellbildung, in der Hoffnung, dass der jeweilige nicht determinierte Vorgang einen determinierten Kern enthält, der sich in entsprechenden determinierten Gleichungssystemen exakt behandeln läßt. Aber selbst in diesem Fall bleibt ein nicht determinierter Rest übrig, möglicherweise ein ziemlich großer, und das gesamte Modell ist damit notwendigerweise unsicher. Ohne nun hier viele Beispiele anzuführen: Real gesehen führt an der Statistik sehr oft kein Weg vorbei, wobei übrigens auch hier von Modellbildung gesprochen wird, nämlich statistischer.

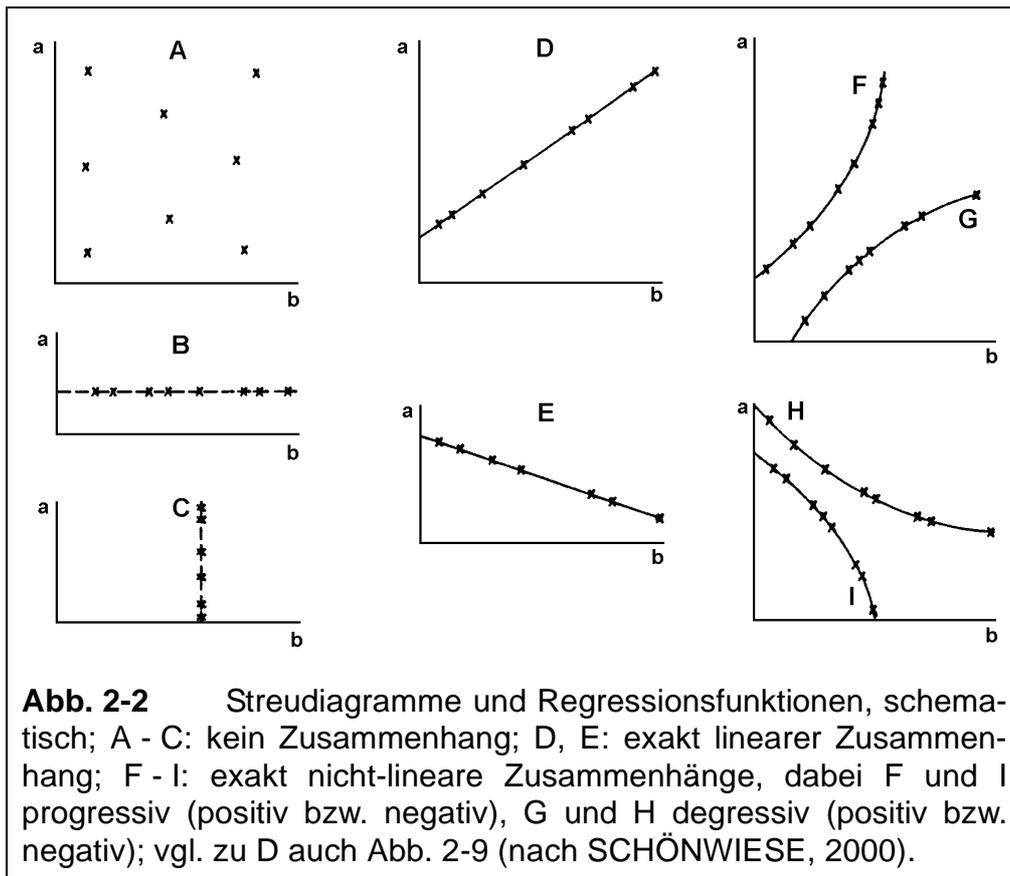
## 2.2 Zweidimensionale Korrelationsrechnung

Die einfachste Situation bei der Frage nach nicht determinierten und somit statischen Zusammenhängen ist die, dass nur eine Einfluß- und nur eine Wirkungsgröße existiert; dann stellen wir die Frage nach *zweidimensionalen Zusammenhängen*. Diese Frage ist in zwei Teilfragen aufzuschlüsseln:

- Wie gut bzw. schlecht ist der Zusammenhang?
- Welcher Art ist der Zusammenhang?

Die erste Frage beantwortet die *Korrelationsrechnung* in der Form von Korrelationskoeffizienten, welche die Güte des Zusammenhangs quantitativ abschätzen. Die zweite Frage, die stets nach der ersten angegangen werden sollte, beantwortet die *Regressionsrechnung*, welche die Art des Zusammenhangs in Form einer Gleichung angibt (wird im nächsten Abschnitt behandelt).

Häufig ist es sinnvoll, sich mittels eines *Streudiagramms* einen ersten Überblick über zweidimensionale Zusammenhänge zu verschaffen, wobei wir die beiden zu vergleichenden Größen mit  $a$  und  $b$  bezeichnen (und lassen dabei sogar offen, welche die Einfluß- und welche die Wirkungsgröße ist). Tragen wir die zuordenbaren Wertepaare  $a_i$  und  $b_i$ ,  $i = 1, 2, \dots, n$  ( $n =$  Stichprobenumfang) in ein solches Diagramm ein, vgl. Abb. 2-2, so gibt es zwei Extremfälle: Die Daten streuen völlig ungleichmäßig (Abb. 2-2, A), oder sie ordnen sich exakt entlang einer Geraden an (Abb. 2-2, D). Im ersten Fall muß vermutet werden, dass überhaupt kein Zusammenhang besteht, im zweiten Fall ist er exakt-linear (bei beliebiger, auch negativer Geradensteigung). Falls die jeweiligen Stichproben genügend groß sind, worauf unter dem Stichwort "Test" noch eingegangen wird, ist die weitere Betrachtung des Falls A i.a. sinnlos, im Fall D liegt möglicherweise ein determinierter Zusammenhang vor, der der weiteren statistischen Analyse nicht bedarf. In der Praxis wird aber irgendein Fall zwischen A und D gegeben sein, d.h. die Daten „streuen“ mehr oder weniger um eine Gerade (der allgemeinere Fall „Kurve“ wird anschließend behandelt).



Wir vermuten dann einen mehr oder weniger guten zweidimensionalen linearen Zusammenhang, dessen Güte durch den sog. *Produkt-Moment-Korrelationskoeffizienten* nach PEARSON geschätzt wird (SACHS, 1997a; SCHÖNWIESE, 2000):

$$r_{ab} = s_{ab} / (s_a s_b) \quad (2-1)$$

mit  $s_{ab} = 1/(n-1) \sum (a_i' b_i')$  = Kovarianz

und  $s_a^2 = 1/(n-1) \sum a_i'^2$  = Varianz der Daten  $a_i$ ,

$s_b^2$  entsprechend der Daten  $b_i$ ,

dies jeweils in der Quadratwurzel  $s_a$  bzw.  $s_b$  als Standardabweichung bezeichnet;  $a_i' = a_i - \bar{a}$  = Abweichungen vom arithmetischen Mittelwert  $\bar{a}$ ;  $b_i'$  entsprechend und  $n$  wiederum Stichprobenumfang (= Anzahl der Wertepaare  $\{a_i, b_i\}$ ; einfache Zahlenbeispiele dazu siehe SCHÖNWIESE, 2000). Auf der Grundlage dieser Definitionen kann sich  $r_{ab}$  zwischen den Grenzen

$$-1 \leq r_{ab} \leq +1 \quad (2-2)$$

bewegen (wobei exakt -1 bzw. +1 einen deterministischen Zusammenhang und exakt 0 gar keinen Zusammenhang bedeutet). In der Form

$$B_{ab} = r_{ab}^2 \quad (2-3)$$

genannt Bestimmtheitsmaß, gibt er die gemeinsame („erklärte“) Varianz der Größen  $a$  und  $b$  an, üblicherweise prozentual ausgedrückt, was eigentlich aussagekräftiger als  $r$  ist. Kein Zusammenhang (Abb. 2-2, A;  $r_{ab} = 0$  bzw.  $B_{ab} = 0\%$ ) gilt im übrigen auch für die Sonderfälle, in denen entweder alle Werte  $a_i$  oder  $b_i$  konstant sind (Abb. 2-2, B und C).

Wichtig ist nun, dass die Berechnung von  $r_{ab}$  bzw.  $B_{ab}$  eine Reihe von Voraussetzungen enthält, die erst zum Teil genannt worden sind. Vollständig lauten sie:

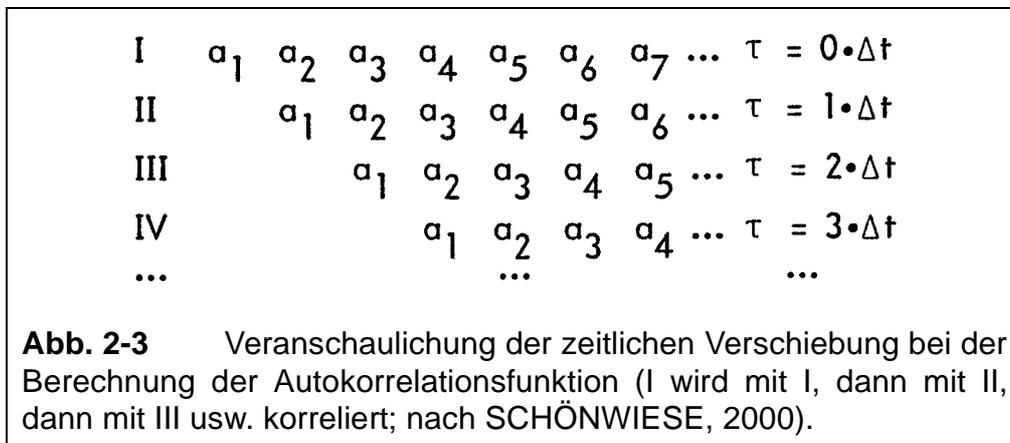
- zweidimensionaler Zusammenhang, also keine weiteren Einflußgrößen;
- linearer Zusammenhang;
- zumindest annähernde Normalverteilung der Daten  $a_i$  und  $b_i$ ;
- keine Autokorrelation (Abhängigkeit in sich) der Daten  $a_i$  bzw.  $b_i$ ;
- genügend große Stichprobenumfänge  $n$ .

Zunächst soll weiterhin angenommen werden, dass die erstgenannte Voraussetzung erfüllt ist. Ist der Zusammenhang nicht-linear (vgl. Abb. 2-2, F - I) oder/und liegen deutliche Abweichungen von der Normalverteilung vor (entsprechende Tests siehe z.B. SACHS, 1999; SCHÖNWIESE, 2000), muß statt (2-1) die *Rangkorrelationsrechnung* angewandt werden. Natürlich wird sich dabei im Streudiagramm wiederum nicht exakt eine bestimmte Kurve ergeben, sondern die Daten werden mehr oder weniger darum herum streuen. Rangkorrelationskoeffizienten sind von mehreren Autoren angegeben worden; am bekanntesten sind die Formeln nach SPEARMAN sowie KENDALL (siehe SCHÖNWIESE, 2000; dort auch Rechenbeispiele). Da (2-1) in solchen Fällen den Zusammenhang unterschätzt, ist dies sozusagen nicht besonders gefährlich.

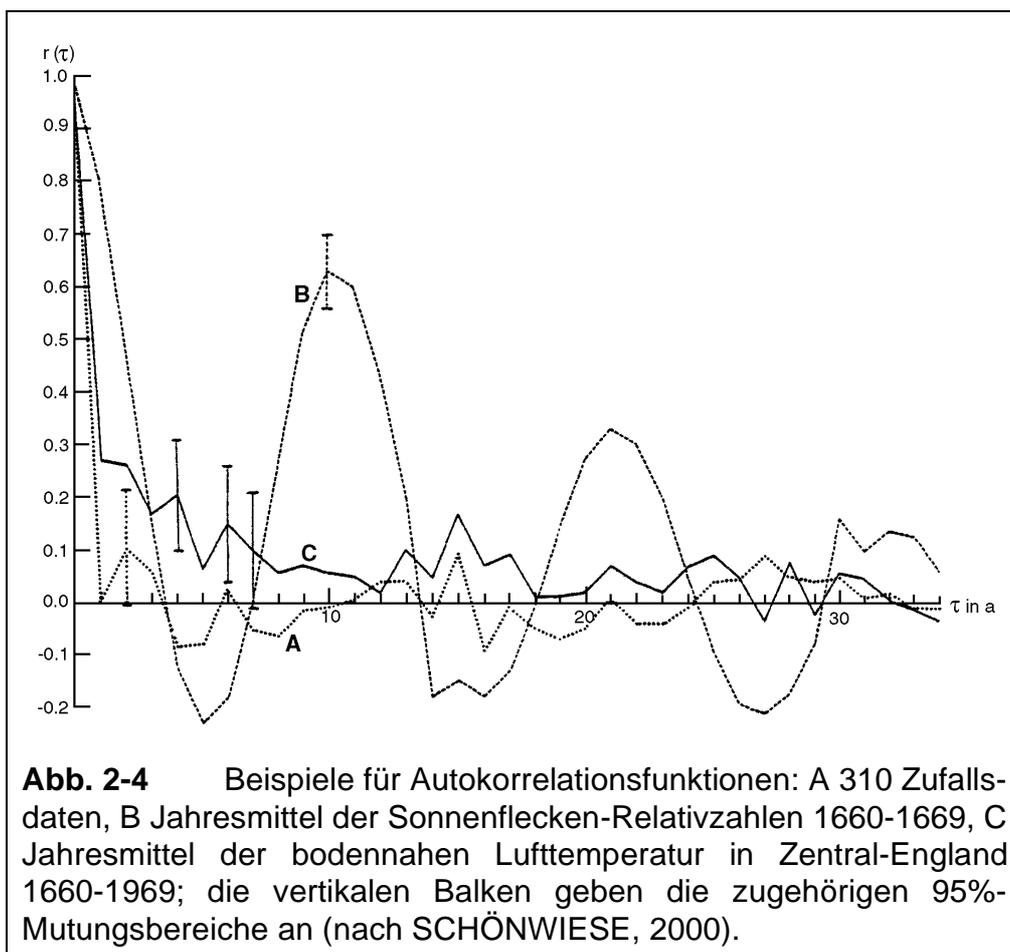
Zudem kann versucht werden, den Zusammenhang zu „linearisieren“, beispielsweise indem entweder die Daten  $a_i$  oder  $b_i$  oder beide logarithmisiert werden ( $^{10}\log a_i$  oder  $\ln a_i; b_i$  entsprechend). Ist dies erfolgreich, kann mit den entsprechend modifizierten (hier logarithmierten) Werten wie im linearen Fall gerechnet werden. In der Praxis kommen allerdings noch viele weitere nicht-lineare Zusammenhänge in Frage (eine Liste ausgewählter Regressionsgleichungen dazu folgt im Kapitel 2.3). Zu prüfen ist bei nicht-linearen Zusammenhängen, ob sie wenigstens monoton sind. Ist das nicht der Fall, d.h. treten bei der vermuteten Kurve, welche die Art des Zusammenhangs repräsentiert (näheres s.u., Regressionsrechnung), relative Maxima bzw. Minima auf, ist auch die Rangkorrelationsrechnung nicht mehr anwendbar und es muß zur wesentlich komplizierteren Berechnung der *Transinformation* übergegangen werden (SCHÖNWIESE, 2000).

Unter Autokorrelation versteht man eine Abhängigkeit der Daten in sich. Am einfachsten ist das zu verstehen, wenn die Daten in *Zeitreihenform* vorliegen, d.h. bezogen auf äquidistant fortschreitende Zeitpunkte oder Zeitintervalle (Beispiele: Beobachtung einmal pro Tag, z.B. Aktienkurs; wöchentlich akkumulierte Werte, z.B. Stromverbrauch; Monatsmittelwerte, z.B. der bodennahen Lufttemperatur in Potsdam usw.). Es wird dann die entsprechende Zeitreihe zunächst (gemäß (2-1)) mit sich selbst korreliert, was  $r_{aa} = r_A = 1$  (bzw.  $B_A = 100\%$ ) ergeben muß, dann aber gemäß dem Schema Abb. 2-3 um einen Zeitschritt gegenüber sich selbst verschoben (Zeitverschiebung  $\tau = 1\Delta t$  mit  $\Delta t =$  Zeitschritt), dann  $\tau = 2\Delta t$  usw.; man erhält dann eine Autokorrelationsfunktion  $r_A(\tau)$ , die je nach  $\tau$  i.a. unterschiedliche Werte annimmt;

vgl. Beispiele Abb. 2-4. Der Bereich, in dem  $r_A$  signifikant (was noch zu prüfen ist) oberhalb des Wertes „0“ verbleibt, wird als *Persistenzintervall* bezeichnet; in dieser Zeit besteht eine gewisse Erhaltungsneigung. Relative Maxima weisen auf zyklische Varianz hin, auf die später noch kurz eingegangen wird.



Sowohl die Berücksichtigung der Autokorrelation bei der statistischen Schätzung von Zusammenhängen als auch die Frage, ob denn die verwendeten Stichprobenumfänge groß genug sind, ist ein Problem der *Signifikanzprüfung*. Ganz allgemein versteht man in der Statistik darunter die Frage, ob irgendwelche Feststellungen, hier die



Errechnung irgendeines Korrelationskoeffizienten, auf Zufall beruht oder aber eine Systematik des dahinterstehenden Prozesses repräsentiert. Zum Beispiel und in anderen Worten könnte danach gefragt werden, ob ein errechneter Wert von  $r_{ab} = 0,3$  (entsprechend  $B = 9\%$ ) Zufall ist, d.h. nach wesentlicher Vergrößerung der Stichprobe sich vielleicht als  $r_{ab} \approx 0$  entpuppt oder nicht.

Spätestens an dieser Stelle kommt die *Wahrscheinlichkeit* ins Spiel, die in der Statistik von fundamentaler Bedeutung ist. Es wird nämlich nun, in gewisser Analogie zu einem mathematischen Widerspruchsbeweis, entschieden, ob die Nullhypothese {Korrelation = 0} oder die Alternativhypothese {Korrelation  $\neq$  0} richtig ist. Da es aber in der Statistik nie sichere Entscheidungen gibt, muß noch festgelegt werden, mit welcher Wahrscheinlichkeit diese Entscheidung getroffen werden soll. In diesem Zusammenhang heißt diese Wahrscheinlichkeit Signifikanzniveau  $S_i$ , z.B.  $S_i = 95\%$ ; häufig wird dazu auch der komplementäre Wert, nämlich die Irrtumswahrscheinlichkeit  $\alpha$  angegeben, in diesem Fall z.B.  $\alpha = 5\%$  oder meist in der Form  $\alpha = 0.05$ .

Die Entscheidung erfolgt in Form eines statistischen Tests, in diesem Fall nach der Formel

$$t = \sqrt{(n-2)/(1-r^2)} \quad (2-4)$$

d.h., es muß dieser  $t$ -Wert berechnet werden, um dann mit dem Tabellenwert einer  $t$ -Verteilung in Abhängigkeit vom gewählten  $S_i$  und der sog. Zahl der Freiheitsgrade  $\Phi$  (in diesem einfachen Fall gilt  $\Phi = n - 2$ ) verglichen zu werden. Ist der  $t$ -Wert nach (2-4) größer als der betreffende Tabellenwert, wird die Alternativhypothese, also signifikante Korrelation, vermutet, andernfalls die Nullhypothese. Beispielsweise findet man mit  $r_{ab} = 0,3$  und einem Stichprobenumfang von  $n = 10$  (Datenpaaren) einen  $t$ -Wert von 0,89. Aus einer  $t$ -Tabelle (z.B. SACHS, 1999; SCHÖNWIESE, 2000) entnimmt man für  $\Phi = n - 2 = 8$  und für z.B.  $S_i = 95\%$ : 1,86. Damit ist der errechnete Wert deutlich kleiner als der Tabellenwert und das Testergebnis negativ: keine signifikante Korrelation. Die gleiche Rechnung mit  $n = 40$  ergibt jedoch einen  $t$ -Wert von 1,94 ( $> 1,68$ ) und somit die Vermutung einer auf dem Niveau  $S_i = 95\%$  signifikanten Korrelation. Die Umkehr des Testentscheids ist in diesem Fall offenbar aufgrund des deutlich größeren Stichprobenumfangs erreicht worden (weitere Beispiele siehe SCHÖNWIESE, 2000).

Dieser Test des Korrelationskoeffizienten erlaubt auch die Berücksichtigung von Autokorrelationen, und zwar durch die sog. reduzierte Zahl der Freiheitsgrade

$$\Phi_r = \Phi(1 - r_{A1}r_{A2})/(1 + r_{A1}r_{A2}) \quad (2-5)$$

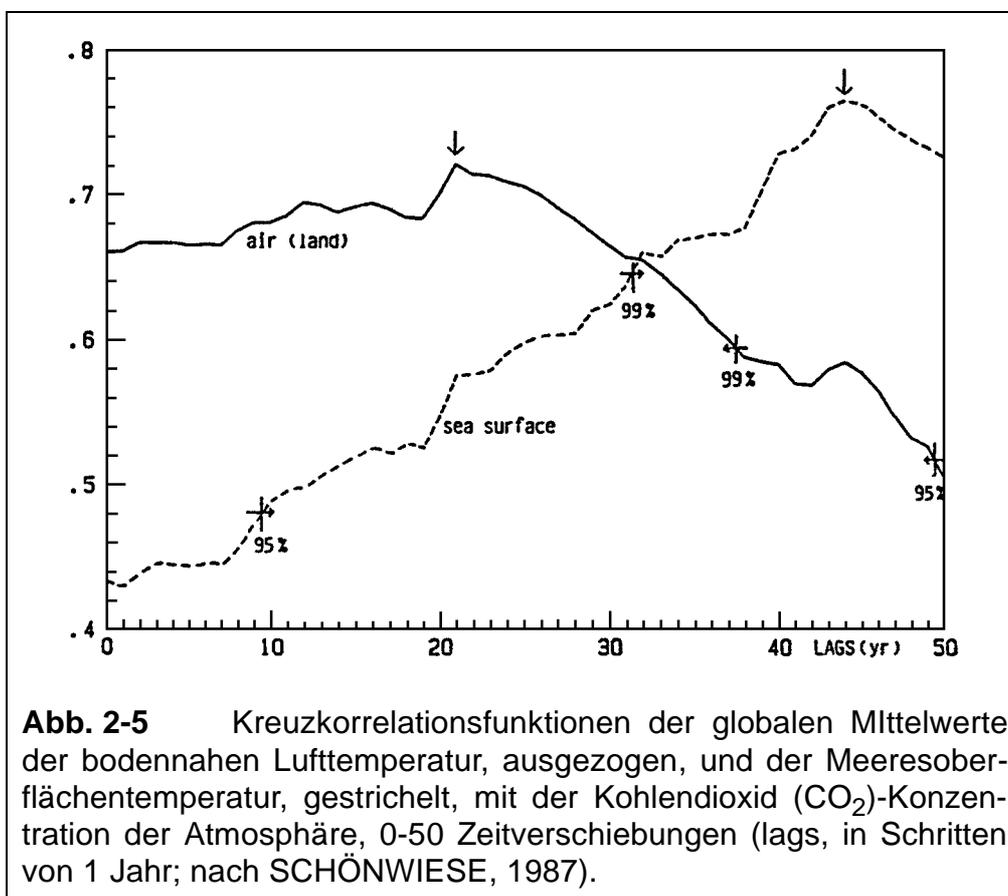
wobei  $r_{A1}$  und  $r_{A2}$  die Autokorrelationskoeffizienten der beiden betrachteten Zeitreihen sind.

Eine Alternative zum  $t$ -Test auf Signifikanz der Korrelation ist die Mutungsbereichsschätzung. Das bedeutet, dass ein Unsicherheitsintervall gesucht ist, in dem sich  $r_{ab}$  bei Variation des Stichprobenumfangs bewegen kann. Dieser Bereich lautet:

$$\pm z_\alpha(1 - r_{ab}^2)/\sqrt{n-1} \quad (2-6)$$

mit  $z_\alpha$  = Parameter der Standardnormalverteilung (Tabellen dazu siehe z.B. SACHS, 1999; SCHÖNWIESE, 2000). Für  $\text{Si} = 95\%$  ist  $z_\alpha = 1,96$ ; daraus folgt für das obige Beispiel mit  $n = 10$  eine Unschärfe von  $\pm 0,59$ , mit  $n = 40$  von  $\pm 0,29$ . Diese Alternative ist offenbar wesentlich kritischer als der  $t$ -Test; denn erst ab ungefähr  $n = 40$  unterscheidet sich der Korrelationskoeffizient  $0,3$  auf dem Niveau  $\text{Si} = 95\%$  gerade eben von  $0$ , ist also entsprechend signifikant. Außerdem gibt es in dieser Variante auch Abschätzungsformeln für die Rangkorrelationsrechnung (Details und Beispiele s. SCHÖNWIESE, 2000).

Das Prinzip der Zeitverschiebung, das im Zusammenhang mit der Autokorrelation genannt worden ist, läßt sich natürlich auch auf unterschiedliche Zeitreihen anwenden. In diesem Fall spricht man von der *Kreuzkorrelationsrechnung* und erhält in Abhängigkeit von der Zeitverschiebung  $\tau$  die Kreuzkorrelationsfunktion  $r_K(\tau)$ , die angibt, wie groß der Korrelationskoeffizient dabei jeweils ist. Abb. 2-5 bringt dafür ein

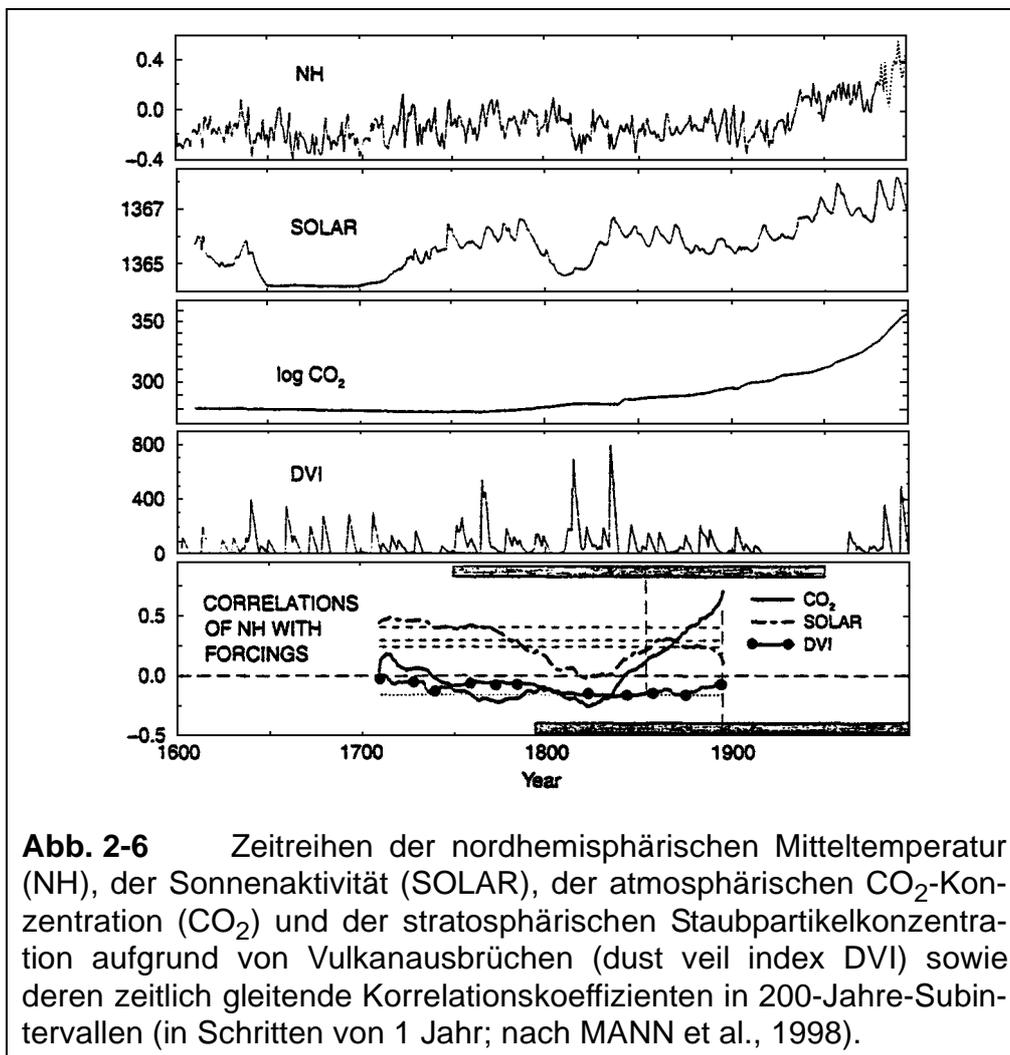


**Abb. 2-5** Kreuzkorrelationsfunktionen der globalen Mittelwerte der bodennahen Lufttemperatur, ausgezogen, und der Meeresoberflächentemperatur, gestrichelt, mit der Kohlendioxid ( $\text{CO}_2$ )-Konzentration der Atmosphäre, 0-50 Zeitverschiebungen (lags, in Schritten von 1 Jahr; nach SCHÖNWIESE, 1987).

Beispiel. Es zeigt, dass das Maximum der Korrelation zwischen der  $\text{CO}_2$ -Konzentration der Atmosphäre und den globalen Mittelwerten der bodennahen Lufttemperatur erst bei einer Zeitverzögerung von 22 Jahren, im Fall der Meeresoberflächentemperatur sogar erst nach 44 Jahren erreicht wird (physikalisch nicht unplausibel, aber sicherlich mit erheblicher Unschärfe belastet). Was weitere Anwendungen betrifft, nämlich die Fouriertransformation der Autokorrelationsfunktion ( $\rightarrow$  spektrale Varianzanalyse, sog. Autospektrum) bzw. der Kreuzkorrelationsfunktion ( $\rightarrow$  Kreuzspektrum mit seinen verschiedenen Komponenten) soll hier nur noch die aus der letztgenannten Methodik hervorgehende *Kohärenzanalyse* erwähnt sein, welche den Korrelationskoeffizienten aufgeschlüsselt nach den Frequenzanteilen angibt, um

festzustellen, ob die Zusammenhänge eher im relativ hochfrequenten (kurzperiodischen) oder relativ tieffrequenten (langperiodischen) oder Bereich bestimmter zyklischer Varianz etabliert ist (hierzu siehe u.a. SCHLITGEN und STREITBERG, 1999; SCHÖNWIESE, 2000; zu alternativen Methoden mit Hilfe der Informationsentropie siehe OLBERG und RAKOCZI, 1984).

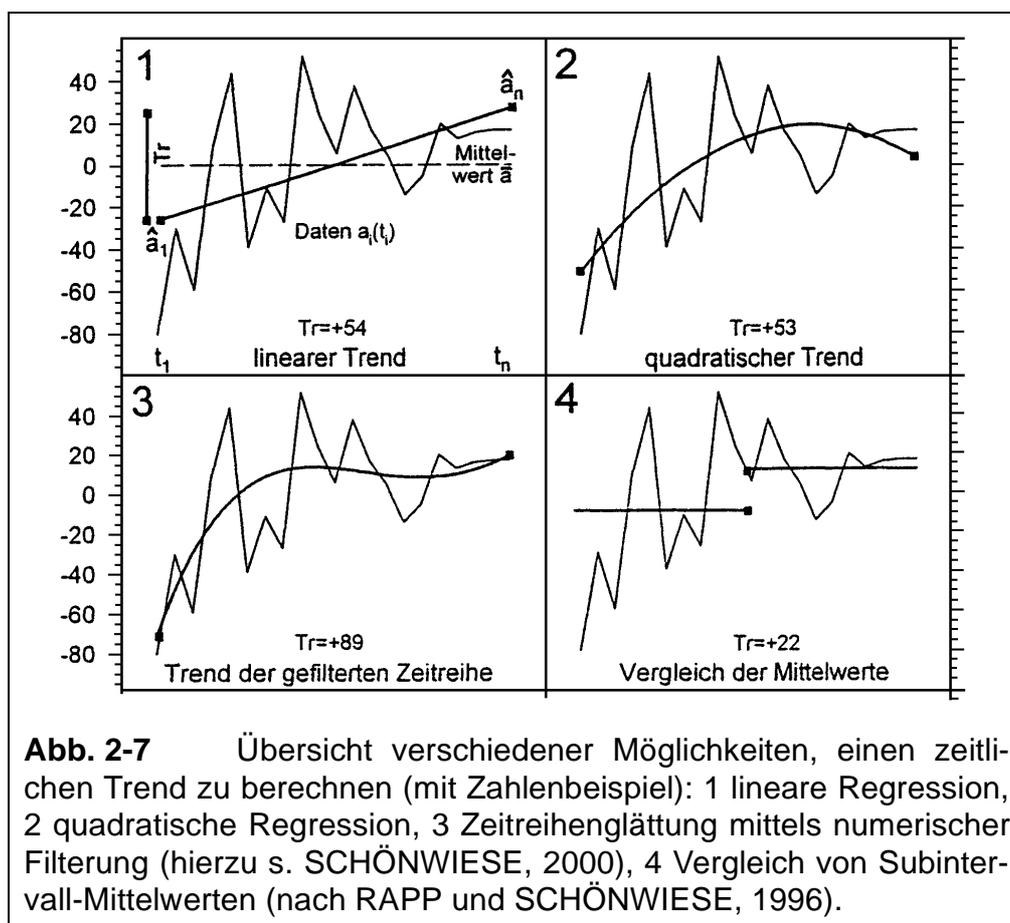
Ein wieder anderer Aspekt, der generell in der statistischen Zeitreihenanalyse aufschlußreich ist, ist die *zeitlich gleitende Korrelationsrechnung*. Dabei wird in Orientierung an das insgesamt erfaßte Zeitintervall ein kleineres Subintervall definiert, dies sozusagen an den Anfang des Gesamtzeitintervalls gesetzt und schrittweise jeweils um den Zeitschritt  $\Delta t$  verschoben, in dem die Zeitreihen vorliegen, bis man damit am Ende des Gesamtzeitintervalls angelangt ist. Ein Beispiel dazu ist in Abb. 2-6 wiedergegeben, wo die nordhemisphärisch gemittelte bodennahe Lufttemperatur 1610-1980



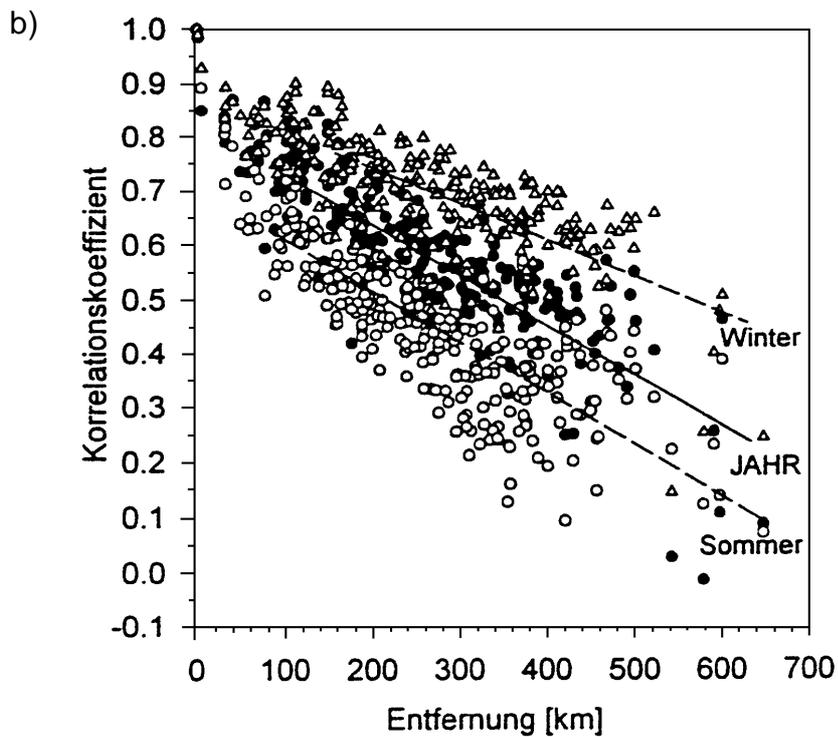
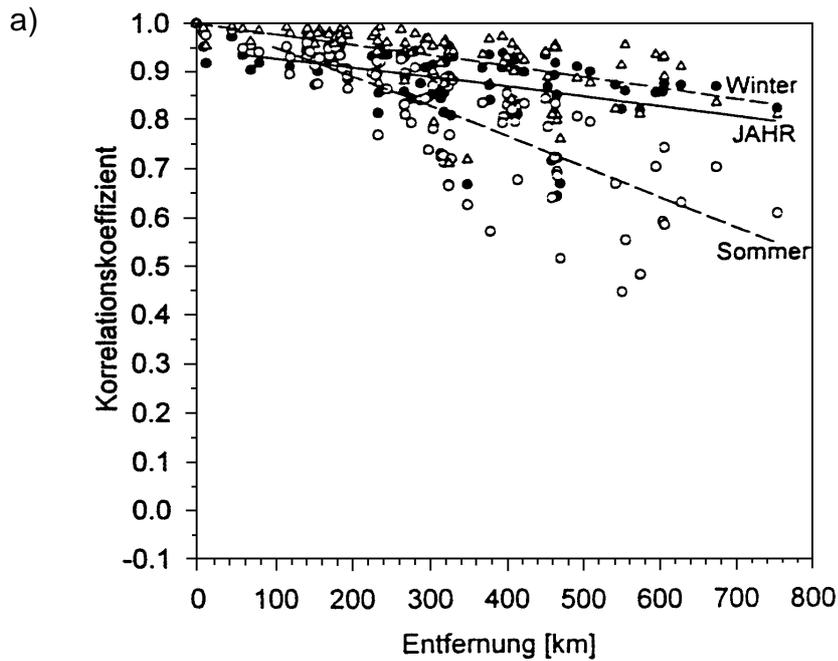
**Abb. 2-6** Zeitreihen der nordhemisphärischen Mitteltemperatur (NH), der Sonnenaktivität (SOLAR), der atmosphärischen CO<sub>2</sub>-Konzentration (CO<sub>2</sub>) und der stratosphärischen Staubpartikelkonzentration aufgrund von Vulkanausbrüchen (dust veil index DVI) sowie deren zeitlich gleitende Korrelationskoeffizienten in 200-Jahre-Subintervallen (in Schritten von 1 Jahr; nach MANN et al., 1998).

in 200-Jahre-Subintervallen zeitlich gleitend mit Indexwerten der Sonnenaktivität, des explosiven Vulkanismus und der CO<sub>2</sub>-Konzentration der Atmosphäre korreliert ist. Man erkennt, dass im letzten Jahrhundert die CO<sub>2</sub>-Korrelation systematisch angestiegen und dominierend geworden ist, während davor, insbesondere in der frühen Zeit, die solare Korrelation vorgeherrscht hat.

Eine weitere Anwendung der zweidimensionalen Korrelationsrechnung ist die *Trendanalyse*, die nichts anderes als eine Korrelation einer Zeitreihe mit der Zeit ist. Dazu zeigt Abb. 2-7 zunächst einmal die verschiedenen Möglichkeiten der Trendberechnung, die natürlich trotz gleichem Datensatz zu unterschiedlichen Ergebnissen führen können. Da hier weniger der Vergleich der Subintervall-Mittelwerte (Teilbild 4), sondern eher die lineare (Teilbild 1) bzw. nicht-lineare (Teilbilder 2 und 3) Trendberechnung von Interesse ist und dies in besonders engem Zusammenhang mit der Regressionsrechnung steht, soll es erst dort (Kapitel 2.3) erörtert werden.



Schließlich soll noch die Anwendung der zweidimensionalen Korrelationsrechnung auf die Repräsentanzanalyse betrachtet werden. Es gibt zwar auch zeitliche Repräsentanzbetrachtungen; meist versteht man aber darunter die Frage nach räumlichen Zusammenhängen. So ist in Abb. 2-8a eine Graphik zu sehen, in der die Korrelationskoeffizienten von Temperaturzeitreihen an unterschiedlichen Stationen als Funktion der Stationsentfernung aufgetragen sind. Man sieht, dass insbesondere bei Winter- bzw. Jahresdaten die Repräsentanz recht gut ist; denn erst ab einer Stationsentfernung von reichlich 1000 km fällt  $r$  auf Werte von unter 0,7 (entspr.  $B$  unter 50%) ab. Ganz anders ist das beim Niederschlag, vgl. Abb. 2-8b, wo dieser Abfall bei den Sommerdaten bereits bei Stationsentfernungen von etwa 50 km an der Fall ist.



**Abb. 2-8** a) Repräsentanzanalyse von jährlichen und jahreszeitlichen ( $\Delta$  Winter = Dez.-Febr.,  $\circ$  Sommer = Juni-Aug.,  $\bullet$  Jahresdaten) deutschen Temperaturdaten 1951-1990, 95 Stationen, Referenzstation Frankfurt a.M., in Form der Korrelation in Abhängigkeit von der Stationsentfernung (nach RAPP und SCHÖNWIESE, 1996). b) Ähnlich Abbildung a, jedoch Niederschlagssummen 1891-1990, 250 Stationen.

## 2.3 Zweidimensionale Regressionsrechnung

Ist die zweidimensionale lineare Korrelationsrechnung zufriedenstellend durchgeführt, kann mittels der Regressionsrechnung die zugehörige *Regressionsgerade* berechnet werden, welche die Art des Zusammenhangs angibt. Diese hat die Form

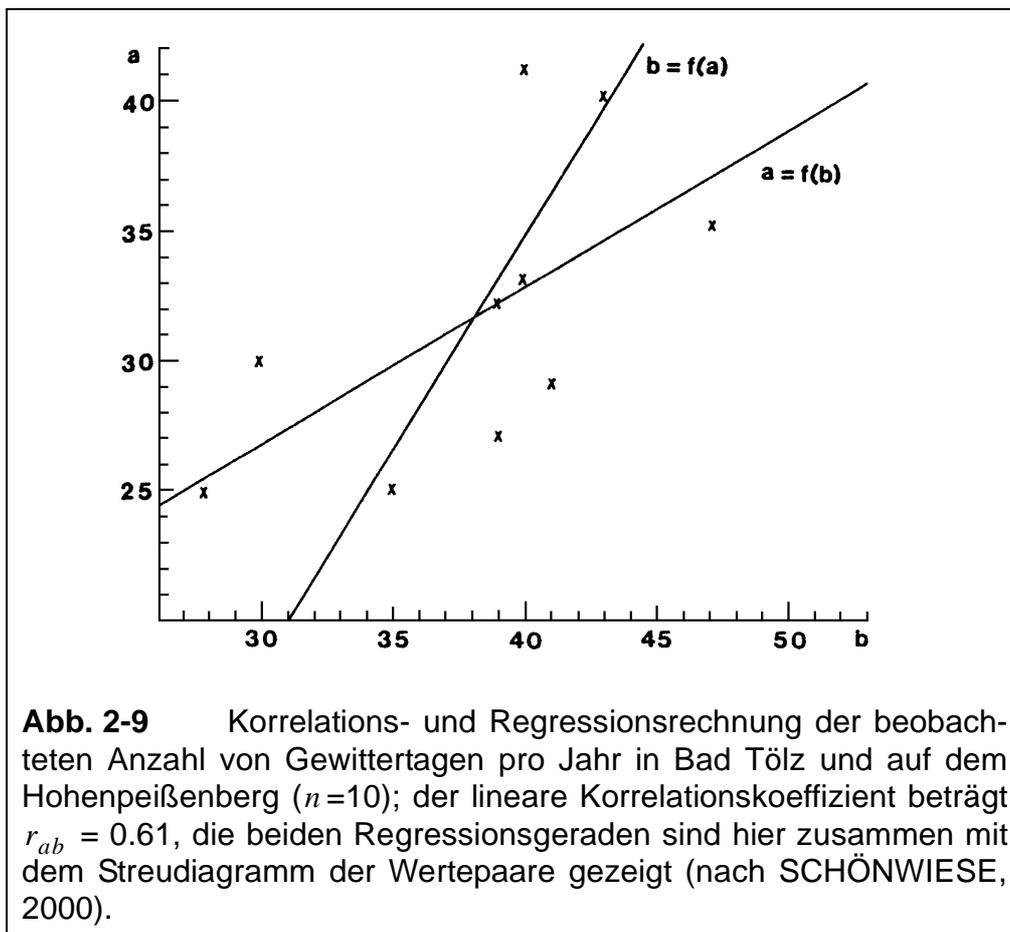
$$\hat{a} = A + Bb \quad (2-7)$$

mit  $A$  = Achsenabschnitt und  $B$  Geradensteigung (die auch negativ sein kann). Die Berechnung von  $A$  und  $B$  kann nach unterschiedlichen Formeln erfolgen. Hier sei nur die Berechnung über die Entwicklung der sog. Normalgleichungen genannt, weil sich diese Methodik problemlos auch auf nicht-lineare sowie mehr ( $>2$ )-dimensionale Zusammenhänge erweitern läßt. Dabei wird die vermutete Grundgleichung, hier (2-7), summiert und anschließend mit der Einflußgröße (unabhängigen Variablen), hier  $b$ , multipliziert; das ergibt

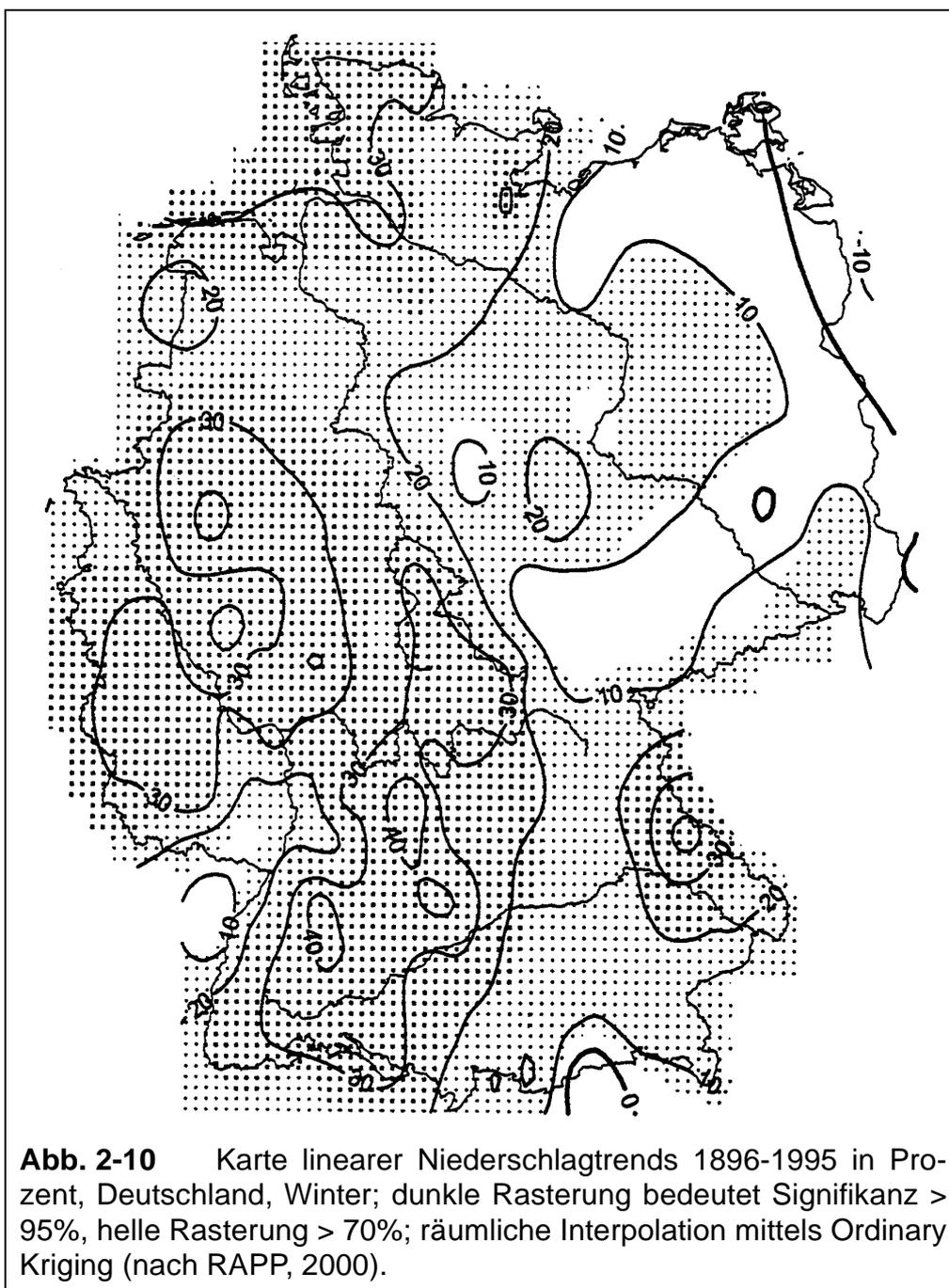
$$\sum a_i = An + B \sum b_i \quad (2-8)$$

$$\sum a_i b_i = A \sum b_i + B \sum b_i^2 \quad (2-9)$$

und somit zwei Gleichungen für die beiden Unbekannten  $A$  und  $B$  (Details und Alternativen siehe SCHÖNWIESE, 2000). Das mathematische Prinzip, das hinter diesen



Berechnungen steht, ist die sog. *Methode der kleinsten Quadrate*; d.h. die Gleichungen erfüllen die Bedingung, dass die quadratischen Abweichungen der Daten  $\hat{a}_i$  (Wirkungsgröße = abhängige Variable) minimal sind. Vertauscht man  $a$  und  $b$  in (2-7) und folglich auch in (2-8) und (2-9), erhält man i.a. eine andere Geradengleichung, siehe Beispiel Abb. 2-9 (wobei in diesem Fall  $n = 10$  und  $r_{ab} = + 0,61$  ist). Die beiden Regressionsgeraden schneiden sich unter einem Winkel, der indirekt ein Maß für die Korrelation ist (senkrecht, falls  $r = 0$ , Identität falls  $|r| = 1$ ).



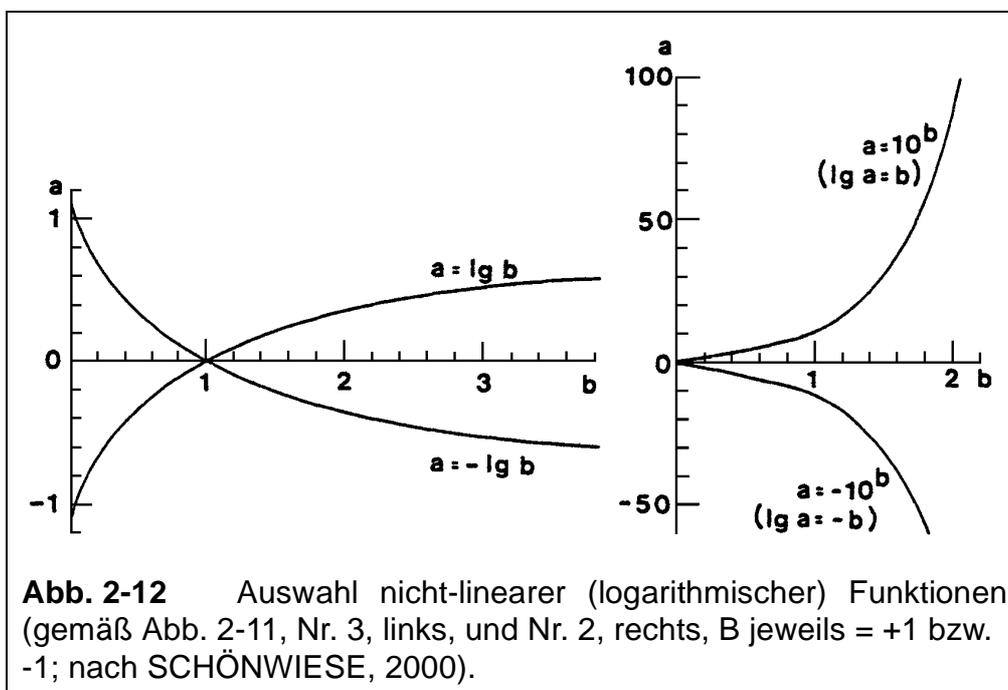
Bei der Anwendung auf Trendanalysen ist es häufig von Interesse, diese auch räumlich differenzierend zu berechnen. Liegen so für unterschiedliche Orte unterschiedliche Trendwerte vor, sind weitergehend räumliche Interpolationsverfahren gefragt (z.B. sog. Kriging), die zu Isolinien-Darstellungen führen. Auf solche Interpolations-

techniken soll hier nicht eingegangen (s. dazu SCHÖNWIESE und RAPP, 1997), wohl aber ein klimatologisches Beispiel linearer Trendanalysen gezeigt werden, und zwar des Winterniederschlags in Deutschland, vgl. Abb. 2-10 (nach RAPP, 2000; vgl. dazu auch RAPP und SCHÖNWIESE, 1996; SCHÖNWIESE und RAPP, 1997). Die Signifikanzprüfung erfolgt am besten, da parameterfrei (d.h. ohne dass Normalverteilungen vorausgesetzt werden) mit Hilfe des MANN-KENDALL-Trendtests (Methodik siehe SNEYERS, 1990; SCHÖNWIESE und RAPP, 1997).

Nr.	Regressionsgleichung	Zugehörige Normalgleichungen
1	$\hat{a} = A + Bb$	$\sum a_i = An + B\sum b_i$ $\sum a_i b_i = A\sum b_i + B\sum b_i^2$
2	$\lg \hat{a} = A + Bb$	$\sum \lg a_i = An + B\sum b_i$ $\sum (b_i \lg a_i) = A\sum b_i + B\sum b_i^2$
3	$\hat{a} = A + B \lg b$	$\sum a_i = An + B\sum \lg b_i$ $\sum (a_i \lg b_i) = A\sum \lg b_i + B\sum (\lg b_i)^2$
4	$\lg \hat{a} = A + B \lg b$	$\sum \lg a_i = An + B\sum \lg b_i$ $\sum (\lg a_i \lg b_i) = A\sum \lg b_i + B\sum (\lg b_i)^2$
5	$\hat{a} = AB^b$ bzw. $\lg \hat{a} = \lg A + (\lg B) b$	$\sum \lg a_i = (\lg A)n + \lg B \sum b_i$ $\sum (b_i \lg a_i) = (\lg A) \sum b_i + (\lg B) \sum b_i^2$
6	$\hat{a} = A + Bb^2$	$\sum a_i = An + B\sum b_i^2$ $\sum a_i b_i = A\sum b_i + B\sum b_i^3$

Anmerkung: Nr. 6 modifizierbar durch  $a = A + Bb^3$ ;  $a = A + Bb^4$ ; usw.; außerdem können die Steigungen (Koeffizienten B) auch negativ sein.

**Abb. 2-11** Auswahl nicht-linearer Regressionsgleichungen mit zugehörigen Normalgleichungen (nach SCHÖNWIESE, 2000).



Im nicht-linearen Fall der Regression wird ganz analog zum linearen vorgegangen. Abb. 2-11 bringt dazu eine Auswahl von Funktionen und zugehörigen Normalglei-

chungen, Abb. 2-12 die Grundformen des Kurvenverlaufs bei logarithmischen Beziehungen. Eine Alternative dazu sind die schon genannten Linearisierungsverfahren. In der Praxis wird man oft nicht gleich wissen, welche nicht-lineare Funktion am besten geeignet ist; dann bleibt nichts anderes als „Probieren“ übrig und es wird schließlich das Regressionsmodell gewählt, das die beste Korrelation aufweist. Auch dazu gibt es allerdings noch eine Alternative, nämlich die Berechnung von  $r = |\sqrt{B}|$  indirekt aus dem Vergleich der Datenwerte mit der vermuteten Regressionskurve. Die betreffende Formel, die übrigens ganz allgemein (auch für beliebige mehrdimensionale Zusammenhänge) anwendbar ist, lautet:

$$B_{ab} = \{\text{erklärte Varianz/Gesamtvarianz}\} = \sum (\hat{a}_i - \bar{a})^2 / \sum a_i'^2, \quad (2-10)$$

wobei  $\hat{a}_i$  die Daten der Regressionsbeziehung  $\hat{a}_i = f(b_i)$  sind.

## 2.4 Mehrdimensionale Korrelations- und Regressionsrechnung

Kommt zunächst zur Einflußgröße  $b$  eine weitere, hier  $c$  genannt, hinzu, so ist prinzipiell zu überlegen, ob diese weitere Größe als Störfaktor oder tatsächlich als weiterer Einfluß aufzufassen ist. Im ersten Fall ist die *partielle*, im zweiten Fall die *multiple* Korrelations- und Regressionsrechnung anzuwenden.

Beispielsweise kann es sein, dass nicht  $a$  mit  $b$ , sondern  $a$  mit  $c$  und  $b$  mit  $c$  korreliert sind. In diesem Fall führt die partielle Korrelationsrechnung zu einem Korrelationskoeffizienten  $r_{part} \approx 0$  und es ist eine Scheinkorrelation aufgedeckt, falls vorher  $r \neq 0$  gewesen ist. Es kann aber auch  $|r_{part}| > |r|$  herauskommen,  $c$  somit tatsächlich eine Störung sein, welche die Korrelation  $r_{ab}$  mehr oder weniger verschleiert. Die Formel für den *dreidimensionalen partiellen Korrelationskoeffizienten* lautet

$$r_{part}(ab \cdot c) = (r_{ab} - r_{ac}r_{bc}) / \sqrt{[(1 - r_{ac}^2)(1 - r_{bc}^2)]}, \quad (2-11)$$

wobei diese Schreibweise bedeutet, dass  $c$  als Störfaktor aufgefaßt und daher ausgeschaltet wird;  $r_{ab}$  usw. sind die üblichen zweidimensionalen Korrelationskoeffizienten (vier- und noch höherdimensionale Fälle ziemlich kompliziert; siehe dazu SCHÖNWIESE, 2000). Die Signifikanzprüfung des partiellen Korrelationskoeffizienten kann wie im zweidimensionalen Fall durch einen  $t$ -Test erfolgen. Auch die genannten Mutungsbereichabschätzungen sind auf die partielle Korrelationsrechnung übertragbar.

Der weitaus häufiger verwendete *lineare multiple Korrelationskoeffizient* lautet im dreidimensionalen Fall (quadratisch und daher hinsichtlich des Vorzeichens unbestimmt)

$$B_{mult} = r_{mult}^2(a \cdot bc) = (r_{ab}^2 + r_{ac}^2 - 2r_{ab}r_{ac}r_{bc}) / (1 - r_{bc}^2) \quad (2-12)$$

wobei diese Schreibweise bedeutet, dass  $a$  zugleich mit  $b$  und mit  $c$  korreliert wird. Für vier- und noch höherdimensionale Fälle ist allerdings die Form (2-10) praktischer. Dabei erhält man die zugehörige lineare Regressionsgleichung am besten wie-

der über die Entwicklung der Normalgleichungen, die im dreidimensionalen Fall für die Grundgleichung

$$\hat{a} = A + Bb + Cc \quad (2-13)$$

lauten (wieder Summierung, dann Multiplikation mit  $b$ , dann mit  $c$ )

$$\sum a_i = An + B\sum b_i + C\sum c_i \quad (2-14)$$

$$\sum a_i b_i = A\sum b_i + B\sum b_i^2 + C\sum b_i c_i \quad (2-15)$$

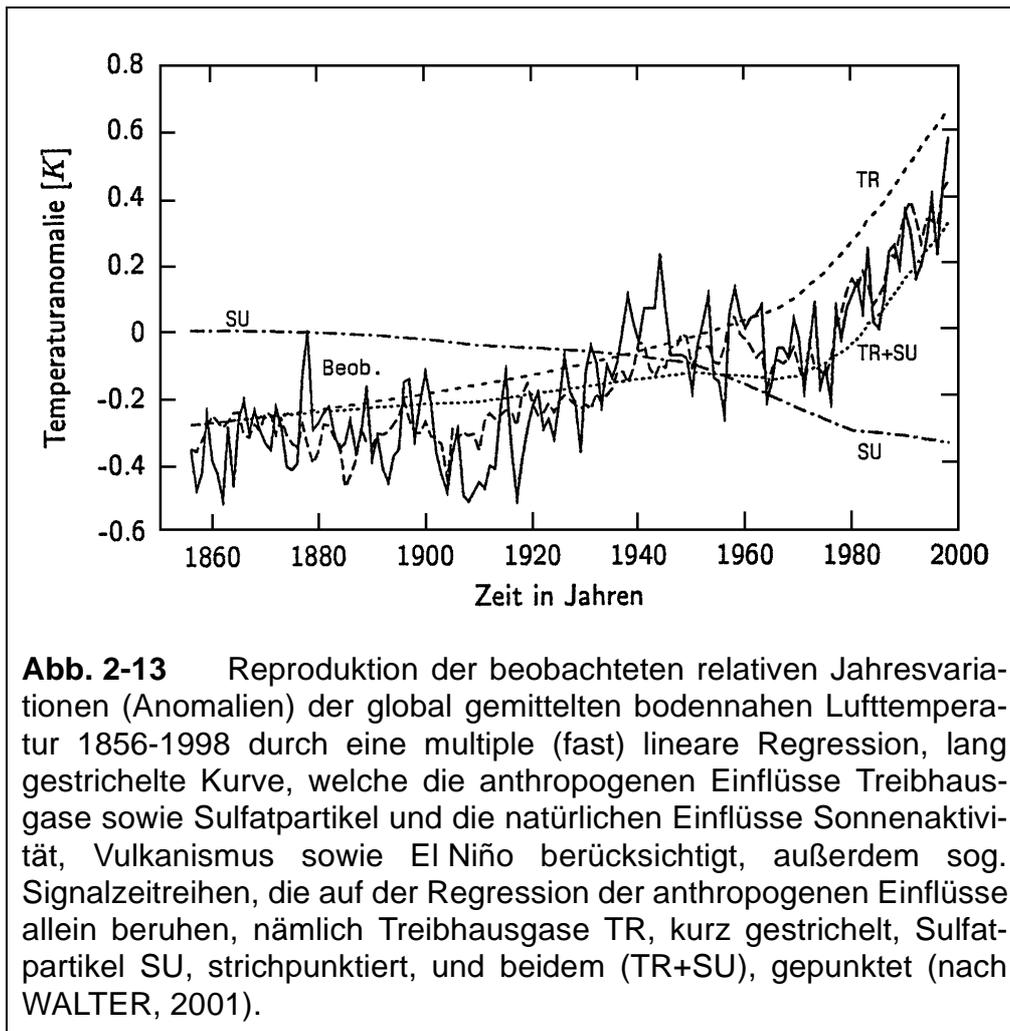
$$\sum a_i c_i = A\sum c_i + B\sum b_i c_i + C\sum c_i^2 \quad (2-16)$$

(somit 3 Gleichungen für die drei Unbekannten  $A$ ,  $B$  und  $C$ ). Nach diesem Schema lassen sich nun Normalgleichungen für beliebig viele Einflußgrößen entwickeln, wobei es sich offenbar um die Lösung linearer Gleichungssysteme handelt, und die zugehörigen multiplen Korrelationskoeffizienten nach (2-10) berechnen. Die Signifikanzprüfung erfolgt im >2-dimensionalen Fall durch einen  $F$ -Test (Details und Beispiele siehe SCHÖNWIESE, 2000).

Bei sehr vielen Einflußgrößen ist es sinnvoll, erst alle zweidimensionalen Korrelationskoeffizienten  $r_{ab}$ ,  $r_{ac}$ ,  $r_{ad}$  usw. zu berechnen und bei der multiplen Korrelation mit dem Maximum dieser Werte zu beginnen, im zweiten Schritt die zweithöchste Korrelation hinzuzunehmen usw.; dabei ist zu entscheiden, bei welcher jeweils erreichten erklärten Varianz  $B_{mult}$  das Ergebnis als befriedigend angesehen werden soll. Die verbleibende unerklärte Varianz sollte dann allerdings, im Fall des linearen multiplen Regressionsmodells, normalverteilt sein. Im Fall von Zeitreihenanalysen sollte dieses Residuum darüber hinaus stationär sein (üblicherweise Mittelwert-, Varianz- und Autokorrelationsstationarität zu prüfen, wobei stationär bzw. praktikabler quasistationär bedeutet, dass diese statistischen Größen zeitlich (annähernd) invariant sind).

Bei allen Korrelations- und Regressionsrechnungen, insbesondere aber bei den multiplen, muß bedacht werden, ob die dabei vermuteten Ursache-Wirkung-Beziehungen auch einen sinnvollen, bei naturwissenschaftlichen Fragestellungen physikalischen bzw. physikochemischen Hintergrund aufweisen. Aufgrund solcher Betrachtungen kann es sinnvoll sein, die eine oder andere lineare Beziehung durch eine nicht-lineare zu ersetzen (falls dadurch auch  $B_{mult}$  signifikant ansteigt).

In Abb. 2-13 ist ein Beispiel zu sehen, in dem die zeitlichen Variationen der global gemittelten bodennahen Lufttemperatur (somit Zeitreihe) mit der äquivalenten atmosphärischen  $\text{CO}_2$ -Konzentration (somit näherungsweise alle sog. Treibhausgase erfaßt), der Sulfatpartikelkonzentration der unteren Atmosphäre, der Sonnenaktivität, dem Vulkanismus und dem El-Niño-Phänomen (jeweils durch entsprechende Parameter-Zeitreihen repräsentiert) multipel korreliert und in Form einer fast-linearen multiplen Regression reproduziert sind (Treibhausgas-Temperatur-Beziehung logarithmisch, ansonsten linear). Der Vergleich der so erhaltenen Regressionszeitreihe mit den beobachteten Temperaturdaten, jeweils Jahresmittelwerte, zeigt, dass diese Reproduktion recht zufriedenstellend ausfällt; die Varianzerklärung beträgt



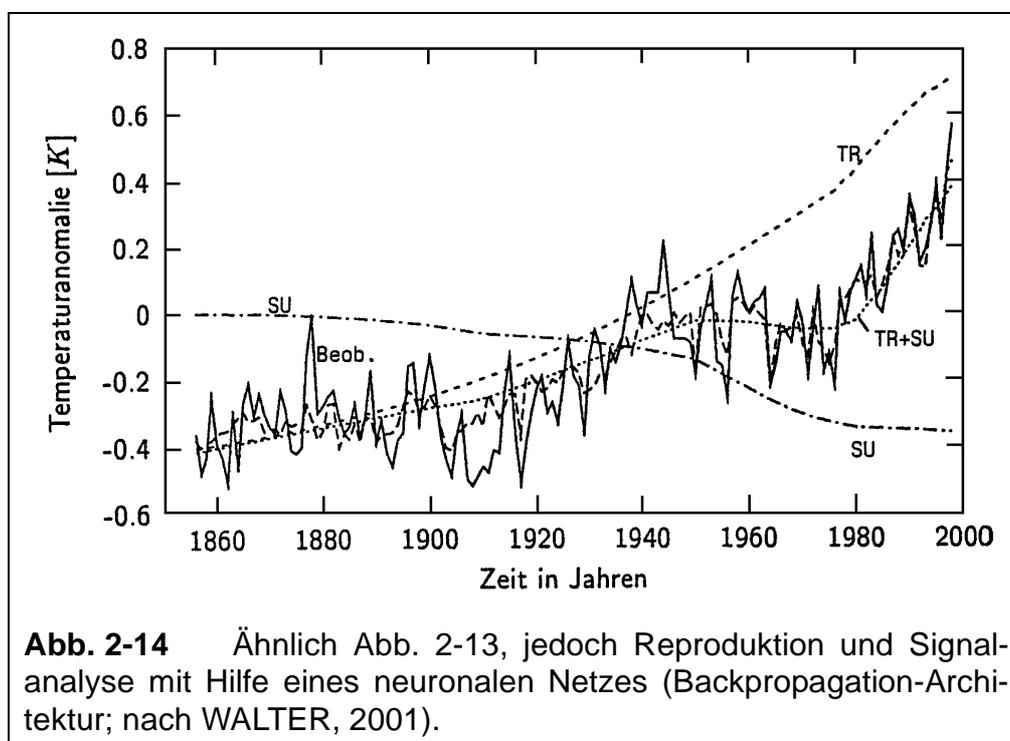
dabei  $B = 75\%$ . Außerdem sind noch die Regressionszeitreihen allein für den Einfluß der Treibhausgase, der Sulfatpartikel, beide anthropogen, und deren Summe gezeigt, die sog. Signalzeitreihen (weitere Details dazu siehe WALTER, 2001; vgl. dazu auch WALTER et al., 1998; GRIESER et al., 2000).

## 2.5 Ausblick und Wertung

In dieser kurzen Zusammenschau konnten nur einige grundlegende, eher konventionelle Analysetechniken von Zusammenhängen behandelt werden. Nicht nur bezüglich der Details, auch hinsichtlich fortgeschrittener Methoden muß auf die Literatur verwiesen werden (z.B. CHATFIELD, 1996; OLBERG und RAKOCZI, 1984; RINNE, 1997; SACHS, 1999; SCHLITTGEN und STREITBERG, 1999; SCHÖNWIESE, 2000; VON STORCH und ZWIERS, 1999; TAUBENHEIM, 1969). Zu den fortgeschrittenen Methoden gehören vor allem die *EOF-Techniken* (empirische Orthogonalfunktionen), welche mögliche Abhängigkeiten zwischen den Einflußgrößen eliminieren, einschließlich deren Anwendungen wie die *Hauptkomponenten- und Faktorenanalyse*, weiterhin neuronale Netze.

Bei den *neuronalen Netzen*, auf die hier noch kurz eingegangen werden soll, handelt es sich um eine Art multiple Regressionsrechnung, bei der beliebig viele Einfluß- und

Wirkungsgrößen zugelassen sind, die Zusammenhänge in einer Art verdecktem Training nicht-linear (meist werden sog. sigmoide Funktionen unterlegt) und unter Zulassung von Wechselwirkungen zwischen den Einflußgrößen ermittelt werden (siehe z.B. BRAUSE, 1995; SCHÖNWIESE, 2000; WALTER, 2001). Zumeist handelt es sich um sog. Black-Box-Methoden, die nur das Ergebnis, nicht aber die Art der Zusammenhänge (im Gegensatz zur Regressionsrechnung) verraten. Daher ist es wichtig, aber auch bei der Regressionsrechnung generell ratsam, ggf. zu analysierende Zeitreihen in ein erstes Subintervall, in dem die Beziehungen ermittelt werden, und ein anschließendes Verifikationsintervall zu unterteilen, in dem geprüft wird, ob die Beziehungen stationär sind. Abb. 2-14 zeigt dafür ein Beispiel, und zwar anhand der gleichen Datensätze wie Abb. 2-13; die angewandte neuronale Netzwerk-Architektur ist hier das sog. Backpropagation (eine häufig angewandte Standardform, bei dem eine sozusagen rückgerichtete Fehlerkorrektur erfolgt (sog. überwachtes Lernen; näheres siehe o.a. Literatur). Die Varianzklärung beträgt in diesem Fall  $B = 84\%$ .



Praktisch alle hier behandelten Methoden können mit Hilfe von Software-Paketen in der EDV genutzt werden. Jedoch gehen daraus nicht immer die notwendigen Voraussetzungen und „Fußangeln“ hervor; auch die zugehörigen Testverfahren fehlen oft, obwohl sie die Ergebnisse erst absichern. Daher ist es unumgänglich, sich wenigstens über die wichtigsten Grundlagen der jeweils angewandten statistischen Analysemethodik anhand geeigneter Literatur zu informieren. Nicht weniger wichtig ist das schon erwähnte Prinzip, nicht allein formal-statistisch vorzugehen, sondern sich zu vergegenwärtigen, welche Prozesse jeweils durch die Statistik formal approximiert werden.

Bei klimatologischen Zeitreihenanalysen, insbesondere der Frage, wie sich anthropogene von natürlichen Ursachen im großräumigen Klimageschehen voneinander abgrenzen lassen, sind beispielsweise die Strahlungsantriebe wichtig, die global und troposphärisch gemittelt bestimmten externen Einflüssen auf das Klimasystem zuge-

ordnet werden können (IPCC, HOUGHTON et al., 2001). Den daraus ersichtlichen physikalischen Grundprinzipien darf die Statistik nicht widersprechen, die ja stets nur formal die Frage der Analyse von Zusammenhängen angehen kann, nicht wirklich ursächlich. Eine solche bedachte Vorgehensweise hilft, der eingangs genannten Skepsis gegenüber statistischen Methoden zu begegnen.

## 2.6 Literatur

- Brause, R. 1995: Neuronale Netze (2. Aufl.). Teubner, Stuttgart.
- Chatfield, C. 1996: The Analysis of Time Series (5. ed.). Chapman and Hall, London.
- Grieser, J., Staeger, T., Schönwiese, C.-D. 2000: Statistische Analyse zur Früherkennung globaler und regionaler Klimaänderungen aufgrund des anthropogenen Treibhauseffektes. Bericht Nr. 103, Inst. Meteorol. Geophys. Univ. Frankfurt a.M.
- Houghton, J. T., et al. (eds.): Climate Change 2001. The Scientific Basis (Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)). Univ. Press, Cambridge.
- Mann, M. E., Bradley, R. S., Hughes, M. K. 1998: Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392**, 779-787.
- Olberg, M., Rakoczi, F. 1984: Informationstheorie in Meteorologie und Geophysik. Akademie Verlag, Berlin.
- Rapp, J. 2000: Konzeption, Problematik und Ergebnisse klimatologischer Trendanalysen für Europa und Deutschland. Bericht Nr. 212, Deut. Wetterdienst, Offenbach.
- Rapp, J., Schönwiese, C.-D. 1996: Atlas der Niederschlags- und Temperaturtrends in Deutschland 1891-1990 (2. Aufl.). *Frankfurter Geowiss. Arb.*, Serie B, Band 5, Frankfurt a.M.
- Rinne, H. 1997: Taschenbuch der Statistik. H. Deutsch, Frankfurt a.M.
- Sachs, L. 1999: Angewandte Statistik (9. Aufl.). Springer, Berlin.
- Schlittgen, R., Streitberg, B. H. J. 1999: Zeitreihenanalyse (8. Aufl.). Oldenbourg, München.
- Schönwiese, C.-D. 1987: Observational assessments of the hemispheric and global climate response to increasing greenhouse gases. *Contr. Atmosph. Physics*, **60**, 48-64.
- Schönwiese, C.-D. 2000: Praktische Statistik für Meteorologen und Geowissenschaftler (3. Aufl.). Borntraeger, Stuttgart.
- Schönwiese, C.-D., Rapp, J. 1997: Climate Trend Atlas of Europe - Based on Observations 1891-1990. Kluwer, Dordrecht.
- von Storch, H., Zwiers, F. 1999: Statistical Analysis in Climate Research. Univ. Press, Cambridge.
- Sneyers, R. 1990: On the Statistical Analysis of Series of Observations. World Meteorol. Org. (WMO), Tech. Note No. 143, Publ. No. 415, Geneva.
- Taubenheim, J. 1969: Statistische Auswertung geophysikalischer und meteorologischer Daten. Geest & Portig, Leipzig.
- Walter, A. 2001: Zur Anwendung neuronaler Netze in der Klimatologie. Bericht Nr. 218, Deut. Wetterdienst, Offenbach.

Walter, A., Denhard, M., Schönwiese, C.-D. 1998: Simulation of global and hemispheric temperature variations and signal detection studies using neural networks. *Meteorol. Z.*, N.F., **7**, 171-180.



### **3 Zur Berücksichtigung der Persistenz in meteorologischen Zeitreihen** *Peter C. Werner*

#### **3.1 Einleitung**

Die mathematische Statistik setzt bei der Anwendung ihrer Methoden voraus, dass die Werte der betrachteten Datenkollektive voneinander unabhängig sind. Diese Voraussetzung ist in der Regel für meteorologische Zeitreihen nicht erfüllt. Um dennoch diese Daten statistisch auswerten zu können, sind bestimmte Bedingungen einzuhalten, die im weiteren behandelt werden sollen. Im Zentrum dieser Betrachtungen steht die Persistenz (oft auch als Erhaltungsneigung bezeichnet). Sie beschreibt die Abhängigkeit eines Ereignisses von zeitlich davor liegenden (bzw. räumlich benachbarten) Ereignissen. (Aus dem Lateinischen kommend bedeutet die Persistenz nichts anderes als das Verharren eines Zustandes.) Damit ist die Persistenz selbst ein statistischer Parameter zur Charakterisierung der Zeitreihe. Es werden deshalb im Kapitel 3.2 verschiedene Möglichkeiten ihrer Schätzung aufgeführt.

Da meteorologische Zeitreihen oftmals einen instationären Prozess widerspiegeln, sind in diesen Fällen die geschätzten statistischen Parameter nicht zeitunabhängig, was auch für die Persistenz gilt. Die zeitliche Variabilität der Persistenz kann damit u.a. der Aufdeckung von Klimaänderungen dienen. Hierfür werden Beispiele gezeigt.

Verallgemeinernd kann die Untersuchung der Persistenz methodisch in zwei Anwendungsgebiete eingeteilt werden:

##### 1. Statistik:

- Prüfung auf Unabhängigkeit der Werte in Stichproben
- Bestimmung der Redundanz
- Ermittlung der effektiven Werteanzahl für statistische Tests

##### 2. Prozessbeschreibung:

- Charakterisierung des Andauerverhaltens
- Vorhersagbarkeit
- Beschreibung charakteristischer Prozessabschnitte

Abschließend noch ein Wort zur Ursache für die Persistenz in meteorologischen Zeitreihen. In der Atmosphäre haben wir es mit Phänomenen zu tun, die jeweils charakteristische Raum- und Zeitskalen besitzen. Damit gehören Messwerte, die an einem Raumpunkt zu verschiedenen Zeitpunkten innerhalb der charakteristischen Zeitskala bzw. die zu einem Zeitpunkt an mehreren Raumpunkten innerhalb der charakteristischen Raumskala entnommen werden, zu der entsprechenden Erscheinung und sind deshalb in gewisser Weise ähnlich. Da sich aber in der Atmosphäre generell mehrere Phänomene überlagern, ist die Erhaltungsneigung in Zeitreihen oft verdeckt und muss daher mittels geeigneter statistischer Verfahren sichtbar gemacht werden.

## 3.2 Methoden zur Schätzung der Persistenz

Die Persistenz findet sich in unterschiedlichen statistischen Begriffen wieder. Typische Vertreter dafür sind: die Autokorrelation, die Erhaltungszahl, Andauer (mittlere Andauer, Andauerhäufigkeitsverteilung) und Sequenzlänge, die nachfolgend näher beschrieben werden.

### a) Autokorrelation

*Frage: Mit welcher Wahrscheinlichkeit hängt ein bestimmter Meßwert  $x(t_j)$  des Prozesses  $X(t)$  zum Zeitpunkt  $t_j$  davon ab, dass dieser Prozeß zum Zeitpunkt  $t_i$  (mit  $t_i < t_j$ ) den Wert  $x(t_i)$  gehabt hat?*

Dabei handelt es sich um die bedingte Wahrscheinlichkeit:  $W_{x(t_i)} x(t_j)$

$x(t_j)$  und  $x(t_i)$  sind in einer zweidimensionalen Verteilungsfunktion darstellbar:

Diese Funktion wird Autokovarianzfunktion genannt:

$$C(\tau) = E\{[x(t) - E(x)][x(t + \tau) - E(x)]\} \quad (3-1)$$

mit

$$\tau = t_j - t_i \quad (3-2)$$

Bei einem stationären Zufallsprozeß entspricht die Autokovarianzfunktion an der Stelle  $\tau = 0$  der Dispersion:

$$C(0) = E\{[x(t) - E(x)]^2\} = D^2(x) \quad (3-3)$$

Wird die Autokovarianzfunktion mittels der Dispersion normiert, erhält man die Autokorrelationsfunktion:

$$K(\tau) = C(\tau)/C(0) \quad (3-4)$$

Im allgemeinen Fall ist  $\tau$  kontinuierlich und damit sind auch die Autokovarianzfunktion und die Autokorrelationsfunktion kontinuierliche Funktionen.

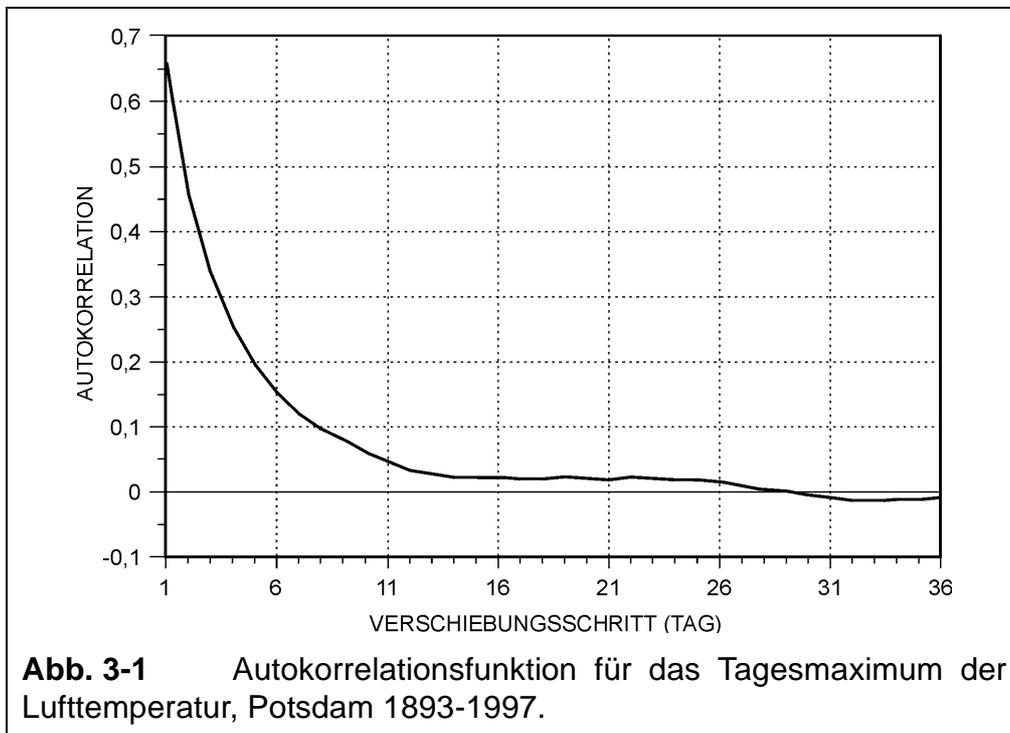
In der Praxis liegt aber das Beobachtungsmaterial in diskreter Form als Zahlenreihe von  $N$  äquidistanten Ablesungen mit dem Abstand  $\Delta t$  vor, so dass  $C(\tau)$  nur für diskrete Verschiebungsschritte  $\tau = r\Delta t$  ( $r = 0, 1, 2, \dots, m$ ) bestimmt werden kann:

$$C(r\Delta t) = C_r = 1/(n - r - 1) \sum_{j=1}^{n-r} (x_j - \mu)(x_{j+r} - \mu) \quad (3-5)$$

Für die Schätzung der Autokovarianzfunktion verwendet man in der Regel statt des Erwartungswertes  $\mu$  den Stichprobenmittelwert

$$C_r = 1/(n-r-1) \sum_{j=1}^{n-r} (x_j - \bar{x})(x_{j+r} - \bar{x}) \quad (3-6)$$

Um eine hinreichend gute Schätzung für  $C$  zu erhalten (es müssen genügend Werte in die Schätzung eingehen), sollte  $r$  nicht größer werden als 10% von  $n$ . Ein Beispiel ist in Abb. 3-1 dargestellt.



Streng genommen gelten die bisherigen Herleitungen zur Autokovarianz- bzw. Autokorrelationsfunktion und der Verzicht auf die Momente höherer Ordnung nur für ein bestimmtes Modell eines stochastischen Prozesses, das Modell des **stationären Gaußschen Zufallsprozesses**.

Die Voraussetzung für die Schätzung können aber etwas modifiziert werden:

- Die Verteilung für den Zufallsprozeß ist eine Gaußsche Normalverteilung. Wenn dies nicht der Fall ist, kann das eventuell durch eine Transformation der Werte erreicht werden.
- Können die Schwankungen der Werte oder der Anteil der statistischen Fluktuationen (Rauschen) an den Werten als ein stationärer Zufallsprozeß angesehen werden, dann ist auch das Modell einsetzbar.

Anmerkung: Langperiodische Schwankungen oder Trends im Vergleich zur betrachteten Zeitskala verfälschen das Ergebnis, d.h. die Schätzwerte der Autokorrelation fallen zu hoch aus. Das bedeutet, dass diese Variationen vor der Autokorrelations-schätzung aus der Datenreihe entfernt werden müssen. Zum Beispiel muss bei der Betrachtung von Tageswerten der Jahrgang eliminiert werden.

Von Interesse ist außerdem, für welche  $\tau$  der Wert der Autokorrelationsfunktion signifikant von Null verschieden ist. Dazu gibt SCHÖNWIESE (2000) ein Mutungsintervall  $M$  an:

$$M(\tau) = K(\tau) \pm q(\alpha)(1 - K^2(\tau)) / (n - 1)^{1/2} \quad (3-7)$$

mit  $q(\alpha)$  - Quantile der Normalverteilung, für z.B.  $\alpha = 0.05$  oder  $0.01$

Wenn  $M(\tau) \leq 0$ , dann ist an der Stelle  $\tau$  die Autokorrelationsfunktion  $K(\tau)$  nicht signifikant von Null verschieden.

### b) Erhaltungszahl nach BARTELS

Um die Persistenz in einer Zeitreihe anschaulich darstellen zu können, leitete BARTELS (1935) mit Hilfe der Autokorrelationsfunktion die Erhaltungszahl  $B$  ab. Sie gibt an, wieviel aufeinanderfolgende Werte im statistischen Sinne als voneinander abhängig angesehen werden können.

$$B = 1 + 2 \sum_{r=1}^{n-r-1} ((n-r)/n)K(r) \quad (3-8)$$

Dabei sollte das  $r$  nur bis zu der Stelle laufen, bei der  $K(r+1)$  das erste Mal nicht mehr signifikant von Null verschieden ist. Damit vermeidet man auch einen Vorzeichenwechsel in den Werten der Autokorrelationsfunktion, was ansonsten eine völlig falsche Schätzung von  $B$  ergeben würde.

### c) Andauer

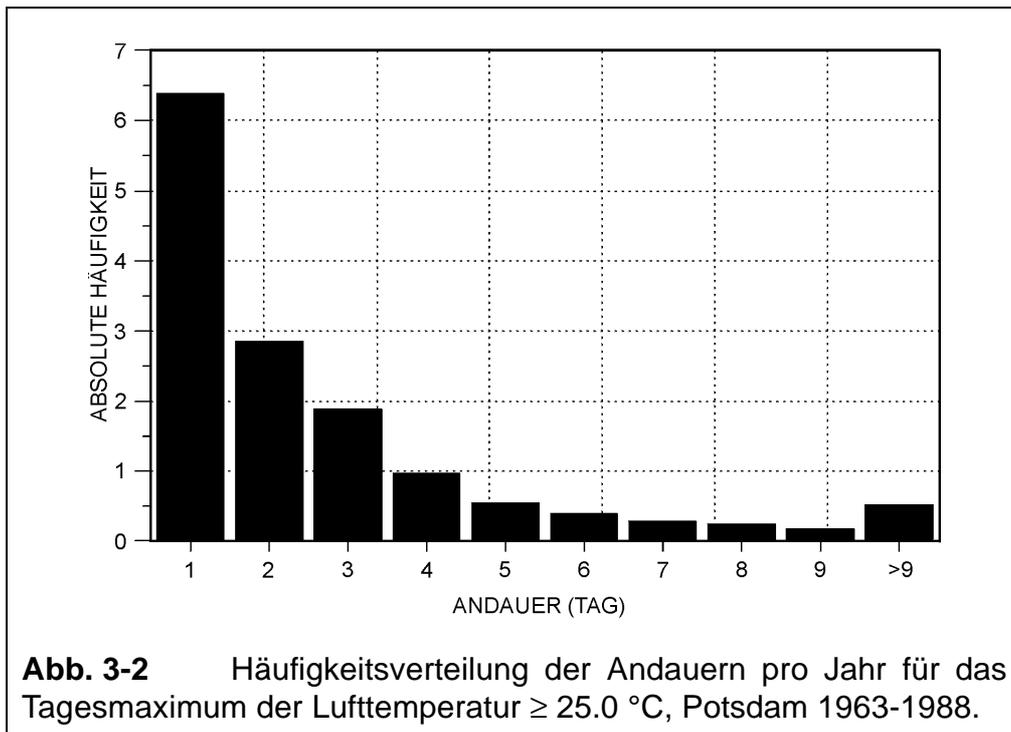
Mit dem Begriff Andauer beschreibt man, wie lange (wieviel Zeitschritte) ein Zustand in der Zeitreihe beibehalten wird. Häufig werden dabei die Längen von Schwellenwertüber- oder -unterschreitungen betrachtet, oder auch das Verharren zwischen zwei Grenzwerten. Dabei kann der Zustand durch mehrere Parameter beschrieben werden. Andere Definitionen des Zustands sind durchaus möglich (z.B. Phänomen vorhanden oder nicht vorhanden).

Bei der Bestimmung der Andauer geht man nun wie folgt vor:

- Aufstellen einer Zeitreihe mit den Werten 1 = Zustand eingetreten und 0 = Zustand nicht eingetreten
- Auszählen der Längen 1, 2, 3, usw. der Perioden mit dem Wert 1
- Aufstellung einer Häufigkeitsverteilung  $h_i$  zu den Längen  $D_i$  für  $k$  Klassen
- Schätzung von Parametern mittels der Häufigkeitsverteilung, z.B. Mittlere Andauer  $\bar{D}$ , minimale und maximale, Quantilwerte

$$\bar{D} = \frac{\sum_{i=1}^k h_i D_i}{\sum_{i=1}^k h_i} \quad (3-9)$$

Ein Ergebnis dieser Art der Andauerberechnung zeigt die Abb. 3-2.



#### d) Sequenzen

Mit der Sequenzanalyse (siehe auch WERNER et al., 1999) untersucht man die Abfolge charakteristischer Zustände eines Prozesses. Dafür werden entsprechende Zustände definiert und die Zeitreihe in Sequenzen (oder auch Tupel genannt) dieser Zustände zerlegt. Danach ist es möglich die Wahrscheinlichkeiten  $P$  der ausgewählten Sequenzen zu berechnen. (Enthalten die Sequenzen keine Zustandswechsel erhält man wieder Andauerwahrscheinlichkeiten.) Diese Wahrscheinlichkeiten  $P$  gehen in die Berechnung der dynamischen Entropie  $H_i$  ein

$$H_i = P(x_1, \dots, x_i) \ln(P(x_1, \dots, x_i))^{-1} \text{ mit } i = \text{Länge der Sequenz.} \quad (3-10)$$

Mittels der dynamischen Entropie kann bedingte Entropie  $h_i$ , die ein Maß für die Vorhersageunsicherheit ist, bestimmt werden

$$h_i = H_{i+1} - H_i \quad (3-11)$$

Der Vorteil der Sequenzanalyse ist, dass nicht nur die Erhaltungsneigung gleicher Zustände sondern auch Übergangswahrscheinlichkeiten betrachtet werden können. So kann beispielsweise die Frage beantwortet werden: Wie groß ist die Wahrschein-

lichkeit dafür, dass nach  $k$  Tagen mit dem Zustand A der Zustand B eintritt. Der Nachteil dieses Verfahrens ist, dass man für eine zuverlässige Schätzung eine große Datenmenge benötigt. Als Faustformel gilt:

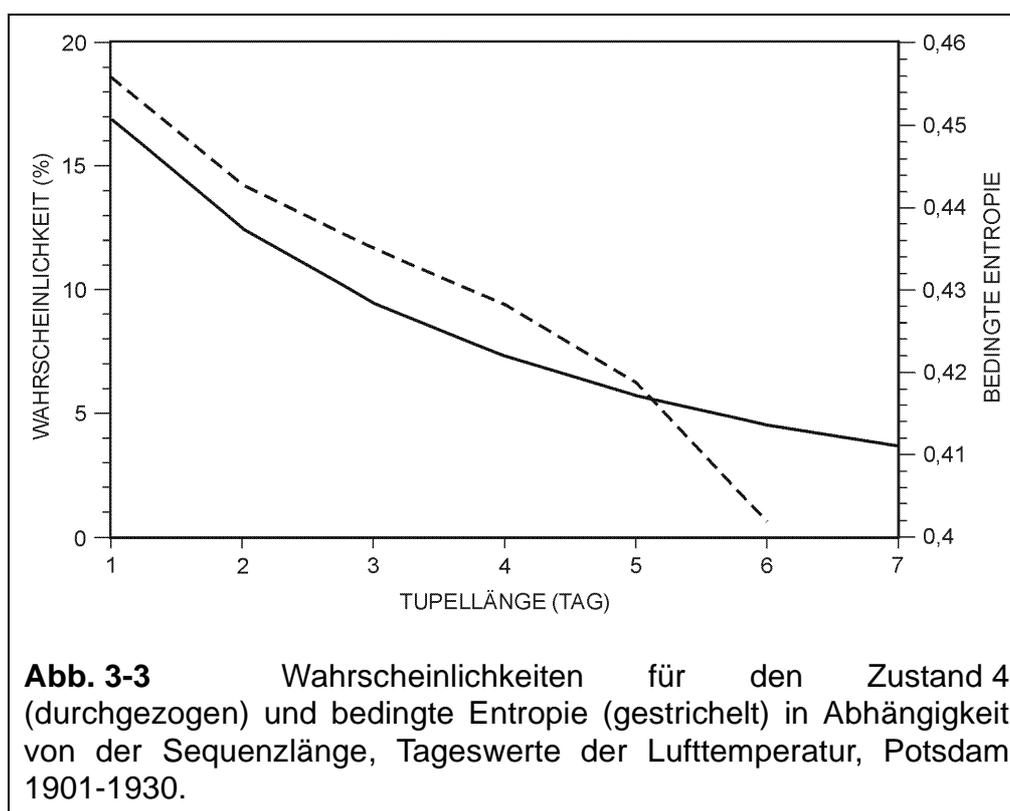
$$n = mz^i \quad (3-12)$$

mit  $n$  - benötigte Werteanzahl für die Schätzung  
 $m$  - minimal gewünschte Häufigkeit für eine der gewählten Sequenzen  
 $z$  - Anzahl der Zustände  
 $i$  - Länge der längsten gewählten Sequenz

Beispiel:  $m = 3$ ; 5 Zustände; längste Sequenz = 5 Tage

$$n = 3 \times 5^5 = 9375$$

Da es sich hier um Tageswerte handelt, wären Werte von mehr als 25 Jahren notwendig.



Für die Definition der Zustände (s. Abb. 3-3) wurden die Tageswerte der Lufttemperatur (Tagesmaximum, -minimum und -mittel) einer Clusteranalyse unterworfen und in 5 Cluster eingeteilt. Zwischen der Tupellänge 5 und 6 ist ein deutlicher "Knick" in der Kurve der bedingten Entropie zu sehen. Das weist darauf hin, dass ihre Schätzung nur bis zu einer Länge von 5 Tagen sinnvoll ist.

### 3.3 Berücksichtigung der Persistenz in statistischen Tests

Die statistischen Tests haben als Voraussetzung die Unabhängigkeit der Werte, die für die Schätzung des zu testenden Parameters verwendet wurden. Da die Werte in meteorologischen Zeitreihen diese Bedingung in der Regel nicht erfüllen, muss man diesen Umstand bei der Durchführung von Tests bzw. bei der Interpretation ihrer Ergebnisse berücksichtigen.

Methoden dafür:

1. Zufällige Auswahl einer Menge an Meßwerten aus der vorliegenden Stichprobe. Dabei muß die zufällig ausgewählte Menge deutlich kleiner als die der Stichprobe sein.  
Nachteil: Vergrößerung der statistischen Unsicherheit durch Verringerung der Werteanzahl.
2. Auswahl aller Werte mit dem Abstand  $n\Delta t$  voneinander. Der Abstand  $n\Delta t$  ist dabei definiert als der Wert, bei dem die Autokorrelationsfunktion (Autokovarianzfunktion) auf einen für das Problem statistisch nicht mehr signifikanten Wert abgesunken ist.  
Nachteil: Vergrößerung der statistischen Unsicherheit durch Verringerung der Werteanzahl.
3. Berechnung einer effektiven Werteanzahl  $N_{eff}$   
In die Berechnung der statistischen Parameter gehen alle Meßwerte ein. Deren Unabhängigkeit wird indirekt bei der Festlegung der Freiheitsgrade für die entsprechenden statistischen Tests durch die Vorgabe einer effektiven Werteanzahl berücksichtigt. Diese effektiven Werteanzahlen berechnen sich wie folgt (bei Tests für):

Mittelwert:

$$N_{effM} = \frac{n}{1 + 2 \sum_{r=1}^{r_0} ((n-r)/n)K(r)} \quad (3-13)$$

$K$  = Autokorrelationsfunktion

$n$  = Werteanzahl

$r_0$  = größte Schrittweite, bei der  $K$  noch nicht signifikant von Null verschieden ist

Streuung:

$$N_{effs} = \frac{n}{1 + 2 \sum_{r=1}^{r_0} ((n-r)/n)K^2(r)} \quad (3-14)$$

Korrelationskoeffizient:

$$N_{effr} = \frac{n}{1 + 2 \sum_{r=1}^{r_m} K_x(r)K_y(r)} \quad (3-15)$$

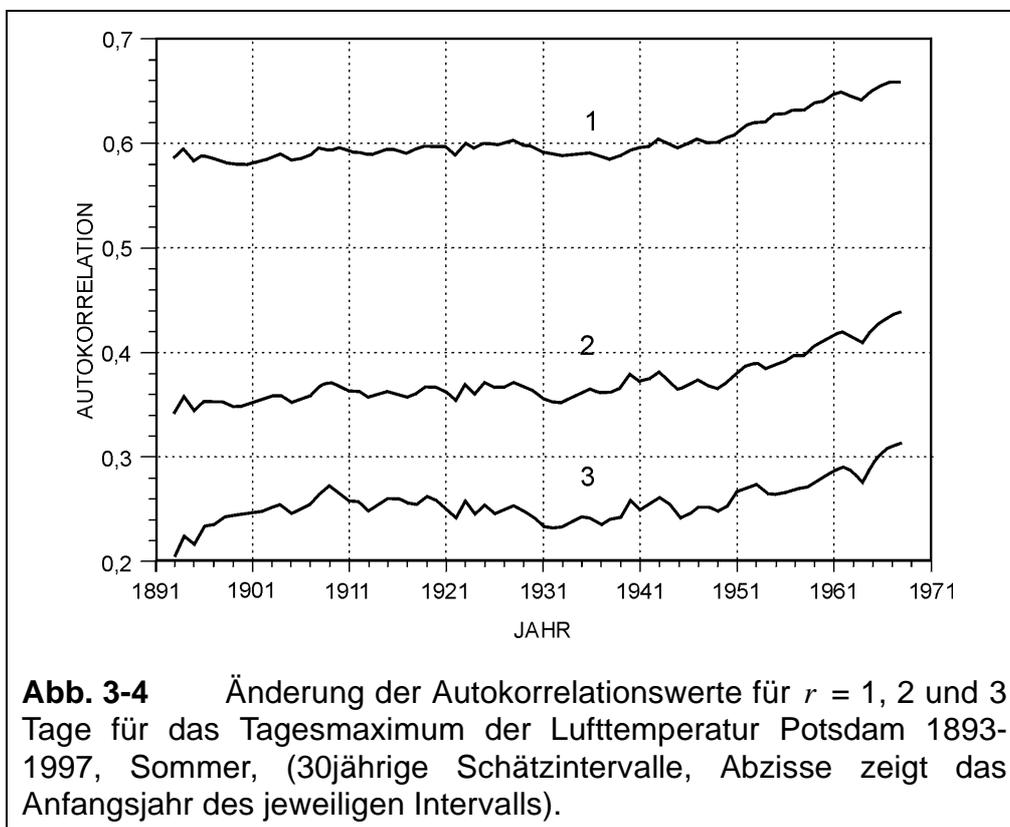
Die hier aufgeführten Beziehungen wurden von OLBERG (1972) und TAUBENHEIM (1974) abgeleitet.

### 3.4 Zeitliche Änderungen der Persistenz

In der Meteorologie haben wir es meist mit instationären Zeitreihen zu tun. Das heißt, dass auch die sie beschreibenden Parameter zeitabhängig sind. Nehmen wir wenigstens für Teilabschnitte der Zeitreihe eine Quasi-stationarität an, so können für diese Abschnitte Parameterschätzungen vorgenommen werden. Wendet man die gleitende Fenstertechnik (Fenster = Teilabschnitt) an, kann man die zeitliche Änderung des geschätzten Parameters beschreiben, wie z.B. die der Persistenz. Im folgenden sollen dafür einige Beispiele gezeigt werden.

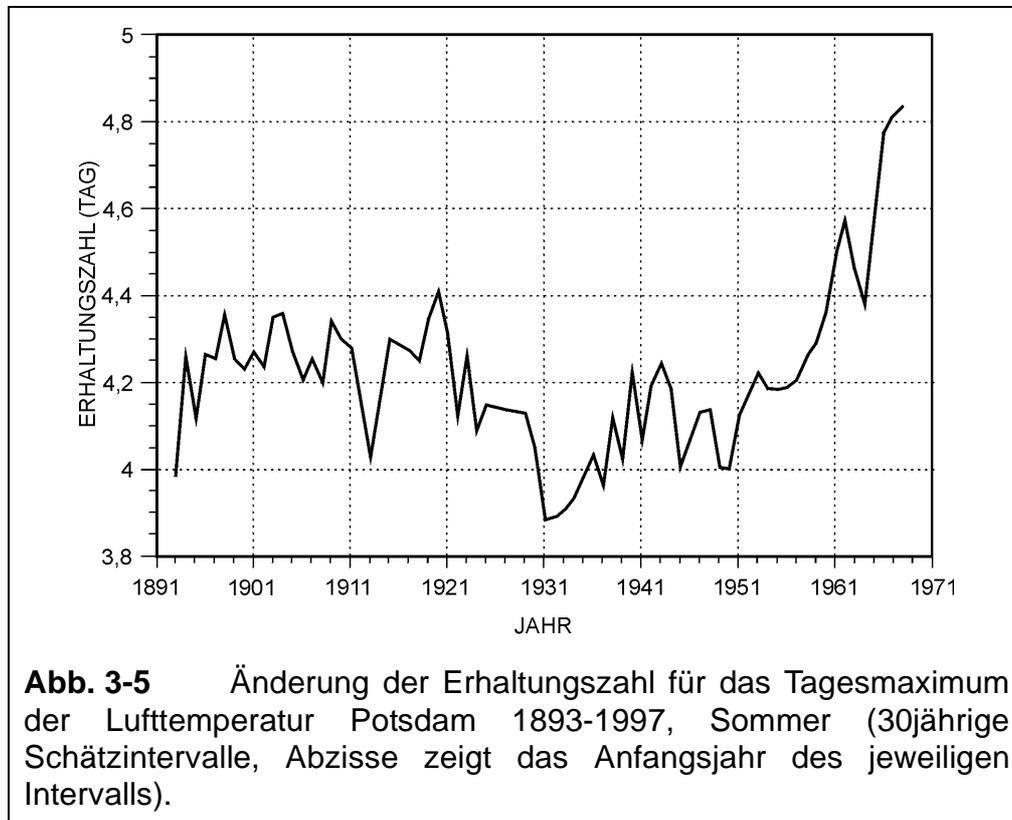
#### Autokorrelation

In Abb. 3-4 ist unschwer zu erkennen, dass die Erhaltungsneigung im Laufe der Zeit zunimmt, was ganz besonders für die letzten Abschnitte des Gesamtzeitraumes zutrifft.



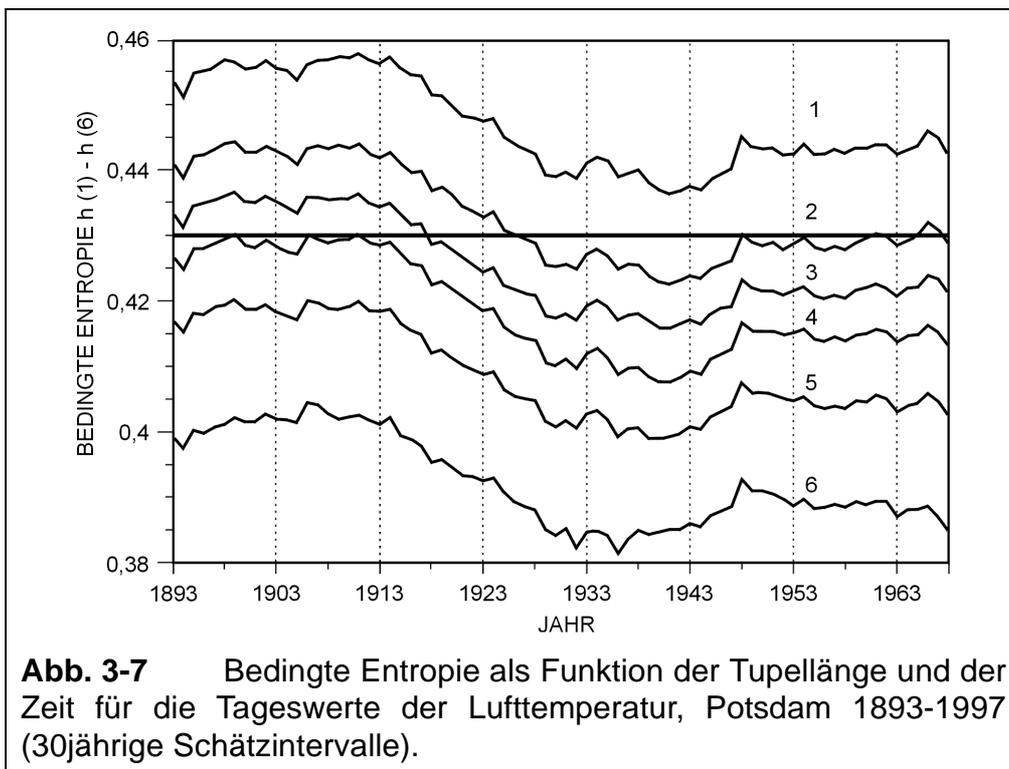
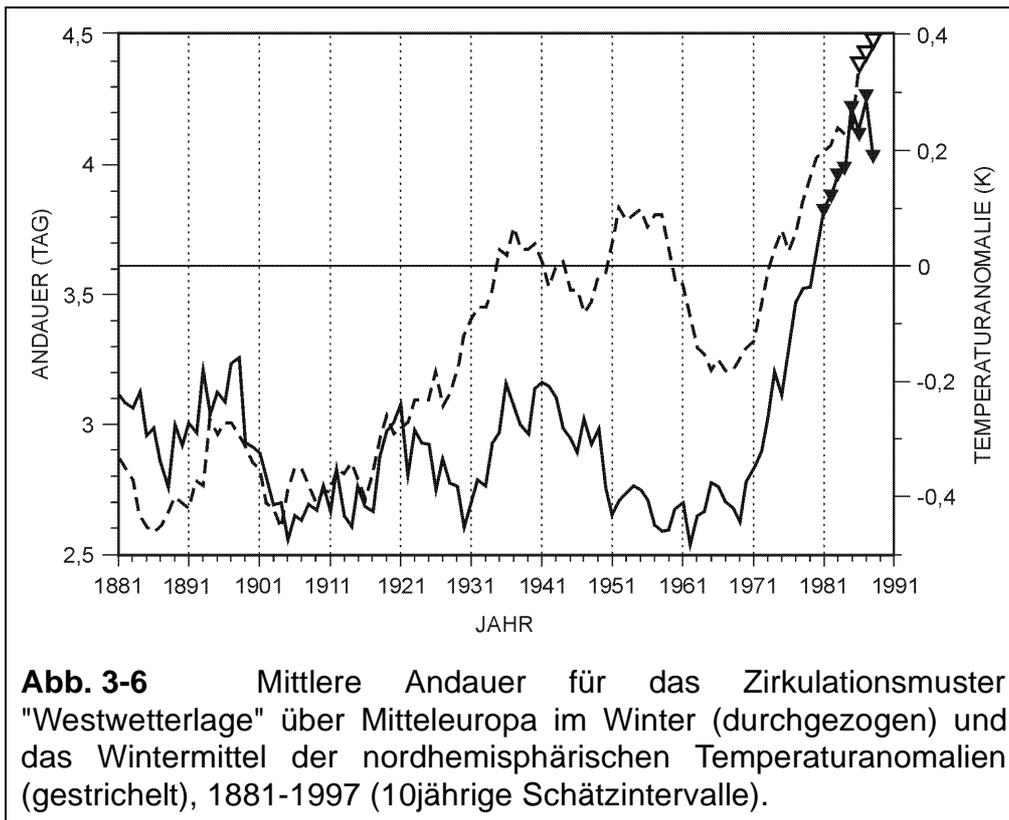
## Erhaltungszahl nach BARTELS

Bei diesem Parameter tritt noch deutlicher als bei den ersten 3 Werten der Autokorrelationsfunktion zu Tage die starke Zunahme der Erhaltungszahl am Ende des Zeitraums (siehe Abb. 3-5).

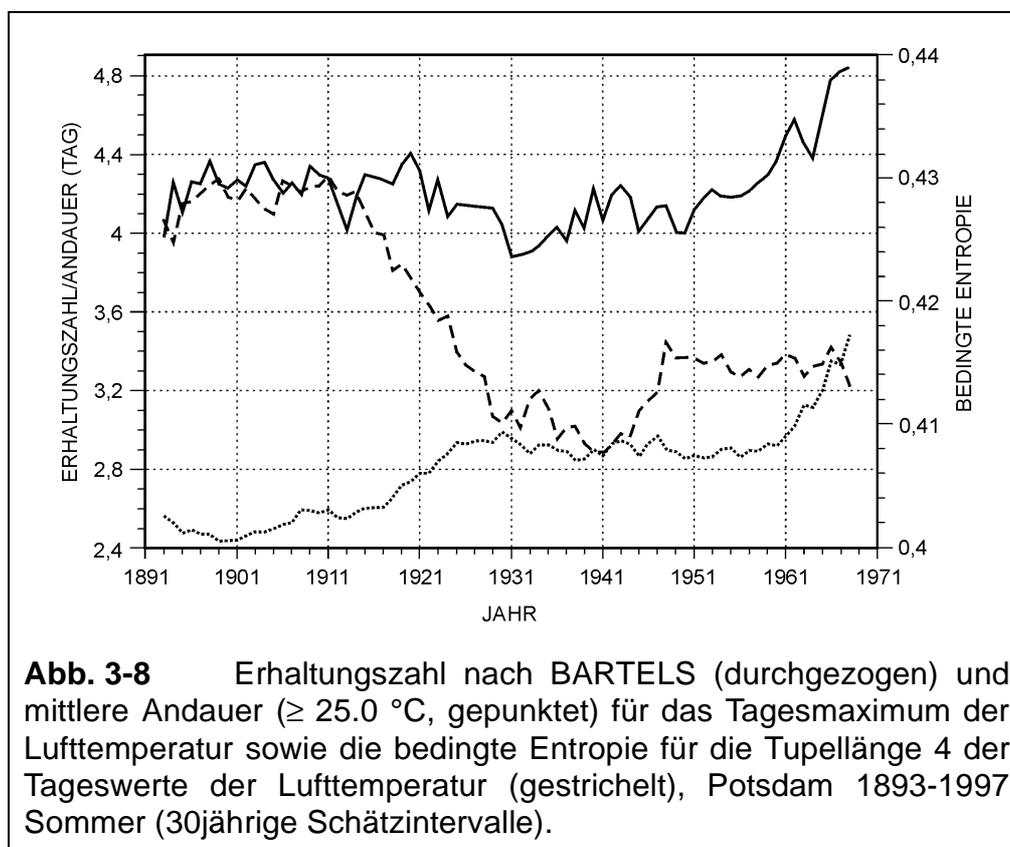


Mit Hilfe der Andauerbetrachtung konnte in diesem Fall ein Zusammenhang zwischen der globalen Erwärmung und der Zirkulationsänderung in den letzten Dekaden des 20. Jahrhunderts nachgewiesen werden (Abb. 3-6).

Man erkennt in Abb. 3-7, dass die Vorhersagesicherheit bis in die 30/40er Jahre deutlich zunimmt, um sich danach wieder etwas zu verschlechtern. (Bestimmung der Zustände siehe Kapitel 3.2 d.) Beispiel (waagerechte Linie): Die Unsicherheit, dass nach 4 Tagen eines Zustandes dieser auch noch am 5. Tag eintritt, beträgt am Anfang des Untersuchungszeitraumes ca. 43%. Am Ende des Zeitraums hat dieser Wert bereits die Tupellänge 2 erreicht, während die Unsicherheit für die Tupellänge 4 deutlich unter 42% liegt. Das heisst, dass die Vorhersagesicherheit (1 - bedingte Entropie) zugenommen hat.



Insgesamt zeigen die Beispiele, dass die Persistenz für eine ganze Reihe von atmosphärischen Prozessen im 20. Jahrhundert angewachsen ist. Abb. 3-8 dokumentiert diese Feststellung noch einmal zusammenfassend.



### 3.5 Zusammenfassung

- Es gibt verschiedene Möglichkeiten, die Persistenz zu beschreiben bzw. zu schätzen und sie mit Maßzahlen zu belegen.
- Die Persistenz muss bei der Anwendung statistischer Tests berücksichtigt werden.
- In vielen meteorologischen Zeitreihen ist die Persistenz nicht zeitunabhängig (Instationarität). Sie kann deshalb der Aufdeckung von Klimaänderungssignalen dienen.

### 3.6 Literatur

- Bartels, J. 1935: Zur Morphologie geophysikalischer Zeitfunktionen. *S.-B. Preuß. Akad. Wiss.*, **30**, 504-522.
- Olberg, M. 1972: Bemerkungen zur Verwendung der effektiven Anzahl unabhängiger Stichprobenwerte bei statistischen Testverfahren in der Meteorologie und Geophysik. *Gerl. Beitr. Geophys.*, **83**, 2/3, 121-128.
- Schönwiese, C. D. 2000: *Praktische Statistik für Meteorologen und Geowissenschaftler*. 3. Auflage, Bebr. Borntraeger, Berlin, Stuttgart, 298 S.

- Taubenheim, J. 1974: Zur Berücksichtigung der Autokorrelation bei statistischen Tests von Durchschnitten, Streuungen und überlagerten Stichtagen. *Gerl. Beitr. Geophys.*, **81**, 1/2, 57-64.
- Werner, P. C., Gerstengarbe, F.-W., Ebeling, W. 1999: Changes in Probability of Sequences, Exit Time Distribution and Dynamical Entropy in the Potsdam Temperature Record. *Theor. Appl. Climatol.*, **62**, 3-4, 125-132.

## 4 Robuste statistische Verfahren zur Auswertung mangelbehafteter Daten *Jens Taubenheim*

### 4.1 Mängelbehaftete Daten

Die Mehrzahl statistischer Verfahren, die in Handbüchern und Softwarepaketen angeboten werden, geht von der stillschweigenden Voraussetzung aus, dass alle in die Auswertungsprozedur eingehenden Quelldaten in gleicher Weise und Qualität gewonnen und erfaßt worden sind. Das verlangt nicht nur, dass für jede Beobachtung ein eindeutiger Zahlenwert als Meßwert vorliegt, sondern vor allem, dass die von den "zufälligen" Störeinflüssen herrührende Streuung der Meßwerte *homoszedastisch* ist, d.h. für die stochastische Streuung (griechisch  $\sigma\kappa\epsilon\delta\alpha\sigma\iota\varsigma$ ) *aller* Beobachtungen *eine einheitliche Fehlerverteilungsfunktion* gilt.

Diese Voraussetzungen sind zweifellos erfüllbar bei der Durchführung und Auswertung planmäßiger Labor- und Feldversuche. Bei Daten von Naturobjekten und -prozessen, insbesondere historischer Art, wie sie in der Klimaforschung und verwandten Gebieten auszuwerten sind, können diese Voraussetzungen aber keineswegs immer gewährleistet werden, sondern mehr oder weniger häufig *mängelbehaftete Daten* folgender Art auftreten:

- **Ausreißer**, drastisch von der Mehrheit der anderen Beobachtungsdaten abweichende Meßwerte, verursacht durch Fehlfunktionen des Meßgeräts oder des Systems der Datenaufzeichnung bzw. -übermittlung (z.B. falsche oder fehlende Dezimalpunkte), unter Umständen aber auch durch zeitlich begrenzte externe Einwirkungen auf den Meßvorgang (Unwetter, Erdbeben u.ä.);
- **Überschreitungen** der Meßbereichsgrenzen, so dass das Beobachtungsergebnis durch einen Zahlenwert mit vorgesetztem Symbol  $>$  oder  $<$  qualifiziert werden muß;
- **Hetero-Szedastizität** im allgemeinen, wenn das Datenmaterial aus verschiedenen Quellen mit unterschiedlicher Fehlerstreuung stammt.

Es liegt auf der Hand, dass beliebig fehlerhafte, sogar irreführende Aussagen resultieren können, wenn solche mangelbehafteten Daten unbeachtet in eine der statistischen Standard-Prozeduren (insbesondere bei automatischer Verarbeitung) einbezogen werden. Für die schwere Auswirkung, die bereits einige wenige Ausreißer auf die Ermittlung eines Trends aus einer (oft relativ kurzen) Beobachtungszeitreihe haben können, sind Beispiele häufig zu finden bzw. leicht zu simulieren.

Es darf nicht unerwähnt bleiben, dass die statistischen Standard-Auswertungsverfahren auch noch eine andere stillschweigende Voraussetzung enthalten, deren Nichtbeachtung zu irreführenden Interpretationen Anlaß geben kann: Die Voraussetzung der Unabhängigkeit (Abwesenheit von Autokorrelation) in Datenzeitreihen. Dieses Problem liegt jedoch außerhalb der im Folgenden zu behandelnden Thematik.

## 4.2 Robuste statistische Auswertungsverfahren

Der triviale Weg, mangelbehaftete oder -verdächtige Daten einfach von vornherein aus der statistischen Prozedur auszuschneiden, stellt keine annehmbare Lösung des Problems dar, weil dies eine vorherige Begutachtung (visuelle Inspektion oder automatisches Screening) der Daten voraussetzt, die ein subjektives Entscheidungskriterium in die Auswertung hineinträgt. Das kann bei offenkundigen drastischen Ausreißern unbedenklich sein, nicht aber bei generell stark streuenden Meßdaten, wo eine subjektive Eliminierung von Meßwerten, die vielleicht nur scheinbar 'Ausreißer' sind, den Verlust wichtiger Information bedeuten kann. Zweifellos ganz unzulässig ist es, Beobachtungswerte mit <- oder >-Symbol zu ignorieren, d.h. als "nicht gemessen" zu betrachten. (Bekanntlich verzögerte ein aus vorgefaßten Erwartungen festgelegter Meßbereich seinerzeit die Entdeckung des antarktischen Ozonloch-Effekts!). Hierbei ist es auch kein befriedigender Ausweg, lediglich das Symbol "<" oder ">" zu ignorieren, den dazugehörigen Zahlenwert aber in die Statistik einzubeziehen.

Aus diesen Gründen ist es empfehlenswert, zur Auswertung von Beobachtungsdaten (hinsichtlich Lokalisierung, Korrelationen, Trend u. dgl.) statistische Verfahren und Kenngrößen anzuwenden, die *resistent* gegenüber dem verfälschenden Einfluß mangelbehafteter Daten sind und es somit erlauben, *alle* vorhandenen Daten ohne vorherige subjektive Zensur in die Auswertung einzubeziehen. Die Aufgabe ist, aus einem vorliegenden Datensatz von  $N$  Beobachtungen (Meßwerten, Wertepaaren oder Werte- $n$ -tupeln) solche statistischen Kenngrößen abzuleiten, die die Eigenschaft haben, unverändert und interpretierbar zu bleiben, wenn ein Teil der ursprünglichen ("unverfälschten") Datenwerte durch Mängel der oben beschriebenen Art entstellt (engl. *contaminated*) wird. Statistische Verfahren und Kenngrößen, die das leisten, nennt man *robust*.

Ansätze zur systematischen Entwicklung robuster Verfahren, vor allem in den USA und in der Schweiz, reichen schon mehr als 30 Jahre zurück (HUBER, 1964; HAMPPEL, 1971), und heute gibt es bereits ein umfangreiches Angebot an Handbüchern und Sammelwerken, wovon das unten angefügte Literaturverzeichnis nur einen kleinen Ausschnitt erfassen kann.

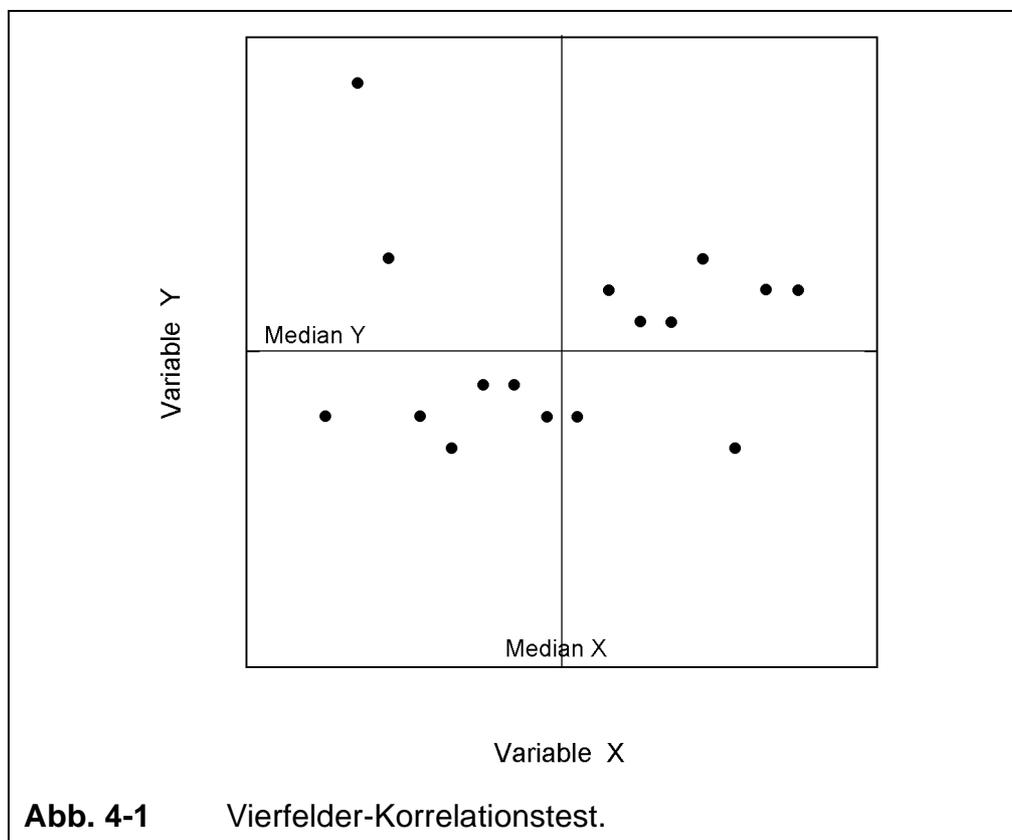
Als Kriterium der Robustheit einer statistischen Kenngröße dient die *Breakdown-Zahl*  $\varepsilon$ . Bei einem Datensatz von  $N$  Beobachtungen gibt sie an, dass die robuste statistische Kenngröße ihren Wert unverändert behält, so lange die Anzahl von mangelbehafteten Daten kleiner ist als  $\varepsilon \cdot N$ . Zweifellos haben daher robuste Kenngrößen Breakdown-Werte, die im Bereich  $0 < \varepsilon \leq \frac{1}{2}$  liegen. Demgegenüber haben die nicht-robusten Kenngrößen bei statistischen Standard-Verfahren (z. B. der arithmetische Mittelwert) im allgemeinen die Breakdown-Zahl  $\varepsilon = 0$ . In der Literatur findet man dementsprechend auch die Bezeichnungen "*outlier-resistant*" und "*high-breakdown*" als Synonyme für *robust*.

Als exemplarisches Beispiel einer robusten statistischen Kenngröße zur Lokalisierung der "Mitte" eines Datensatzes erweist sich der altbekannte *Median*, dessen Breakdown-Zahl  $\varepsilon = \frac{1}{2}$  beträgt. Seine Robustheit gegenüber den oben beschriebenen Kategorien mangelbehafteter Daten ist augenfällig, so dass sie keiner weiteren Erläuterung bedarf. Die Aussagekraft des Medians läßt sich durch Tests beurteilen

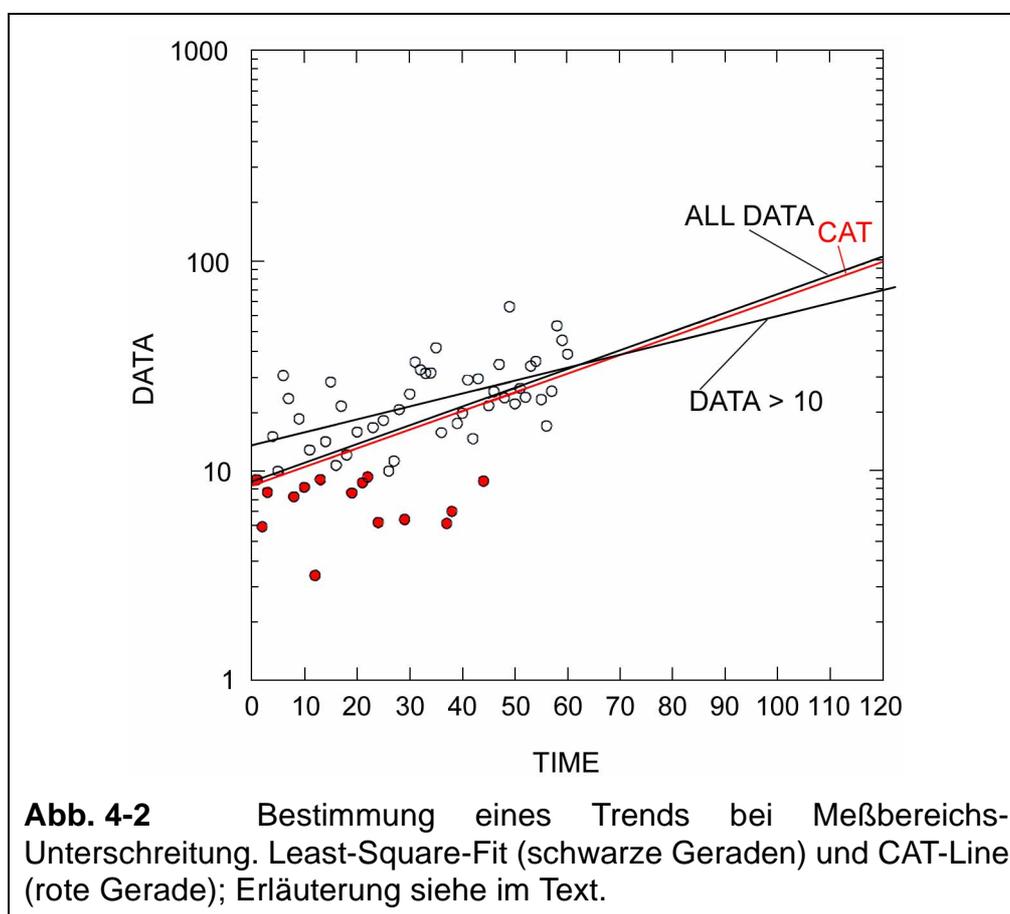
(beispielsweise den Zeichentest, siehe z.B. TAUBENHEIM, 1969), wohingegen der "klassische" arithmetische Mittelwert (Breakdown  $\varepsilon = 0$ ) schon durch einen einzigen Ausreißer verändert und in seiner Brauchbarkeit für Standard-Signifikanztests beschädigt wird.

### 4.3 Robuste Verfahren zur Korrelations- und Regressionsanalyse

Da die Ermittlung von Korrelations- und Regressionsbeziehungen besonders empfindlich durch mangelbehaftete Daten verfälscht oder behindert werden kann, hat man frühzeitig versucht, die robusten Eigenschaften des Medians hierfür zu nutzen. Einfachstes Beispiel ist die *Vierfelder- oder Quadranten-Korrelation* (s. TAUBENHEIM, 1969, p. 181ff), die in Abb. 4-1 kurz verdeutlicht wird: Das bloße Auge erkennt in den 16 Datenpunkten  $y_i(x_i)$  bereits eine mit wachsendem  $x$  ansteigende Tendenz, aber einige Ausreißer führen dazu, dass die übliche Korrelationsrechnung auf einen Korrelationskoeffizienten führt, dessen Wert genau Null ist. Teilt man jedoch durch den Median der  $x_i$  und den Median der  $y_i$  das Korrelogramm in vier Quadranten auf, dann entsteht eine gewöhnliche 2x2-Kontingenztabelle, wofür in diesem Fall definitionsgemäß die unter der Nullhypothese ("keine Korrelation") zu erwartenden Besetzungszahlen aller vier Felder gleich sein müssen, und zwar  $q_0 = N/4$ . Führt man nun mit den tatsächlichen Besetzungszahlen  $q$  den ganz gewöhnlichen  $\chi^2$ -Test (Freiheitsgrad 1) aus, dann erhält man  $\chi^2 = \sum(q - q_0)^2 / q_0 > 3,84$ , was klar besagt, dass die Nullhypothese auf dem 5%-Signifikanzniveau verworfen werden kann. Dieser Quadrantenkorrelationstest ist also ein robustes Verfahren, mit dem Breakdownwert  $\varepsilon = 1/2$ .



Nicht ganz so einfach ist es, ein robustes Verfahren der *Regression* anzugeben, d.h. eine Gerade bzw. eine Kurve vorgewählter nichtlinearer Form an einen Datensatz, der eine "Punktwolke" im zwei- oder mehrdimensionalen Raum bildet, in bestmöglicher Weise anzupassen, ohne durch mangelbehaftete Daten verfälscht zu sein. Die klassische GAUSSsche Least-Square-Anpassung (LS), bei der die Summe der quadrierten Abstände von der Regressionskurve (Residuen) minimiert wird, kann nicht beide Forderungen gleichzeitig erfüllen, weil einerseits die Residuen von Ausreißerwerten infolge ihrer Quadrierung mit zu hohem Gewicht eingehen und damit das Ergebnis dominieren, und weil andererseits bei Meßbereichsüberschreitungen die nicht zahlenmäßig angebbaren Residuen gänzlich ignoriert werden müssen. Abb. 4-2 veranschaulicht, wie sich eine Meßbereichs-Unterschreitung auf die Extrapolation eines (logarithmischen) Trends auswirkt: Die schwarze Gerade "All Data" ist die Least-Square-Regressionsgerade aus allen tatsächlichen Datenwerten, dagegen ergibt sich die Gerade "Data > 10", wenn alle Meßwerte, die < 10 sind (rot markierte Punkte), als "nicht gemessen" aus der LS-Trendberechnung ausgeschlossen werden. Auf die rote Gerade "CAT" werden wir weiter unten zurückkommen.



Um das Verfahren der Regression bzw. Kurvenanpassung robust zu machen, besteht ein naheliegender Weg darin, an Stelle der *Summe* der quadrierten Residuen deren *Median* zu minimieren, weil im Median die meßbereichsüberschreitenden Beobachtungen voll berücksichtigt werden können, und Ausreißer nicht mehr das Ergebnis dominieren. Dieses robuste *Least-Median-of-Squares-(LMS-)*Verfahren, wie es von ROUSSEUW und LEROY (1987) eingehend behandelt worden ist, eignet sich für die einfache Regression  $Y_i(x_i)$  ebenso wie für mehrfache Regression

$Y_i, (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})$ , erfordert aber relativ hohen Rechenaufwand: Zur einfachen linearen Regression in einem Datensatz von  $N$  Wertepaaren  $(y_i, x_i)$  ("Punktwolke") zieht man durch jeweils zwei dieser Punkte genau eine "Test"-Gerade  $Y(x)$ , mißt dann die Residuen (Abstände)  $\delta = (y - Y)$  der übrigen  $N - 2$  Punkte von dieser Geraden, und bestimmt den Medianwert der quadrierten Residuen  $\delta^2$ . (Als mögliche Variante des Verfahrens kann man an Stelle der Quadrate auch die Absolutbeträge  $|\delta|$  benutzen). Dies wiederholt man mit möglichst vielen Testgeraden, die man auf solche Weise durch die Punktwolke legen kann (maximal sind  $N(N - 1)/2$  Testgeraden möglich) und adoptiert dann die Testgerade mit dem kleinsten  $\delta^2$ -Median als die gewünschte *robuste Regressionsgerade*.

Zur Mehrfachregression mit  $p$  unabhängigen Variablen werden aus den  $N$  Beobachtungsdaten  $(y_i, x_i^{(1)}, \dots, x_i^{(p)})$ ,  $(i = 1, 2, \dots, N)$ , "elementare Unter-Datensätze" (*elementary subsets*) gebildet, die aus jeweils  $(p + 1)$  Beobachtungsdaten bestehen. Durch jedes solche Subset ist wiederum jeweils genau eine Testgerade im  $(p + 1)$ -dimensionalen Raum festgelegt, zu der man die Residuen  $\delta$  aller übrigen  $(N - p - 1)$  Beobachtungsdaten bestimmt. Die Testgerade mit dem kleinsten  $\delta^2$ -Medianwert liefert die gewünschte robuste (Mehrfach-)Regressionsgerade.

Da die Anzahl der möglichen Punktepaare bzw. elementaren Subsets durch die Binomialkoeffizienten  $N$  über  $(p + 1)$  (Symbol:  $C_N^{p+1}$ ) gegeben wird, würde beim LMS-Verfahren die erforderliche Menge algebraischer Operationen einen nicht mehr zu bewältigenden Umfang annehmen, wenn tatsächlich sämtliche Test-Subsets berechnet werden müßten. ROUSSEUW und LEROY (1987) zeigen jedoch, dass das gar nicht nötig ist, denn es würde für das Auffinden der LMS-Regressionsgerade schon ausreichen, dass sich unter den gebildeten Test-Subsets nur hinreichend viele befinden, die "gut", d.h. nicht durch Datenmängel kontaminiert, sind. Man kann mit Hilfe der Bernoulli'schen Formel leicht die Wahrscheinlichkeit dafür berechnen, dass unter  $m$  Subsets (aus jeweils  $p + 1$  Daten) *mindestens ein Subset "gut"* ist. Sie beträgt

$$W\{\text{mindestens 1 von } m \text{ Subsets ist "gut"}\} = 1 - [1 - (1 - \mu)^{p+1}]^m.$$

Hierin ist  $\mu$  der Anteil ( $0 < \mu < 1/2$ ) der mangelbehafteten Daten an der Gesamtzahl der  $N$  Beobachtungen. Für die Forderung  $W = 0.95$  geben ROUSSEUW und LEROY (1987, p. 199) eine Tabelle der "ausreichenden Anzahl  $m_{suff}$  von Subsets" für  $p$ -fache Regression an, wovon hier folgender Auszug einen Eindruck vermitteln soll:

$p$	1	2	3	4	$\geq 5$
$m_{suff}$	1000	1500	2000	2500	3000

(sofern nicht  $C_N^{p+1}$  selbst schon kleiner als diese Zahlen ist).

Wichtig ist, dass das robuste LMS-Verfahren nicht nur für lineare, sondern auch für nichtlineare Regression benutzbar ist. Neuerdings ist sogar gezeigt worden, dass es unter bestimmten Bedingungen auch zu einer robusten Frequenzschätzung (LMS-Anpassung sinusoidaler Kurven) verallgemeinert werden kann (SMYTH und HAWKINS, 2000).

#### 4.4 Neue Entwicklungen durch das Paradigma "Datentiefe"

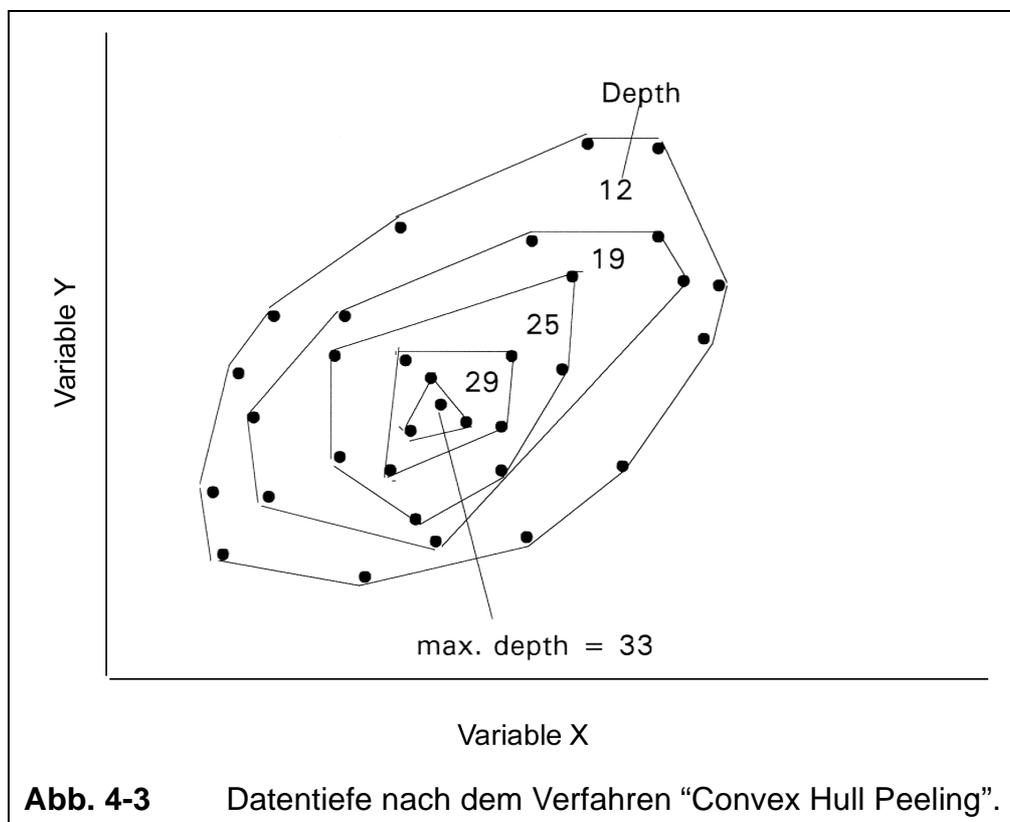
Das LMS-Verfahren hat das Regressionsproblem robuster gemacht durch die Verwendung des Medians an Stelle des Mittelwerts der Residuenquadrate, bleibt aber, neben dem Nachteil einer abnorm langsamen Konvergenzrate (und damit einer schwachen statistischen Effizienz) bei relativ hohem Rechenaufwand, begrifflich unbefriedigend wegen des Zwittercharakters zwischen Ordnungs-Statistik (Median) und metrischer Statistik (Residuenberechnung).

Zu einer begrifflich konsequenteren Behandlung zeichnen sich neuerdings Perspektiven ab, die an den Begriff der *Datentiefe* (*data depth*) anknüpfen, der bereits 1975 von TUKEY eingeführt wurde (ausführlich behandelt in einem Review-Artikel von LIU et al., 1999):

In einem der Größe nach geordneten Datensatz von  $N$  Meßwerten,  $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N$ , lokalisiert man einen Zahlenwert  $x$  durch seine *Tiefe* (*location depth*, "*ldepth*"), die definiert wird als die natürliche Zahl

$$ldepth(x) = \inf \{ \#(x_i < x), \#(x_i > x) \},$$

wobei das Symbol  $\#$  "Anzahl der Daten" bedeutet, und  $\inf\{a, b\}$  das Kurzsymbol ist für "die kleinere der zwei Zahlen  $a$  und  $b$ ". Man erkennt unmittelbar, dass der Median des Datensatzes die maximale Tiefe repräsentiert, nämlich  $ldepth_{max} = N/2$ .



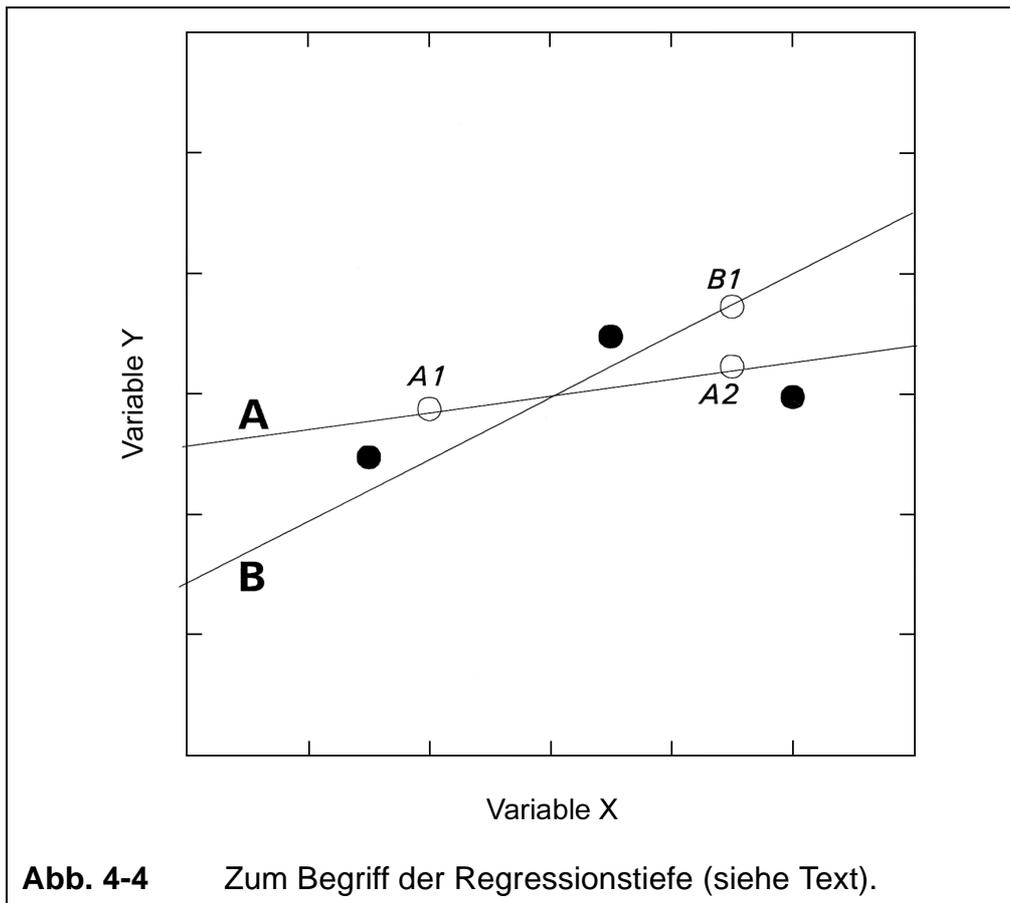
Diese (robuste!) Lokalisierung kann auf mehrdimensionale Datensätze verallgemeinert werden, die eine Punktwolke im 2-, 3-, ... ,  $p$ -dimensionalen Raum bilden. Eine einfache Vorgehensweise dazu ist die sog. *Convex Hull Peeling Depth* (BARNETT, 1976), die in Abb. 4-3 für den 2-dimensionalen Fall dargestellt ist. Man umschließt die ganze Punktwolke mit einer geschlossenen Hüllkurve, von der lediglich verlangt wird, dass sie überall konvex ist (hier ein die äußersten Datenpunkte berührender Polygonzug, dessen Innenwinkel alle  $\leq 180^\circ$  sein müssen) und definiert damit den "äußeren" Halbraum, dem man die Tiefe Null zuschreibt. Fortschreitend werden um die Datenpunkte im "inneren" Halbraum weitere solche konvexen Hüllkonturen "eingenestet", so dass eine zwiebelschalenartige Struktur entsteht. Die *location depth*, die jeweils der ganzen "Zwiebelschale" zwischen zwei Hüllkonturen zugeschrieben wird, ist die Anzahl der durch die "obere" der beiden Konturen berührten Datenpunkte plus aller "außerhalb" davon liegenden Punkte. So nimmt die Tiefe schrittweise nach innen zu und erreicht ihr Maximum innerhalb der innersten Hüllkontur.

Für 3- und mehrdimensionale Datensätze ist die Vorgehensweise (mit konvexen Hüll- und Hyperflächen) völlig analog. Zu anderen Varianten der Konstruktion der *location depth* siehe den Reviewartikel von LIU et al. (1999).

#### 4.5 Die Regressionstiefe nach ROUSSEEUW und HUBERT (1999)

Vor wenigen Jahren ist es nun gelungen, das Datentiefe-Paradigma auch auf das Problem der robusten Regression zu übertragen. Das Ziel des Verfahrens besteht darin, diejenige Gerade (oder andere vorgewählte Kurve) zu finden, die in die gegebenen Daten "optimal und robust eingebettet" ist, analog wie das Maximum der *Location depth* denjenigen Punkt anzeigt, der "optimal und robust eingebettet" ist. Diese "Einbettungstiefe" ist ja interpretierbar als die Datenmenge, die "beseitigt" werden muß, um einen Punkt aus dem (inneren) Halbraum endlicher Tiefe in den (äußeren) Halbraum der Tiefe Null zu schaffen (Anschauliches Bild: Weg einer Obstmade aus dem Kern eines Apfels ins Freie!). Die entscheidende Idee für eine zur *Location depth* analoge *Regression depth* besteht in folgender Definition: Bei einem (zweidimensionalen) Datensatz  $y_i(x_i)$  wird als Regressionstiefe einer Geraden  $Y(x)$  die Anzahl derjenigen Datenpunkte bezeichnet, die "beseitigt" werden müßten, um diese Gerade durch Rotation um einen ihrer Punkte in eine vertikale Position (d.h. parallel zur y-Achse) schwenken zu können.

Abb. 4-4 veranschaulicht diese Definition an einem Datensatz von drei Meßwerten (schwarz gefüllte Punkte): Die Gerade "B" hat die Tiefe Null (ROUSSEEUW UND HUBERT nennen dies einen "Nonfit"), denn man kann sie um den Punkt B1 (als Pol) entgegen dem Uhrzeigersinn in eine vertikale Position schwenken, ohne einen der drei Datenpunkte wegnehmen zu müssen. Dagegen hat die Gerade "A" die Tiefe 1, denn sie kann in eine vertikale Position nur dann gebracht werden, wenn entweder der rechte Datenpunkt beseitigt wird, um sie im Uhrzeigersinn um den Drehpol A1 zu schwenken, oder der linke Datenpunkt beseitigt wird, um sie entgegen dem Uhrzeigersinn um den Pol A2 schwenken zu können. Die Regressionstiefe *rdepth* ist also stets eine natürliche Zahl.



Ausgehend von dem von ROUSSEEUW und HUBERT definierten Begriff der Regressionstiefe ist nun ein Konzept der robusten Regression möglich, das der robusten Lokalisierung mittels Datentiefe adäquat ist: Man kann in einem iterativen Prozeß bei jeder gegebenen Punktwolke diejenige Gerade auffinden, die sich in *maximaler Regressionstiefe* befindet, was als die bestmögliche robuste Anpassung einer Geraden an die Punktwolke anzusehen ist. Diese maximale Regressionstiefe liegt im allgemeinen im Bereich

$$N/3 \leq \max rdepth \leq (N + 2)/2 .$$

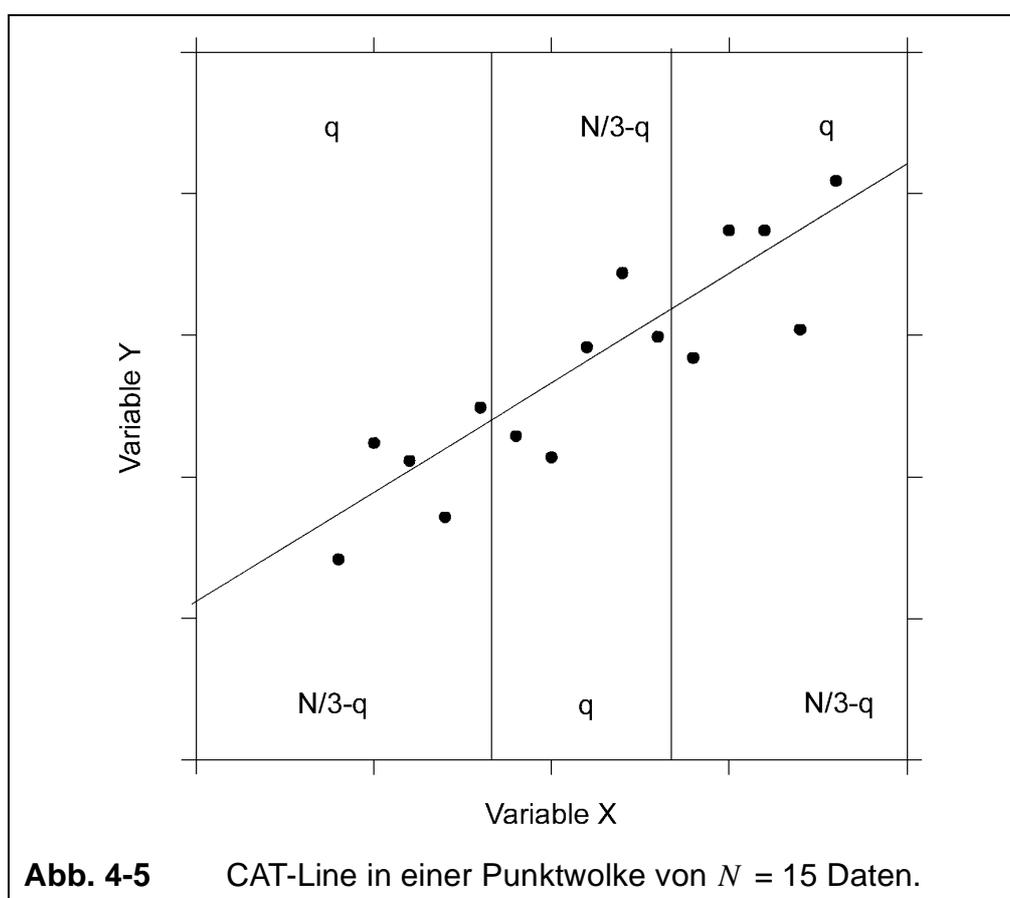
Das Konzept der maximalen Regressionstiefe ist nicht auf die einfache Regression beschränkt, sondern ermöglicht auch die robuste Regression mit mehreren unabhängigen Variablen durch iterative Auffindung von Ebenen bzw. Hyperebenen maximaler Regressionstiefe im  $p$ -dimensionalen Raum.

#### 4.6 Die CAT-Line

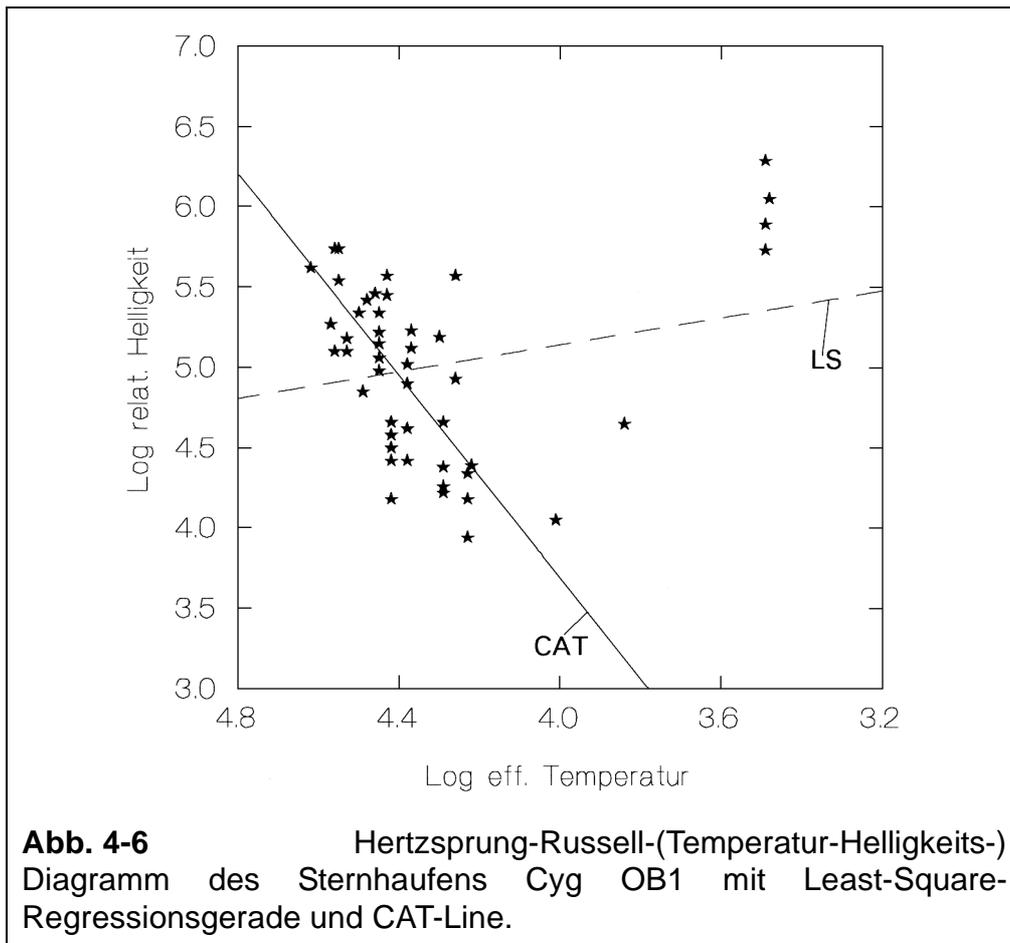
HUBERT und ROUSSEEUW (1998) haben für die einfache lineare Regression folgendes sehr praktikable Verfahren zur Konstruktion einer Regressionsgeraden angegeben, die eine Regressionstiefe von  $rdepth \geq N/3$  hat, wie es auch der oben beschriebenen robusten Bestanpassung entspricht. Dazu wird die  $y_i(x_i)$ -Punktwolke nach der Größe der  $x_i$ -Werte in ein linkes ( $L$ ), ein mittleres ( $M$ ) und ein rechtes ( $R$ ) Drittel geteilt (wenn die Anzahl  $N$  nicht durch 3 teilbar ist, wird Symmetrie hergestellt, indem dem mittleren Teil  $M$  ein Datenpunkt mehr bzw. ein Datenpunkt weniger als

den Teilen  $L$  und  $R$  zugeordnet wird). Durch einen in der Regel rasch konvergierenden Algorithmus wird dann diejenige Gerade aufgefunden, die die Forderung erfüllt, dass *sowohl* in der (kombinierten) Datenmenge  $L \cup M$  *als auch* in der (kombinierten) Datenmenge  $M \cup R$  jeweils *gleich viele* Datenpunkte oberhalb wie unterhalb der Geraden liegen. Die so definierte Gerade wird *CAT-Line* genannt, weil sie alle drei Datendrittel durchschneidet (*cuts all thirds*).

Die CAT-Line *läßt sich stets konstruieren*, gleichgültig welche Gestalt die Punktwolke hat. Die Abb. 4-5 veranschaulicht eine CAT-Line für  $N = 15$  und läßt zugleich erkennen, dass in den jeweiligen Dritteln selbst im allgemeinen *nicht* gleich viele Datenpunkte oberhalb und unterhalb zu liegen brauchen, sondern  $q$  bzw.  $(N/3 - q)$ , mit  $0 \leq q \leq N/3$ .



Die Konstruktion der CAT-Line ist mit einer Breakdown-Zahl von  $\varepsilon = 1/3$  *robust* gegenüber allen Arten mangelbehafteter Daten. Ein anschauliches Beispiel bietet die Abb. 4-6, die das Temperatur-Helligkeits-Korrelogramm der Sterne eines Sternhaufens im Sternbild des Schwans wiedergibt (Daten entnommen aus ROUSSEUW und LEROY, 1987). Die mit dem Standard-Verfahren der Least-Square-(LS)-Anpassung berechnete Regressionsgerade erlaubt keine sinnvolle Interpretation, dagegen zeigt die robuste CAT-Line, dass für die Mehrheit der Sterne eine sinnvolle Korrelation besteht. Die vier "Ausreißer" sind hier keine "grobe Meßfehler", sondern signalisieren die Existenz einer astrophysikalisch von der Mehrheit verschiedenen Kategorie (Riesensterne).



Besonders bewährt sich die Robustheit der CAT-Line bei dem oben in Abb. 4-2 dargestellten Beispiel der Meßbereichs-Unterschreitung: Der Anteil der Beobachtungen mit dem Ergebnis " $Y < 10$ " ist geringer als die Breakdownzahl  $1/3$ . Daher kann man auch dann, wenn die in Abb. 4-2 als rote Punkte markierten Beobachtungen an Stelle ihrer tatsächlichen Zahlenwerte nur durch die "mangelhafte" Angabe " $< 10$ " vertreten werden, die robuste CAT-Line (rote Gerade in Abb. 4-2) eindeutig konstruieren, die den tatsächlichen Trend der Daten wiedergibt.

#### 4.7 Literatur

- Barnett, V. 1976: The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A*, **139**, 319-354.
- Hampel, F. R. 1971: A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887-1896.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. 1986: *Robust Statistics: The Approach Based on Influence Functions*. New York: J. Wiley.
- Huber, P. J. 1964: Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73-101.
- Huber, P. J. 1981: *Robust Statistics*. New York: J. Wiley.
- Hubert, M., Rousseeuw, P. J. 1998: The Catline for deep regression. *J. Multivar. Analysis*, **66**, 270-296.

- Liu, R. Y., Parelius, J. M., Singh, K. 1999: Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statist.*, **27**, 783-840.
- Martin, R. D. 1981: Robust methods in time series. In: *Applied Time Series Analysis II* (ed. by D. F. Findley). New York: Academic Press.
- Rousseeuw, P. J., Hubert, M. 1999: Regression depth. *J. Amer. Statist. Assoc.*, **94**, 388-402.
- Rousseeuw, P. J., Leroy, A. M. 1987: Robust Regression and Outlier Detection. New York: J. Wiley.
- Smyth, G. K., Hawkins, D. M. 2000: Robust frequency estimation using elemental sets. *J. Comput. Graph. Statistics*, **9**, 196-214.
- Stahel, W., Weisberg, S. (eds.) 1991: Directions in Robust Statistics and Diagnostics I. New York etc.: Springer.
- Stromberg, A. J. 1993: Computation of high breakdown nonlinear regression parameters. *J. Amer. Statist. Assoc.*, **88**, 237-244.
- Taubenheim, J. 1969: Statistische Auswertung geophysikalischer und meteorologischer Daten. Leipzig: Akadem. Verlagsges. Geest & Portig.
- Tukey, J. W. 1975: Mathematics and the picturing of data. In: *Proceedings of the 1975 International Congress of Mathematicians*, Vol. 2, 523-531, Vancouver.



## 5 Validation von Modellen

Andreas Hense

### 5.1 Einleitung

Die Validation biogeophysikalischer Modelle ist ein inhärent statistisches Problem, da sowohl das reale als auch das simulierte System i.a. ein stochastisches System ist. Dies ergibt sich zum einen aus der Hochdimensionalität, für die die statistische Physik die Wahrscheinlichkeitsbeschreibung von Zuständen vorgibt. Zum anderen ist aus der Theorie der chaotischen, dynamischen Systeme (ECKMANN und RUELLE, 1985) bekannt, dass auch bei wenigen Freiheitsgraden eine adäquate Beschreibung über Wahrscheinlichkeitsdichten möglich ist. Da alle Modelle im Vergleich zur Realität immer nur niedrigdimensional sind, ist insbesondere das letzte Argument entscheidend für die Stochastizität von Modellsystemen.

Bevor auf die eigentlichen Probleme und Verfahren zur Modellvalidation eingegangen wird, sei zunächst eine kurze Zusammenfassung der notwendigen theoretischen Hintergründe vorangestellt. Die Gliederung dieses Beitrages ist dann wie folgt

1. Zufallsvariablen, Modelle und Daten, Wahrscheinlichkeitsdichten (pdf), Randwahrscheinlichkeitsdichten, bedingte Wahrscheinlichkeiten, Schätzen
2. Darstellung von Modell- und Datenvergleichen, Differenzen, Häufigkeiten
3. Beschreibungsmöglichkeiten der gemeinsamen pdf von Modell und Daten, RMS, skill scores

### 5.2 Theorie

Über das oben angeführte Argument zur stochastischen Betrachtung von Modellen hinaus, wollen wir folgendes Axiom annehmen:

*Alle Daten (Modell- und Beobachtungsdaten) und auch alle Zusatzinformationen sind Realisierungen von Zufallsvariablen (Bayesische Sichtweise).*

Dies trifft u.a. auch auf solche anscheinend deterministischen Parameter eines Modells zu wie Gitter, Art der Diskretisierung etc. Aber so ist z.B. die Wahl des 0°-Meridians als Greenwich-Meridian eine zufällige Entscheidung gewesen. Aber jede andere Lage des Zentralmeridians bedeutet bei einem diskreten Modell eine geänderte Darstellung der Land-Meer-Verteilung oder der Orographie. Damit "sehen" alle Modelle, in der die Land-Meer-Verteilung mit identischer Auflösung diskretisiert wird, nur eine Realisierung der Küstenlinienform, von der seit mehr als 20 Jahren bekannt ist, dass sie eine Zufallsvariable ist (MANDELROT, 1980).

### 5.2.1 Univariate reelle Zufallsvariable

Man definiert als univariate reelle ZVA  $X$  im Intervall  $[a, b]$  (Stichprobenraum) die Menge:

$$X = \{(x, F(x)), x \in [a, b]\} \quad (5-1)$$

wobei  $F$  eine monoton steigende (nicht notwendigerweise streng monoton steigende) Funktion ist, die das Intervall  $[a, b]$  auf das Intervall  $[0, 1]$  abbildet und folgendes Grenzverhalten zeigt:

$$\lim_{x \rightarrow \infty} F(x) = 1 \quad (5-2)$$

und

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad (5-3)$$

Ist  $F(x)$  auch noch stetig differenzierbar  $f(x) = F'(x)$ , so kann man/frau alternativ auch die Menge

$$X = \{(x, f(x)), x \in [a, b]\} \quad (5-4)$$

als eine reelle ZVA definieren.  $F(x)$  heißt Verteilungsfunktion der ZVA  $X$  und  $f(x)$  die Wahrscheinlichkeitsdichtefunktion. (engl. probability density function pdf). Die Wahrscheinlichkeit während der Realisierung eines Experimentes eine Zahl in Intervall  $[c, d]$  zu finden, folgt aus der Verteilungsfunktion als:

$$P([c, d]) = F(d) - F(c) \quad (5-5)$$

bzw. aus der Dichtefunktion zu

$$P([c, d]) = \int_c^d f(t) dt. \quad (5-6)$$

Insbesondere erhält man

$$\int_a^b f(t) dt = 1 \quad (5-7)$$

aus der Definition für  $F$ . Die pdf  $f(x)$  kann man/frau dann als die Wahrscheinlichkeit interpretieren, mit der eine Realisierung einen Wert liefert, der in dem infinitesimalen Intervall  $[x, x + dx]$  liegt.

### 5.2.2 Multivariate reelle ZVA's

Sei  $\Omega \in R^n$  ein vernünftiges Gebiet (oder gutartiges Gebiet: d.h. man kann alle Integrationen ausführen). Ein Element dieses Gebietes sei durch den  $n$ -dimensionalen

Vektor  $\vec{x}$  gekennzeichnet. Sei  $f$  eine Funktion von  $\Omega$  in  $R$ , die positiv semidefinit ist und für die gilt

$$\begin{aligned} f(\vec{x}) &\geq 0 \\ \int_{\Omega} f(\vec{x}) d\vec{x} &= 1 \end{aligned} \quad (5-8)$$

Dann heißt die Menge

$$\vec{X} = \{(\vec{x}, f(\vec{x})), \vec{x} \in \Omega\} \quad (5-9)$$

eine n-dimensionale (multivariate) ZVA.

Aus der gemeinsamen pdf aller ZVA  $X_1, \dots, X_n$  kann man verschiedene andere pdf noch definieren. Die Randverteilung  $r$  bzgl.  $X_i$  erhält man durch folgendes Integral

$$r(x_i) = \int \dots \int f(\vec{x}) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \quad (5-10)$$

Die bedingte pdf für alle  $X_j$  bei festem  $X_i$  ist definiert als

$$c(x_1 \dots x_{i-1} x_{i+1} \dots x_n | x_i) = \frac{f(\vec{x})}{r(x_i)} \quad (5-11)$$

Betrachten wir zwei Variablen  $m, b$ , so kann man entsprechend zwei bedingte pdf's und zwei Rand-pdf definieren, die durch das berühmte Bayestheorem miteinander verbunden sind

$$c(m|b)r(b) = c(b|m)r(m) \quad (5-12)$$

### 5.2.3 Modell-Beobachtungsvergleich

Der Vergleich von Beobachtungen  $B = \{b, b \in S_b, f_r(b)\}$  mit entsprechenden Modellen  $M = \{m, m \in S_m, f_r(m)\}$  ist nun im ganz allgemeinen Fall immer eine Aussage über die gemeinsame Zufallsvariable  $Z = (b, m) = \{\vec{z}, \vec{z} \in S_b \times S_m, f(m, b)\}$  mit Hilfe der gemeinsamen pdf  $f(m, b)$ .

Eine Aussage erübrigt sich, wenn Beobachtungen und Modelle von einander unabhängig sind. Dann gilt zwangsläufig

$$c(m|b) = r(m) \quad (5-13)$$

oder

$$f(m, b) = r(m)r(b) \quad (5-14)$$

Jede Abweichung von  $f$  bzgl. des Produktes der Randverteilungen zeigt, dass es eine Beziehung zwischen den beiden Variablen geben muss.

Die Bayesische Betrachtung des Modellvalidationsproblems wird im Abschnitt 5.3 diskutiert.

#### 5.2.4 Schätzen

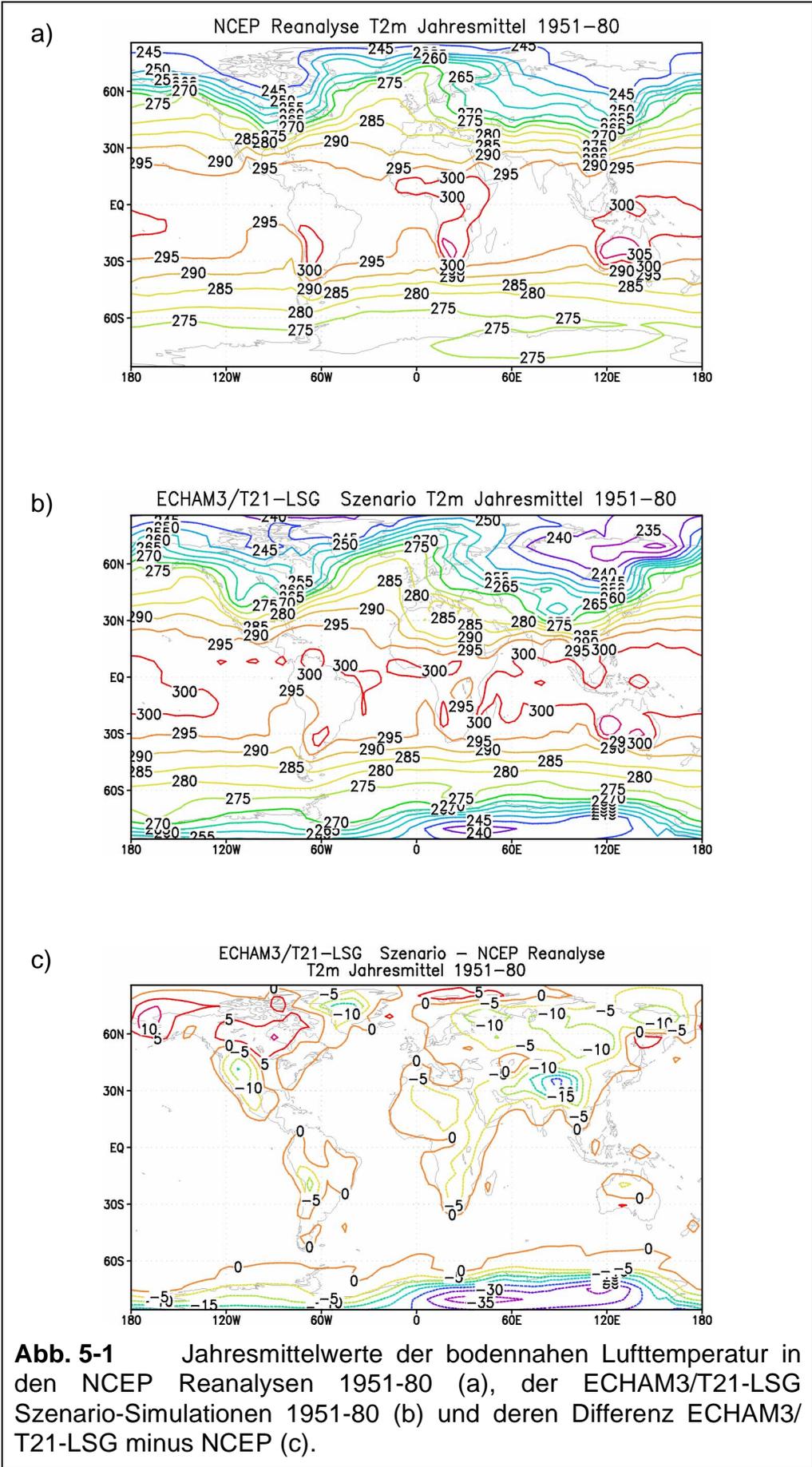
Da die ZVA Mengen sind, kann man mit diesen Grössen nicht rechnen. Jedes Experiment in den Beobachtungen und jede Simulation erzeugt Realisierungen von ZVA, von denen die pdf in den allermeisten Fällen unbekannt ist. Wird die Form der pdf vorgegeben (z.B. Normalverteilung, log-Normal, Gamma, Binomial, Poisson etc) so wird die ZVA durch die (unbekannten) Parameter der pdf-Funktion festgelegt. In jedem Fall muss aber anhand der Realisierungen (Stichprobe) die Eigenschaften der ZVA geschätzt werden (Punktschätzer). Da Realisierungen aber immer nur unvollständige Informationen liefern können wegen der endlichen Grösse der Stichprobe, treten auch immer Fehler bei der Eigenschaftsschätzung auf, die wiederum quantifiziert werden müssen. Nur die Kombination von Punktschätzungen mit den entsprechenden Fehlern macht eine statistische Untersuchung sinnvoll.

#### 5.2.5 Darstellung von Ergebnissen

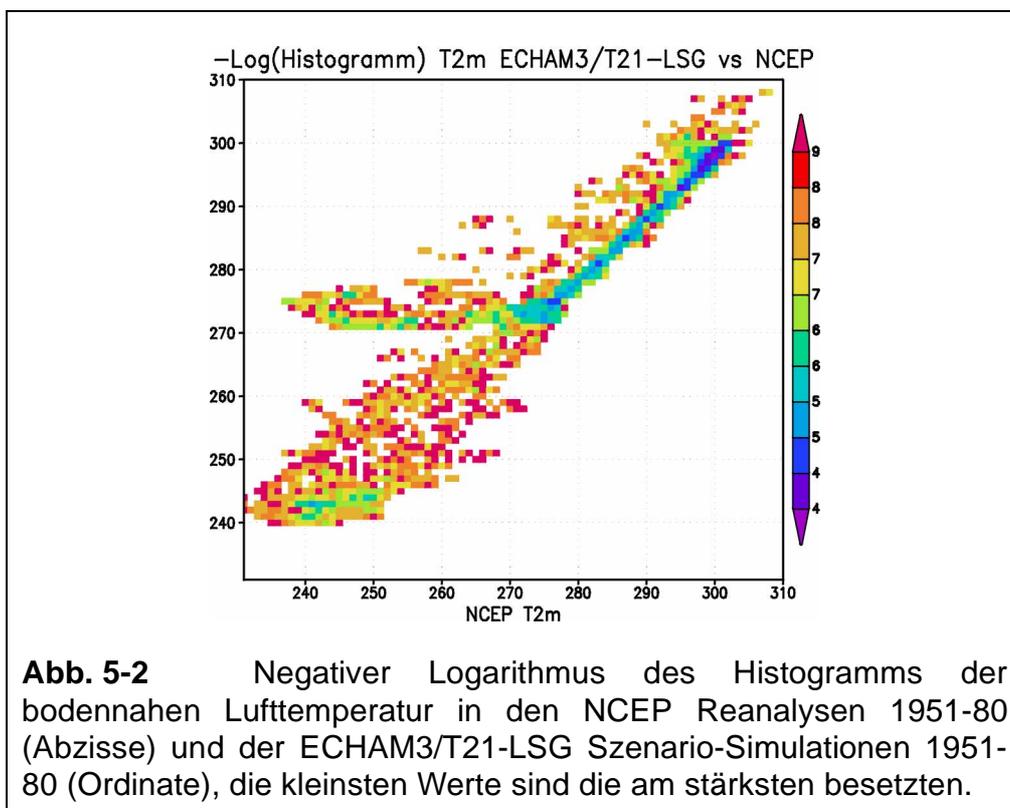
Aber bereits ohne Anwendung von Statistik, kann beim Vergleich von Modell und Beobachtungen sehr viel "Unfug" erfolgen. Denn bereits die Präsentation von Ergebnissen kann in einer Form erfolgen, die es Aussenstehenden wie Insidern nicht ermöglicht, quantitativ ein Modell mit den Beobachtungen zu vergleichen.

Zu einer nachvollziehbaren Präsentation gehören, dass

1. Vergleiche von Beobachtungen und Simulationen immer nur als Differenzdarstellungen bzw. als Differenz und Mittelwert erfolgen. Hierbei verliert man keine Information, da die Differenz und Mittelwertbildung eine orthogonale Transformation der ursprünglichen Daten ist. In Abb. 5-1 werden die mittlere bodennahe Lufttemperatur (2 m Temperatur) aus den NCEP Reanalysen und aus den ECHAM3-LSG Szenario-Simulationen für den Zeitraum 1951-80 verglichen. Während die globalen Strukturen der Felder (Abb. 5-1a und Abb. 5-1b) gut übereinstimmen, sieht man erst in der Differenz (Abb. 5-1c), dass die Kontinente in ECHAM3 systematisch zu kalt sind (dies könnte auch einen Fehler der Reanalyse beinhalten, wenn die bodennahen Lufttemperaturen in den NCEP Daten systematisch zu hoch sind).
2. Vergleiche von Steigungen (z.B. spectral slopes) oder sogar Ableitungen höherer Ordnung nur durch Berechnung derselben verglichen werden können.
3. statt Streudiagramme, die die Korrelation zwischen Beobachtungen und Modelldaten darstellen sollen, tunlichst Häufigkeitsdiagramme verwendet werden, die immer eine Schätzung der gemeinsamen Wahrscheinlichkeitsdichte pdf zwischen Modell und Beobachtungen ist. Streudiagramme liefern ein (schlechtes) Maß für den Träger der pdf, liefern aber keinerlei Anhaltspunkte für die innere Struktur der



pdf. Selbst bei relativ geringen Stichprobenumfängen lassen moderne Methoden (z.B. Kernel-Schätzer) eine vernünftige Schätzung zu. Als Beispiel diene wieder der Vergleich der bodennahen Lufttemperaturen zwischen NCEP Reanalysen und dem ECHAM3/T21-LSG Szenario Lauf. In Abb. 5-2 ist der negative Logarithmus des Histogramm-Schätzers der gemeinsamen Wahrscheinlichkeitsdichte der beiden Datensätze geplottet. Man erkennt eine deutliche lineare Struktur, die die Ähnlichkeit der beiden Felder charakterisiert mit dem bereits erwähnten systematischen Fehler zu niedriger Temperaturen in ECHAM3/T21-LSG (die kleinsten Werte sind die am stärksten besetzten). Der abzweigende Teil mit konstanten ECHAM3/T21-LSG Temperaturen auf Gefrierpunktsniveau bei variierenden NCEP Werten ist auf das Bodenmodul zurückzuführen, das bei Energieeintrag der obersten Bodenschicht die Gefrierpunktstemperatur zuweist, solange Eis oder Schnee vorhanden ist und die Energie als Schmelzwärme aufgebraucht wird.



### 5.2.6 Modell-Daten Vergleiche

Ein systematischer Vergleich zwischen Modell und Beobachtungen basiert auf einer Auswertung der gemeinsamen pdf (deswegen auch der Vorschlag im letzten Kapitel, statt Streudiagramme bereits die pdf zu schätzen). Hier sei zunächst die univariate Betrachtung gewählt, um die Methoden und die Bezeichnungen einzuführen. Die Präsentation orientiert sich an der Veröffentlichung von MURPHY und WINKLER (1987).

Wir betrachten eine Vorhersage  $m$  und eine Beobachtung  $b$  z.B. die Modelldaten gemittelt über vier um eine Station benachbarte Gitterpunkte mit den Messungen dieser Station. Wenn man nun aus der Schätzung der gemeinsamen pdf's  $p(m, b)$  die bedingte pdf  $c(m|b)$  bestimmt, so sagt man, dass  $m$  nicht diskriminierend ist, wenn

die bedingte Wahrscheinlichkeit für  $m$  bei zwei unterschiedlich gegebenen Beobachtungen  $b_1, b_2$  identisch ist  $c(m|b_1) = c(m|b_2)$ . Die Vorhersage  $m$  ist perfekt diskriminierend, wenn die bedingte Wahrscheinlichkeit für  $m$  bei gegebenen  $b$  Null ist, solange  $m \neq b$  aber bei  $m = b$  den Wert 1 annimmt:  $c(m|b) = \delta(m-b)$  ( $\delta(x)$  ist die Dirac'sche Deltafunktion).  $m$  ist perfekt kalibriert, wenn der bedingte Erwartungswert für die Beobachtung bei gegebenen  $m$   $E(b|m)$  gleich dem gegebenen Modellwert  $m$  ist  $E(B|M) = \int bc(b|m)db = m$ .

Andere Parameter, mit denen die gemeinsame pdf beschrieben wird, sind die Korrelation  $corr(M, B)$  und der mittlere quadratische Fehler  $MSE$ . Voraussetzung für eine statistisch sinnvolle Anwendung dieser Größen (d.h. die Schätzer liefern eine sinnvolle Beschreibung der pdf) ist die Normalverteilung der Beobachtungen und Simulationen.

Dabei ist auch zu beachten, dass

1. die Korrelation  $corr(M, B)$  insensitiv gegen Amplitudenfehler aber sensitiv gegen Fehler in der Phasenbeziehung zwischen  $m$  und  $b$  ist, da die Kovarianz mit der Standardabweichung normiert wird.

$$cov(B, M) = \iint (b - E(B))(m - E(M))f(b, m)dbdm$$

$$var(B) = \iint (b - E(B))^2 f(b, m)dbdm$$

$$corr(B, M) = \frac{cov(B, M)}{\sqrt{var(B)var(M)}}$$

2. deshalb der mittlere quadratische Fehler  $MSE$

$$MSE(B, M) = \iint (b - m)^2 f(b, m)dbdm$$

$$= (E(B) - E(M))^2 + var(B) + var(M) - 2corr(B, M)\sqrt{var(B)var(M)}$$

ein weiteres notwendiges Maß ist, in das sowohl Amplituden- als auch Phasenfehler eingehen.

Betrachtet man standardisierte ZVA mit  $E(B) = E(M) = 0$ ,  $Var(B) = Var(M) = 1$ , so erhält man mit

$$MSE(B, M) = 2(1 - corr(B, M))$$

die Beziehung zwischen dem Korrelationskoeffizienten und dem  $MSE$ .

Verschwindet die Korrelation  $Corr(B, M) = 0$  so ist  $MSE(B, M) = 2$ . Somit sind Zufallsvorhersagen (Vorhersagen durch Auswürfeln) durch einen  $MSE = 2$  charakterisiert. Eine andere Art der Vorhersage ist die Wahl  $m = E(B)$ , d.h. die Klimavorhersage. Damit ergibt sich  $\rightarrow MSE(B, M) = Var(B) = 1$  bzw. als Korrelationskoeffizient  $\rightarrow corr(B, M) = 0.5$ . Berechnet man also einen Korrelationskoeffizienten von 0.5 zwi-

schen Beobachtungen und Vorhersagen, so ist das Vorhersagemodell nicht besser als die Klimavorhersage. Dementsprechend sind auch die üblichen Tests für Korrelationskoeffizienten, die typischerweise die Nullhypothese  $Corr(B, M) = 0$  überprüfen, bei Modellvalidationen unerheblich. Ein sehr viel aussagekräftigerer Test ist die Überprüfung der Nullhypothese  $Corr(B, M) > 0.5$ .

Aus diesem Grund konstruiert man Validationsmaße relativ zu einer Kontrollvorhersage. So ist der Brier skill score das Verhältnis der  $MSE$  Werte zwischen  $B, M$  relativ zu einer Kontrollvorhersage  $C$ , die man z.B. als die Klimavorhersage wählen kann.

$$S_{MCB} = \frac{MSE(C, B) - MSE(M, B)}{MSE(C, B)} \quad (5-15)$$

Ein Brier skill score  $S_{MCB} = 1$  erhalten wir mit  $MSE(M, B) = 0$ , also bei einer perfekten Modellvorhersage. Ist dagegen die Vorhersage  $M$  kein Fortschritt gegenüber der Klimavorhersage  $C$ , so ist  $MSE(C, B) = MSE(M, B)$  und somit  $S_{MCB} = 0$ . Ist sogar der mittlere quadratische Fehler der Vorhersage grösser als die Klimavorhersage (d.h. die Klimavorhersage ist besser), so wird der Brier skill score negativ.

Bei kategorisierten Vorhersagen ( $n$  Klassenvorhersagen diskreter Werte), kann man die volle Wahrscheinlichkeitsbeschreibung ausnützen, indem die entsprechenden  $n^2$  vielen gemeinsamen Wahrscheinlichkeiten von Beobachtung und Vorhersage durch Auszählen geschätzt werden. Daraus lassen sich ebenfalls skill scores berechnen, die die Struktur dieser  $n^2$  Wahrscheinlichkeiten relativ zu einer gemeinsamen pdf beschreiben, in der Beobachtungen und Vorhersagen unabhängig sind. Dies definiert z.B. den Heidke skill score, im folgenden am Beispiel von zwei Klassen  $n = 2$  dargestellt.

	b=1	b=0	
m=1	f(1,1)	f(0,1)	r(m=1)
m=0	f(1,0)	f(0,0)	r(m=0)
	r(b=1)	r(b=0)	1

Der Heidke skill score wird dann definiert als

$$S = \frac{p_c - p_e}{1 - p_e}$$

$$p_c = f(1, 1) + f(0, 0)$$

$$p_e = r(b = 1)r(m = 1) + r(b = 0)r(m = 0)$$

$$S \in \left[ -\frac{p_e}{1 - p_e}; +1 \right]$$

Hierbei ist  $p_e$  die Wahrscheinlichkeit bei unabhängigen Beobachtungen.

## 5.2.7 Felder und korrelierte Zeitreihen

Im letzten Abschnitt wurde die Validation von univariaten Vorhersagen und Beobachtungen (z.B. einzelne Stationen) betrachtet. Allerdings werden in Modellen (insbesondere in atmosphärischen und ozeanischen Modellen) oft Felder an Gitterpunkten oder als Spektralkoeffizienten bestimmt. Damit ist es auch notwendig, sich Gedanken über die Validation von Feldern zu machen.

Dazu betrachten wir Felder von diskreten Variablen (z.B. an Gitterpunkten) als (hochdimensionale) Vektoren  $\vec{z}$  von normalverteilten Zufallsvariablen (es gibt fast keine anderen parametrischen multivariaten Verteilungsfunktionen als die Normalverteilung NV) mit der Dimension  $q$  (= Anzahl der Gitterpunkte).

$$f(\vec{z}) \sim \exp\left(-\frac{1}{2}(\vec{z} - \vec{\mu})^T \Sigma^{-1}(\vec{z} - \vec{\mu})\right) \quad (5-16)$$

Die multivariate NV wird durch den Vektor des Mittelwertes  $\vec{\mu}$  und durch die inverse Kovarianzmatrix  $\Sigma^{-1}$  vollständig beschrieben. Diese Werte müssen aus den Daten sowohl für die Beobachtungen als auch für die Vorhersagen geschätzt werden. Das Problem des Schätzens ist dabei die inverse Kovarianzmatrix  $\Sigma^{-1}$ . So ist es i.A. nicht möglich die Inverse zu bestimmen, da die geschätzte Kovarianzmatrix  $\Sigma$  nicht vom vollen Rang ist. Dies tritt dann auf, wenn der Stichprobenumfang kleiner ist als die Dimension  $q$ . Aber auch schon bei Stichprobenumfängen, die nur wenig grösser sind als die Dimension  $q$  treten Probleme auf, da dann zwar die geschätzte Kovarianzmatrix den vollen Rang hat und ein unverzerrter Schätzer für die Kovarianzmatrix der Grundgesamtheit ist, die inverse Matrix aber eine gegebenenfalls grosse Verzerrung aufweist, so dass die geschätzte NV sehr grosse Fehler haben kann. Dies erfordert eine Dimensionsreduktion z.B. durch EOF, physikalisch motivierte Muster oder hierarchische Basen wie Fourierkoeffizienten globaler oder lokaler Natur (Wavelets). (Ein ähnliches Problem tritt auch bei univariaten aber korrelierten Zeitreihen auf.)

## 5.3 Validation of model predictions and analysis of measurements - a unified Bayesian view

### 5.3.1 Introduction

A scientific sound validation of models is constraint by two main aspects

- observations of ozone concentrations are more or less point values or averages for very small volumes while simulations produce data which are representative for much larger volumes in the order of the resolution of the model ( $\vartheta (\Delta x \times \Delta y \times \Delta z)$ ) probably even larger due to the numerical truncation errors.
- the validation is a comparison of two stochastic variables. The observed concentrations are influenced by random observation errors but also by the turbulent fluctuations of the advecting flow, within a network of stations one has to consider additionally the effect of aliasing, the misinterpretation of unresolved scales being part of the resolved scales. The simulations do produces stochastic variables due

to modeled non-linear interaction of the different scales of the advecting flow. In the ideal case a simulation of the ozone concentrations will be done with a large-eddy-resolving model (which is at present not feasible due to limited computer resources) which per-se simulates the turbulence that is a stochastic state.

In view of the last point the validation of a forecast  $F$  in view of the data  $D$  is formally equivalent to an analysis of the joint probability density function (pdf)  $p(F, D)$ . However, for a specific event  $F$  one could prepare a large sample of forecasts  $f$  even with a variety of models  $m$  but in general we have only **one** realisation  $d$  of the data  $D$  available. Then one could ask how probable the forecasts are in view of the given data which is a conditional probability density function  $cond(F|D)$ . The joint pdf and the condition pdf are connected through the celebrated Bayes Theorem which states:

$$p(F, D) = cond(F|D)p(D) \quad (5-17)$$

where  $p(D)$  is the marginal distribution of the data. Of course, the Bayes Theorem can be written in the symmetrical way for a conditional pdf  $cond(D|F)$

$$p(F, D) = cond(D|F)p(F) \quad (5-18)$$

such that we can combine both equations to the form which is mostly used in Bayesian inference

$$cond(F|D) \sim cond(D|F)p(F) \quad (5-19)$$

That is the probability of the forecasts in view of the data ( "given  $d$ " ) is proportional to the probability of the forecasts  $p(F)$  the so called prior probability and the conditional probability of the data given the forecast (the so called likelihood). The conditional pdf  $cond(F|D)$  is also named the posterior. Obviously the priori is a measure for the confidence of the forecast model. In fact, since different forecast models might be available, it is itself again a conditional probability namely  $p(F) = cond(F|m)$  a probability of a certain forecast given a model  $m$ . We could rewrite the last equation using the Bayes theorem again as

$$cond(F|D, m) \sim cond(D|F)cond(F|m)p(m) \quad (5-20)$$

The likelihood  $cond(D|F)$  is a statement about the quality of data including the process of obtaining the data i.e. the measurement process since it states upon a given forecast  $F = f$  what is the probability of finding certain data  $D$ . The Bayes theorem now states that from the data quality and from the confidence of the model we can make inferences about the probability of the forecasts. If we sum or integrate Eq. (5-20) over all possible forecasts  $F$  one obtains the evidence for the data  $D$  given the model  $m$

$$cond(D|m) = \int cond(F|D, m)dF \sim \int cond(D|F)cond(F|m)dF \quad (5-21)$$

A posterior probability for the model can be calculated invoking once more the Bayesian theorem

$$cond(m|D) \sim cond(D|m)p(m) = \left(\int cond(D|F)cond(F|m)dF\right)p(m) \quad (5-22)$$

which could be evaluated provided a prior is known for the model. The least restrictive formulation is to assume equal probability for all models

$$p(m) \sim const \quad (5-23)$$

which leads to the statement that the evidence of the data given the model can not be distinguished from the evidence of the model given the data.

These points have been made in several recent papers on detecting climate signals (TOL and DE VOS, 1998; HASSELMANN, 1998). A very clear paper where several of the above and below arguments have been taken from is presented by LEROY (1998). One of the standard text books where these ideas are described in a more general way is BERGER (1985).

The first point of the above two remarks - the comparison of small vs. large volume averaged values - can be addressed with the same technique. We have to find a data model which filters from the local, pointwise information a spatially coherent, volume averaged new "data" field.

Suppose we have an a-priori information on such a data field which we will now name  $F$ . The quality of this first guess (as it is called) is described by the prior pdf  $p(F)$ . This first guess could be climatology or a well educated guess from some specialists. The conditional pdf  $cond(D|F)$  which includes the sampling of the data can be multiplied with the priori pdf. Then the "data model" can be defined as that forecast  $F$  which maximises the posterior pdf

$$cond(F|D = d) \stackrel{!}{=} max \quad (5-24)$$

### 5.3.2 Univariate example

In the above section we described the abstract case of the Bayesian technique. For a real application one has to define the random variates  $F$  and  $D$  or their realisations  $f, d$  respectively. As a simple test case we will consider the univariate case at first. One might think about forecasts  $F$  with a box model and observations  $D$  at a single site. Formally we define the random variate as the following sets

$$\begin{aligned} F &= [f, f \in \mathbb{R}, g(f)] \\ D &= [d, d \in \mathbb{R}, h(d)] \end{aligned} \quad (5-25)$$

where  $d, f$  are real number associated with the random process (i.e. the concentrations), the interval of definition and the  $g, h$  the function, which describe the pdf. For simplicity we will assume Gaussian random variates (which can be sometimes questioned, since e.g. concentration or precipitation are positive definit,  $f, d \in \mathbb{R}^+$ ).

$$p(F) = g(f) = \frac{1}{\sqrt{2\pi}\sigma_f} \exp\left(-\frac{1}{2} \frac{(f-\mu_f)^2}{\sigma_f^2}\right) \quad (5-26)$$

Next we have to define the likelihood  $cond(D|F)$ . We will assume again a Gaussian statistics. Then we get

$$cond(D|F) = \tilde{g}(d|f) = \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left(-\frac{1}{2} \frac{(f-d)^2}{\sigma_d^2}\right) \quad (5-27)$$

The parameters  $\mu_f, \sigma_f$  describe the model quality, that is the average and the standard deviation around it which the modeller expect for his/her model. The second equation (5-27) describes the errors arising from the measurement if the observer applies his/her instruments to the field  $f$  such that  $\sigma_d$  is the standard error of the measurement. Note that the measurements produce unbiased  $E(D) = E(F)$  observations. Multiplying both pdf's gives the posterior

$$cond(F|D) \sim \frac{1}{2\pi\sigma_f\sigma_d} \exp\left(-\frac{1}{2} \left( \frac{(f-\mu_f)^2}{\sigma_f^2} + \frac{(f-d)^2}{\sigma_d^2} \right)\right) \quad (5-28)$$

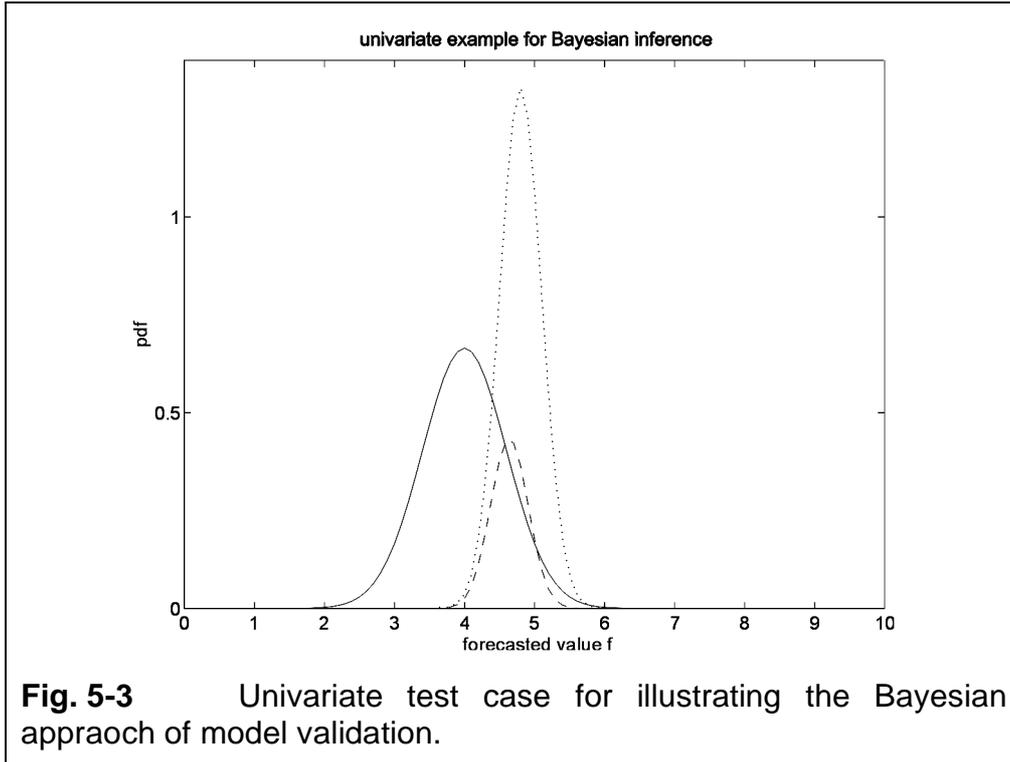
Applying some algebra it is easy to show that the product is again a Gaussian with mean  $e$  and variance  $\sigma_e^2$

$$\begin{aligned} cond(F|D) &\sim \frac{1}{2\pi\sigma_f\sigma_d} \exp\left(-\frac{1}{2} \left( \frac{(f-e)^2}{\sigma_e^2} \right)\right) \exp\left(-\frac{1}{2} \frac{(d-\mu_f)^2}{\sigma_d^2 + \sigma_f^2}\right) \\ e &= \left( \frac{d}{\sigma_d^2} + \frac{\mu_f}{\sigma_f^2} \right) \left( \frac{1}{\sigma_d^2} + \frac{1}{\sigma_f^2} \right)^{-1} \\ \sigma_e^2 &= \left( \frac{1}{\sigma_d^2} + \frac{1}{\sigma_f^2} \right)^{-1} \end{aligned} \quad (5-29)$$

This result is identical to the classical regression analysis of finding the least-squares estimator for optimal averaging two variate with different variances. The better the variable is measured the more weight it will get for the average but every variate no matter how "bad" it is will reduce the variance  $\sigma_e^2$ . Integrating the posterior Eq. (5-29) over all  $f$  will give the evidence of the data given the model or - assuming a constant prior for the model  $p(m)$  - the evidence for the model in view of the data.

$$p(m|D) \sim p(D|m) \sim \frac{\sigma_e}{\sqrt{2\pi}\sigma_f\sigma_d} \exp\left(-\frac{1}{2} \frac{(d-\mu_f)^2}{\sigma_f^2 + \sigma_d^2}\right) \quad (5-30)$$

Either the model is improbable with respect to the data or vice versa if  $(d-\mu_f)^2$  is large compared to the sum of variances  $\sigma_f^2 + \sigma_d^2$ . The following figure (Fig. 5-3) should



illustrate the above derivation. The full line indicates the prior - the pdf  $P(F)$  - while the dotted curve represent the likelihood, the conditional pdf  $cond(D|F)$ . The posterior as the product of these two pdf's is shown with a dashed line. It is clear that the posterior is sharper than the two other pdf's, the conditional variance of the forecasts given the data is smaller than either of the input pdf's. On the other hand the evidence - a measure of the maximum value of the posterior - is low ( $\sim 0.3$  in the present case) indicating that in view of the variances the distance between model and data is comparatively large.

### 5.3.3 Multivariate method

The univariate approach has its merits in describing the method. However, for real applications a multivariate approach is necessary. The forecast at a single grid point is not independent from its neighbours due to the model equations and also measurements at different points are correlated.

We define a  $q$  dimensional model vector  $\vec{f}$ . The components of this vector are the modeled concentrations at every grid point labeled in an appropriate way. The data are also collected in a  $m$  dimensional vector  $\vec{d}$ . For most applications the relation  $m \ll q$  holds.

Assuming a multivariate Gaussian process for the model the prior pdf for the forecasts is given by

$$p(F) = \frac{1}{\sqrt{2\pi^q \det \Sigma_f}} \exp\left(-\frac{1}{2}(\vec{f} - \vec{\mu}_f)^T \Sigma_f^{-1} (\vec{f} - \vec{\mu}_f)\right) \quad (5-31)$$

where  $\vec{\mu}_f$  is the expectation of the forecasted field and  $\Sigma_f$  the variance - covariance matrix of the forecast describing the precision of the forecast (model). The determinant of  $\Sigma_f$  ( $\det\Sigma_f$ ) gives a typical value for this precision.

The conditional pdf  $cond(D|F)$  has to take into account that the measurements of the data do not only introduce observational errors but also a sampling of the field  $\vec{f}$  takes place. We will assume that this sampling process is a linear process, the data can be represented by linear combinations of the forecasted values overlayed with a random process. Again assuming a Gaussian random measurement noise we can write the likelihood as

$$cond(D|F) = \frac{1}{\sqrt{2\pi^m \det\Sigma_o}} \exp\left(-\frac{1}{2}(\vec{d} - M\vec{f})^T \Sigma_o^{-1} (\vec{d} - M\vec{f})\right) \quad (5-32)$$

Here the matrix  $M$  represents the linear measurement operator and  $\Sigma_o$  is the covariance matrix of the respective observation errors.

The posterior  $cond(F|D)$  can be calculated from the product of the prior (5-31) and the likelihood (5-32). With the help of some matrix algebra we obtain

$$cond(F|D) = \frac{1}{\sqrt{2\pi^{m+q} \det\Sigma_o \det\Sigma_f}} \exp\left(-\frac{1}{2}(\vec{f} - A^{-1}\vec{b})^T A(\vec{f} - A^{-1}\vec{b})\right) \exp\left(-\frac{1}{2}\mathcal{A}\right) \quad (5-33)$$

$$A = \Sigma_f^{-1} + M^T \Sigma_o^{-1} M$$

$$\vec{b} = \Sigma_f^{-1} \vec{\mu}_f + M^T \Sigma_o^{-1} \vec{d}$$

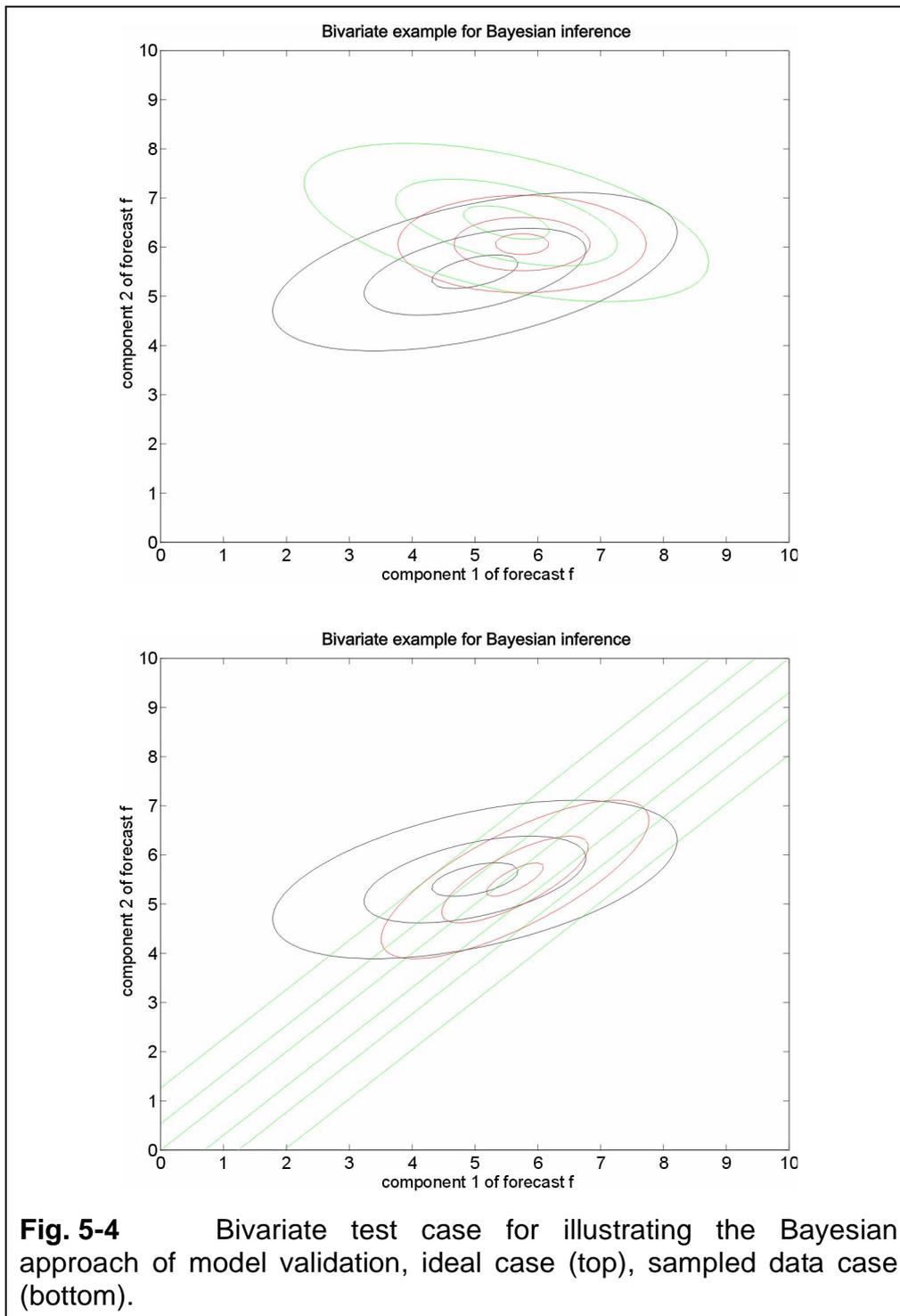
$$\mathcal{A} = (\vec{d} - MA^{-1}\vec{b})^T \Sigma_o^{-1} (\vec{d} - MA^{-1}\vec{b}) + (A^{-1}\vec{b} - \vec{\mu}_f)^T \Sigma_f^{-1} (A^{-1}\vec{b} - \vec{\mu}_f)$$

The expected (and also the most probable) forecast is given by  $A^{-1}\vec{b}$  which is obviously the multivariate extension of the above univariate example. The conditional covariance matrix of the forecasts given the data  $\vec{d}$  is  $A^{-1}$ . The evidence of the model (given the data) again assuming a flat priori for the model's probability  $p(m)$  is equivalent to the evidence for the data (given the model). It is found by integrating the posteriori over all possible forecasts which leads to

$$cond(m|D) = \frac{1}{\sqrt{2\pi^m}} \sqrt{\frac{\det(A^{-1})}{\det\Sigma_o \det\Sigma_f}} \exp\left(-\frac{1}{2}\mathcal{A}\right) \quad (5-34)$$

The evidence for the model given the data or the accuracy of the model prediction given the data is low if  $\mathcal{A}$  is large or vice versa.

In Fig. 5-4 the multivariate method is illustrated for the two dimensional case with two examples. The upper figure shows the result for an ideal case. We have a two dimensional model as well as two dimensional observations with no sampling involved.  $M$  is the identity matrix. The correlation between the model components is 0.8 while for the observations it is -0.5. The black contours are for the model prior  $p(F)$  and the green



ones for the data likelihood  $cond(D|F)$ . The red contours indicate the posterior  $cond(F|D)$ . Evidence for the model given the data is  $\sim 0.05$ , which is of course due to the large difference of the expectations of the prior and the likelihood. The lower figure shows the result in case of a single observation which is a linear combination of the forecast values. Again black contours are for the prior, green ones for the likelihood and red ones for the posterior. Evidence for the model is similar to the ideal case. The examples shows that even for the case with reduced data information the optimal combination of forecast and data will lead to a confined posterior result.

### 5.3.4 Numerical analysis of observations

The numerical analysis of observations can be obtained from the results of the last chapter by some simple reinterpretations and the application of the maximum principle stated in the introductory section.

We will associate the mean forecast field  $\vec{\mu}_f$  with the so called first guess. It can be derived from

- climatology
- the forecast of a comprehensive forecast model although this would require repeated runs of this model
- the forecast of some intermediate model e.g. a simple two dimensional advective - diffusive model.

If we define the data model as the most probable forecasts of the posterior we can directly read off the resulting equations from those in Eq. (5-33)

$$\vec{f}_{fd} = A^{-1}\vec{b} = (\Sigma_f^{-1} + M^T \Sigma_o^{-1} M)^{-1} (\Sigma_f^{-1} \vec{\mu}_f + M^T \Sigma_o^{-1} \vec{d}) \quad (5-35)$$

which is a linear combination of the first guess and the data.

The covariance matrix for this analysed field is given by the matrix  $A$

$$\Sigma_{fd} = (\Sigma_f^{-1} + M^T \Sigma_o^{-1} M)^{-1} \quad (5-36)$$

The major task for the analysis scheme is to define the (simple) forecast model. If we consider two dimensional, horizontal advection with a given flow field  $\vec{v}$  the variable  $f$  should follow

$$\frac{d}{dt} f = \frac{\partial}{\partial t} f + \vec{v} \cdot \nabla f = Q \quad (5-37)$$

This equation can be discretized in time in an Eulerian or Lagrangian sense. If we choose the Lagrangian version, we have to calculate backward trajectories  $i = 1, q$  over a time step  $\Delta t$

$$\begin{aligned} \frac{d}{dt} \vec{r}_i &= \vec{v}(\vec{r}_i) \\ \vec{r}_i(t = t_e) &= \vec{e}_i \\ t &\in [t_e - \Delta t, t_e] \end{aligned} \quad (5-38)$$

where  $\vec{e}_i$  is the grid point, where the trajectory ends and where we would like to produce the analysis. If we integrate Eq. (5-37) in the Lagrangian version along each trajectory we get

$$\mathbf{f}(e_i, t_e) - \mathbf{f}(r = a_i, t_e - \Delta t) = \Delta t \int_{a_i}^{e_i} Q d\mathbf{r} \quad (5-39)$$

The variable at the starting points of the trajectories  $\rho(a_i, t = t_e - \Delta t)$  have to be calculated from grid point variables at time  $t = t_e - \Delta t$  by interpolation. If we arrange the variables at the grid points as components in a  $q$  dimensional vector  $\mathbf{f}$

$$\begin{aligned} \mathbf{f}^T(t_e) &= (\mathbf{f}(e_1), \dots, \mathbf{f}(e_q)) \\ \mathbf{f}^T(t_e - \Delta t) &= (\mathbf{f}(e_1, t_e - \Delta t), \dots, \mathbf{f}(e_q, t_e - \Delta t)) \\ \mathbf{Q} &= \Delta t \left( \int_{a_1}^{e_1} Q d\mathbf{r}, \dots, \int_{a_q}^{e_q} Q d\mathbf{r} \right) \\ \mathbf{f}(t_e) &= P\mathbf{f}(t_e - \Delta t) + \mathbf{Q} + \mathbf{n} \end{aligned} \quad (5-40)$$

The matrix  $P$  summarizes the interpolation, the vector  $\mathbf{Q}$  contains the integrals of the source term along each trajectory and the vector  $\mathbf{n}$  represent the discretization errors as well as all other omitted processes. Also the covariance matrix  $\Sigma_f$  can be calculated as

$$\Sigma_f = E((\mathbf{f}(t_e) - E(\mathbf{f}(t_e)))(\mathbf{f}(t_e) - E(\mathbf{f}(t_e)))^T) = P\Sigma_a P^T \quad (5-41)$$

where  $\Sigma_a$  is the covariance matrix of the concentrations at time  $t_a = t_e - \Delta t$ . If we assume the covariance matrix of the observations  $\Sigma_o$  to be known and associate  $\mu_f$  with  $\mathbf{f}(t_e)$  we can estimate the data model from Eq. (5-35) with its respective covariance matrix Eq. (5-36). Now if we identify  $\mathbf{f}(t_e - \Delta t)$  with the data model at time  $t_e - \Delta t$

$$\mathbf{f}|_d(t_e - \Delta t) = \mathbf{f}(t_e - \Delta t) \quad (5-42)$$

and its covariance matrix  $\Sigma_a$  with the conditional matrix  $\Sigma_{f|d}$  at the previous analysis time, the whole system is closed except for some initial value at time  $t = 0$ . Of course the whole method is equivalent to the well known Kalman filtering although the equations for the Kalman filtering are often written in a slightly modified way (see e.g. LYSTER et al., 1997).

## 5.4 References

- Berger, J. O. 1985: Statistical decision theory and Bayesian analysis. Springer, 617 pp.
- Eckmann, J. P., Ruelle, D. 1985: Ergodic-theory of chaos and strange attractors. Rev. Mod. Phys., **57** (3), 617-656.
- Hasselmann, K. 1998: Conventional and Bayesian Approach to Climate Change Detection and Attribution. *Quarterly Journal of the Royal Meteorological Society*, 1998 (in press), also MPI Report No. 260.
- Mandelbrot, B. 1980: The fractal geometry of Nature.

- Murphy, A. H., Winkler, R. L. 1987: A General Framework for Forecast Verification. *Month. Weath. Rev.*, Vol. 115, **7**, pp. 1330-1338.
- Tol, R. S. J., de Vos, A. 1998: A Bayesian statistical analysis of the enhanced greenhouse effect. *Climatic Change*, **38**, 87-112.
- Leroy, S. S. 1998: Detecting climate signals: some Bayesian aspects. *J. of Climate*, **11**, 640-651.
- Lyster, P. M., Cohn, S. E., Menard, R., Chang, L.-P., Lin, S.-J., Olsen, R. G. 1997: Parallel implementation of a Kalman filter for constituent data assimilation. *Mon. Weath. Rev.*, **125**, 1674-1686.

## **6 How uncertain are regional climate change scenarios? Examples for Europe and the Alps**

*Dimitrios Gyalistras*

### **6.1 Introduction**

Climatologists are confronted with the task to produce useful information for decision making and planning in view of global climate change. Information on likely climatic changes is required at the spatial scale at which most climate impacts occur and are most clearly perceived by the public, i.e. at a scale of a few to a few thousands of kilometers. This scale is termed here as the "regional scale", in contrast to the global scale at which Global Climate Models (GCM) operate.

The needed information is normally provided in the form of scenarios. These are internally consistent, quantitative, more or less plausible prescriptions of possible future space-time evolutions of the climate system (IPCC-TGCIA, 1999). Climate scenarios are a means to account for uncertainty due to "unknowable" knowledge, such as the future global socio-economic development. At the same time, each individual scenario is subject to substantial uncertainties due to limitations in the available knowledge, data and computing power (see also CARTER et al., 1999).

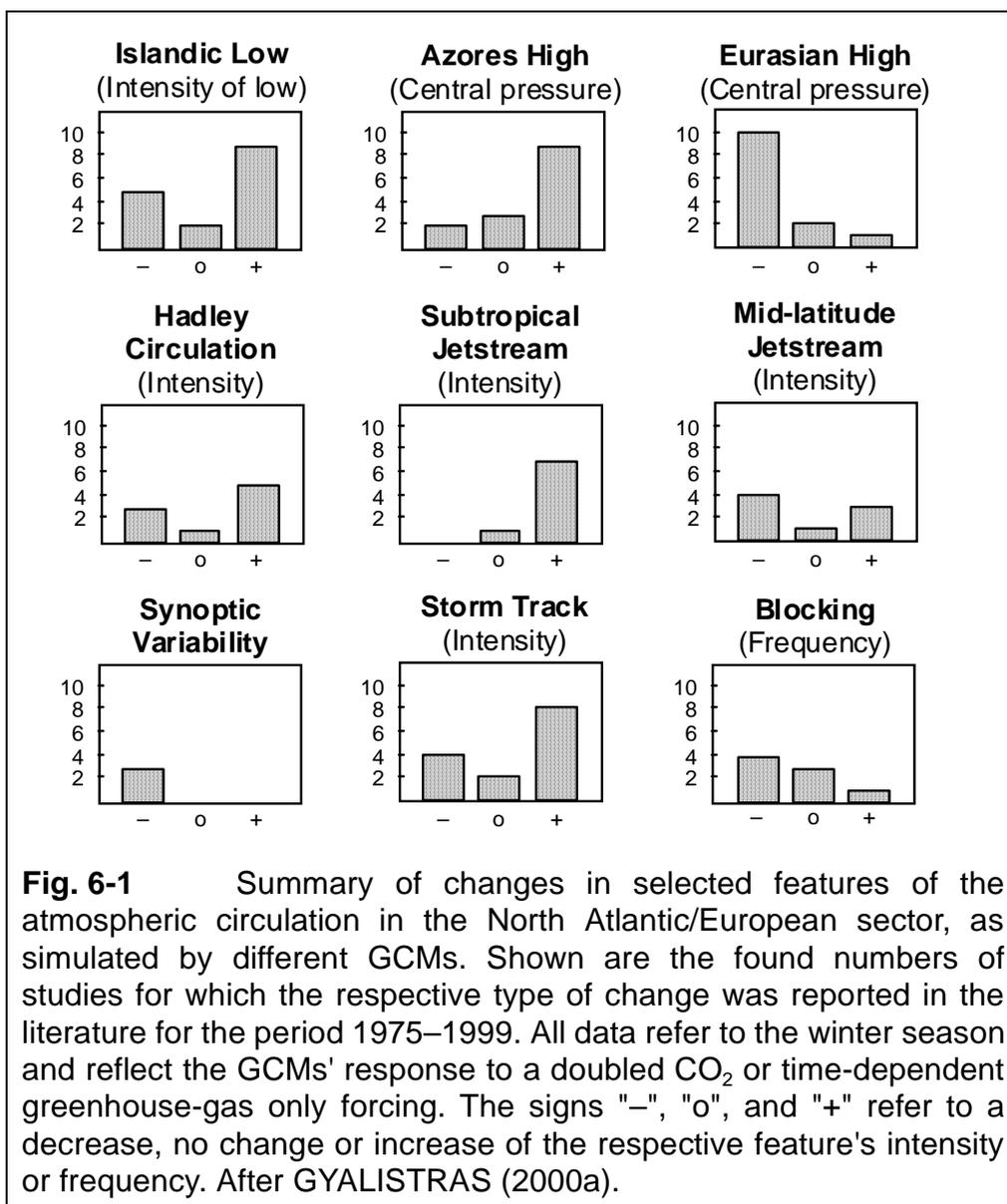
In this context the questions arise: How large is the uncertainty extant in the presently available scenarios of future climatic change for a particular region such as Europe or the European Alps? And how should one deal with that uncertainty in the context of climatic impact assessments?

The present paper addresses these questions based on the results from two large review studies that are dealing with regional climate change scenarios for the North Atlantic/European sector and the Alps. In the next two sections the main results of the two studies are presented. Then these results and their implications for climate and climate impact research are discussed. The paper ends with some conclusions and recommendations for future research.

### **6.2 Scenarios of atmospheric circulation changes for the North Atlantic and Europe**

The first review study (GYALISTRAS, 2000a) focused on possible future changes in the atmospheric circulation in the North Atlantic/European sector as simulated by GCMs. The study examined the specialist literature from ca. 1975 until the end of 1999.

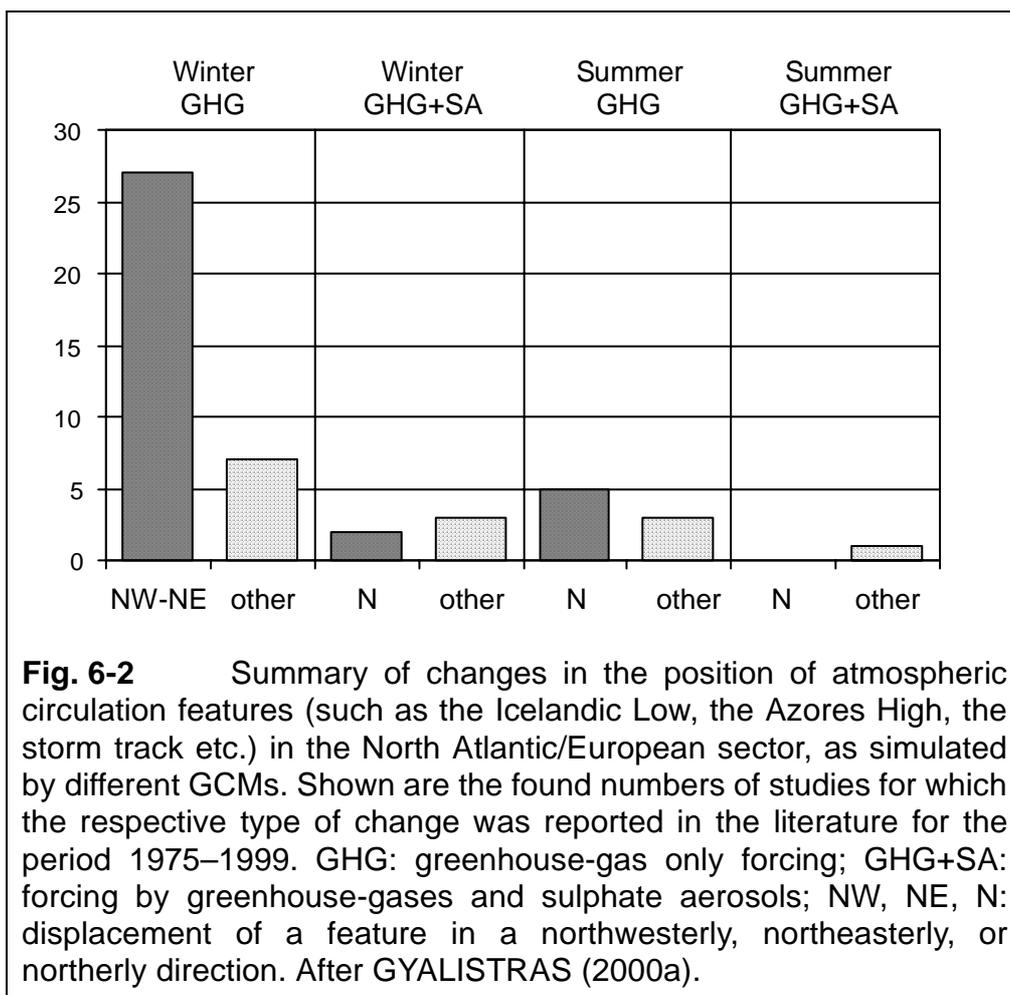
Relevant results were found in 50 out of initially selected 140 publications that dealt with the simulation of climatic changes by GCMs. The data base finally analyzed consisted of 33 simulations that have been carried out at 7 modelling centers using 24 different GCMs.



The review showed that the GCMs do not agree on the sign of the putative changes for many important circulation features. In particular, opposite changes were reported for features related to wintertime storm activity, namely the intensities of the Icelandic Low, the mid-latitude jetstream and the storm track under a greenhouse-gas only forcing; some agreement, however, was obtained with regard to a possible intensification of the Azores and Eurasian High pressure systems and the subtropical jetstream (Fig. 6-1).

The models seem also to agree upon a general northward shift of circulation patterns, at least for wintertime and under a greenhouse-gas only forcing (Fig. 6-2).

The review revealed some major knowledge gaps: from a total of 151 statements that were found on possible future circulation changes only 43 statements applied to the summer, only a few ones to the transition seasons, and only 23 statements referred to simulations that considered in addition to a greenhouse-gas forcing a forcing by aerosols (cf. Fig. 6-2).



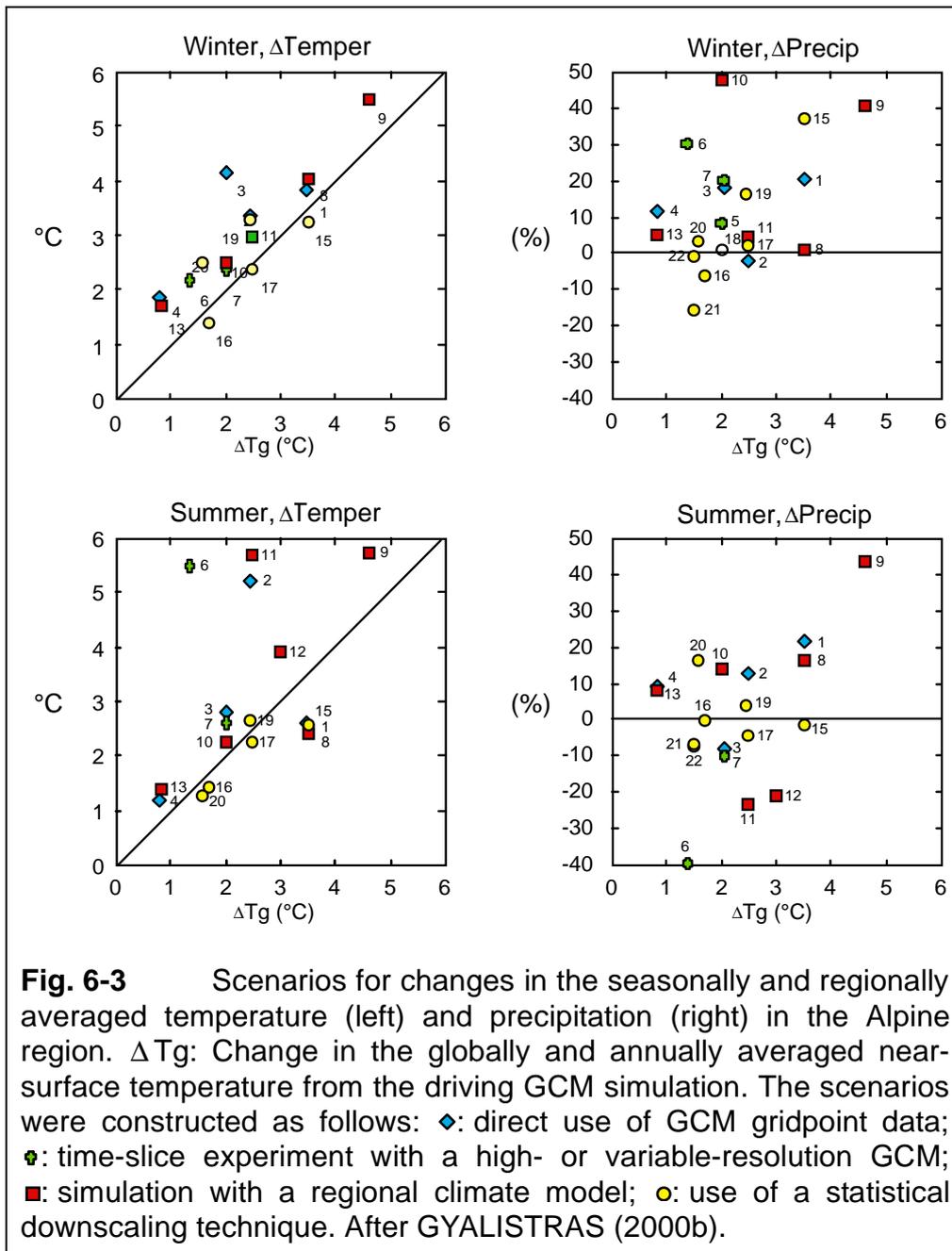
### 6.3 Scenarios of temperature and precipitation changes for the European Alps

In a second review paper, GYALISTRAS (2000b) analyzed the extant scenarios dealing with possible future temperature and precipitation changes in the European Alps. The review considered 21 scenarios which were published during the period from 1992 to early 2000.

The scenarios represented all major approaches to estimate regional climatic changes from GCMs: 4 scenarios were derived based on the direct use of GCM gridpoint data, 3 scenarios were obtained from time-slice experiments with high- or variable-resolution GCMs, 7 scenarios stemmed from simulations with regional climate models, and 8 scenarios were obtained with the aid of statistical downscaling techniques.

The scenarios showed a surprisingly large spread for the possible change in the seasonally and regionally (entire Alpine region) averaged temperature under a given global mean temperature change in the driving GCM simulations (Fig. 6-3, left).

For precipitation the scenarios did not agree on the sign of the future change. This was found to be the case for all seasons, except perhaps for winter (Fig. 6-3, right).



Further the review showed that the uncertainties introduced due to the use of different forcing scenarios, GCMs and regionalization procedures were in the same order of magnitude.

Similarly to the European circulation scenarios, the Alpine scenarios were found to reflect but a limited sample of GCMs and global forcing scenarios: All regional scenarios were produced from only 8 global simulations with 6 different GCMs. The newest GCM simulations use to produce the scenarios dated back to 1997. Only 5 scenarios took the aerosol forcing into account, and this was always the case based on one and the same global simulation.

## 6.4 Which scenarios are most trustworthy?

The future global socio-economic development is fundamentally uncertain such that it is not possible to attach any objective probabilities to individual global forcing scenarios or the associated climate change projections (NAKICENOVIC et al., 2000). Nevertheless, given a specific forcing scenario the question arises which of several available regional climate scenarios should be trusted more and which less.

GCMs are being constantly improved such that usually scenarios that are based on newer versions of the same climate model can be considered more trustworthy than older ones. However, beyond that, the intercomparison of scenarios which were derived using different global models or regionalization techniques poses many difficulties.

One main reason lies in the complexity of the climate system. As a consequence, the models may produce diverging projections due to subtle differences in the evaluation of the balance between different, counteracting factors. The contradictory changes obtained with regard to some European circulation features (Fig. 6-1) provide an example:

The diverging results have to do with the fact that most GCM scenario runs show a decrease in the equator-to-pole temperature gradient near the earth's surface and an increase in the upper troposphere. This leads to a decrease in baroclinicity at lower levels and an increase aloft. In some models the first effect dominates, thus leading to a less stormy climate, whereas in other models the second effect is more important, thus leading to the opposite result.

A second reason relates to the fact that under global warming most GCMs simulate a general increase in the water vapour content of the atmosphere. This contributes on the one hand to a more efficient poleward energy transport by the atmosphere, but on the other hand it also increases the latent energy available for the formation of cyclones. In some GCMs the first effect is responsible for a slowing down the circulation and a reduction in the number of cyclones. In other models, however, the second effect dominates and gives rise to an increased number and/or intensity of cyclones (see also the more detailed discussion in GYALISTRAS, 2000a).

Similar problems occur when one attempts to compare the outcome of different regionalization techniques. In some instances the differences between the various regional scenarios can be traced back to plausible causes, such as major differences in the driving GCM simulations. However, in many cases the complexity of the regionalization procedures, as well as limitations of the observational or modelled data base make a more detailed evaluation and comparison of the available scenarios very difficult (GYALISTRAS, 2000b).

In some instances the reasons for the simulated changes are not even properly understood. For example, the result shown in Fig. 6-3 can be explained to some extent by the above-mentioned reduction of the equator-to-pole temperature gradient near the earth's surface, and possibly also by the asymmetric warming of the two hemispheres. Some theoretical arguments can also be put forward (see discussion in

GYALISTRAS, 2000a). However, to our knowledge, the causes for a possible northward shift of the circulation have not been investigated in more detail up to now.

Methodical problems also aggravate the assessment of the scenarios. For instance many climate change simulations extend only over relatively short periods of time, which makes it difficult to distinguish systematic effects from natural variability in the simulations; or the considered forcings, temporal and spatial windows, and analysis procedures vary strongly across studies (GYALISTRAS, 2000a, b).

In summary, due to several fundamental and methodical problems, our ability to distinguish between more and less realistic regional scenarios for a given global forcing is at present very limited. Every GCM or regionalization technique has its specific advantages and limitations, and different models may evaluate the regional effects of the same basic changes in the climate system in equally plausible, yet widely differing ways.

In this situation the following two extreme possibilities have to be envisaged with regard to regional climate change, or at least selected aspects of it:

1. The contradictory scenarios are all equally appropriate because the long-term regional-scale response of the climate system is fundamentally indeterminable.
2. The true system's regional response is actually robust, but several of the used models or procedures are wrong.

The reality could lie somewhere inbetween: future changes in, say, the intensity of the North Atlantic/European storm track or the Alpine summertime precipitation could indeed depend very sensitively on the details of the future radiative forcing, the dynamics of small-scale processes, chance events, or a combination of several of these factors. At the same time some models may simply be too coarse or too incomplete to capture a possible deterministic signal correctly.

Extended model validation and sensitivity studies, as well as rigorous comparisons of models and regionalization techniques could help to resolve these issues and, ultimately, to determine the most realistic scenarios under a given global forcing.

## **6.5 Explosion of uncertainty in impact studies?**

Does the uncertain climate change information of the kind shown above necessarily lead to an explosion of uncertainty when one is interested in studying possible, specific impacts of climate change? We argue that this does not have to be so:

Firstly, because uncertainties in the projection of future climate vary strongly by parameter, region and time of the year, such that no general statements can be made. For example, as discussed above, present-day GCMs yield contradictory results with regard to possible changes in the frequency or intensity of mid-latitude storms over the North Atlantic and Europe. However, a northward shift of atmospheric circulation patterns (Fig. 6-3) would suggest for the Alps – which are located south of the tail end

of the storm track – a less stormy climate, quite independently from a possible increase or decrease of storminess in more northern regions.

A second reason why uncertainties do not necessarily have to pile up is that the impacted system may be relatively insensitive to climate change, or at least to the changes portrayed in the relevant scenarios of future climate change. For example, extended simulations with a detailed snow model showed that snow cover in Switzerland at locations below ca. 2700 m.a.s.l. is generally temperature-limited. As a result, sensitivity studies with the snow model yielded for these locations always a decrease of snow depth and duration, independently of the exact formulation of the warming scenario used to drive the model (GYALISTRAS et al., in prep.).

Of course, most systems impacted by climate are very complex, such that their response has to be evaluated for each region and impact of interest anew, and this may lead to other results than in the above example.

## **6.6 Should the main goal be to reduce the uncertainties?**

A second question that arises in view of the large uncertainties in the scenarios is: Should the main goal of climate and climate impact researchers be to reduce these uncertainties?

I think that the answer should be no, for several reasons: Firstly, because due to the complexity of the socio-economic and climatic systems the potential for surprise is, and will always remain, very large. Accordingly, strategies for dealing with an uncertain climatic future have to be elaborated anyway. Secondly, because the available time and means for adaptation to climate change are limited, such that there is a need for action in spite of the many uncertainties present. And last but not least, because non-climatic factors can be often as important as climatic influences when a specific impact is being considered.

The last point is nicely illustrated by the study of BÜRKI (2000): According to his study, Swiss winter tourism industry makes substantial investments in order to maintain or increase its supply for downhill skiing facilities. At the same time it was found that tourists are likely to respond to less favourable snow conditions with reduced demand for skiing. This could lead to a disastrous competition between ski resorts, independent of the details of future climate change (BÜRKI, 2000). In this case it seems more important to focus on the development and implementation of sound adaptation strategies rather than to put the main effort into a further reduction of purely climatological uncertainties.

## **6.7 Conclusions**

Except for the sign of the temperature change and perhaps a general northward shift of major atmospheric circulation patterns the climate's evolution in Europe and the Alps must be considered very uncertain. However, given a sufficiently large number of

scenario studies it is at least possible to give a quantitative estimate of the uncertainty range (Fig. 6-3).

There are major gaps in our knowledge, e.g. regarding the climatic effects of sulfate aerosols, the modelling of the global and regional climate, and the form and magnitude of regional climatic changes in the transition seasons systems. The lack of relevant information is probably even more acute for world regions which have not been studied as intensively as Europe.

There are also major methodical problems which indicate a clear need for systematic intercomparisons of models, scenario construction methods and scenarios. Corresponding projects should explore a wide range of global forcing scenarios, climate models, and regionalization procedures using standardized analysis tools.

Climatologists and climate impact analysts should not only aim at minimizing uncertainties, but also at maximizing the robustness of their results. The production of robust results requires that a wide range of climate scenarios is constructed and used to study the likely impacts. Robust results are per definition relatively insensitive to the input assumptions used to generate them and may therefore show very large uncertainty bounds.

Nevertheless, the establishment of large uncertainties does not necessarily present a trivial result because it is impossible to predict the outcome of a complex, state-of-the-art, quantitative assessment before this assessment has actually been carried out. In the end, the value of uncertain information must also be judged by the way it is perceived and used outside the field of climate research.

## 6.8 References

- Bürki, R. 2000: Klimaänderung und Anpassungsprozesse im Wintertourismus. PhD thesis, Department of Geography, University of Zurich, 206 pp. (Also available as Publikation der Ostschweizerischen Geographischen Gesellschaft, Neue Folge, Heft 6).
- Carter, T. R., Hulme, M., Viner, D. (eds.) 1999: Representing uncertainty in climate change scenarios and impact studies. ECLAT-2 Report No. 1, Helsinki workshop, 14-16 April, 1999, CRU, Norwich, UK, 128 pp.
- Gyalistras, D. 2000a: Wie wird sich die atmosphärische Zirkulation im Raum Nordatlantik-Europa verändern? In: Wanner, H., Gyalistras, D., Luterbacher, J., Rickli, R., Salvisberg, E. & Schmutz, C.: *Klimawandel im Schweizer Alpenraum*. vdf, Hochschulverlag AG an der ETH Zürich, pp 163-184.
- Gyalistras, D. 2000b: Klimaszenarien für den Alpenraum und die Schweiz: Neuester Stand und Vergleich. In: Wanner, H., Gyalistras, D., Luterbacher, J., Rickli, R., Salvisberg, E. & Schmutz, C.: *Klimawandel im Schweizer Alpenraum*. vdf, Hochschulverlag AG an der ETH Zürich, pp 197-235.
- Gyalistras, D., Wahrenberger, C., Rohrer, M., Lorenzi, D., Bürki, R., Abegg, B.: Derivation of future snow cover scenarios for Switzerland using spatial and temporal

downscaling of GCM results in combination with a hourly snow cover model (in preparation).

IPCC-TGCIA 1999: Guidelines on the use of scenario data for climate impact and adaptation assessment. Version 1. Prepared by Carter, T.R., Hulme, M. and Lal, M. Intergovernmental Panel on Climate Change, Task Group on Scenarios for Climate Impact Assessment, 69 pp.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grübler, A., Jung, T. Y., Kram, T., La Rovere, E. L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Raihi, K., Roehrl, A., Rogner, H.-H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., Dadi, Z. 2000: IPCC Special Report on Emissions Scenarios. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 599 pp.



## Liste der Autoren

PD Dr. habil. Friedrich-Wilhelm Gerstengarbe  
Potsdam-Institut für Klimafolgenforschung, Telegrafenberg A51,  
Postfach 60 12 03, 14412 Potsdam  
E-mail: [gerstengarbe@pik-potsdam.de](mailto:gerstengarbe@pik-potsdam.de)  
<http://www.pik-potsdam.de>

Prof. Dr. Christian-D. Schönwiese  
Institut für Meteorologie und Geophysik, Johann Wolfgang Goethe-Universität,  
Frankfurt a. Main, Postfach 11 19 32, 60054 Frankfurt a. Main,  
E-mail: [schoenwiese@meteor.uni-frankfurt.de](mailto:schoenwiese@meteor.uni-frankfurt.de),  
<http://www.rz.uni-frankfurt.de/IMGF/meteor/klima/>

PD Dr. habil. Peter C. Werner  
Potsdam-Institut für Klimafolgenforschung, Telegrafenberg A51,  
Postfach 60 12 03, 14412 Potsdam  
E-mail: [PeterC.Werner@pik-potsdam.de](mailto:PeterC.Werner@pik-potsdam.de)  
<http://www.pik-potsdam.de>

Prof. Dr. Jens Taubenheim  
Institut für Meteorologie der Freien Universität Berlin  
Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin

Prof. Dr. Andreas Hense  
Meteorologisches Institut der Universität Bonn, Auf dem Hügel 20, 53121 Bonn  
E-mail: [ahense@uni-bonn.de](mailto:ahense@uni-bonn.de)  
<http://www.meteo.uni-bonn.de>

Dr. Dimitrios Gyalistras  
Climatology and Meteorology, Institute of Geography, University of Bern,  
Hallerstr. 12, CH-3012 Bern, Switzerland  
E-mail: [gyalistras@giub.unibe.ch](mailto:gyalistras@giub.unibe.ch)  
<http://www.giub.unibe.ch/~gyalistr>



PIK Report-Reference:

- No. 1 3. Deutsche Klimatagung, Potsdam 11.-14. April 1994  
Tagungsband der Vorträge und Poster (April 1994)
- No. 2 Extremer Nordsommer '92  
Meteorologische Ausprägung, Wirkungen auf naturnahe und vom Menschen beeinflusste Ökosysteme, gesellschaftliche Perzeption und situationsbezogene politisch-administrative bzw. individuelle Maßnahmen (Vol. 1 - Vol. 4)  
H.-J. Schellnhuber, W. Enke, M. Flechsig (Mai 1994)
- No. 3 Using Plant Functional Types in a Global Vegetation Model  
W. Cramer (September 1994)
- No. 4 Interannual variability of Central European climate parameters and their relation to the large-scale circulation  
P. C. Werner (Oktober 1994)
- No. 5 Coupling Global Models of Vegetation Structure and Ecosystem Processes - An Example from Arctic and Boreal Ecosystems  
M. Plöchl, W. Cramer (Oktober 1994)
- No. 6 The use of a European forest model in North America: A study of ecosystem response to climate gradients  
H. Bugmann, A. Solomon (Mai 1995)
- No. 7 A comparison of forest gap models: Model structure and behaviour  
H. Bugmann, Y. Xiaodong, M. T. Sykes, Ph. Martin, M. Lindner, P. V. Desanker, S. G. Cumming (Mai 1995)
- No. 8 Simulating forest dynamics in complex topography using gridded climatic data  
H. Bugmann, A. Fischlin (Mai 1995)
- No. 9 Application of two forest succession models at sites in Northeast Germany  
P. Lasch, M. Lindner (Juni 1995)
- No. 10 Application of a forest succession model to a continentality gradient through Central Europe  
M. Lindner, P. Lasch, W. Cramer (Juni 1995)
- No. 11 Possible Impacts of global warming on tundra and boreal forest ecosystems - Comparison of some biogeochemical models  
M. Plöchl, W. Cramer (Juni 1995)
- No. 12 Wirkung von Klimaveränderungen auf Waldökosysteme  
P. Lasch, M. Lindner (August 1995)
- No. 13 MOSES - Modellierung und Simulation ökologischer Systeme - Eine Sprachbeschreibung mit Anwendungsbeispielen  
V. Wenzel, M. Kücken, M. Flechsig (Dezember 1995)
- No. 14 TOYS - Materials to the Brandenburg biosphere model / GAIA  
Part 1 - Simple models of the "Climate + Biosphere" system  
Yu. Svirezhev (ed.), A. Block, W. v. Bloh, V. Brovkin, A. Ganopolski, V. Petoukhov, V. Razzhevaikin (Januar 1996)
- No. 15 Änderung von Hochwassercharakteristiken im Zusammenhang mit Klimaänderungen - Stand der Forschung  
A. Bronstert (April 1996)
- No. 16 Entwicklung eines Instruments zur Unterstützung der klimapolitischen Entscheidungsfindung  
M. Leimbach (Mai 1996)
- No. 17 Hochwasser in Deutschland unter Aspekten globaler Veränderungen - Bericht über das DFG-Rundgespräch am 9. Oktober 1995 in Potsdam  
A. Bronstert (ed.) (Juni 1996)
- No. 18 Integrated modelling of hydrology and water quality in mesoscale watersheds  
V. Krysanova, D.-I. Müller-Wohlfeil, A. Becker (Juli 1996)
- No. 19 Identification of vulnerable subregions in the Elbe drainage basin under global change impact  
V. Krysanova, D.-I. Müller-Wohlfeil, W. Cramer, A. Becker (Juli 1996)
- No. 20 Simulation of soil moisture patterns using a topography-based model at different scales  
D.-I. Müller-Wohlfeil, W. Lahmer, W. Cramer, V. Krysanova (Juli 1996)
- No. 21 International relations and global climate change  
D. Sprinz, U. Luterbacher (1st ed. July, 2nd ed. December 1996)
- No. 22 Modelling the possible impact of climate change on broad-scale vegetation structure - examples from Northern Europe  
W. Cramer (August 1996)

- No. 23 A methode to estimate the statistical security for cluster separation  
F.-W. Gerstengarbe, P.C. Werner (Oktober 1996)
- No. 24 Improving the behaviour of forest gap models along drought gradients  
H. Bugmann, W. Cramer (Januar 1997)
- No. 25 The development of climate scenarios  
P.C. Werner, F.-W. Gerstengarbe (Januar 1997)
- No. 26 On the Influence of Southern Hemisphere Winds on North Atlantic Deep Water Flow  
S. Rahmstorf, M. H. England (Januar 1977)
- No. 27 Integrated systems analysis at PIK: A brief epistemology  
A. Bronstert, V. Brovkin, M. Krol, M. Lüdeke, G. Petschel-Held, Yu. Svirezhev, V. Wenzel (März 1997)
- No. 28 Implementing carbon mitigation measures in the forestry sector - A review  
M. Lindner (Mai 1997)
- No. 29 Implementation of a Parallel Version of a Regional Climate Model  
M. Kücken, U. Schättler (Oktober 1997)
- No. 30 Comparing global models of terrestrial net primary productivity (NPP): Overview and key results  
W. Cramer, D. W. Kicklighter, A. Bondeau, B. Moore III, G. Churkina, A. Ruimy, A. Schloss, participants of "Potsdam '95" (Oktober 1997)
- No. 31 Comparing global models of terrestrial net primary productivity (NPP): Analysis of the seasonal behaviour of NPP, LAI, FPAR along climatic gradients across ecotones  
A. Bondeau, J. Kaduk, D. W. Kicklighter, participants of "Potsdam '95" (Oktober 1997)
- No. 32 Evaluation of the physiologically-based forest growth model FORSANA  
R. Grote, M. Erhard, F. Suckow (November 1997)
- No. 33 Modelling the Global Carbon Cycle for the Past and Future Evolution of the Earth System  
S. Franck, K. Kossacki, Ch. Bounama (Dezember 1997)
- No. 34 Simulation of the global bio-geophysical interactions during the Last Glacial Maximum  
C. Kubatzki, M. Claussen (Januar 1998)
- No. 35 CLIMBER-2: A climate system model of intermediate complexity. Part I: Model description and performance for present climate  
V. Petoukhov, A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, S. Rahmstorf (Februar 1998)
- No. 36 Geocybernetics: Controlling a rather complex dynamical system under uncertainty  
H.-J. Schellnhuber, J. Kropp (Februar 1998)
- No. 37 Untersuchung der Auswirkungen erhöhter atmosphärischer CO<sub>2</sub>-Konzentrationen auf Weizenbestände des Free-Air Carbon dioxide Enrichment (FACE) - Experimentes Maricopa (USA)  
Th. Kartschall, S. Grossman, P. Michaelis, F. Wechsung, J. Gräfe, K. Waloszczyk, G. Wechsung, E. Blum, M. Blum (Februar 1998)
- No. 38 Die Berücksichtigung natürlicher Störungen in der Vegetationsdynamik verschiedener Klimagebiete  
K. Thonicke (Februar 1998)
- No. 39 Decadal Variability of the Thermohaline Ocean Circulation  
S. Rahmstorf (März 1998)
- No. 40 SANA-Project results and PIK contributions  
K. Bellmann, M. Erhard, M. Flechsig, R. Grote, F. Suckow (März 1998)
- No. 41 Umwelt und Sicherheit: Die Rolle von Umweltschwellenwerten in der empirisch-quantitativen Modellierung  
D. F. Sprinz (März 1998)
- No. 42 Reversing Course: Germany's Response to the Challenge of Transboundary Air Pollution  
D. F. Sprinz, A. Wahl (März 1998)
- No. 43 Modellierung des Wasser- und Stofftransportes in großen Einzugsgebieten. Zusammenstellung der Beiträge des Workshops am 15. Dezember 1997 in Potsdam  
A. Bronstert, V. Krysanova, A. Schröder, A. Becker, H.-R. Bork (eds.) (April 1998)
- No. 44 Capabilities and Limitations of Physically Based Hydrological Modelling on the Hillslope Scale  
A. Bronstert (April 1998)
- No. 45 Sensitivity Analysis of a Forest Gap Model Concerning Current and Future Climate Variability  
P. Lasch, F. Suckow, G. Bürger, M. Lindner (Juli 1998)

- No. 46 Wirkung von Klimaveränderungen in mitteleuropäischen Wirtschaftswäldern  
M. Lindner (Juli 1998)
- No. 47 SPRINT-S: A Parallelization Tool for Experiments with Simulation Models  
M. Flechsig (Juli 1998)
- No. 48 The Odra/Oder Flood in Summer 1997: Proceedings of the European Expert Meeting in  
Potsdam, 18 May 1998  
A. Bronstert, A. Ghazi, J. Hladny, Z. Kundzewicz, L. Menzel (eds.) (September 1998)
- No. 49 Struktur, Aufbau und statistische Programmbibliothek der meteorologischen Datenbank am  
Potsdam-Institut für Klimafolgenforschung  
H. Österle, J. Glauer, M. Denhard (Januar 1999)
- No. 50 The complete non-hierarchical cluster analysis  
F.-W. Gerstengarbe, P. C. Werner (Januar 1999)
- No. 51 Struktur der Amplitudengleichung des Klimas  
A. Hauschild (April 1999)
- No. 52 Measuring the Effectiveness of International Environmental Regimes  
C. Helm, D. F. Sprinz (Mai 1999)
- No. 53 Untersuchung der Auswirkungen erhöhter atmosphärischer CO<sub>2</sub>-Konzentrationen innerhalb des  
Free-Air Carbon Dioxide Enrichment-Experimentes: Ableitung allgemeiner Modelllösungen  
Th. Kartschall, J. Gräfe, P. Michaelis, K. Waloszczyk, S. Grossman-Clarke (Juni 1999)
- No. 54 Flächenhafte Modellierung der Evapotranspiration mit TRAIN  
L. Menzel (August 1999)
- No. 55 Dry atmosphere asymptotics  
N. Botta, R. Klein, A. Almgren (September 1999)
- No. 56 Wachstum von Kiefern-Ökosystemen in Abhängigkeit von Klima und Stoffeintrag - Eine  
regionale Fallstudie auf Landschaftsebene  
M. Erhard (Dezember 1999)
- No. 57 Response of a River Catchment to Climatic Change: Application of Expanded Downscaling to  
Northern Germany  
D.-I. Müller-Wohlfel, G. Bürger, W. Lahmer (Januar 2000)
- No. 58 Der "Index of Sustainable Economic Welfare" und die Neuen Bundesländer in der  
Übergangsphase  
V. Wenzel, N. Herrmann (Februar 2000)
- No. 59 Weather Impacts on Natural, Social and Economic Systems (WISE, ENV4-CT97-0448)  
German report  
M. Flechsig, K. Gerlinger, N. Herrmann, R. J. T. Klein, M. Schneider, H. Sterr, H.-J. Schellnhuber  
(Mai 2000)
- No. 60 The Need for De-Aliasing in a Chebyshev Pseudo-Spectral Method  
M. Uhlmann (Juni 2000)
- No. 61 National and Regional Climate Change Impact Assessments in the Forestry Sector  
- Workshop Summary and Abstracts of Oral and Poster Presentations  
M. Lindner (ed.) (Juli 2000)
- No. 62 Bewertung ausgewählter Waldfunktionen unter Klimaänderung in Brandenburg  
A. Wenzel (August 2000)
- No. 63 Eine Methode zur Validierung von Klimamodellen für die Klimawirkungsforschung hinsichtlich  
der Wiedergabe extremer Ereignisse  
U. Böhm (September 2000)
- No. 64 Die Wirkung von erhöhten atmosphärischen CO<sub>2</sub>-Konzentrationen auf die Transpiration eines  
Weizenbestandes unter Berücksichtigung von Wasser- und Stickstofflimitierung  
S. Grossman-Clarke (September 2000)
- No. 65 European Conference on Advances in Flood Research, Proceedings, (Vol. 1 - Vol. 2)  
A. Bronstert, Ch. Bismuth, L. Menzel (eds.) (November 2000)
- No. 66 The Rising Tide of Green Unilateralism in World Trade Law - Options for Reconciling the  
Emerging North-South Conflict  
F. Biermann (Dezember 2000)
- No. 67 Coupling Distributed Fortran Applications Using C++ Wrappers and the CORBA Sequence  
Type  
Th. Slawig (Dezember 2000)
- No. 68 A Parallel Algorithm for the Discrete Orthogonal Wavelet Transform  
M. Uhlmann (Dezember 2000)

- No. 69 SWIM (Soil and Water Integrated Model), User Manual  
V. Krysanova, F. Wechsung, J. Arnold, R. Srinivasan, J. Williams (Dezember 2000)
- No. 70 Stakeholder Successes in Global Environmental Management, Report of Workshop,  
Potsdam, 8 December 2000  
M. Welp (ed.) (April 2001)
- No. 71 GIS-gestützte Analyse globaler Muster anthropogener Waldschädigung - Eine sektorale  
Anwendung des Syndromkonzepts  
M. Cassel-Gintz (Juni 2001)
- No. 72 Wavelets Based on Legendre Polynomials  
J. Fröhlich, M. Uhlmann (Juli 2001)
- No. 73 Der Einfluß der Landnutzung auf Verdunstung und Grundwasserneubildung - Modellierungen  
und Folgerungen für das Einzugsgebiet des Glan  
D. Reichert (Juli 2001)
- No. 74 Weltumweltpolitik - Global Change als Herausforderung für die deutsche Politikwissenschaft  
F. Biermann, K. Dingwerth (Dezember 2001)
- No. 75 Angewandte Statistik - PIK-Weiterbildungsseminar 2000/2001  
F.-W. Gerstengarbe (Hrsg.) (März 2002)