

**Theoretische Meteorologie IIIa:
Statistik und Stochastik**

J. Egger

THEORETISCHE METEOROLOGIE IIIA: STATISTIK UND STOCHASTIK	1
Vorbemerkung	3
1. Grundbegriffe	4
1.1. Nachweis des Treibhauseffekts?	4
1.2 Grundbegriffe	5
1.3 Verteilungen	10
1.4 Übergangswahrscheinlichkeiten	21
2. Statistische Schätzungen und Tests	24
2.1. Stichprobenverteilung und Schätzung	25
2.2. Prinzip der maximalen Wahrscheinlichkeit (maximum likelihood).	26
2.3. Statistische Signifikanztests	27
3. Stochastische Prozesse und Differentialgleichungen	32
3.1. Grundbegriffe	32
3.2. Deckschicht des Ozeans	34
3.3. Die Liouville-Gleichung	35
3.4. Zufallsgang und Langevin-Gleichung	37
3.5. Fokker-Planck-Gleichung	38
3.6 Diskrete Markovprozesse	43
3.7. Autokovarianz stochastischer Prozesse	44
4. Zeitreihenanalyse	45
4.1. Fourieranalyse	46
4.1.b) Fourierintegrale	49
4.1. c) Stichprobenfehler und Verfälschung (Aliasing)	52
4.2. Powerspektrum eines stochastischen Prozesses	53
4.3 Kreuzkorrelation und Kreuzspektrum	59
4.4. Empirische Orthogonalfunktionen (EOF)	67
4.6. Hauptschwingungsmuster (POP)	72
5. Stochastische Vorhersage	76
5.1. Regressionsmodelle	77
5.2. Analogvorhersage	80
5.3. Statistisch-dynamische Vorhersage	82
5.4. Monte-Carlo-Vorhersage	83
6. Epilog: Nachweis des Treibhauseffekts?	87

Vorbemerkung

Statistische Verfahren gehören zum Grundbestand physikalischer Methoden. Ihr Einsatz im nichtatomaren Bereich wird insbesondere erzwungen durch Meßfehler, deren Einfluß auf das Resultat einer Messung abgeschätzt werden muß und durch die Endlichkeit des Stichprobenumfangs. Da man nur endlich viele Messungen ausführen kann, deren Ergebnisse sowohl aufgrund der Meßfehler als auch wegen störender Einflüsse alle etwas voneinander abweichen, muß die Unsicherheit des Resultats durch statistische Verfahren eingegrenzt werden. Davon werden auch Vorhersagen betroffen. Wenn der Zustand eines Systems nicht genau bekannt ist, darf eine exakte Vorhersage nicht erwartet werden. Folglich sollten die Vorhersagegleichungen stochastischer Natur sein und eine Vorhersage der Ungenauigkeit des Resultats einschließen. Diese Bemerkungen treffen, wie gesagt, auf die gesamte Physik zu, doch im besonderen Maße auf die Meteorologie. In weiten Bereichen der Physik hat man die Möglichkeit Fehlergrenzen zu senken, indem man das Experiment oft wiederholt und störende Nebeneffekte ausblendet. In der Meteorologie, die es ja mit einem turbulenten Medium zu tun hat, hat man diese Möglichkeit fast nie. Üblicherweise werden nicht Experimente durchgeführt, in denen der Experimentator den Ablauf weitgehend kontrolliert, sondern die Messung sucht einen unkontrollierten Vorgang zu erfassen, bei dem Wiederholung ausgeschlossen ist. Kein Frontdurchgang gleicht dem nächsten, so daß die Unsicherheit bis in die Definition der Phänomene reicht. Nur gelegentlich kann man eine große Zahl von Beobachtungen ausführen und lange Zeitreihen anlegen. Es versteht sich, daß bei der Auswertung dieser Beobachtungsreihen die Statistik einen zentralen Platz einnehmen sollte. Auch ist die Festlegung des Anfangszustandes für eine numerische Wettervorhersage mit beträchtlichen Unsicherheiten behaftet. Es versteht sich, daß die Gleichungen der meteorologischen Dynamik diesen Unsicherheiten Rechnung tragen sollten.

Die Vorlesung gibt eine Einführung in die Grundkonzepte statistischer Methoden, mit denen man den geschilderten Schwierigkeiten entgegentritt. Dabei werden im ersten Kapitel Grundbegriffe erläutert und wichtige Verteilungen von Zufallsgrößen präsentiert, die auch auf meteorologische Größen angewandt werden können. Im zweiten Kapitel werden Entscheidungshilfen bereitgestellt etwa für Fragen der Art, ob der vergangene letzte Januar anomal kalt war. Bis zu diesem Punkt handelt die Vorlesung nur von der Datenbearbeitung. Dem müßte gleichberechtigt ein Abschnitt gegenüberstehen, in dem dargetan wird, wie man statistische Elemente in die Gleichungen der Theoretischen Meteorologie einbringt, die ja in den bisherigen Vorlesungen nur in ihrer deterministischen Fassung benutzt worden sind. Doch kann die Theorie stochastischer Differentialgleichungen (Kapitel 3) nur sehr summarisch behandelt werden, da eine halbwegs adäquate Darstellung den Rahmen dieser Vorlesung sprengen würde. Den Hauptteil des 4. Kapitels nimmt die Anwendung der Fourieranalyse auf stochastische Prozesse ein. Schließlich wird im letzten Kapitel das Gebiet der stochastischen Vorhersage berührt.

1. Grundbegriffe

1.1. Nachweis des Treibhauseffekts?

Aufgrund von Modellrechnungen ist man zu dem Schluß gelangt, daß der Ausstoß von Treibhausgasen in der industrialisierten Welt zu einer Erwärmung der Atmosphäre führen sollte. Die Frage ist, ob sich solch eine Erwärmung auch nachweisen läßt. In Abb.1. sind die beobachteten Änderungen der globalen Temperatur der Atmosphäre von 1860 bis heute entsprechenden Modellrechnungen gegenübergestellt. Das Modell ist so angelegt, daß der Anstieg der Temperatur eindeutig auf die Zunahme der Treibhausgase zurückgeht. Dabei sei nicht auf die Problematik eingegangen, die mit der Auswertung der Beobachtungen verknüpft ist und auch nicht auf die mit der Modellierung verknüpften Probleme. Man sieht, daß die Atmosphärentemperatur seit der Jahrhundertwende um ca. 0.5° C angestiegen ist. Der Anstieg liegt eher etwas unter den errechneten. Die Frage ist, ob dieser beobachtete Anstieg durch die zusätzlichen Treibhausgase hervorgerufen worden ist. Zunächst sieht es so aus als habe man hier die klassische Situation des Vergleichs von Theorie und Beobachtung. Wenn etwa anhand der Gesetze der Himmelsmechanik eine Sonnenfinsternis vorhergesagt wird, so ist die Verifizierung dieser Vorhersage eine einfache Sache. Man könnte geneigt sein, die ja passable Übereinstimmung von Theorie und Beobachtung auch in diesem Sinne zu deuten. Hier muß einen allerdings bedenklich stimmen, daß das Modell keine einzige der vielen seit 1860 beobachteten Schwankungen der Temperatur erfaßt, sondern bestenfalls den gesamten langsamen Anstieg der Temperatur. Warum soll man gerade dieser Aussage des Modells glauben? Hier kommt die Eigenheit des Klimasystems ins Spiel, daß Details der Entwicklung nicht vorhersagbar sind. Bekanntlich hat die Wettervorhersage bei etwa fünf Tagen eine Grenze. Ebenso ist es unmöglich langfristige Temperaturschwankungen, wie in Abb. 1 zu sehen, vorherzusagen oder auch, nachdem sie aufgetreten sind, numerisch nachzuvollziehen.

Wo kommen hier Wahrscheinlichkeit und Statistik zum Zuge? Die Antwort muß sein, daß in einem System, das nicht vorhersagbar ist, nur mit statistischen Methoden Aussagen gemacht werden können. Insbesondere würde man für das Treibhausproblem gern die (hypothetische) Strategie anwenden, daß man die Klimaentwicklung, ausgehend von diversen Anfangszuständen für 1860, oft wiederholt, dafür aber keine Treibhausgase in die Atmosphäre entläßt. Dadurch erhielte man eine große Reihe von Temperaturkurven wie in Abb. 1. Diese würden ebenfalls beträchtliche Schwankungen aufweisen, eben "natürliche" Variabilität. Die Frage, die sich mit statistischen Methoden auch entscheiden ließe, wäre dann so zu stellen: wie groß ist die Wahrscheinlichkeit, daß bei willkürlicher Wahl einer Kurve aus dem Ensemble der Entwicklungen ohne zusätzliche Treibhausgase der mittlere Temperaturanstieg so groß ist wie in Abb. 1. Ist

diese Wahrscheinlichkeit gering, so wird man zu dem Schluß gedrängt, daß der beobachtete Anstieg tatsächlich durch die zusätzliche Emission der Treibhausgase hervorgerufen ist. Theorie und Modellbau hätten bei dieser Entscheidung eine unterstützende Rolle, da sie es ja sind, die einem sagen können, wie groß der zu erwartende Effekt in einem System ist, das keine zusätzliche Variabilität hat.

Es versteht sich, daß diese Strategie nicht verfolgt werden kann, man also versuchen muß, Aussagen über die natürliche Variabilität des Klimasystems aus Daten wie in Abb. 1 gezeigt oder aus Modellrechnungen zu gewinnen. Dazu mehr im Epilog.

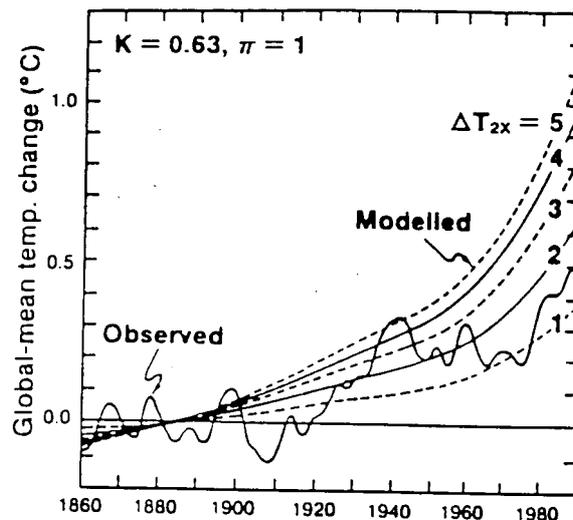


Abb. 1. Beobachtete und mit Hilfe eines einfachen Klimamodells (Wigley and Raper, Nature, 330) berechnete globale Mitteltemperatur für die Jahre 1860-1990. Die Berechnungen unterscheiden sich in der Wahl eines Modellparameters

1.2 Grundbegriffe

Betrachtet sei ein Prozeß, der durch die Anweisung

$$x_{n+1} = ax_n + b + \text{„Münzwurf“} \quad (1.1)$$

beschrieben wird. Dabei sind $0 < a < 1$ und b Konstanten. Bei jedem Schritt, der die Variable x_n zum Schritt n mit der Variablen x_{n+1} zum Schritt $n+1$ verbindet, wird eine Münze geworfen. Kommt das Wappen, so ist bei „Münzwurf“ $\gamma > 0$ einzutragen, kommt die Zahl, so $-\gamma$. Anhand von (1.1) lassen sich eine Reihe von Grundbegriffen erläutern:

1. Ereignis: man kann etwa als Ereignis S_m definieren, daß x in das Intervall $m \leq x \leq m+1$ fällt. Das Werfen der Münze ist ebenfalls ein Ereignis.

2. Zufallsvariable: x_n ist eine Zufallsvariable, die jedem Ereignis eines Münzwurfs eine Zahl zuordnet. Ebenso ist „Münzwurf“ eine Zufallsvariable, die jedem Wurf die Zahl γ oder $-\gamma$ zuordnet.

3. Wahrscheinlichkeit: man kann nach der Wahrscheinlichkeit $p(n; S_m)$ fragen, daß beim Wurf n das Ereignis S_m eintritt, daß also $m \leq x_{n+1} \leq m + 1$. Man kann dies vom speziellen Zeitschritt n entkoppeln, indem man nach der Wahrscheinlichkeit $p(S_m)$ fragt, daß bei einem beliebig gewählten Schritt das Ereignis S_m eintritt. Man kann $p(S_m)$ messen, indem man in einer langen Serie die relative Häufigkeit

$$f_m = \frac{M}{N} \quad (1.2)$$

feststellt, wobei N die Zahl der Schritte und M die Zahl der Ereignisse S_m ist.

4. Wahrscheinlichkeitsdichte: der stetigen Variablen x kann man eine Wahrscheinlichkeitsdichte $f(x)$ zuordnen, so daß

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x, t) \quad (1.3)$$

die Wahrscheinlichkeit angibt, daß x zur Zeit t ins Intervall $x_1 \leq x \leq x_2$ fällt.

Ferner muß für alle Zeit gelten

$$\int_{-\infty}^{+\infty} f(x, t) dx = 1. \quad (1.4)$$

5. Stationarität: eine Zufallsvariable oder ein Zufallsprozeß heißen stationär, wenn die zugehörigen statistischen Kerngrößen nicht von der Zeit abhängen. So ist, wie sich zeigen wird, (1.1) im wesentlichen ein stationärer Prozeß. Die Wahrscheinlichkeitsdichte $f(x)$ für die stationäre Variable x hängt nicht von der Zeit ab.

6. Erwartungswert: für die stetige Variable x ist der Erwartungswert

$$E(x, t) = \int_{-\infty}^{\infty} xf(x, t) dx. \quad (1.5)$$

Diesen Erwartungswert errechnet man für (1.1), indem man den Operator E auf (1.1) anwendet:

$$\int_{-\infty}^{+\infty} f(x) \{x_{n+1} = ax_n + b + \text{„Münzwurf“}\} dx. \quad (1.6)$$

Der Münzwurf ist zufällig, so daß von diesem Term kein Beitrag kommen kann.

Dann folgt bei Stationarität:

$$E(x) = b/(1-a) = \mu \quad (1.7)$$

Man kann leicht sehen, daß (1.1) zumindest für große n stationär ist. Man kann zwar eine Anfangsverteilung wählen, so daß $E(x_0) \neq \mu$. Doch findet man sofort aus (1.1) mit (1.6), daß

$$E(x_{n+1} - \mu) = aE(x_n - \mu). \quad (1.8)$$

Da $0 < a < 1$ folgt, daß der Erwartungswert mit a^n sich μ nähert. Für große n ist also tatsächlich $E(x_n) = \mu$. Andererseits zeigt (1.8), daß der Prozeß (1.1) für kleine n instationär ist, sofern $E(x_0) \neq \mu$.

Ordnet man dem Ereignis S_m den Wert

$$\hat{x}_{(m)} = m + 1/2$$

zu, so kann man einen Erwartungswert auch für die diskrete Variable \hat{x} festlegen:

$$E(\hat{x}_m) = \sum_{m=-\infty}^{+\infty} p(S_m) \hat{x}_m \quad (1.9)$$

Wenn $g(x)$ eine Funktion von x ist, so ist

$$E(g(x)) = \int g(x)f(x)dx \quad (1.10)$$

Auch gilt

$$E(g_1(x) + g_2(x)) = E(g_1) + E(g_2)$$

7. Die Varianz: sie ist festgelegt gemäß

$$E((x-\mu)^2) = \sigma^2 = \int f(x)(x - \mu)^2 dx \quad (1.11)$$

Um die Varianz für (1.1) zu erhalten, quadriert man (1.1):

$$\begin{aligned} x_{n+1}^2 &= a^2 x_n^2 + 2abx_n + b^2 + (\text{Münzwurf})^2 \\ &\quad + 2(ax_n + b) \text{Münzwurf} \end{aligned} \quad (1.12)$$

Wenn man darauf den Operator E anwendet, so hat man zu beachten, daß der quadratische Münzwurfterm stets den Betrag γ^2 hat und daß der letzte Term keinen Beitrag liefert. Also folgt im stationären Fall

$$E(x^2)(1-a^2) = 2ab\mu + b^2 + \gamma^2 \quad (1.13)$$

und damit

$$E((x-\mu)^2) = (2ab\mu + b^2 + \gamma^2)/(1-a^2) - \mu^2. \quad (1.14)$$

Mit (1.7) folgt

$$\sigma^2 = \gamma^2/(1 - a^2) \quad (1.15)$$

Die Varianz ist proportional zur Varianz γ^2 des Antriebs und ist umso größer je näher a an 1 rückt.

8. Autokovarianz: man interessiert sich oft dafür wie sich eine Variable $x(t)$ typischerweise im Lauf der Zeit ändert. Dazu berechnet man die Autokovarianz

$$E((x(t) - \mu)(x(t+\tau) - \mu)) = \int f(x)(x(t) - \mu)(x(t+\tau) - \mu) dx, \quad (1.16)$$

wobei Stationarität vorausgesetzt wird. Die Größe τ heißt Verschiebung (lag). Der Prozeß (1.1) ist diskret in der Zeit, so daß hier

$$E((x_n - \mu)(x_{n+1} - \mu)) \quad (1.17)$$

als Autokovarianz zu definieren ist (l ganze Zahl). Sie läßt sich aus (1.1) errechnen. Man multipliziert (1.1) mit x_n und bilde

$$E\{x_n x_{n+1} = ax_n^2 + bx_n + \text{Münzwurf } x_n\}. \quad (1.18)$$

Es folgt wegen $E(x_n^2) = \sigma^2 + \mu^2$

$$E(x_n x_{n+1}) = a(\sigma^2 + \mu^2) + b\mu \quad (1.19)$$

und so

$$E((x_n - \mu)(x_{n+1} - \mu)) = a\sigma^2 + b\mu + (a - 1)\mu^2$$

Mit (1.15) und (1.7) erhält man

$$E((x_n - \mu)(x_{n+1} - \mu)) = \frac{a\gamma^2}{(1 - a^2)} = a\sigma^2 \quad (1.20)$$

als Autokovarianz mit Verschiebung 1. Es folgt sofort

$$E((x_n - \mu)(x_{n+1} - \mu)) = a^l \sigma^2 \quad (1.21)$$

Die Autokovarianz wird für wachsende Verschiebung l immer kleiner. Doch ist die Autokovarianz immer positiv. Dies muß nicht immer so sein. Abb. 2 zeigt den Southern Oscillation Index (SOI), der hier als Differenz des monatlichen Bodendrucks in hPa zwischen Darwin in Nordaustralien und Tahiti definiert ist. Die Zeitreihe reicht von 1870 bis zur Gegenwart. Sie ist unruhig, zeigt aber, daß dieser Index über einige Jahre hinweg das gleiche Vorzeichen hat. Die Autokorrelation des SOI läßt sich aufgrund der Daten in Abb. 2 gut schätzen. Dabei ist die

$$\text{Autokorrelation} = E((x_n - \mu)(x_{n+1} - \mu))/\sigma^2, \quad (1.22)$$

also die Autokovarianz normiert mit der Varianz. Die Autokorrelation hat für $l = 0$ automatisch den Wert 1 und fällt in Abb. 3 relativ langsam ab. Eine Nullstelle wird nach etwa einem Jahr erreicht, dann findet man einen negativen Wert. Dieser langsame

Abfall einer Korrelation ist für eine meteorologische Variable erstaunlich.

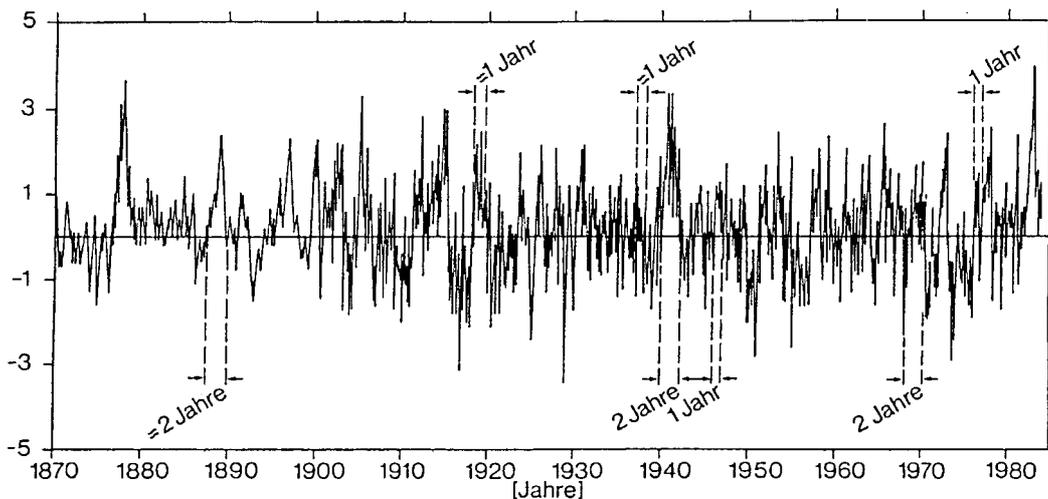


Abb. 2 Southern Oscillation Index SOI in hPa als monatliche Bodendruckdifferenz zwischen Darwin und Tahiti.

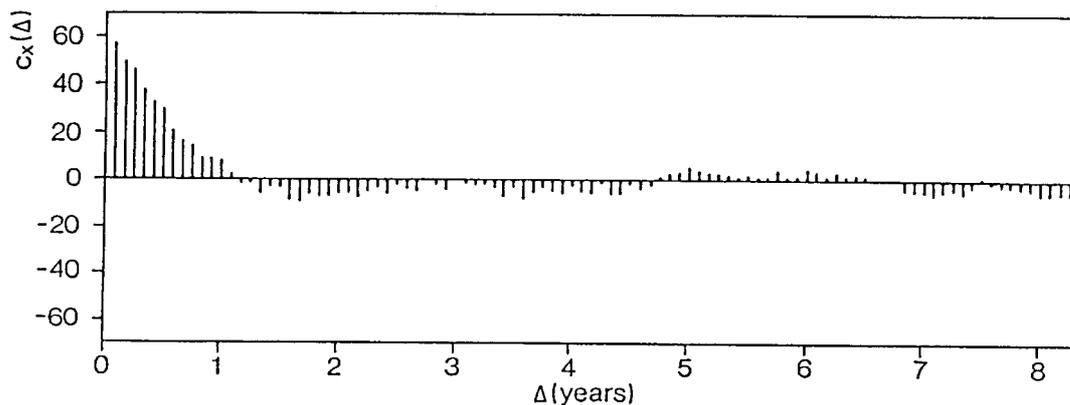


Abb. 3. Autokorrelation des SOI x 100. (Abb. 2 + 3 aus v. Storch und Zwiers 1998)

9. Die >Einzelninformation< I_k soll den Neuigkeitswert des Ereignisses E_k messen, sofern es bei einem Experiment eintritt. Ein eingetretenes Ereignis hat desto weniger Neuigkeitswert, je wahrscheinlicher es vorher war. I_k sollte also eine monoton abnehmende Funktion der Wahrscheinlichkeit p_k des Ereignisses sein. Ferner möchte man gern, daß der Neuigkeitswert des aus zwei unabhängigen Ereignissen kombinierten Ereignisses gleich der Summe ihrer Neuigkeitswerte ist. Dies führt auf die Definition der Information:

$$I_k = -\ln p_k. \quad (1.23)$$

Der Erwartungswert

$$H = \sum_{k=1}^K p_k I_k = -\sum_{k=1}^K p_k \ln p_k \quad (1.24)$$

wird als Maß der Information eingeführt und Entropie genannt. Beim Würfel ergibt sich die Entropie zu

$$H = -\sum_{k=1}^6 \frac{1}{6} \ln \frac{1}{6} = \ln 6$$

Einer Wahrscheinlichkeitsdichte $f(x)$ kann man ebenfalls eine Entropie

$$H = -\int_{-\infty}^{+\infty} f(x) \ln f(x) dx \quad (1.25)$$

zuordnen.

1.3 Verteilungen

In Abb. 4 ist als Beispiel einer aus Daten gewonnenen Häufigkeitsverteilung die Verteilung der Temperaturdifferenz Δt zwischen je zwei Stationen in Ungarn dargestellt, wobei einmal (Punkte) zwei nahe gelegene Stationen (Distanz $d=4$ km) und einmal (Kreuze) zwei weit voneinander entfernte ($d=364$ km) Stationen gewählt wurden. Wie zu erwarten, treten bei $d=364$ km auch große Temperaturunterschiede mit einiger Häufigkeit auf. De facto sind in Abb. 4 relative Häufigkeiten f_k dargestellt, doch kann man diese als Schätzwerte von Wahrscheinlichkeiten akzeptieren und gemäß (1.12) jeder von den Verteilungen eine Entropie zuordnen. Es ergibt sich $H(d=4\text{ km}) = 0,51$, $H(d=364\text{ km}) = 1,07$. Die Entropie für das weit auseinandergelegene Paar ist größer als die für die benachbarten Stationen, weil für $d=4$ km die Information durch eine Beobachtung nahezu immer klein ist, da dort fast immer $\Delta t \approx 0$.

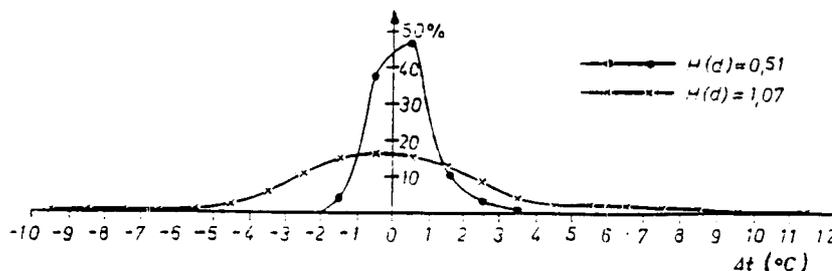


Abb. 4. Häufigkeitsverteilung des Temperaturunterschiedes für Stationspaare, einmal (Punkte) mit Abstand 4 km, einmal (Kreuze) mit Abstand (364 km).

Beobachtete Häufigkeitsverteilungen können nahezu beliebig aussehen, doch gibt es viele Fälle, wo zumindest im Limes großer Versuchszahl Verteilungen herauskommen, die sich mit Hilfe einer Formel beschreiben lassen. Die wichtigsten Fälle seien kurz vorgestellt:

a) Gleichverteilung.

Die konstante Dichtefunktion (Gleichverteilung) schreibt jedem Wert x im Intervall $a \leq x \leq b$ dieselbe Dichte $f(x) = 1/(b-a)$ zu ($f = 0$ für alle Punkte außerhalb des Intervalls).

b) Binomialverteilung:

Wenn p_1 die Wahrscheinlichkeit ausdrückt, daß in einem bestimmten Versuch ein Elementarereignis oder auch Ereignis eintritt und entsprechend $p_2 = 1-p_1$ die Wahrscheinlichkeit, daß das Ereignis nicht eintritt, dann ist die Wahrscheinlichkeit, daß das Ereignis in einer Serie von N Versuchen genau j -mal eintritt

$$p(p_1, N) = \binom{N}{j} p_1^j p_2^{N-j}, \quad j = 0, 1, 2, \dots, N. \quad (1.26)$$

Ein klassisches Beispiel für die Binomialverteilung bildet das Werfen einer Münze wie in (1.1). Die Wahrscheinlichkeit, daß das Wappen kommt, ist $p_1 = 0.5$. Der Mittelwert ist $p_1 N$ und die Varianz beträgt $N p_1 (1 - p_1)$.

c) Poisson-Verteilung:

Setzt man in der Binomialverteilung $N p_1 = \lambda$ konstant und betrachtet den Fall großer N (also kleiner Wahrscheinlichkeit p_1), so ergibt sich die Poissonverteilung

$$p(j/\lambda) = \lambda^j e^{-\lambda} / j! \quad (1.27)$$

Wenn also die Wahrscheinlichkeit $p_1 = \lambda/N$ eines Ereignisses sehr klein ist, dann beschreibt die Poisson-Verteilung in etwas besser faßlicher Form als die Binomial-Verteilung die Wahrscheinlichkeit, daß das seltene Ereignis in einer langen Serie j -mal auftritt.

d) Gaußverteilung (Normalverteilung)

Mit die wichtigste kontinuierliche Verteilung ist die Gaußverteilung

$$f(x) = 1/(\sigma\sqrt{2\pi}) \exp(-(x-\mu)^2/2\sigma^2) \quad (1.28)$$

, die vollständig durch die beiden Parameter μ und σ bestimmt ist (Symbol $N(\mu, \sigma)$). In Abb. 5 ist die Funktion $\sigma f(x)$ dargestellt und die zu f gehörige kumulierte Verteilungsfunktion $F(x)$

$$\frac{dF}{dx} = f(x) \quad (1.29)$$

wobei $F(-\infty) = 0$ gesetzt ist. Entsprechend hat man stets $F(+\infty) = 1$.

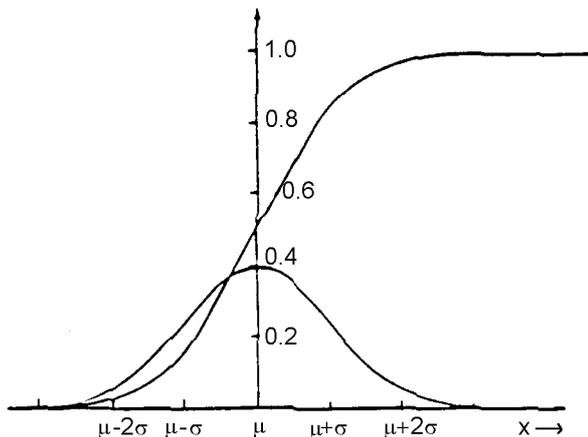


Abb. 5 $\sigma f(x) = \sigma N(\mu, \sigma)$ und die zugehörige kumulierte Verteilungsfunktion.

Eigenschaften von $N(\mu, \sigma)$:

1. Symmetrie bezüglich $x = \mu$
2. Maximum $f = (\sqrt{2\pi}\sigma)^{-1} \approx 0.4$ für $\sigma = 1$ und $x = \mu$
3. Die Standardabweichung ist durch die Abszisse der Wendepunkte gegeben.
4. Bei großen Stichprobenumfängen liegen etwa 90% aller Beobachtungen zwischen 1.64σ . Im Intervall $-0.676 \leq x \leq 0.676$ liegen 50% aller Werte. Man sieht mit Hilfe von

$$\sqrt{2\pi}^{-1} \int_0^{\infty} \exp(-x^2 / 2) dx = \frac{1}{2} \quad (1.30)$$

, daß

$$E(N(\mu, \sigma)) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{+\infty} x \exp\left(-\frac{1}{2}(x-\mu)^2 / \sigma^2\right) dx = \mu \quad (1.31)$$

also der Parameter μ der Erwartungswert von x (Mittelwert) ist.

Ferner gilt

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2\sigma^2} dx = \sigma^2 \quad (1.32)$$

, so daß σ die Standardabweichung ist.

Die Wahrscheinlichkeit $p(-\eta \leq x \leq \eta)$, daß x bei zufälliger Ziehung im Intervall $-\eta \leq x \leq \eta$ zu liegen kommt ist in Tab. 1 angegeben für die Verteilung $N(0,1)$. Hat man eine Verteilung $N(\mu, \sigma)$, so muß das Intervall um μ zentriert werden und es muß η durch $\eta\sigma$ ersetzt werden.

$p(-\eta \leq x \leq \eta)$	α : Fläche von f außerhalb	η
0.683	0.317	1.00
0.950	0.050	1.96
0.990	0.010	2.58
0.997	0.003	3.00

Tabelle 1

Die Fouriertransformierte der Normalverteilung

$$G(k) = \int_{-\infty}^{+\infty} N(\mu, \sigma) \exp(-ikx) dx = \exp\left(-\frac{1}{2}\sigma^2 k^2 - i\mu k\right) \quad (1.33)$$

ist wieder eine Exponentialfunktion, deren Exponent eine nun in k quadratische Form ist.

Die große Bedeutung der Normalverteilung ist wohl eine Folge des zentralen Grenzwertsatzes, der besagt daß die Verteilung $f(x)$ der Summe

$$x = \frac{1}{N} \sum_{i=1}^N x_i$$

von unabhängigen, identisch verteilten Zufallsvariablen x_i schnell mit zunehmendem N gegen eine Normalverteilung strebt. Dabei wird verlangt, daß die x_i dieselben μ und σ besitzen. Die x_i brauchen nicht normal verteilt zu sein.

Wenn $E(x_i) = \mu$ und $E((x_i - \mu)^2) = \sigma^2$ so gilt

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_i - \mu) / (\sigma / N^{1/2}) \sim N(0, 1) \quad (1.34)$$

Die Entropie (1.25) einer Verteilung mit Mittelwert μ und Varianz σ^2 nimmt genau dann ihren Maximalwert an, wenn $f = N(\mu, \sigma)$ normal verteilt ist.

Beweis: Variationsrechnung mit Lagrangeschen Multiplikatoren ergibt als Bedingung für Maximalwert

$$\frac{\partial}{\partial f} (-f \ln f + \lambda_1 f + \lambda_2 x f + \lambda_3 (x - \mu)^2 f) = 0.$$

Also:

$$f(x) = \exp(\lambda_1 - 1 + \lambda_2 \mu) \exp(\lambda_2 (x - \mu) + \lambda_3 (x - \mu)^2).$$

Die Bedingung (1.4) erzwingt $\lambda_2 = 0$ und führt wegen (1.30) auf

$$\int_{-\infty}^{+\infty} f(x) dx = \exp(\lambda_1 - 1) \sqrt{\frac{\pi}{-\lambda_3}} = 1.$$

Die Bedingung

$$\int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = -\exp(\lambda_1 - 1) \frac{1}{2\lambda_3} \sqrt{\frac{\pi}{-\lambda_3}} = \sigma^2$$

liefert $\lambda_3 = -\frac{1}{2\sigma^2}$ und folglich

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}(x - \mu)^2/\sigma^2\right).$$

Wenn also von einer Verteilung Mittelwert und Standardabweichung bekannt sind, so ist der Erwartungswert der Information für diese Verteilung höchstens so groß wie der der zugehörigen Normalverteilung. Die Normalverteilung läßt maximale Information erwarten.

Normalverteilungen werden oft >beobachtet< und noch öfter in theoretischen Modellen benutzt. Dafür sei die Gleichung

$$\frac{dT}{dt} + \lambda T = 0 \tag{1.35}$$

als Beispiel herangezogen. Man habe zur Zeit $t = 0$ für T eine Normalverteilung:

$$f(T,0) = \frac{1}{\sqrt{2\pi\sigma_0}} \exp(-(T)^2/2\sigma_0^2)$$

, wobei T die Abweichung von einer Mitteltemperatur und σ_0 eine bekannte Standardabweichung ist. Gesucht ist die Lösung für $t > 0$. Nun folgt aus (1.35), daß

$$\frac{dT^2}{dt} + 2\lambda T^2 = 0.$$

Man wird also vermuten, daß $\sigma^2 = \sigma_0^2 \exp(-2\lambda t)$. In der Tat heißt die Lösung, wie in Kap. 3.3 gezeigt wird,

$$f(T,t) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-T^2/2\sigma^2). \tag{1.36}$$

Die Verteilung bleibt also normal, aber die Standardabweichung nimmt mit der Zeit ab, da es sich um einen gedämpften Vorgang handelt. Der Maximalwert $f(0,t) = 0.4 \exp(\lambda t)/\sigma_0$ nimmt mit der Zeit zu, die Verteilung wird also immer höher und schmaler.

e) Bivariate Verteilungen

Man führt in Analogie zu den eindimensionalen Verteilungen auch Verteilungen für mehrere Variablen ein. Abb. 6 zeigt eine gemessene Verteilung der zonalen und vertikalen Windkomponenten am Garchinger Turm.

Das Paradebeispiel einer bivariaten Dichte ist natürlich die bivariate Normalverteilung mit

$$f(x,y) = 1/(2\pi\sigma_x\sigma_y (1 - \rho_{xy}^2)^{1/2}) \exp\left(-\frac{1}{2(1-\rho_{xy}^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right). \quad (1.37)$$

Diese Verteilung enthält 5 Parameter. Wenn $\rho_{xy} \neq 0$, so sind x und y abhängig, wenn $\rho_{xy} = 0$, so heißen x und y unabhängige Zufallsvariable. Wenn $\rho_{xy} = 0$, so gilt

$$f(x) = \int_{-\infty}^{\infty} f(x,y) dy = N(\mu_x, \sigma^2).$$

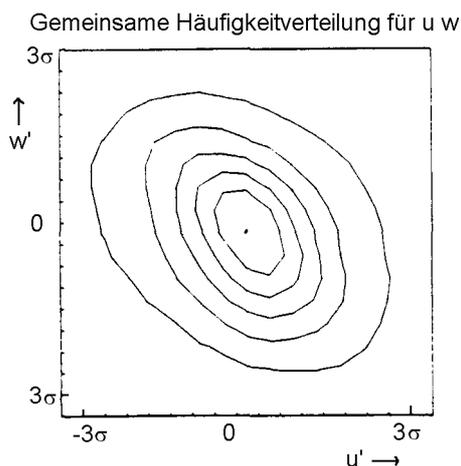


Abb. 6. Bivariate Häufigkeitsverteilung des Windes am Garchinger Turm (Hennemuth), die zeigt, daß positive (negative) Abweichungen u' des Zonalwinds bevorzugt mit negativen (positiven) Werten der Vertikalbewegung w' einhergehen.

Die Isolinien der Verteilung (1.37) sind Ellipsen mit Mittelpunkt μ_x, μ_y . Man kann die Variablen x, y durch Drehung und Streckung durch neue Variable x', y' ersetzen, die unabhängig sind.

Der Begriff der Abhängigkeit soll noch genauer gefaßt werden. Man habe zwei diskrete Zufallsvariable x, y und betrachte die Wahrscheinlichkeit, daß x den Wert x_1 annimmt unter der Bedingung, daß y den Wert y_1 annimmt. Diese bedingte Wahrscheinlichkeit wird angeschrieben als $p(x_1 | y_1)$. Die bivariate Wahrscheinlichkeitsdichte von x und y kann dann geschrieben werden

$$p(x, y) = p(y) p(x | y) = p(x) p(y | x) \quad (1.38)$$

Wenn nun das Eintreten von x in keiner Weise durch y beeinflusst ist, so ist $p(x, y) = p(x)$. In diesem Fall heißen x und y unabhängig. Analog gilt für bivariate stetige Wahrscheinlichkeitsdichten

$$f(x, y) = f(y) f(x | y) \quad (1.39)$$

und

$$f(x, y) = f(x) f(y) \quad (1.40)$$

falls x und y unabhängig sind.

Eine besonders wichtige Kenngröße bivariater Verteilungen ist die Kovarianz

$$E[(x - \mu_x)(y - \mu_y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \quad (1.41)$$

Wenn zwei Variable unabhängig sind, so verschwindet ihre Kovarianz. Für die bivariate Normalverteilung ergibt sich die Kovarianz zu $\sigma_x \sigma_y \rho_{xy}$. Der Korrelationskoeffizient ist eine normierte Kovarianz:

$$\rho_{xy} = \frac{E((x - \mu_x)(y - \mu_y))}{\left(E((x - \mu_x)^2) E((y - \mu_y)^2)\right)^{1/2}} \quad (1.42)$$

wobei nun μ_x, μ_y in Analogie zu (1.27) die zu jeder Variablen gehörigen Erwartungswerte und Varianzen sind:

$$\mu_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy \quad (1.43)$$

etc.

Normalverteilungen können auch für viele Variable

$$\mathbf{x} = (x_1, \dots, x_m)$$

definiert werden. Sei

$$\boldsymbol{\mu} = (\mu_1 \dots \mu_m)$$

der Vektor der Mittelwerte und

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} & \dots \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad (1.44)$$

die Kovarianzmatrix der Variablen, dann ist

$$N(\mathbf{x}) = (2\pi)^{-m/2} |\mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \cdot \mathbf{V}^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu})\right) \quad (1.45)$$

die zu diesem Satz von Variablen gehörige Normalverteilung.

f) Stichprobenverteilungen

Daten sind Zufallszahlen. Mißt man eine Größe X in n unabhängigen Experimenten, so erhält man die in der Regel verschiedenen Werte x_1, \dots, x_N . Der Satz der Daten $x_1 \dots x_N$ wird als Stichprobe bezeichnet. Aus einer solchen Stichprobe sind in der Regel die Dichtefunktion $f(x)$ oder andere Größen zu schätzen. Nun sei man etwa am Mittelwert interessiert. Dieser Mittelwert $\bar{x} = \sum_{i=1}^N x_i/N$ ist selber eine Zufallsvariable und hat eine

Verteilung, da ja die N Messungen beliebig oft wiederholt werden könnten und man jedesmal einen anderen Mittelwert erhielte. Diese Verteilung nennt man die Stichprobenverteilung des Mittelwerts.

Sei speziell x normal verteilt, also $f(x) = N(\mu, \sigma)$, so legt schon der zentrale Grenzwertsatz nahe, daß \bar{x} auch normal verteilt sein muß. Es gilt

$$f(\bar{x}) = \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{N})} \exp\left[-\frac{N(\bar{x} - \mu)^2}{2\sigma^2}\right] = N(\mu, \sigma/\sqrt{N}). \quad (1.46)$$

Das sollte auch dann eine gute Schätzung sein, wenn x nicht normal verteilt ist. Man sieht, daß mit Zunahme von N die Unsicherheit sinkt. Stichprobenverteilungen können nun dazu benützt werden, um Wahrscheinlichkeitsaussagen etwa über den Mittelwert zu machen. Man habe z.B. neun Messungen x_i , ($N=9$), wobei bekannt sei, daß x normal verteilt ist mit $\sigma = 1$. Dabei sei μ nicht bekannt. Dann weiß man von (1.46), daß \bar{x} normal verteilt ist mit Varianz $\frac{1}{9}$. Dann sagt Tab. 1, daß der ermittelte Wert \bar{x} mit 95% Wahrscheinlichkeit im Bereich

$$\mu - 1.96/\sqrt{9} < \bar{x} < \mu + 1.96/\sqrt{9}$$

liegt. Will man also Mittelwerte meteorologischer Variablen aus einer Stichprobe schätzen, so kommt man auf die angegebene Weise zu Aussagen über die Entfernung der geschätzten Mittelwerts vom wahren. Allerdings kennt man die Varianz gemeinhin nicht. So stellt sich die Aufgabe, die Varianz einer Zufallsvariablen anhand einer Stichprobe zu schätzen. Die χ^2 -Verteilung beschreibt die Stichprobenverteilung für eine Summe von Quadraten. Seien x_i wieder N unabhängige Meßwerte aus einer $N(0,1)$ -Gesamtheit, so kann man nach der Stichprobenverteilung

$$\chi_N^2 = \sum_{i=1}^N x_i^2 \quad (1.47)$$

fragen. Diese wird dadurch wichtig, daß man die Varianz von x ja nur über (1.47) schätzen kann. (1.47) heißt χ^2 -Verteilung mit N Freiheitsgraden. Die Lösung dieses Problems lautet

$$f_{\chi_N^2}(x) = \frac{1}{2^{N/2} \Gamma(N/2)} x^{(N/2-1)} \exp(-x/2) \quad (0 \leq x < \infty) \quad (1.48)$$

In Abb. 7 ist die Verteilung dargestellt. Für kleine N ist die Verteilung stark asymmetrisch, nähert sich aber für große N einer Normalverteilung an, wie nach dem zentralen Grenzwertsatz zu erwarten. Für $N = 1$ erreicht f einen unendlich hohen Wert.

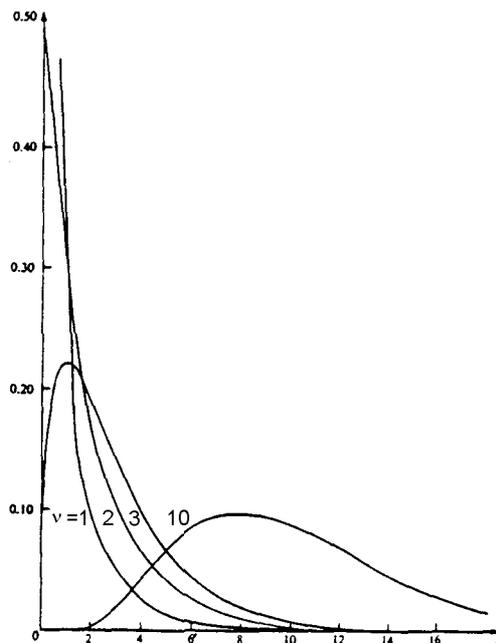


Abb. 7 χ^2 -Verteilung für $N = 1, 2, 3, 10$. Der Parameter v entspricht N .

Man habe N Beobachtungen x_i einer Variablen. Dann ist

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1.49)$$

eine Schätzung der Varianz. Seien die x_i aus $N(0,1)$, dann ist die geschätzte Varianz χ^2 -verteilt mit $v = N - 1$ Freiheitsgraden. Daß es nur $N-1$ Freiheitsgrade sind, ist einleuchtend, weil ja $\sum_{i=1}^N (x_i - \bar{x}) = 0$, so daß nur $N-1$ unabhängige Größen vorliegen. Dennoch könnte man in (1.49) statt durch $N-1$ auch durch N teilen. Dies wäre dann eine andere Schätzung und es wäre nachzuprüfen, welche besser ist. Wenn die x_i aus $N(\mu, \sigma)$ stammen, so ist natürlich

$$(N-1)s^2/\sigma^2 = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1.50)$$

χ^2 -verteilt mit $N-1$ Freiheitsgraden. Genau wie beim Mittelwert kann man nun Wahrscheinlichkeiten

$$p\left\{x_v\left(\frac{\alpha}{2}\right) \leq \frac{vs^2}{\sigma^2} \leq x_v\left(1 - \frac{\alpha}{2}\right)\right\} = 1 - \alpha \quad (1.51)$$

dafür angeben, daß der Schätzwert (1.49) eben in dieses Intervall fällt. Zu einer konkreten Auswertung benötigt man Tabellen. Man schreibt (1.51) um zu

$$p\left\{\frac{v}{x_v(1-\alpha/2)} \leq \frac{\sigma^2}{s^2} \leq \frac{v}{x_v(\alpha/2)}\right\} = 1 - \alpha. \quad (1.52)$$

Man findet in Abb. 8 die zugehörige Tabelle als Graphik. Beispiel: man habe 20 Beobachtungen aus einer Population mit $N(\mu, \sigma^2)$. Dann ist $v = 19$ und man liest aus Abb. 8 ab, daß

$$P\{0.58 \leq \frac{\sigma^2}{s^2} \leq 2.11\} = 0.95.$$

Man würde also erwarten, daß bei vielen Wiederholungen in 95% aller Fälle die Größe σ^2/s^2 im Bereich von 0.58 bis 2.11 zu liegen kommt oder $0.47 \sigma^2 < s^2 < 1.72\sigma^2$.

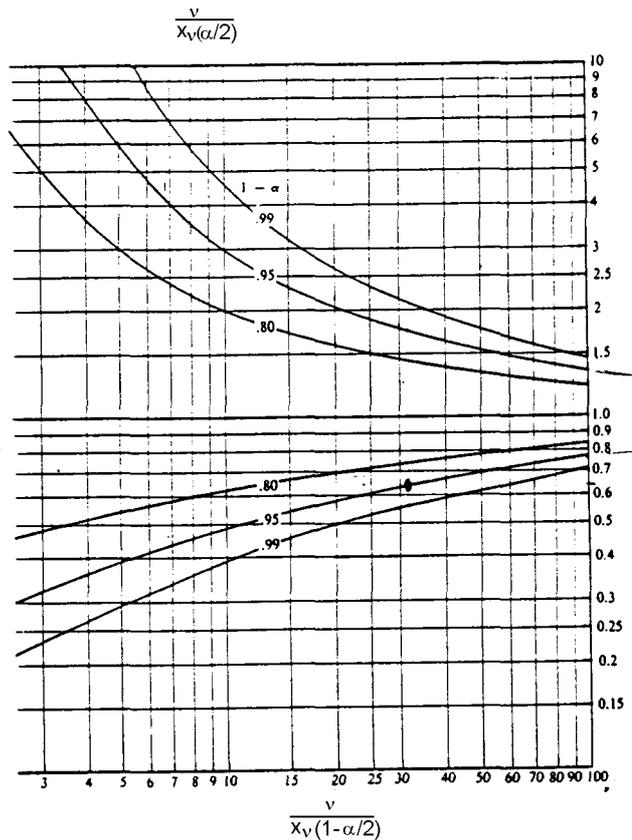


Abb. 8 Graphische Präsentation der Schranken in (1.52)

In vielen Fällen kennt man σ^2 nicht. Dennoch läßt sich in diesem Fall eine Stichprobenverteilung für den Mittelwert ableiten. Man definiert die Variable

$$T_v = \sqrt{N} (\bar{x} - \mu)/s, \quad v = N - 1. \quad (1.53)$$

Auch hier können Wahrscheinlichkeiten

$$p(-t_v(1 - \frac{\alpha}{2}) < T_v \leq t_v(1 - \frac{\alpha}{2})) = 1 - \alpha$$

dafür angegeben werden, daß (1.53) in ein bestimmtes Intervall fällt. Man lese die entsprechenden Werte von der beigefügten Tabelle 2 ab, wobei hier die Rubriken für den „zweiseitigen Test“ einschlägig sind.

Irrtumswahrscheinlichkeit α für den zweiseitigen Test									
FG \ α	0,50	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001
1	1,000	3,078	6,314	12,706	31,821	63,657	318,309	636,619	6366,198
2	0,816	1,886	2,920	4,303	6,965	9,925	22,327	31,598	99,992
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924	28,000
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610	15,544
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869	11,178
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959	9,082
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7,885
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041	7,120
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781	6,594
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587	6,211
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437	5,921
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318	5,694
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221	5,513
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140	5,363
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073	5,239
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015	5,134
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965	5,044
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922	4,966
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883	4,897
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850	4,837
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819	4,784
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792	4,736
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767	4,693
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745	4,654

Tabelle 2: Signifikanzschranken der Student-Verteilung. FG: Freiheitsgrade. Die vollständige Tabelle findet sich in Sachs: Angewandte Statistik.

1.4 Übergangswahrscheinlichkeiten

Gegeben sei ein System, das im Lauf der Zeit verschiedene diskrete Zustände annehmen kann. Zum Beispiel kann man Tage nach der relativen Sonnenscheindauer in Klassen einteilen. Wenn man insgesamt $i = 1, \dots, I$ Klassen definiert, so kann man dann nach der Wahrscheinlichkeit p_{ij} fragen, daß auf einen Tag der Klasse i einer der Klasse j folgt. Man nennt p_{ij} eine Übergangswahrscheinlichkeit. Nach Spirkel und Ries, QJRMS 1986 sei eingeteilt

Zustand	1	0%	$\leq RS \leq 5\%$
	2	5%	$\leq RS \leq 30\%$
	3	30%	$\leq RS \leq 60\%$
	4	60%	$\leq RS \leq 90\%$
	5	90%	$\leq RS \leq 100\%$.

Tabelle 3 zeigt für Weihenstephan sowohl die Wahrscheinlichkeiten p_i für die Zustände als auch die Übergangswahrscheinlichkeiten p_{ij} . Man sieht unter anderem, daß die Persistenz an "schlechten" Tagen ($i = 1$) im Januar hoch ist, auch, daß die p_{ij} nicht symmetrisch sind: $p_{ij} \neq p_{ji}$ im allgemeinen. Man kann so eine Tabelle zur Wettervorhersage benutzen. Da man die Bedeckung i an dem Tag kennt, an dem die Vorhersage ausgegeben werden soll, kennt, wird man für den nächsten Tag diejenige Bedeckung j vorhersagen für die p_{ij} am größten ist.

TABLE 3. UNCONDITIONAL PROBABILITY VECTOR $\{p_i(n)\}$ AND TRANSITION MATRIX $\{p_{i,j}(n)\}$ FOR JANUARY ($n = 15$) AND FOR JULY ($n = 198$). THE COLUMN INDEX REFERS TO THE OLD STATE, THE ROW INDEX i TO THE NEW STATE.

i	p_i	$p_{i,1}$	$p_{i,2}$	$p_{i,3}$	$p_{i,4}$	$p_{i,5}$
			January			
1	0.47	0.57	0.44	0.45	0.31	0.22
2	0.20	0.18	0.26	0.20	0.23	0.11
3	0.14	0.12	0.15	0.15	0.20	0.19
4	0.11	0.09	0.11	0.14	0.15	0.17
5	0.07	0.04	0.04	0.06	0.12	0.31
			July			
1	0.10	0.31	0.13	0.10	0.06	0.01
2	0.23	0.33	0.38	0.20	0.19	0.07
3	0.22	0.19	0.25	0.27	0.22	0.16
4	0.27	0.16	0.19	0.30	0.30	0.33
5	0.19	0.02	0.06	0.13	0.23	0.43

Tabelle 3

Man kann den Zustand der atmosphärischen Zirkulation durch den Begriff der Großwetterlage beschreiben, also typischen Strömungsmustern, denen sich die Atmosphäre ja annähert. Solche Muster sind z. B. von Vautard (JAS 1990) bestimmt worden und in Abb. 10 dargestellt. Man hat die blockierende Lage (BL) mit dem typischen Hoch-Tief Dipol, die zonale Lage (ZO) mit ausgeprägtem Westwindband über dem Atlantik, die Antizyklone über Grönland (GA) und den atlantischen Rücken (AR). Freilich kann man für viele Tage den Zustand der Atmosphäre nicht eindeutig solch einer Großwetterlage zuordnen und dementsprechend gibt es auch keine Übergangsmatrix p_{ij} . Es lassen sich dann nur bevorzugte Übergänge feststellen. So gibt es bevorzugt Übergänge $ZO \rightarrow BL$, $ZO \rightarrow AR$, $BL \rightarrow GA$ dagegen sehr selten $BL \rightarrow ZO$, $ZO \rightarrow GA$. Diese Situation ist in Abb. 9 wiedergegeben.

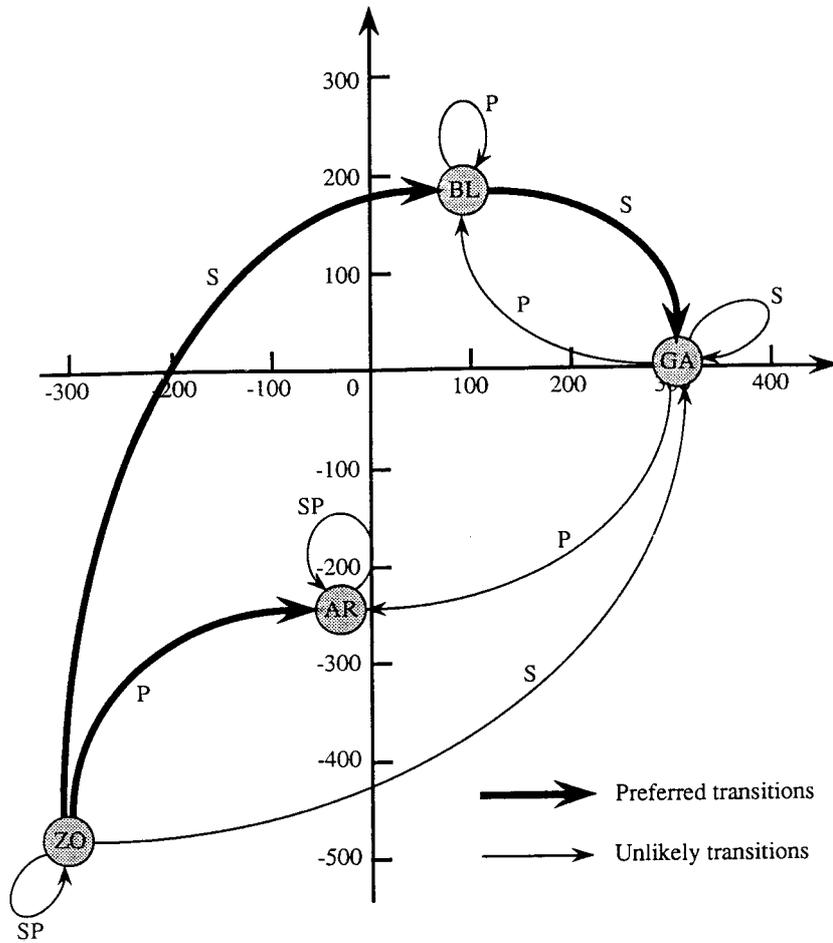


Abb. 9 Übergangsdiagramm für die in Abb. 10 aufgeführten Wetterlagen

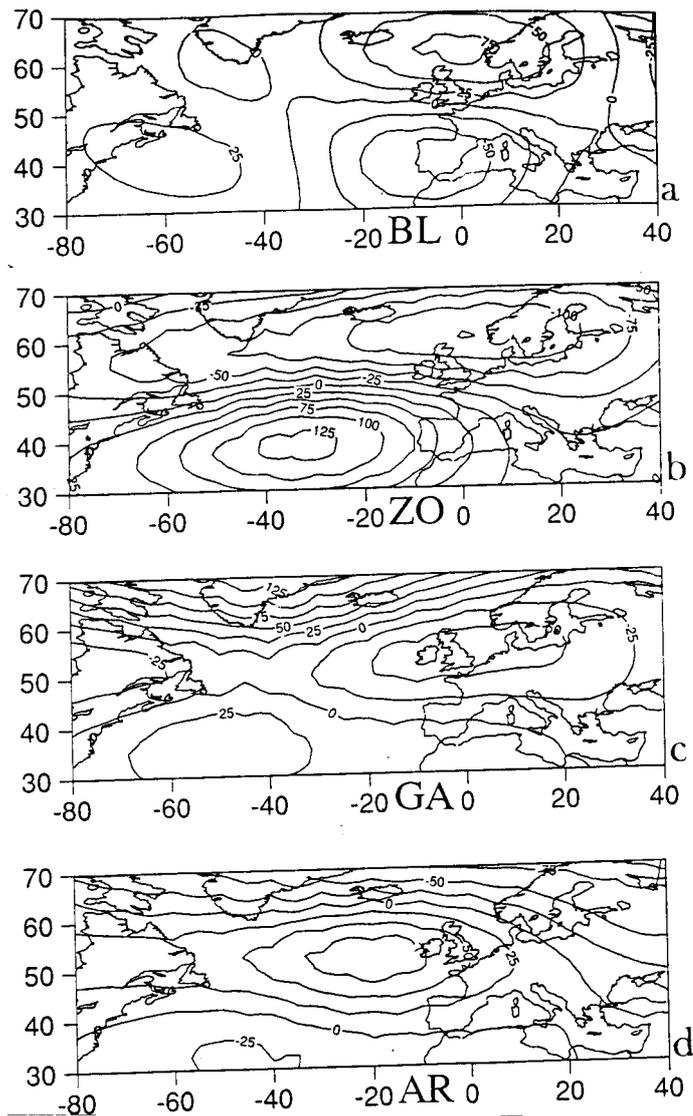


Abb. 10 Abweichung (m) der Höhe der 700mb-Fläche vom Langzeitmittel für die Großwetterlagen blockierendes Hoch (BL), Zonalzustand (ZO), Hoch über Grönland (GA) und Rücken über dem Atlantik (AR) (Vautard, 1990).

2. Statistische Schätzungen und Tests

Wahrscheinlichkeitsaussagen beziehen sich auf zukünftige Ereignisse. Statistische Tests beziehen sich auf vergangene Ereignisse. Wenn bereits Beobachtungen x_1, \dots, x_N einer Zufallsvariablen x vorliegen, so möchte man diese nutzen. Dazu noch ein Münzbeispiel: bisher wurden Aussagen über Erwartungswerte etc. gemacht, also ehe die Münze geworfen wurde. Sei aber nun eine Münze 15 mal geworfen worden und 12 mal die Zahl erschienen. Die Frage ist, ob die Münze gut geprägt, so daß beide Seiten mit gleicher Wahrscheinlichkeit erscheinen. Man fragt also nach der Wahrscheinlichkeit, daß die Zahl 12 mal kommt unter 15 Würfeln. Diese ist (siehe(1.26)) 0.018. Dies macht es unwahrscheinlich, daß die Münze richtig geprägt ist.

2.1. Stichprobenverteilung und Schätzung

Man habe an einem Ort aus vergangenen Jahren Tagesmitteltemperaturen für den ersten Januar und zwar aus $N=9$ Jahren:

$$\begin{aligned} x_1 &= 1.73^{\circ}\text{C}, x_2 = 3.81^{\circ}\text{C}, x_3 = 3.12^{\circ}\text{C} \\ x_4 &= 3.00^{\circ}\text{C}, x_5 = 3.48^{\circ}\text{C}, x_6 = 1.68^{\circ}\text{C} \\ x_7 &= 3.64^{\circ}\text{C}, x_8 = 4.91^{\circ}\text{C}, x_9 = 0.39^{\circ}\text{C} \end{aligned} \quad (2.1)$$

Gefragt ist nach dem wahren Mittelwert der Temperatur für den ersten Januar. Zunächst kann man annehmen, daß die Messungen unabhängig sind und normalverteilt $N(\mu, \sigma)$, wobei weder μ noch σ bekannt sind. Ferner muß ein Schätzwert für μ errechnet werden. Es ist sicher vernünftig

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.2)$$

als Schätzwert für μ zu nehmen ($\bar{x} = 2.86$). Man nimmt an, daß \bar{x} normal verteilt ist mit Mittelwert μ und Standardabweichung σ/\sqrt{N} (siehe (1.46)). Dann hat die Zufallsvariable

$$t_v = \sqrt{N} (\bar{x} - \mu)/s \quad (2.3)$$

mit $s^2 = \sum(x_i - \bar{x})^2 / (N - 1)$ eine Student-t-Verteilung mit $v = N - 1$ Freiheitsgraden. Dann kann die Wahrscheinlichkeit

$$p \left\{ -t_v \left(1 - \frac{\alpha}{2} \right) < \frac{\sqrt{N}(\bar{x} - \mu)}{s} \leq t_v \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha \quad (2.4)$$

, daß t_v in diesem Intervall zu liegen kommt, aus Tab. 2 abgelesen werden (zweiseitiger Test). Man findet für $v = 8$ und $\alpha = 0.05$ den Wert $t_v = 2.306$. Folglich besteht eine Wahrscheinlichkeit $1 - \alpha$ (0.95), daß der wahre Mittelwert μ vom Schätzwert (2.2) weniger als $t_v \left(1 - \frac{\alpha}{2} \right) \frac{s}{\sqrt{N}}$ entfernt ist. Man sagt, daß der $100(1 - \alpha)\%$ Vertrauensbereich für μ gegeben ist durch $\bar{x} \pm t_v \left(1 - \frac{\alpha}{2} \right) \frac{s}{\sqrt{N}}$.

Im Beispiel ist $s = 0.93$, also ist der 95% Vertrauensbereich $2.86 \pm 2.31 \cdot 0.93/3$, also $2.15 \leq \mu \leq 3.57$. Aus (2.4) wird klar, daß der Vertrauensbereich umso enger ausfällt, je größer N und je kleiner s .

Ganz analog kann man einen Vertrauensbereich für Varianzen angeben.

2.2. Prinzip der maximalen Wahrscheinlichkeit (maximum likelihood).

In 1.3f und 2.1 wurden Schätzfunktionen für Mittelwert und Varianz angegeben. Die Methode der maximalen Wahrscheinlichkeit gestattet ebenfalls die Schätzung von Parametern. Dabei gibt man sich eine Wahrscheinlichkeitsverteilung vor. Der Parameter, der in dieser Verteilung auftritt, wird dann durch optimale Anpassung der Verteilung an einen Satz von Messungen bestimmt.

Das Verfahren sei für einen Parameter λ einer vorgegebenen Dichte erläutert $f(x, \lambda)$. Nun hat man die Meßwerte x_1, \dots, x_N , so daß die Wahrscheinlichkeit für das Auftreten gerade dieser Meßwerte

$$L(\lambda) = \prod_{i=1}^N f(x_i, \lambda)$$

ist. $L(\lambda)$ heißt Likelihood-Funktion. Der Parameter λ wird nun so gewählt, daß L maximal wird. Damit stellt man sicher, daß das Meßresultat bei gegebener Dichtefunktion f maximale Wahrscheinlichkeit hat.

Beispiel: Man habe die Lebensdauer t_i von Zyklonen gemessen. Die vermutete Verteilung ist

$$f(t, \lambda) = t \exp(-t/\lambda) / \lambda^2.$$

Man zeigt leicht, daß $\int_0^{\infty} f dt = 1$.

Dann ist

$$L(\lambda) = \prod_{i=1}^N t_i \exp(-t_i/\lambda) / \lambda^2$$

, so daß das Maximum sich ergibt für

$$\frac{dL}{d\lambda} = 0.$$

Man erleichtert sich die Aufgabe, wenn man anstelle von L die Größe $\ln L$ betrachtet.

Dann gilt

$$\frac{d \ln L}{d\lambda} = 0 = \sum_{i=1}^N (t_i / \lambda^2 - 2 / \lambda)$$

und es folgt

$$\lambda = \frac{1}{2N} \sum_{i=1}^N t_i .$$

Es ist also die mittlere Hälfte der Lebensdauer von Zyklonen gerade gleich dem optimalen Parameter.

Man kann dieses Verfahren natürlich auch auf die Normalverteilung anwenden. Wenn x_i die Meßwerte einer Normalverteilung mit Standardabweichung σ sind, so kann den Parameter μ aus

$$L(\mu) = \prod_{i=1}^N \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

bestimmen, indem man verlangt, daß

$$\frac{d}{d\mu} \ln L = \sum_{i=1}^N (x_i - \mu) = 0$$

Das führt auf (2.2).

2.3. Statistische Signifikanztests

a) Parametrische Tests

Es ist die Frage zu entscheiden, ob eine Hypothese über statistische Parameter wahr oder falsch ist. Man kann z.B. fragen, ob eine Stichprobe von Messungen x_1, \dots, x_N mit der Hypothese verträglich ist, daß die Messungen zu einer Zufallsvariablen mit $N(\mu_0, \sigma^2)$ gehören. Die Schritte bei einem Signifikanztest sind folgende:

- 1.) Eine Nullhypothese H_0 muß formuliert werden. Etwa für den Datensatz (2.1) könnte H_0 lauten, daß die Temperaturen normal verteilt sind mit Mitteltemperatur $\mu = T_0$.
- 2.) Eine Alternativhypothese ist zu formulieren. Man könnte etwa $\mu \neq T_0$ ansetzen. Die Nullhypothese soll gegen die Alternativhypothese getestet werden.
- 3.) Es muß entschieden werden, anhand welcher Funktion $\varphi(x_1, \dots, x_N)$ der Beobachtungen (Prüfgröße) die Hypothese getestet werden soll. Wenn man die Varianz kennt, so ist \bar{x} die beste Größe für den Test im Beispiel. Wenn nicht, so ist es (2.3).
- 4.) Man hat die Stichprobenverteilung für diese Funktion anzugeben. Im Beispiel ist es die Student-t-Verteilung mit $N-1$ Freiheitsgraden.
- 5.) Aufgrund von 2) und 4) kann man den Stichprobenraum für die Funktion φ in einen kritischen Bereich und den komplementären Akzeptanzbereich einteilen, wobei eine

Irrtumswahrscheinlichkeit α (klein, meist 0.05) anzugeben ist. Der kritische Bereich ist so zu wählen, daß die Wahrscheinlichkeit $p(\varphi \text{ liegt im kritischen Bereich} / H_0 \text{ ist wahr}) = \alpha$ ist.

6.) Der Signifikanztest besteht darin, daß die Nullhypothese dann verworfen wird, wenn $\varphi(x_1, \dots, x_N)$ im kritischen Bereich zu liegen kommt. Andernfalls wird die Nullhypothese nicht verworfen.

In unserem Beispiel wird der kritische Bereich durch

$$t = \frac{\sqrt{N}(\bar{x} - T_0)}{s} < -t_{N-1}\left(1 - \frac{\alpha}{2}\right)$$

bzw.

$$t > t_{N-1}\left(1 - \frac{\alpha}{2}\right)$$

abgesteckt. Das heißt, daß der kritische Bereich für \bar{x} gegeben ist durch

$$\begin{aligned} \bar{x} &< (T_0 - st_{N-1}(1 - \alpha/2)/\sqrt{N}) \\ \text{bzw. } \bar{x} &> (T_0 + st_{N-1}(1 - \alpha/2)/\sqrt{N}). \end{aligned}$$

Sei etwa $T_0 = 2^\circ\text{C}$, so daß der kritische Bereich (für $\alpha = 0.01$) $\bar{x} < 0.95$, $\bar{x} > 3.05$ ist. Da $\bar{x} = 2.86^\circ\text{C}$ folgt, daß H_0 anzunehmen ist. Die Aussage $\mu = 2.2^\circ\text{C}$ kann allerdings genausowenig zurückgewiesen werden. Dagegen könnte eine Hypothese $T_0 = 0^\circ\text{C}$ zurückgewiesen werden.

Der diskutierte Fall steht als Beispiel für einen zweiseitigen Test, da beide Möglichkeiten $\mu > T_0$ und $\mu < T_0$ zugelassen sind. Hätte man dagegen die Alternativhypothese $\mu > T_0$ formuliert, so wäre dies auf einen einseitigen Test hinausgelaufen. Dann würde der kritische Bereich gegeben sein durch $t > t_{N-1}(1-\alpha)$ also

$$\bar{x} > (T_0 + st_{N-1}(1-\alpha)/\sqrt{N}).$$

Ein weiterer wichtiger Signifikanztest bezieht sich auf die Signifikanz von Korrelationen. Man habe den Schätzwert

$$\rho = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2\right)^{1/2}} \quad (2.6)$$

eines Korrelationskoeffizienten ρ_{xy} (siehe 1.42), wobei bekannt sei, daß $f_{xy}(x,y)$ bivariat normal ist. Dann bilde man die Prüfgröße

$$T_v^2 = \rho^2(N-2)/(1-\rho^2) \quad (2.7)$$

T_v sollte t-verteilt sein und die Korrelation gilt als akzeptiert, wenn $T_v > t_{N-2}(1-\alpha)$. Einfacher ist es die beigefügte Tabelle 4 zu konsultieren.

FG	Zweiseitiger Test			Einseitiger Test		
	5 %	1 %	0,1 %	5 %	1 %	0,1 %
1	0,9969	A*	B*	0,9877	0,9995	C*
2	0,9500	0,9900	0,9990	0,9000	0,9800	0,9980
3	0,8783	0,9587	0,9911	0,805	0,934	0,986
4	0,811	0,917	0,974	0,729	0,882	0,963
5	0,754	0,875	0,951	0,669	0,833	0,935
6	0,707	0,834	0,925	0,621	0,789	0,905
7	0,666	0,798	0,898	0,582	0,750	0,875
8	0,632	0,765	0,872	0,549	0,715	0,847
9	0,602	0,735	0,847	0,521	0,685	0,820
10	0,576	0,708	0,823	0,497	0,658	0,795
11	0,553	0,684	0,801	0,476	0,634	0,772
12	0,532	0,661	0,780	0,457	0,612	0,750
13	0,514	0,641	0,760	0,441	0,592	0,730
14	0,497	0,623	0,742	0,426	0,574	0,711
15	0,482	0,606	0,725	0,412	0,558	0,694
16	0,468	0,590	0,708	0,400	0,543	0,678
17	0,456	0,575	0,693	0,389	0,529	0,662
18	0,444	0,561	0,679	0,378	0,516	0,648
19	0,433	0,549	0,665	0,369	0,503	0,635
20	0,423	0,537	0,652	0,360	0,492	0,622
21	0,413	0,526	0,640	0,352	0,482	0,610
22	0,404	0,515	0,629	0,344	0,472	0,599
23	0,396	0,505	0,618	0,337	0,462	0,588
24	0,388	0,496	0,607	0,330	0,453	0,578
25	0,381	0,487	0,597	0,323	0,445	0,568

Tabelle 4: Prüfung des Korrelationskoeffizienten ρ auf Signifikanz gegen Null. Die Nullhypothese $\rho = 0$ wird abgelehnt, wenn $|\rho|$ den für die gewählte Irrtumswahrscheinlichkeit und den Freiheitsgrad angegebenen Wert überschreitet. Aus Sachs: Angewandte Statistik.

In vielen Fällen tritt eine weitere Schwierigkeit auf, indem oft ganze Variablenfelder etwa mit einem Index korreliert werden. Abb. 11 zeigt die Korrelation des Southern Oscillation Index mit der 700h Pa-Höhe für den Winter. Die Punkte markieren diejenigen Gitterpunkte, für die die Korrelation auf dem 95% Niveau signifikant ist, wenn der Signifikanztest gemäß Tabelle 4 für jeden Punkt für sich durchgeführt wird. Die Frage ist nun, ob dieses Muster durch Zufall entstanden sein kann. Wenn die 700hPa Werte an den jeweiligen Gitterpunkten unabhängig voneinander wären, so würde man bei 936 Punkten erwarten, daß etwa 5% der Punkte zufällig eine so hohe Korrelation zeigen würden. In Abb. 11 sind es 11%, doch folgt daraus noch nicht, daß hier eine signifikante Korrelation vorliegt, da ja benachbarte Punkte sicherlich nicht unabhängig sind. Eine Möglichkeit, sich hier Klarheit zu verschaffen, besteht in der Anwendung von 'Monte-Carlo'-Verfahren. Man kann zum Beispiel die Zeitreihe des SOI ersetzen durch Zufallszahlen und diese Zufallsreihe mit dem 700hPa-Wintermittel gerade so korrelieren wie es in Abb. 11 für den SOI gemacht wurde. Das Resultat ist in Abb. 12 zu sehen. Für die angegebenen Daten ist $\rho = 0.37$ die Signifikanzschwelle. Man sieht große Bereiche, in denen $|\rho| > 0.37$. Dies muß oft wiederholt werden, um sicher zu sein, daß auch das

in Abb. 12 gezeigte Bild nicht höchst unwahrscheinlich ist. Es ergibt sich (Livezey and Chen, 1983, MWR), daß die Hypothese, die hohen Korrelationen in Abb. 11 seien zufällig, auf dem 95% Niveau nicht zurückgewiesen werden kann.

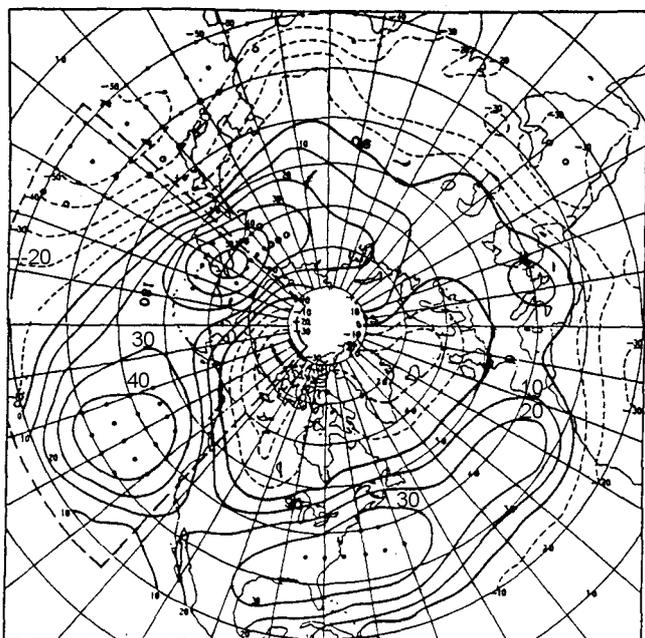


Abb. 11. Correlation of winter season averages of a Southern Oscillation Index (SOI) and 700 mb heights in hundreds. Negative and positive isopleths are shown as thin dashed lines and solid lines, respectively, the two separated by heavy solid zero lines. A subarea approximating the Pacific basin, used for the experiment summarized in Fig. 5c, is enclosed by heavy dashed line. Points on a 5 x 5 degree latitude - longitude grid at which correlations are 95 % significant are indicated by solid dots; open dots represent additional points significant with a more liberal test.

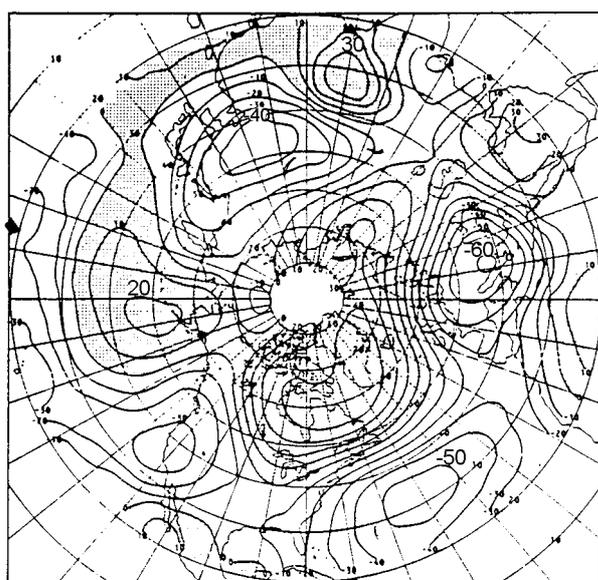


Abb. 12 Correlation in hundreds of winter season averages of 700 mb heights with Gaussian noise. Shading denotes positive values.

b) Nichtparametrische Tests

Bei parametrischen Tests macht man Voraussetzungen über die Verteilungen der Kollektive und wählt meist Parameter dieser Verteilungen als Prüfgrößen. Bei nichtparametrischen Tests macht man keine Voraussetzung über die Verteilung. Das oben beschriebene Monte-Carlo-Testverfahren ist hier schon einschlägig, da man dabei keine Aussagen über die Verteilung des SOI braucht.

Der Kolmogoroff-Smirnoff-Test sei hier als Beispiel eines nichtparametrischen Tests in Form eines Kochrezepts angeführt. Zu entscheiden ist die Frage, ob zwei unabhängige Stichproben derselben Grundgesamtheit entstammen oder nicht.

Man habe zwei Stichproben x_i, y_i , die gleich der Größe nach angeordnet seien

x_i : 0.6 1.2 1.6 1.7 1.7 2.1 2.8 2.9 3.0 3.2
 y_i : 2.1 2.3 3.0 3.1 3.2 3.2 3.5 3.8 4.6 7.2

Die erste Stichprobe habe den Umfang n_1 , die zweite n_2 . Im Beispiel ist $n_1 = n_2 = 10$. Die Frage ist, ob die x_i und y_i aus derselben Gesamtheit stammen, wobei über die Verteilung dieser Gesamtheit nicht spekuliert werden soll. Man teilt zunächst die Daten in Klassen ein und bestimmt die Häufigkeiten n_x, n_y :

Klasse	0-0.9/	1-1.9/	2.0-2.9/	3.0-3.9/	4.0-4.9/	5.0-5.9/	6.0-6.9/	7.0-7.9
n_x	1	4	3	2	0	0	0	0
n_y	0	0	2	6	1	0	0	1

Man bilde die Summenhäufigkeiten F_x, F_y

F_x : 1/10, 5/10, 8/10, 10/10, 10/10, ...

F_y : 0/10, 0/10, 2/10, 8/10, 9/10, 9/10, 9/10, 10/10

Dann suche man das Maximum der Differenzen $F_x - F_y$

Die Größe $D = \max |F_x - F_y|$ ist die Prüfgröße des Tests. Deren Verteilung ist tabelliert. Für mittlere bis große Stichprobenumfänge gilt ($n_1 + n_2 > 35$)

$$D_{(\alpha)} = K_{(\alpha)} \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} \quad (2.8)$$

, wobei K eine von der Irrtumswahrscheinlichkeit abhängige Konstante ist:

α	0.2	0.15	0.10	0.05	0.01	0.001
K	1.07	1.14	1.22	1.36	1.63	1.95

(2.9)

Erreicht oder überschreitet D den kritischen Wert D_{α} , so liegt ein signifikanter Unterschied vor. Für kleinere Stichprobenumfänge muß man Tabellen zuziehen. Dazu etwa für $n_1 = n_2$

D_n	$n = n_1 = n_2$	10	15	20	25	30	
α	0.05	7/10	8/15	9/20	10/25	11/30	(2.10)
α	0.01	8/10	9/15	11/20	12/25	13/30	

In unserem Beispiel ist $D = 6/10$. Das erreicht den kritischen Wert nicht ($\alpha = 0.05$) - also kann man die beiden Gesamtheiten nicht unterscheiden. Die x_i und y_i können zur selben Gesamtheit gehören.

3. Stochastische Prozesse und Differentialgleichungen

Im Prinzip sind die Differentialgleichungen der Meteorologie und Ozeanographie stochastische Differentialgleichungen, da die Variablen Zufallsvariable sind und viele der Koeffizienten Zufallsgrößen. Dieser Tatbestand wird normalerweise ignoriert und man tut so, als hätte man deterministische Gleichungen mit fest gegebenen Koeffizienten. Die numerische Wettervorhersage bietet das Paradebeispiel für diese Situation. Dabei wird angenommen, daß man zu Vorhersagebeginn die Werte aller Variablen an den jeweiligen Gitterpunkten der Modelle genau kennt, wiewohl doch bekannt ist, daß diese Werte mit einiger Unsicherheit behaftet sind, man den Variablen also Wahrscheinlichkeitsverteilungen zuschreiben müßte und deren Entwicklung berechnen. So würde man automatisch zu stochastischen Wettervorhersagen kommen, die Vorhersagen mit gewissen Wahrscheinlichkeiten belegen. Die Ausführung solcher stochastischer Vorhersagen scheitert an der Größe des Rechenaufwandes, der dafür getrieben werden müßte. Dennoch sollen hier wenigstens Grundkonzepte der Theorie der stochastischen Differentialgleichungen angedeutet werden, die in einfachen Fällen auch zum Tragen kommen können.

3.1. Grundbegriffe

Stochastische Differentialgleichungen beschreiben stochastische Prozesse $X(t)$, für die sich genau wie für eine Zufallsvariable eine Wahrscheinlichkeitsdichte $f_X(X,t)$ angeben läßt. In Analogie zu (1.5) wird ein Erwartungswert eingeführt und nach der Autokovarianz gefragt:

$$\gamma_{xx}(t_1, t_2) = E((x(t_1) - \mu(t_1))(x(t_2) - \mu(t_2))). \quad (3.1)$$

Bei einem stationären Prozeß ist $\mu(t_1) = \mu(t_2) = \mu$ und γ_{xx} hängt nur von der Verschiebung $\tau = t_2 - t_1$ ab. Die Autokorrelationsfunktion ist

$$\rho_{xx}(t_1, t_2) = \gamma_{xx}(t_1, t_2) / (\sigma(t_1) \sigma(t_2)), \quad (3.2)$$

Es gibt auch stochastische Prozesse $x(t_1)$, die nur an diskreten Zeitpunkten definiert sind. Ein Beispiel wurde in 1.2 gegeben.

Bei einem stochastischen Prozeß $x(t)$ sind in aller Regel die Werte $x(t_i)$ zu verschiedener Zeit nicht voneinander unabhängig. Die Wahrscheinlichkeit, daß x zur Zeit t_1 im Intervall $[x_1, x_1 + Dx_1]$ sich befindet und zur Zeit t_2 im Intervall $[x_2, x_2 + Dx_2]$ wird geschrieben

$$f(x_1, t_1; x_2, t_2) dx_1 dx_2. \quad (3.3)$$

Die Größe $f(x_1, t_1; x_2, t_2)$ läßt zunächst keinen Schluß zu, ob sich Zustände aus dem Intervall $[x_2, x_2 + Dx_2]$ ins Intervall $[x_1, x_1 + Dx_1]$ begeben haben. Fehlt insbesondere eine statistische Abhängigkeit zu verschiedenen Zeiten, so hat man

$$f(x_1, t_1; x_2, t_2) = f(x_1, t_1) f(x_2, t_2) \quad (3.4)$$

und somit für $t_1 \neq t_2$:

$$\gamma_{xx}(t_1, t_2) = 0. \quad (3.5)$$

Ist schließlich noch $f(x, t) = f(x)$ stationär und normal verteilt, so nennt man diesen Prozeß "weißes Rauschen".

Für die bedingte Wahrscheinlichkeit gilt

$$f(x_2, t_2 | x_1, t_1) = \frac{f(x_2, t_2; x_1, t_1)}{f(x_1, t_1)} \quad (3.6)$$

so daß bei Unabhängigkeit mit $f(x_2, t_2 | x_1, t_1) = f(x_2, t_2)$ die Beziehung (3.4) folgt.

Eine besonders einfache Form zeitlicher Abhängigkeit hat man bei Markovprozessen, für die gilt

$$f(x_n, t_n | x_{n-1}, t_{n-1}; \dots, x_1, t_1) = f(x_n, t_n | x_{n-1}, t_{n-1}). \quad (3.7)$$

Die linke Seite von (3.7) gibt die bedingte Wahrscheinlichkeit an, daß zur Zeit t_n der Wert x_n eintritt, wenn zu den vorhergehenden Zeiten t_{n-1} bis t_1 die Werte x_{n-1} bis x_1 eingetreten waren. Für einen Markovprozeß gilt dann, daß die Übergangswahrscheinlichkeit unabhängig von der Vorgeschichte in dem Sinne ist, daß man nur den Wert x_{n-1} zur Zeit t_{n-1} kennen muß. Die Kenntnis der Vorgeschichte ist zur Berechnung der Übergangswahrscheinlichkeit $x_{n-1} \rightarrow x_n$ nicht nötig.

Insbesondere gilt für $t_3 > t_2 > t_1$ für einen Markovprozeß

$$\begin{aligned} f(x_3, t_3; x_2, t_2; x_1, t_1) &= f(x_3, t_2 \mid x_2, t_2) f(x_2, t_2; x_1, t_1) \\ &= f(x_3, t_3 \mid x_2, t_2) f(x_2, t_2 \mid x_1, t_1) f(x_1, t_1) \end{aligned} \quad (3.8)$$

Durch Integration über x_2 erhält man daraus die Chapman-Kolmogorovgleichung

$$f(x_3, t_3; x_1, t_1) = \int f(x_3, t_3 \mid x_2, t_2) f(x_2, t_2 \mid x_1, t_1) dx_2. \quad (3.9)$$

Diese dient normalerweise als Basis für die Ableitung der im folgenden aufgeführten stochastischen Differentialgleichungen. Hier sei ein relativ anschaulicher und schneller, dafür aber ungenauer Weg zur Ableitung bevorzugt. Dieser wird anhand eines besonders einfachen Beispiels besprochen, nämlich der Temperaturgleichung für die Deckschicht des Ozeans.

3.2. Deckschicht des Ozeans

Sei h die Tiefe der Deckschicht und T die Abweichung der Wassertemperaturen in der Deckschicht von einem klimatologischen Mittelwert. Ferner sei T nicht von der Vertikalcoordinate z abhängig. Dann läßt sich eine Bilanzgleichung

$$\frac{\partial}{\partial t} (hT) = f(T, T_a, q, v_a, v_o, \text{Strahlung}) \quad (3.10)$$

aufstellen, in der T_a die Lufttemperatur, q die Feuchte, v_a der Wind und v_o die Strömungsgeschwindigkeit im Ozean ist. Diese muß eine stochastische Differentialgleichung sein, da man normalerweise die "Antriebe" v_a , q , T_a , Strahlung nicht genau kennt und auch nicht die Ozeanströmung. Dies bleibt sie auch, wenn man (3.10) extrem vereinfacht, indem man Transporte im Ozean vernachlässigt, h als konstant ansetzt, Schwankungen der Strahlung übersieht etc., um auf

$$\frac{dT}{dt} = -\lambda T + \lambda T_a \quad (3.11)$$

zu kommen, wobei

$$\lambda = C(1 + B) \rho_a c_{pa} |\mathbf{v}_a| / (\rho_o c_o h)$$

(C Widerstandskoeffizient; B Bowenratio; c spez. Wärme; $\lambda^{-1} \approx 2$ Monate). Offensichtlich ist auch λ eine Zufallsgröße, doch wird im folgenden λ für konstant gehalten. Dennoch ist (3.11) eine stochastische Gleichung, da ja T_a eine Zufallsgröße ist und somit T eine Zufallsvariable. Die Frage lautet dann: Gegeben die Verteilung $f_a(T_a, t)$, welche Verteilung $f(T, t)$ hat T ? Natürlich ist die Deckschicht des Ozeans nicht das einzige System, auf das sich eine Gleichung vom Typ (3.11) anwenden läßt. Um diese Frage beantworten zu können, empfiehlt es sich, zunächst homogene Gleichungen ohne Antrieb

zu studieren und zu fragen, wie sich in diesem einfacheren Fall Verteilungen vorhersagen lassen.

3.3. Die Liouville-Gleichung

Die Lösung der zu (3.11) gehörigen homogenen Gleichung

$$\frac{dT}{dt} + \lambda T = 0 \quad (3.12)$$

ist bekannt, wenn $T = T_0$ zur Zeit $t = 0$ gegeben ist:

$$T = T_0 \exp(-\lambda t) \quad (3.13)$$

Diese Klarheit geht verloren, wenn zur Zeit $t = 0$ nur die Wahrscheinlichkeitsdichte $f(T,0)$ bekannt ist. Um dieses Problem zu lösen, wechselt man in einen Phasenraum, in dem die Variable des Problems, eben T , Koordinate ist. In diesem Raum gibt das Integral

$$\int_{T_1}^{T_2} f(T,t) dT \quad (3.14)$$

die Wahrscheinlichkeit an, daß sich die Temperatur der Deckschicht im Intervall $T_1 \leq T \leq T_2$ befindet. Um nun eine Vorhersagegleichung für $f(T,t)$ abzuleiten, kann man ganz in Analogie zur Ableitung der Kontinuitätsgleichung vorgehen. Genau wie $\rho = \rho_a$ die Dichte der Luft ist, die angibt, wieviel Masse pro Volumeneinheit vorhanden ist, so ist $f(T,t)$ die Zustandsdichte im Phasenraum, die einem die Zahl der Zustände pro Einheitsintervall der Temperatur angibt. Genau wie Masse nicht verloren gehen kann, so können Zustände nicht verloren gehen. Genau wie

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$$

die zu ρ gehörige Kontinuitätsgleichung ist, so muß für f gelten

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial T} (u_T f) = 0, \quad (3.15)$$

wobei u_T die Geschwindigkeit ist, mit der sich die Zustände im Phasenraum bewegen. Ganz analog zu

$$\mathbf{v} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right) \text{ muß gelten}$$

$$u_T = \frac{dT}{dt} = -\lambda T. \quad (3.16)$$

Damit folgt die Vorhersagegleichung für f , die Liouville-Gleichung,

$$\frac{\partial f}{\partial t} - \frac{\partial}{\partial T} (\lambda T f) = 0. \quad (3.17)$$

Man sieht insbesondere, daß

$$\frac{d}{dt} \int_{-\infty}^{+\infty} f dT = 0$$

wenn $fT = 0$ für $T \rightarrow \pm \infty$. Die Zahl der Zustände ändert sich also nicht.

Sei zur Zeit $t = 0$

$$f(T, 0) = N_0 \exp(-T^2/a_0) \quad (3.18)$$

, wobei $N_0 = (\pi a_0)^{-1/2}$ (siehe 1.36). Dann ist die Lösung von (3.17)

$$f(T, t) = N_0 \exp(\lambda t - T^2(\exp(2\lambda t))/a_0). \quad (3.19)$$

Da $\exp(2\lambda t)$ im Lauf der Zeit zunimmt, bedeutet dies, daß die Standardabweichung immer kleiner, die Verteilung also immer enger um den Nullpunkt konzentriert wird. Dies ist einleuchtend: (3.12) sagt ja, daß Abweichungen vom Mittelwert ausgedämpft werden, also $T \rightarrow 0$ für $t \rightarrow \infty$. Gibt man also eine Gaußverteilung mit $\sigma^2 = a_0/2$ für $t = 0$ vor, so muß σ im Lauf der Zeit abnehmen.

Selbstverständlich läßt sich (3.17) verallgemeinern. Sei

$$\frac{dq_i}{dt} = Q_i(q_j) \quad (3.20)$$

ein Satz von Differentialgleichungen für die Variablen q_i ($i=1, \dots, N$). Insbesondere kann (3.20) die Gleichungen eines numerischen Wettervorhersagemodells symbolisieren, wobei q_i dann die Gitterpunktwerte der meteorologischen Variablen sind ($N = \text{Zahl der Gitterpunkte} \cdot \text{Zahl der Variablen}$). Der zugehörige Phasenraum mit den Koordinaten q_i ist N -dimensional und die Liouville-Gleichung lautet

$$\frac{\partial f}{\partial t} = - \frac{\partial}{\partial q_i} (Q_i f) \quad (3.21)$$

, wobei die Summationskonvention gilt. Anhand von (3.21) wird sofort klar, warum stochastische Vorhersagen im strikten Sinne nicht durchführbar sind. Der Phasenraum ist viel zu groß, um mit heutigen Rechnern eine annähernd genaue Integration von (3.21) zu erlauben.

3.4. Zufallsgang und Langevin-Gleichung

Es gilt nun, für den inhomogenen Term in (3.11) eine Formulierung zu finden, die in etwa die Zufälligkeit atmosphärischer Temperaturreihen nachbildet. Das einfachste Modell für eine solche Formulierung leitet sich aus dem wohlbekannten Zufallsgang (random walk) ab. Dabei nimmt man in (3.11) zunächst an, daß der Term $-\lambda T$ vernachlässigt werden kann, so daß

$$dT = \lambda T_a dt. \quad (3.22)$$

Man nimmt weiter an, daß sich T über das Zeitinkrement Dt um $\pm DT$ ändert, wobei $|DT|$ konstant ist. Das heißt, daß sich T_a von Inkrement zu Inkrement sprunghaft ändern kann, jedoch sind nur die Werte $\pm DT/\lambda Dt$ zugelassen. Positive und negative Änderungen sollen gleich wahrscheinlich sein. Dies entspricht genau dem „Münzwurf“ in (1.1), wobei jetzt $\gamma = DT$.

Sei s die Zahl der positiven Änderungen und nDt die zurückgelegte Zeit. Dann ist $T_s = (2s-n)DT$ die Temperatur, wenn $T = 0$ anfänglich. Die Temperatur ist natürlich eine Zufallsvariable, der man eine Verteilung zuordnen kann. Die Wahrscheinlichkeit, $T = T_s$ nach n Zeitschnitten zu finden ist

$$p(s, n) = \binom{n}{s} 2^{-n} \quad (3.23)$$

entsprechend (1.26). Aus (3.23) folgt:

$$E(T) = \sum_{s=0}^n T_s p(s, n) = 0 \quad (3.24)$$

Die Varianz ergibt sich zu

$$E(T_s^2) = \sum_{s=0}^n T_s^2 p(s, n) = nDT^2 = DT^2 t/Dt \quad (3.25)$$

Den Beweis von (3.25) kann durch vollständige Induktion erbracht werden (siehe auch 1.3b). Die Varianz wächst wie bei einem diffusen Prozeß linear mit der Zeit. Stattet man also das Deckschichtmodell mit einem solchen Antrieb aus und verzichtet man auf Dämpfung, so wird die Varianz der Meerestemperatur $\sim t$ ansteigen.

Den Wienerprozeß $W(t)$ kann man als Limes des Zufallsgangs mit Dt , $DT^2 \rightarrow 0$, $DT^2/Dt = \alpha$, konstant, sehen. Die Dichte von W sei f_w , so daß

$$E \{W(t)\} = \int_{-\infty}^{+\infty} f_w W(t) dw.$$

Die Integration geht hier über ein Ensemble möglicher Realisierungen von w . Für einen Wienerprozeß muß gelten

$$E(W(t+Dt) - W(t)) = 0 \quad (3.26)$$

$$E((W(t' + Dt) - W(t')) (W(t''+Dt)-W(t''))) = Dt\delta(t'-t''). \quad (3.27)$$

Man sieht, daß der Wienerprozeß eine Verallgemeinerung des Zufallsgangs ist. (3.26) drückt aus, daß positive Änderungen genau so wahrscheinlich sind wie die entsprechenden negativen. (3.27) sagt, daß Änderungen, die zu verschiedenen Zeiten erfolgen, unkorreliert sind. Schließlich leitet man aus (3.27) ab

$$E ((W(t+Dt) - W(t))^2) = Dt. \quad (3.28)$$

Die Varianz eines Wienerprozesses wächst linear mit der Zeit an. Damit läßt sich nun ansetzen für (3.11)

$$dT = -\lambda T dt + Q^{1/2} dW \quad (3.29)$$

als Modellgleichung für die Deckschicht, wobei nun

$$\lambda T_a dt = Q^{1/2} dW. \quad (3.30)$$

Mit Q wird die Intensität des Antriebs charakterisiert (Einheit $K^2 s^{-1}$).

Mit (3.29) nimmt man an, daß sich die Temperatur der Atmosphäre im Sinne eines Zufallsgangs ändert. Dies ist an sich keine sehr gute Annahme, da ja die Atmosphäre Persistenz hat. Doch sind die ozeanischen Zeitskalen so lang, daß (3.29), wie sich zeigen wird, durchaus gute Resultate liefert. (3.29) trägt den Namen Langevingleichung.

3.5. Fokker-Planck-Gleichung

Die Frage ist nun, wie der Antrieb in (3.29) die Dichte f beeinflusst. Ein heuristisches Vorgehen führt rasch zum Ziel. Man denke sich die Temperaturachse in diskrete Intervalle der Breite DT unterteilt und T_j sei die Temperatur im Intervall mit Index j . Dann ist $f(T_j, t)DT$ die Wahrscheinlichkeit, zur Zeit t die Temperatur T_j zu finden. Dann muß ohne dW gelten

$$f(T_j, t + Dt)DT = f(T_j, t) DT + (\lambda T_{j+1} f(T_{j+1}, t) - \lambda T_j f(T_j, t))Dt. \quad (3.31)$$

(3.31) entspricht offensichtlich der Liouville-Gleichung (3.17), wobei in der Formulierung von (3.31) $T_{j+1} > T_j > 0$ angenommen wurde. Der Sinn von (3.31) ist klar: Zustände im Intervall gehen verloren, weil die Dämpfung die Temperaturen abnehmen läßt. Dieser Verlust wird durch $-\lambda T_j f(T_j, t)$ beschrieben. Es werden aber auch Zustände dazukommen, da Zustände aus dem Intervall $j+1$ durch Kühlung ins Intervall j gelangen. Daher der Term $+\lambda T_{j+1} f(T_{j+1}, t)$. Schaltet man nun einen Wienerprozeß dazu, wie in (3.29), so verliert dadurch das Intervall j Zustände nach beiden Seiten und in gleichem Maße, da ja durch den Zufallsgang Sprünge nach beiden Seiten möglich und gleich wahrscheinlich sind. Also hat man Verluste $\sim -2f(T_j, t)DT$, aber auch Gewinne der Form $f(T_{j+1}, t)DT$ und $f(T_{j-1}, t)DT$, die durch Sprünge aus den Nachbarintervallen ins Intervall j hervorgerufen werden. Somit

$$DT f(T_j, t+Dt) = f(T_j, t)DT + \lambda f(T_{j+1}, t)T_{j+1}Dt - \lambda f(T_j, t)T_jDt + \frac{Q}{2\alpha} (f(T_{j-1}, t) - 2f(T_j, t) + f(T_{j+1}, t))DT \quad (3.32)$$

, wobei der Faktor $Q/2\alpha$ nicht weiter begründet sei. Die einzelnen Übergänge sind in Abb. 13 dargestellt.

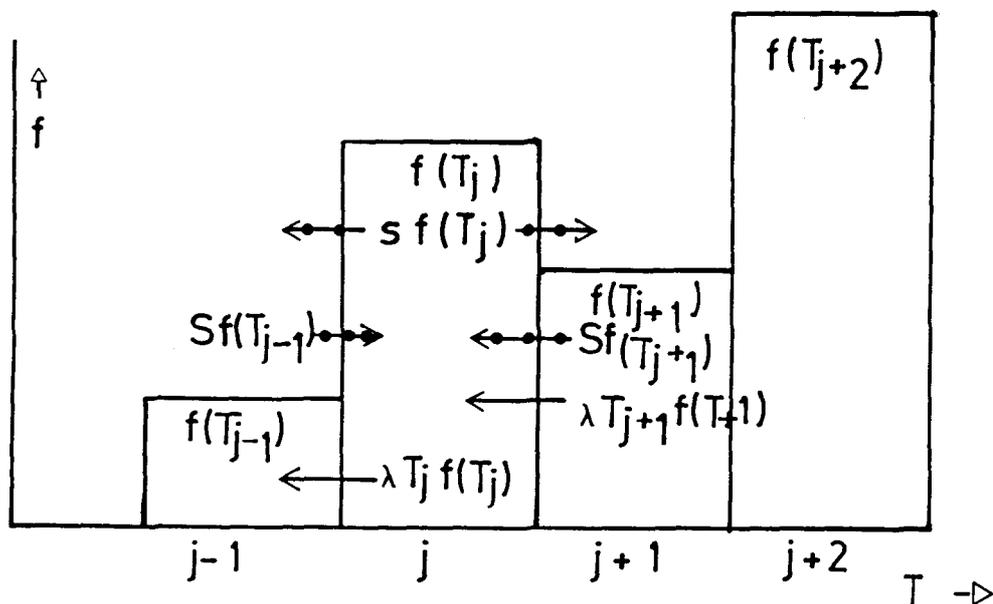


Abb. 13 Die Abbildung dient der Veranschaulichung von (3.32). Dargestellt sind die Zu- und Abflüsse von Zuständen für das Intervall j . Dabei ist $s = Q/(2\alpha)DT$. Die mit dem Sprungprozeß verbundenen Übergänge sind punktiert.

Durch Grenzübergang folgt die Fokker-Planck-Gleichung ($DT, Dt \rightarrow 0$)

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial T}(\lambda f T) + \frac{Q}{2} \frac{\partial^2 f}{\partial T^2} \quad (3.33)$$

, für die sich, wie gesagt, auch eine präzise Ableitung finden läßt. Für uns ist wichtig, daß sich ein Antrieb von der Form eines Wienerprozesses auf die Entwicklung der Wahrscheinlichkeitsdichte wie ein Diffusionsprozeß auswirkt. Man darf vermuten, daß solch ein Antrieb der durch (3.19) beschriebenen Verengung der Verteilung entgegenwirkt. In der Tat findet man jetzt eine stationäre Verteilung. Für $\frac{\partial f}{\partial t} = 0$ folgt aus (3.33)

$$\frac{\partial}{\partial T}(\lambda f T) + \frac{Q}{2} \frac{\partial^2 f}{\partial T^2} = 0 \quad (3.34)$$

und so

$$\lambda f T + \frac{Q}{2} \frac{\partial f}{\partial T} = 0. \quad (3.35)$$

Also gilt

$$f = \left(\frac{\pi Q}{\lambda} \right)^{-1/2} \exp(-T^2 \lambda / Q). \quad (3.36)$$

Man hat nun

$$2\sigma^2 = Q/\lambda. \quad (3.37)$$

Je größer die Intensität des Antriebs, desto breiter ist die Verteilung, je größer λ , desto schmaler.

Als Beispiel für die Anwendung von (3.33) sei ein nulldimensionales Klimamodell diskutiert, das stochastische Übergänge von Eiszeiten zu gemäßigten Klimaten aufweist (nach Fraedrich, QJRMS 1978).

Man betrachte das globale Klimasystem, wo

$$R \downarrow = \frac{1}{4} \mu I_0 (1 - \alpha_p)$$

die Einstrahlung ist (I_0 Solarkonstante; α_p planetare Albedo; μ Variationsparameter für die Solarkonstante) und

$$R \uparrow = \varepsilon \sigma T^4$$

die Abstrahlung (σ = Stephan-Boltzmann-Konstante; ε Emissivität; T mittlere Temperatur). Ferner wird eine Rückkopplung zwischen Strahlung und Eisbedeckung angenommen:

$$\alpha_p = a - bT.$$

Mit Zunahme der Temperatur sinkt die planetare Albedo ($a, b > 0$). Da man nicht möchte, daß die planetare Albedo den Wert 1 erreicht oder gar überschreitet, wird, natürlich etwas willkürlich, die Obergrenze $\alpha_p = 0.75$ festgesetzt. Diesen Wert kann α nicht überschreiten und α hängt folglich für tiefe Temperaturen nicht von T ab. Der Faktor c in der globalen Bilanzgleichung

$$c \frac{dT}{dt} = \{-\varepsilon\sigma T^4 + \frac{1}{4} \mu I_0 b T + \frac{1}{4} \mu I_0 (1-a)\} \quad (3.38)$$

legt die Größe der durch die Energieumsätze ausgelösten Temperaturänderungen fest. Für $\alpha_p = 0.75$ enthält die Bilanz (3.38) in T keinen linearen Term. Die Gleichgewichtszustände $\bar{T} (\frac{dT}{dt} = 0)$ für (3.38) lassen sich relativ leicht ermitteln und sind in Abb. 14 aufgetragen. Man erhält als Funktion von μ für $\mu > 0.97$ drei Lösungen \bar{T} . Eine davon (\bar{T}_{jetzt}) entspricht dem heutigen Klima. Sie ist stabil in dem Sinne, daß $T \rightarrow \bar{T}$, wenn man in (3.38) eine Anfangstemperatur vorgibt, die nur wenig von \bar{T} entfernt ist. Die zweite Lösung ist instabil. Die dritte dagegen ($\bar{T} = \bar{T}_{\text{Eis}} = 215\text{K}$) ist wieder stabil und entspricht einer eisbedeckten Erde. Für die gilt $\alpha_p = 0.75$. Für $\mu < 0.96$ gibt es nur eine Lösung, eben die eisbedeckte Erde.

Wir fügen nun auf der rechten Seite von (3.38) einen stochastischen Antrieb hinzu, der etwa den Einfluß des Wetters auf die Strahlungsbilanz darstellen möge. Dann hat man

$$dT = (-C_1 T^4 + C_2 T - C_3) dt + \sqrt{Q} dW \quad (3.39)$$

, wobei die Konstanten $C_i > 0$ unmittelbar aus (3.38) folgen. Für die Dichteverteilung f

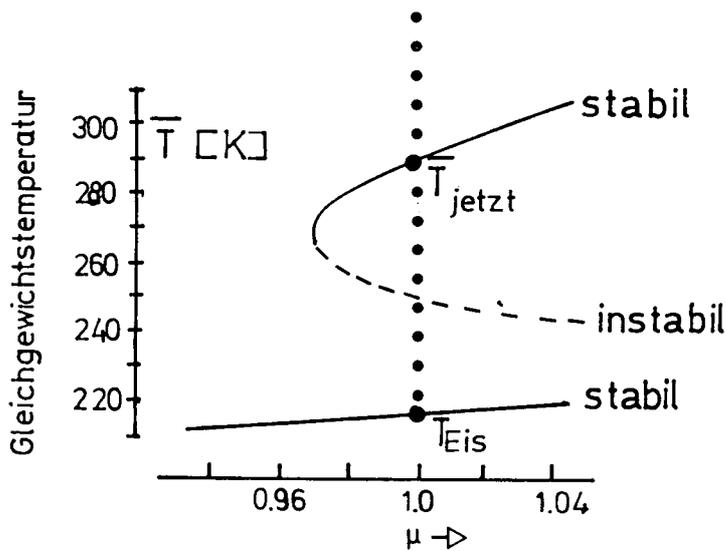


Abb. 14 Gleichgewichtstemperaturen \bar{T} des nulldimensionalen Klimamodells, wobei μ die relative Intensität der Solarstrahlung ist. $\eta = 1$ entspricht der Gegenwart.

der Klimazustände folgt sofort:

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial T} ((-C_1 T^4 + C_2 T - C_3) f) + \frac{Q}{2} \frac{\partial^2 f}{\partial T^2} \quad (3.40)$$

Für $\alpha_p = 0.75$ ist (3.40) zu ersetzen durch

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial T} ((-C_1 T^4 + I_o \mu / 16.) f) + \frac{Q}{2} \frac{\partial^2 f}{\partial T^2}$$

Sucht man eine Gleichgewichtsverteilung ($\frac{\partial}{\partial t} = 0$), so hat man

$$-f(-C_1 T^4 + C_2 T - C_3) + \frac{Q}{2} \frac{\partial f}{\partial T} = 0.$$

Also

$$\frac{\partial}{\partial T} \ln f = \frac{2}{Q} (-C_1 T^4 + C_2 T - C_3). \quad (3.41)$$

Für eine qualitative Diskussion genügt es, sich den groben Verlauf der rechten Seite von (3.41) klarzumachen, der ja einfach der Tendenz von T ohne stochastischen Antrieb entspricht. Man kann ihn unmittelbar aus Abb. 14 (s. Abb.15) entnehmen, wo ja die Gleichgewichtszustände die Nullstellen sind. Der Kurvenverlauf in Abb. 15 basiert auch auf der hier nicht ausgeführten Stabilitätsanalyse. So ist $\frac{dT}{dt} < 0$ für $\bar{T} > T_{\text{jetzt}}$, d.h.

Störungen werden ausgedämpft. Folgt schließlich

$$f \sim \exp(\int (-C_1 T^4 + C_2 T - C_3) dT). \quad (3.42)$$

Der Verlauf von

$$E = \int (-C_1 T^4 + C_2 T - C_3) dT \quad (3.43)$$

läßt sich aus Abb. 15 ablesen. Für $T < \bar{T}_{\text{Eis}}$ muß E anwachsen, um für $T = \bar{T}_{\text{Eis}}$ ein Maximum zu erreichen. Dann sinkt E ab, um am zweiten instabilen Gleichgewichtszustand ein Minimum zu erreichen. Folgt schließlich ein zweites Maximum bei \bar{T}_{jetzt} . Die zugehörige Verteilung der Klimazustände ist durch Maxima bei den "Eiszeiten" und bei Klimaten vom jetzt beobachteten Typ gekennzeichnet, doch haben etwa Zustände dazwischen auch eine endliche Wahrscheinlichkeit des Auftretens. Insbesondere ist in so einem Modell ein Übergang vom jetzigen Klima in eine Eiszeit möglich.

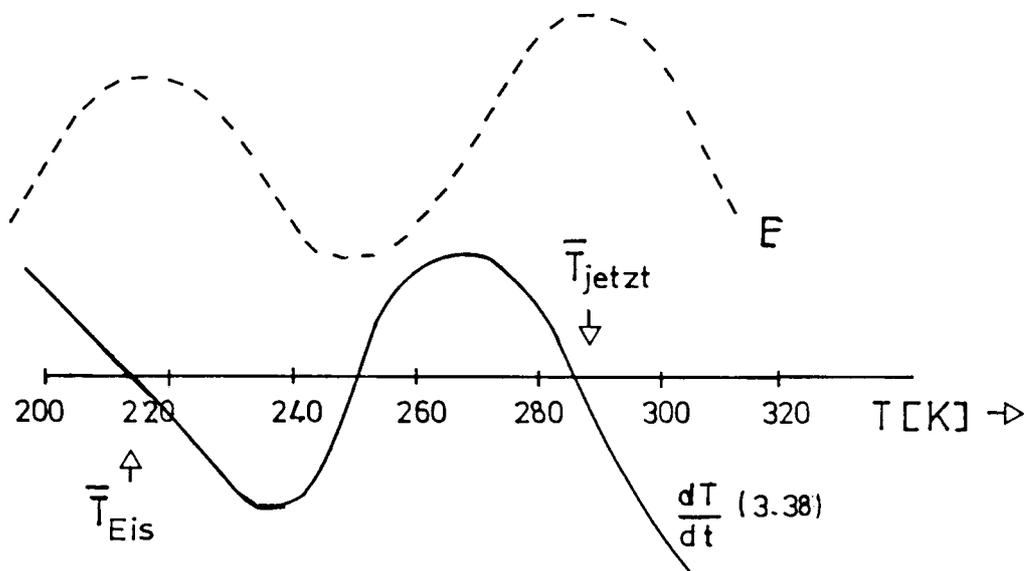


Abb. 15 Schematische Darstellung der Ableitung der Wahrscheinlichkeitsverteilung f für das Klimamodell. Die durchgezogene Kurve gibt die Tendenz $\frac{dT}{dt}$ des Modells ohne stochastischen Antrieb, wie sie aus (3.31) folgt und etwa längs der gepunkteten Kurve in Abb. 14 abgelesen werden könnte. Die strichlierte Kurve stellt E dar und wegen $f \sim \exp E$ deutet sie auch den Verlauf von f an.

3.6 Diskrete Markovprozesse

Die Temperatur T als Lösung von (3.29) stellt einen Markovprozeß vor, da ja die Kenntnis von T zu einem Zeitpunkt ausreicht, um T für alle späteren Zeitpunkte zu berechnen. Die Fokker-Planck-Gleichung (3.33) ist die zugehörige Vorhersagegleichung erster Ordnung in der Zeit f

Man betrachtet oft ein diskretes Analogon zu (3.29). Man führt also den Grenzübergang $Dt, DT^2 \rightarrow 0$ nicht aus. Sei $t = nDt$, Dt endliches Inkrement, i Zeitindex, dann läßt sich (3.29) in endlichen Differenzen schreiben

$$T_{n+1} - T_n = -\lambda T_n Dt + Q^{1/2} (W_{n+1} - W_n). \quad (3.44)$$

Also

$$T_{n+1} = aT_n + Z_n \quad (3.45)$$

, wobei

$$0 < a = 1 - \lambda Dt < 1 \quad (3.46)$$

zu wählen ist und Z_n dem Inkrement des Wienerprozesses entspricht. Man nennt (3.45) einen diskreten autoregressiven Prozeß erster Ordnung. Mit $E(Z_i) = 0$, $E(Z_{i+1}Z_i) = 0$, $E(Z_i^2) = \sigma_z^2$, findet man sofort $E(T_i) = 0$. Dieser Prozeß ist stationär und durch Quadrieren von (3.45) erhält man wegen $E(T_i Z_i) = 0$:

$$\sigma_T^2 = \sigma_z^2 / (1 - a^2) \quad (3.47)$$

Man sieht insbesondere, daß für $a = 1$ ($\lambda = 0$) keine Lösung existiert (s. (3.25)). Es versteht sich, daß (3.44) nahezu äquivalent ist zu (1.1). Der einzige Unterschied, sieht man vom Term b ab, besteht darin, daß bei (1.1) mit γ die Amplitude des Antriebs fixiert ist, bei (3.44) nur dessen Varianz.

3.7. Autokovarianz stochastischer Prozesse

Es sei (3.11) in die allgemeine Form

$$\frac{dX}{dt} + \lambda X = Z(t) \quad (3.48)$$

gebracht. Sei $Z(t)$ stationär und $E(Z) = 0$. Gesucht ist ein Zusammenhang zwischen der Autokovarianz $\gamma_{zz}(\tau)$ des Antriebs und der der Lösung, also γ_{xx} . Man hat

$$X = \int_{-\infty}^t \exp(-\lambda(t-u)) Z(u) du. \quad (3.49)$$

Mit $\tau' = t - u$ läßt sich (3.49) umschreiben zu

$$X(t) = \int_{-\infty}^{+\infty} h(\tau') Z(t-\tau') d\tau' \quad (3.50)$$

, wobei

$$\begin{aligned} h(\tau') &= \exp(-\lambda\tau') \text{ für } \tau' \geq 0 \\ h(\tau') &= 0 \text{ für } \tau' < 0 \end{aligned} \quad (3.51)$$

Offensichtlich gibt $h(\tau)$ das Gewicht an, mit dem der Antrieb zur Zeit $t-\tau$ in die Lösung eingeht. Dann ist

$$\gamma_{xx}(\tau) = E(X(t)X(t+\tau)). \quad (3.52)$$

Also

$$\gamma_{xx}(\tau) = \int_0^{\infty} \int_0^{\infty} h(\tau')h(\tau'')E(Z(t-\tau')Z(t+\tau-\tau''))d\tau'd\tau'' \quad (3.53)$$

und

$$\gamma_{xx}(\tau) = \int_0^{\infty} \int_0^{\infty} h(\tau')h(\tau'')\gamma_{zz}(\tau+\tau'-\tau'')d\tau'd\tau''. \quad (3.54)$$

Da $h(\tau)$ bekannt ist, kann man γ_{xx} aus γ_{zz} bestimmen. Sei nun $Z(t)$ weißes Rauschen mit $\gamma_{zz}(\tau) = \sigma_z^2\delta(\tau)$ so folgt

$$\begin{aligned} \gamma_{xx}(\tau) &= \sigma_z^2 \int_0^{\infty} h(\tau')h(\tau+\tau')d\tau' \\ &= \sigma_z^2 \exp(-\lambda|\tau|)/2\lambda. \end{aligned} \quad (3.55)$$

Die Autokovarianz von γ ist symmetrisch bezüglich $\tau = 0$. Wiewohl der Antrieb weiß ist, gibt es eine statistische Abhängigkeit von $X(t)$ und $X(t-|\tau|)$ zu früheren Zeitpunkten. Die Autokovarianz nimmt exponentiell mit der Zeit ab.

Für den diskreten autoregressiven Prozeß (3.45) findet man (siehe auch 1.19)

$$\gamma_{TT}(i) = \sigma_T^2 a^{|i|} \quad (i=0, \pm 1, \pm 2, \dots). \quad (3.56)$$

4. Zeitreihenanalyse

Man habe eine Beobachtungsreihe für eine Variable $x = x(t)$ als Funktion der Zeit. Dabei sei x eine Zufallsgröße. Dann nennt man die Beobachtung $x(t)$ eine Zeitreihe (time series). Parallel zur Beobachtung führt man die Zufallsvariable $X(t)$ und deren statistische Eigenschaften ein. Die beobachtete Zeitreihe $x(t)$ wird als Realisierung x von $X(t)$ angesehen. $X(t)$ und die zugehörigen Wahrscheinlichkeitsverteilungen beschreiben einen stochastischen Prozeß: Etwas abweichend von der bisherigen Schreibweise wird im folgenden die Konvention eingeführt, daß stochastische Variable X groß geschrieben werden, die zugehörigen Beobachtungswerte x klein.

Beispiel: Beobachtung der Zonalgeschwindigkeit $u(t)$ am Garchinger Turm. Dieser Vorgang ist gewiß ein stochastischer Prozeß und u eine Zufallsgröße. Entsprechend sieht man $u(t)$ als Realisierung eines Zufallsprozesses $U(t)$. Würde man an einem anderen Tag unter "gleichen" äußeren Bedingungen wieder eine Messung durchführen, so würde man eine weitere Realisierung $u(t)$ erhalten. etc. Diese Beobachtungen sind dann Realisierungen der unendlich vielen Zeitreihen, die $U(t)$ annehmen kann.

Das Ziel der Zeitreihenanalyse ist die Beschreibung des stochastischen Prozesses anhand der vorliegenden Beobachtungen. Um dieses Ziel wenigstens näherungsweise erreichen zu können, nimmt man zumeist an, daß der vorliegende stochastische Prozeß stationär ist. Eine wichtige Technik ist dabei die Fourieranalyse, doch werden zusätzlich oft Verfahren der Mustererkennung eingesetzt.

4.1. Fourieranalyse

Ziel ist die Anwendung der Fourieranalyse auf stochastische Prozesse. Zunächst seien die Methoden der klassischen Fourieranalyse rekapituliert.

4.1. a) Diskrete Fourierreihen

Das Signal $s(t)$ sei an den diskreten Zeitpunkten $t_r = r Dt$ bekannt (Dt etwa Meßintervall), die den gesamten Meßzeitraum $(-\frac{T}{2} \leq t \leq \frac{T}{2})$ in $N = T/Dt$ Abschnitte teilen, N sei gerade ($N = 2n$). Dann lassen sich die bekannten $2N$ Meßwerte s_r darstellen durch

$$s_r = A_0 + 2 \sum_{m=1}^{n-1} (A_m \cos 2\pi m f rDt + B_m \sin 2\pi m f rDt) + A_n \cos 2\pi n f rDt; \quad (4.1)$$

$(r = n, \dots, 0, 1, \dots, n-1)$

mit $f = 1/NDt$ als Basisfrequenz und rDt als Zeit. Man hat $N = 2n$ Meßwerte S_r und ebenso N Fourierkoeffizienten. Die Basisfunktionen sind orthogonal

$$\sum_{r=-n}^{n-1} \sin(2\pi k r/N) \cos(2\pi m r/N) = 0$$

$$\sum_{r=-n}^{n-1} \sin(2\pi k r/N) \sin(2\pi m r/N) = \begin{cases} 0., & k \neq m \\ N/2, k & k = m \neq \{0, n\} \\ 0., & k = m = \{0, n\} \end{cases} \quad (4.2)$$

$$\sum_{r=-n}^{n-1} \cos(2\pi kr/N) \cos(2\pi mr/N) = \begin{cases} 0., & k \neq m \\ N/2, & k = m \neq \{0, n\} \\ N., & k = m = \{0, n\} \end{cases}$$

und man erhält durch Multiplikation von (4.1) durch entsprechende Basisfunktionen und Summation

$$A_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r \cos \frac{2\pi mr}{N} \quad (4.3)$$

$$B_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r \sin \frac{2\pi mr}{N}$$

für $m = 0, 1, \dots, n$. A_0 ist der Mittelwert.

Beispiel: $N = 2$; $n = 1$

$$s_{-1} = -1; s_0 = 1$$

Aus (4.3) folgt $A_0 = 0$; $A_1 = 1$; $B_1 = 0$.

Gelegentlich schreibt man (4.1) in der Form

$$s_r = A_0 + 2 \sum_{m=1}^{n-1} R_m \cos(2\pi mfrDt + \phi_m) + R_n \cos(2\pi nfrDt) \quad (4.4)$$

, wobei

$$R_m = (A_m^2 + B_m^2)^{1/2}; \phi = \arctan(-B_m/A_m). \quad (4.5)$$

R_m (ϕ_m) wird die Amplitude (Phase) der m -ten Harmonischen genannt.

Der mittlere quadratische Wert (mittlere >power<) von s_r ist

$$\sigma_p^2 = 1/N \sum_{r=-n}^{n-1} s_r^2. \quad (4.6)$$

Unter Benutzung der Orthogonalitätseigenschaften der Basisfunktionen findet man, daß

$$1/N \sum_{r=-n}^{n-1} s_r^2 = R_0^2 + 2 \sum_{m=1}^{n-1} R_m^2 + R_n^2. \quad (4.7)$$

Die Varianz

$$\sigma^2 = 1/N \sum_{r=-n}^{n-1} (s_r - R_0)^2 = 2 \sum_{m=1}^{n-1} R_m^2 + R_n^2 \quad (4.8)$$

ist die mittlere quadratische Abweichung vom Mittelwert R_0 . Dargestellt werden die Beiträge der einzelnen Harmonischen zur mittleren power gern in Form eines Fourierlinienspektrums (Periodogramm), wobei einfach für jede Harmonische die power aufgetragen wird.

In Abb. 16 ist ein Periodogramm von weißem Lärm gezeigt, wobei $N = 240$. Die Frequenz 0.5 ist die „Nyquistfrequenz“, also die Frequenz, die einer vollen Schwingung über zwei Intervalle entspricht. Es ist offensichtlich, daß das Periodogramm von weißem Lärm unbrauchbar ist.

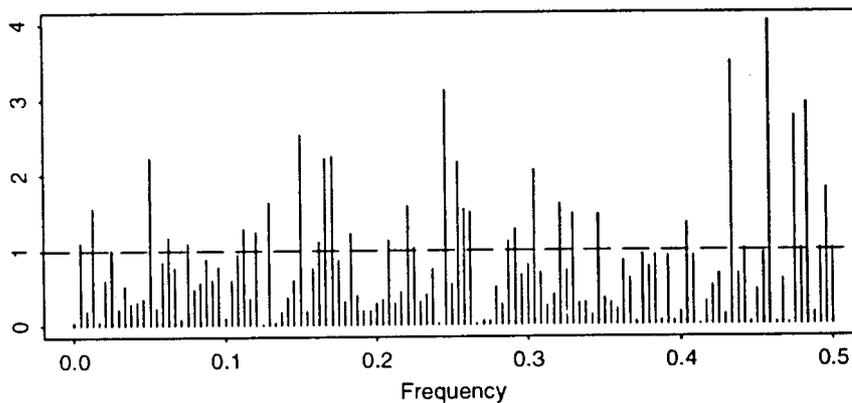


Abb. 16 Periodogramm für weißen Lärm der Länge $N = 240$. Die strichlierte Kurve entspricht der Spektraldichte des weißen Lärms.

Gelegentlich empfiehlt sich eine komplexe Schreibweise mit komplexen Amplituden $S_m = R_m \exp(i\phi_m) = A_m - i B_m$. Dann kann man (4.1) schreiben

$$s_r = \sum_{m=-n}^{n-1} S_m \exp(2\pi i m t_r / Dt N) \quad (4.9)$$

mit $s_{-m} = s_m^*$ (konjugiert komplex). Es gilt dann

$$1/N \sum_{r=-n}^{n-1} s_r^2 = \sum_{m=-n}^{n-1} |S_m|^2 \quad (4.10)$$

, wobei die Fourierkoeffizienten gemäß

$$S_m = 1/N \sum_{r=-n}^{n-1} s_r \exp(-2\pi i m r / N) \quad (4.11)$$

ausgewertet werden.

4.1.b) Fourierintegrale

Soll s nicht nur an diskreten Punkten im Intervall $-T/2 \leq t \leq T/2$ dargestellt werden, sondern kontinuierlich, so kann man (4.11) umschreiben zu

$$S_m = 1/NDt \sum_{r=-n}^{n-1} s_r Dt \exp(-2\pi imrDt/NDt). \quad (4.12)$$

Wenn nun $Dt \rightarrow 0$ und $N \rightarrow \infty$ so, daß $NDt = T$, dann hat man $rDt \rightarrow t$, $s_rDt \rightarrow sdt$. Damit geht (4.12) über in das Integral

$$S_m = 1/T \int_{-T/2}^{T/2} s(t) \exp(-2\pi imt/T) dt \quad (4.13)$$

und

$$s(t) = \sum_{m=-\infty}^{\infty} S_m \exp(2\pi imt/T). \quad (4.14)$$

Das Fourierintegral ergibt sich schließlich beim Übergang zu unendlich ausgedehnten Intervallen $-\infty \leq t \leq \infty$. Man kann zu diesem Zweck (4.14) umschreiben zu

$$s(t) = \sum_{m=-\infty}^{m=\infty} (TS_m) \exp(2\pi imt/T) \frac{1}{T}$$

Im Limes $T \rightarrow \infty$ geht $1/T$ über in df , m/T in f und TS_m in $S(f)$. Dann gilt

$$s(t) = \int_{-\infty}^{+\infty} S(f) \exp(2\pi ift) df = \int_{-\infty}^{+\infty} \tilde{S}(\omega) \exp(i\omega t) d\omega \quad (4.15)$$

und

$$(2\pi) \tilde{S}(\omega) = S(f) = \int_{-\infty}^{+\infty} s(t) \exp(-2\pi ift) dt \quad (4.16)$$

$$\int_{-\infty}^{+\infty} s^2(t) dt = \int_{-\infty}^{+\infty} |S(f)|^2 df$$

wobei $\omega = 2\pi f$.

Die Fouriertransformierte $|S(f)|$ beschreibt die Verteilung der "Signalintensität" auf die Frequenzen und kann als Dichtefunktion interpretiert werden. Wenn s z.B. in K angegeben ist, so hat $S(f)$ die Dimension K s. Die Größe $|S(f)|^2 = \Gamma(f)$ heißt Powerspektrum.

Besonders wichtig ist die Tatsache, daß sich Faltungen gut transformieren lassen. In (3.50) war so eine Faltung aufgetreten. Allgemein nennt man

$$F(t) = \int_{-\infty}^{+\infty} h(u)x(t-u)du \quad (4.17)$$

die Faltung von h und x oder (3.50) die Faltung von h und Z . Durch Transformation von (3.48) erhält man

$$i\omega X(f) + \lambda X(f) = Z(f)$$

wobei $X(f)$, $Z(f)$ die Transformierten von $X(t)$, $Z(t)$ sind. Dann folgt

$$X(f) = Z(f)/(\lambda + i\omega)$$

Andererseits ist

$$h(f) = 1/(\lambda + i\omega)$$

so daß die Fouriertransformierte $X(f)$ der Faltung (3.50) gerade gleich dem Produkt der Fouriertransformierten von h und Z ist. Das gilt allgemein. Für die obige Faltung (4.17) hat man

$$F(f) = h(f) \cdot X(f) \quad (4.18)$$

Wichtige Beispiele von Fouriertransformierten:

$$\begin{aligned} \text{a)} \quad s(t) &= \exp(-|t|\lambda) \\ S(f) &= 2\lambda/(\lambda^2 + (2\pi f)^2) \end{aligned} \quad (4.19)$$

$S(f)$ hat ihr Maximum bei $f = 0$, ist symmetrisch bezüglich $f = 0$ und fällt $\sim 1/f^2$ zu hohen Frequenzen hin ab.

b) Delta-Funktion: die Delta-Funktion ist so definiert, daß

$$\delta(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}$$

und

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1.$$

Man stelle sich am Ursprung ein Rechteck mit Höhe a und Breite $2b$ vor. Sei $2ab = 1$. Dann ist die Fouriertransformierte dieses Rechtecks $s(t)$:

$$S(f) = \sin 2\pi fb / 2\pi fb. \quad (4.20)$$

Geht nun b gegen Null, so hat man $S(f) \rightarrow 1$. Die Transformierte einer Delta-Funktion ist also eine Konstante. Alle Frequenzen sind gleich stark an der Repräsentation der Deltafunktion beteiligt. Umgekehrt gilt, daß die Transformierte einer Konstanten eine Deltafunktion ist. Hat man eine Abfolge

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nDt)$$

von Deltafunktionen, so ist deren Transformierte ebenfalls eine Folge äquidistanter Deltafunktionen

$$S(f) = 1/Dt \sum_{n=-\infty}^{+\infty} \delta(f - n/Dt). \quad (4.21)$$

Bis hierher wurde angenommen, daß Daten beliebig verfügbar sind. In der Praxis ist das nicht der Fall. Zumeist kennt man von einer Funktion $s(t)$ nur die Werte in einem Intervall $-T/2 \leq t < T/2$. Will man dennoch die Fouriertransformierte von s bestimmen, so wird die Beschränkung der Kenntnis der Funktion auf jenes Intervall zu Verzerrungen bei der Berechnung der Transformierten führen. Man definiert das Fenster (data window)

$$w(t) = \begin{cases} 1, & |t| \leq T/2 \\ 0, & |t| > T/2 \end{cases}. \quad (4.22)$$

Sei nun $s(t)$ ein Signal, das für den Bereich $-\infty < t < \infty$ definiert ist, so kann das tatsächlich gemessene Signal durch $s_T = s(t) w(t)$ dargestellt werden. Die Transformierte eines Produkts ist die Faltung der Transformierten von s und w , also wird f_1 mit f_2 gefaltet:

$$f_1(u) * f_2(u) = \int_{-\infty}^{+\infty} f(\hat{u}) f(u - \hat{u}) d\hat{u}$$

$$s_T(f) = \int_{-\infty}^{+\infty} S(g) W(f-g) dg \quad (4.23)$$

, wobei das spektrale Fenster (spectral window) $W(f)$ eben die Transformierte des Fensters w ist, in unserem Falle also

$$W(f) = T \sin(\pi f T) / (\pi f T) \quad (4.24)$$

(siehe Fouriertransformierte eines Rechtecks!). Abb. 17 gibt anhand von Beispielen einen Eindruck von der Situation. Das perfekte Spektrum $S(f)$ ist rechts oben gegeben. Man hat drei Linien bei den Frequenzen f_0, f_1, f_2 . In a) ist $f_2 - f_1 = 1/T$ gewählt. Dann werden die beiden Linien mit Frequenzen f_1 bzw. f_2 nicht aufgelöst und erscheinen in s_T als eine Linie. Verdoppelt man T (b), so werden beide Linien aufgelöst. Will man also

zwei Linien trennen, so muß die Intervallgröße (Meßdauer) $T \geq (f_2 - f_1)^{-1}$ sein. Sind die Fenster nicht rechteckig (c), so sollte man $T \geq 2/(f_2 - f_1)$ verlangen.

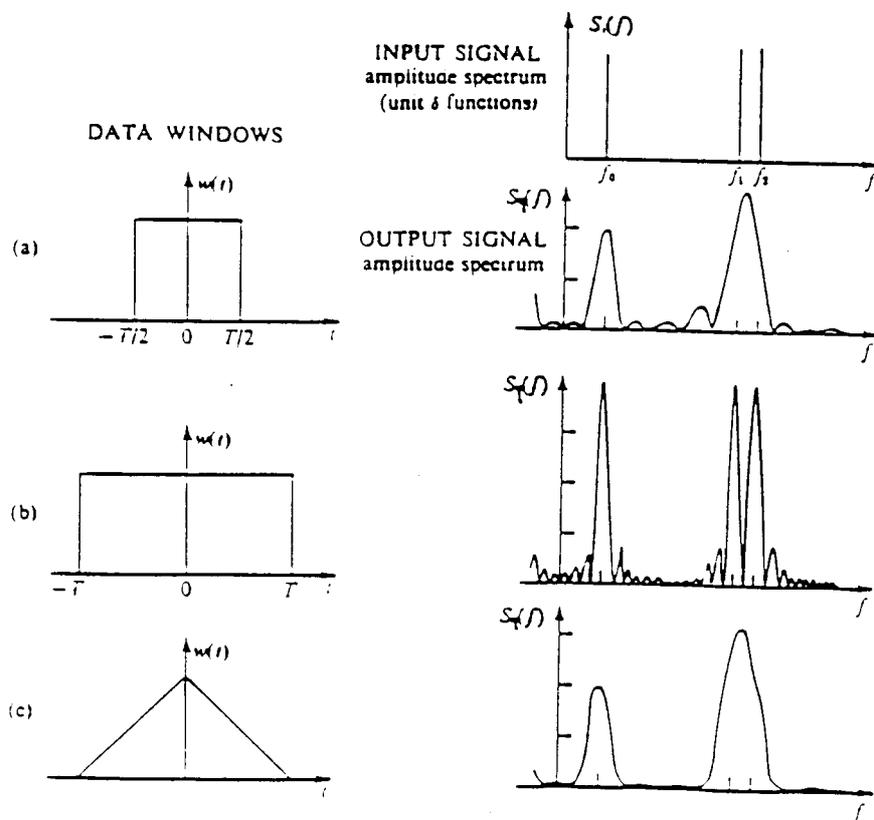


Abb. 17 Korrektes Spektrum als Eingabe; darunter die Spektren, die sich ergeben, wenn diese Eingabe durch die entsprechenden Fenster betrachtet wird.

4.1. c) Stichprobenfehler und Verfälschung (Aliasing)

Während eines Meßvorgangs werden kontinuierliche Signale $s(t)$ in Intervallen Dt abgelesen, so daß eine diskrete Reihe s_i von Meßwerten entsteht. Diese diskrete Meßreihe s_i kann man sich entstanden denken als Produkt

$$s_i(t) = s(t) i(t) \quad (4.25)$$

des tatsächlichen Signals mit einer Abfolge von Deltafunktionen

$$i(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nDt). \quad (4.26)$$

Wieder gilt

$$S_i(f) = \int_{-\infty}^{+\infty} S(f-g) I(g) dg \quad (4.27)$$

, wobei (siehe (4.21))

$$I(g) = 1/Dt \sum_{n=-\infty}^{+\infty} \delta(g-n/Dt). \quad (4.28)$$

Also

$$S_i(f) = 1/Dt \sum_{n=-\infty}^{+\infty} S(f-n/Dt). \quad (4.29)$$

Weiß man nun, daß $S(f) = 0$ wenn $f > 1/2Dt$, dann trägt nur $n = 0$ zu dem Summanden in (3.29) bei. Man kann dann $S(f)$ einfach dadurch erhalten, daß man $S_i(f)$ mit $H(f)$ multipliziert

$$H(f) = \begin{cases} Dt, & |f| \leq 1/2Dt \\ 0, & |f| > 1/2Dt \end{cases} \quad (4.30)$$

Da Multiplikation im Frequenzraum einer Konvolution im Zeitraum entspricht, so gilt

$$s(t) = \sum_{i=1}^N \int_{-\infty}^{+\infty} \sin(\pi u/Dt) / (\pi u/Dt) s_i(t-u) du. \quad (4.31)$$

Dabei entspricht der Sinusterm dem Rechteck (4.30). Die Funktion $\sin(\pi u/Dt)/(\pi u/Dt)$ ist der geeignete Filter, um das kontinuierliche Signal $s(t)$ aus dem diskreten s_i zu interpolieren.

Ist aber $S(f) \neq 0$ für Frequenzen $> 1/2Dt$, so treten in (4.29) Summanden mit $n \neq 0$ auf. Diese müssen als Verfälschung (alias) angesehen werden. Die Frequenz $f_N = 1/(2Dt)$ heißt die Nyquistfrequenz.

Ein krasses Beispiel: $S(t) = \cos(2\pi t/T)$ werde mit dem Intervall $Dt = T$ abgelesen. Dann ist $s_i(t) = 1$ und man bekommt eine völlige Verfälschung.

4.2. Powerspektrum eines stochastischen Prozesses

Bei der Bestimmung des Powerspektrums eines stochastischen Prozesses hat man mit speziellen Schwierigkeiten zu rechnen. Zunächst muß man annehmen, daß man Kenntnis des stochastischen Prozesses nur für das Intervall $-T/2 \leq t \leq T/2$ hat. Dann ist

$$c_{xx}(\tau) = \begin{cases} \frac{1}{T} \int_0^{T-|\tau|} (X(t) - \bar{X})(X(t+|\tau|) - \bar{X}) dt, & 0 \leq |\tau| \leq T \\ 0, & |\tau| > T \end{cases} \quad (4.32)$$

eine vernünftige Schätzung der Autokovarianz $\gamma_{xx}(\tau)$. Gelegentlich wird $1/T$ durch $1/(T-|\tau|)$ ersetzt.

Den Erwartungswert $E[c_{xx}(\tau)]$ von (4.32) erhält man direkt aus (4.32) mit $E[X(t)] = 0$:

$$E[c_{xx}(\tau)] = E\left[\frac{1}{T} \int_0^{T-|\tau|} X(t)X(t+\tau)dt\right] \quad (4.33)$$

$$= \frac{1}{T} \int_0^{T-|\tau|} \gamma_{xx}(\tau)dt = \begin{cases} \gamma_{xx}(\tau) \left(1 - \frac{|\tau|}{T}\right), & 0 \leq |\tau| \leq T \\ 0, & |\tau| > T \end{cases}$$

Will man das Powerspektrum der Zeitreihe $X(t)$ eines stochastischen Prozesses bestimmen, so kann man nicht so wie bei einem determinist. Signal vorgehen. Es hat dann keinen Sinn, die Zeitreihe einfach in den Frequenzraum zu transformieren.

Im deterministischen Fall hat man in Analogie zu (4.10) bzw. (4.16):

$$\frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt = \sum_{m=-\infty}^{\infty} |X_m|^2$$

Für $T \rightarrow \infty$ erhält man die Varianz

$$\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt = \lim_{T \rightarrow \infty} \sum_{m=-\infty}^{\infty} (T|X_m|^2) \frac{1}{T} = \int_{-\infty}^{+\infty} \Gamma(f)df$$

wobei eben

$$\Gamma(f) = \lim_{T \rightarrow \infty} T|X_m|^2$$

das Powerspektrum ist. Wegen

$$X_m = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \exp(-2\pi imt / T) dt$$

kann man die Funktion $T|X_m|^2$ umschreiben zu

$$T|X_m|^2 = C_{xx}(f) = \frac{1}{T} \left| \int_{-T/2}^{T/2} x(t) \exp(-2\pi ift) dt \right|^2$$

Die Größe $C_{xx}(f)$ nennt man das Stichprobenspektrum und $\Gamma(f) = \lim_{T \rightarrow \infty} C_{xx}(f)$.

Im Fall einer diskreten Zeitreihe hat man

$$C_{xx}(f) = \frac{Dt}{N} \left| \sum_{r=-n}^{n-1} x_r \exp(-2\pi i f r Dt) \right|^2$$

Bei einem stochastischen Prozeß hat man, wie schon auf den Periodogrammen in Abb. 17 zu ersehen ist, keine Garantie, daß $C_{xx}(f)$ gegen einen Grenzwert $\Gamma(f)$ konvergiert, wenn die Intervallfläche T zunimmt.

Beispiel: Von einer Normalverteilung werden 400 Zufallsgrößen gezogen und als diskrete Zeitreihe angeordnet: x_i , $i = 1 \dots 400$. Dann wird das Stichprobenspektrum $C_{xx}(f)$ berechnet, wobei zunächst nur die ersten 50 Werte benutzt werden, dann die ersten 100, 200 und schließlich alle 400. Es wird also die Intervalllänge, für die C_{xx} bestimmt wird, vergrößert. Abb. 18 zeigt das Stichprobenspektrum für $N = 50$, und $N = 100$ wobei 0.5 die Nyquistfrequenz bedeutet. Das wahre Spektrum des Prozesses ist bekannt und wird durch die Gerade wiedergegeben. Man sieht, daß das Stichprobenspektrum für $N = 100$ genauso unruhig ist, wie für $N = 50$. Steigert man auf $N = 200$ oder $N = 400$ so tritt keine Verbesserung ein.

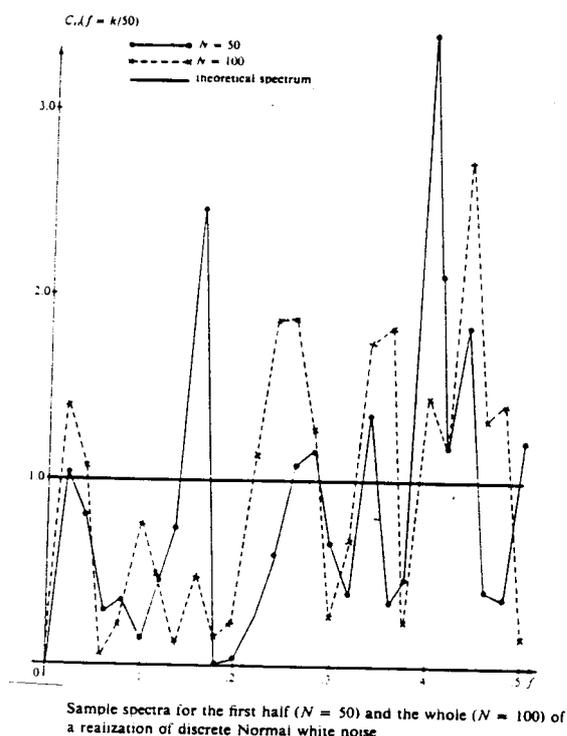


Abb. 18

Will man das Powerspektrum eines stochastischen Prozesses schätzen, so muß man dessen Autokorrelation heranziehen. Die folgende Beziehung gilt zwar allgemein, also nicht nur für stochastische Prozesse, doch muß sie für stochastische Prozesse auf jeden Fall verwendet werden. Die Definition des Stichprobenspektrums liefert

$$C_{xx}(f) = 1/T \int_{-T/2}^{T/2} x(t) \exp(-2\pi i f t) dt \int_{-T/2}^{T/2} x(t') \exp(+2\pi i f t') dt'. \quad (4.35)$$

Die Transformation $\tau = t - t'$, $v = t'$ liefert

$$C_{xx}(f) = \int_0^T \frac{1}{T} \left[\int_{-T/2}^{T/2-\tau} x(v)x(v+\tau)dv \exp(-2\pi if\tau) \right] d\tau \\ + \int_{-T}^0 \left[\frac{1}{T} \int_{-T/2-\tau}^{T/2} x(v)x(v+\tau)dv \right] \exp(-2\pi if\tau) d\tau \quad (4.36)$$

Mit der Definition der Autokovarianzfunktion hat man

$$C_{xx}(f) = \int_{-T}^T c_{xx}(\tau) \exp(-2\pi if\tau) d\tau; \quad -\infty \leq f \leq \infty. \quad (4.37)$$

Das Stichprobenspektrum ist die Fouriertransformierte der Autokovarianz der Stichprobe. Dies gilt allgemein. Man kann (4.37) umkehren und erhält

$$c_{xx}(\tau) = \int_{-\infty}^{\infty} C_{xx}(f) \exp(2\pi if\tau) df, \quad -T \leq \tau \leq T \quad (4.38)$$

und, insbesondere, für $\tau = 0$:

$$c_{xx}(0) = s_T^2 = \int_{-\infty}^{\infty} C_{xx}(f) df. \quad (4.39)$$

Man kann also am Stichprobenspektrum C_{xx} ablesen wie die Power über die Frequenzen verteilt ist.

Im diskreten Fall gilt

$$C_{xx}(f) = Dt \sum_{k=-(N-1)}^{(N-1)} c_{xx}(k) \exp(-2\pi ifkDt), \quad -1/2Dt \leq f < 1/2Dt \quad (4.40)$$

und

$$c_{xx}(\tau) = \int_{-1/2Dt}^{1/2Dt} C_{xx}(f) \exp(2\pi if\tau) df, \quad -NDt \leq \tau < Ndt. \quad (4.41)$$

Damit läßt sich das Powerspektrum für einen stochastischen Prozeß festlegen. Man habe eine Zeitreihe $x(t)$ für das Intervall $-T/2 \leq t \leq T/2$, die als Realisierung eines stochastischen Prozesses angesehen werden kann mit der zugehörigen Zufallsgröße $X(t)$. Man bildet den Erwartungswert für (4.37)

$$E(C_{XX}(f)) = \int_{-T}^T E(c_{XX}(\tau)) \exp(-2\pi i f \tau) d\tau \quad (4.42)$$

, was wegen (4.33) ergibt

$$E(C_{XX}(f)) = \int_{-T}^T \gamma_{XX}(\tau) (1 - |\tau|/T) \exp(-2\pi i f \tau) d\tau. \quad (4.43)$$

Für $T \rightarrow \infty$ erhält man das Powerspektrum

$$\Gamma_{XX}(f) = \lim_{T \rightarrow \infty} E(C_{XX}(f)) = \int_{-\infty}^{+\infty} \gamma_{XX}(\tau) \exp(-2\pi i f \tau) d\tau. \quad (4.44)$$

Damit hat man eine Vorschrift zur Berechnung des Powerspektrums eines stochastischen Prozesses. Durch Umkehrung bekommt man wie üblich

$$\gamma_{XX}(\tau) = \int_{-\infty}^{+\infty} \Gamma_{XX}(f) \exp(2\pi i f \tau) df \quad (4.45)$$

und

$$\gamma_{XX}(0) = \sigma_X^2 = \int_{-\infty}^{+\infty} \Gamma_{XX}(f) df.$$

Man kann also aus dem Powerspektrum ablesen, wie die Varianz von $X(t)$ über die Frequenzen verteilt ist.

Im diskreten Fall gilt analog

$$\Gamma_{XX}(f) = DT \sum_{k=-\infty}^{\infty} \gamma_{XX}(k) \exp(-2\pi i k f D t), \quad -1/2Dt \leq f < 1/2Dt \quad (4.46)$$

und

$$\gamma_{XX}(k) = \int_{-1/2Dt}^{1/2Dt} \Gamma_{XX}(f) \exp(2\pi i f k D t) df, \quad k = 0, \pm N. \quad (4.47)$$

Beispiele:

a) Powerspektrum von weißem Lärm

Die Autokovarianzfunktion von weißem Lärm ist $Z(t)$ ist $\gamma_{ZZ} = \sigma_Z^2 \delta(\tau)$, da ja aufeinanderfolgende Werte von Z definitionsgemäß nicht miteinander korreliert sind. Strikt gesprochen ist weißer Lärm als stochastischer Prozeß nicht zugelassen, da die

Varianz $\gamma_{ZZ}(0)$ unendlich groß ist. Man erhält aus (4.44) sofort $\Gamma_{ZZ} = \sigma_Z^2$. Das Powerspektrum von weißem Lärm ist eine Konstante. Man erinnere sich an Abb. 18, wo weißer Lärm zugrunde liegt, wo aber die üblichen Verfahren zur Berechnung von Powerspektren gescheitert waren.

b) Powerspektrum für lineare Prozesse

Durch (3.55) ist die Autokovarianz eines linearen Prozesses mit Antrieb Z gegeben. Folglich kann man über (4.44) das Powerspektrum errechnen

$$\Gamma_{XX}(f) = \int_{-\infty}^{+\infty} \gamma_{XX}(\tau) \exp(-2\pi i f \tau) d\tau \quad (4.48)$$

$$\begin{aligned} &= \int_{-\infty}^{+\infty} \exp(-2\pi i f \tau) \int_0^{\infty} \int_0^{\infty} h(\tau') h(\tau'') \gamma_{ZZ}(\tau + \tau' - \tau'') d\tau' d\tau'' d\tau \\ &= \int_0^{\infty} h(\tau') \exp(2\pi i f \tau') d\tau' \int_0^{\infty} h(\tau'') \exp(-2\pi i f \tau'') d\tau'' \int_{-\infty}^{+\infty} \gamma_{ZZ}(y) \exp(-2\pi i f y) dy \quad (4.49) \end{aligned}$$

, wobei $y = \tau + \tau' - \tau''$. Es folgt

$$\Gamma_{XX}(f) = H^*(f) H(f) \Gamma_{ZZ}(f) = |H(f)|^2 \Gamma_{ZZ}(f). \quad (4.50)$$

Dies ist ein wichtiges Resultat: das Powerspektrum der Lösung einer linearen Differentialgleichung gewinnt man, indem man das Powerspektrum des Antriebs mit $|H(f)|^2$ multipliziert. Man hätte dieses Resultat anhand des Faltungssatzes vermuten können, doch wäre dies insofern unsicher gewesen, da es Fouriertransformierte eines stochastischen Prozesses nicht gibt.

Für den autoregressiven Prozeß erster Ordnung (3.45) findet man

$$\Gamma_{XX}(f) = \frac{\sigma_Z^2}{(2\pi f)^2 + a^2}. \quad (4.51)$$

Das Maximum der Power liegt bei $f = 0$, also bei langsamen Veränderungen (rotes Spektrum). Dagegen ist die Power gering für hohe Frequenzen. Dies ist einleuchtend, (3.50) ist ja die Differentialgleichung für ein angefachtetes System mit linearer Dämpfung. Die Dämpfung rascher Veränderungen (große f) ist effektiver als die von langsamen.

Etwa das Powerspektrum der Temperatur in der Deckschicht des Ozeans läßt sich mit Hilfe von (4.51) berechnen. Man weiß, daß das Powerspektrum des Flusses fühlbarer und latenter Wärme an der Ozeanoberfläche nahezu weiß ist. Damit ist (4.51) direkt anwendbar mit $\lambda = a$. Abbildung 19 zeigt das Powerspektrum der Meerestemperatur am Wetterschiff India gemäß Beobachtungen und gemäß (4.51), wobei sowohl σ_Z als auch

λ so gewählt wurden, daß eine optimale Übereinstimmung von Theorie und Beobachtung eintritt. Die Deckung ist recht gut.

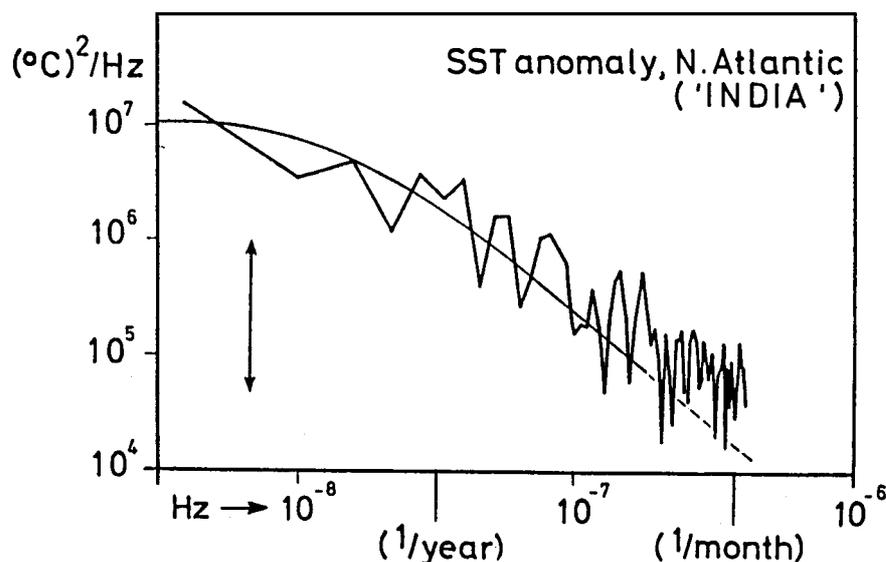


Abb. 19 Powerspektrum der Meerestemperatur am Wetterschiff India gemäß Messungen und gemäß (4.51).

4.3 Kreuzkorrelation und Kreuzspektrum

Während bisher nur Zeitreihen betrachtet wurden, die auf einer Variablen fußen, so soll nun die Situation untersucht werden, wo zwei Variable gemeinsam zu analysieren sind. Man möchte herausfinden, ob und wie die beiden Variablen zusammenhängen. Vom Konzept her treten hier keine neuen Probleme auf.

Beobachtet sei eine bivariate Zeitreihe $\{x_1(t), x_2(t)\}$, die als Realisierung eines bivariaten stochastischen Prozesses $\{X_1(t), X_2(t)\}$ gilt. Der Prozeß sei stationär, so daß die ersten und zweiten Momente dieser Zeitreihen nicht von t abhängen. Es gelte $E[X_i(t)] = \mu_i$ und die Autokovarianzen sind

$$\gamma_{X_1X_1}(\tau) = E[(X_1(t) - \mu_1)(X_1(t + \tau) - \mu_1)] \quad (4.52)$$

, analog $\gamma_{X_2X_2}(\tau)$.

Die Kreuzkovarianzfunktionen sind

$$\gamma_{X_1X_2}(\tau) = E[(X_1(t) - \mu_1)(X_2(t + \tau) - \mu_2)] \quad (4.53)$$

$$\gamma_{X_2X_1}(\tau) = E[(X_2(t) - \mu_2)(X_1(t + \tau) - \mu_1)]. \quad (4.54)$$

Der Einfachheit halber wird geschrieben: $\gamma_{X_1X_1} = \gamma_{11}$, $\gamma_{X_1X_2} = \gamma_{12}$ etc.

Es gilt $\gamma_{ii}(0) = \text{Varianz}(X_i(t)) = \sigma_{X_i}^2$, $i = 1, 2$

$$\begin{aligned}\gamma_{ii}(\tau) &= \gamma_{ii}(-\tau) \\ \gamma_{12}(\tau) &= \gamma_{21}(-\tau)\end{aligned}\quad (4.55)$$

Die Kovarianz ist im allgemeinen nicht symmetrisch bezüglich $\tau = 0$. Analog definiert man die Kreuzkorrelationsfunktion

$$\rho_{12}(\tau) = \frac{\gamma_{12}(\tau)}{(\gamma_{11}(0)\gamma_{22}(0))^{1/2}} = \frac{\gamma_{12}(\tau)}{\sigma_1\sigma_2}. \quad (4.56)$$

Ganz in Analogie zur univariaten Situation schätzt man die Kreuzkovarianzfunktion, wenn X_1 und X_2 keine Mittelwerte haben, gemäß:

$$c_{X_1X_2}(\tau) = \begin{cases} \frac{1}{T} \int_{-T/2}^{T/2-\tau} X_1(t)X_2(t+\tau)dt, & 0 \leq \tau \leq T \\ \frac{1}{T} \int_{-T/2-\tau}^{T/2} X_1(t)X_2(t+\tau)dt, & -T \leq \tau \leq 0 \end{cases}. \quad (4.57)$$

Bildet man die Erwartungswerte, so folgt

$$E(c_{X_1X_2}(\tau)) = \left(1 - \frac{|\tau|}{T}\right) \gamma_{X_1X_2}(\tau). \quad (4.58)$$

Das heißt, daß die Stichprobenkreuzkovarianzfunktion γ_{12} unterschätzt wird, um so mehr, je größer τ .

Das Kreuzspektrum

Hat man zwei Zeitreihen $x_1, x_2(t)$ im Intervall $-T/2 \leq t \leq T/2$ und seien $X_i(f)$ die Fourier-transformierten ($i = 1,2$) von x_i . Das Stichprobenpowerspektrum für jede Variable war definiert worden zu

$$C_{X_iX_i}(f) = T|X_i(f)|^2. \quad (4.59)$$

Dann liegt es nahe, das Stichprobenkreuzspektrum zu

$$C_{X_1X_2}(f) = TX_1^*X_2(f) \quad (4.60)$$

festzulegen (Stern konjugiert komplex).

Durch Aufspaltung nach Amplitude und Phase kann man noch zusätzliche Information gewinnen. Man kann schreiben

$$X_i(f) = A_i(f) \exp(i\phi_i(f)) \quad (4.61)$$

, wobei die Amplitude $A_i(f)$ eine positive und gerade Funktion von f ist, während die Phasenfunktion ungerade ist ($\phi(f) = -\phi(-f)$).

Das Kreuzspektrum ergibt sich zu

$$C_{x_1x_2}(f) = A_1(f)A_2(f) \exp(i(\phi_2 - \phi_1)) \quad (4.62)$$

verkürzt geschrieben

$$C_{12}(f) = A_{12}(f) \exp(i\phi_{12}(f)). \quad (4.63)$$

Die Kovarianz zwischen zwei Zeitreihen kann beschrieben werden durch das Stichprobenamplitudenkreuzspektrum

$$A_{12}(f) = T A_1(f) A_2(f) \quad (4.64)$$

und das Phasenspektrum

$$\phi_{12}(f) = \phi_2(f) - \phi_1(f). \quad (4.65)$$

Das Phasenspektrum zeigt einem, ob die eine Zeitreihe bei gegebener Frequenz gegenüber der anderen in der Phase voreilt oder nachhinkt. Sehr gebräuchlich ist eine alternative Darstellung für (4.60). Da (4.63) komplex ist, kann man schreiben

$$C_{12}(f) = L_{12}(f) - i Q_{12}(f) \quad (4.66)$$

, wobei

$$L_{12}(f) = A_{12}(f) \cos \phi_{12}(f), \quad Q_{12}(f) = -A_{12}(f) \sin \phi_{12}(f) \quad (4.70)$$

und

$$A_{12}^2(f) = L_{12}^2(f) + Q_{12}^2(f), \quad \phi_{12}(f) = \arctan \frac{Q_{12}}{L_{12}} \quad (4.71)$$

Bedenkt man, daß $X_i = A_i(\cos \phi_i + i \sin \phi_i)$ so sieht man alsbald, daß das Kospektrum $L_{12}(f)$ die Kovarianz zwischen beiden Cosinustermen und beiden Sinustermen angibt, also die Kovarianz zwischen den Termen, die in Phase sind, während das Quadraturpektrum Q_{12} die Kovarianz zwischen den Cosinus- und Sinustermen angibt. Gern wird auch das Kohärenzspektrum

$$C_o(t) = \left(L_{12}^2 + Q_{12}^2 \right)^{1/2} / (L_{11} \cdot L_{22})^{1/2} \quad (4.72)$$

betrachtet.

Als Beispiel diene der Fall, wo

$$x_i = A_i \cos(2\pi f_0 t + \phi_i)$$

also zwei monochromatische Wellen mit verschiedener Amplitude und Phase vorliegen. Dann gilt für die Fouriertransformierte für das Intervall T

$$X_i(f) = \frac{A_i}{2T} \left(\exp(i\phi_i) \frac{\sin \pi(f - f_0)T}{\pi(f - f_0)} + \exp(-i\phi_i) \frac{\sin \pi(f + f_0)T}{(f + f_0)\pi} \right). \quad (4.73)$$

Damit bekommt man wie bisher die Powerspektren

$$C_{x_i x_i}(f) = T X_i(f)^2$$

, die sich für $T \rightarrow \infty$ zu

$$\frac{1}{4} A_i^2 (\delta(f - f_0) + \delta(f + f_0))$$

ergeben. Wie man es erwarten sollte, hat das Powerspektrum bei $f + f_0$ je eine Spitze und verschwindet sonst. Das Kreuzspektrum ergibt sich für $T \rightarrow \infty$ zu

$$C_{x_1 x_2}(f) = \frac{1}{4} A_1 A_2 (\exp(-i(\phi_2 - \phi_1)) \delta(f + f_0) + \exp(i(\phi_2 - \phi_1)) \delta(f - f_0)).$$

Sind beide Wellen in Phase, also $\phi_1 = \phi_2$, so gibt das Kreuzspektrum de facto das Produkt beider Amplituden. Beachtet man $\exp(i(\phi_2 - \phi_1)) = \cos(\phi_2 - \phi_1) + i \sin(\phi_2 - \phi_1) = \sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 + i \sin \phi_2 \cos \phi_1 - i \sin \phi_1 \cos \phi_2$ etc., so ergibt sich

$$L_{12}(f) = (A_1 \cos \phi_1 A_2 \cos \phi_2 + A_1 \sin \phi_1 A_2 \sin \phi_2) (\delta(f + f_0) + \delta(f - f_0)) / 4$$

$$Q_{12}(f) = (A_1 \cos \phi_1 A_2 \sin \phi_2 - A_1 \sin \phi_1 A_2 \cos \phi_2) (-\delta(f + f_0) + \delta(f - f_0)) / 4$$

Es beschreibt also in der Tat das Kospektrum die Kovarianz der Komponenten, die in Phase sind, und das Quadratspektrum die Kovarianz zwischen Sinus- und Cosinus Komponenten.

Bisher wurde stillschweigend angenommen, daß x_1 und x_2 nicht Realisierungen eines stochastischen Prozesses sind. Ist das nicht der Fall, so hat man wie für den univariaten Fall auch jetzt das Problem, daß Kreuzspektren, die von einer Stichprobe ermittelt werden, nicht konvergieren, auch wenn T beliebig groß wird. Der Ausweg, der zu beschreiten ist, ist derselbe wie im univariaten Fall. Man berechnet die Spektren aus den Kovarianzfunktionen. Es gilt (Nachweis analog zum univariaten Fall):

$$\lim E(C_{x_1x_2}(f)) = \Gamma_{x_1x_2}(f) = \int_{-\infty}^{+\infty} \gamma_{x_1x_2}(\tau) \exp(-2\pi i f \tau) d\tau \quad (4.74)$$

Das Kreuzspektrum ist die Fouriertransformierte der Kreuzkorrelationsfunktion. Setzt man

$$\lambda_{12} = 0.5 (\gamma_{12}(\tau) + \gamma_{12}(-\tau)) \quad (4.75)$$

$$\psi_{12} = 0.5 (\gamma_{12}(\tau) - \gamma_{12}(-\tau))$$

und geht damit in (4.74), so ergibt sich

$$\Gamma_{12}(f) = \Lambda_{12}(f) - i\Psi_{12}(f) \quad (4.76)$$

, wobei sich das Kospektrum zu

$$\Lambda_{12}(f) = \int_{-\infty}^{+\infty} \lambda_{12}(\tau) \cos 2\pi f \tau d\tau \quad (4.77)$$

ergibt und das Quadraturpektrum zu

$$\Psi_{12}(f) = \int_{-\infty}^{+\infty} \psi_{12}(\tau) \sin 2\pi f \tau d\tau. \quad (4.78)$$

Man kann via

$$\Gamma_{12}(f) = \alpha_{12}(f) \exp(i\phi_{12}(f)) \quad (4.79)$$

ein Amplitudenspektrum

$$\alpha_{12}(f) = \left(\Lambda_{12}^2(f) + \Psi_{12}^2(f) \right)^{1/2} \quad (4.80)$$

und ein Phasenspektrum

$$\phi_{12}(f) = \arctan \left(-\frac{\Psi_{12}(f)}{\Lambda_{12}(f)} \right) \quad (4.81)$$

einführen.

Als weitere Beispiele seien die Kreuzspektren einfacher diskreter linearer Prozesse betrachtet. Für einen diskreten Prozeß hat man das Kreuzspektrum in Analogie zu (4.78)

$$\Gamma_{12}(f) = T \sum_{j=-\infty}^{+\infty} \gamma_{12}(j) \exp(-2\pi i f j). \quad (4.82)$$

Im folgenden sei $E(Z_1) = 0$, $E(Z_2) = 0$, $E(Z_1^2) = \sigma_1^2$, $E(Z_2^2) = \sigma_2^2$ und $T = 1$.

Man habe zunächst den Prozeß

$$X_{2k} = Z_{2k}, X_{1k} = Z_{1k}.$$

Dann ist $\gamma_{12}(j) = E(Z_{1k} Z_{2k+j}) = 0$.

Dies heißt entsprechend zu (4.75), (4.76), daß $\lambda_{12} = \psi_{12} = 0$ und daß somit Ko- und Quadraturspektrum verschwinden.

Etwas interessanter ist der Prozeß

$$X_{2k} = Z_{2k} + \beta_1 Z_{1k}$$

$$X_{1k} = Z_{1k}$$

, also

$$X_{2k} = \beta_1 X_{1k} + Z_{2k}$$

Dann ist

$$\gamma_{12}(0) = \beta_1 \sigma_1^2$$

und

$$\gamma_{12}(j) = 0$$

für $j \neq 0$. Daraus folgt mit (4.82) und (4.75)

$$\Gamma_{12}(f) = \beta_1 \sigma_1^2$$

$$\Lambda_{12}(f) = \beta_1 \sigma_1^2$$

$$\Psi_{12}(f) = 0.$$

Wenn also zwei Prozesse nur zu je gleichen Zeiten korreliert sind, so ist das Kospektrum eine Konstante, das Quadraturspektrum muß dagegen verschwinden.

Sei

$$X_{2k} = Z_{2k} + \beta_1 Z_{1k-d}$$

$$X_{1k} = Z_{1k}$$

, also

$$X_{2k} = Z_{2k} + \beta_1 X_{1k-d}$$

, so daß X_2 durch X_1 mit Verschiebung d beeinflußt wird (d natürliche Zahl). Dann folgt

$$\gamma_{12}(f) = \begin{cases} \beta_1 \sigma_1^2, & j = d \\ 0 & \text{sonst} \end{cases}$$

Also,

$$\Gamma_{12}(f) = \beta_1 \sigma_1^2 \exp(-2\pi i f d).$$

Das Kreuzspektrum ist jetzt frequenzabhängig und man hat:

$$\lambda_{12}(j) = \beta_1 \sigma_1^2 / 2, \quad |j| = d$$

$$\psi_{12}(j) = +\beta_1 \sigma_1^2 / 2, \quad j = \pm d$$

Also

$$\Lambda_{12}(f) = \beta_1 \sigma_1^2 \cos 2\pi f d$$

$$\Psi_{12}(f) = \beta_1 \sigma_1^2 \sin 2\pi f d$$

und das Amplitudenspektrum ist mit

$$\alpha_{12}(f) = \beta_1 \sigma_1^2$$

eine Konstante, während das Phasenspektrum

$$\phi_{12}(f) = -2\pi f d$$

ist. Dies sollte so sein, da man bei der Frequenz f gerade fd Zyklen braucht, um die Verschiebung d zu überbrücken.

Beispiele:

a) Turbulenzmessungen an einem Mast (Hennemuth, 1978)

Gemessen wurde der mittlere Wind \bar{u} und die Abweichungen u' , w' . Die zeitlichen Spektren sind in räumliche verwandelt, gemäß der "Taylorhypothese", daß die Wirbel in der Grenzschicht mit der Geschwindigkeit \bar{u} vorbeidriften. Dann gilt zwischen Frequenz und Wellenzahl $k = 2\pi/L$

$$f = \bar{u} k / 2\pi$$

b) Ebenso können bei Autokorrelationen Verschiebungen τ in Abständen r zum Meßpunkt gemäß

$$\tau = u/r$$

umgerechnet werden. Für die angegebenen Messungen liegt \bar{u} bei 3ms^{-1}

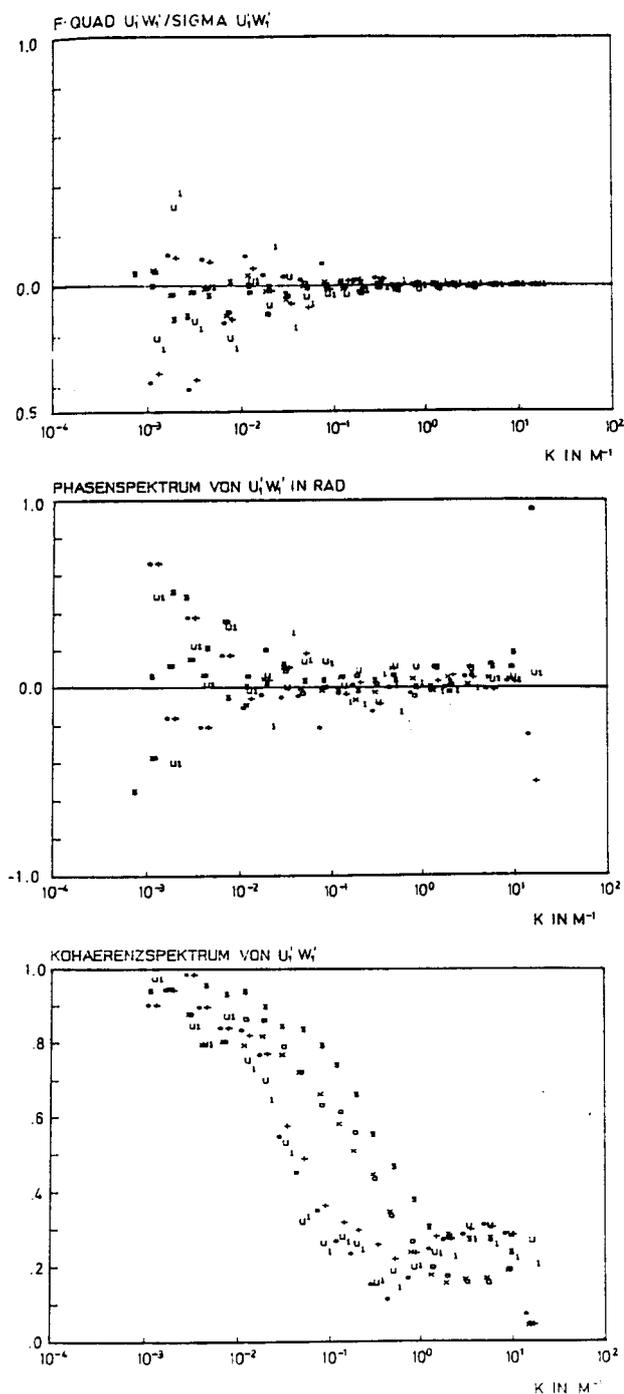


Abb. 20

Quadraturspektrum

von u' und w' ,

normiert mit

$\bar{\epsilon}_{uw}$ und multi-

pliziert mit der

Frequenz (oben),

Phasenspektrum

von u' und w'

(Mitte) und Ko-

härenzspektrum

nach (3.18)

von u' und w' für

0.8 m Höhe gegen

die Wellenzahl K_1 .

Abb. 20 Das Quadraturspektrum von u und w ist hier mit f multipliziert, um den hochfrequenten Anteil besser herauszuheben, und normiert. Es ist keine rechte Struktur erkennbar. Dagegen zeigen Phasen- und Kohärenzspektrum ein deutliches Maximum bei niederen Frequenzen $f \sim 10^{-3} - 10^{-2}\text{s}^{-1}$. Dies heißt, daß die abwärts gerichteten Impulstransporte $u'w' < 0$ in der Grenzschicht vor allem durch relativ langsame Fluktuationen verursacht werden.

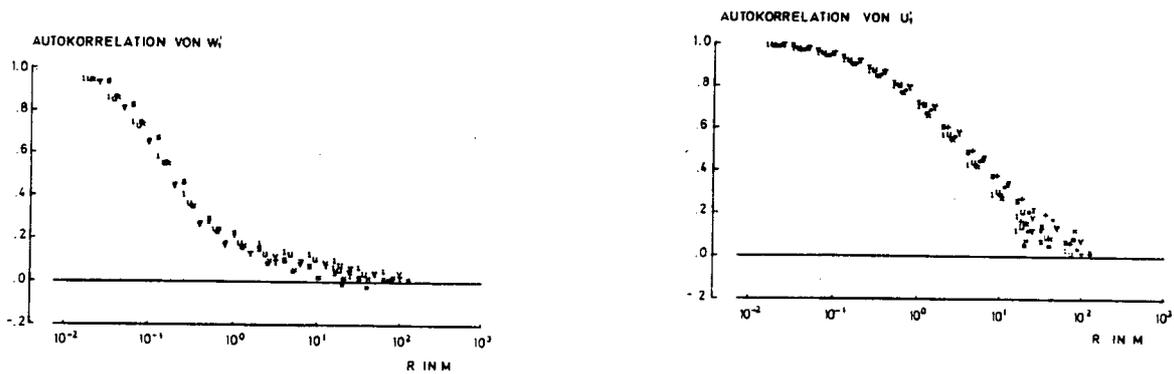


Abb. 21 Autokorrelationsfunktion von u' und w' in 0.8 m Höhe. Die Autokorrelationsfunktion fällt für u' (w') für Verschiebungen von 30 s(0.3 s) auf Null ab.

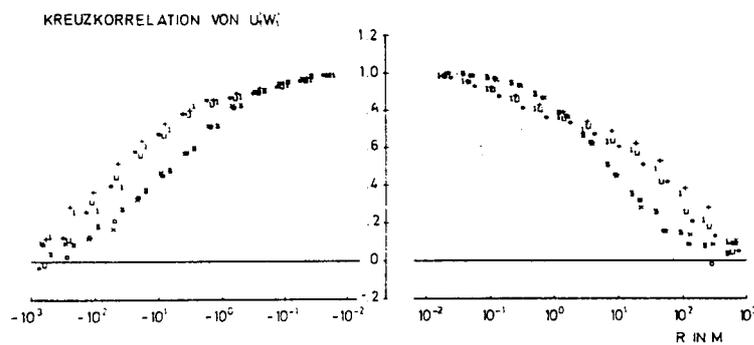


Abb. 22 Kreuzkorrelationsfunktion ρ_{uw} in 0.8 m Höhe. Im Vergleich zur Autokorrelation von \bar{u} sieht man einen langsameren Abfall mit τ . Die Funktion ist nahezu symmetrisch bez. $\tau = 0$.

4.4. Empirische Orthogonalfunktionen (EOF)

Die Fourieranalyse ist wenig geeignet, die typische räumliche Struktur von Datenfeldern herauszuarbeiten. Man habe zum Beispiel monatliche Mittel des Bodendrucks p_s im Pazifik über 30 oder 40 Jahre hin vorliegen, wobei auch die Rohdaten $p_{si}(t_n)$ an den Stationen mit Index i ($i=1\dots l$) vorliegen müssen (t_n Zeit; n Monatsindex). Man möchte die beträchtliche Information, die in diesen Karten enthalten ist, komprimieren, so daß man nicht alle Bodendruckfelder durchsehen muß, sondern Hauptmuster der Variabilität des Bodendrucks von Monat zu Monat erhält. Die empirischen Orthogonalfunktionen stellen unter günstigen Umständen solche Hauptmuster dar.

Kernpunkt des Verfahrens ist die Erstellung der Kovarianzmatrix

$$\mathbf{C} = \overline{p_{si}p_{sj}} = C_{ij} \quad (4.83)$$

, wobei der Querstrich eine Zeitmittlung bedeutet:

$$C_{ij} = \frac{1}{N} \sum_{n=1}^N p_{si}p_{sj} \quad (4.84)$$

und alle p_{si} Abweichungen sind:

$$\sum_{n=1}^N p_{si} = 0. \quad (4.85)$$

Der Index n bezeichnet die Zeitpunkte, an denen Messungen des Bodendrucks vorliegen. Er ist bei den p_{si} nicht angegeben. Die Kovarianzmatrix C ist reell, symmetrisch und positiv definit. Deshalb besitzt sie I reelle, positive Eigenwerte λ_k . Die zu den Eigenwerten gehörigen Eigenvektoren (in unserem Beispiel Bodendruckfelder) \mathbf{e}_{ki} (k fest; $i=1\dots I$)

$$C_{ij}\mathbf{e}_{kj} = \lambda_k\mathbf{e}_{ki} \quad (4.86)$$

bilden ein Orthonormalsystem

$$\mathbf{e}_{ki}\mathbf{e}_{ji} = \delta_k^j \quad (4.87)$$

, das vollständig ist (Summationskonvention). Das heißt, daß die ursprünglichen Datenfelder sich als Linearkombinationen der Eigenvektoren darstellen lassen:

$$p_{sj} = \sum_{i=1}^I a_i\mathbf{e}_{ij} \quad (4.88)$$

Die Muster $\mathbf{e}_k = (\mathbf{e}_{ki}, i = 1\dots I)$ nennt man EOF. Es gibt gerade so viele EOFs, wie es Datenpunkte gibt. Entwickelt man nun zu verschiedenen Zeiten die Bodendruckfelder nach den EOF, so ändern sich natürlich die EOFs nicht, wohl aber die Entwicklungskoeffizienten a_i . Die Bestimmung der Entwicklungskoeffizienten beruht auf der Orthogonalität der EOF:

$$a_i = p_{sj}\mathbf{e}_{ij} \quad (4.89)$$

Damit wäre noch wenig gewonnen, doch gilt, daß durch die Größe des Eigenwerts λ_k der Varianzanteil bestimmt wird, den der Eigenvektor mit Index k zur Erklärung (Repräsentation) der p_{si} beisteuert. Hat man also wenige EOF mit großen Eigenwerten und viele mit kleinen, so reichen diese wenigen Eigenvektoren schon aus, um eine gute Darstellung der Bodendruckfelder zu ermöglichen. Folglich muß man nur diese EOF betrachten, um eine gute Information über die hauptsächlichsten Schwankungsformen der Bodendruckfelder zu erhalten.

Die Anwendung der Technik sei an einem extremen Beispiel demonstriert. Man habe nur zwei Stationen. Die Druckwerte seien

$$p_{s1} = p_0\sin\omega t$$

$$p_{s2} = -p_0 \sin \omega t \quad (4.90)$$

(p_0 , ω konstant). Dann ist

$$\mathbf{C} = \frac{p_0^2}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

, wobei die Zeitmittlung durch Integration über eine Schwingungsperiode ersetzt wird.

Es folgt $\lambda_1 = 2$, $\lambda_2 = 0$ und

$$\begin{aligned} e_{11} &= 2^{-1/2} & e_{21} &= 2^{-1/2} \\ e_{12} &= -2^{-1/2} & e_{22} &= 2^{-1/2}. \end{aligned}$$

Der zweite Eigenvektor ist wegen $\lambda_2 = 0$ bedeutungslos und der erste Eigenvektor stellt die Bodendruckverteilungen dar, bei denen die Druckanomalie an der ersten Station gerade das entgegengesetzte Vorzeichen zu den an der zweiten hat. Dieses ist das Grundmuster aller Druckschwankungen, das man natürlich auch aus (4.90) direkt hätte ersehen können.

Ein interessanteres Beispiel ergibt sich bei der Untersuchung von Bodendruck und Meerestemperatur im Pazifik. Es liegen Daten für die Monatsmittelwerte an Gitterpunkten für 28 Jahre vor. Abb. 23a zeigt die Standardabweichungen, die beim Bodendruck ein deutliches Maximum aufweisen (Blockierende Hochs?). Davis (JPO, 1978) hat die EOFs getrennt für Bodendruck und Meerestemperatur berechnet. Dargestellt sind in Abb. 23 die ersten sechs EOFs der Anomalien der Monatsmittel des Bodendrucks im Pazifik ebenso die ersten acht EOFs der Anomalien der Monatsmittel der Meeresoberflächentemperatur. Ferner zeigt Abb. 23 den Anteil der Varianz der durch die meisten EOFs jeweils erfasst wird. Man sieht, daß sechs Druck-EOFs bereits über 80 % der Varianz erfassen. Man sieht an Abb. 23, daß die räumliche Struktur der EOFs mit zunehmender Nummer kleinteiliger wird. Die eigentlich interessante Aufgabe besteht hier darin, daß man nun eine Abhängigkeit zwischen EOFs von Druck und Temperatur sucht.

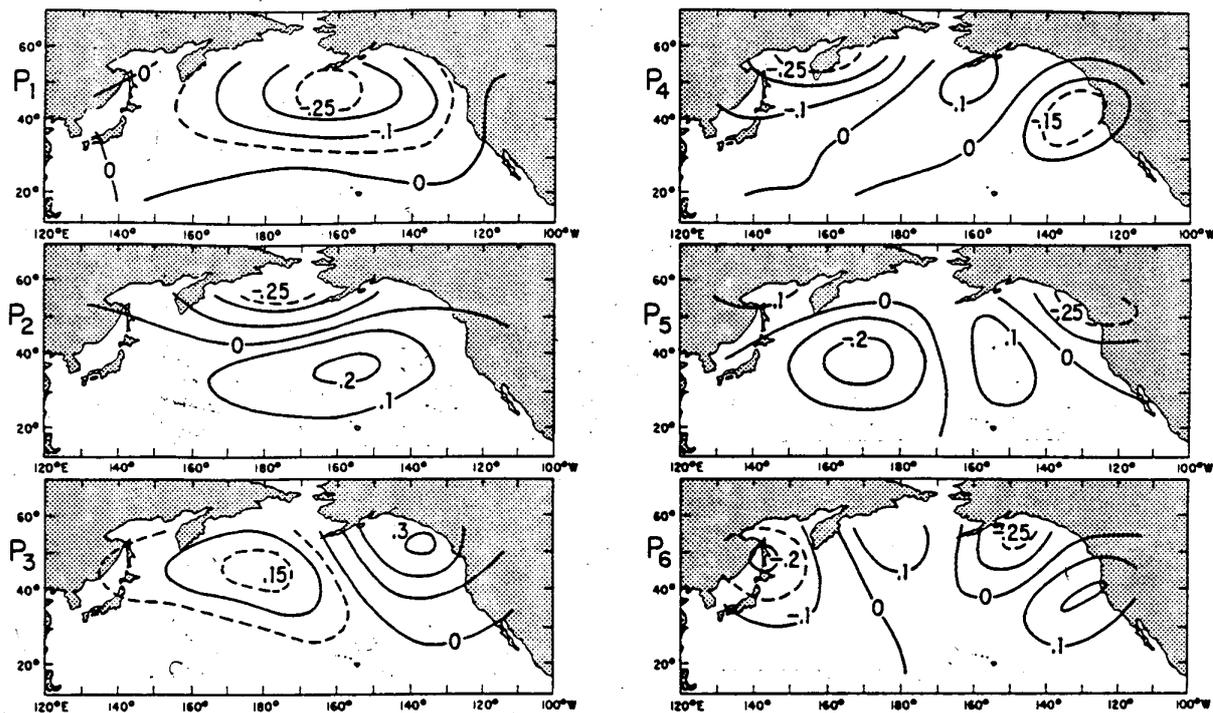
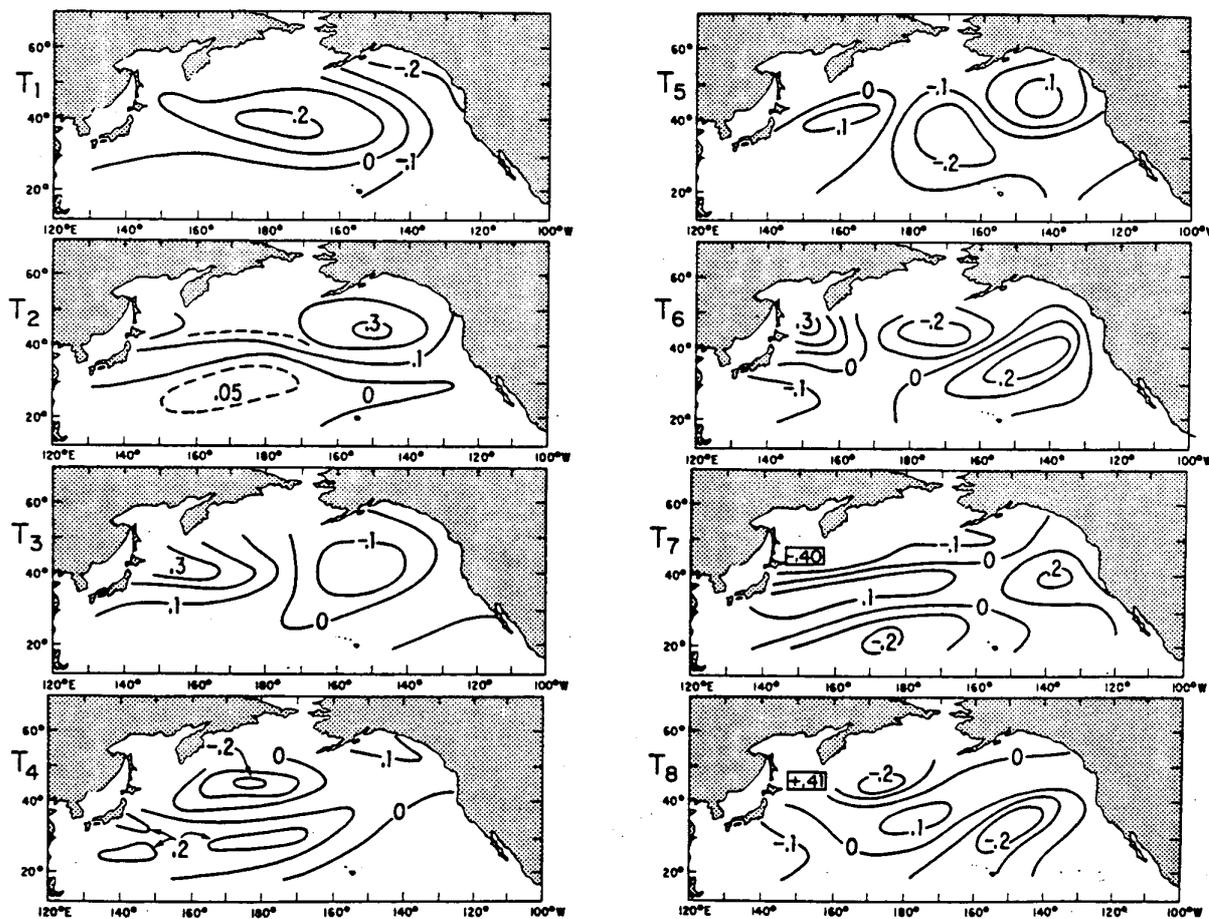


FIG. The six principal empirical orthogonal functions, P_1 - P_6 , describing SLP anomalies.



The eight principal empirical orthogonal functions, T_1 - T_8 , describing SST anomalies.

Wendet man das Verfahren auf die täglichen Niederschlagssummen an den Klimahauptstationen Südbayerns an, liefert dies für EOF1 (64%) eine ziemlich homogene Verteilung mit etwas angehobenen Amplituden in Alpennähe, für EOF2 (6.8%) einen deutlichen Nordsüdgegensatz. Dies heißt, daß der Niederschlag in Südbayern im Großteil der Fälle gleichmäßig auf Südbayern verteilt ist. (Abb. 24)

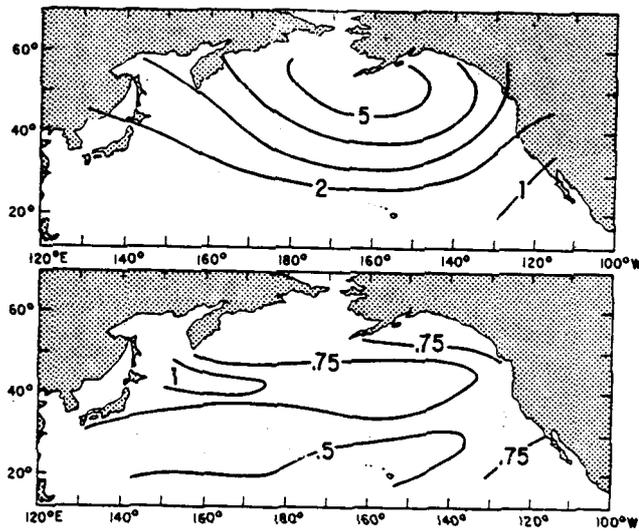


FIG Standard deviation of SLP anomalies (mb), upper, and SST anomalies ($^{\circ}\text{C}$), lower. Anomalies are departures from monthly normal values. The anomaly variances are averaged over all months of the 28-year record.

Abb. 23a

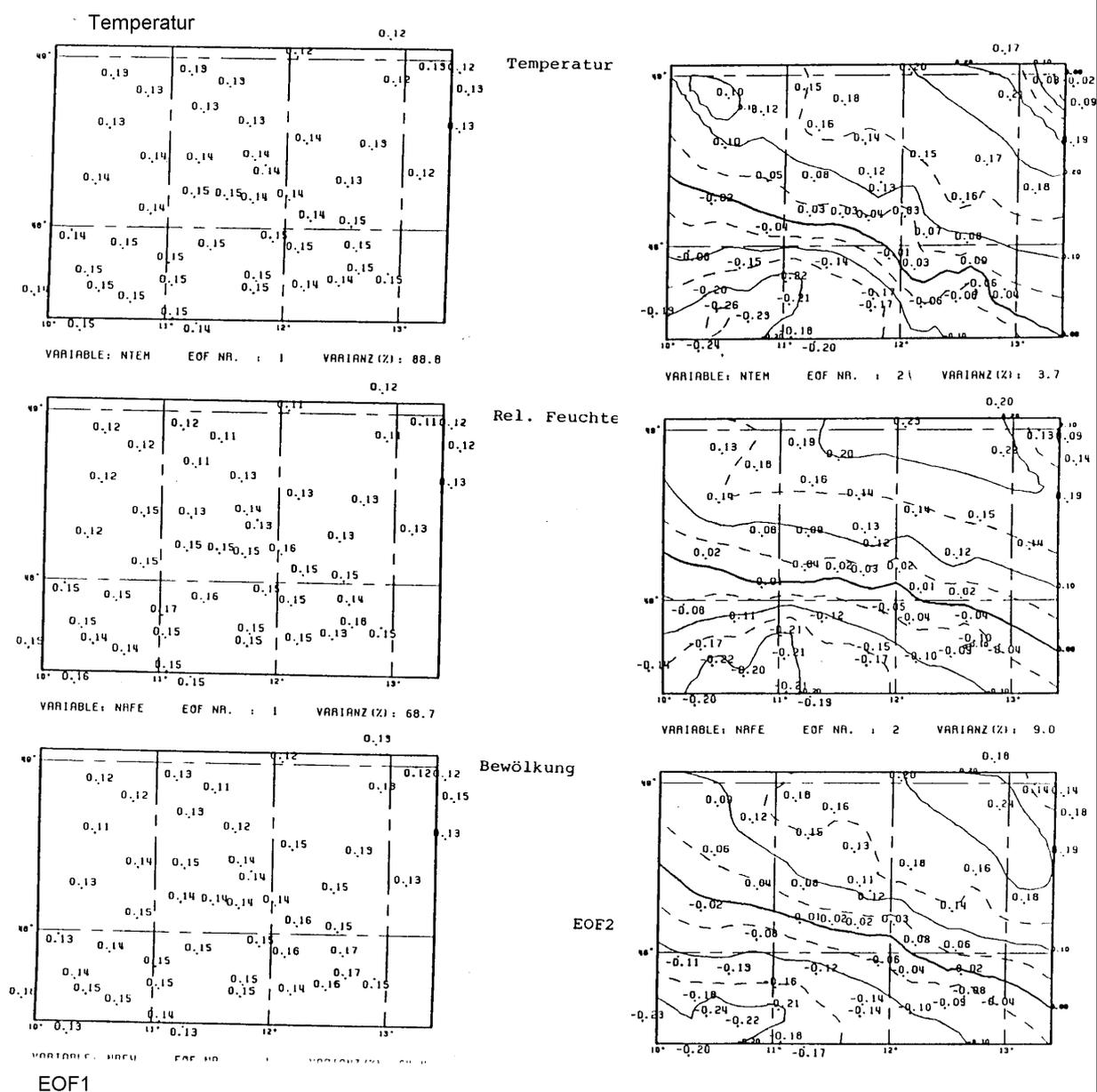


Abb. 24: EOF1 und EOF2 von Temperatur, relative Feuchte und Bewölkung in Oberbayern, wobei hier die Zahlenwerte angegeben sind, die den einzelnen Stationen zugeordnet sind.

4.6. Hauptschwingungsmuster (POP)

Die empirischen Orthogonalfunktionen dienen einem dazu (zumindest in der hier vorgestellten Form), vor allem ortsfeste Schwingungsmuster in Daten zu entdecken. Zur Aufdeckung wandernder Wellenmuster dient die POP-Analyse (Principal Oscillation Pattern). Genau wie bei der EOF-Analyse habe man einen Datensatz $p_i(t_k)$, wobei die Daten an Stationen mit Index i zu Zeiten $t_k = kDt$ vorliegen. Man postuliert nun, daß die Daten p_i durch einen autoregressiven Prozeß erster Ordnung \hat{p}_i approximiert werden können. Mit dem Vektor $\hat{\mathbf{p}} = (\hat{p}_i)$ verlangt man

$$\hat{\mathbf{p}}(t_{k+1}) = \mathbf{A}\hat{\mathbf{p}}(t_k) + \mathbf{n} \quad (4.91)$$

, wobei \mathbf{n} ein Rauschen ist und \mathbf{A} eine zu bestimmende Matrix mit konstanten Koeffizienten. Man verlangt, wie üblich, daß der Fehler

$$\sum_i (\mathbf{p}_i - \hat{\mathbf{p}}_i)_{t_{k+1}}^2 = \sum_i \left(\mathbf{p}_i(t_{k+1}) - \sum_r \mathbf{A}_{ir} \hat{\mathbf{p}}_r(t_k) \right)^2 \quad (4.92)$$

minimal werden soll. Das ergibt, bei Differentiation nach den \mathbf{A}_{ir} , die Bedingungen

$$\left(\mathbf{p}_i(t_{k+1}) - \sum_r \mathbf{A}_{ir} \hat{\mathbf{p}}_r(t_k) \right) \hat{\mathbf{p}}_j(t_k) = 0. \quad (4.93)$$

Diese Gleichung wird über alle Zeiten (soweit verfügbar) gemittelt unter der Annahme $\hat{\mathbf{p}}_j = \mathbf{p}_j$. Damit hat man

$$\mathbf{A}_{ir} = \overline{\mathbf{p}_i(t_{k+1})\mathbf{p}_j(t_k)} \cdot \overline{(\mathbf{p}_r(t_k)\mathbf{p}_j(t_k))}^{-1} \quad (4.94)$$

, so daß die Matrix \mathbf{A} in (4.91) bekannt ist. Die weitere Auswertung verläuft nun so, daß die Eigenwerte λ_n und die zugehörigen Eigenvektoren \mathbf{q}_n über

$$\mathbf{A}\mathbf{q}_n = \lambda_n\mathbf{q}_n \quad (4.95)$$

bestimmt werden. Die Eigenvektoren \mathbf{q}_n nennt man POPs. Ist dies geschafft, kann man jeden Datensatz $\mathbf{p}(t_k)$ entwickeln nach den Eigenvektoren.

$$\mathbf{p}(t_k) = \sum_n \gamma_n(t_k) \mathbf{q}_n \quad (4.96)$$

Eingesetzt in (4.91) unter Verwendung von (4.95) erhält man ohne Beachtung des Rauschens

$$\sum_n \gamma_n(t_{k+1}) \mathbf{q}_n = \sum_n \gamma_n(t_k) \lambda_n \mathbf{q}_n. \quad (4.97)$$

Für die Entwicklungskoeffizienten gilt folglich

$$\gamma_n(t_{k+1}) = \lambda_n \gamma_n(t_k). \quad (4.98)$$

Also

$$\gamma_n(t_k) = \lambda_n^k \gamma_n(0). \quad (4.99)$$

Wie bei den EOF ist es auch hier so, daß ein POP \mathbf{q}_n umso wichtiger ist (um so mehr Varianz erklärt), je größer $|\lambda|$.

Zunächst sei ein Beispiel gegeben, wo die POPs direkt berechnet werden können. Man habe zwei Meßpunkte, von denen die Datenreihen

$$\begin{aligned} p_1 &= \sin \omega t \\ p_2 &= \cos \omega t \end{aligned} \quad (4.100)$$

vorliegen. Die POP-Gleichung (4.91) mit $t = kDt$ lautet

$$\begin{aligned} p_1(t_{k+1}) &= A_{11}p_1(t_k) + A_{12}p_2(t_k) \\ p_2(t_{k+1}) &= A_{21}p_1(t_k) + A_{22}p_2(t_k). \end{aligned} \quad (4.101)$$

Nun ist

$$\begin{aligned} p_1(t_{k+1}) &= \sin \omega(k+1)Dt \\ &= \sin \omega t \cos \omega Dt + \sin \omega Dt \cos \omega t \\ p_2(t_{k+1}) &= \cos \omega(k+1)Dt = \\ &= \cos \omega t \cos \omega Dt - \sin \omega t \sin \omega Dt. \end{aligned}$$

Multiplikation von (4.101) mit je $p_1(t_k)$, und $p_2(t_k)$ und Zeitintegration über eine Periode liefert

$$\begin{aligned} A_{11} &= \cos \omega Dt = A_{22} \\ A_{12} &= \sin \omega Dt = -A_{21}. \end{aligned}$$

Die Eigenwertgleichung lautet dann

$$\lambda^2 - 2\lambda \cos \omega Dt + 1 = 0.$$

Folgt

$$\lambda_{1,2} = \cos \omega Dt \pm i \sin \omega Dt = \exp(\pm i \omega Dt).$$

Die zugehörigen Eigenvektoren (POPs) sind

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Da

$$\operatorname{Re}[\mathbf{q}_1 \exp(i \omega t)] = \begin{pmatrix} \sin \omega t \\ \cos \omega t \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

folgt, daß die Beobachtungen in diesem Fall ausnahmslos durch POP_1 wiedergegeben werden. Die Nützlichkeit der POP-Analyse kann natürlich anhand eines solchen Beispiels nicht demonstriert werden, wohl aber ihre technische Durchführung.

In meteorologischen Anwendungen der POP-Technik ist der Aufwand gemeinhin wesentlich höher. Als Beispiel sei hier eine POP-Analyse des Geopotentials beschrieben. Die Geopotentialwerte liegen in 7 Druckflächen vor. Dann wird für jedes Niveau angesetzt.

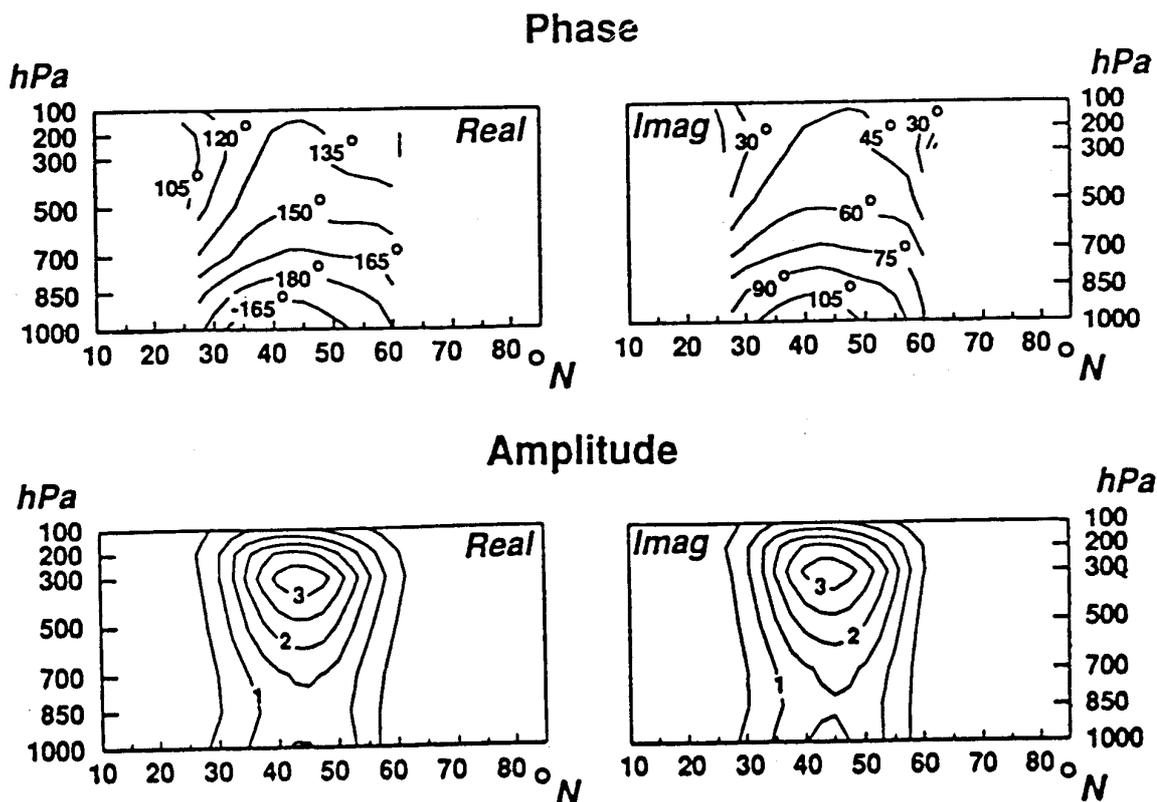
$$\phi(\lambda, \varphi, p) = \sum C_k(\varphi, p) \cos(k\lambda) + S_k(\varphi, p) \sin(k\lambda)$$

Die Daten liegen in 31 Breiten φ_j und eben 7 Druckniveaus vor. Für jede zonale Wellenzahl k hat man somit einen Datenvektor

$$\mathbf{p} = \begin{Bmatrix} C_K(\varphi_i, \varphi_j) \\ S_K(\varphi_i, \varphi_j) \\ \vdots \end{Bmatrix}$$

der $31 \times 7 \times 2 = 434$ Werte enthält. Für einen so großen Datenvektor hätte man große Probleme, statistisch signifikante POPs zu erhalten. In praxi muß noch eine Komprimierung der Daten vorgenommen werden, die aber hier übergangen sei. Aus Abb. 25 sind einige Charakteristika des wichtigsten POPs zu ersehen.

Das Hauptschwingungsmuster wandert bei zonaler Wellenzahl 8 mit einer Phasengeschwindigkeit von 11ms^{-1} in mittleren Breiten nach Osten. Dabei fällt die Amplitude mit einer Rate von 8.6 Tagen für eine Abnahme um den Faktor e . Das Phasendiagramm zeigt, daß die Welle nach Westen mit der Höhe sich zurückneigt. Die Amplitude ist in der oberen Troposphäre am größten. Das Wellenmuster variiert nur geringfügig mit der geographischen Länge. All dies legt die Vermutung nahe, man habe hier das Gegenbild eines linear instabilen Modes wie er etwa im Eady-Modell auftritt. Allerdings gibt die lineare Theorie an, daß die Moden wachsen, doch die POP-Moden sind grundsätzlich gedämpft.



The most significant POP of a POP analysis of twice-daily geopotential height in the Northern Hemisphere for the three winters (DJF) 84/85 to 86/87 and for zonal wave number 8. The oscillation period is 4 days, and the e -folding time is 8.6 days. The real part P^1 (left) and imaginary part P^2 (right) of the POP P are represented by phase patterns Ξ^1 and Ξ^2 (top) and amplitude patterns A^1 and A^2 (bottom). The evolution of the pattern is defined by the sequence ... \rightarrow Imag \rightarrow Real \rightarrow -Imag \rightarrow -Real \rightarrow Imag \rightarrow ...

The absolute values of the amplitude patterns are arbitrary.

Abb. 25

5. Stochastische Vorhersage

Bei Modellen der statistischen Vorhersage kann man zumindest zwei Grundtypen unterscheiden. Die Situation sei am Beispiel der Wettervorhersage erläutert. Diese wird ja gemeinhin deterministisch betrieben in dem Sinne, daß ein wohldefinierter Anfangszustand zu genau einer Vorhersage führt. Berücksichtigt man jedoch, daß der Anfangszustand nur ungenau bekannt ist und daß wohl auch einige Modellparameter stochastische Größen sind, so wird man, wie in Kapitel 3 ausgeführt, im Idealfall die Fokker-Planck-Gleichung zu lösen haben, in der Praxis müssen Näherungen reichen, zu denen auch Monte-Carlo-Vorhersagen gehören. Dem gegenüber verzichtet die zweite Methode weitgehend auf die physikalischen Kenntnisse, die in die Ableitung der Gleichungen der Wettervorhersage eingegangen sind und zieht ausschließlich Daten zu Rate,

die in der Vergangenheit gewonnen wurden. Dies wird gemeinhin im Rahmen eines Regressionsmodells durchgeführt. Analogverfahren gehören ebenfalls hierher. Im Abschnitt 1.5 (Übergangswahrscheinlichkeiten) wurden bereits statistische Vorhersagemöglichkeiten besprochen.

5.1. Regressionsmodelle

Es sei y die Größe, die vorherzusagen ist, also der Prädiktand, und die x_k seien die gewählten Prädiktoren, also die Größen, von denen man annimmt, daß sie die für eine Vorhersage von y nötige Information enthalten. Dabei seien nur Abweichungen betrachtet, so daß $E(y) = E(x_i) = 0$. Es ist nicht nötig, daß alle x_k zum selben Zeitpunkt vorliegen. Bei einem Regressionsmodell postuliert man einen linearen Zusammenhang

$$y = \phi_1 x_1 + \dots + \phi_K x_K + Z \quad (5.1)$$

, wobei die ϕ_i die Modellkonstanten sind und Z ein Restterm. Kommt es nun zur Vorhersage, so liegen die x_k vor und man schätzt y gemäß

$$\hat{y} = \sum_{k=1}^K \phi_k x_k . \quad (5.2)$$

Die Koeffizienten ϕ_i werden nach Gauß so bestimmt, daß das Fehlerquadrat $\overline{(y - \hat{y})^2}$ minimal wird, wobei der Querstrich eine Mitteilung über sämtliche Datensätze y , \hat{y} der Vergangenheit bedeutet. Das Vorhersagemodell (5.2) wird umso besser sein, je mehr Paare vorliegen. Wir wollen im folgenden die Annahme machen, daß unendlich viele Realisierungen vorliegen. Dies vereinfacht die Rechnung, ferner wird das Resultat davon nicht wirklich berührt.

Minimierung von

$$F = \overline{(y - \sum \phi_k x_k)^2} \quad (5.3)$$

führt auf

$$\frac{\partial F}{\partial \phi_i} = \overline{(y - \sum \phi_k x_k) x_i} = 0 . \quad (5.4)$$

Nun ist $\overline{y x_i}$ ($i=1 \dots K$) der Kovarianzvektor, der den Prädiktanden mit den Prädiktoren verknüpft und $\overline{x_k x_i}$ ($k, i = 1 \dots K$) ist die Kovarianzmatrix der Prädiktoren. Sei nun $\overline{(x_k x_i)}^{-1}$ das Element am Ort (k, i) der zu $\overline{x_k x_i}$ inversen Matrix. Dann hat man

$$\phi_k = \sum_{i=1}^K \overline{y x_i} \overline{(x_k x_i)}^{-1} \quad (5.5)$$

als Bestimmungsgleichung für die Modellkoeffizienten. Es läßt sich für endliche Datenkollektive, wo also — die Mitteilung über die vorhandenen Daten darstellt, zeigen, daß auch dann (5.5) die beste Lösung liefert. Die gegebene Ableitung ist völlig analog zu der eines POP-Modells. Letzteres ist insofern spezieller als alle x_k zum selben Zeitpunkt definiert sind. Es ist allgemeiner weil nicht nur eine Variable vorausgesagt wird.

Der autoregressive Prozeß erster Stufe (3.45)

$$T_{n+1} = aT_n + Z_n \quad (5.6)$$

(n Zeitindex) kann mit der Zuordnung $y = T_{n+1}$, $\phi_1 = a$, $x_1 = T_n$, $k = 1$ ebenfalls als Regressionsmodell gedeutet werden, wie natürlich auch ein autoregressiver Prozeß zweiter Stufe

$$T_{n+1} = aT_n + bT_{n-1} + Z_n \quad (5.7)$$

Gemäß (5.5) schätzt man für (5.6)

$$a = \overline{T_{n+1}T_n} / \overline{T_n^2} \quad (5.8)$$

, wobei die Kovarianzen in (5.8) aus Beobachtungen zu erhalten sind. Man erhält etwa für das Deckschichtmodell $a \sim (2-4 \text{ Monate})^{-1}$.

Damit ist die Vorhersagegleichung

$$T_{n+1} = aT_n$$

optimal angepaßt und kann für Vorhersagen der Meerestemperatur verwendet werden. Als relativ kompliziertes Anwendungsbeispiel sei ein Satz von Regressionsmodellen von R. Davis (JPO, 1976) präsentiert, mit deren Hilfe Monatsmittelwerte der Meeresoberflächentemperaturen im Pazifik aus dem Bodendruck im vorhergehenden Monat bestimmt werden sollen oder umgekehrt der Bodendruck aus den Meerestemperaturen (SST). Dazu werden zunächst die EOFs beider Felder bestimmt (s. Kapitel 4.5). Der Bodendruck kann entwickelt werden

$$p_{sj} = \sum_{i=1}^I a_i e_{ij} \quad (5.9)$$

(siehe (4.88) und ebenso die Meerestemperatur

$$T_j = \sum_{i=1}^I \theta_i s_{ij} \quad (5.10)$$

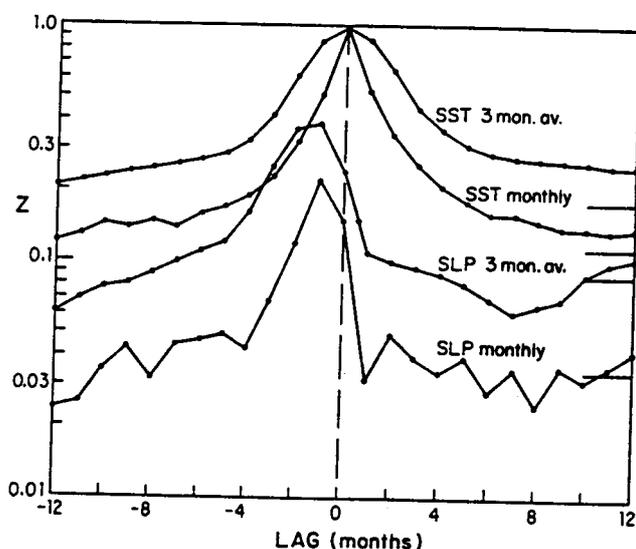
, wobei $e_{ij}(s_{ij})$ der Wert der EOF mit Index i am Gitterpunkt j ist. Ein Vorhersagemodell verknüpft nun die EOF-Koeffizienten a_i in einem Monat (n) mit einem EOF-Koeffizienten θ_i im folgenden Monat ($n+1$)

$$\theta_i(n+1) = \sum_{k=1}^I \phi_{ik} a_k(n) \quad (5.11)$$

Dabei wird $I = 10$ benutzt. Nun werden die ϕ_{ik} für $L \leq 10$ gemäß (5.5) bestimmt. Somit hat man ein Vorhersagemodell für die Meerestemperatur. Dies kann für größere Vorhersageschnitte ($t+2$, $t+3$ etc.) und auch für negative Schritte ($t-1$), also vorhergehende Monate, wiederholt werden. Ferner kann man versuchen, die Meerestemperatur im folgenden Monat aus der Meerestemperatur selber vorherzusagen, also anzusetzen

$$\theta_i(n+1) = \sum \phi_{ik} \theta_k(n). \quad (5.12)$$

Resultate sind in Abb. 27 dargestellt, worin Z ein Gütemaß in dem Sinn ist, daß $Z=1$ für ein perfektes Modell gilt. Man sieht zunächst (SST monthly), daß eine Vorhersage der Meerestemperatur in künftigen Monaten aufgrund der Meerestemperatur im Vorhersagemonat selbst ganz gut möglich ist. Das darf nicht überraschen. Gleichung (3.11) sagt dies ja gerade.



Skill of estimating SST and SLP in month $t+\text{lag}$ using as data ten SST modes from month t . The skill index Z , described in the text, measures the fraction of the entire spatial field correctly estimated. The top two curves refer to estimating one and three month averages of SST; the lower two curves are SLP estimation skill. The horizontal lines at the right are the artificial skill expected when there is no true skill.

Abb. 26

Interessant ist dagegen das Resultat, daß Bodendruckanomalien in vorhergehenden Monaten ($\tau < 0$) gut mit den Anomalien der Meerestemperatur zur Zeit t verknüpft sind.

Dies legt den Schluß nahe, daß Anomalien der Meerestemperatur durch Anomalien der atmosphärischen Zirkulation in den vorhergehenden Monaten erzeugt werden.

Die Untersuchungen von Davis haben nicht eigentlich auf die Vorhersage künftiger Anomalien der Meerestemperatur abgezielt, auch wenn (5.11) durchaus für diesen Zweck verwandt werden könnte, sondern auf die Aufdeckung kausaler Zusammenhänge. So sei nun als weiteres Beispiel ein Regressionsmodell vorgestellt, das zur Niederschlagsvorhersage eingesetzt wird. Dabei ist y in (5.1) die Wahrscheinlichkeit, daß es etwa in den nächsten 12 Stunden in Melbourne regnen wird. Die Prädiktoren sind x_1 = Bodendruck, x_2 = Taupunktdifferenz, x_3 = Komponente des Zonalwinds jeweils in Melbourne. Man beachte, daß alle x_i Abweichungen von Mittelwerten sind. Da in Melbourne natürlich eine gewisse Chance für Niederschlag auch dann besteht, wenn alle $x_i = 0$, muß (5.1) um einen klimatologischen Term α erweitert werden:

$$y = \alpha + \phi_1 x_1 + \phi_2 x_2 + \phi_3 x_3 \quad (5.13)$$

Die Prädiktoren werden für die Stunde vom Vorhersagebeginn ausgewertet. Als Gütemaß dient

$$B = K^{-1} \sum_{k=1}^K (\delta_k - y_k)^2 \quad (5.14)$$

, wobei K die Zahl der Vorhersagetage ist, $\delta_k = 0$ (trocken), $\delta_k = 1$ (Niederschlag). Die statistische Vorhersage erzielt $B = 0.14$, ein numerisches Wettervorhersagemodell $B = 0.18$, die Meteorologen vom Dienst $B = 0.17$ (mehr Details in Friedrich und Leslie, MWR 1987).

Man kann das Modell (5.13) verbessern, indem man die Aussage eines verfügbaren numerischen Vorhersagemodells Y ($Y = 0$ kein Regen, $Y = 1$ Regen) mit (5.13) zu einer Wahrscheinlichkeitsvorhersage

$$Y' = ay + (1-a)Y \quad (5.15)$$

kombiniert, wobei a wieder über Minimierung des Fehlerquadrats bestimmt wird. Man findet dann $B = 0.11$.

5.2. Analogvorhersage

Die Analogvorhersage beruht wie die Regressionsmethode ausschließlich auf Nutzung von in der Vergangenheit angesammelten Daten. Doch wird hier keine Vorhersagegleichung eingesetzt, sondern man sucht in seinem Archiv nach Situationen, die im Vorhersagemoment ähnlich sind. Die aus der Vergangenheit bekannte Entwicklung dieser Lagen wird dann als Vorhersage benutzt. Für die Wettervorhersage bedeutet dies, daß

man für die gerade beobachtete Lage, in der Vergangenheit eine Analoglage sucht, die der jetzt beobachteten möglichst ähnlich ist. Die Vorhersage besteht dann darin, daß man die Entwicklung der Analoglage zum nächsten Tag etwa als Vorhersage für morgen benutzt. Dieses Verfahren hat sich bei der Wettervorhersage nicht bewährt, da es offensichtlich hinreichend gute Analoglagen in den Archiven nicht gibt. Doch sieht es so aus, als ob die Analogvorhersage für die Langfristvorhersage derzeit das einzige Verfahren ist, das überhaupt Vorhersagegüte aufweist.

In einem fast trivialen Beispiel werde ein System betrachtet, das nur zwei Variable (x,y) hat. Man habe ferner ein perfektes Archiv, das einem für $t < \tau$ sagt, daß $x = \sin\omega t$, $y = \cos\omega t$. Man befindet sich nun im Zeitpunkt $t = t_0 > \tau$ und stellt fest, daß $x = x_0 = \sin\omega t_0$, $y = y_0 = \cos\omega t_0$. Analoge werden gesucht, indem man nach Minima des Abstands

$$d^2 = (x-x_0)^2 + (y-y_0)^2$$

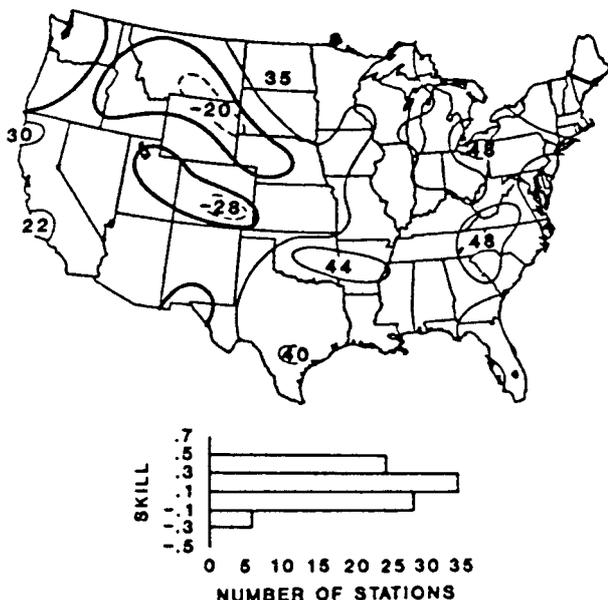
im Archiv Ausschau hält. All diese Minima liegen natürlich bei $t_i = t_0 - i2\pi/\omega$. Man hat noch nachzusehen, welche Lagen zu den Zeiten $t_i + Dt$ eingetreten waren, wobei Dt das Vorhersageintervall ist:

$$x = \sin(\omega(t_i + Dt)) = \sin\omega(t_0 + Dt), \quad y = \cos\omega(t_0 + Dt).$$

Diesen Wert verwendet man als Vorhersage.

In der Praxis der Langfristvorhersage ist das wesentlich schwieriger, schon weil man nicht von vorneherein weiß, für welche Variablen man die Vorhersage machen soll. Hier ist zumindest eine Kompression der Daten durch eine EOF-Analyse vonnöten. Details seien hier übergangen (siehe z.B. Livezey and Barnston, JGR 93, 10955ff.) und nur eine Erfolgsmeldung gezeigt. In Abb. 27 ist das Gütemaß $S = (C-E)/(T-E)$ für 3 Monatsvorhersagen eines Datenvektors für die USA angegeben, der Information insbesondere über die 700-mb Höhe enthält. Dabei ist C die Zahl der korrekten Vorhersagen (innerhalb gewisser Grenzen), E die Zahl der Vorhersagen, die man zufällig erwarten müßte und T ist die Gesamtzahl der Fälle. Die Vorhersagen werden je im November für den Winter gemacht. Man sieht erstaunliche Erfolge in den östlichen USA, doch die mittlere Güte liegt nur bei ca. 0.12. Ob derlei Vorhersagen von Nutzen sind, wird sich wohl erst zeigen müssen.

WINTER SELECTED MODEL LOCAL SKILL



Spatial (map) and statistical (histogram) distributions of local skill (S) \times 100 for the winter model ($n = 9$, $N = 34$, $L = 10$). Contour interval is 20. Areas with values in excess of 25 are significant at less than the 5% level.

Abb. 27

5.3. Statistisch-dynamische Vorhersage

Im Rahmen eines statistisch-dynamischen Modells versucht man nicht die gesamte Wahrscheinlichkeitsverteilung vorherzusagen, wie es ja die Liouville-Gleichung erlaubt, sondern nur einige Momente der Variablen, zumeist Mittelwert und Varianz. Im Prinzip lassen sich die entsprechenden Gleichungen aus der Liouville-Gleichung ableiten. So erhält man aus (3.17) sofort durch Multiplikation mit T und Integration

$$\frac{\partial}{\partial t} \mu - \int_{-\infty}^{+\infty} \frac{T \partial}{\partial T} (\lambda T f) dT = 0. \quad (5.16)$$

Es folgt die Vorhersagegleichung für den Mittelwert μ

$$\frac{\partial \mu}{\partial t} + \lambda \mu = 0. \quad (5.17)$$

Dieses Beispiel ist, da (3.11) linear ist, nahezu trivial. Bei nichtlinearen Gleichungen, also bei Wettervorhersage, geht man im Prinzip genau so vor, doch geht dann die Trivialität verloren. Sei

$$\frac{dx_i}{dt} + \sum a_{ij} x_j + \sum b_{ijk} x_j x_k = 0 \quad (5.18)$$

der Satz der Modellgleichungen eines Wettervorhersagemodells, wobei die x_i die Werte der Variablen an Gitterpunkten bedeuten mögen, dann lautet die Liouvillegleichung

$$\frac{\partial f}{\partial t} - \sum_i \frac{\partial}{\partial x_i} \left(f \sum_j a_{ij} x_j + f \sum_{jk} b_{ijk} x_j x_k \right) = 0. \quad (5.19)$$

Durch Multiplikation mit x_i und Integration über den Phasenraum ϕ gewinnt man mit

$$\begin{aligned} \mu_i &= \int_{\phi} f x_i dx \\ \sigma_{ij}^2 &= \int_{\phi} (x_i - \mu_i) (x_j - \mu_j) f dx \end{aligned} \quad (5.20)$$

$$\frac{\partial}{\partial t} \mu_i + \sum_j a_{ij} \mu_j + \sum_{jk} b_{ijk} (\mu_j \mu_k + \sigma_{jk}) = 0$$

, wobei immer unterstellt wird, daß f für große Werte von $|x_i|$ hinreichend rasch abklingt, damit etwa $x_i x_j f \rightarrow 0$ für $|x| \rightarrow \infty$. Das Modell (5.18) ist deterministisch. Man schreibt zur Zeit $t = 0$ Anfangswerte x_{i0} vor und erhält eine eindeutige Vorhersage. Dagegen muß man in (5.20) auch zur Zeit $t=0$ die Kovarianzen σ_{jk} kennen, die sich ja gerade aus Meßfehlern und vor allem aus Unsicherheiten in der Festlegung des Anfangszustandes ableiten. Natürlich werden (5.20) und (5.18) identisch, wenn $\sigma_{jk}=0$, wenn es also keine Unsicherheiten gibt. Insgesamt sollte (5.20) die besseren Vorhersagen liefern. Doch ist zu beachten, daß die Kovarianzen σ_{jk} ja ihrerseits von der Zeit abhängen, daß man also Prognosegleichungen für alle σ_{jk} braucht. Diese sind zwar relativ leicht aus (5.19) abzuleiten, doch erhöht sich damit die Zahl der Modellgleichungen enorm. Ferner hat man das Problem, daß in den Vorhersagegleichungen für σ_{jk} Tripelkovarianzen vom Typ

$$\int_{\phi} f (x_i - \mu_i) (x_j - \mu_j) (x_k - \mu_k) dx$$

auftauchen, für die eigentlich auch Vorhersagegleichungen abgeleitet werden müßten. Man hilft sich in praktischen Fällen mit der Annahme, daß alle Variablen normal verteilt sind. Dann verschwinden die Tripelkorrelationen. Das kann jedoch leicht dazu führen, daß negative Werte für die σ_{jk} vorhergesagt werden. Insgesamt hat sich diese Technik nicht durchgesetzt.

5.4. Monte-Carlo-Vorhersage

Das Vorhersagemodell sei vom Typ (5.18): zur Zeit $t = 0$ sei ein Satz von Anfangswerten x_{i0} verfügbar. Man weiß, daß diese Anfangswerte mit Fehlern behaftet sind und daß man anstelle von festgelegten Anfangswerten eine Wahrscheinlichkeitsdichte $f(x_1, x_2, \dots, x_k)$ vorgeben müßte, deren Entwicklung anhand der zu (5.18) gehörigen Liouville-

gleichung (5.19) zu bestimmen ist. Diesem Wissen kann man, wie oben erwähnt, nicht durch die Tat entsprechen, einmal, weil man die Fehler nicht kennt, zum andern, weil die Integration der Liouville-Gleichung bei weitem zu aufwendig ist. Einen gewissen Ausweg bietet die Monte-Carlo-Vorhersage, bei der anstelle einer einzigen Vorhersage zum Anfangszustand x_{i_0} ein Ensemble von Vorhersagen zu Anfangszuständen $x_{i_0} + \delta_j x_{i_0}$ gemacht wird, wo δ eine kleine Störung bedeutet, und j der Index des Ensemblemitglieds ist. Wenn diese Störungen geschickt gewählt sind und man ferner ein hinreichend großes Ensemble hat, so besteht Hoffnung, daß die Entwicklung des Ensembles in der Zeit wenigstens näherungsweise die Entwicklung von f beschreibt und Information darüber liefert, ob die Anfangslage x_{i_0} bei kleinen Störungen stark abweichende Vorhersagen liefert. Tut sie das, so weiß man, daß die vom Anfangszustand x_{i_0} startende Vorhersage, eben die 'normale' Vorhersage, nur mit Vorsicht zu benutzen ist. Bleiben dagegen alle Mitglieder des Ensembles im Vorhersagezeitraum nahe beisammen, so ist dies ein Indiz dafür, daß die normale Vorhersage zuverlässig sein sollte. Das Problem besteht dann in der geschickten Wahl der Störungen. Dies sei an einem besonders einfachen Beispiel erläutert. Die Vorhersagegleichungen seien

$$\frac{dx_1}{dt} + \alpha x_2 = 0 \quad (5.21)$$

$$\frac{dx_2}{dt} + \alpha x_1 = 0$$

($\alpha > 0$) und die Vorhersage für den Anfangszustand x_{10}, x_{20} lautet

$$x_1 = (x_{10} - x_{20}) \exp(\alpha t)/2 + (x_{10} + x_{20}) \exp(-\alpha t)/2 \quad (5.22)$$

$$x_2 = -(x_{10} - x_{20}) \exp(\alpha t)/2 + (x_{10} + x_{20}) \exp(-\alpha t)/2.$$

Die zu (5.21) gehörige Liouvillegleichung lautet

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x_1}(\alpha x_2 f) + \frac{\partial}{\partial x_2}(\alpha x_1 f) \quad (5.23)$$

und schreibt sich mit $u_1 = -\alpha x_2, u_2 = -\alpha x_1$:

$$\frac{\partial f}{\partial t} + u_1 \frac{\partial f}{\partial x_1} + u_2 \frac{\partial f}{\partial x_2} = 0.$$

Somit gilt im Phasenraum

$$\frac{df}{dt} = 0. \quad (5.24)$$

Sei nun $x_{10} = x_{20} = 0$ und die "korrekte" aber "unbekannte" Anfangsverteilung gegeben durch

$$f = \pi^{-1}; x_1^2 + x_2^2 \leq 1$$

$$f = 0; \text{sonst} \tag{5.25}$$

Dann läßt sich die Liouville-Gleichung wegen (5.24) leicht integrieren. Zunächst erhält man aus (5.22)

$$x_1 \sinh \alpha t + x_2 \cosh \alpha t = x_{20}$$

$$x_1 \cosh \alpha t + x_2 \sinh \alpha t = x_{10} \tag{5.26}$$

als Zusammenhang zwischen Startpunkt und späterem Punkt einer Trajektorie.

Beschreiben nun die Ausgangspunkte (x_{10}, x_{20}) den Kreis $x_{10} = \cos \varphi$, $x_{20} = \sin \varphi$ ($0 \leq \varphi \leq 2\pi$), so wird dieser Kreis abgebildet auf die Ellipse

$$(x_1^2 + x_2^2) (\sinh^2 \alpha t + \cosh^2 \alpha t) + 4x_1 x_2 \sinh \alpha t \cosh \alpha t = 1, \tag{5.27}$$

deren große Halbachse die Länge $\exp(\alpha t)$ hat und in die Richtung $x_1 = -x_2$ weist, während die dazu orthogonale kleine Halbachse die Länge $\exp(-\alpha t)$ hat (s. Abb. 27). Dies heißt, daß Abweichungen vom Nullpunkt rasch (Zeiten $> \alpha^{-1}$) wachsen, wenn sie in der Nähe der großen Halbachse liegen, dagegen abnehmen, wenn sie in der Nähe der kleinen Halbachse liegen. Die normale Vorhersage, die vom Ursprung ausgeht, liefert $x_1 = x_2 = 0$ für alle Zeit und ist nicht sonderlich verläßlich.

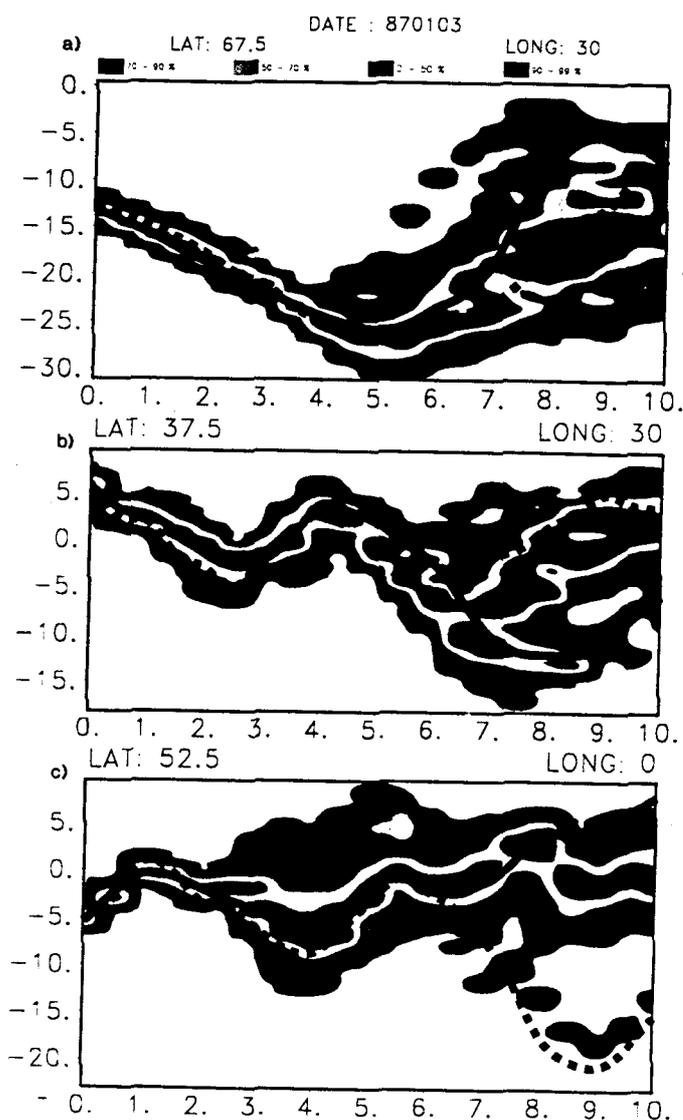
In einer Monte-Carlo-Vorhersage hat man nun $\delta_j x_{i0}$ zu wählen. Ungeschickt ist die Wahl $\delta_j x_{10} \sim \delta_j \cdot x_{20}$, denn eine solche Wahl liefert eine Konvergenz der Lösung gegen $x_1 = x_2 = 0$, also zur Normalvorhersage hin. Dagegen liefert $\delta_j x_{10} \sim -\delta_j x_{20}$ rasche Divergenz von der Normalvorhersage und zeigt an, daß der Normalvorhersage wenig Vertrauen entgegenzubringen ist. In Abb. 27 sind die Trajektorien für einige Mitglieder eines Monte-Carlo-Ensembles eingetragen. Für diejenigen Fälle, wo $\delta x_{10} \sim \delta x_{20}$ läuft die Vorhersage zur Normalvorhersage hin, dagegen hat man für andere durchaus die erwünschte Zunahme der Abweichung. Man könnte also anhand dieses kleinen Ensembles bereits den richtigen Schluß ziehen, daß die Normalvorhersage nicht sonderlich verläßlich ist.

Das Problem läßt sich systematisch angehen, indem bezüglich $x_{10} = x_{20} = 0$ eine Stabilitätsanalyse von (5.21) durchführt. Man findet, natürlich, die Eigenwerte $\pm \alpha$ und die zugehörigen Eigenvektoren $\begin{pmatrix} +1, \mp 1 \\ +, \mp 1 \end{pmatrix}$. Eine geschickte Wahl der $\delta_j x_{i0}$ wird die Störungen so anordnen, daß sie den ersten, instabilen Eigenvektor repräsentieren, also $\delta x_{10} = -\delta x_{20}$.

In aktuellen Fällen der numerischen Vorhersage ist die Durchführung derartiger Stabilitätsrechnungen aufwendig, doch machbar.

Abb. (28; ECMWF) zeigt Vertrauensbereiche der Vorhersage, die auf einem Ensemble von 32 Vorhersagen basieren, für drei verschiedene Gitterpunkte. Die durchgezogene Kurve gibt die Normalvorhersage, die gepunktete Kurve stellt die Analyse dar, also den tatsächlichen eingetretenen Verlauf der 850-hPa Temperatur.

In allen Fällen weicht das Ensemble in den ersten drei Tagen relativ wenig von der Analyse und vom Normalverlauf ab, doch dann beginnt sich die Stärke der Monte-Carlovorhersagemethode zu zeigen. Am Gitterpunkt b weicht die aktuelle Entwicklung nach 6 Tagen stark von der Normalvorhersage ab, doch deutet dies das Ensemble auch klar an. Dagegen erfaßt für Gitterpunkt C nur eine Vorhersage aus 32 die wirkliche Entwicklung, während alle anderen Mitglieder des Ensembles in die Irre gehen. Immerhin hätte dieser eine "Ausreißer" eventuell als Warnung dienen können.



a-c. Confidence intervals for the ensemble forecast from 3 January 1987 of 850 hPa temperature. Shown at three grid points throughout the forecast range. Contours shown are 99%, 90%, 70% and 50%.

Abb. 28

6. Epilog: Nachweis des Treibhauseffekts?

Man könnte erwarten, daß nun, am Ende des Kurses, die Methoden bereitstehen, die erlauben zu entscheiden, ob und in welchem Ausmaß sich der Treibhauseffekt nachweisen läßt. Dieser Erwartung kann hier nicht ganz entsprochen werden, doch beginnen diese Methoden Form zu gewinnen. In Hasselmann (MPI-Report 88, 1992) wird der nötige Rahmen aufgestellt, die Anwendung ist im Gange. Doch soll hier der Versuch unternommen werden, wenigstens die Hauptzüge einer solchen Teststrategie anzugeben, die vom Konzept her relativ einfach ist.

Normalerweise stellt man die Klimaentwicklung durch den Vektor $\mathbf{X}(t)$ dar im Sinne des Modells (5.18), wobei der Vektor \mathbf{X} etwa die Werte aller Variablen an den Gitterpunkten eines Analysennetzes oder eines Modells ausgibt. Die Zeit schreitet in diskreten Schritten voran. Zur Diagnose einer Klimaänderung wird man nicht alle Variablen heranziehen können, sondern sich auf einen niedrigdimensionalen Satz von Variablen beschränken. So basiert die Diskussion in (1.1) nur auf einer einzigen Variablen. Ferner ist es günstig, wenn man ein Raumzeitnetz betrachtet, so daß der "totale" Klimavektor ψ nicht mehr von der Zeit abhängt. In diesem Vektor sind die Werte aller zur Analyse zugelassenen Variablen zu allen verfügbaren Zeitpunkten gespeichert. So enthielte der Klimavektor für Abb. 1 die Komponenten $(T_{1860}, T_{1861}, \dots)$.

Nun wird angenommen, daß sich der Klimavektor ψ aufspalten läßt gemäß

$$\psi = \psi^s + \tilde{\psi} \quad (6.1)$$

wobei ψ^s das, durch externen Antrieb erhöhte (Treibhausgasproduktion) und gesuchte Signal ist und $\tilde{\psi}$ die natürliche Variabilität des Klimas repräsentiert. Ferner nimmt man an, daß ψ einen statistisch stationären Prozeß darstellt und daß man die Kovarianzmatrix

$$\mathbf{C} = \overline{(\tilde{\psi}\tilde{\psi})} \quad (6.2)$$

kennt, wobei der Querstrich hier eine Mittelung über ein Ensemble meint (siehe 1.1; die Bestimmung von \mathbf{C} anhand der einzigen Realisierung des Ensembles, die wir kennen, stellt natürlich ein Problem dar).

Die entscheidende Voraussetzung ist nun, daß man von der Theorie her mindestens das Muster der erwarteten Klimaänderung kennt:

$$\psi^s = \mathbf{c}g \quad (6.3)$$

Für Abb. 1 entspricht \mathbf{g} im wesentlichen einem exponentiellen Anstieg. Dann reduziert sich das Problem darauf, daß man eine optimale Detektorvariable

$$\mathbf{d} = (\mathbf{f} \cdot \boldsymbol{\psi}) \quad (6.4)$$

eingführt, wobei \mathbf{f} der sogenannte "Fingerabdruck" (fingerprint) ist, für den das Signal-Rausch-Verhältnis

$$R^2 = (d^s)^2 / \overline{\tilde{d}^2} \quad (6.5)$$

maximal wird. Nur bei solcher Maximierung kann man hoffen, daß das errechnete Klimasignal sich als statistisch signifikant erweisen wird. In (6.5) ist

$$d^s = (\mathbf{f} \cdot \boldsymbol{\psi}^s) \quad (6.6)$$

$$\tilde{d} = (\mathbf{f} \cdot \tilde{\boldsymbol{\psi}}) \quad (6.7)$$

und

$$\mathbf{d} = (\mathbf{f} \cdot \boldsymbol{\psi}) = d^s + \tilde{d}. \quad (6.8)$$

Damit ist d^s die Projektion des Signals auf den Fingerabdruck und \tilde{d} die Projektion des natürlichen Klimas auf \mathbf{f} . Statt R^2 zu maximieren kann man auch $\overline{\tilde{d}^2}$ minimieren mit der Nebenbedingung, daß $(d^s)^2$ konstant ist. Man hat also nach einem Minimum von

$$\overline{(\mathbf{f} \cdot \tilde{\boldsymbol{\psi}})^2} + \lambda c^2 (\mathbf{f} \cdot \mathbf{g})^2 \quad (6.9)$$

durch Variation von \mathbf{f} zu suchen. Dabei kann c^2 in λ aufgenommen werden. Also muß gelten

$$\mathbf{C} \cdot \mathbf{f} + \lambda (\mathbf{f} \cdot \mathbf{g}) \mathbf{g} = 0. \quad (6.10)$$

Das ist eine lineare Gleichung für \mathbf{f} . Eine Lösung findet sich durch

$$\mathbf{f} = \mathbf{C}^{-1} \cdot \mathbf{g} \quad (6.11)$$

mit

$$\lambda = -(\mathbf{g} \cdot \mathbf{C} \cdot \mathbf{g})^{-1}. \quad (6.12)$$

Damit ist der Fingerabdruck festgelegt und man hat nun die Klimaänderung abzuschätzen gemäß

$$\boldsymbol{\psi}^{\ddot{a}} = c_2 \mathbf{g}. \quad (6.13)$$

Die Konstante c_2 legt man dadurch fest, daß bei Projektion auf \mathbf{f} das tatsächliche Klima herauskommen soll:

$$c_2 = \frac{d}{(\mathbf{f} \cdot \mathbf{g})} \quad (6.14)$$

Nun hat man noch die Signifikanz des Signals d^s nachzuweisen. Dies wird man gemäß 2.3 durchführen können unter der Annahme, daß \tilde{d} eine Normalverteilung hat.

Ein Beispiel mache das Verfahren etwas anschaulicher. Die einzige Klimavariablen $T(^{\circ}\text{C})$ sei für zwei Jahre bekannt.

$$\psi_1 = T_1 = 0.5$$

$$\psi_2 = T_2 = -0.5$$

Man habe eine Theorie, die eine Klimaänderung

$$g_1 = -1, g_2 = 1$$

vorhersagt. Ferner sei $\overline{\psi_1^2} = 2, \overline{\psi_1\psi_2} = -1, \overline{\psi_2^2} = 2$. Dann liefert (6.11) $f_1 = -1/3, f_2 = 1/3$ und man hat $d = -\frac{1}{3}$ so daß $c_2 = -\frac{1}{2}$. Damit ist die analysierte Klimaänderung (0.5, -0.5), die in diesem trivialen Fall genau mit der Beobachtung übereinstimmt.

In Abb. 29 ist der optimale Fingerabdruck gezeigt, der sich bei einer Klimasimulation mit erhöhtem Treibhausgas und Aerosol ergibt (aus Hegerl et al. 1997; Climate Dyn.). Man würde dann eine leichte Abkühlung über dem Atlantik, Erwärmung über den Kontinenten und stärkere Erwärmung in der Südhemisphäre als in der Nordhemisphäre sehen. Das Signal/Rauschverhältnis ist dabei von der Größenordnung 3-6 mit Signifikanzen von gut über 90 Prozent. Die Antwort scheint also zu sein, daß die beobachtete Klimaänderung tatsächlich durch die Treibhausgase verursacht ist.

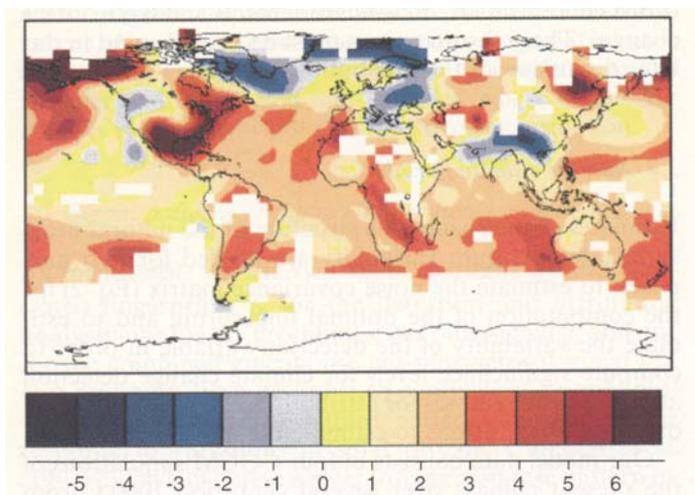


Abb. 29. Optimaler Fingerabdruck bei Klimasimulation mit erhöhtem Treibhausgas und Aerosol. Man sieht eine Abkühlung im Nordatlantik und größere Erwärmungsraten in der Südhemisphäre.