

**Über die Statistik der
Maximawerte und Wiederkehrintervalle
in langzeitkorrelierten Systemen**

Promotionsarbeit im Fach Physik

vorgelegt von

JAN F. EICHNER
geb. in Freiburg i. Br.

Institut für Theoretische Physik
Justus-Liebig-Universität Gießen

26. Januar 2006

Inhaltsverzeichnis

1	Einleitung	1
2	Langzeitkorrelationen in der Zeitreihenanalyse	3
2.1	Statistische Grundlagen der Zeitreihenanalyse	4
2.2	Langzeitkorrelationen und ihre Detektionsmethoden	7
2.2.1	Autokorrelationsfunktion	7
2.2.2	Fourier-Transformation und Power-Spektrum	8
2.2.3	Trendbereinigende Fluktuationsanalyse (DFA)	11
2.3	Langzeitkorrelationen in der Natur	15
2.3.1	Taglich gemessene Temperaturzeitreihen	17
2.3.2	Rekonstruierte und historische Zeitreihen	22
2.4	Erzeugung langzeitkorrelierter Datensatze	25
2.4.1	Fourier-Filtering Methode	26
2.4.2	Iterative Methode zur Erhaltung der Ausgangsverteilung (Schreiber-Verfahren)	28
3	Extremwertstatistik I: Wiederkehrintervalle	31
3.1	Klassische Theorie der Wiederkehrintervalle	31
3.1.1	Definition von Wiederkehrintervallen und Quantilen	32
3.1.2	Wiederkehrperioden bei verschiedenen Verteilungsfunktionen	33
3.1.3	Poisson-Statistik bei unkorrelierten Systemen	35
3.2	Statistik der Wiederkehrzeiten bei Langzeitkorrelationen	36
3.2.1	Wiederkehrperiode und Varianz	37
3.2.2	Verteilungsdichte der groen Wiederkehrintervalle	39
3.2.3	Verteilungsdichte der kleinen Wiederkehrintervalle	45

3.2.4	Korrelationseigenschaften der Reihen von Wiederkehrintervallen	48
3.2.5	Bedingte Wiederkehrintervalle	53
3.2.6	Bedingte Verteilungsdichten der Wiederkehrintervalle	58
3.2.7	Bedingte Latenzzeiten	62
4	Extremwertstatistik II:	
	Statistik der Maximawerte	67
4.1	Klassische Theorie der Extremwerte	69
4.1.1	Das klassische Modell	69
4.1.2	Drei Typen von Extremwertverteilungen	71
4.1.3	Generalisierte Extremwertverteilung GEV	73
4.1.4	Verteilungen im Anziehungsbereich der Gumbel-Verteilung	74
4.1.5	Reale Systeme und endliche Reihenlängen	75
4.2	Statistik der Maximawerte bei Langzeitkorrelationen	76
4.2.1	Verteilungen der Maxima	77
4.2.2	Korrelationen in den Maxima	82
4.2.3	Bedingte Maxima	83
4.2.4	Bedingte Maximaverteilungen und bedingte Überschreitungswahrscheinlichkeiten	86
4.2.5	Einfluss der Langzeitkorrelationen auf Jahrhundertereignisse	90
5	Zusammenfassung und Ausblick	97
A	Tabelle	101
B	Beweis zur Verteilung der Wiederkehrzeiten	103
B.1	Zero-Level Crossing nach Newell und Rosenblatt	103
B.2	Erweiterung auf q -Level Crossing (POT Methode)	107

Abbildungsverzeichnis

2.1	Autokorrelationsfunktion künstlicher korrelierter Zufallszahlen ohne Trend . . .	8
2.2	Autokorrelationsfunktion langzeitkorrelierter Zufallszahlen mit Trend	9
2.3	Power-Spektrum kurzzeitkorrel., langzeitkorrel. und trendüberlagerter Daten . .	11
2.4	Poynomiale Approximation des Profils auf Fenstern der Breite s	13
2.5	DFA künstlicher Reihen, unkorreliert und langzeitkorreliert	13
2.6	Beispiele langzeitkorrelierter Zeitreihen	15
2.7	Histogramm der α -Werte verschiedener weltweiter Temperaturreihen	16
2.8	Vier DFA-Beispiele der australischen Temperaturstationen	18
2.9	Geografische Lagen der australischen Temperaturstationen	19
2.10	Histogramm der α -Werte der australischen Temperaturstationen	19
2.11	α -Werte der australischen Stationen als Funktion der südlichen Breite	20
2.12	DFA2 Kurven 10 sibirischer und 10 nordamerikanischer Temperaturreihen . .	21
2.13	Skalenexponent α als Funktion des Abstands vom Meer	21
2.14	Historische Wasserstände der Pegnitz	22
2.15	Rekonstruierte und historische hydrometeorologische Meßreihen	24
2.16	DFA rekonstruierter und historischer Daten	25
2.17	Fourier-Filtering Methode zur Erzeugung langzeitkorrelierter Daten	27
2.18	Diagramm der einzelnen Iterationsschritte des Schreiber-Verfahrens	28
2.19	Qualitätstest von Langzeitkorrelationen in künstlichen Daten	30
3.1	Definition von Wiederkehrintervallen	32
3.2	Definition von Quantilen	33
3.3	Vier Verteilungsfunktionen künstlicher Daten	34
3.4	Wiederkehrintervallreihen unkorrelierter und langzeitkorrelierter Daten	37
3.5	Wiederkehrperioden	38

3.6	Standardabweichung der Wiederkehrintervalle	39
3.7	Verteilungsdichte der Wiederkehrintervalle	40
3.8	Gestreckt exponentielle Funktion	41
3.9	Verteilungsdichte der Wiederkehrzeiten vier verschiedener Grundverteilungen	43
3.10	Finite-Size Effekte in den Verteilungsdichten der Wiederkehrintervalle	43
3.11	DFA und Verteilungsfunktion der Wiederkehrzeiten echter Daten	44
3.12	Verteilung der kleinen Wiederkehrzeiten	46
3.13	Verteilung der Wiederkehrzeiten der Länge 1	47
3.14	Autokorrelationsfunktion der Wiederkehrzeiten	48
3.15	Finite-Size Untersuchung der Autokorrelationen bei Wiederkehrzeiten	49
3.16	Autokovarianzfunktion der Wiederkehrintervalle der vier Verteilungen	50
3.17	Fluktuationsanalyse der Wiederkehrzeiten	51
3.18	Fluktuationsanalyse der Wiederkehrintervalle Gauss-verteilter Daten	52
3.19	Bedingte Wiederkehrzeiten	53
3.20	Bedingte Wiederkehrperioden	54
3.21	Bedingte Wiederkehrperioden der vier Verteilungen	55
3.22	Bedingte Wiederkehrzeiten als Funktion von γ	57
3.23	Bedingte Wiederkehrperioden der vier Verteilungen als Funktion von γ	58
3.24	Verteilungsdichtefunktion der bedingten Wiederkehrintervalle	59
3.25	Bedingte Verteilungsdichtefunktion der Wiederkehrzeiten echter Daten	60
3.26	Quotient aus bedingter u. unbedingter Verteilungsdichte der Wiederkehrzeiten	61
3.27	Definition der bedingten Latenzzeiten	62
3.28	Bedingte Latenzzeiten künstlicher Daten	63
3.29	Bedingte Latenzzeiten als Funktion von r_0 und x	64
3.30	Bedingte Latenzzeiten als Funktion von γ	65
4.1	Die drei Extremwertverteilungen: Gumbel, Fréchet und Weibull	72
4.2	Täglicher, jährlicher und maximaler Niederschlag in Oxford	74
4.3	Definition von Maximawerten (Blockmaxima)	76
4.4	Verteilung der Maxima, unkorreliert	78
4.5	Verteilung der Maxima, korreliert	80
4.6	Überschreitungswahrscheinlichkeit, unkorreliert und korreliert	81
4.7	Vergleich der Maxima-Selektion bei korrelierten und unkorrelierten Daten . .	81

4.8	Autokorrelationsfunktion der Maximawerte	82
4.9	Finite-Size Untersuchungen an der Autokorrelationsfunktion der Maxima	83
4.10	Definition bedingter Maxima	84
4.11	Mittlere bedingte Maxima	85
4.12	Mittlere bedingte Minima des Nil und Maxima der Moberg-Reihe	86
4.13	Bedingte Verteilungsdichte und Überschreitungswahrscheinlichkeiten	87
4.14	Bedingte Überschreitungswahrscheinlichkeiten	88
4.15	Relative bedingte Überschreitungswahrscheinlichkeiten	89
4.16	Drei Definitionen 100-jähriger Ereignisse	90
4.17	Vergleich der Histogramme von hundertjährigen Ereignissen	92
4.18	Hundertjährige Ereignisse als Funktion des Korrelationsexponenten	93
4.19	Bedingte hundertjährige Ereignisse	94
B.1	Zero-Level Crossing beim kontinuierlichen Gauss-Prozess	104

Kapitel 1

Einleitung

In den letzten zehn Jahren hat man in Mitteleuropa eine Zunahme extremer Naturereignisse beobachtet. So hatten wir in Deutschland vier große Überflutungen in nur acht Jahren: die Oderflut (1997), das Pfingsthochwasser in Bayern (1999), die Elbeflut (2002) und das Augusthochwasser in Bayern (2005). Drei der letzten sieben Winter brachten im Alpenraum weit überdurchschnittlich viel Schnee, was zu vielen tragischen Lawinenunglücken geführt hat. Aber auch das Wetter hat in den letzten fünfzehn Jahren Rekorde gebrochen. So war der Rekordsommer 2003 der wärmste Sommer und der Winter 1989/90 der mildeste Winter, der seit Aufzeichnung der Temperaturen in Deutschland gemessen wurde, während der aktuelle Winter (2005/06) eine Kältewelle mit rekordverdächtigen Tiefsttemperaturen über Ost- und Mitteleuropa brachte. Diese Häufungen extremer Klimaereignisse beschränken sich nicht auf Europa. Während in einem durchschnittlichen Jahr etwa 13 bis 15 tropische Stürme über dem Atlantik entstehen, von denen sich im Mittel 8 bis 9 zur Hurrikanstärke entwickeln, waren es im letzten Sommer (2005) 27 tropische Stürme, davon 22 Hurrikane.

Eines haben alle diese extremen Ereignisse gemein: Sie sind in kurzen Phasen häufiger aufgetreten, als man es statistisch erwarten würde. Man neigt intuitiv dazu, die Häufung dieser extremen Naturereignisse als *nicht*-zufällig anzusehen.

In der öffentlichen Diskussion werden diese extremen Klimaphänomene schnell der globalen Klimaerwärmung zugeschrieben. Dass der aktuelle Klimawandel und seine Auswirkungen (unabhängig davon, ob er natürlichen Ursprungs, antropogenen Ursprungs oder eine Mischung aus beidem ist) zu einer größeren Variabilität des Klimas und zum häufigeren Auftreten extremer Wetterereignisse führen kann, ist unumstritten [Latif 2004, Grassel 2005]. Schaut man aber in lange gemessene oder rekonstruierte Datenreihen, die mehrere Jahrhunderte zurück reichen, oder sieht man in historischen Überlieferungen und Geschichtsbüchern nach, so findet man, dass extreme Ereignisse wie Dürren und Fluten oder strenge Winter auch schon damals in bestimmten Epochen gehäuft auftraten und in anderen wiederum fast gar nicht. Häufungen extremer Ereignisse gab es somit schon in Zeiten, in denen der Mensch noch keinen Einfluss auf das Klima nahm. Es muss also auch einen natürlichen Mechanismus geben, der das gehäufte Auftreten von extremen Naturphänomenen (zumindest im Mittelalter) erklären kann.

Mittels der Extremwertstatistik versucht man heutzutage die Eintrittswahrscheinlichkeiten solcher Extremereignisse möglichst präzise abzuschätzen, um durch geeignete Maßnahmen

die Risiken einer Katastrophe zu minimieren. In der klassischen Extremwertstatistik geht man davon aus, dass extreme Ereignisse, die zeitlich weit genug auseinander liegen, statistisch unabhängig sind. In der Praxis ist diese statistische Unabhängigkeit unrealistisch. Ein Beispiel soll dies erklären. Aus eigener Erfahrung weiß man, dass - egal ob Sommer oder Winter - beispielsweise auf einen sehr warmen Tag wohl wieder ein sehr warmer Tag folgen wird und auf einen kalten Tag wohl eher wieder ein kalter Tag. Dies hängt mit den sogenannten Großwetterlagen zusammen, den "Hochs" und "Tiefs", von denen im täglichen Wetterbericht die Rede ist. Sie halten einen Wetterzustand über mehrere Tage bis hin zu Wochen stabil und ermöglichen Wettervorhersagen. Diese Erhaltungsneigung, auch Persistenz genannt, kann eine charakteristische Eigenschaft eines vermeintlich zufälligen Prozesses sein. Analysen von langen Temperaturzeitreihen zeigen, dass solche Kopplungen (Korrelationen) noch weit über die Zeitskalen von Großwetterlagen hinaus bis hin zu mehreren Dekaden reichen können [Koscielny 1998a]. Solche lang anhaltenden Korrelationen werden Langzeitkorrelationen genannt und zeichnen sich durch eine Potenzgesetz-artig abfallende Autokorrelationsfunktion aus.

Neben den atmosphärischen Temperaturzeitreihen findet man Langzeitkorrelationen in den unterschiedlichsten Prozessen in der Natur, wie zum Beispiel in hydrologischen Abflusszeitreihen, in Meeresoberflächentemperaturen, in biomedizinischen Reihen von Herzschlagintervallen, in DNA-Sequenzen, sogar in Volatilitäten von Aktienkursen und im Internetverkehr [Bunde 1994, Bunde 2001, Bunde 2002b].

Im Rahmen dieser Arbeit wurde der Einfluss von Langzeitkorrelationen auf die Extremwertstatistik untersucht. Dabei habe ich mich auf zwei Gebiete der Extremwertstatistik konzentriert: zum einen die Statistik der Wiederkehrintervalle (POT-Methode) und zum anderen die Statistik der Maximawerte (Gumbel-Statistik). Die Arbeit ist wie folgt aufgebaut. In Kapitel 2 gebe ich eine Übersicht der statistischen Grundlagen und stelle Methoden zur Korrelationsanalyse vor. Verschiedene gemessene und rekonstruierte Temperaturzeitreihen werden beispielhaft auf Langzeitkorrelationen untersucht und ihre Resultate mit denen anderer Gruppen verglichen. Im selben Kapitel werden numerische Verfahren beschrieben, mittels derer ich mir künstliche langzeitkorrelierte Datensätze zu unterschiedlichen Verteilungen erzeugt habe. Kapitel 3 beginnt mit der klassischen Beschreibung der Statistik der Wiederkehrintervalle (Poisson-Statistik). Es folgen die Ergebnisse zu den Auswirkungen von Langzeitkorrelationen auf die Intervalle anhand simulierter und echter Datenreihen. Die Form der Verteilungsfunktion der Wiederkehrintervalle wird ausführlich diskutiert (auch im Anhang). Zusätzlich werden ausgiebige Finite-Size Untersuchungen und Korrelationsanalysen an den Reihen der Wiederkehrintervalle durchgeführt. Es wird außerdem gezeigt, wie Langzeitkorrelationen zur Clusterbildung extremer Ereignisse führen können und wie man mittels bedingter Wiederkehrintervalle zu einer verbesserten Risikoabschätzung gelangt. Kapitel 4 beschreibt zunächst den Aufbau der klassischen Extremwertstatistik und das Extremwerttheorem. Danach werden die Ergebnisse der umfassenden numerischen Untersuchungen zum Einfluss von Langzeitkorrelationen auf die Maximawerte und ihrer Verteilung vorgestellt. Es wird gezeigt, wie mittels bedingter Maximawerte und bedingter Überschreitungswahrscheinlichkeiten Risikoabschätzungen in der Maximastatistik auf langzeitkorrelierten Systemen verbessert werden können. Das Ende von Kapitel 4 beschäftigt sich mit dem Korrelationseinfluss auf die Abschätzung von Jahrhundertereignissen, wie sie in der praktischen Hydrologie üblich ist. Kapitel 5 fasst die Ergebnisse aus Kapitel 3 und 4 zusammen und gibt einen kurzen Ausblick über mögliche Auswirkungen anderer statistischer Abhängigkeiten auf die Extremwertstatistik.

Kapitel 2

Langzeitkorrelationen in der Zeitreihenanalyse

Die Dynamik komplexer (hochdimensionaler) Systeme ist aufgrund der großen Anzahl an Freiheitsgraden und oft auch nicht-linearer Kopplungen nicht exakt beschreibbar [Alligood 1996, Lam 1997]. Man beschränkt sich auf die Untersuchung geeigneter Messgrößen, welche ein dimensionsreduziertes System darstellen, und akzeptiert, dass diese Systeme zufälligen und zufällig erscheinenden Fluktuationen unterworfen sind und nur noch Wahrscheinlichkeitsaussagen zulassen. Die Untersuchung dieser *stochastischen* Fluktuationen ist eine zentrale Aufgabe der Zeitreihenanalyse [Chatfield 1982, Kendall 1990, Schlittgen 1990, Box 1994, Brockwell 1996]. Statistische Abhängigkeiten wie Korrelationen können ausgenutzt werden, um Wahrscheinlichkeitsaussagen zu verbessern. Eine besondere Form dieser Abhängigkeiten bilden die *Langzeitkorrelationen*, auch *Persistenz* genannt [Hurst 1965, Mandlbrot 1987, Bunde 1994, Bunde 2001]. Prozesse bzw. Zufallsgrößen sind *persistent*, wenn sie dazu neigen, ihren augenblicklichen Zustand bzw. Wert beizubehalten, das heißt auf große Werte folgen eher große Werte und auf kleine Werte eher kleine Werte. Persistenz kann dadurch über gewisse Zeiträume zu erheblichen Abweichungen vom langzeitlichen, globalen Mittelwert führen, was auf kurzen Zeitskalen bei der Unterscheidung zwischen Trends und Korrelationen Probleme bereitet [Feder 1988, Peng 1992, Kantelhardt 2001a, Cohn 2005]. Zur sicheren Korrelationsanalyse sind moderne Methoden der statistischen Physik notwendig.

Dieses Kapitel soll einen Überblick über Eigenschaften von Langzeitkorrelationen und deren Auftreten in der Natur geben. Dazu werden zuerst die statistischen und mathematischen Grundlagen vorgestellt und die Eigenschaften von Langzeitkorrelationen beschrieben. Danach werden kurz die Methoden vorgestellt, mittels derer man Langzeitkorrelationen bestimmen kann. Der nächste Abschnitt zeigt beispielhaft Ergebnisse zu von mir auf Langzeitkorrelationen untersuchten meteorologischen und hydrologischen Zeitreihen.

Da die Analyse echter Daten zur Studie von Langzeitkorrelationen aufgrund der oft unzureichenden Statistik nicht ausreicht und außerdem Kurzzeitkorrelationen bei kurzen Datensätzen einen störenden Einfluss haben, empfiehlt es sich, künstlich erzeugte Daten mit realistischen Langzeitkorrelationsstrukturen zu erzeugen und zu studieren. Man benötigt dazu numerische Methoden, mit denen man sich beliebig viele Daten mit vorgegebenen Korrelationseigenschaften erzeugen kann. Zwei Methoden [Makse 1996, Schreiber 1996], die von mir zur Erzeugung

künstlicher langreichweitig korrelierter Zufallszahlen verwendet wurden, werden am Ende dieses Kapitels beschrieben.

2.1 Statistische Grundlagen der Zeitreihenanalyse

Die Analyse komplexer dynamischer Systeme basiert auf der Aufzeichnung von geeigneten Messgrößen x (Variablen) über bestimmte Zeitintervalle. Diese Messreihen werden als *Zeitreihen* bezeichnet [Chatfield 1982, Kendall 1990, Schlittgen 1990, Box 1994, Brockwell 1996]. Viele Variablen sind nur an diskreten Zeitpunkten definiert, wie zum Beispiel monatliche Gehaltsauszahlungen, Ernteerträge oder Bundesligaergebnisse. Andere Variablen werden durch Aggregation über eine gewisse Zeitdauer erzeugt. So zum Beispiel Tagesniederschlag, jährliche Passagier-Meilen einer Fluggesellschaft oder das Bruttosozialprodukt. Eine Messgröße, die zu allen Zeitpunkten definiert ist, liefert eine kontinuierliche Messreihe wie beispielsweise Temperaturen an einem festen Ort, die Position eines Flugzeugs oder Pegelstände eines Flusses. Aus praktischen Gründen werden viele kontinuierliche Zeitreihen diskretisiert. Bei Temperaturen und Flüssen betrachtet man zum Beispiel Tagesmaxima, Tagesminima oder Tagesmittelwerte. Man erhält so wieder eine Zeitreihe $(x_i)_{i=1}^N$, bestehend aus einer endlichen Anzahl N von Messpunkten, der “neuen” Länge der Zeitreihe.

Die Messgrößen x sind häufig Beschränkungen unterworfen und verschiedene Messwerte x_i in einer Zeitreihe $(x_i)_{i=1}^N$ treten unterschiedlich häufig auf. Ihr genauer Wert erscheint zwar zufällig, gehorcht aber einer zugrunde liegenden Verteilungsfunktion $P(x)$, welche oft nicht exakt bekannt ist, aber durch ein Histogramm einer hinreichend großen Stichprobe (z. B. einer langen Zeitreihe) recht gut angenähert werden kann. Dabei sind Aussagen über *selten* auftretende Ereignisse unweigerlich schlechter als Aussagen über *häufig* auftretende Ereignisse.

Verteilungen von Messwerten werden unterschieden in *Verteilungsfunktionen* und deren Dichten, den *Verteilungsdichtefunktionen*¹ [Sachs 1974, Papoulis 1984, Papula 1994]. Die Dichten stellen hierbei das normierte Histogramm einer hinreichend großen Stichprobe da, während man die Verteilungsfunktionen durch das Integral über die normierten Dichten erhält und damit Aussagen über die Wahrscheinlichkeiten für das Auftreten von Ereignissen gewisser Größe oder Mindestgröße treffen kann. In dieser Arbeit werden alle Verteilungsdichten mit P bezeichnet, die zugehörigen Verteilungsfunktionen in Kapitel 3 mit W und in Kapitel 4 mit G bzw. $E = 1 - G$.

Um aus einer Verteilungsfunktion eine Wahrscheinlichkeitsaussage ableiten zu können, muss die zugehörige Dichtefunktion normiert sein [Sachs 1974], das heisst

$$\int_{-\infty}^{\infty} P(x)dx = 1. \quad (2.1)$$

Der *Erwartungswert* einer Verteilung ist definiert als das erste Moment dieser Verteilung [Sachs 1974, Papoulis 1984, Papula 1994]

$$\bar{x} := \int_{-\infty}^{\infty} xP(x)dx. \quad (2.2)$$

¹In der Literatur werden oft auch die Bezeichnungen *Wahrscheinlichkeitsverteilung* und *Wahrscheinlichkeitsdichte* verwendet

Ist die Verteilungsdichte $P(x)$ nicht analytisch bekannt, so kann man den *arithmetischen Mittelwert*

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.3)$$

einer Stichprobe mit hinreichendem Umfang N als numerische Näherung für den Erwartungswert [Sachs 1974, Papula 1994] betrachten,

$$\langle x \rangle \simeq \bar{x}. \quad (2.4)$$

Aus den höheren Momenten

$$\overline{x^n} = \int_{-\infty}^{\infty} x^n P(x) dx \quad (2.5)$$

lassen sich Aussagen über Symmetrie, Schiefe und Wölbung der Verteilungsdichtefunktion gewinnen [Sachs 1974, Papula 1994]. Eine zentrale Rolle nimmt dabei das zweite Moment $\overline{x^2}$ ein. Aus erstem und zweitem Moment erhält man die *Varianz*

$$var_x := \int_{-\infty}^{\infty} (x - \bar{x})^2 P(x) dx = \int_{-\infty}^{\infty} (x^2 - 2\bar{x}x + \bar{x}^2) P(x) dx = \overline{x^2} - \bar{x}^2 \quad (2.6)$$

als Maß für die Breite der Verteilungsdichtefunktion [Sachs 1974, Papula 1994]. Ist die Verteilungsdichtefunktion analytisch nicht bekannt, muss man auf eine alternative numerische Definition, die auf einer Stichprobe vom Umfang N basiert, zurückgreifen

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2, \quad (2.7)$$

wobei ähnlich wie beim arithmetischen Mittel bei hinreichend großer Stichprobe N die Näherung

$$\sigma_x^2 \simeq var_x \quad (2.8)$$

erfüllt ist. Die Größe σ_x wird als *Standardabweichung* bezeichnet und macht eine Angabe über die (symmetrische) Streuung der Stichprobenwerte x_i um das arithmetische Mittel $\langle x \rangle$ [Sachs 1974, Papula 1994]. Da in dieser Arbeit der Schwerpunkt auf numerischen Untersuchungen mit riesigen Stichproben liegt, kann hier die Annahme $\langle x \rangle = \bar{x}$ und $\sigma_x^2 = var_x$ gemacht werden.

Eine Größe zur Beschreibung von Abhängigkeiten der Zufallsvariablen untereinander ist die *Kovarianz* [Sachs 1974, Brockwell 1996]. Sie ist definiert als das mittlere Produkt² aus zwei mittelwertbereinigten Werten x_i und x_{i+s} im Abstand s ,

$$cov_x(s) := \langle (x_i - \langle x \rangle)(x_{i+s} - \langle x \rangle) \rangle \equiv \frac{1}{N-s} \sum_{i=1}^{N-s} (x_i - \langle x \rangle)(x_{i+s} - \langle x \rangle). \quad (2.9)$$

Bei statistisch unabhängigen Werten mit gleicher Verteilungsfunktion (“White Noise”) ist die Kovarianz für alle $s > 0$ im Mittel Null. Für $s = 0$ ist sie identisch zur Varianz,

$$cov_x(0) \equiv \sigma_x^2. \quad (2.10)$$

²Oft wird der Vorfaktor als $1/N$ angegeben, was im Falle $s \ll N$ keinen bedeutenden Unterschied macht.

Der Quotient aus Kovarianz und Varianz wird als *Autokorrelationsfunktion*

$$C(s) := \frac{\text{cov}_x(s)}{\text{cov}_x(0)} \quad (2.11)$$

bezeichnet [Sachs 1974, Brockwell 1996]. Aus dieser Definition folgt direkt $C(0) = 1$. Eine Autokorrelationsfunktion unterschiedlich von Null für $s > 0$ kennzeichnet statistische Abhängigkeit. Die Art ihres Abfalls definiert den Einfluss von vergangenen Ereignissen auf zukünftige mit zeitlichem Abstand s . Korrelierte Prozesse zeigen demnach eine Art *Gedächtnis* oder auch *Memory-Effekt*. Bei langzeitkorrelierten Reihen fällt die Autokorrelationsfunktion $C(s)$ nur sehr langsam (algebraisch) auf Null ab. Mehr dazu im nächsten Abschnitt.

Eine wichtige Eigenschaft eines Prozesses oder einer Zeitreihe ist die *Stationarität*. Ein stationärer stochastischer Prozess befindet sich in einem gewissen statistischen Gleichgewicht, womit gewährleistet sein soll, dass die statistischen Eigenschaften zeitlich konstant bleiben und sprunghafte Veränderungen von Mittelwert oder Varianz nicht auftreten. Man sagt, ein Prozess ist *stark stationär*, wenn für verschiedene Stichproben x_1, x_2, \dots, x_N und $x_{1+t}, x_{2+t}, \dots, x_{N+t}$ die Verteilungsfunktion und die Kovarianz im gesamten Prozess identisch und somit unabhängig von t sind [Chatfield 1982, Kendall 1990, Box 1994, Brockwell 1996, v.Storch 1999]. Ein Prozess (x_i) wird als *schwach stationär* bezeichnet, wenn Mittelwert $\langle x \rangle$, Varianz σ_x^2 und Kovarianz $\langle (x_i - \langle x \rangle)(x_{i+s} - \langle x \rangle) \rangle$ dieselben endlichen Werte annehmen und somit unabhängig von t sind [Chatfield 1982, Kendall 1990, Brockwell 1996, v.Storch 1999]. Für den praktischen Gebrauch ist *schwache* Stationarität vollkommen ausreichend. Starke Stationarität ist in der Praxis nicht so einfach nachzuweisen, da die zugrunde liegenden Verteilungsfunktionen oft nicht bekannt sind und die endlichen Stichproben meist nur brauchbare Ergebnisse für Mittelwert und Varianz liefern. In dieser Arbeit soll mit Stationarität immer die *schwache* Stationarität gemeint sein.

Stationarität ist oft eine Frage des zeitlichen Maßstabs [Brockwell 1996]. Eine tägliche Temperaturzeitreihe über wenige Monate ist nichtstationär aufgrund des saisonalen Verlaufs. Hingegen kann dieselbe Reihe über einen Zeitraum von vielen Jahrzehnten wieder stationär erscheinen, da die Saison auf großen Zeitskalen keinen Einfluss auf Mittelwert und Varianz nimmt. Ähnlich verhält es sich mit Korrelationen. Da Korrelationen eine Anordnungseigenschaft sind und keinen Einfluss auf die Verteilungsfunktion haben, verletzen sie die Stationarität nicht, solange die Stichproben nur groß genug ausfallen. In zu kurzen Ausschnitten können Korrelationen, insbesondere Langzeitkorrelationen, nur schwer von *Trends* unterschieden werden [Bunde 1994, Kantelhardt 2001a, Cohn 2005]. Trends, auch als *Nichtstationaritäten* bezeichnet, führen zu systematischen Abweichungen in Mittelwert oder Varianz und werden durch äußere oder innere Einflüsse verursacht [Sachs 1974, Kendall 1990, Brockwell 1996]. Bei den traditionellen statistischen Methoden in der Korrelationsanalyse und der Extremwertstatistik geht man davon aus, dass die zugrundeliegenden Prozesse stationär sind. Die Annahme stationären Verhaltens ist jedoch für viele natürliche Zeitreihen unrealistisch. Trotzdem ist es wünschenswert zu wissen, ob es sich bei der Reihe um eine trendbehaftete oder langzeitkorrelierte oder eine Reihe mit beidem handelt, und wie stark die jeweiligen Komponenten in den Reihen vertreten sind.

2.2 Langzeitkorrelationen und ihre Detektionsmethoden

Das Verhalten der Autokorrelationsfunktion Gl. (2.11) kann grob in drei Fälle unterteilt werden [Bunde 1994, Bunde 2001]:

- Für *unkorrelierte* Datensätze $(x_i)_{i=1}^N$ gilt aufgrund des Mittelwertabzugs in (2.9)

$$C(s) = 0 \quad \forall \quad s > 0. \quad (2.12)$$

- Im Falle *kurzreichweitiger* Korrelationen gibt es eine charakteristische *Korrelationszeit* oder *Korrelationslänge* s_\times . Der Zerfall der Korrelationen kann in vielen Fällen etwa durch einen exponentiellen Abfall von $C(s)$ beschrieben werden,

$$C(s) \sim e^{-\frac{s}{s_\times}}, \quad (2.13)$$

wobei die (endliche) Korrelationslänge definiert ist als

$$s_\times = \lim_{L \rightarrow \infty} \int_0^L C(s) ds. \quad (2.14)$$

- Bei *langreichweitigen* Korrelationen divergiert die Korrelationslänge s_\times und es gibt keine charakteristischen Zeitskalen. Dies ist zum Beispiel der Fall, wenn die Korrelationen durch ein abfallendes Potenzgesetz

$$C(s) \sim s^{-\gamma}, \quad 0 < \gamma < 1 \quad \text{und} \quad s \geq 1, \quad (2.15)$$

beschrieben werden können. Der Exponent γ wird als *Korrelationsexponent* bezeichnet.

Der Korrelationsexponent γ gibt die Stärke der Langzeitkorrelationen an: γ nahe bei 0 bedeutet starke Langzeitkorrelationen und γ nahe bei 1 schwache Langzeitkorrelationen. Der Fall $\gamma \geq 1$ wird zu den Kurzzeitkorrelationen gezählt, da die Korrelationslänge Gl. (2.14) für $s > 0$ wieder endlich ist. Ein γ -Wert kleiner oder gleich 0 bedeutet *nichtstationäres* Verhalten. Die Korrelationsfunktion würde mit anwachsendem Abstand s immer größer werden und der Mittelwert der Reihe könnte im Verlauf der Reihe nicht konstant bleiben. In dieser Arbeit interessiert aber ausschließlich der Fall $0 < \gamma < 1$.

2.2.1 Autokorrelationsfunktion

Die *Autokorrelationsfunktion* $C(s)$ wurde bereits in Gl. (2.11) im vorherigen Abschnitt vorgestellt. Sie beschreibt Zwei-Punkt-Korrelationen zum zeitlichen Abstand s einer stationären Zeitreihe [Brockwell 1996, v.Storch 1999],

$$C_x(s) = \frac{1}{\sigma_x^2 N} \sum_{i=1}^{N-s} (x_i - \langle x \rangle)(x_{i+s} - \langle x \rangle) \quad (2.16)$$

Oft findet man auch Definitionen der Kovarianzfunktion, in denen der Vorfaktor $1/(N-s)$ lautet, anstatt $1/N$. Da in der Praxis aber immer $s \ll N$ gilt, spielt diese Feinheit hier keine Rolle.

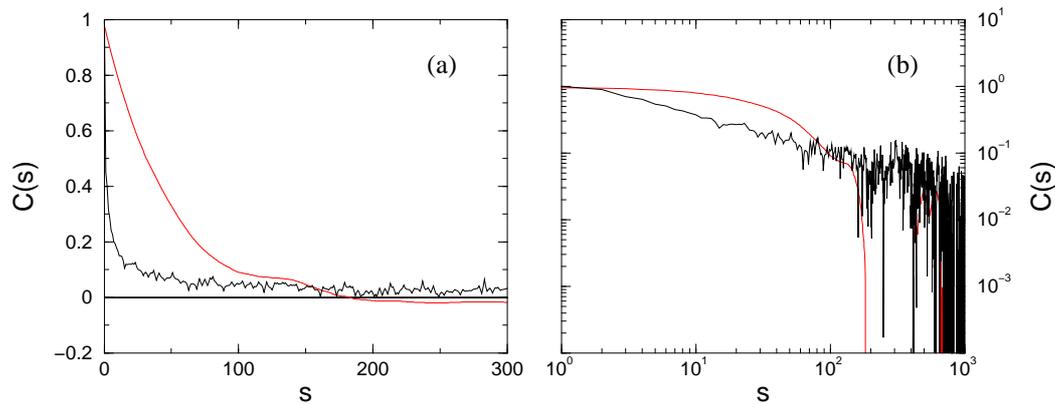


Abbildung 2.1: Autokorrelationsfunktionen künstlich erzeugter Zeitreihen bestehend aus 50000 Einträgen. Rote Linie: Kurzzeitkorrelationen mit einer Korrelationszeit von $s_{\times} = 100$; schwarze Linie: Langzeitkorrelationen mit dem Exponenten $\gamma = 0.6$. (a) lineare Darstellung, (b) doppelt-logarithmische Darstellung.

Abbildung 2.1 zeigt $C(s)$ für künstlich erzeugte kurzzeit- und langzeitkorrelierte Zufallszahlen mit Varianz $\sigma_x^2 = 1$. Die Reihen sind relativ lang (50000 Einträge), weshalb die Kurven recht glatt erscheinen. In der Praxis hat man es meist mit kürzeren Reihen zu tun, was zu einem weitaus verrauschteren Bild führt. Die Korrelationszeit s_{\times} beträgt in der kurzzeitkorrelierten Reihe 100. Bis zu Skalen $s \simeq 100$ ist $C_{kurz}(s)$ noch deutlich größer Null. In der doppelt-logarithmischen Darstellung (b) wird dies am starken Abfall im Bereich $s > 100$ deutlich. Die langzeitkorrelierte Reihe hat einen Korrelationsexponenten von $\gamma = 0.6$, was sich in der doppelt-logarithmischen Darstellung in der Steigung widerspiegelt. Zwar scheint in der linearen Darstellung (a) $C_{lang}(s)$ schneller abzufallen als $C_{kurz}(s)$. Dies täuscht jedoch, wie im doppelt-logarithmischen Plot (b) deutlich wird. Im Bereich großer Skalen liegt $C_{lang}(s)$ wieder deutlich über $C_{kurz}(s)$.

Nichtstationaritäten und die endliche Länge realer Zeitreihen machen es oft schwer oder sogar unmöglich, aus $C(s)$ das korrekte Korrelationsverhalten, d. h. im Falle langreichweitiger Korrelationen den tatsächlichen Korrelationsexponenten γ als Steigung in einem doppelt-logarithmischen Plot zu ermitteln. In Abbildung 2.2 ist $C(s)$ für langzeitkorrelierte Zufallszahlen ($\gamma = 0.6$) ohne bzw. mit einem schwachen linearen Trend abgebildet. In der doppelt-logarithmischen Darstellung (b) scheinen sich trendbehaftete und trendfreie Reihe nur in der effektiven Steigung zu unterscheiden. Ein Trend macht die Steigung flacher und man bekommt einen falschen Korrelationsexponenten. Die Autokorrelationsfunktion eignet sich zur Bestimmung des Korrelationsverhaltens nur bei sehr langen, stationären Zeitreihen.

2.2.2 Fourier-Transformation und Power-Spektrum

Betrachtet man die diskreten Werte einer Zeitreihe als Stützpunkte auf einer kontinuierlichen Kurve, so kann man mit Hilfe der *Fourier-Transformation* [Press 1997] Information über vorhandene Frequenzen (sozusagen Berge und Täler) im Verlauf dieser Kurve und damit auch Information über das Korrelationsverhalten der Zeitreihe gewinnen [Press 1997]. Man be-

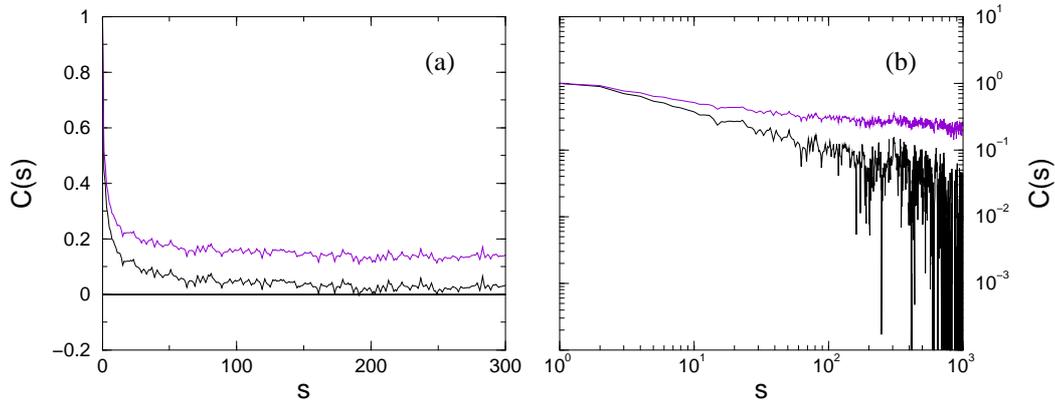


Abbildung 2.2: Autokorrelationsfunktionen künstlich erzeugter langzeitkorrelierter Zeitreihen ($\gamma = 0.6$) bestehend aus 50000 Einträgen. Schwarze Linie: ohne Trend; violette Linie: mit linearem Trend.

zeichnet diese Form der Korrelationsanalyse auch als *Spektralanalyse* [Press 1997]. Liegt dem erzeugenden Prozess der Zeitreihe eine Periodizität zugrunde, so ist sie im *Frequenzspektrum* (auch *Power-Spektrum* genannt) als Peak zu erkennen. Ist ein Prozess unkorreliert, so kommen alle Frequenzen gleichhäufig vor und das Frequenzspektrum ist ein Plateau. Ist ein Prozess korreliert, so kommen Frequenzen unterschiedlicher Größe unterschiedlich häufig vor. Im Falle von Langzeitkorrelationen liefert das Frequenzspektrum ähnlich wie die Autokorrelationsfunktion auch einen Potenzgesetzverlauf, wie in diesem Abschnitt gezeigt wird. Der analytische Zusammenhang zwischen Frequenzspektrum und Korrelationsverhalten ist auch für die im Abschnitt 2.4 vorgestellten Methoden zur Erzeugung von Langzeitkorrelationen von großer Bedeutung und soll hier ausführlich erklärt werden.

Zur Einfachheit und ohne Beschränkung der Allgemeinheit hat in diesem Abschnitt der stochastische Prozess (x_i) den Mittelwert $\langle x_i \rangle = 0$ und die Varianz $\sigma_x^2 = 1$. Die Fourier-Transformierte eines solchen stationären stochastischen Prozesses (x_i) der Länge N lautet dann [Papoulis 1984, Brockwell 1996]

$$F(f) = \frac{1}{\sqrt{N}} \sum_{j=1}^N x_j e^{-\frac{2\pi i}{N} f j}, \quad f \text{ durchläuft alle Frequenzen,} \quad (2.17)$$

wobei sie durch die Wurzel ihrer Länge normiert ist.

Das Frequenzspektrum $P(f)$ ist durch das Betragsquadrat der Fourier-Transformierten definiert [Press 1997] und lässt sich auf folgende Art als Fourier-Transformierte der Autokorrelationsfunktion darstellen, auch *Wiener-Khintchine-Theorem*³ genannt [WKT, Gottman 1981, Rangarajan 2000]:

$$P(f) \quad \equiv \quad F(f) \cdot F^*(f) \quad (2.18)$$

³Streng genommen gilt das Wiener-Khintchine-Theorem nur für die Definition der Autokorrelationsfunktion mit Vorfaktor $1/N$ anstatt $1/(N-s)$. Da aber in der Praxis immer $s \ll N$ gilt, spielt diese Feinheit keine Rolle.

$$\begin{aligned}
&= \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^N x_j x_k e^{-\frac{2\pi i}{N} f(j-k)} \\
&= \frac{1}{N} \sum_{j=1}^N x_j^2 + \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^{j-1} x_j x_k \left(e^{-\frac{2\pi i}{N} f(j-k)} + e^{\frac{2\pi i}{N} f(j-k)} \right) \\
&\stackrel{j-k=s>0 \text{ und (2.16)}}{=} C(0) + \frac{1}{N} \sum_{s=1}^N NC(s) \left(e^{-\frac{2\pi i}{N} fs} + e^{\frac{2\pi i}{N} fs} \right) \\
&= C(0) + 2 \sum_{s=1}^N C(s) \cos(2\pi fs/N) \tag{2.19}
\end{aligned}$$

Für Langzeitkorrelationen folgt mittels der Proportionalität aus Gleichung (2.15)

$$P(f) \sim \sum_{s=1}^N s^{-\gamma} \cos(2\pi fs/N). \tag{2.20}$$

Um diesen Ausdruck weiter zu vereinfachen, ist folgende Taylorentwicklung

$$\left(1 - e^{\frac{2\pi i}{N} f}\right)^{\gamma-1} = \sum_{s=0}^{\infty} A_s^{-\gamma} e^{\frac{2\pi i}{N} fs} \tag{2.21}$$

mit den Koeffizienten

$$A_s^{-\gamma} \begin{cases} := 1 & \text{für } s = 0 \\ = \frac{(1-\gamma)(2-\gamma)\dots(s-\gamma)}{s!} \simeq \frac{s^{-\gamma}}{\Gamma(1-\gamma)} & \text{für } s \geq 1 \end{cases} \tag{2.22}$$

hilfreich (angegeben in [Rangarajan 2000]) und man erhält

$$\sum_{s=1}^{\infty} s^{-\gamma} e^{\frac{2\pi i}{N} fs} = \Gamma(1-\gamma) \left(\left(1 - e^{\frac{2\pi i}{N} f}\right)^{\gamma-1} - 1 \right). \tag{2.23}$$

Da die Koeffizienten in (2.22) für große s immer kleiner (also unbedeutender) werden und für (2.20) nur der Realteil von (2.23) benötigt wird, gilt

$$\operatorname{Re} \sum_{s=1}^{\infty} s^{-\gamma} e^{\frac{2\pi i}{N} fs} \simeq \sum_{s=1}^N s^{-\gamma} \cos(2\pi fs/N). \tag{2.24}$$

Im Falle von Langzeitkorrelationen tragen nur sehr niedrige Frequenzen wesentlich zum Spektrum bei, weshalb der Ausdruck in der großen Klammer um $f \simeq 0$ entwickelt werden kann. Die Entwicklung liefert (nach [Rangarajan 2000])

$$\begin{aligned}
\sum_{s=1}^N s^{-\gamma} \cos(2\pi fs/N) &\simeq \Gamma(1-\gamma) \operatorname{Re} \left(\left(1 - e^{\frac{2\pi i}{N} f}\right)^{\gamma-1} - 1 \right) \\
&= \Gamma(1-\gamma) \left((2\pi f/N)^{\gamma-1} \cos(\pi(\gamma-1)/2N) \right). \tag{2.25}
\end{aligned}$$

Für den wesentlichen Verlauf des Frequenzspektrums (2.20) gilt also

$$\begin{aligned}
P(f) &\sim 2 \cdot \Gamma(1-\gamma) \cos(\pi(\gamma-1)/2N) (2\pi f/N)^{\gamma-1} \\
&\sim f^{\gamma-1}. \tag{2.26}
\end{aligned}$$

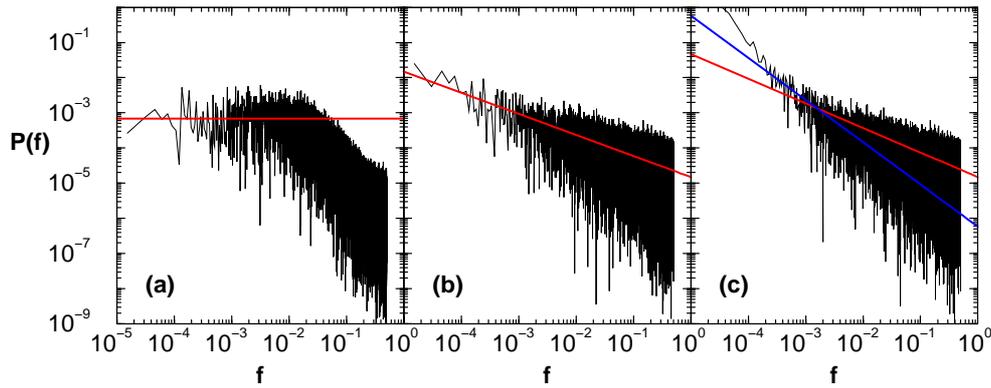


Abbildung 2.3: Beispiele zur Fourier-Analyse: (a) Frequenzspektrum $P(f)$ einer künstlichen kurzzeitkorrelierten Reihe aus 65536 Einträgen in doppelt-logarithmischer Darstellung. Man sieht ein Abflachen im Bereich niedriger Frequenzen (rote Linie), da dort die Korrelationen aufbrechen. (b) Frequenzspektrum einer künstlichen langzeitkorrelierten ($\gamma = 0.4$) Reihe gleicher Länge. Der rote Fit liefert eine Steigung von $-\beta = -0.6$, was einem $\gamma = 0.4$ entspricht. (c) Frequenzspektrum einer künstlichen langzeitkorrelierten ($\gamma = 0.4$) Reihe mit einem schwachen linearem Trend. Der rote durchgezogene Fit liefert eine Steigung von $-\beta = -0.65$. Im Bereich niedriger Frequenzen (entspricht großen Zeitskalen) könnte man sogar eine deutlich steilere Steigung vermuten (grüne Linie).

Das Potenzgesetzverhalten der Autokorrelationsfunktion $C(s)$ bei Langzeitkorrelationen mit Skalenexponent γ zeigt sich also auch in der Spektralanalyse in Form eines Potenzgesetzverhaltens des Spektrums $P(f)$ mit Skalenexponent β

$$P(f) \sim f^{-\beta} \quad \text{mit} \quad \beta = 1 - \gamma. \quad (2.27)$$

Die folgenden drei Beispiele sollen nun die Grenzen dieser Methode zur Bestimmung von Langzeitkorrelationen aufweisen.

In Abbildung 2.3(a) ist das Spektrum der **F**ast-**F**ourier-**T**ransformation (FFT) [Press 1997] für künstlich erzeugte kurzzeitkorrelierte Daten der Länge $N = 2^{16} = 65536$ zu sehen. Im Bereich niedriger Frequenzen (großer Wellenlängen) flacht es deutlich ab und wird zu einem Plateau. Der Exponent β , hier als Steigung aufgrund der doppelt-logarithmischen Auftragung, geht gegen Null, da es auf diesen Skalen keine Korrelationen mehr gibt. Abbildung 2.3(b) zeigt das Spektrum einer gleichlangen Reihe, die jedoch mit einem Korrelationsexponenten $\gamma = 0.4$ langzeitkorreliert ist. Der Fit (rote Linie) liefert den korrekten Skalenexponenten $\beta = 0.6$. Liegt dieser langzeitkorrelierten Reihe noch zusätzlich ein linearer Trend auf, wie in Abbildung 2.3(c), so sieht man, dass es im Bereich niedriger Frequenzen nicht mehr möglich ist, den korrekten Exponenten β und daraus folgend γ zu bestimmen. Die durch einen Trend überlagerte, langzeitkorrelierte Reihe ist mittels der Fourier-Transformation nicht zufriedenstellend auf Korrelationen in den Fluktuationen analysierbar.

2.2.3 Trendbereinigende Fluktuationsanalyse (DFA)

Eine Methode, die zur Bestimmung von Langzeitkorrelationen schon bei deutlich geringeren Reihenlängen sehr erfolgreich eingesetzt werden kann, ist die trendbereinigende Fluktuationsanalyse (englisch **D**etrended **F**luctuation **A**nalysis, DFA) [Peng 1994, Koscielny 1998a,

Kantelhardt 2001a]. Sie bildet eine Erweiterung der Fluktuationsanalyse (FA) [Feder 1988] und ist in [Kantelhardt 2001a] ausführlich beschrieben.

Die vorherigen Methoden haben alle das Problem, dass ihre Ergebnisse durch Trends stark beeinträchtigt werden. Es wird also eine Methode zur Korrelationsanalyse benötigt, die gegenüber Trends resistent ist. Die DFA wurde zunächst von Peng et al. in [Peng 1994] und [Buldyrev 1995] zur Analyse von Korrelationen in DNA-Sequenzen eingeführt. Im Laufe der Zeit wurde sie weiterentwickelt [Bunde 2000, Kantelhardt 2001a] und fand Anwendungen unter anderem in der Zeitreihenanalyse von Temperaturdaten [Koscielny 1996, Koscielny 1998a, Eichner 2003, Vjushin 2004], Abflussreihen [Kantelhardt 2003b, Kantelhardt 2006] und Herzrhythmen [Bunde 2000, Kantelhardt 2001b].

Die Vorgehensweise ist zunächst analog zur FA (beide ausführlich beschrieben in [Eichner 2002, Kantelhardt 2001a]). Die kumulierte Reihe (auch das *Profil* genannt)

$$Y_l = \sum_{i=1}^l x_i \quad \text{mit } l = 1, 2, \dots, N \quad \text{und} \quad Y_0 \equiv Y_L \equiv 0 \quad (2.28)$$

wird in Fenster der Breite s unterteilt. In diese Fenster werden polynomiale Approximationen (Gauss'sche Methode der kleinsten Quadrate [Zurmühl 1963]) der Ordnungen $n = 0, 1, \dots$ gelegt und die Abstandsquadrate zwischen Profil $Y_{(\nu-1) \cdot s+i}$ und Fit $p_{n,\nu}(i)$ an jedem Messpunkt im ν -ten Fenster bestimmt (Abbildung 2.4) und es gilt

$$\tilde{Y}_{(\nu-1) \cdot s+i}^2 = (Y_{(\nu-1) \cdot s+i} - p_{n,\nu}(i))^2 \quad \text{mit } i = 1, 2, \dots, s. \quad (2.29)$$

Über diese Abstandsquadrate wird gemittelt und jedem der $2N_s$ Fenster⁴ ein mittleres Fluktuationsquadrat zugeordnet

$$F_n^2(s, \nu) = \frac{1}{s} \sum_{i=1}^s \tilde{Y}_{(\nu-1) \cdot s+i}^2 \quad \text{mit } \nu = 1, 2, \dots, 2N_s. \quad (2.30)$$

Durch die polynomiale Approximation $p_{n,\nu}(i)$ wird das Profil lokal (d. h. im jeweiligen ν -ten Fenster) von polynomialen Trends bis zur Ordnung n bereinigt. Die Originalreihe wird dadurch von Trends bis zur Ordnung $n-1$ bereinigt. Trends, die keinem polynomialen Verlauf gehorchen, lassen sich mit dieser Methode nur näherungsweise bereinigen [Kantelhardt 2001a]. Nun wird das arithmetische Mittel über die Fluktuationsquadrate aller $2N_s$ Fenster der Breite s gebildet und die Wurzel daraus gezogen

$$F_n(s) = \left(\frac{1}{2N_s} \sum_{\nu=1}^{2N_s} F_n^2(s, \nu) \right)^{1/2}. \quad (2.31)$$

Man erhält für die gesamte Reihe eine mittlere Fluktuationsfunktion $F_n(s)$ zu Skalen der Fensterbreite s . Wie in der Autokorrelationsfunktion wird nun die Skala s variiert bis zu einem s_{max} , welches nicht größer als etwa ein Viertel der Gesamtlänge N der Reihe sein sollte. Auch muss ein s_{min} immer größer als die Approximationsordnung n sein, da die Fluktuationen sonst verschwinden würden aufgrund der exakten Lösbarkeit der Approximationsgleichungen.

⁴Es gibt $2N_s$ anstatt N_s Fenster, da eine Reihe auch rückwärts zerlegt werden kann, um möglichst wenig Datenpunkte an den Enden der Reihe zu verlieren.

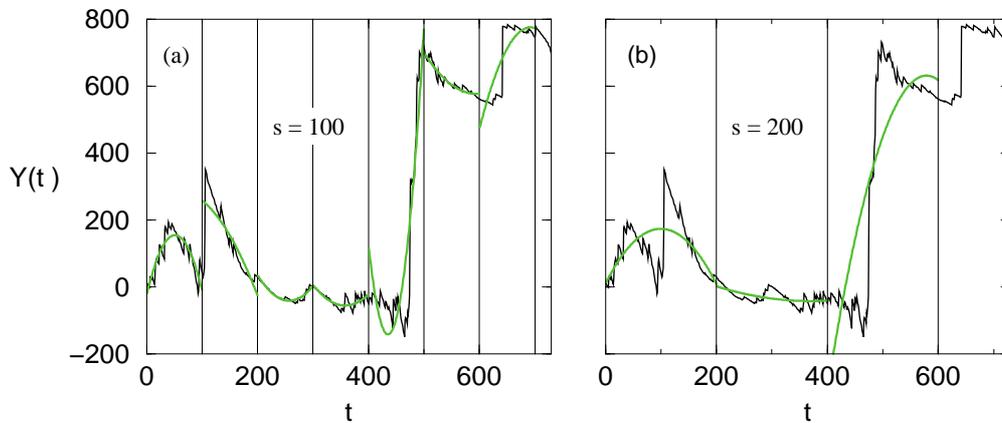


Abbildung 2.4: (a) zeigt einen Ausschnitt aus dem Profil (schwarz), welches in Fenstern der Breite $s = 100$ durch Polynome (grün) der Ordnung $n = 2$ approximiert wird. Die Differenzenquadrate dienen als Fluktuationsquadrate. In (b) ist dasselbe nochmal für die Breite $s = 200$ dargestellt. Man sieht, dass die Approximationen schlechter sind und deshalb die Fluktuationen auf größeren Skalen s zunehmen.

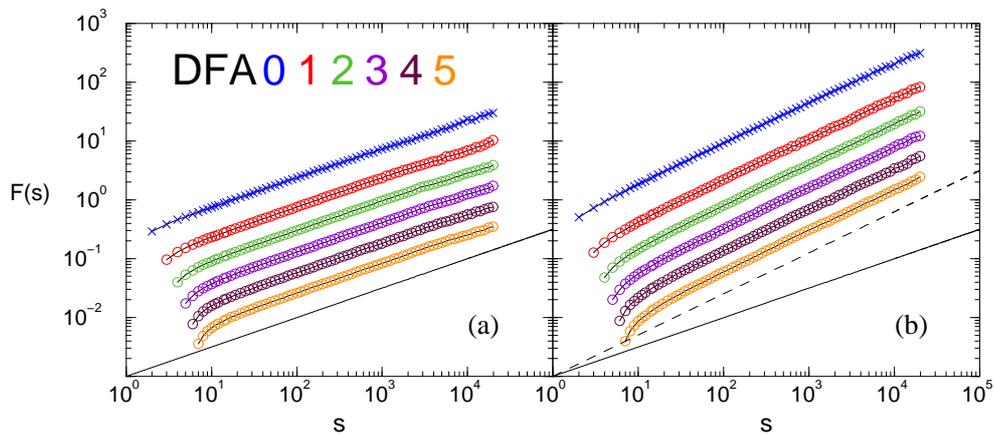


Abbildung 2.5: (a): DFA0 bis 5 einer künstlichen unkorrelierten Datenreihe. (b): DFA0 bis 5 einer mit $\gamma = 0.6$ ($\alpha = 0.7$) langzeitkorrelierten Datenreihe. Man sieht in beiden Plots die überhöhte Steigung auf kleinen Skalen, besonders bei höheren Approximationsordnungen. Die durchgezogene Linie ist eine Referenzlinie mit $\alpha = 0.5$, die gestrichelte mit $\alpha = 0.7$.

Plottet man die Ergebnisse in doppelt-logarithmischer Auftragung, so erhält man die DFA-Kurven, welche den Approximationsgraden n folgend mit DFA0, DFA1, usw. bezeichnet werden - allgemein DFA n . DFA0 ist *nicht* trendbereinigend und für stationäre Reihen vollkommen ausreichend. Aus der Verwandtschaft zur Fluktuationsfunktion der Fluktuationsanalyse FA [Hurst 1965, Feder 1988], die mit dem *Random-Walk*-Modell [Bunde 1994, Fließbach 1995] hergeleitet werden kann und in der nur die Differenzen zwischen den Anfangs- bzw. Endpunkten zweier aufeinander folgender Fenster berücksichtigt werden, folgt die Verbindung zur Autokorrelationsfunktion

$$\begin{aligned}
F_n^2(s) \sim F_{FA}^2(s) &= \langle (Y_{\nu \cdot s} - Y_{(\nu-1) \cdot s})^2 \rangle & (2.32) \\
&\stackrel{(2.28)}{=} \left\langle \left(\sum_{i=1}^s x_{i+(\nu-1) \cdot s} \right)^2 \right\rangle \\
&= \left\langle \sum_{i=1}^s x_{i+(\nu-1) \cdot s} \sum_{j=1}^s x_{j+(\nu-1) \cdot s} \right\rangle \\
&= \sum_{i=1}^s \sum_{j=1}^s \langle x_{i+(\nu-1) \cdot s} x_{j+(\nu-1) \cdot s} \rangle \\
&\stackrel{(2.16)}{\sim} \sum_{i=1}^s \sum_{j=1}^s C(j-i). & (2.33)
\end{aligned}$$

Unter der Annahme der Zeitumkehrinvarianz der Autokorrelationsfunktion, dass heisst $C(j-i) = C(i-j)$, lässt sich Gl. (2.33) weiter umschreiben zu

$$\sum_{i=1}^s \sum_{j=1}^s C(j-i) = \sum_{i=1}^s C(0) + 2 \cdot \sum_{i=1}^s \sum_{j=1}^{i-1} C(j-i). \quad (2.34)$$

Die Doppelsumme in diesem Ausdruck darf als Zweifachintegral betrachtet werden, sodass für $\gamma < 1$ die Proportionalität gilt

$$F_n^2(s) \sim \sum_{i=1}^s \sum_{j=1}^{i-1} C(j-i) \sim \sum_{i=1}^s \sum_{j=1}^{i-1} |j-i|^{-\gamma} \sim |j-i|^{2-\gamma}. \quad (2.35)$$

Für Langzeitkorrelationen findet man in der DFA also auch ein *Skalenverhalten*⁵, d. h. ein Potenzgesetzverhalten der Fluktuationsfunktion $F_n(s)$ mit einem *Skalenexponenten* α , welcher mit dem Korrelationsexponenten γ auf folgende Art verknüpft ist [Feller 1951, Peng 1994, Kantelhardt 2001a]

$$F(s) \sim s^\alpha \quad \text{mit} \quad \alpha = 1 - \frac{\gamma}{2}. \quad (2.36)$$

Ein Beispiel mit künstlichen Daten ist in Abbildung (2.5) gegeben. Für die Untersuchungen an hydrometeorologischen Zeitreihen hat sich die DFA2 als ausreichend stabil erwiesen, weshalb hier in vielen DFA-Plots nur die Kurven für DFA2 gezeigt werden. In klimatischen Zeitreihen gibt es sowohl Langzeitkorrelationen als auch explizite Kurzzeitkorrelationen (Großwetterlagen in Bereichen von ein bis zwei Wochen), die dazu führen, dass der Exponent α auf kleinen Skalen größer ist. Diskussionen über den Einfluss von Trends verschiedener Arten auf die DFA finden sich in [Kantelhardt 2001a].

⁵Skalengesetze sind fundamentale Konzepte bei der Beschreibung fraktaler Strukturen [Bunde 1994].

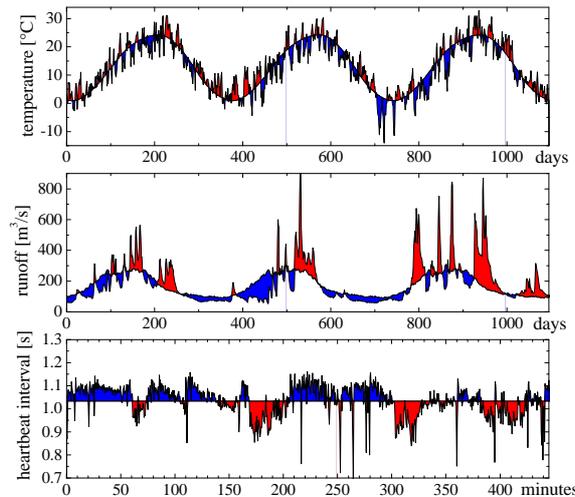


Abbildung 2.6: Drei Beispiele von langzeitkorrelierten Messreihen: typischer Verlauf einer täglich gemessenen Temperaturzeitreihe über drei Jahre (oben), ebenfalls drei Jahre einer täglich gemessene Abflusszeitreihe eines Flusses (mitte) und in sekundlicher Auflösung aufgezeichnete Herzschlagintervalle über einen Zeitraum von etwa acht Stunden (unten). Die Langzeitkorrelationen machen sich in Form der phasenhaften Abweichungen vom Mittelwert (rot: nach oben, blau: nach unten) bemerkbar.

2.3 Langzeitkorrelationen in der Natur

Langzeitkorrelationen treten in vielen Prozessen in der Natur auf. Prominente Beispiele finden sich in vielen hydrologischen Datensätzen (Pegelstände, Abflussreihen) [Mandelbrot 1969, Pelletier 1997, Kantelhardt 2003b], meteorologischen und klimatologischen Zeitreihen (tägliche und jährliche Temperaturzeitreihen, rekonstruierte und historische Temperaturzeitreihen, Meeresoberflächentemperaturen) [Koscielny 1998a, Pelletier 1999, Talkner 2000, Weber 2001, Bunde 2002a, Eichner 2003, Monetti 2003, Bunde 2005]. In der Physik begegnet man Langzeitkorrelationen in der Festkörperphysik [Schlesinger 1987] und auch in Daten von Turbulenzen [Prasad 1988, Bramwell 1998]. In der Medizin findet man Langzeitkorrelationen in Herzschlagintervallen bei verschiedenen Schlafstadien [Peng 1993, Schäfer 1998, Bunde 2000, Ashkenazy 2000, Kantelhardt 2001b, Kantelhardt 2003a], in Nukleotid-Sequenzen der DNA [Peng 1992, Peng 1994, Arneodo 1995, Buldyrev 1995] und neuronalen Signalen [Ivanov 1999]. Langzeitkorrelationen wurden auch in den Volatilitäten ökonomischer Zeitreihen auf dem Aktienmarkt gefunden [Liu 1997, Liu 1999]. Neueste Ergebnisse aus der Erdbebenanalyse zeigen, dass auch hier ein gewisses Langzeitverhalten vorkommt [Corral 2004, Livina 2005]. Abbildung 2.6 zeigt drei Beispiele für natürliche langzeitkorrelierte Messreihen: der typische Verlauf einer täglich gemessenen Temperaturzeitreihe über drei Jahre (oben), ebenfalls drei Jahre einer täglich gemessenen Abflusszeitreihe eines Flusses (Mitte) und Herzschlagintervalle über einen Zeitraum von etwa acht Stunden (unten). In den meteorologischen und hydrologischen Reihen sieht man außerdem deutlich den Jahresgang in Form eines periodischen Schwankens des Mittelwertes. Die Fluktuationen nach oben sind rot gekennzeichnet, die Fluktuationen nach unten sind blau. In rein zufälligen Prozessen sind solche langen blauen und roten Phasen nicht zu erwarten. Sie sind ein erstes Indiz für korrelierte Prozesse.

Abbildung 2.7 zeigt die Ergebnisse sehr umfangreicher DFA-Untersuchungen für vier verschie-

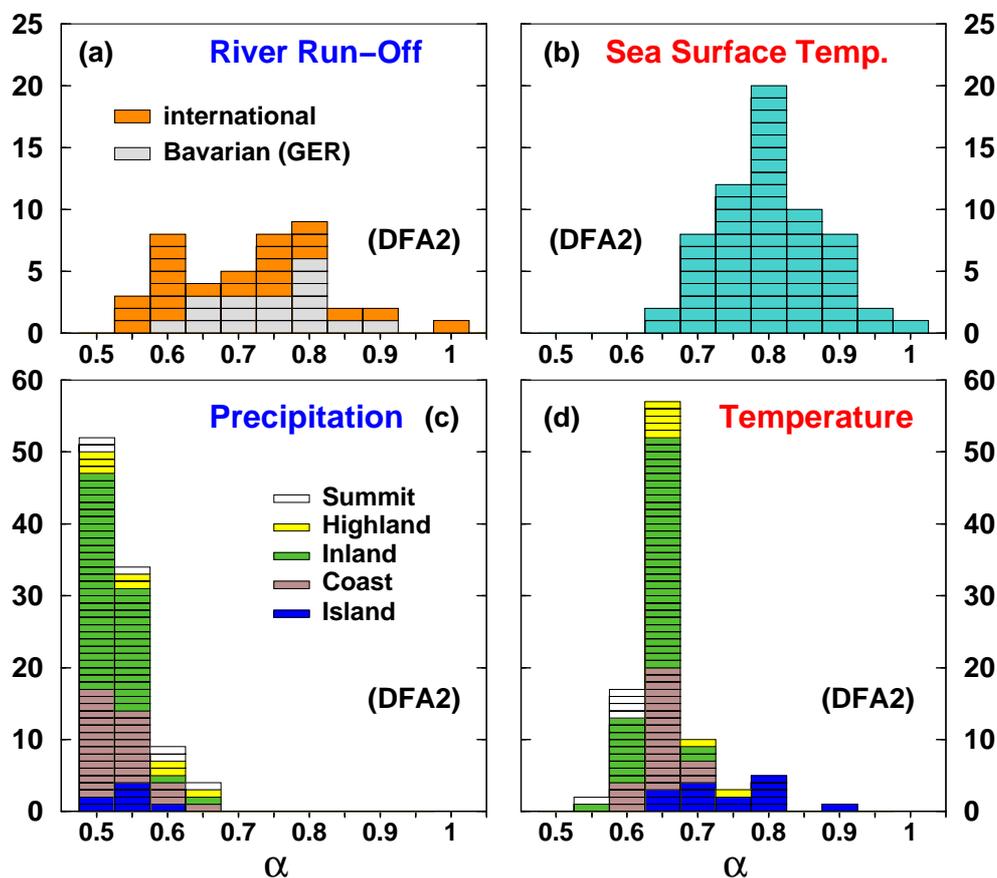


Abbildung 2.7: Vier Beispiele von DFA-Untersuchungen hydrometeorologischer Messreihen. Abbildung (a) zeigt das Histogramm der DFA2-Exponenten von Abflussreihen von 24 internationalen Flüssen und 18 bayerischen Flüssen [Kantelhardt 2006]. Das Korrelationsverhalten ist stark gestreut, von sehr schwach korreliert ($\alpha = 0.55$) bis zur Grenze des nichtstationären Verhaltens ($\alpha = 1.0$). Abbildung (b) zeigt das Histogramm der DFA2-Exponenten von 63 untersuchten Meeresoberflächentemperaturreihen [Monetti 2003]. Die Werte sind mit einem mittleren α von 0.8 sehr stark korreliert. Abbildung (c) zeigt das Histogramm von 100 Niederschlagszeitreihen weltweit verteilter Klimastationen [Rybski 2002], aufgeteilt in fünf geografische Kategorien. Täglicher Niederschlag ist nahezu unkorreliert, nur wenige Stationen zeigen einen α -Wert über 0.55. Die täglichen Lufttemperaturen an diesen Plätzen zeigen ein nahezu universelles Langzeitkorrelationsverhalten mit einem mittleren α von 0.65 [Eichner 2003]. Ausnahmen bilden kleine Inseln (blau), welche aufgrund ihrer Lage eher dem Korrelationsverhalten der Meeresoberflächentemperatur aus (b) folgen.

dene hydrometeorologische Messgrößen. Es sei an dieser Stelle nochmal auf die Verwandtschaft zwischen dem Skalenexponenten α und dem Korrelationsexponenten γ hingewiesen:

$$\alpha = 1 - \gamma/2 \quad \text{bzw.} \quad \gamma = 2 - 2\alpha \quad (2.37)$$

Abbildung 2.7(a) zeigt das Histogramm der DFA2-Exponenten von Abflussreihen von 24 internationalen Flüssen und 18 bayerischen Flüssen [Kantelhardt 2006]. Das Korrelationsverhalten ist stark gestreut, von sehr schwach korreliert ($\alpha = 0.55$) bis zur Grenze des nichtstationären Verhaltens ($\alpha = 1.0$). Abbildung (b) zeigt das Histogramm der DFA2-Exponenten von 63 untersuchten Meeresoberflächentemperaturreihen [Monetti 2003]. Die Daten werden von fixierten Bojen und Schiffen entlang ihrer festen Routen gewonnen [Parker 1994]. Die Werte sind mit einem mittleren α von 0.8 sehr stark korreliert, was auch auf die hohe Wärmekapazität des Wassers zurückzuführen ist. Abbildung (c) zeigt das Histogramm von 100 Niederschlagszeitreihen weltweit verteilter Klimastationen [Rybski 2002, Gell-Mann 2003], aufgeteilt in fünf geografische bzw. topografische Kategorien ([Diercke 1996]). Täglicher Niederschlag ist demnach nahezu unkorreliert, nur wenige Stationen zeigen einen α -Wert über 0.55. Stattdessen findet man in Niederschlagsreihen häufig ein ausgeprägtes Multiskalen-Verhalten⁶ (auch *Multifraktalität* genannt) [Feder 1988, Lovejoy 1995, Tessier 1996, Rybski 2002]. Anders verhält es sich mit den täglichen Lufttemperaturen an diesen Plätzen. Sie zeigen ein nahezu universelles Langzeitkorrelationsverhalten über mehrere Dekaden [Koscielny 1998a, Eichner 2002, Eichner 2003, Bunde 2004b] mit einem mittleren α von 0.65 ± 0.03 , unabhängig von der geografischen Lage auf dem Festland (wie im folgenden Abschnitt noch gezeigt wird). Ausnahmen bilden lediglich kleine Inseln (blau), welche aufgrund ihrer Lage eher dem Korrelationsverhalten der Meeresoberflächentemperaturen aus (b) folgen. Dieses universelle Verhalten der Temperaturen hat einen praktischen Nutzen beim Testen von Klimamodellen [Vjushin 2002, Govindan 2002, Vjushin 2004]. Zeigen die modellierten Daten kein korrektes Langzeitkorrelationsverhalten, so sind die Aussagen und Prognosen, die man aus diesen Modelldaten ableitet, insbesondere wenn es um die Beurteilung von Trends geht, sehr eingeschränkt gültig und für lang reichende Vorhersagen unbrauchbar. Dass die Abflussreihen in 2.7(b) ein so ausgeprägtes Langzeitkorrelationsverhalten zeigen, die Niederschläge in (c) jedoch nicht, ist vermutlich auf Reservoireffekte im Grundwasser und auf die (teilweise) fraktalen Strukturen der Einzugsgebiete von Flüssen zurückzuführen (siehe auch [Mandelbrot 1969, Turcotte 1992, Rodriguez-Iturbe 1997]).

In den folgenden beiden Unterabschnitten werden Langzeitkorrelationsuntersuchungen an täglichen Temperaturreihen von 72 australischen Klimastationen und 10 sibirischen und 10 nordamerikanischen tief im Inland liegenden Klimastationen exemplarisch diskutiert, unter anderem auch, weil wir Schlussfolgerungen anderer Gruppen anzweifeln [Király 2005, Fraedrich 2003, Bunde 2004b]. Außerdem werden rekonstruierte und historische jährliche Zeitreihen hydrometeorologischer Messgrößen vorgestellt, da diese in den kommenden Kapiteln über Extremwertanalysen als praktische Beispiele Verwendung finden.

2.3.1 Tächlich gemessene Temperaturzeitreihen

Das vermutlich prominenteste Beispiel für natürlich auftretende Langzeitkorrelationen findet man in den Fluktuationen täglicher Temperaturdaten [Koscielny 1998a, Eichner 2003]. In

⁶Grob gesagt gehorchen große Fluktuationen einem anderen Skalengesetz als kleine Fluktuationen, Starkniederschläge folgen also einer anderen statistischen Gesetzmäßigkeit als Nieselregen.

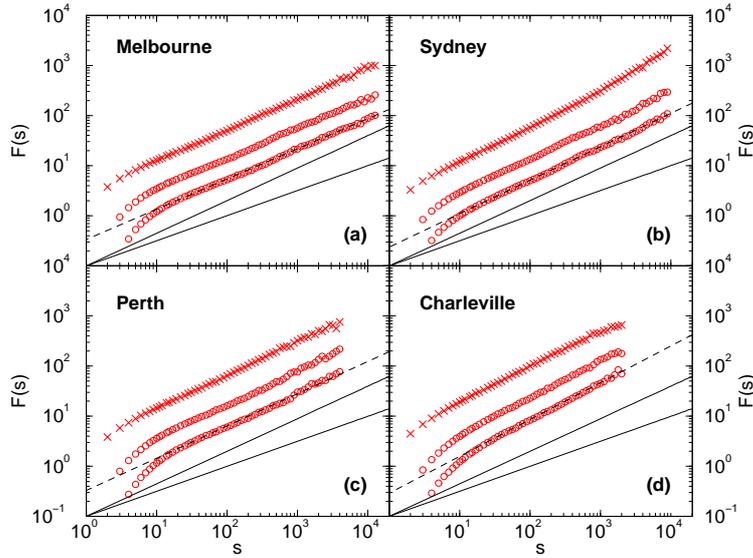


Abbildung 2.8: Vier Beispiele der DFA-Untersuchungen australischer Temperaturreihen. In Melbourne (a) ergibt die DFA2 (unterste rote Kurve) eine Steigung von $\alpha = 0.61$ (gestrichelter Fit). In Sydney (b) und Perth (c) ist der Wert etwas höher, beide $\alpha = 0.67$. Sydney zeigt in der DFA0 (oberste Kurve) einen leichten Trend, was vermutlich auf die Stadterwärmung zurückzuführen ist. Charleville (d) zeigt eine deutlich stärkere Persistenz mit $\alpha = 0.74$. Die unteren schwarzen Linien sind Hilfslinien mit den Steigungen 0.5 (unkorreliert) und 0.65 (langzeitkorreliert).

der Untersuchung täglicher Fluktuationen ist der Saisonverlauf (Jahresgang) jedoch störend. Die Saison s_i in klimatischen Zeitreihen wird im Vorfeld ermittelt und von den Messgrößen abgezogen. Dazu werden Mittelwerte aller Kalendertage gebildet, also beispielsweise ein mittlerer 1. Januar, indem über alle zur Verfügung stehenden 1. Januare gemittelt wird, dann ein mittlerer 2. Januar und so weiter. Man erhält einen mittleren Jahresgang s_i mit

$$s_i = \frac{1}{N_s(i)} \sum_{j=1}^{N_s(i)} x_{i+(j-1) \cdot 365} \quad \text{mit } i = 1, 2, \dots, 365, \quad (2.38)$$

wobei $N_s(1)$ die Anzahl der in der gesamten Reihe vorkommenden 1. Januare ist, $N_s(2)$ die Anzahl aller 2. Januare und so weiter. Aufgrund der jährlichen Periodizität gilt per Definition

$$s_i = s_{i+n \cdot 365} \quad \text{mit } n \in \mathbb{N}. \quad (2.39)$$

Nun wird von jedem Kalendertag der Messreihe der Wert des zugehörigen mittleren Kalendertages abgezogen und man erhält die saisonbereinigte Reihe der täglichen Temperaturen, die weiterhin mit x_i bezeichnet wird,

$$x_i := x_i - s_i \quad \text{mit } i = 1, 2, \dots, N. \quad (2.40)$$

In [Kiraly 2005] haben A. Kiraly und I. M. Janosi 61 australische Temperaturzeitreihen mittels DFA analysiert. Sie finden im Fluktuationsverhalten eine Breitengradabhängigkeit (Nord-Süd-Gefälle in den Fluktuationsexponenten), die dem in [Koscielny 1998a] und [Eichner 2003]

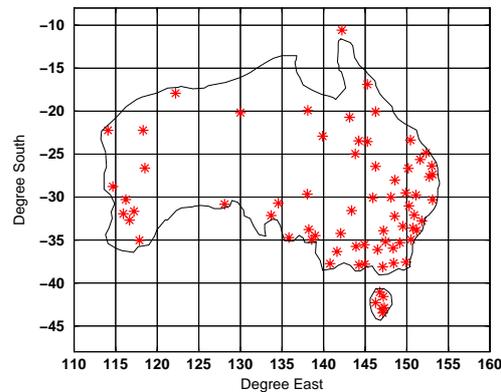


Abbildung 2.9: Die geografischen Lagen der 72 australischen Klimastationen. Die meisten Klimastationen sind in der Nähe von Städten und die meisten Städte sind offensichtlich in Küstennähe. Auffallend sind noch Thursday Island, die Insel ganz im Norden, und Tasmanien, die Inseln ganz im Süden. Das Inland ist sicherlich aufgrund der unwirtlich heißen klimatischen Verhältnisse und daher spärlichen Besiedelung kaum mit Klimastationen bestückt.

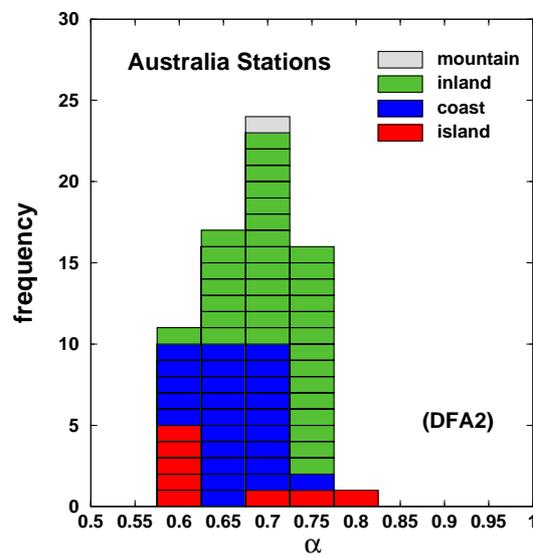


Abbildung 2.10: Histogramm der ermittelten α -Werte der 69 australischen Stationen, deren Fluktuationen ein Skalenverhalten aufweist. 3 Stationen der ursprünglich 72 Datensätze zeigen kein vernünftiges Skalenverhalten, sodass kein Exponent bestimmbar ist. Die Einteilung in vier verschiedene Klimaregionen erfolgte nach [Diercke 1996].

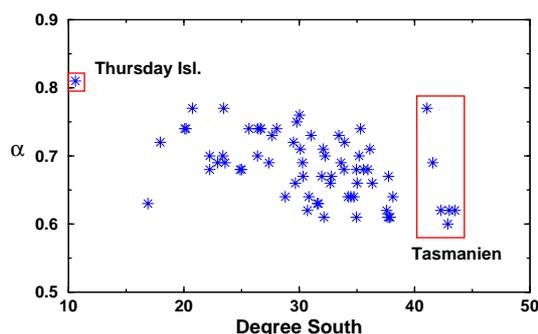


Abbildung 2.11: α -Werte aus dem Histogramm 2.10 als Funktion der südlichen Breite. (Thursday Island und Tasmanien sind rot markiert.) Es ist keine ausgeprägte Nord-Süd-Abhängigkeit der Exponenten zu erkennen. Würde man die Station links unten (Cairns, $\alpha = 0.63$) herausnehmen, würde sich eine leichte Tendenz zeigen, die aber im Rahmen jeder Varianz liegt.

gefundenen universellen Verhalten widerspricht. Da mir ein noch größerer Datensatz (72 australische Temperaturzeitreihen vom National Climate Centre, Bureau of Meteorology, Australien [BoM]) zur Verfügung stand, habe ich die Analyse von A. Kiraly und I. M. Janosi wiederholt. Hier sollen kurz die Ergebnisse meiner DFA2-Analysen als ein Beispiel für Langzeitkorrelationen in der Natur dargestellt werden.

In Abbildung 2.8 sind vier Beispiele von DFA-Kurven australischer Temperaturzeitreihen dargestellt. Gezeigt sind die DFA0-, DFA1- und DFA2-Kurven (von oben nach unten). Die durchgezogenen schwarzen Linien sind Hilfslinien mit den Steigungen 0.5 für unkorreliertes Verhalten und 0.65 für das universelle Langzeitkorrelationsverhalten bei Temperaturreihen. Sydney (b) und Perth (c) geben diese Steigung in ihrer DFA2 (gestrichelte Linien) recht gut wieder (beide $\alpha = 0.67$). Während Melbourne (a) eine etwas niedrigere Steigung ($\alpha = 0.61$) aufweist, zeigt Charleville (d) eine stärkere Persistenz ($\alpha = 0.74$). Auffällig ist ein schwacher Trend in Form des leichten Aufwärtsknicks in der DFA0 bei Sydney (eventuell auf die Stadterwärmung zurückzuführen). Großwetterlagen (z. B. stabile *Hochs*) und *Blocking*-Ereignisse [Charney 1979, Hupf 1998] führen zu starken Kurzzeitkorrelationen, sodass die DFA auf kleinen Skalen bei allen Stationen einen stärkeren Anstieg aufweist, als dies bei künstlich korrelierten Daten in Abbildung 2.5 der Fall ist.

Die Lagen aller 72 australischen Stationen sind in Abbildung 2.9 eingezeichnet. Das Histogramm in Abbildung 2.10 zeigt die Ergebnisse der ermittelten α -Werte von 69 Stationen, kategorisiert nach vier geografischen Zonen. Drei Stationen zeigen kein ausgeprägtes Skalenverhalten, sodass dort kein Exponent zu bestimmen war. Der Peak des Histogramms ist jedoch nicht so scharf ausgeprägt wie in Abbildung 2.7(d). Die α -Werte streuen trotzdem recht dicht um $\alpha = 0.65 - 0.7$. Einziger Ausreisser ist Thursday Islands, eine kleine Insel im Norden Australiens, mit $\alpha = 0.81$. Es ist aber bereits bekannt, dass Inseln ähnlich wie Meeresoberflächentemperaturen aufgrund der höheren Wärmekapazität des Wassers stärkere Persistenz zeigen [Monetti 2003]. Abbildung 2.11 zeigt die Exponenten α als Funktion der südlichen Breite. Es ist kein deutliches Nord-Süd-Gefälle im Verhalten der Exponenten zu erkennen.

In [Fraedrich 2003, Blender 2003] finden K. Fraedrich und R. Blender eine geografische Ab-

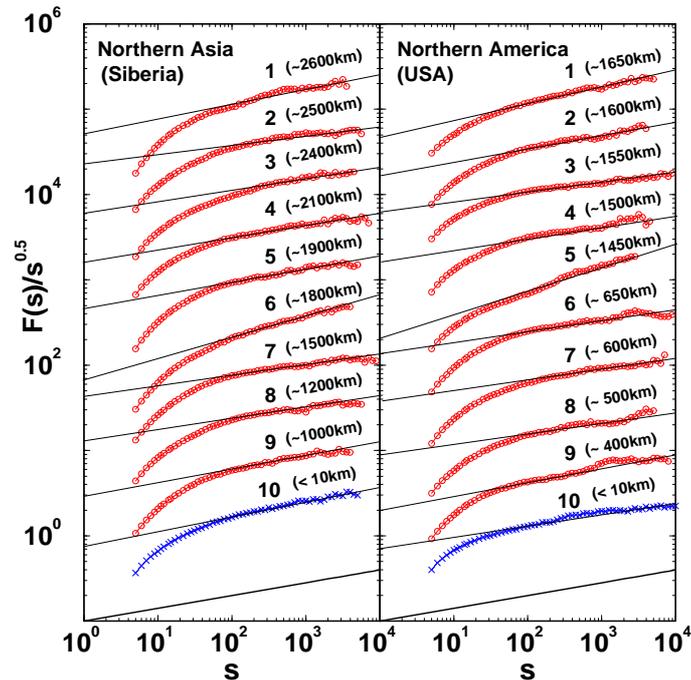


Abbildung 2.12: DFA2 Kurven von 10 sibirischen und 10 nordamerikanischen Temperaturzeitreihen mit ihren geschätzten Abständen zur nächsten Küste. Die jeweils untersten Kurven (blau) stammen von Küstenstationen, die roten Kurven sind aus dem Inland. Die Fluktuationsfunktion wurde durch $s^{0.5}$ dividiert, sodass eine Steigung von $\alpha = 0.5$ einer waagrechten Linie entspricht. Die unterste Linie hat eine Steigung von 0.15, was einem $\alpha = 0.65$ entspricht. Die Stationen und ihre Reihenlängen sind: Sibirien: 1. Urumchi (14600d), 2. Tomsk (43272d), 3. Atbasar (23298d), 4. Chita (36653d), 5. Olekminsk (33493d), 6. Horog (29004d), 7. Swerdlowsk (42705d), 8. Surgut (35438d), 9. Jakutsk (39573d) und 10. Aleksandrowsk (35862d). USA: 1. Huron (20075d), 2. Academy (34643d), 3. Cheyenne (41900d), 4. Gothenburg (36261d), 5. Gunnison (34541d), 6. Spokane (20805d), 7. Winnemucca (41900d), 8. Pendleton (38690d), 9. Tuscon (35405d) und 10. New York (38690d). Daten erhalten vom Potsdamer Institut für Klimafolgenforschung [PIK]. Abbildung nach [Bunde 2004b].

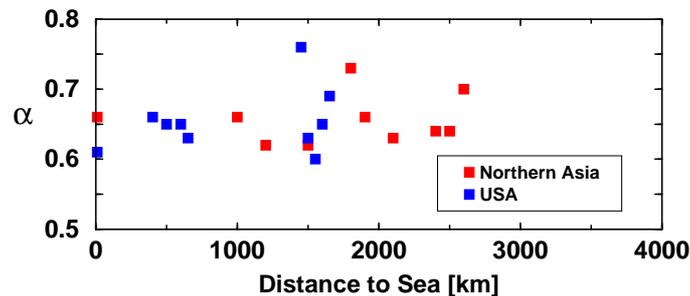


Abbildung 2.13: Skalensexponent α als Funktion des Abstands vom Meer derselben Stationen wie in Abbildung 2.12. Die Abbildung zeigt eindeutig, dass das Korrelationsverhalten für innerkontinentale Stationen nicht auf $\alpha = 0.5$ abfällt und steht damit im Widerspruch zu einer Behauptung in [Fraedrich 2003]. Das Langzeitkorrelationsverhalten mit $\alpha = 0.65 \pm 0.03$ ist auf dem Festland ein universelles Verhalten. Abbildung nach [Bunde 2004b].

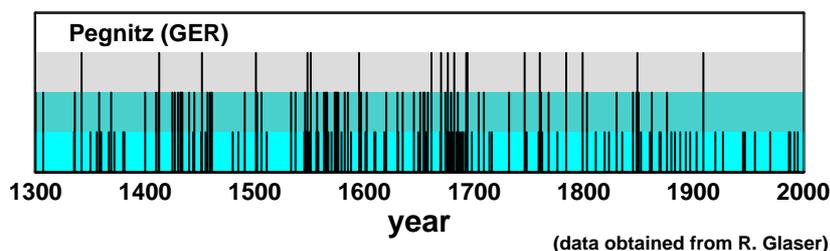


Abbildung 2.14: Historische Hochwässer der Pegnitz bei Nürnberg, von R. Glaser [Glaser 2001] in drei Kategorien eingeteilt. Man kann dieser Reihe ansehen, dass es offensichtlich Epochen (z. B. um 1700 herum) gab, in denen binnen weniger Jahre viele Hochwässer eintraten. Ebenso sieht man in der Reihe auch Phasen, in denen kaum Hochwässer notiert wurden (z. B. vor und nach 1400).

hängigkeit der Langzeitkorrelationen. Stationen, die weit im Inland liegen, zeigen ihrer Meinung nach systematisch niedrigere Skalenexponenten α (sogar bis 0.5, also unkorreliert), als solche in Küstennähe, wo sie ebenfalls $\alpha = 0.65$ finden. Diese geografische Abhängigkeit können wir in unseren Datensätzen nicht sehen [Eichner 2003, Bunde 2004b]. Abbildung 2.12 zeigt die DFA2-Kurven von 10 sibirischen und 10 nordamerikanischen Klimastationen zusammen mit ihren gerundeten Abständen zum nächsten Küstenstreifen. Die roten Kurven stehen für Inland-Stationen, die weit von der Küste entfernt sind. Die blauen Kurven repräsentieren zwei Küstenstationen, die $\alpha = 0.6$ bzw. 0.66 zeigen. Das Skalenverhalten wird für große Skalen bei fast allen Reihen schlechter, was auf die mangelnde Statistik in diesen Skalenbereichen zurückzuführen ist. In Abbildung 2.13 sind die α -Werte als Funktion des Küstenabstands gezeigt. Man sieht hier deutlich, dass es keine systematischen, geografisch bestimmbareren Abweichungen vom mittleren Wert $\alpha = 0.65$ gibt. Die Fluktuationen um diesen Wert erscheinen eher zufällig. Die von K. Fraedrich und R. Blender verwendete sibirische Temperaturreihe stand uns leider nicht zur Verfügung. Trotzdem meinen wir, mit unseren Analysen eine geografische Abhängigkeit von α ausschließen zu können und den universellen Charakter von $\alpha = 0.65 \pm 0.03$ für das Festland zu untermauern [Bunde 2004b].

2.3.2 Rekonstruierte und historische Zeitreihen

Um *saisonale Effekte* zu vermeiden, die insbesondere bei Abflusszeitreihen sehr unterschiedlich ausfallen, sodass sie nicht so einfach zu bereinigen sind wie es bei Temperaturzeitreihen der Fall ist, verwendet man in der angewandten Extremwertstatistik meist Jahresdaten. Da die Statistik gemessener Reihen mit zufriedenstellender Qualität oft nur wenige Jahrzehnte zurückreicht, greift man auf rekonstruierte, modellierte und historische Daten zurück. Die Rekonstruktion liefert dabei meist Jahresextrema oder Jahresmittelwerte, je nach Messgröße und Rekonstruktionsmethode. Historische und auch viele rekonstruierte Zeitreihen haben außerdem den Vorteil, *anthropogene Einflüsse* wie zum Beispiel den Treibhauseffekt, Flussbegradigungen oder Staustufen auszuschließen. Daher sind Trends in diesen Reihen unwahrscheinlicher und Korrelationen kommen deutlicher zum Tragen.

Methoden der Rekonstruktion von Niederschlags-, Abfluss- und auch Temperaturzeitreihen findet man u. a. in [Jones 1998, Briffa 2000, Esper 2002, Noren 2002, Soon 2003, Moberg 2005] und in den dortigen Referenzen. Drei Methoden seien hier kurz erwähnt. Über die Dicke von

Baumringen kann man auf die Feuchtigkeit eines Jahres Rückschlüsse ziehen und so Angaben über den Jahresniederschlag machen. Aus Sedimenten lassen sich Rückschlüsse auf Hochwässer und Niedrigwässer ziehen. Lufteinschlüsse in Eisbohrkernen liefern Information über atmosphärische Zusammensetzungen. Meistens basieren rekonstruierte hydrometeorologische Zeitreihen auf Kombinationen verschiedener Rekonstruktionsmethoden.

Historische Zeitreihen bestehen (sofern keine richtige Buchführung existiert) meist aus extremen Wetterereignissen und Katastrophen [Pfister 1999, Glaser 2001, Christensen 2002, Glaser 2003, Jacobeit 2003, Mudelsee 2003, Mudelsee 2004], da es Ereignisse sein mussten, die sozusagen *erwähnenswert* waren, wie zum Beispiel Überflutungen. Teilweise wird auch auf indirekte Informationen zurückgegriffen. So kann man aus Hungersnöten oft auf ein sehr trockenes Jahr schließen. Abbildung 2.14 zeigt die historische Zeitreihe der Pegnitz bei Nürnberg, wie sie in [Glaser 2001] zu finden ist. Man kann an dieser Reihe bereits erkennen, dass es ganz offensichtlich Epochen gehäufter Hochwässer und ebenso Epochen mit geringen Hochwässern gab. Korrelationsanalysen ergeben für diese Reihe einen α -Wert von $\alpha = 0.63 \pm 0.05$.

Um meine numerischen Resultate zur Statistik der Wiederkehrintervalle (Kapitel 3) und der Maximastatistik (Kapitel 4) mit Beispielen echter Daten (in diesem Fall rekonstruierte und historische Datensätze) zu untermauern, standen mir folgende fünf Zeitreihen zur Verfügung. Sie sind in Abbildung 2.15 zusammen mit ihren Histogrammen dargestellt. Die α -Werte kann man Abbildung 2.16 entnehmen.

- Rekonstruierte Abflussreihe des **Sacramento River** (USA) über 1109 Jahre [Meko 2001]. Der Fluktuationsexponent der DFA-Analyse liefert $\alpha = 0.6$, was über $\gamma = 2 - 2\alpha$ einem Korrelationsexponenten $\gamma = 0.8$ entspricht.
- Rekonstruierte Temperaturreihe der **Baffin Island** (CA) über 1241 Jahre [Moore 2001] mit $\alpha = 0.73$ ($\gamma = 0.55$).
- Rekonstruierte Jahresniederschlagsreihe von **New Mexico** (USA) über 2129 Jahre [Grissino-Meyer 1996] mit $\alpha = 0.8$ ($\gamma = 0.4$).
- Historische Niedrigwasserstände des **Nil** bei Roda nahe Kairo (EGY) über 663 Jahre [Tousson 1925, Beran 1994, Whitcher 2002] mit $\alpha = 0.85$ ($\gamma = 0.3$).
- Rekonstruierte Temperaturreihe der **nördlichen Hemisphäre** nach **M. E. Mann** [Mann 1999] über 981 Jahre mit $\alpha = 0.95$ ($\gamma = 0.1$).

Die Zeitreihen Baffin, Sacramento und New Mexico werden von der NOAA Paleoclimatological Data Bank [NOAA] verwaltet und sind unter <http://www.ngdc.noaa.gov/paleo/recons.html> in Internet zugänglich. Die Mann-Reihe der nördlichen Hemisphäre, die in letzter Zeit auch als "Hockey-Schläger" Kurve bekannt wurde, da in ihr die Erderwärmung der letzten Dekaden im Verhältnis recht stark ausfällt, kann man ebenfalls dort über das Internet bekommen. Die historische Nil-Reihe, die schon R. E. Hurst zur Planung des Assuan-Staudamms verwendete [Hurst 1951, Feder 1988], gibt es unter <http://sunsite.univie.ac.at/statlib/S/beran>. Zu den hier verwendeten historischen und rekonstruierten Reihen findet man weitere Angaben im Netz bei den oben genannten URLs und in den dortigen Literaturhinweisen.

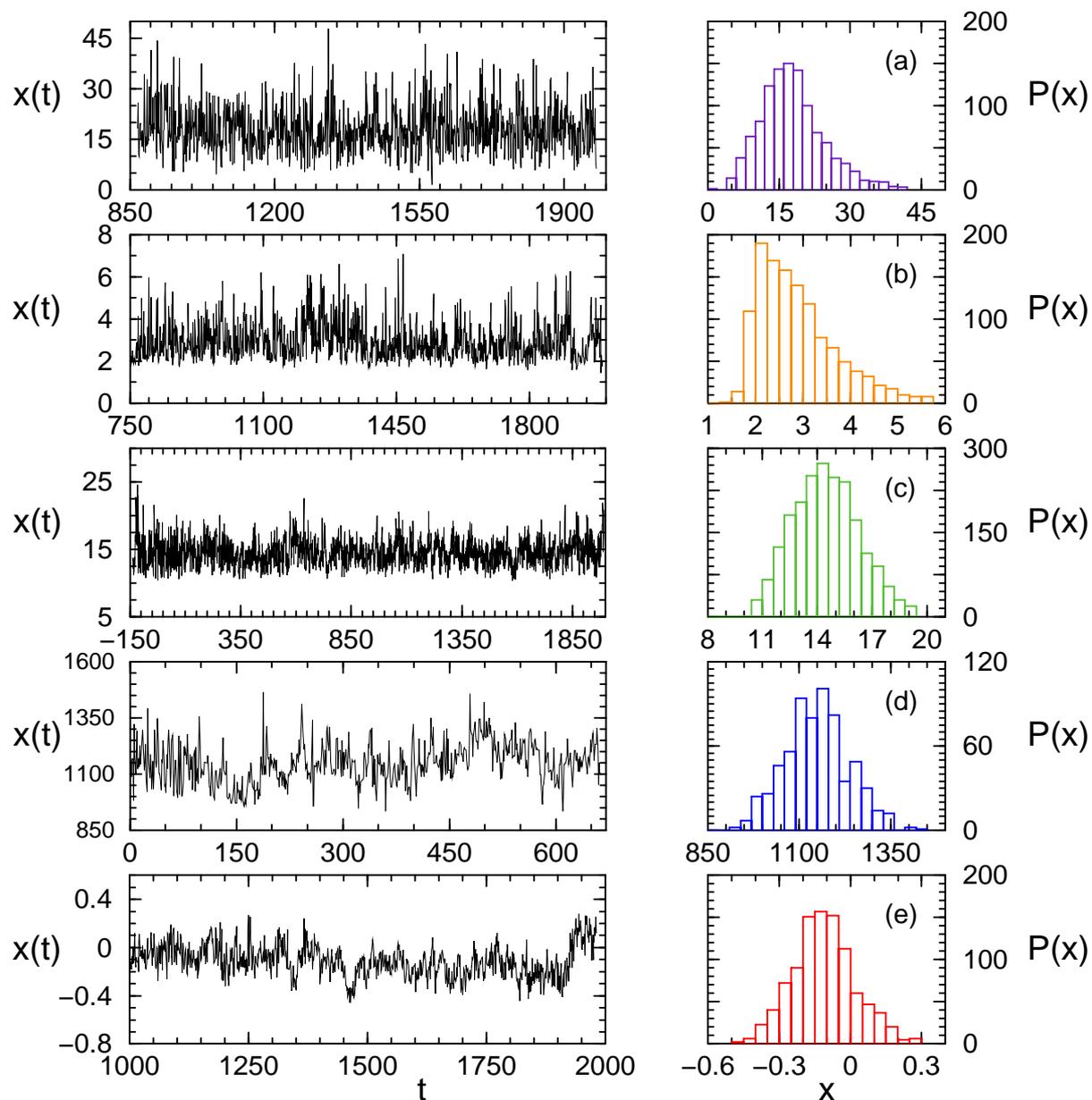


Abbildung 2.15: Fünf rekonstruierte und historische Jahresdatenreihen, links die Reihe, rechts das zugehörige Histogramm der Messwerte. Die Angaben zu den Reihen Sacramento (a), Baffin Island (b), New Mexico (c), Nil (d) und Mann-Reihe (e) befinden sich im Text. Rekonstruktionen von Niederschlags-, Abfluss- und Temperaturreihen erfolgen u. a. über Baumringe, Sedimente und Eisbohrkerne. Man kann am Erscheinungsbild der Messreihen (linke Spalte) schon erahnen, dass die Korrelationen von oben nach unten zunehmen.

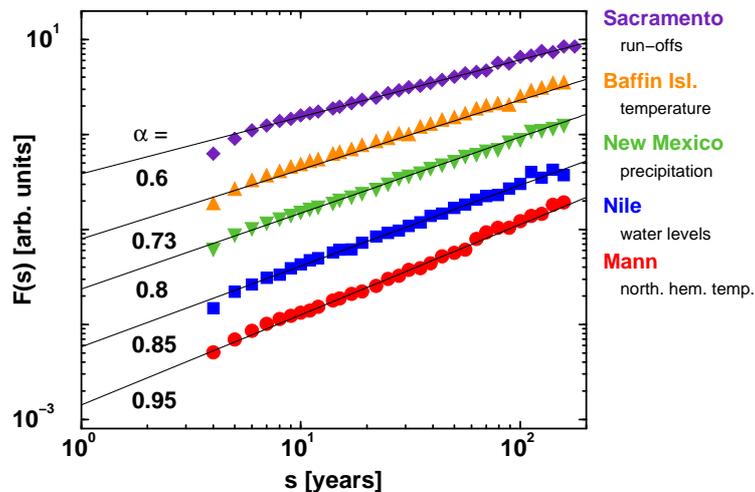


Abbildung 2.16: Die Abbildung zeigt die Fluktationskurven $F(s)$ der DFA2-Analyse historischer und rekonstruierter Jahresdaten mit den daraus ermittelten α -Werten. Angaben zu den Reihen (Sacramento \diamond , Baffin Island \triangle , New Mexico ∇ , Nil \square und Mann-Reihe \circ) befinden sich im Text. Alle Reihen zeigen langzeitkorreliertes Verhalten, welches in doppellogarithmischer Auftragung in Form einer Geraden mit Steigung α erscheint. Die Auftragung von $F(s)$ ist in beliebigen Einheiten (*arb. units* = *arbitrary units*), da nur die Steigungen α der Kurven von Bedeutung sind. Sie sind über $\alpha = 1 - \gamma/2$ mit dem Korrelationsexponenten γ verknüpft. Abbildung nach Bunde05.

Dass es sich bei diesen Reihen um langzeitkorrelierte Reihen handelt, zeigt die DFA-Untersuchung in Abbildung 2.16. Es wurde die oben erklärte DFA2 auf die Datensätze angewendet. Die schwarzen durchgezogenen Linien zeigen die asymptotischen Fits. Ihr potenzgesetzartiger Anstieg (*Skalenverhalten*) ist ein Indikator für Langzeitkorrelationen. Die Steigungen der Kurven liegen alle über 0.5, dem α -Wert für unkorreliertes Verhalten.

2.4 Erzeugung langzeitkorrelierter Datensätze

Viele statistische Eigenschaften kommen erst bei sehr großen Stichprobenumfängen bzw. Reihenlängen zum Vorschein. Insbesondere gilt dies für Langzeitkorrelationen. Jedoch sind genau die interessierenden Eigenschaften in echten Datensätzen oft mit anderen Eigenschaften (z. B. Trends) überlagert. Um den reinen Korrelationseffekt studieren zu können, ist es notwendig, Datensätze zur Verfügung zu haben, deren statistische Eigenschaften (Verteilungsfunktion, Korrelationsstruktur, Trendverhalten, etc.) man exakt kennt, und deren Datensatzlängen man außerdem beliebig regeln kann.

In meinem Fall sind Datensätze mit bekannter Korrelationsfunktion notwendig. Zur Erzeugung von langzeitkorrelierten Zufallszahlen stehen verschiedene Methoden zur Verfügung. Eine sehr bewährte ist die *Fourier-Filtering* Methode [Mandelbrot 1971, Saupe 1988, Peng 1991, Prakash 1992, Makse 1996]. Sie ist jedoch auf Gauss-verteilte Zufallszahlen beschränkt. Eine Erweiterung auf langzeitkorrelierte Zufallszahlen mit faktisch beliebiger Verteilungsfunktion ist ein *Iterationsverfahren* von T. Schreiber und A. Schmitz [Schreiber 1996, Schreiber 2000,

Theiler 1992] (hier *Schreiber-Verfahren* genannt), welches aber auf die *Fourier-Filtering* Methode zurückgreift. Beide werden in den folgenden Unterabschnitten erklärt. Methoden wie das FARIMA-Modell (*Fractional Auto-Regressive Integrated Moving Average*) [Brockwell 1996, Cohn 2005] oder das *Klemes-Modell* [Klemes 1974] wurden in den hier dargestellten Analysen nicht verwendet. FARIMA-Modelle werden häufig zur Simulation echter Datenreihen verwendet, da sie viele Parameter bieten, mit denen man die statistischen Eigenschaften steuern kann. Die Fourier-Filtering Methode hingegen erlaubt es, mit nur einem Parameter (β bzw. γ) die für uns wichtigen Korrelationseigenschaften zu erzeugen.

2.4.1 Fourier-Filtering Methode

Die **Fourier-Filtering Methode** (FFM) zur Erzeugung langzeitkorrelierter Zufallszahlen basiert, wie der Name verrät, auf der Fourier-Transformation und dem Power-Spektrum (siehe Kapitel 2.2.2) [Mandelbrot 1971, Saupe 1988, Prakash 1992, Makse 1996, Press 1997]. Ein Signal oder eine Zeitreihe (x_i) mit den beiden Koordinaten x (*Ort*) und i (*Zeit*) wird durch die diskrete Fourier-Transformation [Brockwell 1996, Press 1997] in ein Signal mit den Koordinaten F (*Amplitude*) und $2\pi f/N$ (*Frequenz*) bzw. f umgeformt (siehe Gl. (2.17)),

$$F(f) = \frac{1}{\sqrt{N}} \sum_{j=1}^N x_j e^{-\frac{2\pi i}{N} f j}.$$

Das Produkt der *Fourier-Transformierten* mit ihrer komplex-konjugierten, $P(f) = F(f)F^*(f)$ bezeichnet man als *Frequenz-Spektrum* oder *Power-Spektrum*. Handelt es sich um ein unkorreliertes Signal (“White Noise”), so zeigt das Spektrum keine charakteristische f -Abhängigkeit und $P(f)$ in Gleichung (2.18) ist ein Plateau im Frequenzraum (siehe langwelliger Bereich in Abbildung 2.3(a)). Langzeitkorrelierte Signale zeigen in $P(f)$ einen Potenzgesetzverlauf (siehe Abbildung 2.3(b)), d. h. $P(f) \sim f^{-\beta}$ mit $0 < \beta < 1$ ($\gamma = 1 - \beta$).

Da die diskrete Fourier-Transformation eine eineindeutige Abbildung des Signals vom Ortsraum in den Frequenzraum ist, gibt es zu dieser Abbildung auch eine Umkehrfunktion, die *Rücktransformation* [Brockwell 1996, Press 1997],

$$x_j = \frac{1}{\sqrt{N}} \sum_{f=1}^N F(f) e^{\frac{2\pi i}{N} f j}. \quad (2.41)$$

Nun kann man ein stationäres, unkorreliertes Signal oder eine stationäre, unkorrelierte Reihe von Zufallszahlen Fourier-transformieren und das Spektrum $P(f)$ mit dem eines langzeitkorrelierten Signals $P_{neu}(f) \sim f^{-\beta}$ mit gewünschtem β ersetzen, indem man die Fourier-Transformierte $F(f)$ mit $\sqrt{P_{neu}(f)}$ durchmultipliziert,

$$F_{neu}(f) = F(f) \sqrt{P_{neu}(f)} \quad \text{für } f = 1, 2, \dots, N. \quad (2.42)$$

Die Rücktransformation liefert nun Gauss-verteilte langzeitkorrelierte Zufallswerte. Abbildung 2.17 skizziert noch einmal die Vorgehensweise in der FFM. Dass dies nur für Gauss-verteilte Zufallszahlen gilt, lässt sich an Gleichung (2.17) daran motivieren, dass sie eine gewichtete Summe der Signal-Werte x_i darstellt und daher nach dem Zentralen Grenzwertsatz [Sachs 1974, Papoulis 1984] die Amplituden $F(f)$ wieder einer Gauss-Verteilung gehorchen.

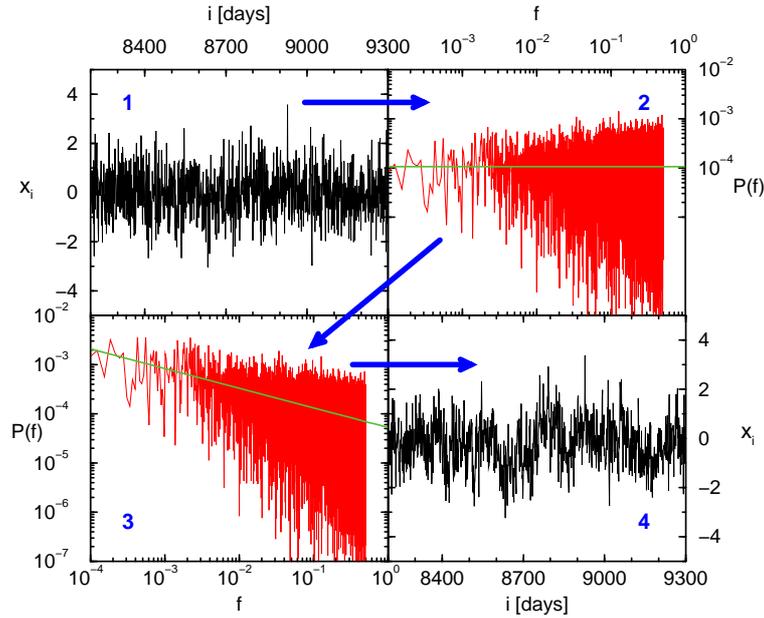


Abbildung 2.17: Skizze der Vorgehensweise in der Fourier-Filtering Methode: 1. Man erzeugt sich unkorrelierte (Gauss-verteilte) Zufallszahlen. 2. Das Frequenzspektrum $P(f)$ ergibt dann ein Plateau. 3. Das Plateau wird durch ein Potenzgesetz-artig abfallendes Spektrum mit Exponent β ersetzt. Dies kann durch Multiplikation des plateauförmigen $P(f)$ mit einem Potenzgesetz der Form $f^{-\beta}$ geschehen. 4. Die Rücktransformation liefert Gauss-verteilte langzeitkorrelierter Daten mit $\gamma = 1 - \beta$.

Dasselbe gilt nun auch für die Rücktransformation in Gleichung (2.41). Waren die ursprünglichen Daten x_i nun *nicht* Gauss-verteilt, so müssen nach dem Zentralen Grenzwertsatz die rücktransformierten Daten einer (annähernden) Gauss-Verteilung gehorchen. Die Methode der Fourier-Transformation und Rücktransformation eignet sich also nur zur Erzeugung langreichweitig korrelierter Gauss-verteilter Zufallszahlen.

Ein Nachteil dieser Methode ist, dass effektiv nur ein kleiner Bruchteil ($< 1\%$) des gesamten Datensatzes zufriedenstellend langzeitkorreliert ist [Peng 1991, Prakash 1992]. Das heisst, die Autokorrelationsfunktion bricht ihr Potenzgesetzverhalten ab einer gewissen Skala s ab (*cutoff*). H. Makse et al. schlagen in [Makse 1996] deswegen eine Modifikation der FFM vor, die die *gesamte* Reihe langzeitkorreliert erscheinen lässt. Dabei wird zuerst die Autokorrelationsfunktion für Langzeitkorrelationen $C(s) \sim s^{-\gamma}$ durch einen Ausdruck ersetzt, der auch bei $s = 0$ definiert ist, aber für große s dasselbe Verhalten zeigt,

$$C(s) := (1 + s^2)^{-\gamma/2}. \quad (2.43)$$

Dieser neue Ausdruck für Langzeitkorrelationen wird nun auf einem Intervall $[-N/2, \dots, N/2]$ definiert und es werden periodische Randbedingungen $C(s) = C(s + N)$ eingeführt. Makse et al. bestimmen analytisch das zugehörige Fourier-Spektrum $P_{neu}(f)$ für Gl. (2.42). Das neue Spektrum hat dasselbe asymptotische Verhalten für kleine Frequenzen f (d. h. lange Wellenlängen), aber einen abrupten *cutoff* bei großen Frequenzen (d. h. kurze Wellenlängen) [Makse 1996]. Dieser *cutoff* verhindert Aliasing-Fehler und dadurch den *cutoff* in den Autokorrelationen nach der Rücktransformation [Makse 1996, Press 1997]. Für die Praxis empfehlen

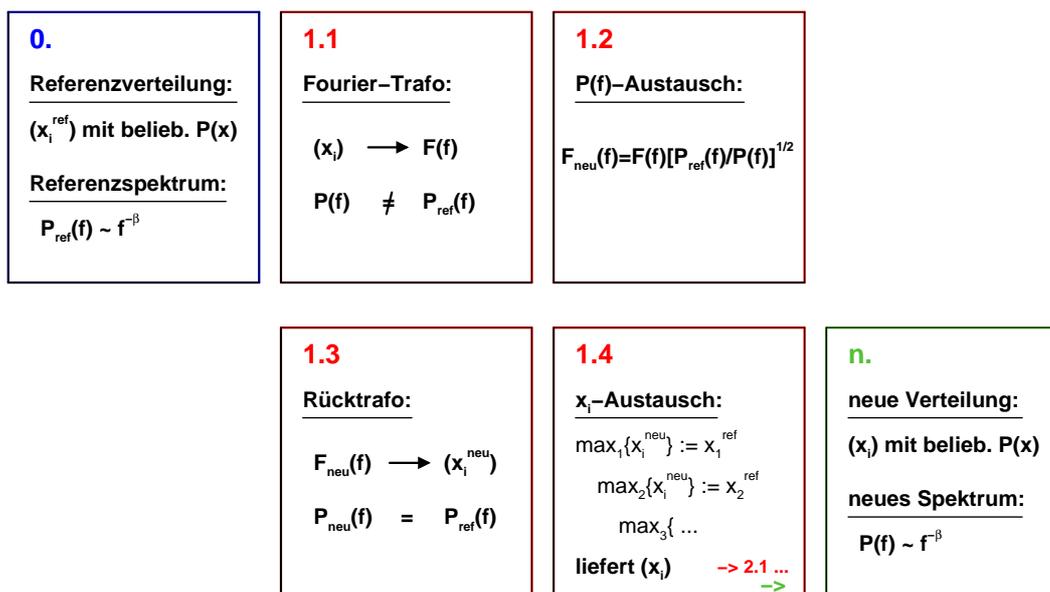


Abbildung 2.18: Ablauf der einzelnen Iterationsschritte des Schreiber-Verfahrens zur Erzeugung langzeitkorrelierter Zufallszahlen mit vorgegebener Verteilungsfunktion. (0.) Man erzeugt zunächst ein Referenzspektrum $P_{ref}(f) \sim f^{-\beta}$ aus langzeitkorrelierten Gauss-Zahlen (z. B. mittels der FFM). Außerdem erzeugt man unkorrelierte Zufallszahlen (x_i) mit der gewünschten Verteilungsdichte $P(x)$, ordnet diese der Größe nach und erhält eine Referenzreihe (x_i^{ref}) . (1.1) Als nächstes nimmt man die unkorrelierte Reihe (x_i) , Fourier-transformiert diese und bestimmt das Spektrum $P(f)$. Dieses Spektrum wird im Schritt (1.2) durch $P_{ref}(f)$ ausgetauscht, indem $F_{neu}(f) = F(f) \sqrt{P_{ref}(f)/P(f)}$ gebildet wird. Die neue Fourier-Transformierte wird in (1.3) rücktransformiert. Die so erhaltene Reihe (x_i^{neu}) besitzt das Referenzspektrum, hat aber jedoch nicht die gewünschte Verteilungsdichte $P(x)$. Diese wird in (1.4) wieder hergestellt, indem mittels *rank ordering* die Werte von (x_i^{neu}) der Größe nach mit den Werten von (x_i^{ref}) ausgetauscht werden. Die neu erhaltene Reihe (wieder (x_i) genannt) hat die gewünschte Verteilungsdichte $P(x)$, aber das Spektrum $P(f)$ ist aufgrund des Wertaustauschs wieder etwas verfälscht worden. Wiederholt man mit dieser neuen Reihe (x_i) die Schritte 1.1 bis 1.4, nähert sich das Spektrum dem Referenzspektrum an. Man kann diese Prozedur nun so oft durchführen (n.), bis das Spektrum der neuen Reihe (x_i) die gewünschte Qualität hat, und dann abbrechen.

Makse et al. die *Fast-Fourier-Transformation* [Press 1997], um die erwähnten Schritte numerisch durchzuführen. Die periodischen Randbedingungen verursachen dieselbe Periodizität in den erzeugten langzeitkorrelierten Daten. Um Reihen mit offenen Randbedingungen zu erzeugen, empfiehlt es sich eine langzeitkorrelierte Reihe mit doppelter Länge $2N$ zu generieren und in der Mitte bei N zu unterteilen.

2.4.2 Iterative Methode zur Erhaltung der Ausgangsverteilung (Schreiber-Verfahren)

Wie bereits erwähnt, hat die Fourier-Filtering Methode zur Erzeugung langzeitkorrelierter Zufallszahlen den Nachteil, dass sie nur Gauss-verteilte Zufallszahlen produzieren kann. Ein Verfahren, mit dem man korrelierte Zufallszahlen beliebiger Ausgangsverteilung und sogar mit beliebigem Power-Spektrum erzeugen kann, ist das iterative Verfahren von T. Schreiber

und A. Schmitz [Schreiber 1996, Schreiber 2000], auch *Schreiber-Test* oder *Improved Surrogate Data Test* genannt. Es basiert auf einer iterativen Fourier-Filtering Methode, die in jedem Iterationsschritt abwechselnd im Ortsraum mittels Rangordnung (*rank ordering*) die Verteilung der Daten korrigiert und im Frequenzraum das Frequenzspektrum anpasst. Ich habe in meiner Arbeit dieses Verfahren verwendet, um zusätzlich zu den Gauss-verteilten Zufallszahlen auch exponentiell verteilte, Log-Normal-verteilte und Potenzgesetz-verteilte langzeitkorrelierte Zufallszahlen zu erhalten.

Zuerst erzeugt man sich eine Reihe Gauss-verteilter Zufallszahlen der Länge N . Diese werden mittels der Fourier-Filtering Methode (Kapitel 2.4.1) zu einem beliebigen vorgegebenen γ langzeitkorreliert. Das aus der Fourier-Transformierten erhaltene Power-Spektrum wird als Referenzspektrum $P_{ref}(f)$ gespeichert. Als nächstes erzeugt man sich eine gleichlange Reihe (x_i) unkorrelierter Zufallszahlen mit der gewünschten Verteilung, z. B. der Exponentialverteilung. Diese Reihe wird mittels eines Sortieralgorithmus, z. B. dem *Quicksort*-Algorithmus [Knuth 1997], der Größe nach sortiert (oft als Rangordnung oder *rank ordering* bezeichnet) und als Referenzreihe (x_i^{ref}) abgespeichert. Man hat nun ein Referenzspektrum für die gewünschten Korrelationen und eine geordnete Referenzreihe mit der gewünschten Verteilung.

Mittels Gl. (2.17) erzeugt man aus der Reihe (x_i) die Fourier-Transformierte $F(f)$ und bestimmt das Power-Spektrum $P(f)$. Dieses Spektrum ist plateauförmig, da die Reihe unkorreliert ist. Nun wird das Spektrum gegen das Referenzspektrum ausgetauscht, indem

$$F_{neu}(f) = F(f) \frac{\sqrt{P_{ref}(f)}}{\sqrt{P(f)}} \sim f^{-\beta/2} \quad (2.44)$$

für alle Frequenzen f gebildet wird. Diese neue Fourier-Transformierte wird nun mittels Gl. (2.41) rücktransformiert. Die so erhaltene Reihe (x_i^{neu}) hat das gewünschte Spektrum, ist aber aufgrund des Zentralen Grenzwertsatzes eher wieder Gauss-förmig verteilt.

Um die gewünschte Verteilung wieder herzustellen, müssen die x^{neu} -Werte gegen die x^{ref} -Werte ausgetauscht werden. An die Position des größten Wertes der x^{neu} kommt der größte Wert der x^{ref} , an die Position des zweitgrößten Wertes der x^{neu} kommt der zweitgrößte Wert der x^{ref} und so weiter⁷. Mit der so erhaltenen Reihe kann die Reihe der (x_i) überschrieben werden. Sie hat nun die gewünschte Verteilung und ist auch deutlich korreliert. Aber aufgrund des Austauschs ist die Potenzgesetz-förmige Korrelationsstruktur wieder etwas abgeschwächt worden.

Der Iterationsprozess schreibt nun vor, dass diese neue Reihe (x_i) wieder Fourier-transformiert wird und die Schritte des Spektromaustauschs, der Rücktransformation und des Wertaustauschs wiederholt werden. Abbildung 2.19 beschreibt in einem Diagramm noch einmal die einzelnen Schritte, die in einer Iteration durchgeführt werden. Die Reihe (x_i) , die man durch die zweite Iteration gewinnt, folgt wieder der gewünschten Verteilung (sie besteht ja exakt aus denselben x^{ref} -Werten), aber ihre Langzeitkorrelationen sind vermutlich immer noch etwas schwächer durch den erneuten Austausch. Dieser Iterationsprozess konvergiert nicht im strengen Sinne, da die Reihen ja endlich sind. Jedoch führt er zu einem Punkt, ab dem jede weitere Iteration keine Umordnung mehr bewirkt, was bedeutet, dass das Power-Spektrum der erzeugten Reihe maximal dicht am Referenzspektrum liegt.

⁷Hierzu empfiehlt es sich, die Reihe (x_i^{neu}) ebenfalls mittels dem *Quicksort*-Algorithmus zu ordnen und die ursprünglichen Positionen zuvor in einer Indexreihe (I_i) zu speichern.

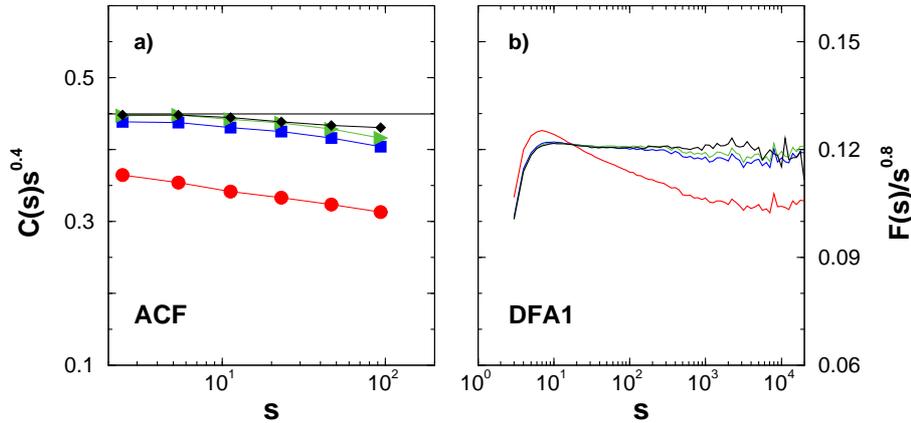


Abbildung 2.19: Qualitätstest der mittels dem Schreiber-Verfahren erzeugten langzeitkorrelierten Daten. Abbildung (a) zeigt die Autokorrelationsfunktion $C(s)$ dividiert durch $s^{(-0.4)}$ von exponentiell verteilten Zufallszahlen nach 1 Iteration (rote Kreise), 10 Iterationen (blaue Quadrate) und 100 Iterationen (grüne Dreiecke) zusammen mit $C(s)$ von Gauss-verteilten Daten (schwarze Rauten). In dieser Auftrugung sollten perfekt korrelierte Daten ein Plateau bilden (durchgezogene schwarze Linie). Der Unterschied zwischen $C(s)$ der Gauss-Zahlen und den iteriert erzeugten exponentiell verteilten Zahlen wird für 1000 Iterationen nicht mehr besser, da das Iterationsverfahren nach der Zusammenstellung der neuen Reihe die Korrelationen wieder leicht verfälscht. Die Abweichungen vom Plateau sind aber in dieser Auftrugung als sehr gering einzuschätzen. Selbst die Gauss-Zahlen zeigen hier Abweichungen. Abbildung (b) zeigt DFA1-Ergebnisse für dieselben Daten wie in (a) (Farben wie in (a)). Die DFA1-Kurve $F(s)$ wurde durch $s^{0.8}$ dividiert, sodass auch hier die Abweichungen von einem Plateauförmigen Verlauf ein Qualitätskriterium bilden. Die schwarze Kurve ist die DFA1 der Gauss-verteilten Referenz-Daten. Man sieht, dass nach 1 Iteration (rot) die Abweichungen auf allen Skalen noch recht groß sind. Nach 10 Iterationen (blau) und 100 Iterationen sind die Abweichungen nur noch auf ganz großen Skalen deutlich. Auch hier kann man nach 1000 Iterationen mit dem Auge keine Verbesserung zu 100 Iterationen ausmachen. Da die Langzeitkorrelationsbestimmung mittels DFA einer gewissen Ungenauigkeit unterworfen ist, ist die Qualität nach 1000 Iterationen absolut ausreichend.

Für meine Untersuchungen habe ich sehr lange Datenreihen erzeugt ($N = 2^{21} \simeq 2 \cdot 10^6$), sodass aufgrund der mit $N \log(N)$ ansteigenden Rechendauer der Zeitpunkt, ab dem keine Umordnung mehr stattfindet, nicht abgewartet werden konnte. Hinzu kommt, dass die ersten 5-10 Schritte den erheblichsten Teil zur Konvergenz des Korrelationspektrums beitragen und sich danach die Konvergenz drastisch verlangsamt. Bei den Ergebnissen in Kapitel 3 und 4 habe ich mich auf 1000 Iterationsschritte pro Reihe beschränkt. Abbildung 2.19 zeigt am Beispiel einer langen Datenreihe den Vergleich von DFA- und Autokorrelationsuntersuchungen mit unterschiedlicher Iterationszahl. Ab ca. 100 Iterationen sind auf den uns interessierenden Skalen mit bloßem Auge keine Verbesserungen mehr zu erkennen. In der Praxis räumt man dem Skalenexponenten der DFA im Mittel einen Fehler von ± 0.02 zu, was man mit dem Auge noch erkennen kann.

Kapitel 3

Extremwertstatistik I: Wiederkehrintervalle

Die Extremwertstatistik hat viele Facetten. Die beiden Schwerpunkte aber bilden die Statistik der Maxima und die Statistik der Wiederkehrintervalle. Während es in der Statistik der Maxima um die Werte extremer Ereignisse geht, behandelt die Statistik der Wiederkehrintervalle, salopp formuliert, die Zeitpunkte des Auftretens extremer Ereignisse. Oder, um es praxisbezogener zu formulieren, während sich die Statistik der Maxima damit befasst, *wie heftig* eine Katastrophe (z. B. Magnitude eines Erdbebens, Pegelstand einer Flut, Rekordtemperatur, etc.) ausfällt, versucht die Statistik der Wiederkehrintervalle darüber Aussagen zu treffen, *wann* eine Katastrophe eintritt oder wie lange sie andauert.

Da ich mich in meiner Promotion chronologisch zuerst mit der Statistik der Wiederkehrintervalle befasst habe, möchte ich auch in dieser Arbeit mit diesem Kapitel beginnen. Die folgenden Unterkapitel beschreiben zunächst die klassische Statistik der Wiederkehrzeiten, das heißt Wiederkehrintervalle in unkorrelierten Systemen, bevor ich dann die Ergebnisse meiner Arbeit zur Statistik in langzeitkorrelierten Systemen darstellen werde.

3.1 Klassische Theorie der Wiederkehrintervalle

Die klassische Theorie der Wiederkehrintervalle ist auch unter dem Namen *Poisson-Statistik* bekannt [Poisson 1837, Haight 1967, Vazquez 2006]. Sie gilt jedoch nur für unkorrelierte Systeme. S.-D. Poisson hatte ursprünglich seine Konzepte der Wahrscheinlichkeitstheorie in der Justizverwaltung eingesetzt [Poisson 1837]. Heute findet sie Anwendung in Verkehrsmodellen, der Unfallstatistik und auch in vielen sozialwissenschaftlichen Fragestellungen und hat sich als kommerziell nützlich erwiesen in der Mobilkommunikation, bei Bestandskontrollen und auch bei vielen versicherungstechnischen Fragestellungen [Haight 1967, Greene 1997, Anderson 2003]. In den folgenden drei Abschnitten werden die Grundlagen der Poisson Statistik für unkorrelierte Systeme erklärt. Dabei wird zunächst definiert, was ein *Wiederkehrintervall* ist (auch *Wiederkehrzeit* genannt). Danach wird der analytische Zusammenhang zwischen *Quantilen* und *Wiederkehrperioden* hergeleitet, bevor es schließlich um die Verteilungsfunktion (*Poisson-Verteilung*) der Wiederkehrzeiten geht.

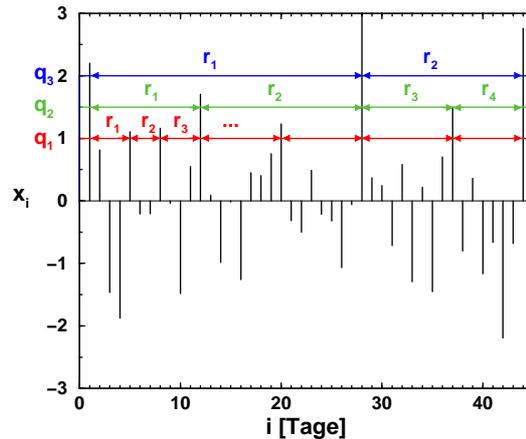


Abbildung 3.1: Definition von Wiederkehrintervallen: Alle Werte einer Zeitreihe (x_i), die eine vorgegebene Schranke q (*Quantil*) überschreiten, werden als Extremwerte betrachtet und definieren Anfangs- und Endpunkt eines Wiederkehrintervalls r . Je größer der q -Wert, desto größer sind auch die r -Werte.

3.1.1 Definition von Wiederkehrintervallen und Quantilen

In Worten ist ein Wiederkehrintervall bzw. eine Wiederkehrzeit die Dauer zwischen zwei aufeinander folgenden Ereignissen, die eine gemeinsame Eigenschaft haben. Solche Eigenschaften können das Überschreiten oder Unterschreiten eines vorgegebenen Schwellwertes sein, oder ein mehrfaches Überschreiten einer Schwelle in einem bestimmten Zeitraum, oder sonst irgendein Kriterium, das man den Ereignissen in einer Zeitreihe zuordnen kann [Naess 1998, Sornette 1997, Shiau 2003, Ehrhardt 2004, Bunde 2005, Altmann 2005, Livina 2005]. In dieser Arbeit beschränke ich mich auf das Überschreiten einer Schwelle [Bunde 2003, Bunde 2004a, Bunde 2004d, Bunde 2005, Eichner 2006b]. Diese Form der Extremwertstatistik wird auch als *POT-Methode* bezeichnet (aus dem Englischen, **P**eaks **O**ver **T**hreshold) [Leadbetter 1983, Smith 1989, Davison 1990, Galambos 1994, Embrechts 1997]. Wie wir im übernächsten Abschnitt sehen werden, folgen die Wiederkehrintervalle der POT-Methode auf unkorrelierten Systemen der Poisson-Statistik.

Gegeben ist also eine Messreihe $(x)_{i=1}^N$ der Länge N . Ein Wiederkehrintervall wird definiert als die Zeitdauer r zwischen zwei Ereignissen x_i und x_{i+r} , die beide über einem Schwellwert q liegen (also $x_i > q$ und $x_{i+r} > q$), wobei dazwischen kein Ereignis über q liegen darf (also $x_{i+j} \leq q$ für $0 < j < r$). Abbildung 3.1 veranschaulicht die Definition für drei verschiedene q -Werte. Die Wiederkehrzeiten r werden im Mittel länger, je größer q gewählt wird. Ebenso nimmt die Anzahl N_r der Wiederkehrzeiten r_j , $j = 1, 2, \dots, N_r$, die aus einer Reihe der Länge N bei gegebenem q bestimmt werden können, mit anwachsendem q ab.

Der Schwellwert q trägt oft den Namen *Quantil* [Gumbel 1958, Sachs 1974, Galambos 1994], da er bestimmt, welcher Anteil (*Quantum*) der Verteilungsdichtefunktion $P(x)$ (siehe Kapitel 2.1) darüber bzw. darunter liegt. So wäre ein 5%-Quantil der q -Wert, der eben nur von 5% aller x -Werte einer Verteilung oder einer Messreihe überschritten wird. Oft ist die Definition auch umgekehrt, so dass dies einem 95%-Quantil entsprechen würde. In dieser Arbeit verwende ich die erste Variante. Abbildung 3.2 veranschaulicht den Zusammenhang zwischen Quantil und Verteilungsfunktion. Das 5%-Quantil einer Gauss-Verteilung hat einen anderen q -Wert

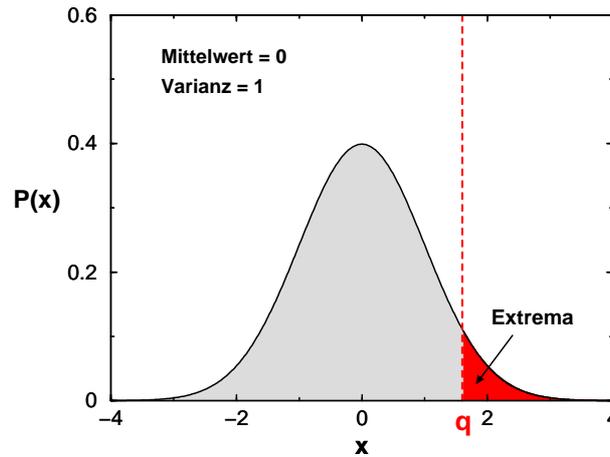


Abbildung 3.2: Definition von Quantilen: Jeder Messwert (x) ist Element einer (oft unbekannt) Verteilungsdichtefunktion $P(x)$. Die Abbildung zeigt eine normierte Gauss-Verteilung mit Mittelwert 0 und Varianz 1. Alle Werte, die nun größer als eine vorgegebene Schranke q (*Quantil*) sind, werden als *Extrema* betrachtet.

als das 5%-Quantil einer anderen Verteilung, z. B. einer Exponentialverteilung.

3.1.2 Wiederkehrperioden bei verschiedenen Verteilungsfunktionen

Das Quantil q steht in einem direkten, aber von Verteilung zu Verteilung unterschiedlichen Zusammenhang mit dem Erwartungswert R_q der Wiederkehrzeiten,

$$\frac{1}{R_q} := \int_q^\infty P(x) dx, \quad (3.1)$$

auch genannt *Wiederkehrperiode* [Smith 1989, Galambos 1994], und gibt die Wahrscheinlichkeit $p = 1/R_q$ an, dass ein Ereignis größer q eintritt. Ist dagegen die normierte Verteilungsdichtefunktion $P(x)$ nicht vollständig bekannt, so verwendet man als numerische Schätzung für den Erwartungswert der Wiederkehrperiode das arithmetische Mittel aller N_r Intervalle, basierend auf einer Messreihe der Länge N ,

$$R_q = \frac{1}{N_r} \sum_{j=1}^{N_r} r_j. \quad (3.2)$$

Aus Konsistenzgründen gilt, sofern keine periodischen Randbedingungen angenommen werden,

$$N \geq \sum_{j=1}^{N_r} r_j = N_r R_q. \quad (3.3)$$

Ist die Stammfunktion von $P(x)$ bekannt, kann man den Wert von R_q für ein vorgegebenes q durch Gl. (3.1) berechnen. Oft wird in der Fragestellung aber der R_q -Wert vorgegeben (z.

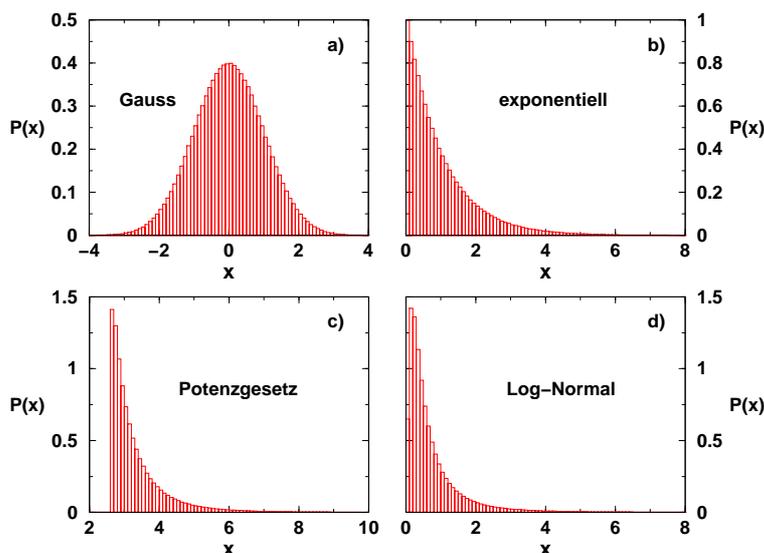


Abbildung 3.3: Skizzen der vier Verteilungsfunktionen, auf denen die künstlichen Daten für meine Analysen basieren. Abbildung (a) Gauss-Verteilung mit Mittelwert $\langle x \rangle = 0$ und Standardabweichung $\sigma_x = 1$, (b) Exponentialverteilung mit $\langle x \rangle = 1$ und $\sigma_x = 1$, (c) Potenzgesetz-Verteilung mit $\langle x \rangle = 3.357$ und $\sigma_x = 1$ und (d) Log-Normal-Verteilung mit $\langle x \rangle = 0.763$ und $\sigma_x = 1$.

B. *hundertjähriges* Hochwasser) und man möchte den zugehörigen Schwellwert q (z. B. den entsprechenden Pegelstand eines Jahrhunderthochwassers) bestimmen.

Für zwei der vier einfachen aber in der Natur recht häufig verbreiteten Verteilungsfunktionen (siehe Abbildung 3.3), auf denen die künstlichen Daten für meine Analysen basieren, sei hier kurz der Zusammenhang hergeleitet. Alle Verteilungen sind auf Varianz $\sigma_x^2 = 1$ normiert.

- **Exponentiell verteilte Daten**

$$P(x) = \frac{1}{\lambda} \exp(-x/\lambda), \quad x \in \mathbb{R}^+ \quad (3.4)$$

$$\frac{1}{R_q} = \int_q^\infty \frac{1}{\lambda} \exp(-x/\lambda) dx = \exp(-q/\lambda) \quad (3.5)$$

$$R_q = \exp(q/\lambda) \quad \text{bzw.} \quad q = \lambda \ln(R_q) \quad (3.6)$$

- **Potenzgesetz-verteilte Daten**

$$P(x) = (\delta - 1)x^{-\delta} \quad \text{für } \delta > 1 \text{ und } x \in \mathbb{R}^+ \quad (3.7)$$

$$\frac{1}{R_q} = \int_q^\infty (\delta - 1)x^{-\delta} dx = q^{-\delta+1} \quad (3.8)$$

$$R_q = q^{\delta-1} \quad \text{bzw.} \quad q = R_q^{\frac{1}{\delta-1}} \quad (3.9)$$

Da es bei der Gauss-Verteilung keinen analytischen Ausdruck für dieses Integral gibt, muss man hier auf tabellierte Werte zurückgreifen, wie sie z. B. in “Bronstein, Taschenbuch der Mathematik” (2. Auflage), Seite 974 [Bronstein 1995] aufgelistet sind. Dasselbe gilt für die vierte Verteilungsfunktion, die Log-Normal-Verteilung.

- **Gauss-verteilte Daten**

$$P(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), \quad x \in \mathbb{R} \quad (3.10)$$

$$\frac{1}{R_q} = \int_q^\infty \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = 1 - \Phi(q) = \frac{1}{2}[1 - \operatorname{erf}(q/\sqrt{2})] \quad (3.11)$$

$$R_q = [1 - \Phi(q)]^{-1}, \quad \text{wobei } \Phi \text{ tabelliert ist} \quad (3.12)$$

- **Log-Normal-verteilte Daten**

$$P(x) = \frac{1}{\sqrt{2\pi x}} \exp[-(\ln x + \mu)^2/2], \quad x \in \mathbb{R}^+ \quad (3.13)$$

$$\frac{1}{R_q} = \int_q^\infty \frac{1}{\sqrt{2\pi x}} \exp[-(\ln x + \mu)^2/2] dx \quad (3.14)$$

$$R_q = [1 - \Phi(\ln q + \mu)]^{-1}, \quad \Phi \text{ tabelliert, } \mu = 0.763 \text{ ist Mittelwert der Verteilung} \quad (3.15)$$

3.1.3 Poisson-Statistik bei unkorrelierten Systemen

Zufallsexperimente mit konstanten Wahrscheinlichkeiten und nur zwei verschiedenen möglichen Ausgängen (z. B. “Münzwurf” oder “Ereignis größer bzw. kleiner Schwellwert”) bezeichnet man als *Bernoulli-Experimente* und werden durch die *Binomialverteilung* [Sachs 1974, Papoulis 1984, Papula 1994, v.Storch 1999] beschrieben. Ein besonderer Fall ergibt sich, wenn die Eintrittswahrscheinlichkeiten p_0 und $p_1 = 1 - p_0$ stark unterschiedlich sind, also $p_0 \ll p_1$. Ereignisse u dieser Art, die also sehr selten auftreten, genügen der diskreten *Poisson-Verteilung*. Sie ergibt sich aus der Binomialverteilung, wenn das Produkt aus Stichprobenumfang n und Eintrittswahrscheinlichkeit p_0 für sehr große n immer noch recht klein bleibt, $\mu = np_0 < 10$ [Papoulis 1984, Fließbach 1995, v.Storch 1999]. Die Wahrscheinlichkeitsdichte für u solcher seltenen Ereignisse in einer Stichprobe vom Umfang n lautet

$$f(u) = \frac{(\mu)^u}{u!} e^{-\mu}. \quad (3.16)$$

Die Poisson-Verteilung kann demnach als Spezialfall der Binomialverteilung bei geringer Eintrittswahrscheinlichkeit angesehen werden. Sie findet Anwendung, wenn es um das Zählen seltener Ereignisse geht. Daher kommt ihr eine besondere Rolle in der Extremwertstatistik, insbesondere in der Statistik der Wiederkehrintervalle zu. Man kann sie wie folgt motivieren.

Betrachtet wird ein beliebiger Ausschnitt der Länge r einer Zeitreihe $(x_i)_{i=1}^N$. Zu vorgegebenem Schwellwert q zählt man nun alle Ereignisse $x_i > q$ im Ausschnitt r . Die Anzahl dieser

Ereignisse soll mit n_q bezeichnet werden. Um die Verteilung aller $x_i > q$ analytisch bestimmen zu können, muss man die Annahmen machen, dass die Eintrittswahrscheinlichkeit p eines solchen Extremereignisses zeitlich konstant bleibt und die einzelnen Ereignisse statistisch unabhängig sind. Nun zerteilt man den Ausschnitt r in n_r gleichgroße Segmente, die klein genug sind, sodass das Eintreten von zwei Extremereignissen in einem Segment vernachlässigbar wird. Damit kann die Eintrittswahrscheinlichkeit für ein Extremereignis in jedem dieser Segmente durch pr/n_r angenähert werden und die n_q -Werte im gesamten Ausschnitt r sind dann näherungsweise binomialverteilt. Betrachtet man nun den Grenzfall $n_r \rightarrow \infty$, also eine beliebige Verfeinerung der Segmente, so wird aus der Binomialverteilung die Poisson-Verteilung [v.Storch 1999]

$$f(n_q) = \frac{(pr)^{n_q}}{n_q!} e^{-pr} \quad \text{für } n_q = 0, 1, 2, \dots \quad (3.17)$$

Für die Wiederkehrintervalle darf es im Ausschnitt r *keine* Überschreitungen geben. Es interessiert nur der Fall $n_q = 0$ bzw. $x_i < q$ im Ausschnitt der Länge r , und damit

$$f(n_q = 0) = e^{-pr}. \quad (3.18)$$

Für das erste Ereignis am Ende des Ausschnitts r muss aber wieder $x_i > q$ gelten, um das Wiederkehrintervall zu beenden. Das heißt, $f(0)$ muss nochmal mit p multipliziert werden. Wie wir vom vorherigem Abschnitt wissen, gilt für die Eintrittswahrscheinlichkeit $p = 1/R_q$ und somit für die Wahrscheinlichkeitsdichte der Wiederkehrzeiten

$$P_q(r) = \frac{1}{R_q} e^{-r/R_q} \quad (3.19)$$

und für die Wahrscheinlichkeitsverteilung

$$W_q(r) = \int_0^r P_q(r') dr' = 1 - e^{-r/R_q}. \quad (3.20)$$

Es soll an dieser Stelle noch einmal betont werden, dass diese Herleitung nicht für korrelierte Daten gilt. Die Verteilung der Wiederkehrzeiten folgt also bei unkorrelierten Daten für vernünftige¹ q -Werte einer Exponentialfunktion, unabhängig von der Verteilungsdichtefunktion $P(x)$ der Daten.

3.2 Statistik der Wiederkehrzeiten bei Langzeitkorrelationen

Eine grundlegende Annahme in der klassischen Extremwertstatistik ist, dass die Ereignisse $x_i > q$ für große q unkorreliert sind, bzw. dass mögliche Korrelationen zwischen Ereignissen, die weit genug von einander entfernt sind, keinen Einfluss mehr haben [Galambos 1978, Leadbetter 1983]. In diesem Fall wird die Statistik der Wiederkehrintervalle für große q -Werte ausschließlich durch die gerade beschriebene Poisson-Statistik bestimmt und die Reihe der Wiederkehrzeiten $(r_j)_{j=1}^{N/R_q}$ ist ebenfalls unkorreliert.

Betrachtet man Abbildung 3.4, in der je eine Reihe von Wiederkehrzeiten langzeitkorrelierter (b) und unkorrelierter (gemischter) Daten (a) gezeigt ist, so fällt sofort auf, dass im korrelierten Fall deutlich größere r -Werte auftreten. Aber es gibt auch Bereiche, in denen viele

¹Mit "vernünftig" ist gemeint, dass q im Definitionsbereich der x -Werte liegen muss, und im Falle diskreter Zeitreihen nicht zu klein sein darf, um Diskretheitseffekte möglichst gering zu halten.

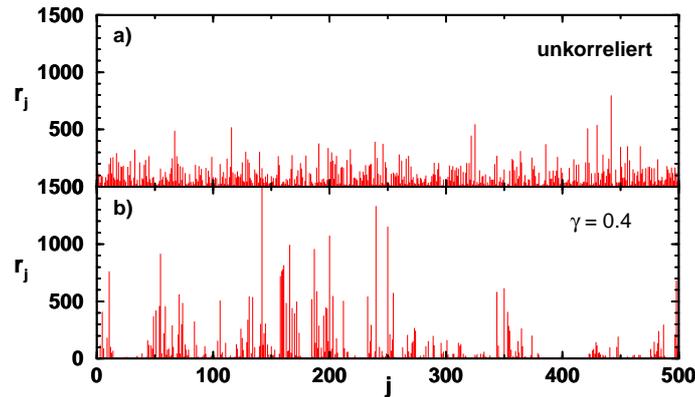


Abbildung 3.4: Wiederkehrintervalle unkorrelierter und langzeitkorrelierter Daten: (a) Reihe der Wiederkehrzeiten einer unkorrelierten (gemischten) Datenreihe zu einem vorgegebenen q , sodass der mittlere r -Wert $R_q = 100$ ist. Kaum ein r -Wert überschreitet die Länge 500. (b) Reihe der Wiederkehrzeiten einer langzeitkorrelierten Datenreihe mit $\gamma = 0.4$ zu selbem R_q . Es gibt deutlich mehr r -Werte von der Größe 500 und größer, aber auch viele Bereiche, in denen die r -Werte so klein sind, dass sie in dieser Auftragung kaum zu erkennen sind.

sehr kleine r -Werte aufeinander folgen. In diesen Bereichen treten extreme Ereignisse gehäuft (in *Clustern*) auf, während die großen r -Werte lange Ruhephasen repräsentieren. Die Ursache hierfür liegt offensichtlich in den Langzeitkorrelationen. In den folgenden Abschnitten wird gezeigt, dass Langzeitkorrelationen in der ursprünglichen Messreihe $(x_i)_{i=1}^N$ nicht nur die Dauer der Wiederkehrzeiten beeinflussen, sondern in der Reihe der Wiederkehrzeiten selbst Korrelationen verursachen. Dieser Effekt ist von großer Bedeutung bei der Vorhersage extremer Ereignisse.

Die folgenden Unterkapitel stellen meine Ergebnisse zum Einfluss der Langzeitkorrelationen auf die Statistik der Wiederkehrintervalle vor. Dabei werden die Ergebnisse an künstlichen langzeitkorrelierten Daten mit zugrundeliegender Gauss- oder Exponentialverteilung, teilweise auch mit Potenzgesetz- oder Log-Normal-Verteilung mit sehr großer Statistik belegt. Fast alle Ergebnisse wurden über 150 Reihen mit je $2^{21} \simeq 2 \times 10^6$ Datenpunkten gemittelt, wenn nicht anders ausgezeichnet. Viele beispielhafte Ergebnisse können an echten und rekonstruierten Datensätzen verifiziert werden, obwohl diese nur einen Bruchteil der Statistik der künstlichen Daten bieten. Um saisonale Effekte auszuschließen eignen sich bei hydrometeorologischen Zeitreihen nur Jahresdaten für Extremwertuntersuchungen. Die hier verwendeten Datensätze wurden in Kapitel 2.3.2 vorgestellt.

3.2.1 Wiederkehrperiode und Varianz

Da Korrelationen nur Einfluss auf die Anordnung von Zahlen in einer Messreihe haben, aber die Form der zugrunde liegenden Verteilungsdichte $P(x)$ unverändert lassen, ist die mittlere Wiederkehrzeit oder Wiederkehrperiode R_q nach Gl. (3.1) unbeeinflusst von Korrelationen. Auch bei unbekannter Verteilungsdichtefunktion $P(x)$, wenn man die Wiederkehrperiode also

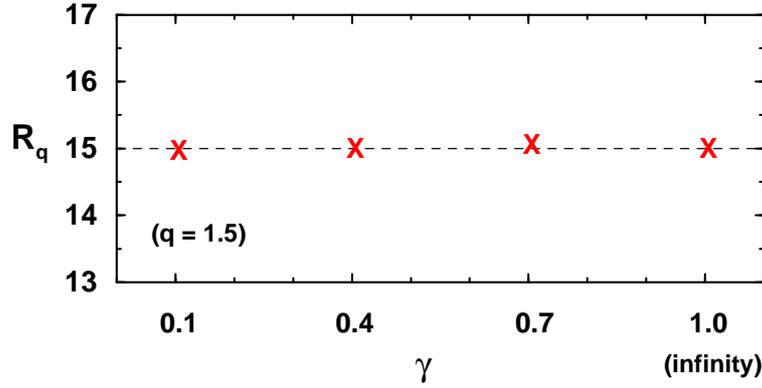


Abbildung 3.5: Wiederkehrperioden als Funktion von γ : Da die Wiederkehrperiode R_q direkt aus dem Quantil bestimmbar ist, also nur von der Lage von q in $P(x)$ abhängt (siehe Bild 3.2), ist R_q unabhängig von der Korrelationsstruktur im Datensatz. Bestimmt man die Wiederkehrperiode numerisch aus einer Stichprobe von Wiederkehrzeiten, so muss im Falle korrelierter Daten gewährleistet sein, dass die Stichprobe groß genug ist.

mittels Gl. (3.2) numerisch bestimmen muss, bleibt R_q unbeeinträchtigt² von Korrelationen, da sowohl für langzeitkorrelierte als auch für gemischte (unkorrelierte) Daten die Anzahl der $x_i > q$ gleich bleibt und damit auch die Anzahl N_r der Wiederkehrzeiten. Es gilt

$$R_q = \frac{1}{N_r} \sum_{j=1}^{N_r} r_j = \frac{N}{N_r}. \quad (3.21)$$

Abbildung 3.5 zeigt die numerischen Werte für R_q zu verschiedenen γ -Werten bei Gaussverteilten Daten mit $q = 2$. Die Unterschiede zwischen den R_q sind ausschließlich auf Randeffekte zurückzuführen, da keine periodischen Randbedingungen angenommen wurden.

Die mittlere Wiederkehrzeit R_q bleibt zwar unbeeinträchtigt durch Langzeitkorrelationen, nicht aber die Varianz σ_r^2 der Wiederkehrzeiten,

$$\sigma_r^2 = \frac{1}{N_r} \sum_{j=1}^{N_r} (R_q - r_j)^2. \quad (3.22)$$

In langzeitkorrelierten Systemen treten deutlich größere Wiederkehrzeiten auf als in unkorrelierten Systemen, weshalb die Varianzen größer werden. Abbildung 3.6 zeigt beispielhaft die Standardabweichungen (σ_r) der Wiederkehrzeiten in Einheiten von R_q zu verschiedenen γ -Werten. Die Standardabweichungen werden mit zunehmender Korrelationsstärke, d. h. mit kleiner werdendem γ , beträchtlich größer. Das Abfallen am linken Ende basiert auf Diskretisierungseffekten, die bei kleinen q auftreten, da der kleinstmögliche r -Wert 1 ist und dadurch $(R_q - r_j)$ in Gl. (3.22) nach oben durch $(R_q - 1)$ beschränkt. Je größer die q -Werte, desto größer ist R_q und dieser Effekt verliert schnell an Bedeutung. Für ganz große q fällt die Standardabweichung wieder ab, was durch einen Finite-Size Effekt bedingt ist. Auf diese Finite-Size Effekte werde ich in den nächsten Abschnitten noch näher eingehen.

²Der Einfluss der nicht-periodischen Randbedingungen wird bei sehr großen Systemen vernachlässigbar.

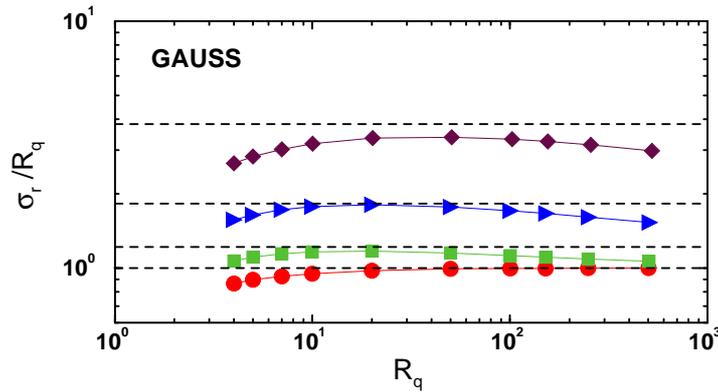


Abbildung 3.6: Standardabweichung σ_r der Wiederkehrzeiten in Einheiten von R_q als Funktion von R_q zu Gauss-verteilt unkorrelierten Daten (rote Kreise) und langzeitkorrelierten Daten mit $\gamma = 0.7$ (grüne Quadrate), 0.4 (blaue Dreiecke) und 0.2 (bordeaux Rauten). Die horizontalen gestrichelten Linien entsprechen den theoretischen Standardabweichungen der zugehörigen gefitteten Verteilungsdichten (siehe nächster Abschnitt). Die Abweichungen bei großen R_q sind auf Finite-Size Effekte zurückzuführen. Die Abweichungen bei kleinen R_q haben ihre Ursache in Diskretheitseffekten da $r \in \mathbb{N}$. Der kleinste auftauchende r -Wert ist 1. In kontinuierlichen Systemen wären aber auch Werte zwischen 0 und 1 zu erwarten. Als Konsequenz ist die maximale Differenz zwischen R_q und dem kleinstmöglichen r -Wert $R_q - 1$ und nicht R_q . Diese Einschränkung der Differenz macht sich bei kleinen R_q ($R_q < 10$) stärker bemerkbar als bei großen R_q . Dadurch wird die Standardabweichung für kleine R_q ebenfalls kleiner.

3.2.2 Verteilungsdichte der großen Wiederkehrintervalle

Wie im obigen Abschnitt schon angedeutet wurde, wird die Verteilung der Wiederkehrzeiten unter dem Einfluss von Langzeitkorrelationen nicht mehr der Poisson-Statistik Gl. (3.19) genügen, da deutlich größere und viel mehr kleine Wiederkehrintervalle auftreten. Da die Gesamtzahl der Wiederkehrzeiten N_r ja wie R_q korrelationsunabhängig ist, muss es im Gegenzug weniger mittlere Wiederkehrzeiten geben. Abbildung 3.7(a) zeigt in halblogarithmischer Auftragung das normierte Histogramm $P_q(r)$ der Wiederkehrzeiten bei Gauss-verteilt langzeitkorrelierten ($\gamma = 0.4$) Daten zu drei verschiedenen q -Werten (also drei R_q). In dieser Auftragung wäre die Poisson-Verteilung eine Gerade. Die dargestellten Kurven weichen aber offensichtlich von einer Geraden ab. Skaliert man die Achsen mit dem jeweiligen R_q , so fallen die Kurven aufeinander und geben einen Datenkollaps wie in Abbildung 3.7(b). Die gemeinsame asymptotische Form dieser Kurven ist eine *gestreckt exponentielle* Funktion [Bunde 2003, Bunde 2004a] und lautet

$$P_q(r) = \frac{a_{\gamma'}}{R_q} \exp[-b_{\gamma'}(r/R_q)^{\gamma'}]. \quad (3.23)$$

Sie ist in Abbildung 3.7(c) als theoretische Kurve (schwarz) eingezeichnet. Dieser Datenkollaps erlaubt es, in stationären langzeitkorrelierten Prozessen von Ergebnissen für kleine q -Werte auf Ergebnisse für große q -Werte, bei denen die Statistik nicht ausreicht, zu extrapolieren. Mischt man die Daten zuvor, d. h. zerstört man die Korrelationsstruktur, so ergibt sich für $P_q(r)$ wieder die Poisson-Verteilung, wie in Abbildung 3.7(d) (schwarze Kreise).

Die γ' -Abhängigkeit der Normierungsparameter $a_{\gamma'}$ und $b_{\gamma'}$ kann aus zwei Normierungsbe-

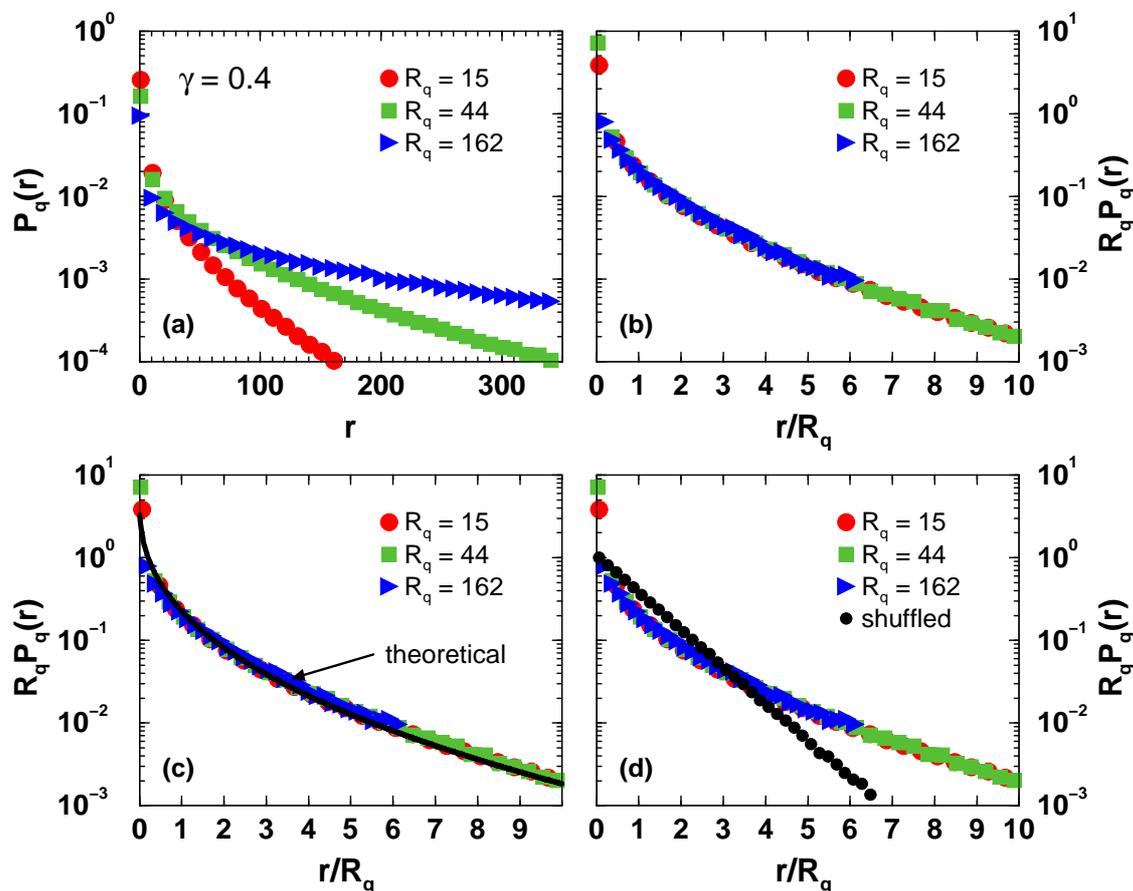


Abbildung 3.7: Verteilungsfunktion der Wiederkehrintervalle. Abbildung (a) zeigt in halblogarithmischer Auftragung das normierte Histogramm $P_q(r)$ der Wiederkehrzeiten r Gauss-verteilter langzeitkorrelierter ($\gamma = 0.4$) Zufallszahlen zu drei verschiedenen Quantilen: $R_q = 15$ entspricht $q = 1.5$ (rote Kreise), $R_q = 44$ entspricht $q = 2.0$ (grüne Quadrate) und $R_q = 162$ entspricht $q = 2.5$ (blaue Dreiecke). Je größer das Quantil q , desto mehr lange Wiederkehrzeiten (große r -Werte) zeigt $P_q(r)$. Abbildung (b) zeigt einen Kollaps der Verteilungsfunktionen zu einer einzigen Kurve, wenn die Auftragung mit R_q skaliert, d. h. wenn man $R_q P_q(r)$ und r/R_q aufträgt. Die resultierende Kurve ist eine gestreckt exponentielle Funktion (siehe Text) und ist einzig vom Korrelationsparameter γ abhängig. In Abbildung (c) ist der theoretische Verlauf einer gestreckt exponentiellen Kurve aus Gl. 3.26 zum Wert $\gamma = 0.4$ eingezeichnet. Die Übereinstimmung ist deutlich. Abbildung (d) zeigt den Verlauf der skalierten Verteilung, wenn die Datenreihen zuvor gemischt wurden, d. h. alle Korrelationen zerstört wurden. Die resultierende Verteilung ist dann eine simple exponentielle Verteilung, die Poisson-Verteilung.

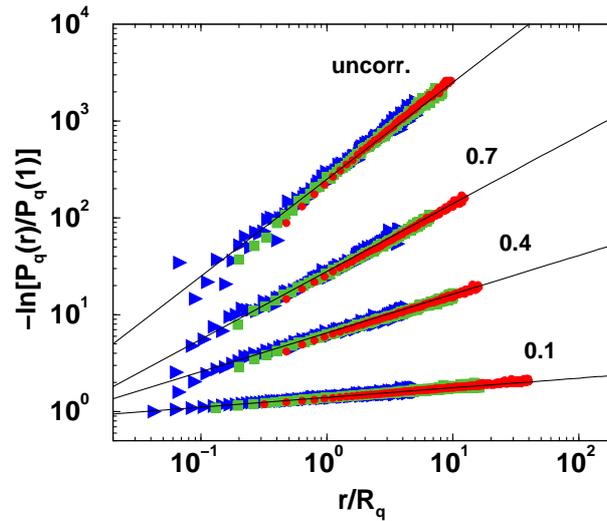


Abbildung 3.8: Die Verteilungsdichtefunktion der Wiederkehrzeiten langzeitkorrelierter Daten ist eine gestreckt exponentielle Funktion, welche nur vom Korrelationsexponenten γ abhängen. Für die Abbildung wurde eine Darstellung gewählt, in der die Exponenten als Steigung der Kurven erscheinen. Gezeigt sind die Verteilungen zu $R_q = 15$ (rote Kreise), 44 (grüne Quadrate) und 162 (blaue Dreiecke) für vier verschiedene Korrelationsexponenten $\gamma = 1, 0.7, 0.4$ und 0.1 (von oben nach unten). Je kleiner γ desto stärker die Korrelationen. Die obere Kurvenschar basiert auf unkorrelierten Daten. Die gefittete schwarze durchgezogene Linie hat die Steigung 1, d. h. die gestreckt exponentielle Funktion geht über in die Poisson-Verteilung. Abbildung nach [Bunde 2004a].

dingungen bestimmt werden, die für $P_q(r)$ gelten müssen,

$$\sum_{r=1}^{\infty} P_q(r) = 1 \quad \text{und} \quad \sum_{r=1}^{\infty} r P_q(r) = R_q. \quad (3.24)$$

Ersetzt man die Summen über die diskreten Werte der Wiederkehrzeiten durch Integrale, kann man das Gleichungssystem mit dem Ansatz (3.23) für $a_{\gamma'}$ und $b_{\gamma'}$ lösen und man erhält

$$a_{\gamma'} = \frac{\gamma' \Gamma(2/\gamma')}{\Gamma^2(1/\gamma')} \quad \text{und} \quad b_{\gamma'} = \frac{\Gamma(2/\gamma')}{\Gamma(1/\gamma')}. \quad (3.25)$$

Abbildung 3.8 zeigt dieselben $P_q(r)$ für vier verschiedene Korrelationsexponenten γ . In der gewählten Auftragung erscheint der unbekannte Exponent γ' als Steigung der $P_q(r)$. Offensichtlich lässt sich γ' mit dem Korrelationsexponenten γ identifizieren ($\gamma' = \gamma$). Gleichung (3.23) lässt sich damit schreiben als

$$P_q(r) = \frac{a_\gamma}{R_q} \exp[-b_\gamma (r/R_q)^\gamma]. \quad (3.26)$$

Eine Tabelle der a_γ und b_γ zu verschiedenen γ -Werten befindet sich im Anhang A in Tabelle A.1.

Die Ergebnisse zur Verteilungsfunktion der Wiederkehrzeiten bei Langzeitkorrelationen besitzen im Fall Gauss-verteilter Zufallszahlen auch ein mathematisches Fundament. Für eine

sehr ähnliche Problemstellung, das sogenannte *zero-level crossing* (das heißt in unserem Fall $q = 0$), konnte von den Mathematiker G. F. Newell und M. Rosenblatt bereits 1962 die gestreckt exponentielle Form als obere Schranke für die Wahrscheinlichkeitsverteilung $W_0(r)$ der Intervalle bewiesen werden [Newell 1962]. Die von ihnen gefundene Wahrscheinlichkeitsverteilung ist von der Form

$$\ln[W_0(r)] \sim -Kr^\gamma, \quad K > 0 \text{ und } \gamma \in (0, 1). \quad (3.27)$$

Das asymptotische Verhalten der Dichtefunktion $P_0(r)$ zeigt für große r ebenfalls einen gestreckt exponentiellen Verlauf³, was mit unseren Ergebnissen übereinstimmt. Der Beweis erklärt jedoch nicht den von uns gefundenen Datenkollaps der Krüven für $q > 0$, den man durch die mit R_q skalierte Auftragung erhält (siehe Abbildung 3.7(b)). Eine ausführliche Form des Beweises mit einer von mir konstruierten Erweiterung auf die POT-Methode (d. h. für beliebige $q > 0$) befindet sich im Anhang B dieser Arbeit. Für Daten anderer Verteilungsdichten $P(x)$, wie zum Beispiel exponentielle, Log-normale oder Potenzgesetz-artige Verteilungsdichten, sind mir keine analytischen Beweise bekannt. 3.9 werden im nächsten Abschnitt diskutiert.

Abbildung 3.9 zeigt $P_q(r)$ in doppelt-logarithmischer Auftragung für zwei verschiedene γ -Werte (rot: $\gamma = 0.4$ und blau: $\gamma = 0.2$) und für drei weitere Verteilung von x -Werten (exponentiell, Potenzgesetz- und Log-Normal-verteilt) zu $R_q = 100$. Auch die drei weiteren Verteilungen liefern einen Verlauf von $P_q(r)$, der durch eine gestreckt exponentielle Form über mehr als eine Dekade gut beschrieben werden kann (Vergleich mit theoretischer, durchgezogener schwarzer Linie). Am besten ist der gestreckt exponentielle Fit bei Gauss-Daten. Es gibt aber offensichtlich auch Abweichungen von diesem Verlauf, die je nach Grundverteilung unterschiedlich stark ausfallen. Die Abweichungen für $r \gg R_q$ sind auf Finite-Size Effekte, also auf die endliche Reihengänge N zurückzuführen. Die Abbildungen 3.10(a) und (b) zeigen für drei verschiedene Reihengänge $N = 2^{21}$, 2^{17} und 2^{13} von Gauss-verteilt langzeitkorrelierten ($\gamma = 0.4$) Daten, dass die Abweichungen von der theoretischen Kurve (durchgezogene schwarze Linie) umso größer werden, je kürzer die Reihen sind und je größer das gewählte Quantil q bzw. die zugehörige Wiederkehrperiode R_q ist. Große q -Werte bzw. R_q -Werte führen zu einer stärkeren N -abhängigen Beschränkung der maximalen r -Werte. So ist der theoretisch maximale r -Wert bei einer Reihengänge von $N = 2^{13} = 8192$ und einem R_q von 250 etwa $N - N/R_q \simeq 8150$. Wie Abbildung (b) zeigt, geht $P_q(r)$ für $N = 2^{13}$ (∇) nur bis etwa $r_{max} \simeq R_q \times 20 = 5000$. Viel größere Werte sind bei einer Reihengänge von $N = 2^{21}$ auch bei mehreren zehntausend Reihen nicht zu erwarten, wodurch bei großen R_q auch die Standardabweichung σ_r reduziert wird (siehe Abbildung 3.6).

Je länger die Reihen werden, umso dichter liegt $P_q(r)$ wieder an der theoretischen Kurve. Der Abbruch für kleine q , also kleine R_q erscheint in Abbildung 3.10(a) von N unabhängig. Hier wurde im Vorfeld ein maximaler r/R_q -Wert festgelegt, der für kleine R_q auch bei kurzen Reihen statistisch noch erreicht wird. Aber auch bei kleinen R_q gilt, je länger die Reihe, umso dichter liegt $P_q(r)$ asymptotisch an der gestreckt exponentiellen Funktion. Die Abweichungen bei kleinen r/R_q -Werten in Abbildung

Abbildung 3.11 zeigt, dass der gestreckt exponentielle Verlauf von $P_q(r)$ auch bei echten bzw. rekonstruierten Datensätzen zu beobachten ist. Abbildung (a) zeigt die DFA2-Ergebnisse der

³Es gilt $P_0(r) = W_0'(r) \sim r^{\gamma-1} e^{-Kr^\gamma} \sim e^{-Kr^\gamma}$ für $r \rightarrow \infty$, d. h., $P(r)$ wird für große r durch den gestreckt exponentiellen Term dominiert.

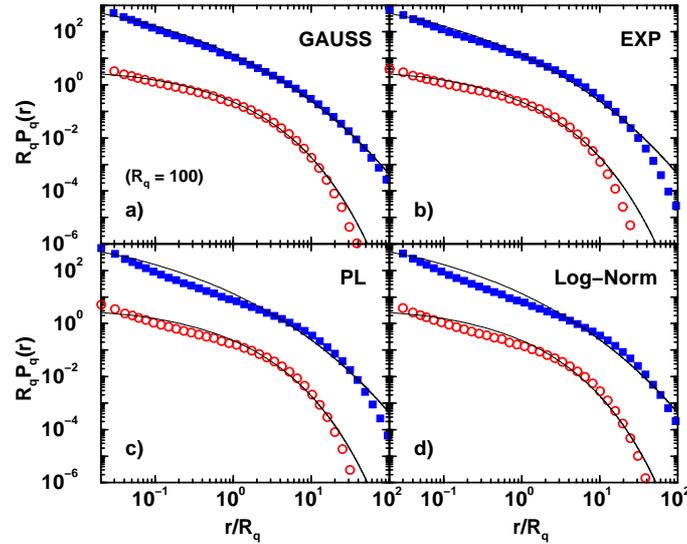


Abbildung 3.9: Verteilungsdichtefunktion $P_q(r)$ skaliert mit der Wiederkehrperiode R_q (in diesem Fall konstant 100) zu zwei verschiedenen Korrelationsexponenten $\gamma = 0.4$ (rote Kreise) und $\gamma = 0.2$ (blaue Quadrate). Die Verteilungsdichten zu $\gamma = 0.2$ (blaue Quadrate) sind um einen Faktor 10 nach oben verschoben, um einen Überlapp mit $\gamma = 0.4$ (rote Kreise) zu vermeiden. Die schwarzen durchgezogenen Linien sind die durch γ definierten entsprechend normierten gestreckt exponentiellen Funktionen. Abbildung (a) zeigt die Ergebnisse für Gauss-verteilte langzeitkorrelierte Ausgangsdaten, (b) für exponentiell verteilte, (c) für Potenzgesetz-verteilte (Englisch **P**ower **L**aw) und (d) für Log-Normal-verteilte Daten. Abweichungen vom gestreckt exponentiellen Verlauf (schwarze Linien) bei großen r/R_q -Werten sind auf Finite-Size Effekte zurückzuführen (siehe dazu Abbildung 3.10). Abweichungen bei kleinen r/R_q zeigen einen *Power-Law* Effekt, der bei Potenzgesetz-verteilten (c) und Log-Normal-verteilten Daten (d) deutlich stärker ausfällt als in (a) oder (b). Abbildung nach [Eichner 2006b].

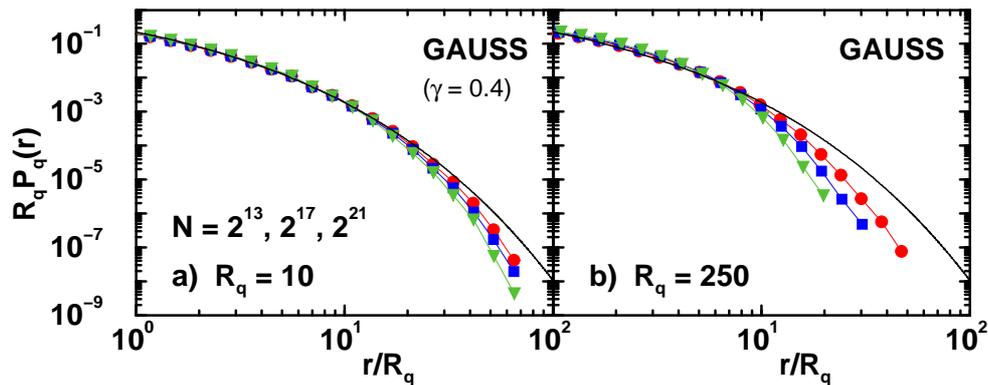


Abbildung 3.10: Verteilungsdichtefunktionen $P_q(r)$ zu gleichen q -Werten aber unterschiedlichen Reihengängen N . Abbildung (a) zeigt $P_q(r)$ für einen niedrigen q -Werte mit $R_q = 10$. Der Unterschied der Histogrammkurven für verschiedene Reihengängen (rot = 2^{21} , blau = 2^{17} und grün = 2^{13}) ist relativ gering. Bei größeren q -Werten in Abbildung (b) ($R_q = 250$) macht sich die Reihengänge stärker bemerkbar. Das vorschnelle Abfallen der Histogramme vom theoretischen gestreckt exponentiellen Verlauf (schwarze Linie) in Abhängigkeit von der Reihengänge N ist ein starkes Indiz für einen Finite-Size Effekt. Abbildung nach [Eichner 2006b].

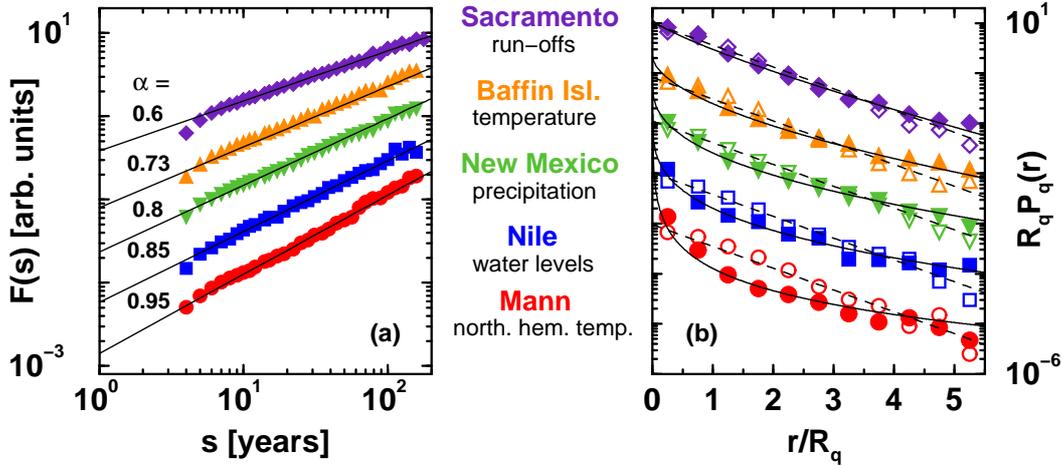


Abbildung 3.11: DFA und Verteilungsfunktion der Wiederkehrzeiten historischer und rekonstruierter Jahresdaten. Die Angaben zu den Reihen (Sacramento \diamond , Baffin Island \triangle , New Mexico ∇ , Nil \square und Mann-Reihe \circ) befinden sich im Kapitel 2.3.2. Abbildung (a) zeigt nochmal die Fluktuationen $F(s)$ aus Abbildung 2.16 der DFA2-Analyse dieser Reihen zusammen mit den daraus ermittelten α -Werten. Alle Reihen zeigen langzeitkorreliertes Verhalten, welches sich in doppellogarithmischer Auftragung in Form einer Geraden mit Steigung α zeigt. Die Auftragung von $F(s)$ ist in beliebigen Einheiten (*arb. units = arbitrary units*), da nur die Steigungen α der Kurven von Bedeutung sind. Sie sind über $\alpha = 1 - \gamma/2$ mit dem Korrelationsexponent γ verknüpft. Abbildung (b) zeigt die Verteilungsfunktion $P_q(r)$ der jeweiligen Reihen (geschlossene Symbole) und der gemischten Reihen (offene Symbole), die zugehörigen gestreckt exponentiellen Fits (durchgezogene schwarze Linien) und die Poisson-Verteilungen (gestrichelte Linien) für die unkorrelierten gemischten Daten. Die $P_q(r)$ in (b) wurden über alle Quantile gemittelt für die $R_q > 3$ Jahre war und sofern das Quantil noch mehr als 50 Wiederkehrzeiten pro Reihe zugelassen hat. Damit sich die Kurven nicht überlappen, wurden die y -Achsen jeweils um einen Faktor 10 verschoben. Abbildung nach [Bunde 2005].

historischen und rekonstruierten Reihen aus Jahresdaten, welche in Abbildung 2.15 vorgestellt wurden (Sacramento, Baffin, New Mexico, Nil und die nördliche Hemisphäre). Alle Kurven zeigen in der DFA Langzeitkorrelationen. Abbildung 3.11(b) zeigt die Verteilungsfunktion $P_q(r)$ der repräsentativen langzeitkorrelierten hydrometeorologischen Zeitreihen (volle Symbole) aus Abbildung (a) im Vergleich mit $P_q(r)$ derselben Daten, wenn man diese vorher mischt, also die Korrelationen zerstört (offene Symbole). Die Kurven der geschlossenen Symbole (ungemischt) passen sehr gut mit den zugehörigen gestreckt exponentiellen Fits (durchgezogene schwarze Linien) überein. Die offenen Symbole stehen für die gemischten Daten und folgen wie erwartet der Poisson-Statistik (exponentieller Abfall, gestrichelte Linien). Große und kleine r/R_q -Werte kommen im ungemischten Fall auf Kosten von mittleren r/R_q deutlich häufiger vor als im gemischten Fall.

In [Altmann 2005] verwenden E. G. Altmann und H. Kantz eine andere Definition für die Wiederkehrintervalle, die aber mit unserer POT-Methode verwandt ist. Zur Verteilungsdichte $P(x)$ der langzeitkorrelierten Gauss-verteilten Zufallszahlen gibt man ein Intervall Δx um einen Werte x_c vor. Eine Wiederkehrzeit r wird nun definiert als die Dauer zwischen zwei aufeinanderfolgenden Ereignissen x_i und x_{i+r} , welche beide *in* dem Bereich Δx um das vor-

gegeben x_c liegen. Völlig analog zu Gl. (3.1) definiert das Integral

$$\frac{1}{R_c} := \int_{x_c - \Delta x}^{x_c + \Delta x} P(x) dx \quad (3.28)$$

den Kehrwert der zugehörigen Wiederkehrperiode. Umgekehrt passt man zu vorgegebenem R_c und festem x_c die Breite Δx an.

Bei festem Korrelationsexponent γ und vorgegebenem R_c (sie verwenden beispielhaft $R_c = 10, 100$ und 1000) finden Altmann und Kantz ebenfalls eine gestreckt exponentielle Verteilungsdichtefunktion, deren Exponent γ_c eine starke Abhängigkeit sowohl von x_c als auch von Δx aufweist. Liegt x_c in der Mitte von $P(x)$, zum Beispiel bei $x_c = 0$, und ist R_c sehr groß (d. h. Δx um x_c sehr klein), so ergibt sich für die Verteilungsdichte der Wiederkehrintervalle ebenfalls eine gestreckt exponentielle Funktion mit $\gamma_c = \gamma$. Wird nun das Intervall Δx vergrößert (d. h. R_c verkleinert), so geht die gestreckt exponentielle Verteilungsdichte über in die Poisson-Statistik für unkorrelierte Systeme ($\gamma = 1$). Lässt man aber in der selben Zeit bei festem Δx das x_c aus der Mitte an die Ränder von $P(x)$ wandern, so geht die Poisson-Verteilung wieder in die gestreckt exponentielle Funktion mit $\gamma_c = \gamma$ über.

Altmann und Kantz kommen daher zu dem Schluss, dass die Korrelationen im wesentlichen von Werten an den Rändern der Verteilungsdichte $P(x)$ getragen werden. Die Ergebnisse aus [Altmann 2005] stehen nicht im direkten Widerspruch zu unseren Ergebnissen, obwohl wir selbst für kleine R_q -Werte, das heisst q nahe bei 0, die gestreckt exponentielle Verteilungsdichte finden. Jedoch werden in unserer Definition auch bei kleinen R_q alle Randwerte x_i berücksichtigt. Leider führen Altmann und Kantz in [Altmann 2005] keine Finite-Size Untersuchungen durch, sodass Effekte dieser Art spekulativ bleiben.

3.2.3 Verteilungsdichte der kleinen Wiederkehrintervalle

Im Bereich $r < R_q$ sieht man für $P_q(r)$ in Abbildung 3.9 Abweichungen vom gestreckt exponentiellen Verlauf. Abbildung 3.12 zeigt nochmal genauer das linke Ende von $P_q(r)$ für die vier verschiedenen Verteilungsfunktionen zu je zwei γ -Werten ($\gamma = 0.2$ und 0.4) und drei großen q -Werten⁴ ($R_q = 150, 250$ und 500). Man kann sehen, dass $P_q(r)$ in (a) und (b) im Bereich $0.01 < r/R_q < 1$ und in (c) und (d) im Bereich $0.1 < r/R_q < 1$ eher einem Potenzgesetz mit einer Steigung von ungefähr $\gamma - 1$ in der doppellogarithmischen Darstellung erscheint (durchgezogene schwarze Linien), als dem gestreckt exponentiellen Verlauf gehorcht (gestrichelte Linien). Dieser *Potenzgesetz-Effekt* ist bei Gauss-verteilten (a) und exponentiell verteilten Zahlen (b) klarer ausgeprägt als bei Potenzgesetz-verteilten (c) oder Log-Normal-verteilten Zahlen (d). Ein geschlossener Ausdruck für die Verteilungsdichte $P_q(r)$ ist also eher von der Form

$$P_q(r) = \begin{cases} \frac{a_\gamma}{R_q} \exp[-b_\gamma (r/R_q)^\gamma] & \text{für } r \geq R_q \\ c_\gamma (r/R_q)^{-\delta} & \text{für } r < R_q, \end{cases} \quad (3.29)$$

wobei δ nahe bei $1 - \gamma$ zu liegen scheint. In den unteren beiden Bildern von Abbildung 3.12 sieht man für $r/R_q < 0.1$ zusätzlich ein starkes Abknicken nach oben. Dieses Abknicken ist eine Abweichung sowohl vom gestreckt exponentiellen Verlauf, als auch vom Potenzgesetzverhalten. Man sieht es etwas schwächer auch in den Abbildungen (a) und (b). Die Ursache

⁴Aufgrund der Auftragung r/R_q sind die Abweichungen nur bei großen R_q -Werten zu bemerken, da nur diese Kurven für $r = 1$ weit genug nach links reichen.

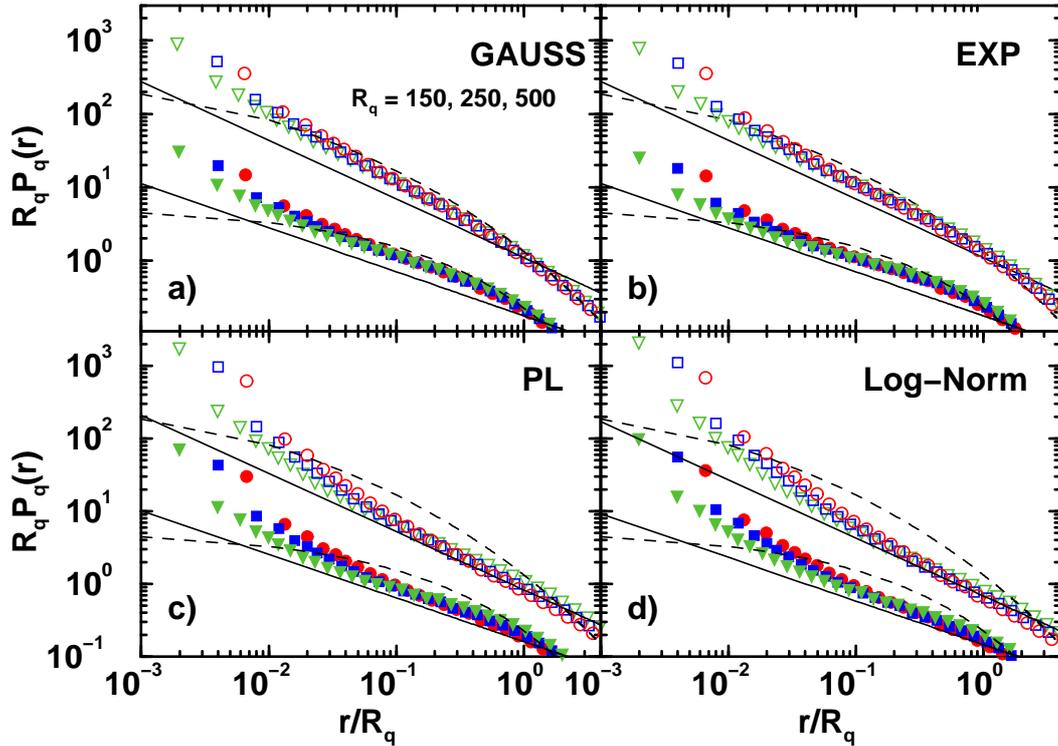


Abbildung 3.12: Verteilungsdichten der kleinen Wiederkehrzeiten. Abbildungen (a)-(d) zeigen das linke Ende der Verteilungsdichte $P_q(r)$ langzeitkorrelierter Daten der vier x -Verteilungen zu drei verschiedenen R_q -Werten, $R_q = 150$ (rote Kreise), 250 (blaue Quadrate) und 500 (grüne Dreiecke) und zwei γ -Werten, $\gamma = 0.2$ (volle Symbole) und 0.4 (offene Symbole), wobei die Kurven für $\gamma = 0.2$ um einen Faktor 10 nach oben verschoben wurden, um Überdeckungen zu vermeiden. Die gestrichelten Linien entsprechen den gestreckt exponentiellen Funktionen, die durchgezogenen Linien sind Potenzgesetze mit der Steigung $1 - \gamma$. In den Abbildungen (a) und (b) zeigt $P_q(r)$ näherungsweise ein Potenzgesetz-Verhalten im Bereich $0.01 < r/R_q < 1$, Abbildungen (c) und (d) zeigen dies nur etwa im Bereich $0.1 < r/R_q < 1$. In den Bereichen für noch kleinere r/R_q sieht man in allen vier Abbildungen einen Aufwärtsknick (in (a) am schwächsten). Die Ursache hierfür liegt in der Diskretheit der Wiederkehrzeiten, welche sich bei kleinen r -Werten bemerkbar macht. Das Potenzgesetz-Regime ist für große R_q deutlicher, da bei großen R_q der Diskretheitseffekt für kleine r/R_q später (d. h. weiter links) einsetzt. Abbildung nach [Eichner 2006b].

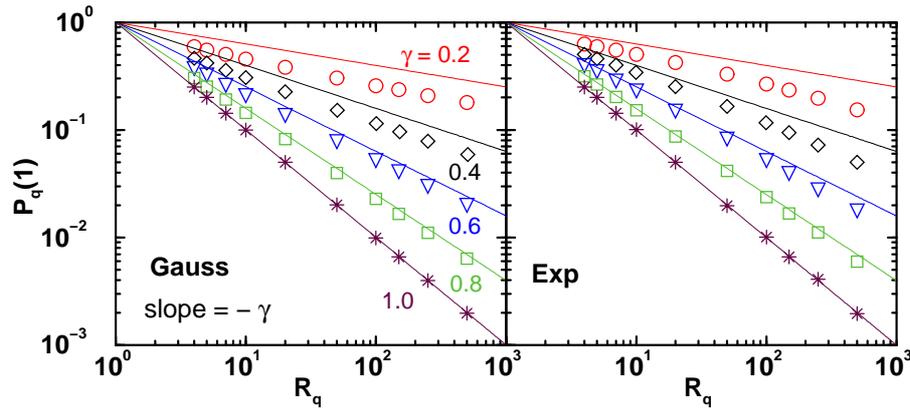


Abbildung 3.13: Verteilung der Wiederkehrzeiten der Länge $r = 1$ als Funktion von R_q bei langzeitkorrelierten Gauss-verteiltern (a) und exponentiell verteilten (b) Daten zu fünf verschiedenen γ -Werten ($\gamma = 1.0, 0.8, 0.6, 0.4$ und 0.2). Die farblich markierten Geraden haben jeweils die Steigung $-\gamma$. Offensichtlich gehorcht die relative Häufigkeit der 1er als Funktion von R_q einem Potenzgesetz mit Exponent $-\gamma$. Dies bedeutet, dass der Diskretheitseffekt in Abbildung 3.12 bei Gauss-Daten und exponentiellen Daten von q bzw. R_q unabhängig ist. Abbildung nach [Eichner 2006b].

hierfür liegt in der Diskretheit der r -Werte, die bei kleinen r -Werten stärker zum Tragen kommt, als bei großen. Für sehr kleine r -Werte ($r = 1, 2, \dots \leq 10$) müsste Gl. (3.29) durch eine diskrete Verteilung ersetzt werden. Die Summen über die diskreten r -Werte in Gl. (3.24) lassen sich außerdem für kleine r -Werte nicht sonderlich gut durch Integrale annähern. Als Folge dieses *Diskretheitseffekts* weichen die Werte von a_γ und b_γ aus Gl. (3.25) von ihrem idealen Wert für Gl. (3.26) ab. Der Effekt ist für $r = 1$ am größten, da hier der Diskretheitseffekt am stärksten ist, weil es keine Zwischenwerte wie $r = 0.5$ oder 0.01 gibt. Insgesamt sind einige der ersten Punkte davon betroffen und landen höher als es die theoretischen Kurven vorschreiben, da sich in ihnen auch die Wahrscheinlichkeiten aufrechnen, die den nicht-ganzzahligen r -Werten zuzuschreiben sind, welche im diskreten System nicht auftreten können, aber im kontinuierlichen Ansatz für $P_q(r)$ in Gl. (3.29) zugelassen werden.

Eine besondere Rolle spielen die Wiederkehrzeiten mit $r = 1$. Sie haben in der Praxis eine große Relevanz, da sie eine Antwort auf die Frage geben, ob das nächste Ereignis wieder ein Extremereignis mit $x > q$ ist. Also z. B. mit welcher Wahrscheinlichkeit im nächsten Jahr wieder mit einer Flut zu rechnen ist. Abbildung 3.13 zeigt die relativen Häufigkeiten der 1er als Funktion von R_q für Gauss- und exponentiell verteilte Zufallszahlen. Da die Bin-Breite im diskreten Fall 1 ist, muss die Wahrscheinlichkeitsdichte $P_q(r)$ nur mit 1 multipliziert anstatt integriert zu werden. Gesetzt den Fall, dass soeben ein Ereignis x größer q stattgefunden hat, gibt der Plot die Wahrscheinlichkeit an, dass der nun folgende r -Wert ein 1er ist. Das heißt im übertragenen Sinne, mit welcher Wahrscheinlichkeit im nächsten Jahr wieder mit einer Katastrophe $x > q$ zu rechnen ist. Im unkorrelierten Fall (Sternchen) ist die Wahrscheinlichkeit für $x > q$ konstant $1/R_q$ für jedes Ereignis, so wie man es in der Abbildung ablesen kann. Im korrelierten Fall ist diese Wahrscheinlichkeit aufgrund der Persistenz größer. Je stärker die Reihe korreliert ist, umso wahrscheinlicher ist mit einem $r = 1$ zu rechnen.

Die Wahrscheinlichkeit für zwei aufeinander folgende Extremereignisse folgt näherungsweise ebenfalls einem Potenzgesetz, aber eher der Art $R_q^{-\gamma}$ (farbige Referenzlinien in Abbildung

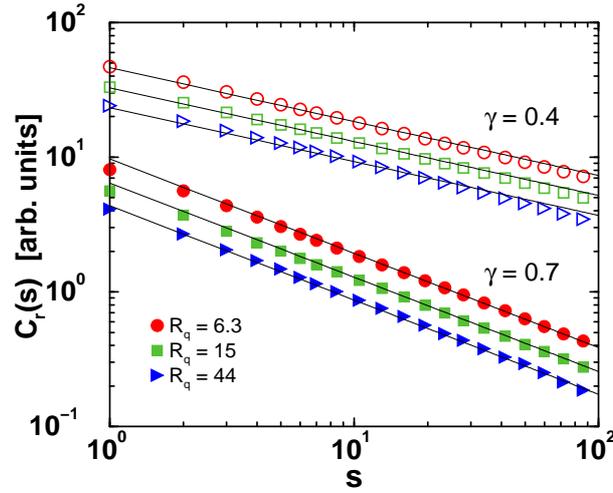


Abbildung 3.14: Autokorrelationsanalyse der Wiederkehrintervalle Gauss-verteilter langzeitkorrelierter Daten. Die x -Werte der offenen Symbole waren mit $\gamma = 0.4$ langzeitkorreliert. Die zugehörigen Reihen der Wiederkehrzeiten (R_q -Werte siehe Legende) zeigen ebenfalls eine Potenzgesetz-artig abfallende Autokorrelationsfunktion mit nahezu identischem γ . Die y -Achse ist in beliebigen Einheiten (*arbitrary units*) aufgetragen, da hier nur das Skalenverhalten interessiert. Die durchgezogenen oberen drei Linien haben die Steigung $-\gamma = -0.4$. Die vollen Symbole zeigen dieselben Kurven, nur zu schwächer korrelierten Ausgangsdaten ($\gamma = 0.7$). Auch hier zeigen die Wiederkehrintervalle dasselbe Abfallen der Autokorrelationsfunktion wie schon die Ausgangsdaten. Die Wiederkehrintervalle langzeitkorrelierter Daten sind also ebenfalls langzeitkorreliert, und zwar in etwa mit demselben Korrelationsexponenten, der schon die Ausgangsdaten charakterisiert. Als Konsequenz folgen in der Reihe der Wiederkehrintervalle auf lange Intervalle eher wieder lange und auf kurze Intervalle eher kurze, und es bilden sich Epochen mit nur wenigen Extremereignissen und Cluster gehäufte Extremereignisse. Abbildung aus [Bunde 2004c].

3.13). Die Abweichungen der Symbole von diesen Werten können bei starken Korrelationen möglicherweise ähnlich begründet werden wie bereits das Verhalten der Varianzen (siehe Abbildung 3.6), also mittels Finite-Size oder Diskretheit. In echten Daten sind Ereignisse, die zeitlich dicht beieinander liegen jedoch oft von Kurzzeitkorrelationen überlagert, sodass ein reiner Langzeitkorrelationseffekt nicht ausreicht, um Vorhersagen für Ereignisse in naher Zukunft zu verbessern. Hat man es aber mit Jahresdaten zu tun, ist der Einfluss von Kurzzeitkorrelationen deutlich abgeschwächt.

3.2.4 Korrelationseigenschaften der Reihen von Wiederkehrintervallen

Aus den Untersuchungen an der Verteilungsfunktion $P_q(r)$ geht hervor, dass im langzeitkorrelierten Fall bedeutend mehr kleine Wiederkehrzeiten ($r < R_q$) und große Wiederkehrzeiten ($r > R_q$) auftreten als im unkorrelierten Fall. Die Verteilungsfunktion (3.29) verrät jedoch nicht, ob die Reihen der Wiederkehrzeiten selbst irgendwelche Korrelationen aufweisen oder nicht. Um mögliche Korrelationen in den Wiederkehrzeiten zu bestimmen, wurden sowohl die

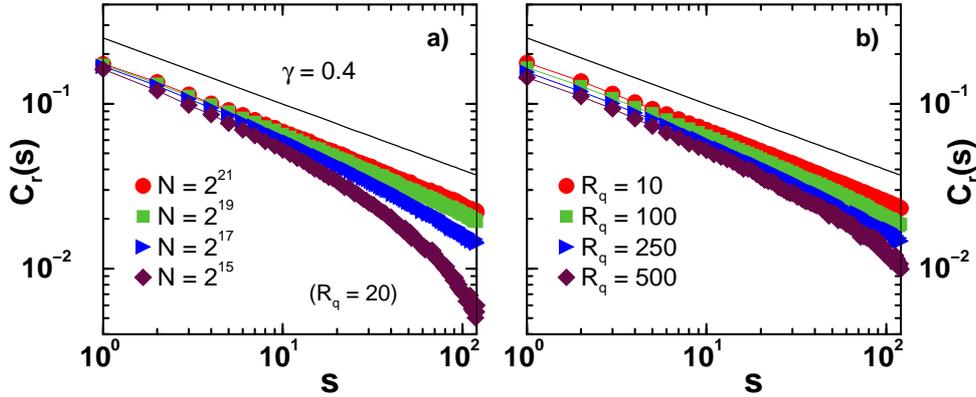


Abbildung 3.15: Finite-Size Untersuchung an der Autokorrelationsfunktion der Wiederkehrzeiten Gauss-verteilter langzeitkorrelierter ($\gamma = 0.4$) Daten. In Abbildung (a) wurde zu $R_q = 20$ die Reihengänge der Datensätze von $N = 2^{21} \simeq 2 \times 10^6$, $2^{19} \simeq 5 \times 10^5$, $2^{17} \simeq 10^5$ bis $2^{15} \simeq 3 \times 10^4$ variiert. Um gleiche Statistik zu gewähren, wurde die Anzahl der Reihen jeweils um einen Faktor vier erhöht; waren es bei $N = 2^{21}$ nur 150 Konfigurationen, über die gemittelt wurde, so waren es bei 2^{15} dafür 9600 Konfigurationen. Je kürzer die ursprüngliche Messreihe ist, desto schneller fällt die Autokorrelationsfunktion vom Potenzgesetz-artigen Verlauf ab. Die Statistik reicht nicht aus, damit $C_r(s)$ das korrekte Korrelationsverhalten wiedergibt. In Abbildung (b) wurde die Reihengänge mit $N = 2^{21}$ konstant gehalten, aber die q -Werte variiert. Die zugehörigen Wiederkehrperioden sind 10, 100, 250 und 500. Auch hier macht sich die reduzierte Anzahl der Wiederkehrzeiten $N_r \simeq N/R_q$ mit größer werdendem R_q durch ein Abfallen vom Potenzgesetzverlauf bemerkbar. Abbildung nach [Eichner 2006b].

Autokorrelationsfunktion (siehe dazu Kapitel 2.2.1)

$$C_r(s) = \frac{\text{cov}_r(s)}{\text{var}_r} = \frac{\langle (r_j - R_q)(r_{j+s} - R_q) \rangle}{\langle (r_j - R_q)^2 \rangle} = \frac{\langle r_j r_{j+s} \rangle - R_q^2}{\langle r_j^2 \rangle - R_q^2} \quad (3.30)$$

als auch die Fluktuationfunktionen DFA0 und DFA1 (siehe Kapitel 2.2.3) der r -Werte bestimmt. Abbildung 3.14 zeigt die Ergebnisse zur Autokorrelationsanalyse bzw. zur Autokovarianzfunktion $\text{cov}_r(s)$ für Gauss-verteilte Zufallszahlen mit zwei γ -Werten ($\gamma = 0.4$ offene Symbole und $\gamma = 0.7$ geschlossene Symbole) und drei verschiedenen q -Werten mit $R_q = 6.3$, 15 und 44. Das lineare Verhalten in der doppel-logarithmischen Auftragung ist ein Indiz für Langzeitkorrelationen innerhalb der r -Werte. Die Steigungen der parallelen Geraden entsprechen den Korrelationsexponenten der r -Werte. Es stellt sich heraus, dass diese wieder nahezu identisch zu γ sind. Die Reihen der Wiederkehrzeiten sind also mit demselben Korrelations-exponent γ langzeitkorreliert wie bereits die ursprüngliche Reihe der x -Werte. Demnach sind lange und kurze Wiederkehrintervalle nicht zufällig angeordnet. Vielmehr folgen auf lange Wiederkehrzeiten eher wieder lange und auf kurze Wiederkehrzeiten eher wieder kurze. Das hat zur Folge, dass sich zum einen Epochen ausprägen, in denen es so gut wie keine extremen Ereignisse gibt, nämlich wenn lange Wiederkehrzeiten dominieren. Andererseits bilden sich auch Epochen, in denen Extremereignisse binnen kurzer Zeit sehr häufig auftreten, also wenn viele kurze r -Werte aufeinander folgen. Es bilden sich dann sogenannte *Cluster* von extremen Ereignissen. Im übertragenen Sinne gibt es dann viele Katastrophen in kurzer Zeit hintereinander.

Aus Kapitel 2.2.1 ist bekannt, dass für eine sinnvolle Autokorrelationsanalyse eine enorme

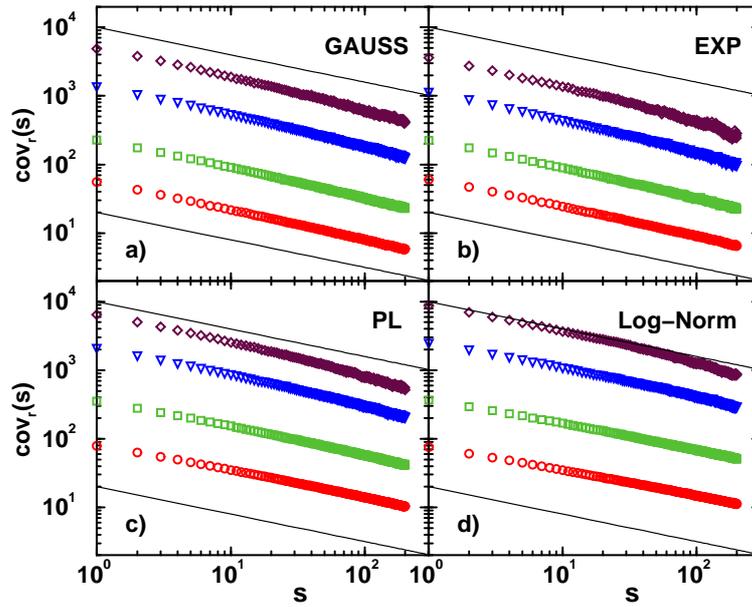


Abbildung 3.16: Autokovarianzfunktion der Wiederkehrintervalle langzeitkorrelierter Daten ($\gamma = 0.4$) der vier Verteilungen (a) Gauss, (b) exponentiell, (c) Potenzgesetz und (d) Log-normal) zu vier Wiederkehrperioden: $R_q = 10$ (rote Kreise), 20 (grüne Quadrate), 50 (blaue Dreiecke) und 100 (bordeaux Rauten). Die durchgezogenen Linien zeigen die Steigung $-\gamma = -0.4$. Die Wiederkehrzeiten aller vier Verteilungen zeigen eine sehr ähnliche, ebenfalls Potenzgesetz-artig abfallende Autokorrelationsfunktion mit nahezu dem selben γ -Wert der ursprünglichen x -Werte. Finite-Size Untersuchungen an Gauss-verteilten Daten sind in Abbildung 3.15 diskutiert worden. Abbildung nach [Eichner 2006b].

Statistik notwendig ist. Da die Reihe der Wiederkehrzeiten um den Faktor R_q kürzer ist als die Originalreihe $(x_i)_{i=1}^N$, bleibt für die Autokorrelationsfunktion nur eine Reihe der Länge $N_r \simeq N/R_q$ übrig. Abbildung 3.15 zeigt den Einfluss der Systemgrößen auf die Autokorrelationsfunktion $C_r(s)$. In (a) wurden zu langzeitkorrelierten ($\gamma = 0.4$) Gauss-Daten mit vier verschiedenen Reihenlängen die Wiederkehrintervalle zu einem q -Wert mit $R_q = 20$ bestimmt. Je kürzer die Reihe ist, desto schneller verlässt $C_r(s)$ das Potenzgesetzverhalten und fällt ab, da die Statistik nicht genug große r -Werte hergibt. In Abbildung (b) ist zur Reihenlänge $N = 2^{21}$ die Autokorrelationsfunktion für vier verschiedene R_q -Werte geplottet. Auch hier fällt $C_r(s)$ mit größer werdendem R_q ab, da eine ausreichende Statistik für sehr große $r \gg R_q$ (auch bei $R_q = 500$) auf einer endlichen Reihe nicht erreicht werden kann. Als Konsequenz können die Korrelationen nicht durchschlagen, die Reihe bleibt auf großen Skalen verrauscht und $C_r(s)$ fällt rapide ab. Um mittels der Autokorrelationsfunktion die Korrelationen der Wiederkehrzeiten vernünftig bestimmen zu können, sind sehr lange Datensätze nötig, wie man sie in der Praxis jedoch so gut wie nie zur Verfügung hat.

Abbildung 3.16 zeigt die Autokovarianzfunktion $cov_r(s)$ der Wiederkehrzeiten für alle vier untersuchten Verteilungsfunktionen zu ebenfalls vier R_q -Werten. Alle Kurven zeigen ein Potenzgesetz-artiges Abfallen mit derselben Steigung und indizieren Langzeitkorrelationen. Die Kurven für $R_q = 100$ scheinen bereits durch Finite-Size Effekte beeinflusst zu sein. Unabhängig von der zugrunde liegenden Verteilungsfunktion der x -Werte zeigen die Wiederkehrintervalle

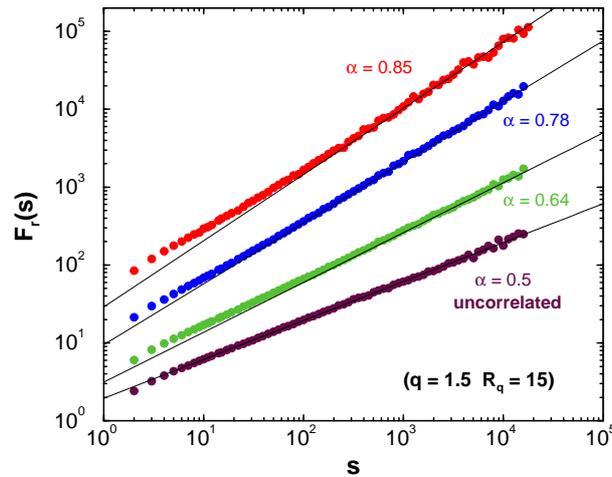


Abbildung 3.17: Fluktuationsanalyse der Wiederkehrzeiten: Um Aussagen über das Korrelationsverhalten der Wiederkehrzeiten auf größeren Zeitskalen zu treffen, ist hier das Ergebnis einer Fluktuationsanalyse (DFA0) für vier verschieden stark langzeitkorrelierte Reihen mit $q = 1.5$ gezeigt ($\gamma = 0.1, 0.4, 0.7$ und 1.0 (unkorreliert)). Die DFA0 (oder auch FA) hat gegenüber der Autokorrelationsfunktion den Nachteil, dass sie auf kleinen Skalen s bei verrauschten Daten nicht das korrekte Korrelationsverhalten wiedergeben kann. Weißes Rauschen ist den Langzeitkorrelationen überlagert, welche erst auf großen Skalen zur Geltung kommen. Die resultierenden Korrelationsexponenten $\gamma = 2 - 2\alpha$ aus der Steigung α der DFA0-Kurven weichen mit zunehmender Korrelationsstärke (also kleinerem γ) stärker vom ursprünglichen γ ab: $\alpha = 0.5$ ($\gamma = 1.0$) steht für unkorreliertes Verhalten, $\alpha = 0.64$ ergibt ein $\gamma = 0.72$, also gering schwächer korreliert als die Ursprungsdaten mit $\gamma = 0.7$, $\alpha = 0.78$ ergibt ein $\gamma = 0.44$ ($\gamma = 0.4$) und $\alpha = 0.85$ ergibt ein $\gamma = 0.3$, was jedoch recht deutlich von $\gamma = 0.1$ abweicht. Trotzdem ist in allen Kurven ein eindeutiges Langzeitverhalten, insbesondere auf großen Skalen auszumachen. Abbildung nach [Gell-Mann 2003].

dasselbe Korrelationsverhalten wie die Ausgangsdaten.

Da Autokorrelationsanalysen starken Finite-Size Effekten unterworfen sind, soll hier noch die DFA Verwendung finden, welche etwas robuster gegenüber knappen Statistiken ist und in der Praxis sehr erfolgreich eingesetzt wird, um Langzeitkorrelationen zu detektieren. Die DFA, vorgestellt in Kapitel 2.2.3, geht außerdem mit ihren Skalen s deutlich weiter als die Autokorrelationsfunktion. Sie ist also für Aussagen über das echte Langzeitverhalten geeigneter als die Autokorrelationsfunktion. Leider hat sie jedoch den entscheidenden Nachteil, dass in ihr das Rauschen nicht wie in der Autokorrelationsfunktion herausgemittelt wird, sondern in der Fluktuationsfunktion den Korrelationen überlagert ist und so auf kleinen Skalen s eher die für unkorrelierte Daten charakteristische Steigung $\alpha = 0.5$ liefert. Hier wollen wir uns daher auf die Betrachtung der großen Skalen beschränken, da die kleinen Skalen schon mittels der Autokorrelationsfunktion ausreichend beschrieben wurden.

Abbildung 3.17 zeigt die DFA0-Ergebnisse für die Reihen der Wiederkehrintervalle mit $R_q = 15$ von langzeitkorrelierten Gauss-Daten mit $\gamma = 0.1, 0.4, 0.7$ und 1.0 (unkorreliert). Zwischen dem Korrelationsexponent γ und dem Fluktuationsexponent α besteht der Zusammenhang $\gamma = 2 - 2\alpha$ (siehe Kapitel 2.2.3). Die gefitteten Steigungen für das rechts-asymptotische Verhalten, die man dem Bild entnehmen kann, liefern numerische γ -Werte, die recht nahe am

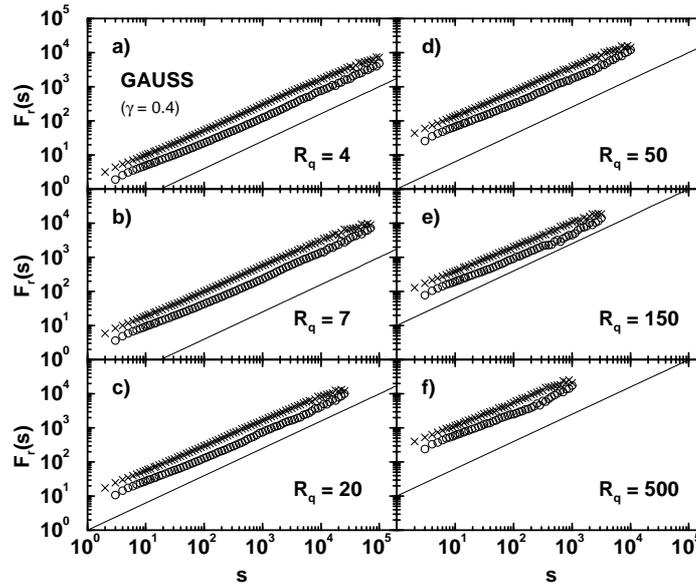


Abbildung 3.18: Fluktuationsanalyse (DFA0: \times , DFA1: \circ) der Wiederkehrintervalle langzeitkorrelierter ($\gamma = 0.4$) Gauss-verteilter Daten zu (a) $R_q = 4$, (b) 7, (c) 20, (d) 50, (e) 150 und (f) 500. Asymptotisch erreichen die DFA1 Kurven in allen 6 Graphen die Steigung $\alpha = 0.8$ (durchgezogene Linien), was einem $\gamma = 2 - 2\alpha = 0.4$ entspricht. Auf kleinen Skalen sind die Reihen der Wiederkehrzeiten recht stark verrauscht, was dort zu einem kleineren Skalensexponent α führt. Für kleine Skalen ist daher die Autokorrelationsfunktion geeigneter als die DFA.

ursprünglichen γ liegen: $\gamma = 1.0$ (steht für unkorreliertes Verhalten), $\gamma = 0.72$ (also gering schwächer korreliert als die Ursprungsdaten mit $\gamma = 0.7$), $\gamma = 0.44$ (anstatt $\gamma = 0.4$) und $\gamma = 0.3$ (was recht deutlich von $\gamma = 0.1$ abweicht). Ein γ -Wert von 0.1 gibt sehr starke Langzeitkorrelationen an und ist bereits dicht am nichtstationären Verhalten einer Reihe. Die Korrelationen der Wiederkehrzeiten können dadurch beeinträchtigt werden.

Der Einfluss der Quantilgröße auf die DFA-Kurven ist in Abbildung 3.18 gezeigt. Dargestellt sind die DFA0- und DFA1-Kurven der Wiederkehrzeiten für sechs verschiedene R_q -Werte von Gauss-verteilter Daten mit $\gamma = 0.4$. Die Verkürzung der Reihenlänge N_r der r -Werte führt dazu, dass das rechts-asymptotische Verhalten durch das Rauschen auf kleinen Skalen ($\alpha \simeq 0.5$) nicht korrekt bestimmt werden kann. In diesen Bereichen eignet sich die Autokorrelationsfunktion besser zur Bestimmung des korrekten Korrelationsverhaltens. Die durchgezogene Linie hat die Steigung $\alpha = 0.8$ (entspricht $\gamma = 0.4$) und man bekommt den Eindruck, dass die Kurven (insbesondere DFA1) sich asymptotisch dieser Steigung tatsächlich annähern. Das zentrale Ergebnis aber ist, dass auch die DFA ein eindeutiges Skalenverhalten liefert, auch wenn die hier bestimmten Exponenten etwas schwächere Korrelationen beschreiben.

Durch die Langzeitkorrelationen, die sich auch in den Wiederkehrintervallen wiederfinden, nimmt die Vergangenheit Einfluss auf die Zukunft. Ein zukünftiges Wiederkehrintervall ist vom Betrag her dem vorhergehenden Intervall ähnlich. Auf einen großen r -Wert folgt eher ein großer, auf einen kleinen r -Wert eher ein kleiner. Dies steht nun völlig im Widerspruch zur klassischen Poisson-Statistik, in der der darauffolgende r -Wert im Mittel immer R_q ist. Dieser Einfluss soll im nächsten Abschnitt geklärt werden.

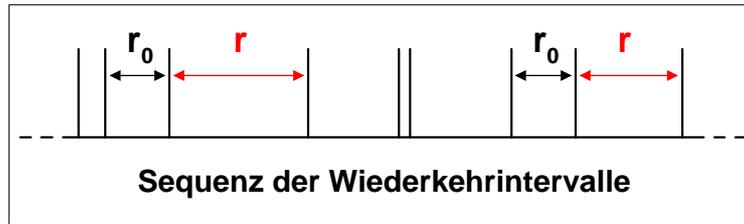


Abbildung 3.19: Definition bedingter Wiederkehrzeiten: Es werden nur diejenigen Wiederkehrzeiten r (rot) betrachtet, welche in der Reihe der Wiederkehrzeiten unmittelbar auf eine Wiederkehrzeit der Länge r_0 folgen.

3.2.5 Bedingte Wiederkehrintervalle

Will man den Effekt der Langzeitkorrelationen zur Verbesserung von Prognosen und Risikoabschätzungen ausnutzen, muss der Einfluss der Vergangenheit auf die Zukunft berücksichtigt werden. Mittels *bedingter* Wiederkehrintervalle kann der Memory-Effekt, d. h. die Abhängigkeit künftiger Wiederkehrintervalle von vorhergehenden Wiederkehrintervallen, quantifiziert werden. Abbildung 3.19 erläutert das Prinzip der bedingten Wiederkehrintervalle. Es werden nur solche r -Werte berücksichtigt, welche in der Reihe der Wiederkehrintervalle unmittelbar auf einen r -Wert einer gewissen Größe r_0 (die “Bedingung”) folgen. Aufgrund der Korrelationen in den Wiederkehrintervallen folgen auf große r_0 eher große r und auf kleine r_0 eher kleine r . Dies wird deutlich, wenn man über die bedingten r -Werte zu vorgegebenem r_0 mittelt, also die *bedingte Wiederkehrperiode* $R_q(r_0)$ bestimmt. $R_q(r_0)$ ist die im Mittel zu erwartende Wiederkehrzeit, wenn die Dauer zwischen den beiden vorherigen Extremereignissen r_0 beträgt.

Abbildung 3.20 zeigt die bedingten Wiederkehrperioden $R_q(r_0)$ als Funktion von r_0 in Einheiten der unbedingten Wiederkehrperiode R_q zu vier verschiedenen q -Werten bei Gaussverteilten Daten mit $\gamma = 0.4$ (volle Symbole) und gemischten, also unkorrelierten Daten (offene Symbole). Der Memory-Effekt der Langzeitkorrelationen wird daran deutlich, dass der bedingte Mittelwert $R_q(r_0)$ für kleine Bedingungen ($r_0 < R_q$) kleiner als R_q ausfällt, und für große Bedingungen ($r_0 > R_q$) mit anwachsendem r_0 sehr stark anwächst und deutlich größer als R_q wird. So folgt auf einen r -Wert der Größe $r_0 = 0.1R_q$ im Mittel ein r -Wert der Größe $R_q(r_0) \simeq 0.75R_q$ und auf einen r -Wert der Größe $r_0 = 10R_q$ im Mittel ein r -Wert der Größe $R_q(r_0) \simeq 2R_q$. Die bedingte Wiederkehrperiode ist also auch ein Maß, um abzuschätzen, ob man sich in einer Ruhephase oder in einem Cluster gehäufte Extremereignisse befindet. Mischt man die Daten zuvor, so werden die Korrelationen zerstört und es gilt $R_q(r_0) \equiv R_q$ (offene Symbole, konstant bei $R_q(r_0)/R_q = 1$). Dass die Kurven der korrelierten Daten bei $r_0/R_q = 1$ bereits über $R_q(r_0)/R_q = 1$ liegen, ist eine Konsequenz der Diskrettheit der r -Werte: der kleinste r -Wert hat den Wert 1 und Zwischenwerte wie 0.5 oder 0.01 kommen nicht vor. Folglich sind die bei $r_0 < R_q$ häufiger auftretenden 1er etwas zu groß und erhöhen $R_q(r_0)$ stärker gegenüber R_q . Dieser Effekt verschwindet für große r_0 -Werte, da hier aufgrund der Korrelationen deutlich weniger 1er auftreten. In der gewählten Darstellung in Einheiten von R_q skalieren die $R_q(r_0)$ -Kurven für verschiedene Quantile q (hier $R_q = 10, 20, 50$, und 100). Dieses Skalenverhalten hat den Vorteil, dass man zur Verbesserung der Statistik bei schlechter Datenlage, wie es in echten Daten meist der Fall ist, über verschiedene q -Werte oder q -Bereiche mitteln kann, um glattere Ergebnisse zu erhalten. Außerdem wurden

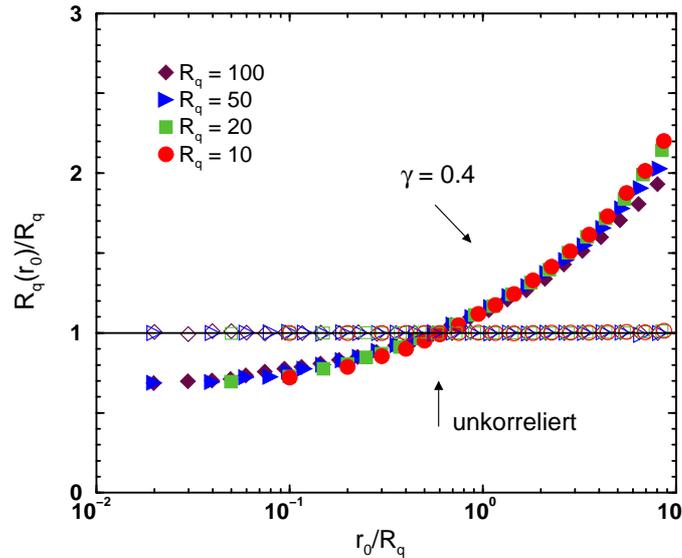


Abbildung 3.20: Bedingte Wiederkehrperioden: Die Abbildung zeigt die arithmetischen Mittelwerte aller bedingten Wiederkehrzeiten zu vier verschiedenen Quantilen ($R_q = 10$ rote Kreise, $R_q = 20$ grüne Quadrate, $R_q = 50$ blaue Dreiecke und $R_q = 100$ braune Rauten) von Gauss-verteilten langzeitkorrelierten Daten ($\gamma = 0.4$, volle Symbole) und den gemischten Daten (unkorreliert, offene Symbole) als Funktion der vorherigen Intervalllänge r_0 . Die bedingten Wiederkehrperioden $R_q(r_0)$ und die Bedingungen r_0 sind in Einheiten von R_q aufgetragen und bringen so die vier Kurven zu einem Datenkollaps. Der Anstieg der Kurven mit größer werdendem r_0 bedeutet, dass auf große r -Werte häufiger große r -Werte folgen, und das Abfallen der Kurven für kleine r_0 unter 1 bedeutet analog, dass auf kleine r -Werte eher wieder kleine r -Werte folgen. Dieses Verhalten ist eine direkte Auswirkung der Langzeitkorrelationen auf die Wiederkehrintervalle und erklärt das korrelationsbedingte Clustern extremer Ereignisse. Abbildung nach [Bunde 2004a].

die $R_q(r_0)$ -Kurven für die Auftragung logarithmisch gebinnt. Das heißt, es wurden bei großen r_0 -Werten mehrere zusammengefasst, da sonst die schlechter werdende Statistik die Kurven sehr stark verrauschen würde. Die Anzahl der r_0 -Werte pro Bin wächst dabei exponentiell an, sodass die logarithmische Auftragung äquidistante Abstände bei großen r_0 -Werten liefert. Diese Form des Binnens wurde auch bei den Histogrammen in den vorherigen Abschnitten angewendet, wenn die Auftragung der x -Achse logarithmisch gewählt war. Die Ergebnisse basieren auf einer Statistik von 150 Konfigurationen mit jeweils 2^{21} Datenpunkten.

Abbildung 3.21 zeigt die bedingten Wiederkehrperioden auch für die anderen drei Verteilungsfunktionen. Die gewählten q -Werte entsprechen $R_q = 5, 10, 50$ und 250 . Man sieht zwar Unterschiede im exakten Verhalten der $R_q(r_0)$ -Kurven, aber bei allen vier Verteilungen ist die Tendenz dieselbe: 1. Nach kurzen Wiederkehrzeiten ist eher mit einer weiteren kurzen Wiederkehrzeit zu rechnen und Extremereignisse können clustern. 2. Nach einer langen Wiederkehrzeit ist eher mit einer langen Wiederkehrzeit zu rechnen und Extremereignisse treten seltener auf. In Abbildung 3.21(e) sind nur vier $R_q(r_0)$ -Kurven für $R_q = 100$ gezeigt, aber jede für eine der vier Verteilungsfunktionen. In dieser doppellogarithmischen Auftragung erscheint das Anwachsen von $R_q(r_0)$ mit r_0 für Potenzgesetz-verteilte Zahlen und Log-normal-verteilte Zahlen fast linear zu sein, was wiederum einem Potenzgesetzverhalten im funktionellen Zusammenhang von $R_q(r_0)$ entsprechen würde.

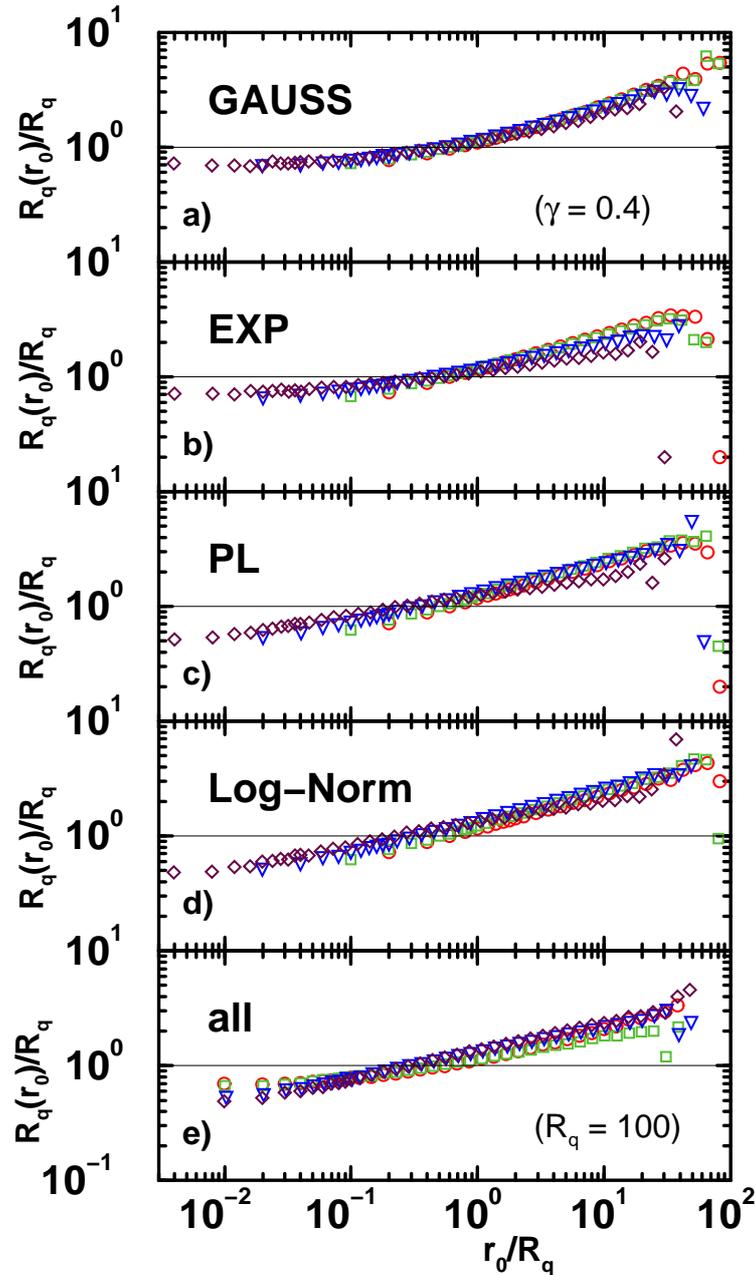


Abbildung 3.21: Bedingte Wiederkehrperioden als Funktionen von r_0 bei langzeitkorrelierten Daten ($\gamma = 0.4$) der vier untersuchten Verteilungen. Abbildung (a) zeigt das Ergebnis für Gauss-verteilte Zahlen zu vier vorgegebenen Wiederkehrperioden $R_q = 5$ (rote Kreise), 10 (grüne Quadrate), 50 (blaue Dreiecke) und 250 (bordeaux Rauten). Der Korrelationseffekt ist deutlich zu erkennen: auf große r_0 sind im Mittel deutlich größere r -Werte als R_q zu erwarten und auf kleine r_0 kleiner r -Werte. Durch die Skalierung mit der unbedingten Wiederkehrperiode R_q werden die vier Kurven zum Kollaps gebracht. Abbildung (b) zeigt dieselben Resultate für exponentiell verteilten Daten, (c) für Potenzgesetz-verteilten Daten und (d) für Log-Normal-verteilte Daten. Alle vier zeigen den korrelationsbedingten Memory-Effekt durch einen sehr ähnlichen Verlauf der Kurven. In Abbildung (e) sind nochmal zu $R_q = 100$ die Kurven der vier Verteilungen (Gauss: rote Kreise, exponentiell: grüne Quadrate, Potenzgesetz: blaue Dreiecke, Log-Normal: bordeaux Rauten) eingezeichnet. Es gibt zwar Unterschiede in den Krümmungen, aber die Tendenz ist bei allen vier Verteilungen mit selbem $\gamma = 0.4$ dieselbe. Abbildung nach [Eichner 2006b].

Um das Clustern und die Ruhephasen auch in oft viel zu kurzen gemessenen oder rekonstruierten Datensätzen studieren zu können, muss man die Statistik verbessern, indem man bestimmte r_0 -Bereiche zusammenfasst. Anstatt r_0 -Werte vorzugeben, kann man die Bedingung auch wie folgt formulieren: Es werden nur die r -Werte berücksichtigt, welche unmittelbar auf einen r_0 -Wert folgen, der größer oder kleiner als ein vorgegebener Schwellwert ist. Als Schwellwert bietet sich R_q oder auch der Median r^* der r -Werte an. Man kann die größeren r -Werte dann mit $+$ und die kleineren mit $-$ indizieren. Für die bedingten Wiederkehrperioden bedeutet das, dass man ein R_q^+ bzw. ein R_q^- erhält. Der Median hat gegenüber R_q als Bedingung den Vorteil, dass sich R_q^+ bzw. R_q^- aus derselben Anzahl an r^+ -Werten wie r^- -Werten zusammensetzen und damit die gleiche statistische Güte haben. Abbildung 3.22(a) zeigt zwei Werte für R_q^+ und R_q^- bei Gauss-verteiltern langzeitkorrelierten Daten ($\gamma = 0.4$, siehe x -Achse) mit dem Median r^* als Bedingung. Die Langzeitkorrelationen lassen die Werte in etwa symmetrisch um $R_q^{+/-}/R_q = 1$ bei etwa $R_q^+/R_q \simeq 1.2$ und $R_q^-/R_q \simeq 0.8$ erscheinen. Variiert man nun γ und damit die Stärke der Korrelationen, so erhält man die beiden Kurven für R_q^+ und R_q^- als Funktion von γ , wie sie in Abbildung (b) gezeigt sind. Die beiden Kurven beruhen für jeden γ -Wert (in 0.05er Schritten) auf 150 Reihen künstlicher Gauss-Daten mit Reihenlängen $N = 2^{21}$. Abbildung (c) zeigt die Resultate für die echten Daten, die bereits in Kapitel 2.3.2 und in Abbildung 3.11 vorgestellt wurden. Man sieht, dass die Werte der bedingten Wiederkehrperioden innerhalb der Fehlerbalken der künstlichen Jahresdaten (schwarze T-förmige Linien) liegen. Um aus den echten Daten trotz der geringen Statistik noch Ergebnisse für R_q^+ bzw. R_q^- zu bekommen, wurden die Ergebnisse für jede Reihe über einen q -Bereich gemittelt, was aufgrund des Datenkollaps in Abbildung 3.20 gerechtfertigt ist. Dabei wurde der q -Bereich jeweils so gewählt, dass der kleinste q -Wert noch eine Wiederkehrperiode von $R_q > 3$ Jahre ergab und der größte q -Wert noch mindestens 50 Wiederkehrzeiten in jeder Reihe lieferte. Abbildung (d) zeigt mit den offenen Symbolen noch die Ergebnisse für die gemischten (also unkorrelierten) echten Daten. In diesem Fall sollte $R_q^+ \simeq R_q^- \simeq R_q$ gelten, was das Bild auch bestätigt.

In Abbildung 3.23 wurden die bedingten Wiederkehrperioden R_q^+ und R_q^- noch erweitert, indem auch der vor-vorhergehende Wert r_{-1} berücksichtigt wurde. Die beiden neuen Größen R_q^{++} bzw. R_q^{--} setzen sich aus all jenen r -Werten zusammen, deren unmittelbarer Vorgänger r_0 und dessen Vorgänger r_{-1} beide größer bzw. kleiner als R_q waren. Es wurde für dieses Bild als Bedingung R_q statt des Medians r^* gewählt, da bei den künstlichen Daten kein so großer Wert auf eine ausgewogene Statistik gelegt werden muss wie bei den echten Daten. Die Statistik kann hier nahezu beliebig verbessert werden. Da der Median kleiner als R_q ist (aufgrund der korrelationsbedingten riesigen Anzahl an r -Werten mit $r = 1$), werden durch die Bedingung $r_0 > R_q$ große r -Werte stärker betont als durch die Bedingung $r_0 > r^*$. R_q^+ wird demnach im stark korrelierten Fall deutlich größer, wenn es mit $r_0 > R_q$ bedingt ist, anstatt mit $r_0 > r^*$. Vergleicht man R_q^+ aus Abbildung 3.23(a) mit R_q^+ Abbildung 3.22(b), so sieht man nochmal den unterschiedlichen Einfluss der Bedingung r_0 . In Abbildung 3.23(a) sind die Ergebnisse von R_q^{++} , R_q^+ , R_q^- und R_q^{--} für Gauss-verteilte Daten zu $R_q = 50$ in Einheiten von R_q gegenüber γ aufgetragen. Der Korrelationseffekt (das Auseinanderlaufen der \pm -Kurven) ist bei den zweifach bedingten Größen R_q^{++} und R_q^{--} noch stärker als bei den einfach bedingten R_q^+ und R_q^- . Eine gemischte Größe, also so etwas wie R_q^{+-} oder R_q^{-+} würde den Memory-Effekt nur geschwächt zeigen, da sie der Erhaltungsneigung der langzeitkorrelierten r -Reihe nicht entspricht. Deshalb wurde einer solchen Größe kein Interesse gewidmet. Die durchgezogenen schwarzen Linien liegen bei 1 und deuten die Lage für den unkorrelierten

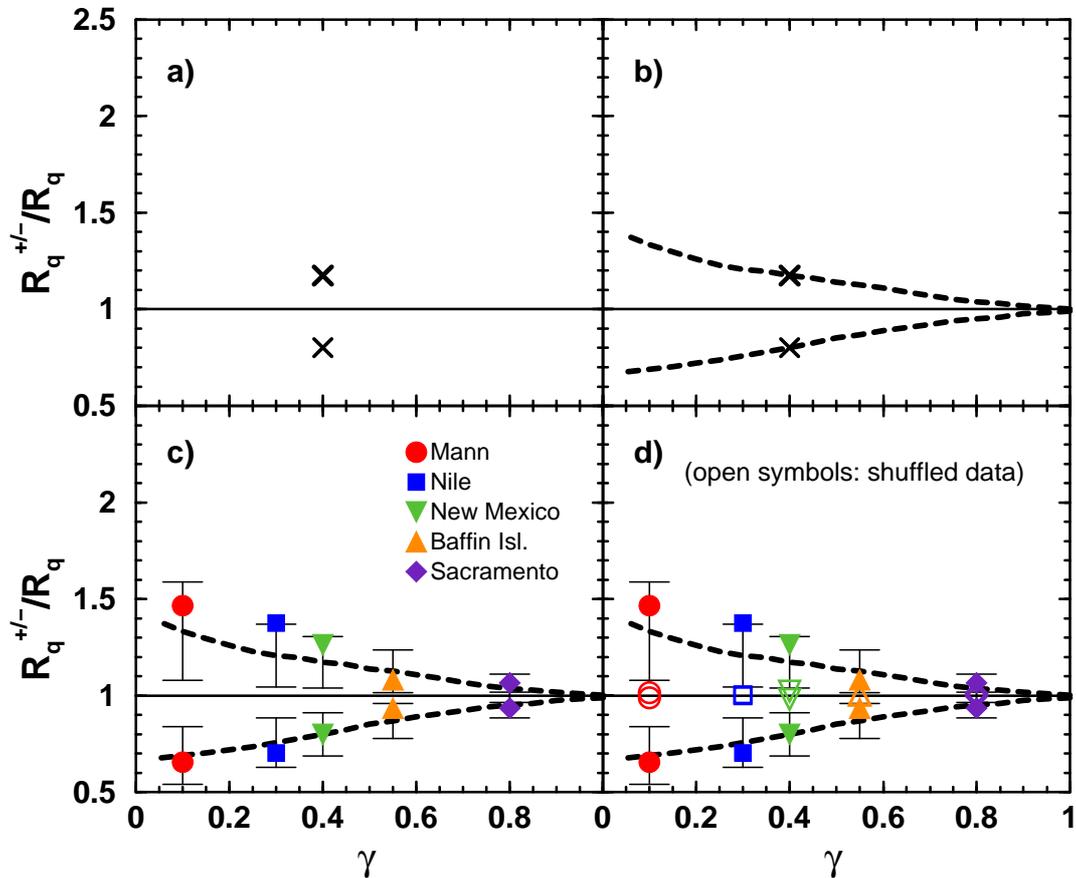


Abbildung 3.22: Bedingte Wiederkehrzeiten $R_q^{+/-}(r_0)$ langzeitkorrelierter Gauss-verteilter (a,b) und echter Daten (c,d) als Funktion vom Korrelationsexponent γ . Aufgrund der geringen Statistik echter Daten haben wir die Reihe der Wiederkehrzeiten unterteilt in die Reihen derjenigen r -Werte, welche unmittelbar auf ein r_0 größer dem Median r^* aller r -Werte folgen (ihr Mittelwert wurde mit R_q^+ bezeichnet) und in die Reihe aller r -Werte, welche unmittelbar auf ein r_0 kleiner dem Median r^* folgen (mit dem Mittelwert R_q^-). Abbildung (a) zeigt für Gauss-verteilter Daten mit $\gamma = 0.4$ die bezüglich des Medians bedingten Wiederkehrperioden R_q^- (unteres Kreuz) und R_q^+ (oberes Kreuz), also in etwa den Mittelwert der linken bzw. rechten Hälfte der Kurven von Abbildung 3.20. Abbildung (b) zeigt das numerische Ergebnis (gestrichelte Kurven) für viele verschiedene γ -Werte. In Abbildung (c) sind die zugehörigen Fehlerbreiten der gestrichelten Kurven aus (b) eingezeichnet und außerdem die Resultate für die echten Datensätze, die schon in Abbildung 3.11 betrachtet wurden (Symbole siehe Legende). Zerstört man die Korrelationen mittels mehrfachen Mischens der echten Datensätze, so können die Bedingungen keinen Einfluss auf die bedingten Wiederkehrperioden haben. R_q^+ und R_q^- sollten nahezu identisch zu R_q sein und der Quotient somit 1. Abbildung (d) zeigt die Resultate für die echten Daten, wenn sie zuvor gemischt wurden (offene Symbole). Abbildung nach [Bunde 2005].

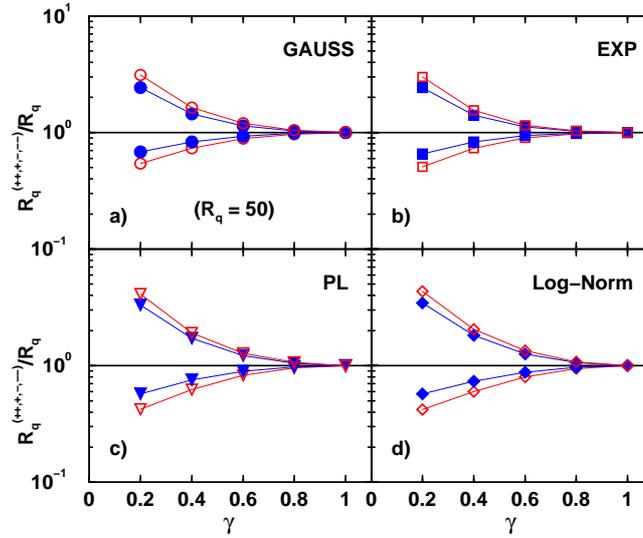


Abbildung 3.23: Bedingte Wiederkehrperioden $R_q^{+/-}$ und $R_q^{++/--}$ in Einheiten von R_q (hier $R_q = 50$) als Funktion des Korrelationsexponenten γ zu den betrachteten vier Verteilungen: (a) Gauss, (b) exponentiell, (c) Potenzgesetz und (d) Log-Normal. R_q^+ (blaue Kreise, oben) bzw. R_q^- (blaue Kreise, unten) bezeichnen die einfach bedingten Wiederkehrperioden, die sich aus den Wiederkehrintervallen r zusammensetzen, welche unmittelbar auf ein Intervall $r_0 > R_q$ (+) bzw. $r_0 < R_q$ (-) folgen. R_q^{++} (rote Kreise, unten) bzw. R_q^{--} (rote Kreise, oben) bezeichnen die doppelt bedingten Wiederkehrperioden, bei denen sowohl der unmittelbar vorhergehende Wert r_0 größer bzw. kleiner R_q ist als auch der vor-vorhergehende Wert r_{-1} . Formal also: $R_q^{++} = R_q(r_0 > R_q \wedge r_{-1} > R_q)$, R_q^{--} analog. Das Auseinanderlaufen der Kurven mit kleiner werdenden γ -Werten ist eine direkte Konsequenz der Langzeitkorrelationen der ursprünglichen x -Werte auf die Reihen der zugehörigen Wiederkehrzeiten r . Im unkorrelierten Fall ($\gamma \geq 1$) degenerieren alle R -Werte zu R_q (horizontale schwarze Linie). Man beachte die halblogarithmische Auftragung. Alle Werte sind über 20 Reihen mit jeweils knapp 42000 r -Werten gemittelt. Abbildung nach [Eichner 2006b].

Fall an. Die Abbildungen (b), (c) und (d) zeigen die Resultate für die anderen drei Verteilungsfunktionen. Der Effekt ist bei Potenzgesetz- und Log-Normal-verteilten Daten in der halblogarithmischen Auftragung sogar stärker als bei Gauss- oder exponentiell-verteilten Daten. Der zu erwartende r -Wert ist bei $\gamma = 0.2$ im zweifach bedingten Fall bis zu einem Faktor 4 größer (++) oder einem Faktor 2.5 kleiner (--) als R_q . Alle Ergebnisse beruhen auf einer Statistik von 20 Reihen mit jeweils 2^{21} Datenpunkten, was bei $R_q = 50$ etwa 42000 r -Werten pro Run entspricht.

3.2.6 Bedingte Verteilungsdichten der Wiederkehrintervalle

Da die Größe $R_q(r_0)$ und ebenso die Größen $R_q^{+/-}$ und $R_q^{++/--}$ aus dem vorherigen Kapitel bedingte Mittelwerte darstellen, also demnach verbesserte Näherungen für den Erwartungswert von Wiederkehrzeiten in langzeitkorrelierten Systemen, gewinnt man aus ihnen keine Wahrscheinlichkeitsaussagen über einzelne mögliche Wiederkehrintervalle. Oft lautet in der Praxis eine relevante Fragestellung, wie wahrscheinlich es ist, dass das nächste Wiederkehrintervall länger oder kürzer als ein vorgegebener Schwellwert r_c ist, der sich deutlich von den

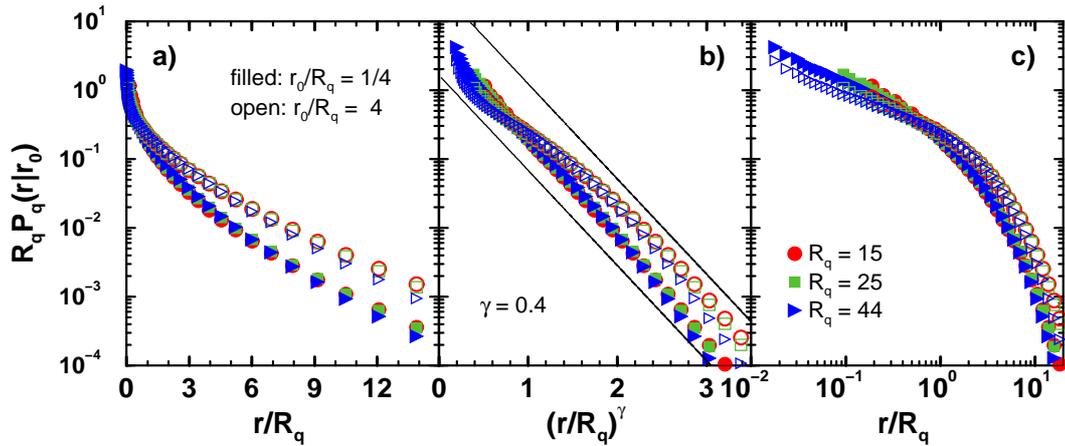


Abbildung 3.24: Verteilungsdichtefunktionen der bedingten Wiederkehrintervalle: Abbildung (a) zeigt für Gauss-Zahlen mit $\gamma = 0.4$ die skalierten Verteilungsdichtefunktionen $R_q P_q(r|r_0)$ zu drei verschiedenen Quantilen ($R_q = 15, 25$ und 44) und zwei verschiedenen Bedingungen, $r_0/R_q = 4$ (offene Symbole) und $r_0/R_q = 0.25$ (volle Symbole). Man sieht in dieser Auftragung deutlich die größere Häufigkeit der großen r/R_q -Werte nach großen r_0/R_q -Werten (offene Symbole über geschlossenen Symbolen). Abbildung (b) zeigt dieselben Kurven nochmal, wobei die x -Achse mit $(r/R_q)^\gamma$ aufgetragen wurde, sodass ein gestreckt exponentieller Verlauf als Gerade erscheint (siehe schwarze Linien). Man sieht, dass die bedingten Verteilungsdichten $P_q(r|r_0)$ rechts-asymptotisch (für große r/R_q -Werte) einen gestreckt exponentiellen Verlauf annehmen (bis auf das Finite-Size-typische Abfallen für große R_q). In Abbildung (c) sind dieselben Kurven in doppelt-logarithmischer Auftragung dargestellt, um den Bereich kleiner r/R_q -Werte zu betonen. Das Verhalten der Kurven bei kleinen r -Werten dreht sich gegenüber (a) um, da auf kleine r_0/R_q -Werte häufiger kleine r/R_q -Werte folgen. Abbildung nach [Bunde 2005].

obigen Mittelwerten unterscheidet. Dazu ist es notwendig, den Verlauf der *bedingten* Verteilungsfunktion der Wiederkehrintervalle zu kennen. Aus ihrer Wahrscheinlichkeitsdichtefunktion $P_q(r|r_0)$ lassen sich die relevanten Größen für Risikoabschätzungen bestimmen. Das erste Moment ist (per Definition) die bedingte Wiederkehrperiode $R_q(r_0)$. Die Wahrscheinlichkeit, einen r -Wert größer oder gleich einem vorgegebenen Schwellwert r_c zu erhalten, wenn der vorherige r -Wert von der Größe r_0 war, ist dann

$$W_q(r \geq r_c|r_0) = \int_{r_c}^{\infty} P_q(r'|r_0) dr'. \quad (3.31)$$

Die Wahrscheinlichkeit für einen r_0 -bedingten r -Wert kleiner als r_c ist $W_q(r < r_c|r_0) = 1 - W_q(r \geq r_c|r_0)$.

Abbildung 3.24 zeigt die bedingten Verteilungsdichtefunktionen $P_q(r|r_0)$ basierend auf allen r -Werten, deren Vorgänger r_0 entweder im Bereich $r_0 \simeq 1/4R_q$ oder $r_0 \simeq 4R_q$ liegen. Dabei wurde dem Bereich um r_0 die Breite $\pm 0.1r_0$ gegeben. In (a) ist die Auftragung halblogarithmisch. Man sieht ein deutliches Auseinanderlaufen der Kurven für große r -Werte. Die Kurven für große r_0/R_q liegen im Bereich großer r/R_q -Werte über den Kurven für kleine r_0/R_q , da sich in den Verteilungsdichten bei großen r_0/R_q korrelationsbedingt mehr große r/R_q -Werte wiederfinden als in den Verteilungsdichten bei kleinen r_0/R_q . In (b) ist auf der x -Achse $(r/R_q)^\gamma$ aufgetragen, da in dieser Darstellung eine gestreckt exponentielle Form als Gerade erscheint. Man sieht, dass auch die bedingten Verteilungsdichtefunktionen asymptotisch

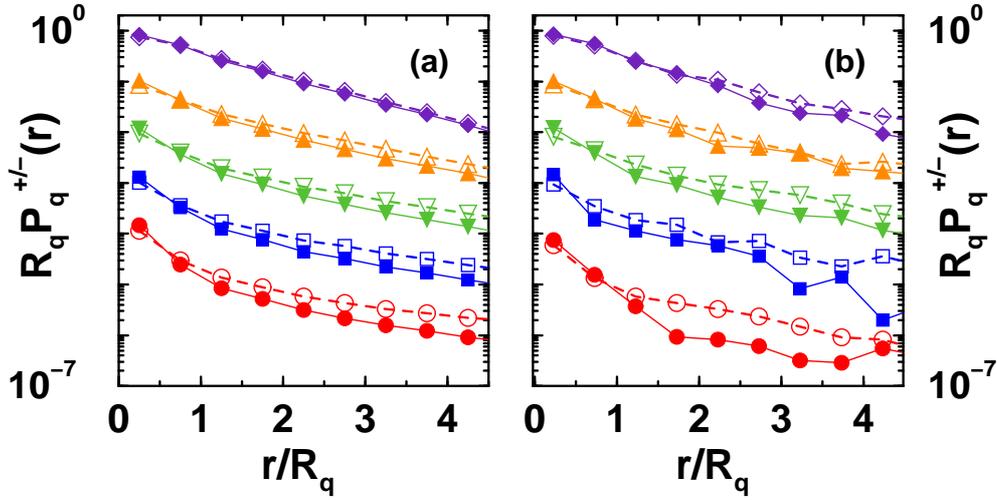


Abbildung 3.25: Bedingte Verteilungsdichtefunktionen der Wiederkehrintervalle künstlicher und echter Daten: Wie im Zusammenhang mit Abbildung 3.22 bereits erklärt, sind bei der Analyse echter Daten mit Rücksicht auf die geringere Statistik die Bedingungen $r_0 > r^*$ ($P_q^+(r)$) bzw. $r_0 < r^*$ ($P_q^-(r)$) sinnvoller. Abbildung (a) zeigt die bedingten Verteilungen der Wiederkehrintervalle künstlicher Gauss-Daten mit denselben Reihenlängen und denselben fünf verschiedenen γ -Werten, welche aus der DFA-Analyse der echten Reihen aus Abbildung 3.22(a) bestimmt wurden (von oben nach unten: $\gamma = 0.8, 0.55, 0.4, 0.3$ und 0.1). Die offenen Symbole stehen für $P_q^+(r)$, die vollen für $P_q^-(r)$. Es wurden hier 10000 Konfigurationen über einen Quantilbereich gemittelt, sodass $R_q > 3$ war und die Anzahl der Wiederkehrzeiten pro Run nicht unter 50 fiel, um eine möglichst gute Statistik der Wiederkehrintervalle zu erreichen. Abbildung (b) zeigt dieselben Resultate für die fünf echten Reihen (Symbole siehe Abbildung 3.22). Man sieht auch hier den Korrelationseffekt im Auseinanderlaufen der bedingten Verteilungen im Vergleich der offenen Symbole mit den geschlossenen Symbolen. Die Abbildung basiert auf Berechnungen von Jan W. Kantelhardt und ist aus [Bunde 2005].

tisch für große r/R_q einen gestreckt exponentiellen Verlauf annehmen. Die Abweichungen bei großen r/R_q -Werten sind denselben Finite-Size Effekten zuzuschreiben wie in Abbildung 3.10. Für große q bzw. große R_q fallen die Kurven früher vom geraden Verlauf einer gestreckt exponentiellen Funktion ab (Vergleich mit durchgezogenen schwarzen Linien). Abbildung (c) zeigt in doppelt-logarithmischer Auftragung den Bereich für kleine r/R_q . Hier dreht sich die Lage der Kurven um. $P_q(r|r_0)$ verläuft für den größeren r_0/R_q -Wert unter der Kurve für den kleineren r_0/R_q -Wert, da aufgrund der Korrelationen im Falle kleiner r_0/R_q -Werte in $P_q(r|r_0)$ auch mehr kleine r/R_q -Werte zu erwarten sind.

Dass dieser Effekt so auch in echten Daten zu sehen ist, bestätigt Abbildung 3.25. Abbildung (a) zeigt die Größen $P_q^+(r)$ und $P_q^-(r)$, welche als Bedingungen $r_0/R_q > r^*/R_q$ und $r_0/R_q < r^*/R_q$ haben (ganz analog zu R_q^+ und R_q^-), wobei r^* wieder der Median aller r -Werte zu gegebenem q -Wert ist. Die zugrunde liegenden Datensätze waren Gauss-verteilt mit denselben Korrelationsexponenten und denselben Reihenlängen wie sie die echten Reihen aus Abbildung 3.11 haben. Es wurde jedoch über 10000 Reihen gemittelt, um glatte Kurven zu bekommen. Die Kurven laufen im Bereich großer r/R_q -Werte auseinander und die Kurve P_q^+ liegt dort über P_q^- . Abbildung zeigt nun die Ergebnisse für die echten Datensätze, wie sie in Abbildung 2.15 und 3.11 verwendet wurden. Der Korrelationseffekt ist auch in den echten Datensätzen

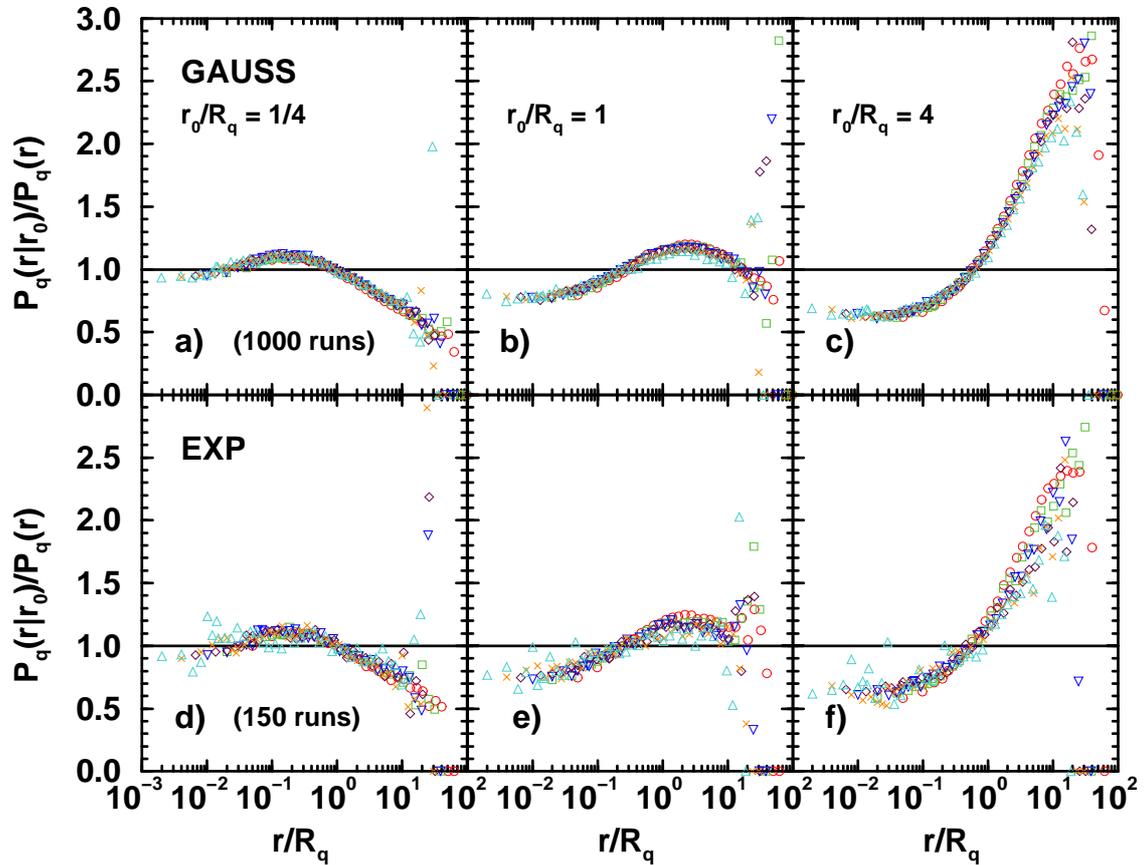


Abbildung 3.26: Quotient aus bedingter Verteilungsdichte $P_q(r|r_0)$ und unbedingter Verteilungsdichte $P_q(r)$ Gauss-verteilter (a,b,c) und exponentiell verteilter (d,e,f) langzeitkorrelierter ($\gamma = 0.4$) Zahlen. Abbildungen (a) und (d) zeigen das Verhältnis für die Bedingung $r_0/R_q = 1/4$, (b) und (e) für $r_0/R_q = 1$ und (c) und (f) für $r_0/R_q = 4$. Der Datenkollaps besteht aus sechs verschiedenen R_q -Werten: $R_q = 20$ (rote Kreise), 50 (grüne Quadrate), 100 (blaue Dreiecke nach unten), 150 (bordeaux Rauten), 250 (orange Kreuze) und 500 (türkise Dreiecke nach oben). Man sieht, wie sich das Maximum mit anwachsendem r_0 nach rechts verschiebt, sogar divergieren zu scheint. Insbesondere die Epochen, in denen lange Ruhephasen aufeinanderfolgen, erzeugen einen großen Unterschied zwischen der bedingten und der unbedingten Verteilungsdichte. Im unkorrelierten Fall ist $P_q(r|r_0)$ unabhängig von r_0 und demnach gleich $P_q(r)$, sodass der Quotient 1 ergibt (durchgezogene schwarze Linie). Die oberen drei Graphen basieren auf einer Statistik von 1000 Konfigurationen mit $N = 2^{21}$ Datenpunkten, die unteren drei Kurven auf 150 Konfigurationen derselben Länge. Der Korrelationseffekt erscheint verteilungsunabhängig und nur von der Bedingung, also dem Einfluss der Vergangenheit auf die (unmittelbare) Zukunft, abhängig zu sein. Abbildung nach [Eichner 2006b].

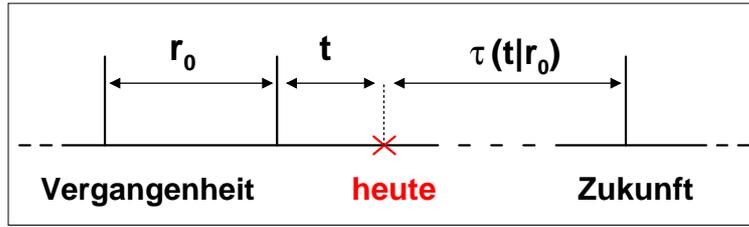


Abbildung 3.27: Definition der bedingten Latenzzeiten $\tau_q(t|r_0)$ bei vorangegangener Ruhephase der Länge t : In vielen praktischen Fragestellungen nach zukünftigen Ereignissen hat man noch die Information der unmittelbaren Vergangenheit seit dem letzten Ereignis. In Systemen mit Abhängigkeiten, also z. B. langzeitkorrelierten Systemen, nimmt diese Dauer t ähnlich wie bereits r_0 Einfluss auf das Eintreten zukünftiger Ereignisse. Die uns jetzt interessierende Größe $\tau_q(t|r_0)$ (bedingte Latenzzeit) setzt sich also zusammen aus den r -Werten, die unmittelbar auf ein r_0 folgen und außerdem eine Mindestlänge t überschritten haben. Die mittlere bedingte Latenzzeit schließt damit alle kurzen Intervalle aus, die seit dem letzten Extremereignis hätten eintreten können.

deutlich zu sehen. Für die Analyse wurde über dieselben q -Bereiche gemittelt wie in Abbildung 3.22.

Abbildung 3.26 zeigt für Gauss- und exponentiell verteilte langzeitkorrelierte Daten das Verhältnis von bedingter und unbedingter Verteilungsdichte der Wiederkehrzeiten zu drei verschiedenen Bedingungen $r_0/R_q = 1/4, 1.0$ und 4 . Die Sequenz der Bilder verdeutlicht nochmal den starken Korrelationseffekt auf die Verteilungsdichtefunktionen und die daraus resultierenden Verschiebungen der Peaks mit anwachsender Bedingung. Während kleine Bedingungswerte eine (leichte) Erhöhung der Wahrscheinlichkeitsdichte bei kleinen r/R_q -Werten verursachen, rutscht der Peak bei großen Bedingungen zu großen r/R_q -Werten und scheint bei sehr großen Bedingungen sogar zu divergieren. Hier führt die endliche Reihenlänge und damit stark beschränkte Länge der maximalen Wiederkehrintervalle in einen Bereich, in dem keine Statistik mehr vorhanden ist, die Kurven brechen ab. Durch die Beschränkung der r -Werte nach unten (kleinster r -Werte ist 1) ist der Effekt bei kleinen r/R_q -Werten zwar schwächer. Er ist aber nicht zu unterschätzen, da diese 1er für das Clustern, also das gehäufte Auftreten von Extremereignissen zuständig sind. Das divergente Erscheinen der großen r -Werte besagt, dass die Ruhephasen, die auf lange Ruhephasen folgen, wieder mit hoher Wahrscheinlichkeit sehr lang werden. In langzeitkorrelierten Systemen sind also nach langen Ruhephasen Extremereignisse keineswegs überfällig. Im Gegenteil, sie sind ein Indiz dafür, dass auch weiterhin eine ruhige kurz- und mittelfristige Zukunft ohne viele Extremereignisse bevorsteht.

3.2.7 Bedingte Latenzzeiten

Die bedingten Wiederkehrintervalle berücksichtigen die Vergangenheit in Form eines vollständigen Wiederkehrintervalls *zwischen* den letzten beiden Extremereignissen. Es wird jedoch nicht berücksichtigt wie lange das letzte Extremereignis bereits zurückliegt. Angenommen, das letzte Extremereignis ($x_i > q$) fand zum Zeitpunkt i statt, dann kann seit dem bereits die Zeit t vergangen sein, ohne dass ein weiteres Extremereignis stattgefunden hat ($x_{i+j} < q$ für $j = 1, 2, \dots, t$). Diese Dauer t zwischen dem jetzigen Zeitpunkt und dem letzten Extremereignis enthält im korrelierten Fall zusätzliche Information, die ebenfalls berücksichtigt werden

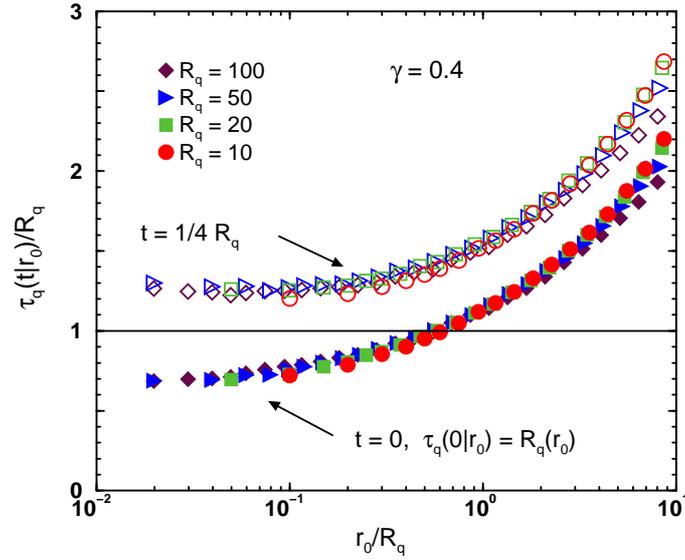


Abbildung 3.28: Bedingte Latenzzeiten $\tau_q(t|r_0)$ Gauss-verteilter langzeitkorrelierter Daten ($\gamma = 0.4$) bei vorangegangener Ruhephase der Länge $t = 0$ (geschlossene Symbole, $\tau_q(t = 0|r_0) = R_q(r_0)$) und $t = R_q/4$ (offene Symbole). Die Kurven für die gewählten Quantile (siehe Legende) werden auch für die Größe $\tau_q(t|r_0)$ durch Skalieren mittels der unbedingten Wiederkehrperiode R_q zum Kollaps gebracht. Die bedingte Latenzzeit (offene Symbole) liegt für die gewählten Parameter $\gamma = 0.4$ und $t/R_q = 0.25$ etwa um den konstanten Wert 0.5 über der bedingten Wiederkehrperiode $R_q(r_0)$ (geschlossene Symbole), d. h. je länger die Ruhephase schon andauert, um so länger muss im Mittel auf das kommende Extremereignis gewartet werden. In langzeitkorrelierten Systemen gilt demnach, je länger die aktuelle Ruhephase schon andauert und je länger die Zeit zwischen den beiden letzten extremen Ereignissen dauerte, um so länger lässt das zukünftige Extremereignis auf sich warten. Die Abweichungen bei großen r_0/R_q -Werten und großen R_q sind auf Finite-Size Effekte zurückzuführen.

muss. Die verbleibende Restdauer bis zum nächsten Extremereignis (Englisch *residual waiting time*) möchte ich hier mit *Latenzzeit* $\tau(t)$ bezeichnen. Berücksichtigt man neben t auch das vorangegangene r_0 , so hat man die *bedingte Latenzzeit* (oder Restwartezeit) $\tau_q(t|r_0)$. Abbildung 3.27 stellt die Definition von $\tau_q(t|r_0)$ graphisch dar. Da wir uns nur für die Mittelwerte von $\tau_q(t|r_0)$ zu vorgegebenem t und r_0 interessieren, steht $\tau_q(t|r_0)$ analog zur bedingten Wiederkehrperiode $R_q(r_0)$. Zum Zeitpunkt $t = 0$ befindet man sich an einem Extremereignis und es gilt $\tau_q(0|r_0) \equiv R_q(r_0)$. Bei bekannter bedingter Verteilungsdichtefunktion $P_q(r|r_0)$ lautet die Definition der bedingten Latenzzeit

$$\tau_q(t|r_0) := \int_t^\infty (r - t)P_q(r|r_0)dr / \int_t^\infty P_q(r|r_0)dr. \quad (3.32)$$

Der Einfluss der bereits verstrichenen Dauer t seit dem letzten Extremereignis ist in Abbildung 3.28 für Gauss-vertelte langzeitkorrelierte Daten ($\gamma = 0.4$) dargestellt. Dort wird die Kurve der bedingten Wiederkehrperiode $R_q(r_0) = \tau_q(0|r_0)$ (volle Symbole) aus Abbildung 3.20 mit der bedingten Latenzzeit $\tau_q(t|r_0)$ für $t = R_q/4$ verglichen. Das Ergebnis zeigt, je länger die aktuelle Ruhephase t anhält und je größer die letzte Wiederkehrzeit r_0 war, umso länger wird die Zeit τ bis zum nächsten Extremereignis dauern. Für $t/R_q = 1/4$ ist die zu erwartende

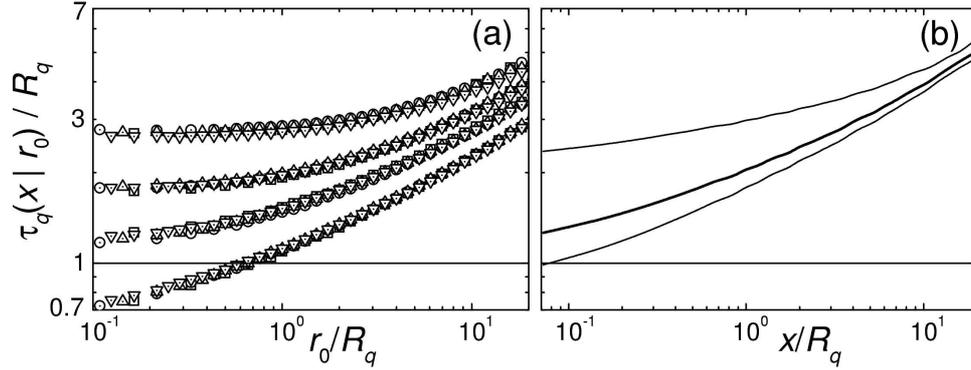


Abbildung 3.29: Bedingte Latenzzeiten $\tau_q(t|r_0)$ als Funktion von r_0 (a) und t (b) in doppelt-logarithmischer Auftragung (Anm.: in dieser Abbildung x statt t). Abbildung (a) zeigt, wie in Abbildung 3.28, das Anwachsen der bedingten Latenzzeit als Funktion von Bedingung r_0 für vier verschiedene t -Werte (von unten nach oben: $t/R_q = 0$ (entspricht $R_q(r_0)$), 0.25, 1 und 4) und vier q -Werte ($q = 1.0, 1.25, 1.5$ und 1.75). Abbildung (b) zeigt die bedingten Latenzzeiten als Funktion der anhaltenden Ruhephase t (hier x). Die obere dünne Linie steht für die Bedingung $r_0/R_q = 8$, die untere für $r_0/R_q = 1/8$. Die mittlere Linie entspricht der Mittelung über alle möglichen Bedingungen und somit einer Größe $\tau_q(t)$. Das Bild zeigt, je länger die Ruhephase t schon andauert, um so länger muss im Mittel auf das kommende Extremereignis gewartet werden. In langzeitkorrelierten Systemen findet man demnach ein eher unintuitives Anwachsen der Latenzzeiten. Das heißt, je länger die aktuelle Ruhephase schon andauert und je länger die Zeit zwischen den beiden letzten extremen Ereignissen dauerte, um so länger lässt das zukünftige Extremereignis auf sich warten. Abbildung entnommen aus [Bunde 2005].

Verlängerung gegenüber R_q , also die Differenz zwischen $R_q(r_0)$ und $\tau_q(R_q/4|r_0)$, etwa konstant $0.5R_q$ und $\tau_q(R_q/4|r_0)/R_q$ liegt sogar für kleine r_0/R_q deutlich über R_q .

Die Berücksichtigung von r_0 in langzeitkorrelierten Systemen führt dazu, dass $\tau_q(R_q/4|r_0)$ genau wie $R_q(r_0)$ mit anwachsendem r_0 ebenfalls anwächst. Der Einfluss von t schließt für die Bestimmung von $\tau_q(R_q/4|r_0)$ nicht nur alle Werte $r < t$ aus, sondern verursacht aufgrund der Langzeitkorrelationen auch noch einen ähnlichen Effekt wie r_0 : Wenn t bereits groß ist, steigt die Wahrscheinlichkeit, dass das nächste Intervall $r = t + \tau_q(t|r_0)$ ebenfalls größer wird. Abbildung 3.29(b), in der $\tau_q(t|r_0)$ als Funktion von t aufgetragen ist, zeigt sogar, dass für große t auch $\tau_q(t|r_0)$ ansteigt. Je größer t ist, desto länger wird auch $\tau_q(t|r_0)$. Entgegen der intuitiven Erwartung, dass nach einer gewissen Ruhephase t nun endlich wieder mit einem Extremereignis zu rechnen sein sollte, zeigt dieses Bild, dass die verbleibende Zeit bis zum nächsten Extremereignis umso länger ist, je länger die aktuelle Ruhephase t bereits anhält. Nach langen Ruhephasen r_0 und anhaltender Ruhephase t wird in langzeitkorrelierten Systemen ein nächstes Extremereignis nicht etwa “überfällig”, sondern lässt mit hoher Wahrscheinlichkeit noch weiterhin auf sich warten. Dieser Effekt ist in unkorrelierten Daten nicht zu sehen.

Abbildung 3.30 zeigt die Latenzzeiten als Funktion von γ mit $t/R_q = 0.25$ für Gauss-Daten (gestrichelte Linien) und für die echten Daten, wie in Abbildung 3.22. Dabei wurde die Größe $\tau_q^{+/-}(t)$ völlig analog zu $R_q^{+/-}$ gebildet (siehe Kapitel 3.2.5). Als Bedingung wurde auch hier aus statistischen Gründen der Median r^* gewählt. Der q -Bereich, über den gemittelt wurde, ist ebenfalls mit dem aus Abbildung 3.22 identisch. Die Ergebnisse der echten Daten liegen (bis auf die Temperaturreihe von Baffin Island) innerhalb der Fehlerbalken. Zum Kontrast sind auch die Ergebnisse der gemischten Daten (offene Symbole) gezeigt. Sie laufen nicht

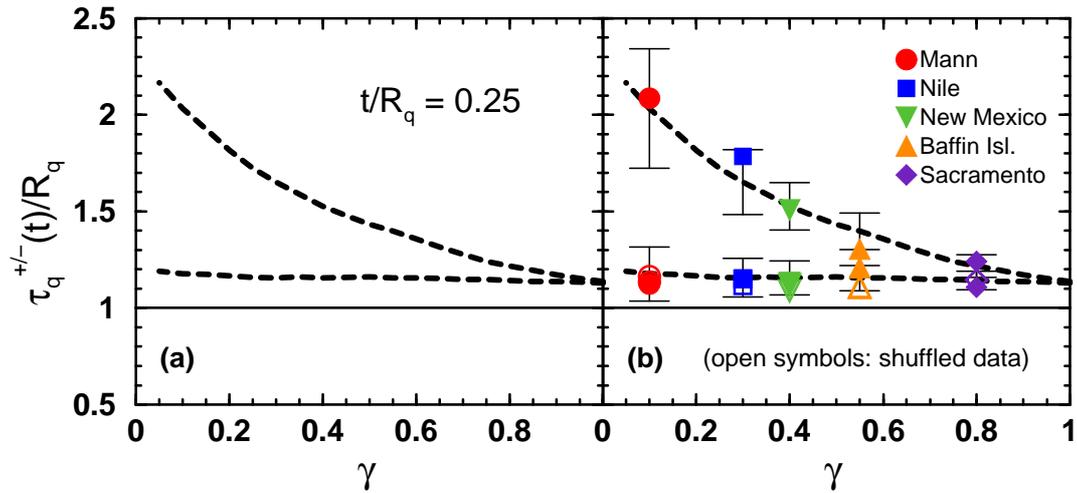


Abbildung 3.30: Bedingte Latenzzeiten $\tau_q^{+/-}(t)$ bei vorangegangener Ruhephase der Länge $t = 1/4R_q$ künstlicher (a) und echter (b) langzeitkorrelierten Daten als Funktion der Korrelationsexponenten γ . Die Abbildung ist völlig analog zur Abbildung 3.22. Abbildung (a) zeigt das Anwachsen von $\tau_q^+(t)$ mit stärker werdenden Korrelationen während $\tau_q^-(t)$ zum gewählten $t/R_q = 0.25$ in etwa konstant bleibt, aber bereits deutlich größer als die unbedingte Wiederkehrperiode R_q . Abbildung (b) zeigt die Resultate für die echten Reihen aus Abbildung 3.11 (volle Symbole) und die zugehörigen gemischten Reihen (offene Symbole). Die Latenzzeiten der gemischten Reihen liegen über R_q , der Quotient für die gemischten Reihen liegt bei etwa 1.1. Abbildung nach [Bunde 2005].

auseinander, liegen aber etwas über der Linie für $\tau_q^{+/-}(t)/R_q = 1$, etwa bei 1.1. Referenz [Sornette 1997] hat sich mit diesem Problem auseinander gesetzt und schreibt es einem Finite-Size Effekt zu.

Kapitel 4

Extremwertstatistik II: Statistik der Maximawerte

Die Extremwertstatistik ist in den letzten 50 Jahren zu einer der wichtigsten Disziplinen der Statistik und der angewandten Wissenschaften geworden [Gumbel 1958, Galambos 1978, Leadbetter 1983, Galambos 1994, Coles 2001]. Sie hat außerdem Einzug gehalten ins Risikomanagement der Finanzmärkte, ins Versicherungswesen und besonders in Rückversicherungen, welche sich am intensivsten mit Naturkatastrophen und Katastrophen jeglicher Art auseinandersetzen müssen [Smith 1996, Embrechts 1997, Embrechts 1999, Danielsson 2000, Reiss 2001]. Aber auch in vielen physikalischen und technischen Bereichen ist sie ein wichtiges Instrument [Bramwell 1998, Bramwell 2000, Raychaudhuri 2001, Dean 2001, Guclu 2004, Majumdar 2004]. Die Extremwertstatistik, d. h. die Statistik der Maximawerte, trifft Wahrscheinlichkeitsaussagen für Ereignisse, die oft extremer sind als alle, die bisher beobachtet wurden. Wenn beispielsweise ein neuer Damm gebaut wird, so soll dieser während seiner Lebensdauer von sagen wir 100 Jahren einen Schutz gegen alle Fluten garantieren, die innerhalb dieser 100 Jahre vorkommen können. Meistens gibt es aber nur verlässliche historische Aufzeichnungen, die nur wenige Jahrzehnte zurückreichen. Um also die Größe einer Flut abzuschätzen, die höchstens alle 100 Jahre auftritt, muss man aus den Aufzeichnungen extrapolieren. Es ist Aufgabe der Extremwerttheorie, ein mathematisches Gerüst zu liefern, mittels dessen diese Abschätzungen so realistisch wie möglich werden. Zum einen, um möglichst hohe Sicherheit zu gewähren, zum anderen aber auch, um unnötige Kosten (z. B. bei einem fälschlich überdimensionierten Bau eines Damms) zu sparen.

Da man in der Praxis so gut wie keine empirischen oder physikalischen Richtlinien zur Verfügung hat, die eine dem Problem angepasste mathematische Extrapolation vorschreiben, bedient man sich asymptotischer Argumente. Dies sieht dann wie folgt aus: x_1, x_2, \dots seien Messwerte, beispielsweise täglich gemessene Pegelstände oder Tageshöchsttemperaturen. Dann ist

$$m^{(R)} = \max\{x_1, x_2, \dots, x_R\} \quad (4.1)$$

der maximale Pegelstand oder die maximale Tageshöchsttemperatur einer R -tägigen Beobachtung. Da das exakte statistische Verhalten der x_i in der Praxis nicht bekannt ist, kann auch keine exakte Aussage über das Verhalten von $m^{(R)}$ getroffen werden. Jedoch kann unter gewissen Bedingungen, die im Folgenden noch erklärt werden, das Verhalten von

$m^{(R)}$ für $R \rightarrow \infty$ durch *Grenzwertgesetze* bestimmt werden. Diese Grenzwertgesetze liefern für die Verteilungen der Maxima Modelle, welche mittels der beobachteten Werte x_i und $m^{(R)}$ angepasst (gefittet) werden und erlauben es, auf unbeobachtete Werte zu extrapolieren [Gumbel 1958, Galambos 1978, Leadbetter 1983]. Diese Methode, auch *Extremwertparadigma* genannt [Coles 2001], hat ganz offensichtlich Schwächen, da man im Rahmen der Extrapolation dem Prozess eine gewisse statistische Glätte (*Stationarität*) unterstellt, welche nur vage gemessen werden kann. Es ist aber bis heute keine bessere Vorgehensweise etabliert, mal abgesehen vom unwissenschaftlichen intuitiven Raten. Die Schwächen der Methode sind:

- Die asymptotischen Modelle und Aussagen gelten prinzipiell nur für unendlich große Systeme. In der Realität sind alle Systeme endlich.
- Die meisten Modelle gelten unter idealisierten Bedingungen, welche in der Realität so garantiert nicht herrschen (z. B. statistische Unabhängigkeit).
- Viele Modelle passen oft besser, wenn gewisse Informationen vernachlässigt werden [Coles 2001], wenn zum Beispiel Ausreißer nicht berücksichtigt werden.

Der letzte Punkt bedarf vielleicht etwas mehr Erklärung. Eine gängige Methode zur Aufzeichnung von Extremdaten ist, dass man nur die Maximalwerte einer bestimmten Periode R , beispielsweise $R = 365$, also Jahresmaxima, notiert. Unter der Annahme, dass R groß genug ist, liefert die asymptotische Argumentation nun ein Modell, welches auf das Verhalten von Jahresmaxima zurückgreift. Jedoch kommen in manchen Jahren auch extremere Ereignisse vor, die nicht typisch für Perioden der Länge 1 Jahr sind, sondern z. B. typisch für 100 Jahre. In der Praxis werden diese Ausreißer häufig aus der Statistik ausgeschlossen mit dem Argument, dass es sich hier nicht um Jahresmaxima handelt, sondern um Jahrhundertmaxima, und diese die Statistik der Jahresmaxima verzerren würden. Wird ein Jahrhundertmaximum fälschlicherweise als Jahresmaximum eingestuft, kann dies in den Modellen zur Folge haben, dass sich die Abschätzungen für Jahrhundertmaxima verschieben. Daher kann diese Vorgehensweise im Risikomanagement fatal sein. Ein Ausschluss von Information, damit ein theoretisches Modell besser passt, käme aber auch einem Anpassen der Realität an das gewählte Modell gleich und nicht dem Umgekehrten. Je nach Fragestellung muss neu entschieden werden, welche Daten sich für welches Modell eignen und welche Aussagen das gewählte Modell zulässt.

In diesem Kapitel studiere ich den Einfluss von Langzeitkorrelationen auf die Statistik der Maximalwerte. Dabei gilt das besondere Augenmerk dem praxisrelevanten Fall eines *endlichen* R -Werts (z. B. $R = 365$ bei Tagesdaten). Die asymptotischen Gesetze der klassischen Extremwerttheorie gelten für $R \rightarrow \infty$, was in der Praxis natürlich nicht zu realisieren ist. Das Kapitel ist wie folgt aufgebaut: Im nächsten Abschnitt beschreibe ich die klassische Extremwerttheorie mit ihren drei Grenzverteilungen. Danach zeige ich in detaillierten Studien, welche Effekte Langzeitkorrelationen auf die Extremwertstatistik haben und wie man sie zur Verbesserung von Risikoabschätzungen und der Vorhersage von Jahrhundertereignissen ausnutzen kann. Dazu kommen auch Ergebnisse von einigen Beispielen echter Datenreihen, wie sie auch schon im Kapitel 3 über die Wiederkehrintervalle gezeigt wurden.

4.1 Klassische Theorie der Extremwerte

Die klassische Theorie der Extremwerte, die bei Hochwasser, Starkniederschlägen, Seewellen und Erdbeben etc. angewendet wird, wurde als erstes von Fréchet (1927) und von Fisher & Tippett (1928) eingeführt [Fréchet 1927, Fisher 1928]. Später entwickelte und verbreitete Gumbel (1958) ihre Anwendung [Gumbel 1958]. Nach der klassischen Extremwert-Theorie geht die Verteilung des größten (bzw. des kleinsten) Wertes von identisch und unabhängig verteilten Zufallsvariablen in eine asymptotische Verteilung über, die nur vom rechten (bzw. linken) Ende der ursprünglichen Verteilung abhängt [Gumbel 1958]. In diesem Kapitel möchte ich die klassische Extremwerttheorie, wie sie von Fréchet, Fisher, Tippett und Gumbel begründet wurde und heute noch Anwendungen hat, kurz erläutern.

4.1.1 Das klassische Modell

In dem der Maxima-Statistik zugrunde liegenden Modell [Gumbel 1958, Leadbetter 1983, Coles 2001] besteht eine Folge $(x_i)_{i=1}^N$ (Stichprobe, Messreihe, Realisation, etc.) aus statistisch unabhängigen Zufallsvariablen. Die Messwerte x_i folgen einer gemeinsamen aber nicht hinreichend gut bekannten Verteilungsdichtefunktion $P(x)$. Die Wahrscheinlichkeit für das Auftreten eines Ereignisses $x = m$ (mit $m \in \mathbb{R}$) lässt sich nicht angeben, da sie das Maß Null hat bzw. physikalisch der Ausdehnung einer Punktmasse entspricht. Stattdessen fragt man zum Beispiel nach der Wahrscheinlichkeit W für das Eintreten eines Ereignisses x *kleiner* einer gegebenen Schranke m

$$W(x \leq m) = F(m) = \int_{-\infty}^m P(x) dx. \quad (4.2)$$

Entsprechend ist die Wahrscheinlichkeit für das Eintreten eines Ereignisses *oberhalb* dieser Schranke

$$W(x > m) = 1 - F(m). \quad (4.3)$$

In der Extremwertstatistik interessiert man sich für das Verhalten des größten Wertes $m^{(R)}$ innerhalb einer Stichprobe $(x_i)_{i=1}^R$ vom Umfang R ,

$$m^{(R)} = \max\{x_1, x_2, \dots, x_R\}. \quad (4.4)$$

Wenn R zum Beispiel die Anzahl der Messungen innerhalb eines Jahres ist, entspricht $m^{(R)}$ dem Jahresmaximum. Die $m^{(R)}$ -Werte folgen nun einer Verteilungsdichtefunktion $P_R(m)$. Die Wahrscheinlichkeit $W(m^{(R)})$ für das Ereignis der Größe $m^{(R)}$ in Gl. (4.4) lässt sich wie $W(x)$ nicht angeben, da es sich hier ebenfalls um eine Wahrscheinlichkeit mit Maß Null handelt. Stattdessen fragt man nach der Wahrscheinlichkeit, dass ein Wert $m^{(R)}$ *kleiner* einer gegebenen Schranke m auftritt. Dies lässt sich theoretisch exakt herleiten, vorausgesetzt die Zufallsvariablen x_i (Messwerte) sind statistisch unabhängig und identisch verteilt, abgekürzt **i.i.d.** (aus dem Englischen für **i**ndependent and **i**dentically **d**istributed, [Sachs 1974]), und folgen einer vollständig bekannten Wahrscheinlichkeitsdichte $P(x)$. Dann gilt für die Wahrscheinlichkeit $G_R(m)$, dass in einer beliebigen Stichprobe vom Umfang R ein Wert $m^{(R)}$ auftaucht, der kleiner als eine gegebene Schranke m ist,

$$G_R(m) = W(m^{(R)} \leq m) \quad (4.5)$$

$$\begin{aligned}
&= W(x_1 \leq m, x_2 \leq m, \dots, x_R \leq m) \\
&\stackrel{\text{(i)}}{=} W(x_1 \leq m) \cdot W(x_2 \leq m) \cdot \dots \cdot W(x_R \leq m) \\
&\stackrel{\text{(ii)}}{=} W(x \leq m) \cdot W(x \leq m) \cdot \dots \cdot W(x \leq m) \\
&= F(m) \cdot F(m) \cdot \dots \cdot F(m) \\
&= [F(m)]^R.
\end{aligned} \tag{4.6}$$

Schritt (i) nutzt die statistische Unabhängigkeit aus, Schritt (ii) die identische Verteilung der x_i . Für die Wahrscheinlichkeit, einen Wert $m^{(R)}$ größer einer Schranke m zu finden, gilt dann

$$W(m^{(R)} > m) = 1 - G_R(m) =: E_R(m). \tag{4.7}$$

Dieser Ansatz hilft in der Praxis leider nicht weiter, da die statistische Unabhängigkeit der Variablen nicht gewährleistet ist und F für große m nicht hinreichend bekannt ist aufgrund der unvollständigen Kenntnis von $P(x)$. Kleine Fehler in der Schätzung von F können dann zu sehr großen Fehlern in F^R führen. Man muss akzeptieren, dass F an den Rändern praktisch unbekannt ist und versuchen, eine möglichst gute Annäherung für $[F(m)]^R$ bzw. $G_R(m)$ direkt zu finden.

Hierbei geht man intuitiv vor. Ein Beispiel soll dies erläutern: Angenommen, man hat eine Methode mittels derer man mit hoher Zuverlässigkeit vorhersagen kann, dass ein bestimmter Fluss innerhalb der nächsten 30 Jahre keine neuen Deiche braucht (d. h., der Wasserpegel bleibt unter einer bestimmten Höhe h). Jeder praktisch veranlagte Mensch würde nun dieselbe Methode auch für Vorhersagen der nächsten 33 Jahre verwenden. In anderen Worten, $G_R(m)$ ändert sich bei einem großen R kaum gegenüber einem etwas größeren R . Da die Verteilung $G_R(m)$ und die Variation von $m^{(R)}$ durch das unbekannte $F(m)$ im Zusammenhang stehen (Gln. ((4.6)), wäre das akkurate mathematische Äquivalent zu obigem "naivem" Ansatz [Galambos 1978], dass nach einer simplen linearen Transformation (Renormierung) mit Konstanten $\alpha_R > 0$ (*Skalenparameter*) und u_R (*Lokalisationsparameter*) die Größe $(m^{(R)} - u_R)/\alpha_R$ mit größer werdendem R mehr und mehr unabhängig von R wird, also gegen einen Wert $M < \infty$ konvergiert.

$$\lim_{R \rightarrow \infty} \frac{m^{(R)} - u_R}{\alpha_R} = \frac{m^{(\infty)} - u_\infty}{\alpha_\infty} = M < \infty. \tag{4.8}$$

Das bedeutet, aus

$$W(m^{(R)} \leq m) = W\left(\frac{m^{(R)} - u_R}{\alpha_R} \leq \frac{m - u_R}{\alpha_R}\right) = G_R(m) \tag{4.9}$$

folgt

$$\lim_{R \rightarrow \infty} W\left(\frac{m^{(R)} - u_R}{\alpha_R} \leq \frac{m - u_R}{\alpha_R}\right) = \lim_{R \rightarrow \infty} G_R(m) = G(M). \tag{4.10}$$

Das besondere hieran ist nun, dass die Verteilung G völlig unabhängig von R und F^R ist. Das heißt, dass bei bekanntem G die Ausgangsverteilung $P(x)$ der x_i keine wichtige Rolle mehr spielt, sofern die Renormierungskonstanten u_∞ und α_∞ auf irgendeine Art bestimmt werden können. Im folgenden möchte ich u_∞ mit u und α_∞ mit α abkürzen und für $G(M)$ die Schreibweise $G\left(\frac{m-u}{\alpha}\right)$ einführen. Gesucht sind als nächstes die Bedingungen, unter denen dieser "naive" Ansatz mathematisch gerechtfertigt ist. Zusätzlich soll die Grenzwertverteilung G und eine Methode angegeben werden, mit der die Konstanten $\alpha > 0$ und u aus G bestimmt werden können.

4.1.2 Drei Typen von Extremwertverteilungen

Das zentrale Ergebnis der klassischen Extremwertstatistik ist das sogenannte *Extremal Types Theorem*, zuerst aufgestellt von Fréchet (1927) und Fisher & Tippett (1928) und später bewiesen von Gnedenko (1943) [Fréchet 1927, Fisher 1928, Gnedenko 1943]. In Worten besagt es, dass es als Grenzverteilungen $G_R(m)$ für Maximawerte nur *drei* mögliche Extremwertverteilungen gibt. Bevor ich diese drei Extremwertverteilungen vorstelle, möchte ich zunächst den Begriff *Maximum-stabil* erläutern (siehe beispielsweise [Gumbel 1958, Leadbetter 1983, Coles 2001]).

Eine Verteilungsfunktion bezeichnet man als *Maximum-stabil*, wenn man zu allen R -Werten einen Lokalisationsparameter u_R und einen Skalenparameter $\alpha_R > 0$ finden kann, so dass

$$G_R(m) = G\left(\frac{m - u_R}{\alpha_R}\right). \quad (4.11)$$

gilt. Das heißt, alle Verteilungsfunktionen G_R der Maxima fallen mit der von R unabhängigen Kurve G zusammen und bilden einen Datenkollaps, sobald man die Maxima m mit u_R und α_R skaliert. Man beachte, dass hier nicht die Konvergenz von G_R gegen G gemeint ist, sondern *Gleichheit* gilt (siehe z. B. Kapitel 1.3 in [Leadbetter 1983]).

Nun besagt ein Theorem (siehe Theorem 1.4.1 auf Seite 10 in [Leadbetter 1983]), dass die einzigen Verteilungen, die diese Eigenschaft erfüllen, zu einer der folgenden drei Verteilungen gehören:

$$G_{\text{I}}\left(\frac{m - u}{\alpha}\right) = \exp\left[-e^{-\frac{m-u}{\alpha}}\right] \quad \text{für} \quad -\infty < \frac{m - u}{\alpha} < \infty \quad (4.12)$$

$$G_{\text{II}}\left(\frac{m - u}{\alpha}\right) = \begin{cases} 0 & \text{für} \quad \frac{m-u}{\alpha} \leq 0 \\ \exp\left[-\left(\frac{m-u}{\alpha}\right)^{-\phi}\right] & \text{für} \quad \frac{m-u}{\alpha} > 0 \end{cases} \quad \text{und mit} \quad \phi > 0 \quad (4.13)$$

$$G_{\text{III}}\left(\frac{m - u}{\alpha}\right) = \begin{cases} \exp\left[-\left(-\frac{m-u}{\alpha}\right)^{\phi}\right] & \text{für} \quad \frac{m-u}{\alpha} \leq 0 \quad \text{und mit} \quad \phi > 0 \\ 1 & \text{für} \quad \frac{m-u}{\alpha} > 0 \end{cases} \quad (4.14)$$

Diese drei Verteilungen sind genau die drei möglichen Extremwertverteilungen. Ihre zugehörigen Verteilungsdichten sind in Abbildung 4.1 skizziert. Jede Maximum-stabile Verteilung ist also eine dieser drei Extremwertverteilungen¹, die häufig folgende Synonyme haben:

- Typ I: Gumbel-Verteilung, 1. Extremwertverteilung oder Doppelexponentialverteilung
- Typ II: Fréchet-Verteilung, 2. Extremwertverteilung oder Log-Gumbel-Verteilung
- Typ III: Weibull-Verteilung oder 3. Extremwertverteilung

Das *Extremal Types Theorem* (siehe Theorem 1.4.2 auf Seite 11 in [Leadbetter 1983] oder Theorem 3.1 auf Seite 46 in [Coles 2001]) besagt nun, dass die Verteilung der Maxima nach Gl. (4.4) aus Stichproben von beliebig aber stationär verteilten i.i.d. Zufallszahlen mit anwachsendem R gegen eine Maximum-stabile Verteilung konvergiert. Es kommt dann nur eine der drei möglichen Extremwertverteilungen als Grenzwertverteilung in Frage. Formal heißt

¹Umgekehrt besagt das Theorem auch, dass jede Extremwertverteilung Maximum-stabil ist. Somit gibt es tatsächlich nur drei Extremwertverteilungen.

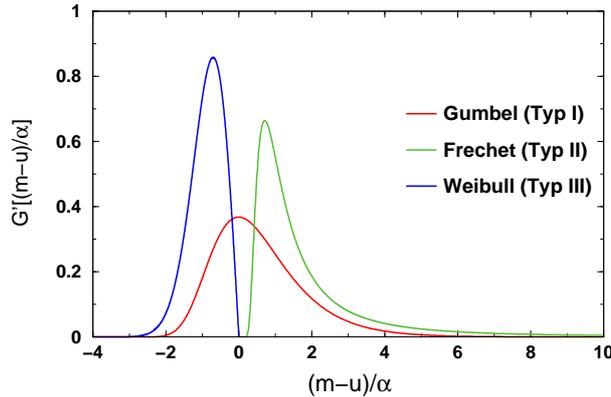


Abbildung 4.1: Dichtefunktionen der drei Extremwertverteilungen: Gumbel (rot), Fréchet (grün) und Weibull (blau). Die Parameter für die Kurven beziehen sich auf die Gln. (4.12), (4.13) und (4.14) und wurden zur Anschauung wie folgt gewählt: $u = 0$, $\alpha = 1$ und $\phi = 1.5$ für Fréchet und $\xi = 2$ für Weibull.

das, dass für $R \rightarrow \infty$ die Folgen der Skalenparameter α_R und Lokalisationsparameter u_R die Wahrscheinlichkeitsverteilung $G_R(m) \rightarrow G\left(\frac{m-u}{\alpha}\right)$ konvergieren lassen, wobei G eine der drei Extremwertverteilungen² ist.

Die Verteilung der Maxima muss also selbst nicht Maximum-stabil sein, sie muss nur gegen eine Maximum-stabile Verteilung für $R \rightarrow \infty$ konvergieren:

$$G_R(m) \simeq G\left(\frac{m-u_R}{\alpha_R}\right) \rightarrow G\left(\frac{m-u}{\alpha}\right) \quad (4.15)$$

Dieser Aspekt ist von großer Wichtigkeit für die Praxis. Dadurch ist es erlaubt, numerisch erhaltene Histogramme mittels einer dieser drei Extremwertverteilungen anzunähern (fitten). Es gilt dann beispielsweise für alle Verteilungen, die im Anziehungsbereich der Gumbel-Verteilung (Typ I) liegen

$$G_R(m) \rightarrow G\left(\frac{m-u}{\alpha}\right) = \exp\left[-e^{-\frac{m-u}{\alpha}}\right] \quad (4.16)$$

und für die zugehörigen Wahrscheinlichkeitsdichten

$$\frac{dG_R(m)}{dm} = P_R(m) \rightarrow P\left(\frac{m-u}{\alpha}\right) = \frac{1}{\alpha} \exp\left[-e^{-\frac{m-u}{\alpha}} - \frac{m-u}{\alpha}\right]. \quad (4.17)$$

Die Verteilungen, die man numerisch gewinnen kann, sind $G_R(m)$ bzw. $P_R(m)$. Diese kann man mittels der Ausdrücke der rechten Seiten von Gln. (4.16) und (4.17) fitten, sofern R groß genug ist. Es sei hier jedoch nochmal darauf hingewiesen, dass die in der Praxis verwendete Größe bei Tagesdaten, nämlich $R = 365$, nicht mit unendlich gleichzusetzen ist.

Die Bestimmung der Fit-Parameter α und u erfolgt entweder iterativ über die *Maximum-Likelihood* Methode [Sachs 1974] oder über folgende zwei Gleichungen

$$\alpha = \frac{\sqrt{6}}{\pi} \sigma_R \quad \text{und} \quad u = \langle m \rangle - \gamma_e \alpha, \quad (4.18)$$

²Umgekehrt ist auch jede Extremwertverteilung Grenzwertverteilung von $G_R(m)$

welche sich mittels der *Methode der Momente* herleiten lassen [Chow 1988, Stöcker 1995, Hosking 1997, DVWK 1999]. Dabei sind σ_R die Standardabweichung und $\langle m \rangle$ der Mittelwert der Maxima und $\gamma_e = 0.577216$ die Euler'sche Konstante. Da sich die Werte der beiden Möglichkeiten zur Bestimmung der Parameter α und u nur recht wenig unterscheiden, habe ich die Gleichungen der zweiten Methode gewählt. Sie finden aufgrund ihrer Einfachheit auch in der Praxis häufige Anwendung [DVWK 1999].

Die Gumbel-Verteilung (Typ I) wurde von Gumbel in großem Maße bei der Hochwasserschätzung angewendet. Sie hat nur zwei Parameter, Lokalisationsparameter u und Skalenparameter α . Der Lokalisationsparameter u ist gleichzeitig der Modalwert (der wahrscheinlichste Wert) der Verteilung. Der Skalenparameter α ist ein Maß für die Streuung, welche nur von der Varianz σ_R^2 der Maxima abhängt.

4.1.3 Generalisierte Extremwertverteilung GEV

Die drei Extremwertverteilungen, Gumbel (Typ I), Fréchet (Typ II) und Weibull (Typ III), lassen sich ebenfalls als Grenzfälle einer einzigen verallgemeinerten Form, der *General Extremevalue Distribution*, kurz GEV, darstellen [Gumbel 1958, Coles 2001]. Die Form der GEV lautet

$$G\left(\frac{m-u}{\alpha}\right) = \exp\left\{-\left[1 + \xi\left(\frac{m-u}{\alpha}\right)\right]^{-1/\xi}\right\}, \quad (4.19)$$

wobei als zusätzlicher Parameter der *Formparameter* ξ hinzugekommen ist. Typ II (Fréchet) und Typ III (Weibull) entsprechen genau den Fällen $\xi > 0$ und $\xi < 0$ [Coles 2001]. Der Fall $\xi = 0$, eigentlich nicht definiert, wird als Fall $\xi \rightarrow 0$ behandelt und liefert als Grenzverteilung Typ I (Gumbel)³ [Coles 2001],

$$G\left(\frac{m-u}{\alpha}\right) = \exp\left[-e^{-\frac{m-u}{\alpha}}\right], \quad (4.20)$$

wobei $-\infty < \frac{m-u}{\alpha} < \infty$. Die drei Grenzverteilungen unterscheiden sich im wesentlichen in ihren rechten Enden (siehe Abbildung 4.1) und lassen sich mittels ξ klassifizieren in

- Typ I (Gumbel-Verteilung): $\xi \rightarrow 0$: "light tail"
- Typ II (Fréchet-Verteilung): $\xi > 0$: "heavy tail"
- Typ III (Weibull-Verteilung): $\xi < 0$: "bounded tail".

Um zu entscheiden, welche der drei Extremwertverteilungen nun im Speziellen Anwendung findet, muss der Formparameter ξ (numerisch) bestimmt werden. Dies ist nicht immer nötig, da es allgemeine Sätze über diverse Ausgangsverteilungen der Ursprungsdaten x_i gibt, welche Aussagen darüber treffen, gegen welche der drei Extremwertverteilungen die Maximaverteilungen konvergieren.

³Man nutzt die Konvergenz $(1 + \xi \frac{m-u}{\alpha})^{-1/\xi} \rightarrow e^{-\frac{m-u}{\alpha}}$ für $\xi \rightarrow 0$.

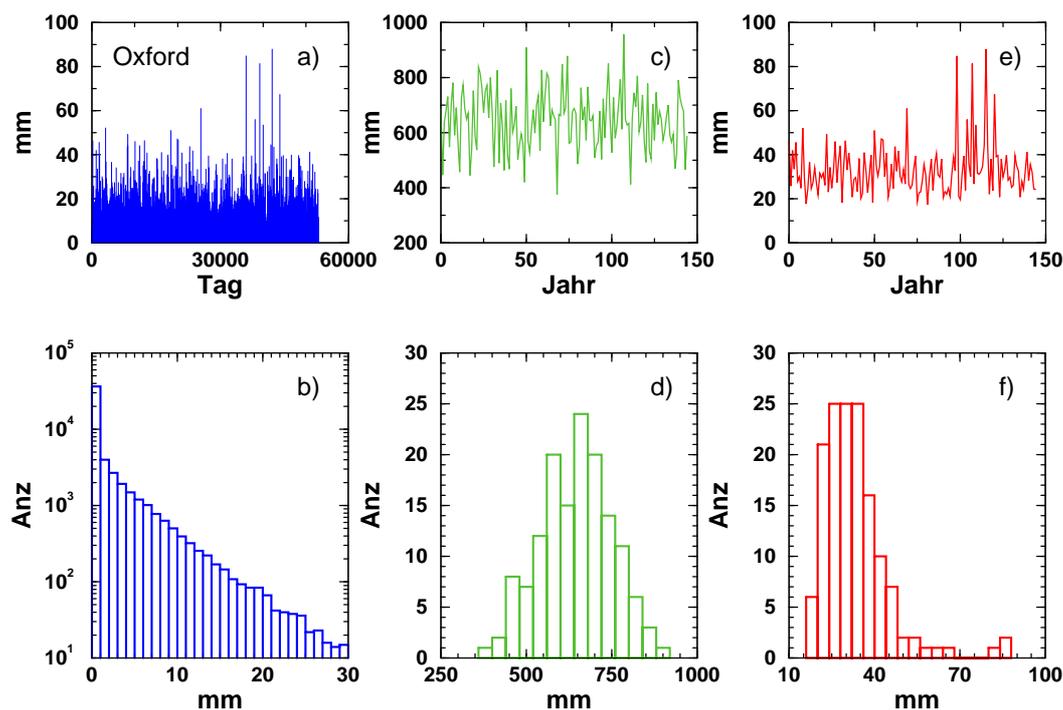


Abbildung 4.2: Zentraler Grenzwertsatz und Extremwert-Theorem visualisiert am Beispiel von Niederschlagsdaten der Stadt Oxford (GB). Abbildung (a) zeigt die in Millimeter (mm) gemessene Reihe täglicher Niederschlagsmengen in Oxford über einen Zeitraum von 144 Jahren (1853 - 1997) (Daten erhalten von D. Rybski [Rybski 2002]). Abbildung (b) zeigt das zugehörige Histogramm in halb-logarithmischer Darstellung. Die lineare Form lässt auf einen exponentiellen Abfall für alle Niederschlagsmengen > 1 mm schließen. Tage ohne Regen (< 1 mm) folgen diesem Schema nicht. Abbildung (c) zeigt die kumulierten Jahresniederschläge, Abbildung (d) das zugehörige Histogramm in linearer Darstellung. Die Form des Histogramms hat bereits deutlich Gauss-Charakter, der Zentrale Grenzwertsatz ist aufgrund der Summation der Tageswerte “sichtbar” geworden. Abbildung (e) zeigt die Jahresmaxima, bestimmt nach der Methode aus der Abbildung 4.3. Das zugehörige Histogramm ist asymmetrischer als das der kumulierten Jahreswerte in (d). Die Rechtslastigkeit ist ein deutliches Indiz für die Annäherung an die Gumbelkurve (Abbildung 4.1 bzw. Typ I aus dem Extremwert-Theorem).

4.1.4 Verteilungen im Anziehungsbereich der Gumbel-Verteilung

Eine der häufigsten Verteilungen in der Natur, insbesondere in Meteorologie und Klimatologie, ist die Gauss-Verteilung, auch *Normalverteilung* genannt. Ihre Häufigkeit hat ihre Begründung im *Zentralen Grenzwertsatz* [Sachs 1974], welcher besagt, dass sich durch Summation oder Mittelung von vielen Zufallszahlen beliebiger Verteilung mit endlichem Mittelwert und endlicher Schwankung (d. h. schwach stationär) als gemeinsame Verteilung der Summenwerte bzw. Mittelwerte die Gauss-Verteilung herauskristallisiert. Diese Eigenschaft eines Prozesses wird auch *Summen-Stabilität* genannt. Ein Beispiel ist der Jahresniederschlag, welcher sich aus kumulierten Tagesniederschlägen zusammensetzt. Während eine Verteilungsfunktion für den Tagesniederschlag noch recht wenig Verwandtschaft mit der Gauss-Kurve aufweist, wird die Ähnlichkeit bei Jahresdaten doch sehr deutlich. Abbildung 4.2 vergleicht die Histogramme der gemessenen Tagesdaten mit den kumulierten Jahresdaten und den Jahresmaxima.

Messgrößen, die sich aus multiplizierten Werten zusammensetzen, sind häufig Log-Normalverteilt [Sachs 1974]. Eine Größe, die sich aus multiplizierten Werten zusammensetzt, wird durch Logarithmieren zu einer normalverteilten Größe, da die Logarithmen der Produkte in Summen zerfallen. Die Eigenschaft eines Prozesses, der zur Log-Normalverteilung führt, wird auch als *Produkt-Stabilität* bezeichnet. Ein Beispiel hierfür sind Gehälter, denn Einkommenserhöhungen erfolgen prozentual (also multiplikativ), sodass große Gehälter stärker erhöht werden als kleine. Auf Dauer bleiben dadurch die vielen kleineren Einkommen niedrig und die wenigen großen wachsen deutlich stärker, was zu einer Schiefe der Verteilung führt (siehe Abbildung 3.3(d)).

In einem der vorherigen Abschnitte wurde bereits die *Maximum-Stabilität* vorgestellt. Prozesse, deren Maxima für beliebig große Intervalle R immer derselben Verteilung gehorchen, sind *Maximum-stabil*. Die einzigen Verteilungsfunktionen, die diese Eigenschaft besitzen, sind die drei Extremwertverteilungen [Leadbetter 1983, Coles 2001]. Aufgrund des Extremwerttheorems konvergieren auch Verteilungsfunktionen von Maxima beliebiger stationärer Prozesse mit anwachsendem R gegen eine dieser drei Extremwertverteilungen.

Es gilt nun zu entscheiden, welche der drei Extremwertverteilungen für den untersuchten Prozess in Frage kommt. In [Leadbetter 1983] sind zwei Theoreme gegeben (Theorem 1.6.1, Seite 16 und Theorem 1.6.2, Seite 17), welche bei bekannter Verteilungsdichte bzw. bei bekannter Verteilungsfunktion Entscheidungskriterien liefern, welcher Typ von Extremwertverteilung zu erwarten ist. Da jedoch in der Praxis die Verteilungsfunktionen nicht bekannt sind, sondern auf Fit-Funktionen beruhen, beschränkt man sich auf das Konvergenzverhalten dieser wenigen Fitfunktionen. Im selben Buch [Leadbetter 1983] wird gezeigt, dass die Gauss-Verteilung, die Exponentialverteilung und die Log-Normalverteilung ebenso wie die Gumbel-Verteilung selber im Einzugsbereich (*domain of attraction*) der Gumbel-Verteilung liegen, während die Potenzgesetz-Verteilung (auch *Pareto-Verteilung* genannt) gegen die Fréchet-Verteilung konvergiert [Leadbetter 1983, Coles 2001]. Alle Aussagen über numerische Ergebnisse der folgenden Kapitel befassen sich mit Daten, die entweder Gauss- oder exponentiell verteilt sind, und damit im Anziehungsbereich der Gumbel-Verteilung liegen. Bei echten Daten ist das Wissen um die Konvergenz nicht gegeben. Trotzdem kann man oft herausfinden, von welcher Art die Verteilung ist und zu welcher der drei Extremwertverteilungen ihre Maxima daher konvergieren. Kumulierte Größen, wie z. B. Jahresniederschläge oder Durchflüsse, sind aufgrund des Zentralen Grenzwertsatzes nahezu Gauss-förmig verteilt und demnach im Anziehungsbereich für Gumbel.

4.1.5 Reale Systeme und endliche Reihenlängen

Alle mathematischen Aussagen und Beweise über das asymptotische Verhalten ($R \rightarrow \infty$) der Maxima setzen unendliche Reihenlängen voraus. Dies ist in realen Systemen natürlich nicht gegeben. Gängige Datensätze von Pegelständen, Abflussdaten, Tagestemperaturen, Niederschlägen oder auch Erdbeben sind selten länger als 150 Jahre. Aufgrund der natürlich vorgegebenen Periodizität von einem Jahr wird R im Falle von Tagesdaten auf $R = 365$ gesetzt. Die Reihe der Tagesdaten wird in Blöcke der Länge 1 Jahr zerteilt und in jedem dieser Blöcke das Jahresmaximum bestimmt. Abbildung 4.3 erläutert diese sogenannte *Blockmaxima*-Methode [Coles 2001].

In der Praxis ist man weniger an der asymptotischen Verteilung $G(\frac{m-u}{\alpha})$ interessiert, als

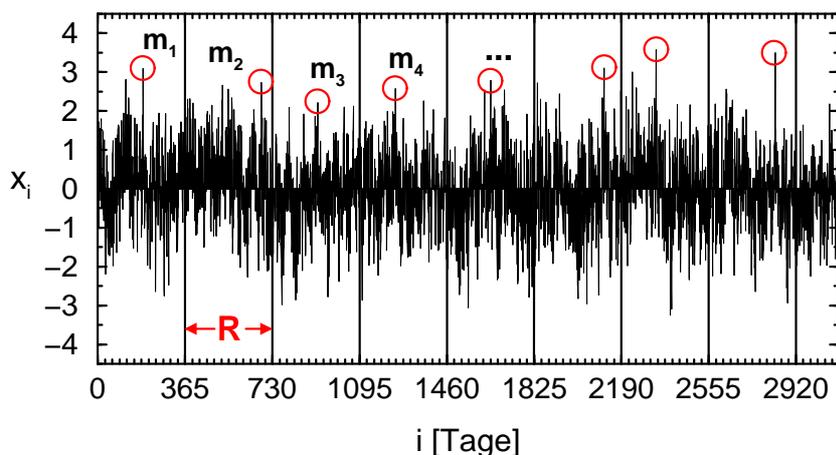


Abbildung 4.3: Definition von Maximawerten einer Zeitreihe (Blockmaxima): eine Zeitreihe aus beispielsweise täglich gemessenen Daten x_i , $i = 1, 2, \dots, N$ wird in Segmente (Blöcke) der Länge R (hier $R = 365$, also Jahre) unterteilt. Die größten Werte in den Segmenten, gekennzeichnet durch die Kreise, bilden die Reihe der Jahresmaximalwerte m_j , $j = 1, 2, \dots, N_R$ mit $N_R = N/R$.

an $G_R(m)$ aus Gl. (4.9). Durch die Beschränkung von R ist die statistische Unabhängigkeit in vielen Datensätzen, insbesondere bei solchen mit Langzeitkorrelationen, nicht mehr gewährleistet. Zwar ist man aufgrund der geringen Statistik der echten Messreihen in der Zuverlässigkeit von Prognosen deutlich beschränkt, andererseits kann aber die statistische Abhängigkeit der Daten auf endlichen Skalen R ausgenutzt werden, um Prognosen über die nahe Zukunft deutlich zu verbessern. Die folgenden Kapitel stellen meine Ergebnisse zum Einfluss von Langzeitkorrelationen auf die Statistik der Maximalwerte, insbesondere auf Jahresmaxima und Jahrhundertmaxima, dar.

4.2 Statistik der Maximawerte bei Langzeitkorrelationen

Das Kapitel ist wie folgt aufgebaut. Als erstes diskutiere ich den Einfluss der Langzeitkorrelationen auf die Verteilungsfunktionen $P_R(m)$ und $G_R(m)$, wobei das besondere Augenmerk bei $R = 365$ liegt, da dies der praxisrelevante Wert (Jahresmaxima) ist. Im nächsten Teil untersuche ich die Autokorrelationen in den Reihen der Maxima und den Einfluss der Systemgrößen auf diese. Danach betrachte ich bedingte Maxima, um den Korrelationseinfluss für verbesserte Vorhersagen auszunutzen. Im letzten Teil dieses Kapitels geht es dann um die in der Hydrologie verwendete Größe HQ_{100} (auch Q_{100} genannt), welche man zur Beschreibung 100-jähriger Ereignisse verwendet. Alle Ergebnisse an künstlichen Daten (Gauss- und exponentiell verteilt) sind mit sehr großer Statistik untermauert. Hinzu kommen Ergebnisse von langzeitkorrelierten echten und rekonstruierten Datensätzen.

Da in diesem Kapitel immer angegeben wird, welcher R -Wert zur Bestimmung der Maxima verwendet wurde, werden die Maxima ab jetzt nur noch mit m bezeichnet. Der Mittelwert

aller Maxima einer Reihe zu gegebenem R wird mit m_R bezeichnet

$$m_R = \langle m \rangle = \frac{R}{N} \sum_{j=1}^{N/R} m_j. \quad (4.21)$$

Die theoretische Größe $G(\frac{m-u}{\alpha})$ spielt von jetzt ab keine wichtige Rolle, es werden nur noch die numerisch bestimmten Verteilungen $P_R(m)$ und $G_R(m)$ betrachtet.

4.2.1 Verteilungen der Maxima

Die Verteilungsfunktion der Maxima hat verschiedene Darstellungen. Während sie formal immer in integrierter Form, also als Wahrscheinlichkeitsverteilung $G_R(m)$, angegeben ist, wird graphisch zumeist die Form der normierten Wahrscheinlichkeitsdichte $P_R(m)$, also die Ableitung der Wahrscheinlichkeitsverteilung gewählt. Aber auch die Dichte hat zwei Darstellungsmöglichkeiten. Die eine ist das einfache normierte Histogramm $P_R(m)$ der tatsächlich gemessenen Maxima. Die andere ist eine reskalierte Form, in der die Maxima um den Lokalisationsparameter u verschoben und in Einheiten des Skalenparameters α angegeben werden. Diese letzte Version hat, wie in den vorherigen Abschnitten bereits erwähnt, den Vorteil, dass die Kurven für verschiedene R übereinander liegen, sofern R groß genug gewählt wurde. Die integrierte Wahrscheinlichkeitsdichte $G_R(m)$, also die Wahrscheinlichkeitsverteilung, basiert auch auf der reskalierten Darstellung, welche formal durch die drei Typen der Extremwertverteilungen beschrieben wird. Die Abbildungen 4.4 und 4.5 vergleichen die numerischen Verteilungen für unkorrelierte und langzeitkorrelierte Daten.

Abbildung 4.4 zeigt numerische Ergebnisse dieser drei Darstellungen der Maxima-Verteilungen für ursprünglich Gauss- (a,c,e) und exponentiell (b,d,f) verteilte unkorrelierte x -Daten. Abbildung (a) zeigt das normierte Histogramm $P_R(m)$ der Maxima m für vier verschiedene R -Werte ($R = 6, 30, 365$ und 1500). Man kann gut erkennen, wie sich mit anwachsendem R die Form der Kurve von der Gauss'schen Glockenkurve wegentwickelt und sich einer rechtslastigen Maxima-Verteilung annähert. Abbildung (b) zeigt dieselben Kurven für ursprünglich exponentiell verteilte x -Daten. Während für $R = 6$ (Kreise) noch die Form der Exponentialverteilung am abrupten linken Anstieg zu erahnen ist, kann man bei den größeren R -Werten nur eine Verschiebung nach rechts beobachten, aber keine Formveränderung mehr. Abbildungen (c) und (d) zeigen dieselben Kurven in der reskalierten Darstellung. Die durchgezogene schwarze Kurve ist der analytische Ausdruck für die Ableitung der Gumbel-Kurve. Man sieht hier, dass die gewählten (realistischen) R -Werte bei den Gauss-Zahlen noch nicht in der Größenordnung sind, in der der Gumbel-Fit die Realität beschreiben könnte. Zwar ist die Kurve für $R = 1500$ (Dreiecke) steiler als die Kurve für $R = 6$, aber beide sind noch deutlich von der Ableitung der Gumbel-Kurve entfernt. Im Falle der exponentiell verteilten Daten liefern die Kurven in dieser Darstellung rein optisch bereits einen perfekten Datenkollaps. Die Abbildungen (e) und (f) zeigen die integrierten reskalierten Histogramme $G_R(m)$ mit einer zweifach logarithmierten y -Achse. Die Gumbel-Kurve $G(\frac{m-u}{\alpha})$ (durchgezogene schwarze Linie) erscheint so als Gerade. Man sieht hier, dass die Konvergenz für die gewählten R -Werte insbesondere bei kleinen und großen Maxima m für Gauss-verteilte Daten noch recht schlecht ist, bei den exponentiell verteilten Daten hingegen ist der Fit exzellent.

Abbildung 4.5 zeigt dieselben Kurven für Gauss- und exponentiell verteilte Daten, die mit $\gamma = 0.4$ langzeitkorreliert sind. Die Kurven in Abbildungen (a) und (b) wirken gegenüber dem

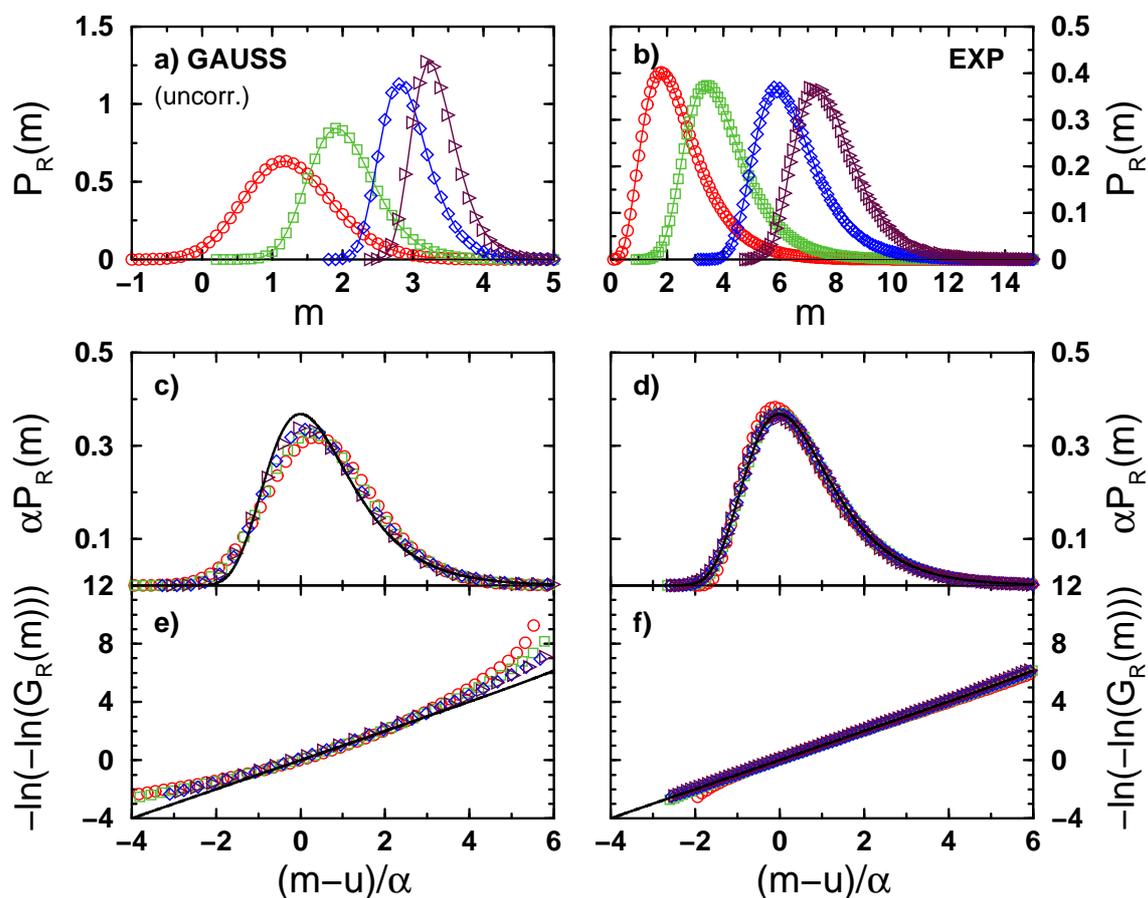


Abbildung 4.4: Verteilungskurven der Maxima aus Segmenten unterschiedlicher Längen R für unkorrelierte Daten x_i mit (a,c,e) Gauss'scher und (b,d,f) exponentieller Verteilungsdichtefunktion $P(x)$. Abbildungen (a,b) zeigen die Verteilungsdichtefunktionen $P_R(m)$ von Maxima zu vier verschiedenen R -Werten: $R = 6$ (\circ), 30 (\square), 365 (\diamond) und 1500 (\triangleright). Skaliert man die m -Achse mit $(m-u)/\alpha$ und multipliziert man $P_R(m)$ mit α , so erhält man einen Kollaps dieser Kurven, siehe (c,d). Die durchgezogene Kurve ist die Gumbel-Verteilungsdichte Gl. (4.17). Abbildungen (e,f) zeigen die zugehörigen integrierten Verteilungen $G_R(m)$ zusammen mit der Gumbel-Verteilung Gl. (4.16). Zu beachten ist, dass die Maxima der exponentiell verteilten Daten deutlich schneller gegen die Gumbelkurve konvergieren, als die der Gauss-verteilten Daten. Abbildung nach [Eichner 2006a].

unkorrelierten Fall in Abbildung 4.4 gedrungener. Der reskalierte Fall zeigt insbesondere für Gauss (c) deutliche Abweichungen vom Gumbel-Fit (durchgezogene Linie). Bei exponentiellen Daten (d) scheinen die Abweichungen nur bei $R = 6$ deutlich zu sein. Dies hat aber eine simple Erklärung: $R = 6$ ist ein sehr kleiner R -Wert, weshalb in $P_6(m)$ und $G_6(m)$ noch das abrupte linke Ende der Exponentialverteilung durchschlägt. Dies ist bei korrelierten Daten stärker als bei unkorrelierten, da im korrelierten Fall auch kleinere Werte als Maxima identifiziert werden⁴. Bei größeren R -Werten scheint dieser Effekt wieder zu verschwinden. Die Darstellungen in (e) und (d) zeigen, dass die Konvergenz gegen die Gumbel-Kurve (durchgezogene Linie) deutlich stärker von der Ausgangsverteilung $P(x)$ bestimmt wird, als von den Langzeitkorrelationen. Langzeitkorrelationen scheinen die Konvergenz nur zu verlangsamen. Da die Gumbel-Verteilung erst für $R \rightarrow \infty$ erreicht werden muss, ist dieses Resultat insbesondere bei langzeitkorrelierten exponentiell verteilten Daten erstaunlich, da R deutlich kleiner als die Korrelationslänge $s_x = \infty$ ist und somit keine statistische Unabhängigkeit der Maxima gewährleistet ist. Dies ist jedoch Voraussetzung für die Gültigkeit des Extremwerttheorems. S. M. Berman hat bereits 1964 in [Berman 1964] gezeigt, dass auch bei langzeitkorreliertem Verhalten die Konvergenz für $R \rightarrow \infty$ gewährleistet ist. Eine abgewandelte Form des Beweises findet sich auch in [Leadbetter 1983]. Jedoch basieren beide Versionen des Beweises für $R \rightarrow \infty$ auch auf der Tatsache, dass dann R nicht mehr kleiner als s_x ist. Wir finden aber, dass die Konvergenz im Falle exponentiell verteilter Daten bereits bei endlichen R -Werten beliebig gut erreicht wird.

In Abbildung 4.6 sind die unskalierten normierten Histogramme $P_R(m)$ für $R = 365$, also für Jahresmaxima von unkorrelierten Daten (rot) und langzeitkorrelierten Daten (blau) übereinandergelegt. Zunächst fällt auf, dass die Korrelationen neben einer leichten Linksverschiebung des Schwerpunkts auch eine deutliche Linksverbreiterung der Histogramme verursachen. Es tauchen in beiden Fällen der hier untersuchten Gauss- und exponentiell verteilten Daten bei Langzeitkorrelationen deutlich mehr kleine Maxima im Histogramm auf. Eine Erklärung hierfür liefert Abbildung 4.7. Die Langzeitkorrelationen in den x_i verursachen Bereiche, in denen sich große x -Werte häufen, und auch Bereiche, in denen sich kleine x -Werte häufen. Es bilden sich regelrechte Berge und Täler in einer langzeitkorrelierten Zeitreihe aus. Wenn man nun die Maxima m aus den Blöcken der korrelierten Daten selektiert, ergeben die Bereiche, in denen die kleineren x -Werte dominieren (Täler), auch eher mal kleinere Maxima, verglichen mit den unkorrelierten Daten, in denen die großen x gleichmäßiger verteilt sind. Diese kleinen m -Werte verursachen die Linksverbreiterung von $P_R(m)$. Die größten Werte jedoch bleiben davon unbeeinflusst, da sie in jedem Fall als Maxima identifiziert werden. Das rechte Ende von $P_R(m)$ bleibt daher ebenfalls so gut wie unbeeinflusst. Dies wird in Abbildung 4.6 verdeutlicht, wenn man statt $P_R(m)$ die Überschreitungswahrscheinlichkeit $E_R(m)$ betrachtet. Sie definiert die Wahrscheinlichkeit, dass ein Wert m auftaucht, der größer als eine vorgegebene Schranke m^* ist,

$$E_R(m^*) := \int_{m^*}^{\infty} P_R(m') dm' = 1 - G_R(m^*). \quad (4.22)$$

Die ermittelten Werte von $E_R(m^*)$ (die hellblauen Flächen unter den Kurven) für ein beliebig gewähltes m^* sind der Bildunterschrift zu entnehmen. Die Werte für langzeitkorrelierte und unkorrelierte Daten unterscheiden sich nur gering, egal ob die ursprünglichen Daten einer Gauss- oder einer Exponentialverteilung folgen.

⁴Die Ursache hierfür liegt an der ausgeprägten ‘‘Berge & Täler’’ Struktur der langzeitkorrelierten Daten. In den Tälern werden dadurch auch kleine Werte als Maxima identifiziert

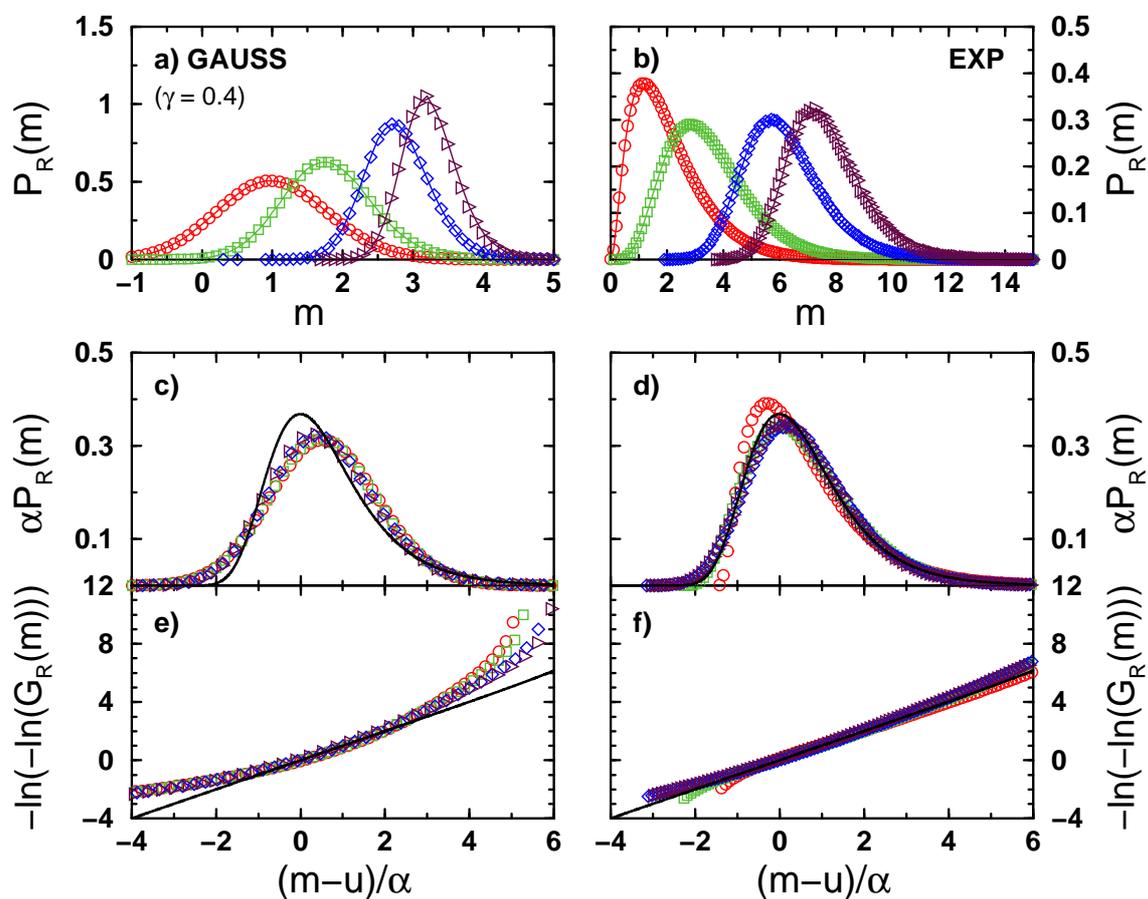


Abbildung 4.5: Verteilungskurven der Maxima aus Segmenten unterschiedlicher Längen R für langzeitkorrelierte Daten ($\gamma = 0.4$) mit (a,c,e) Gauss'scher und (b,d,f) exponentieller Verteilungsdichtefunktion $P(x)$. Zur Bedeutung der Symbole und Beschriftungen, siehe Abbildung 4.4. Die Kurven in (a,b) sind breiter und gedrungener als jene in Abbildung 4.4, da in korrelierten Daten mehr *kleine* Maxima in $P_R(m)$ berücksichtigt werden, als in unkorrelierten Daten. In Abbildung (d) weicht die Kurve für $R = 6$ von der theoretischen Kurve bei unkorrelierten Daten (durchgezogene Linie) am deutlichsten ab. Aufgrund der Korrelationen und des sehr kleinen R -Werts werden auch Werte nahe Null (der abrupten unteren Grenze der Exponentialverteilung) als Maxima identifiziert. Für große R -Werte verschwindet dieser Effekt und die Gumbel-Verteilungsdichte wird gut angenähert. Auch hier ist die Konvergenz der Gauss-verteilten Daten in (e) gegen die Gumbel-Kurve (durchgezogene Linie) deutlich schlechter als bei exponentiell verteilten Daten (f). Abbildung nach [Eichner 2006a].

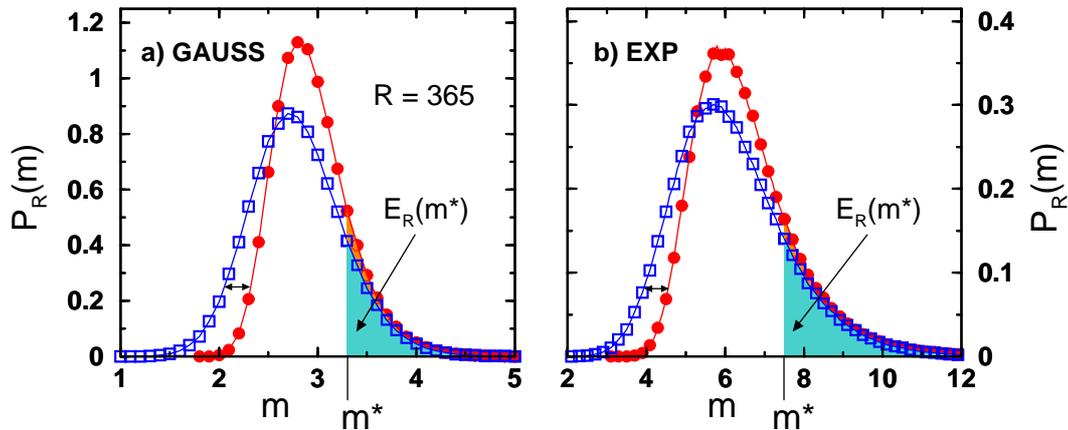


Abbildung 4.6: Vergleich der Verteilungsdichtefunktionen $P_R(m)$ zwischen unkorrelierten (\circ) und langzeitkorrelierten (\square , $\gamma = 0.4$) Gauss-verteilt (a) und exponentiell verteilt (b) Daten für $R = 365$. In beiden Fällen (a,b) wirken sich die Langzeitkorrelationen in Form einer Links-Verbreiterung auf $P_R(m)$ aus. Das rechte Ende der Verteilung bleibt nahezu unbeeinträchtigt. Dadurch wird der Einfluss auf die Überschreitungswahrscheinlichkeit $E_R(m^*)$, also die Wahrscheinlichkeit in der Reihe der Maximawerte einen Wert m größer als eine beliebig gewählte aber hinreichend große Schranke m^* zu finden, insbesondere für große Schranken m^* nur sehr gering. In (a) ist $E_R(3.3) = 0.14$ für unkorrelierte Daten (orange Fläche, durch hellblaue Fläche zum größten Teil verdeckt), für korrelierte hingegen 0.16 (hellblaue Fläche). In (b) ergibt sich analog $E_R(7.4) = 0.18$ für unkorrelierte und 0.20 für korrelierte Daten. Abbildung nach [Eichner 2006a].

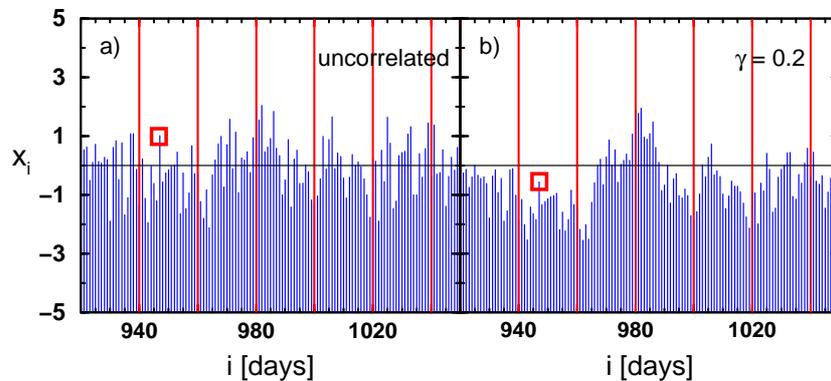


Abbildung 4.7: Vergleich der Maxima-Selektion bei unkorrelierten und langzeitkorrelierten Daten. Beide Bilder zeigen einen Ausschnitt von etwa 130 unkorrelierten (a) und langzeitkorrelierten (b) Tagesdaten. In langzeitkorrelierten Daten treten gegenüber unkorrelierten Daten ausgeprägte Berge und Täler auf, sodass sich bei kleinen Blocklängen R (hier $R = 20$) die kleinsten Maximawerte deutlich unterscheiden. Während in (a) das kleinste Maximum etwa bei 1 liegt (rotes Kästchen), ist in (b) der Wert noch niedriger als -0.5 (rotes Kästchen). Aufgrund der ausgeprägten Täler kommen im korrelierten Fall wesentlich häufiger kleine Maxima vor als im unkorrelierten Fall. Dieser Effekt führt zur Links-Verbreiterung von $P_R(m)$ (siehe Abbildung 4.6). Bei großen R -Werten wird dieser Effekt schwächer: Das Maximum aller 130 Datenpunkte liegt in beiden Ausschnitten etwa bei $i = 980$ und hat einen Wert von etwa 2. Da die Korrelationen nur eine Anordnungs-eigenschaft sind und die statischen Eigenschaften wie Stationarität nicht beeinträchtigen, gleichen sich die m -Werte mit anwachsendem R an.

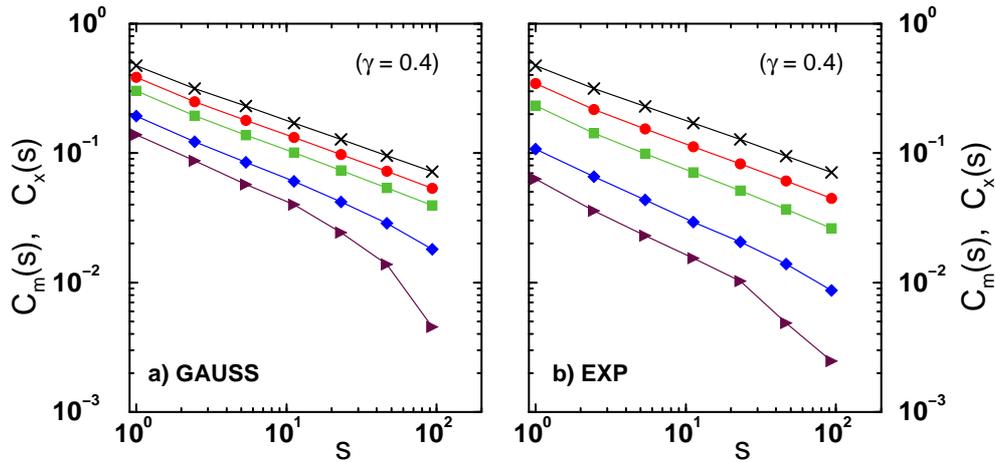


Abbildung 4.8: Autokorrelationsfunktion $C_m(s)$ der Maximalwerte m von (a) Gauss-verteilt und (b) exponentiell verteilten langzeitkorrelierten Daten x mit $\gamma = 0.4$ zu verschiedenen R -Werten, $R = 6$ (o), 30 (□), 365 (◇) und 1500 (▷). Die Autokorrelationsfunktion $C_x(s)$ der Originaldaten x (×) hat eine Steigung von -0.4 , d. h. einen Korrelationsexponenten $\gamma = 0.4$. Abbildung nach [Eichner 2006a].

4.2.2 Korrelationen in den Maxima

Die Verteilungen der Maxima aus dem vorherigen Abschnitt geben keine Information preis, ob die mittels der Blockmaxima-Methode erzeugte Reihe der Maxima $(m_j)_{j=1}^{N/R}$ auch korreliert sind. Hierzu muss die Autokorrelationsfunktion aus Gl. (2.2.1) der Maximareihen bestimmt werden,

$$C_m(s) = \frac{\langle (m_j - m_R)(m_{j+s} - m_R) \rangle}{\langle (m_j - m_R)^2 \rangle}. \quad (4.23)$$

Abbildung 4.8 zeigt die Ergebnisse der Autokorrelationsfunktionen für Maxima bei Gauss- und exponentiell verteilten, mit $\gamma = 0.4$ langzeitkorrelierten Daten. Die oberste Kurve ist die Autokorrelationsfunktion der originalen x -Werte, darunter sind die Kurven $C_m(s)$ für verschiedene R -Werte zu sehen. Der Vergleich der Kurven zeigt, dass die Maximareihen bei langzeitkorrelierten Daten ebenfalls Langzeitkorrelationen in Form eines Potenzgesetz-artigen Abfalls von $C_m(s)$ aufweisen (die bei großen R -Werten jedoch abweichen). Die ermittelten Korrelationsexponenten γ' sind sogar dem γ der ursprünglichen Daten sehr ähnlich. In der Maximareihe folgen demnach auf große m -Werte eher wieder große m -Werte und entsprechend auf kleine m wieder eher kleine m . Dieses Verhalten führt zu einem *Clustern* (Anhäufen) von großen bzw. kleinen Werten in den Maximareihen. Vorhandene Langzeitkorrelationen können somit eine Ursache für gehäuftes Auftreten von Katastrophen sein, aber ebenso auch untypisch lange Ruhephasen erklären. Das Verhalten der Maxima ist dem der Wiederkehrintervallen aus dem Kapitel 3 sehr ähnlich [Bunde 2005].

Was man in Abbildung 4.8 auch sieht, ist ein stärkeres Abfallen von $C_m(s)$ mit größer werdendem R -Wert. Die Ursache hierfür liegt nicht in einem Verlust der Korrelationen je extremer die Ereignisse im Mittel werden. Es ist vielmehr ein Finite-Size Effekt aufgrund der endlichen Reihelängen. Je größer R , um so geringer die Anzahl N/R der Maxima, da N fest vorgegeben ist. Bei einer Reihelänge von $N = 10^6$ und einem R -Wert von z. B. 1000 bleiben

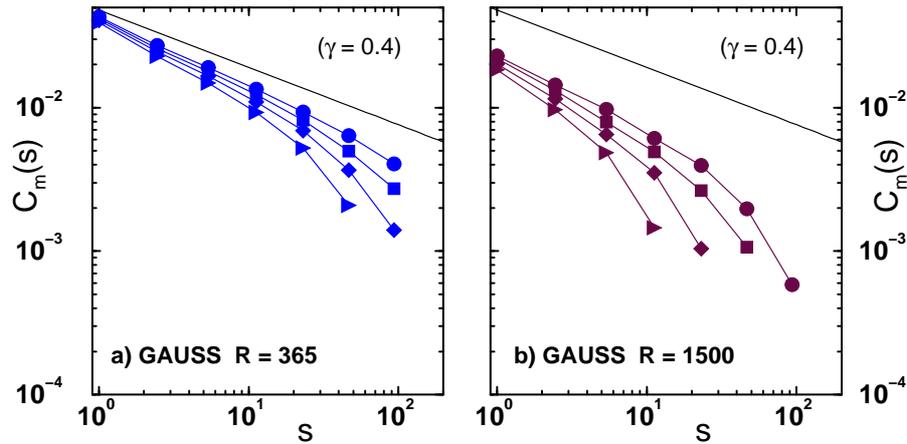


Abbildung 4.9: Finite-Size-Untersuchungen an der Autokorrelationsfunktion $C_m(s)$ der Maximawerte von Gauss-verteiltern langzeitkorrelierten Daten für zwei verschiedene R -Werte ($R = 365$ (a) und 1500 (b)). Die Reihenlängen sind $N = 2^{21}$ (o), 2^{20} (□), 2^{19} (◇) und 2^{18} (▷). Das Abfallen der Steigungen der Kurven gegenüber der durchgezogenen Linie ($-\gamma = -0.4$) nimmt mit der Länge des Systems ab. Diese Systemgrößenabhängigkeit der Korrelationen deutet auf einen Finite-Size Effekt. Abbildung nach [Eichner 2006a].

nur noch $N/R - s < 10^3 = 1000$ Datenpunkte für die Autokorrelationsfunktion übrig. Dies reicht nicht aus, um die Korrelationen auf großen Skalen mittels Autokorrelationsfunktion korrekt bestimmen zu können. Die DFA ist hier zwar robuster, wird jedoch vom recht starken Rauschanteil auf kleinen bis mittleren Skalen überlagert. Dieser Finite-Size Effekt wird umso drastischer, je kürzer die Reihenlänge N ist. Abbildung 4.9 zeigt nochmal $C_m(s)$ für festes R (365 bzw. 1500) aber für unterschiedliche Reihenlängen ($N = 2^{21}$, 2^{20} , 2^{19} und 2^{18}) bei langzeitkorrelierten Gauss-Zahlen. Man erkennt deutlich wie der Finite-Size Effekt mit zunehmender Reihenlänge abnimmt. Das ist ein starkes Indiz dafür, dass in einer ausreichend langen Reihe die Korrelationen wie in den Ausgangsdaten auch bei großen R -Werten in Form eines Potenzgesetz erscheinen. Im Falle unkorrelierter Ausgangsdaten zeigen auch die Maxima keine Korrelationen, d. h. $C_m(s) = 0$ für $s > 0$. Dies ist in der doppellogarithmischen Auftragung nicht darzustellen.

4.2.3 Bedingte Maxima

Aufgrund der Langzeitkorrelationen in der Reihe der Maxima-Werte hängt die Wahrscheinlichkeit, einen m -Wert einer bestimmten Größe zu finden, von der Vergangenheit, d. h. von den vorhergehenden Werten, ab. Insbesondere nimmt der unmittelbar vorherige Wert m_{j-1} in der Maximareihe $(m_j)_{j=1}^{N/R}$ einen besonders starken Einfluss auf den folgenden Wert m_j . Dieser Effekt kann von großer Bedeutung sein bei der Vorhersage extremer Ereignisse und somit bei jeder Form von Risikoabschätzung in langzeitkorrelierten Systemen. Da es im Folgenden nicht um die exakte Position der m -Werte innerhalb der Maximareihe geht, sondern nur um eine quantitative Analyse aufeinanderfolgender Werte, wird m_{j-1} mit m_0 und m_j mit m bezeichnet. Dies wurde bereits völlig analog im Kapitel 3 über die Wiederkehrintervalle getan.

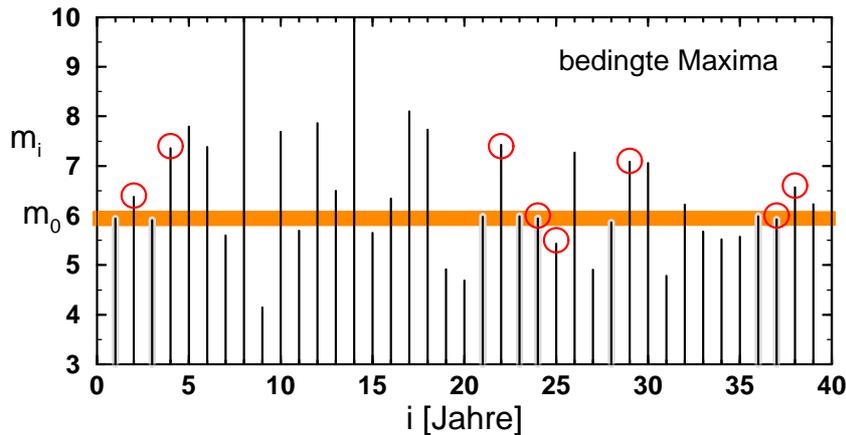


Abbildung 4.10: Definition bedingter Maximawerte. In der Reihe der Maxima werden nur solche m -Werte berücksichtigt, welche unmittelbar auf ein Maximum vom (ungefähren) Wert m_0 folgen (hier $m_0 \simeq 6$, oranges Band). Diese neue Reihe von m -Werten, gekennzeichnet durch Kreise, ist die Reihe der bedingten Maxima.

Den Einfluss der vergangenen Maxima auf zukünftige zu untersuchen, funktioniert mittels *bedingter* Maxima oder auch *konditionierter* Maxima (siehe Abbildung 4.10). In der Reihe der Maxima werden nur solche m -Werte berücksichtigt, welche die Bedingung erfüllen, dass sie unmittelbar auf einen m -Wert folgen, der etwa die Größe m_0 hat. In der Abbildung 4.10 wird m_0 durch ein oranges Band der Breite Δm_0 dargestellt. Dabei lässt sich die Breite Δm_0 so wählen, dass man für jede Bedingung immer noch genügend Werte pro untersuchter Reihe erhält. Analog zum mittleren Maximum m_R einer Reihe wird das mittlere m_0 -bedingte Maximum, also der Mittelwert aus den m_0 -bedingten Maxima, mit $m_R(m_0)$ bezeichnet. Für unkorrelierte Daten ist $m_R(m_0)$ unabhängig von m_0 . Durch Mischen der Originaldaten werden die Korrelationen zerstört und man erhält $m_R(m_0) \simeq m_R$ bis auf statistische Schwankungen, da sich $m_R(m_0)$ nur aus einer Untermenge aller m -Werte zusammensetzt.

In Abbildung 4.11 sind die mittleren bedingten Maxima für Gauss- und exponentiell verteilte langzeitkorrelierte ($\gamma = 0.4$) Daten als Funktion von m_0 dargestellt. Die Breite Δm_0 der Bedingungen wurde so gewählt, dass sich jeder Punkt aus etwa 700 bedingten m -Werten zusammensetzt. Der korrelationsbedingte Memory-Effekt ist deutlich an der Steigung bzw. Krümmung der Kurven zu erkennen. Auf kleine m_0 ($m_0 < m_R$) folgen im Mittel wieder eher kleine m -Werte ($m < m_R$) und auf große m_0 ($m_0 > m_R$) eher wieder große m -Werte ($m > m_R$). Die durchgezogenen horizontalen Linien entsprechen den unbedingten m_R -Werten. Je größer R , um so größer ist natürlich auch m_R , da die m -Werte mit anwachsendem R extremer werden. Der Memory-Effekt wird bei kleiner Blocklänge R stärker betont, er ist aber bei großen R (sogar $R = 1500$, Dreiecke) immer noch zu erkennen. Es gibt außerdem einen deutlichen Unterschied im Krümmungsverlauf der Kurven zwischen (a) Gauss- und (b) exponentiell verteilten Daten bei kleinen R -Werten. Die Gauss-basierten bedingten mittleren Maxima scheinen eine fast lineare Abhängigkeit von m_0 zu haben, während die exponential-basierten mittleren Maxima eine deutliche Krümmung bei kleinen m_0 aufweisen. Dies liegt an der asymmetrischen Form der Exponentialverteilung, welche abrupt bei $x \leq 0$ auf Null abfällt, was dadurch auch $m_{min} = 0$ bedeutet.

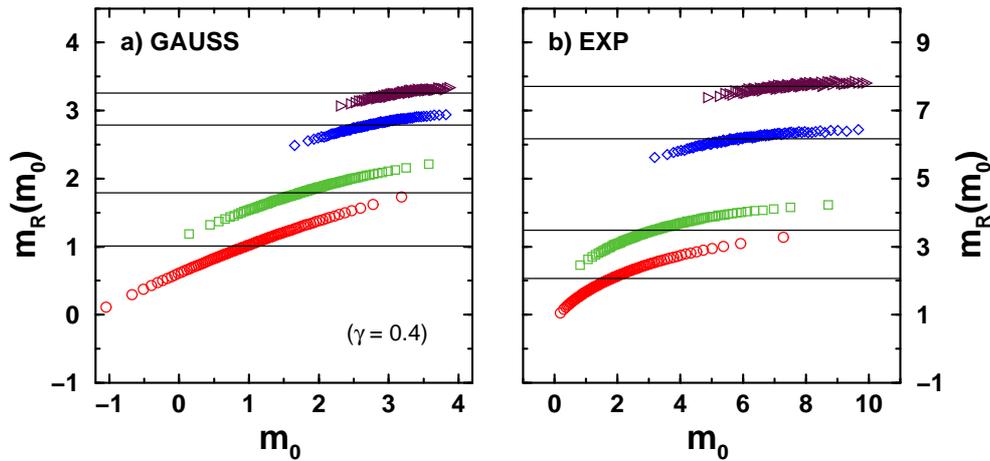


Abbildung 4.11: Mittlere bedingte Maxima $m_R(m_0)$ für $\gamma = 0.4$ und $R = 6$ (\circ), 30 (\square), 365 (\diamond) und 1500 (\triangleright) als Funktion von m_0 für (a) Gauss- und (b) exponentiell verteilte Daten. Die durchgezogenen horizontalen Linien repräsentieren die unbedingten Mittelwerte m_R zu gegebenen R . Beide Abbildungen zeigen den Memory-Effekt in Form von $m_R(m_0) > m_R$ für $m_0 > m_R$, bzw. $m_R(m_0) < m_R$ für $m_0 < m_R$, d. h. auf große m -Werte folgen eher große m und auf kleine m -Werte eher kleine. Die Breite Δm_0 (das orange Band in Abbildung 4.10) wurde so gewählt, dass in jedem der 150 Runs mit $N = 2^{21}$ etwa 700 m -Werte zu jedem m_0 zur Verfügung waren. Abbildung nach [Eichner 2006a].

Dieser Effekt lässt sich auch an echten, langzeitkorrelierten Datenreihen verifizieren. Als Beispiel habe ich zum einen die historische Datenreihe der jährlichen Niedrigwasserstände des Nil (siehe auch Kapitel 2.3.2), gemessen in den Jahren 622 AD bis 1284 AD in Roda, in der Nähe von Kairo (EGY) [Tousson 1925, Beran 1994], und zum anderen die rekonstruierten jährlichen Temperaturen der nördlichen Hemisphäre nach Moberg [Moberg 2005] zwischen den Jahren 1 AD und 1979 AD gewählt. Beide Reihen zeigen eine ähnliche Korrelationsstruktur mit einem Korrelationsexponent $\gamma = 0.3$ (Nil: siehe Abbildung 2.16 in Abschnitt 2.3.2; nördliche Hemisphäre nach Moberg: siehe [Rybski 2006]). Extreme Minima eines Flusses (Trockenheiten) sind in der Bewässerung von ähnlich gravierender Bedeutung wie Hochwasser. Aufgrund der recht symmetrischen Verteilung der Jahresminima des Nil (siehe Abbildung 2.15) lassen sich mit den bisher verwendeten Methoden durch $x_i \rightarrow -x_i$ auch Minima untersuchen. Abbildung 4.12 zeigt die bedingten Minima des Nil (a,b,c) und die bedingten Maxima der Moberg-Temperaturreihe (d,e,f). Der R -Wert steht hier für Jahre (nicht Tage), weshalb $R = 1$ also 365 Tagen entspricht. Um trotz der kurzen Reihen noch eine vernünftige Statistik zu erreichen, wurden nur sechs m_0 -Werte verwendet, wobei Δm_0 so groß war, dass zum einen jeder $m_R(m_0)$ -Wert auf einer gleichen Statistik beruht und zum anderen keine Lücken zwischen den sechs m_0 -Werten entstanden, um soviel Information wie möglich aus den Reihen zu bekommen. Von der Nil-Reihe wurden die letzten 660 Datenpunkte verwendet, von der Moberg-Reihe die letzten 1968, da diese Zahlen gemeinsame Vielfache von $R = 1, 6$ und 12 sind. Die geschlossenen Symbole in Abbildung 4.12 zeigen deutlich eine m_0 -Abhängigkeit, qualitativ ähnlich wie man sie auch bei den künstlichen Daten findet. Die bedingten Mittelwerte sind umso kleiner, je kleiner die Bedingung m_0 ist bzw. umso größer, je größer m_0 ist. Zum Vergleich zeigt die Abbildung auch die bedingten Minima bzw. Maxima der gemischten Reihen, in denen keine Korrelationen vorhanden sind (offene Symbole). Sie sind nahezu

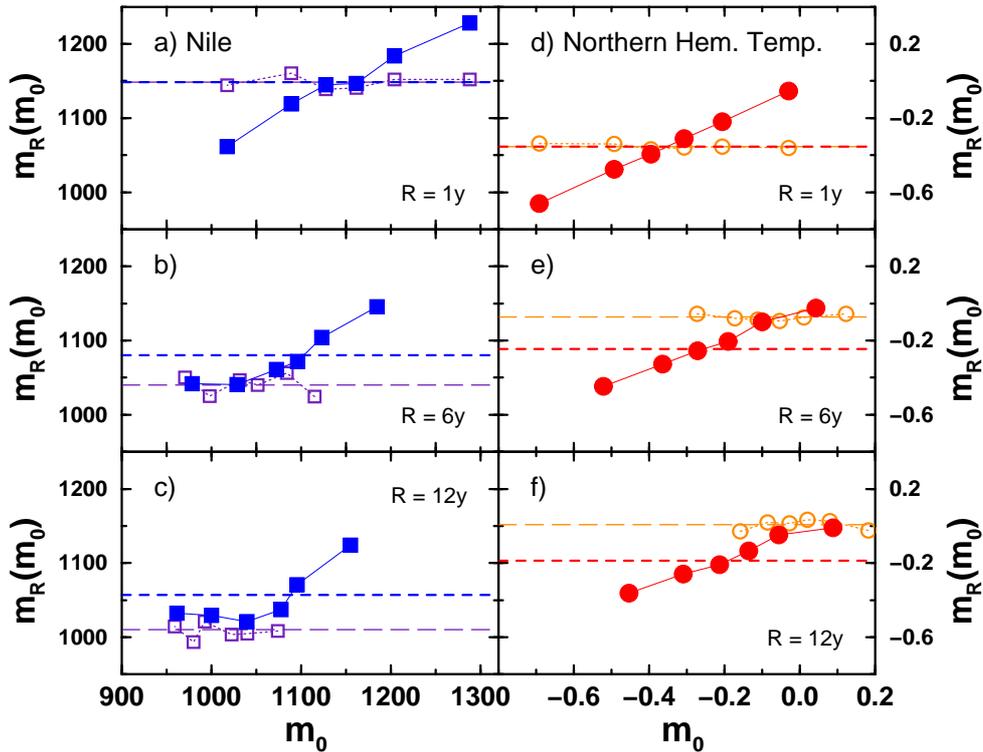


Abbildung 4.12: Mittlere bedingte Minima $m_R(m_0)$ der jährlichen Wasserniedrigststände des Nil (\square) (a,b,c) [Tousson 1925, Beran 1994] und mittlere bedingte Maxima $m_R(m_0)$ der rekonstruierten Jahrestemperaturen der nördlichen Hemisphäre (\circ) (d,e,f) nach Moberg [Moberg 2005] für $R = 1$ Jahr (a,d), 6 Jahre (b,e) und 12 Jahre (c,f). Die gefüllten Symbole zeigen die Resultate für die echten Datenreihen, die offenen Symbole für dieselben Daten, die jedoch zuvor mehrfach gemischt wurden, um alle Korrelationen in den Reihen zu zerstören, weshalb hier die aufsteigende Tendenz mit wachsendem m_0 nicht zu beobachten ist. Die gemischten Daten haben einen kleineren m_0 -Bereich, da sie nicht der korrelationsbedingten Links-Verbreiterung von $P_R(m)$ (im Falle der Minima des Nils einer Rechtsverbreiterung) unterworfen sind (siehe Abbildung 4.6). Die kurz-gestrichelten Linien zeigen die unbedingten mittleren Minima (a,b,c) und Maxima (d,e,f), die lang-gestrichelten Linien dasselbe für die gemischten Daten. Abbildung nach [Eichner 2006a].

plateauförmig und hängen deshalb (wie erwartet) nicht von der Vergangenheit ab.

4.2.4 Bedingte Maximaverteilungen und bedingte Überschreitungswahrscheinlichkeiten

Die Größe $m_R(m_0)$ ist das erste Moment der bedingten Verteilungsdichte $P_R(m|m_0)$

$$m_R(m_0) := \int_{-\infty}^{\infty} m P_R(m|m_0) dm \quad (4.24)$$

und damit ein Erwartungswert. Man kann aus dem zugehörigen numerischen Mittelwert jedoch keine Aussagen über die Häufigkeit anderer Ereignisse treffen. Hierzu benötigt man die bedingte Verteilungsdichte $P_R(m|m_0)$. Die bedingte Verteilungsdichte $P_R(m|m_0)$ ist definiert

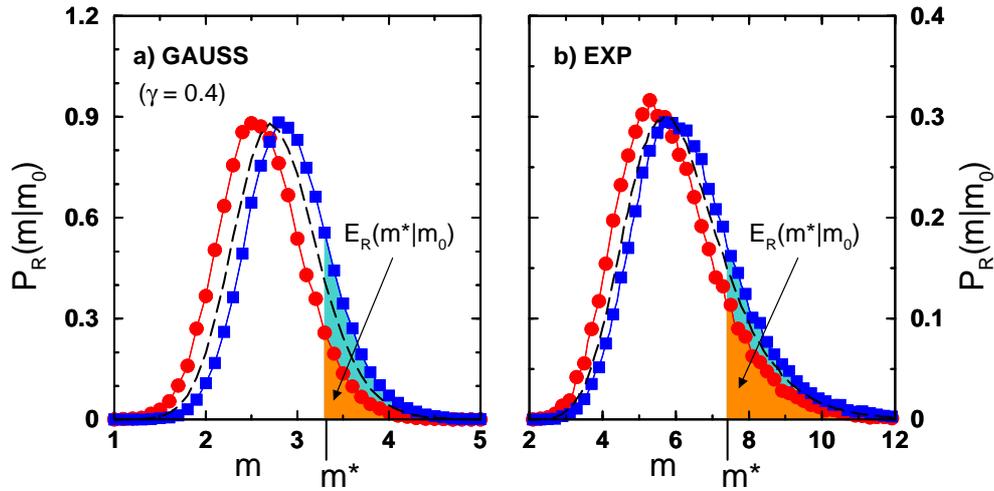


Abbildung 4.13: Bedingte Verteilungsdichtefunktionen $P_R(m|m_0)$ der Maxima langzeitkorrelierter ($\gamma = 0.4$) Gauss-verteilter Daten (a) und exponentiell verteilter Daten (b) für $R = 365$. Abbildung (a) zeigt $P_R(m|m_0)$ für zwei m_0 -Werte, $m_0 = 2.06$ (\circ) und $m_0 = 3.55$ (\square). In Abbildung (b) sind die Werte $m_0 = 4.10$ (\circ) und $m_0 = 8.65$ (\square). Die Breite Δm_0 wurde so gewählt, dass in jedem der 150 Runs mit $N = 2^{21}$ Datenpunkten etwa 700 m -Werte zu jedem m_0 zur Verfügung waren. Die Verschiebung der Kurven ist ein Resultat der Langzeitkorrelationen in den Maximareihen. Der Einfluss auf die (bedingte) Überschreitungswahrscheinlichkeit $E_R(m|m_0)$ ist deutlich zu erkennen: In (a) ergibt sich für $m^* = 3.30$ $E_{365}(3.30|2.06) = 0.08$ (orange Fläche) und $E_{365}(3.30|3.55) = 0.20$ (hellblaue Fläche, teilweise überdeckt durch die orange Fläche). Die analogen Zahlenwerte in (b) für $m^* = 7.35$ sind $E_{365}(7.35|4.10) = 0.14$ und $E_{365}(7.35|8.65) = 0.24$. Die gestrichelten Linien sind die unbedingten Verteilungsdichten $P_R(m)$. Die numerischen Mittelwerte m_R betragen 2.79 für Gauss-Daten und 6.17 für exponentielle Daten. Abbildung nach [Eichner 2006a].

als die Verteilungsdichte aller jener m -Werte, die in der Reihe der Maxima unmittelbar auf einen Wert der Größe m_0 folgen. Um diese Größe numerisch gewinnen zu können, benötigt man eine sehr große Statistik. Abbildung 4.13 zeigt zu langzeitkorrelierten ($\gamma = 0.4$) Gauss- und exponentiell verteilten Daten die Verteilung $P_R(m|m_0)$ für zwei m_0 -Werte. Vergleicht man die Lage von $P_R(m|m_0)$ (rote und blaue Kurven) mit der Lage der unbedingten Verteilungsdichte $P_R(m)$ (gestrichelte Linie), so fällt auf, dass $P_R(m|m_0)$ für kleine m_0 nach links und für große m_0 nach rechts verschoben ist.

Die bedingte Überschreitungswahrscheinlichkeit

$$E_R(m|m_0) = \int_m^\infty P_R(m'|m_0) dm' \quad (4.25)$$

definiert die Wahrscheinlichkeit, in der Reihe der Maximawerte ein Ereignis größer m unmittelbar nach einem Ereignis der Größe m_0 zu finden. $E_R(m|m_0)$ weist eine starke Abhängigkeit von der Bedingung m_0 auf, wie die Beispielswerte in der Bildunterschrift zu Abbildung 4.13 zeigen.

Nun soll der Effekt der Langzeitkorrelationen auf die bedingte Überschreitungswahrscheinlichkeit $E_R(m|m_0)$ für verschiedene Bedingungen m_0 und verschiedene Schwellwerte m quantifiziert werden. Abbildung 4.14 zeigt $E_R(m|m_0)$ für sechs verschiedene m -Werte (Quantile)

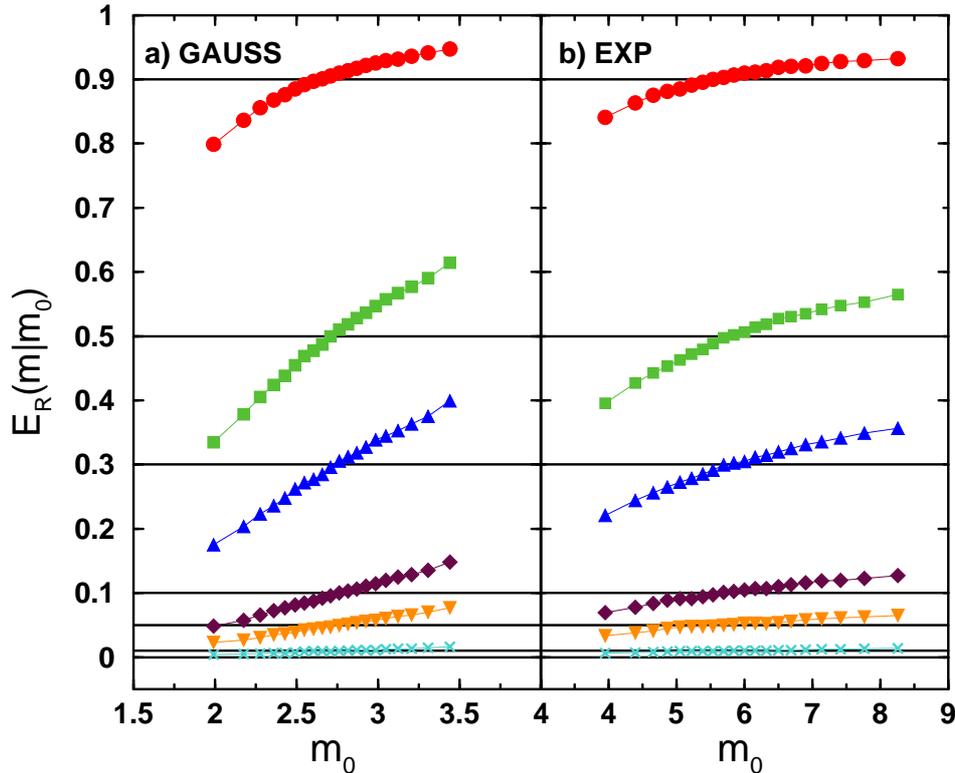


Abbildung 4.14: Bedingte Überschreitungswahrscheinlichkeiten $E_R(m|m_0)$ für Maxima größer als ein gegebener Schwellwert m als Funktion des Bedingungswertes m_0 . Bild (a) zeigt das Ergebnis für Gaussverteilte Daten mit $\gamma = 0.4$ und $R = 365$; Bild (b) zeigt dasselbe für exponentiell verteilte Daten. Die horizontalen Linien stehen für die unbedingten Überschreitungswahrscheinlichkeiten $E_R(m)$, welche unabhängig von m_0 sind. Die sechs m -Werte wurden wie folgt gewählt: Bild (a) $E_R(m = 2.15) = 0.9$ (\circ), $E_R(m = 2.75) = 0.5$ (\square), $E_R(m = 2.95) = 0.3$ (\triangle), $E_R(m = 3.35) = 0.1$ (\diamond), $E_R(m = 3.55) = 0.05$ (∇) und $E_R(m = 3.95) = 0.01$ (\times). Bild (b) $E_R(m = 4.40) = 0.9$ (\circ), $E_R(m = 5.95) = 0.5$ (\square), $E_R(m = 6.70) = 0.3$ (\triangle), $E_R(m = 8.00) = 0.1$ (\diamond), $E_R(m = 8.75) = 0.05$ (∇) und $E_R(m = 10.40) = 0.01$ (\times). Jeder Punkt in den Kurven beruht auf einer Statistik von 500 bedingten Maxima und ist gemittelt über 150 Runs der Länge $N = 2^{21}$. Abbildung nach [Eichner 2006a].

als Funktion von m_0 . Die m -Werte wurden so gewählt, dass die zugehörigen unbedingten Überschreitungswahrscheinlichkeiten $E_R(m)$ (durchgezogene horizontale Linien im Bild) die Werte 0.9, 0.5, 0.3, 0.1, 0.05 und 0.01 haben. Der m -Wert mit $E_R(m) = 0.01$ entspricht also einem 99%-Quantil der Gumbel-Fitkurve und spielt bei der Abschätzung von hundertjährigen Ereignissen eine besondere Rolle (dazu mehr im Kapitel 4.2.5). Auch hier zeigt die Krümmung der Kurven nochmal deutlich den erheblichen Einfluss der Korrelationen. Bei Gaussverteilten Daten mit $\gamma = 0.4$ und einem m -Wert, für den $E_R(m) = 0.5$ ist, variiert die Kurve $E_R(m|m_0)$ bis zu einem Faktor 2 in Abhängigkeit vom vorausgehenden m_0 (für $2 < m_0 < 3.5$). Bei exponentiell verteilten Daten ist dieser Faktor immerhin auch noch 1.5 (für $4 < m_0 < 8.5$). Die unterschiedliche m_0 -Skala bei Gauss- und exponentiellen Daten kommt durch die unterschiedlichen Wertebereiche der Ursprungsverteilungen $P(x)$ zustande.

Abbildung 4.14 lässt den Eindruck erwecken, dass der Memory-Effekt bei mittleren m -Werten

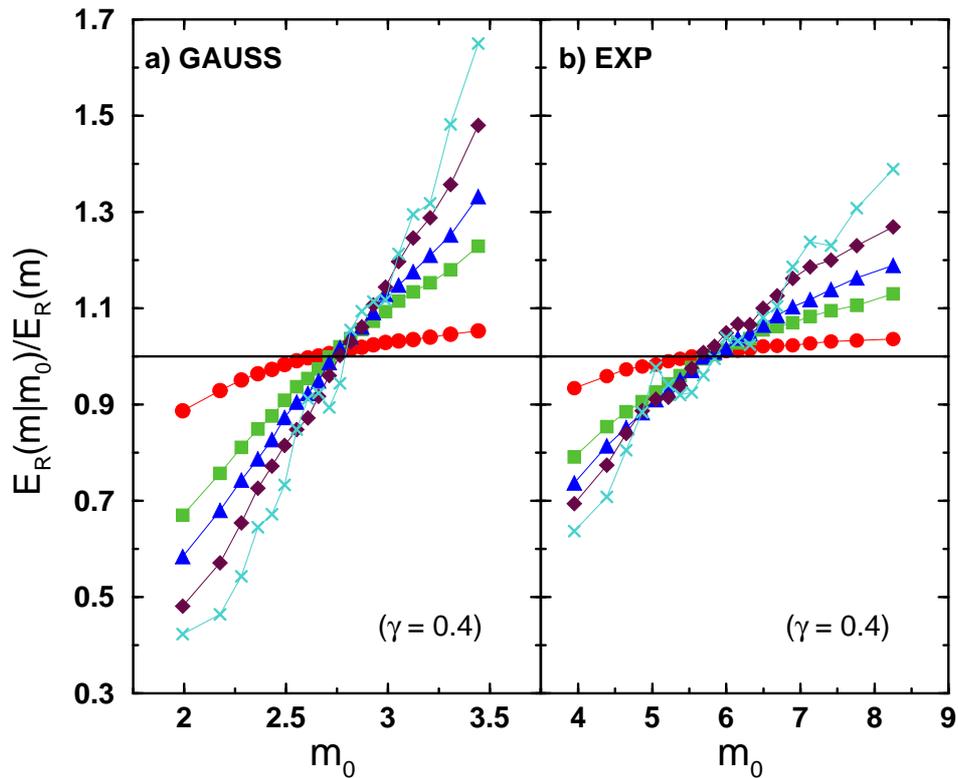


Abbildung 4.15: Quotienten aus der bedingten Überschreitungswahrscheinlichkeit $E_R(m|m_0)$ und der unbedingten Überschreitungswahrscheinlichkeit $E_R(m)$ als Funktion des Bedingungswertes m_0 für (a) Gauss-verteilte Daten und (b) exponentiell verteilte Daten mit $\gamma = 0.4$ und $R = 365$. Symbole und Statistik sind Abbildung 4.14 zu entnehmen. Die Kurven für $E_R(m) = 0.05$ sind hier nicht gezeigt, um ein Überlappen zu vermeiden. Der Effekt der Langzeitkorrelationen scheint für große m -Werte (\times) am stärksten zu sein: $E_R(m|m_0)/E_R(m)$ variiert zwischen 0.4 und 1.7, also etwa um einen Faktor 4 für Gauss-verteilte Daten. Bei exponentiell verteilten Daten ist es immerhin noch ein Faktor größer 2. Abbildung nach [Eichner 2006a].

($E_R(m) = 0.5$) am stärksten zu sein scheint. Für größere m -Werte, also für $E_R(m) \ll 0.5$, scheint er wieder schwächer zu werden. Die Differenz zwischen $E_R(m|m_0)$ und $E_R(m)$ wird umso kleiner, je größer m wird. Daraus könnte man schließen, dass der Korrelationseffekt für große m -Werte verschwindet. Dies ist jedoch nicht der Fall, wie Abbildung 4.15 belegt. In Abbildung 4.15 ist das Verhältnis der bedingten Überschreitungswahrscheinlichkeit $E_R(m|m_0)$ und der unbedingten Überschreitungswahrscheinlichkeit $E_R(m)$ als Funktion von m_0 aufgetragen. Das Bild zeigt, wie der Memory-Effekt sogar zunimmt, je größer m ist, d. h. je extremer ein Ereignis ist. Der Zuwachs scheint bei exponentiell verteilten Daten etwas schwächer zu sein als bei Gauss-verteilten Daten. Die Tendenz ist aber in beiden Bildern (a) und (b) dieselbe. Der Unterschied zwischen der bedingten und der unbedingten Überschreitungswahrscheinlichkeit in Abbildung 4.15 kann bei langzeitkorrelierten Daten ($\gamma = 0.4$) bis zu einem Faktor 2 betragen, wenn man die Vergangenheit m_0 berücksichtigt. Ein so starker Effekt sollte bei Vorhersagen und Risikoabschätzungen in korrelierten Systemen in jedem Falle berücksichtigt werden.

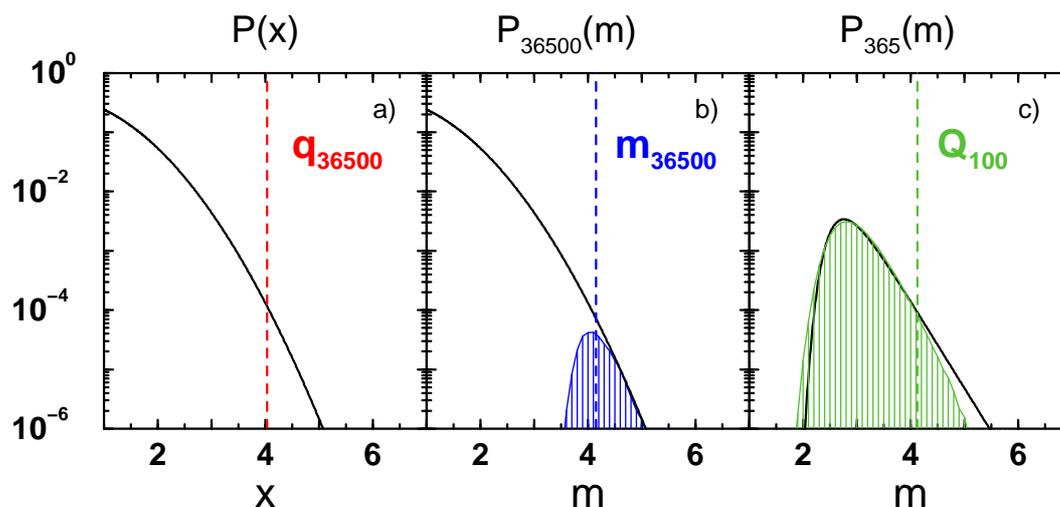


Abbildung 4.16: Drei Definitionen 100-jähriger Ereignisse (vertikale gestrichelte Linien): (a) das $1/36500$ Quantil q_{36500} bei bekannter Verteilungsdichte $P(x)$, (b) das mittlere Jahrhundertmaximum m_{36500} aus dem Histogramm $P_{36500}(m)$ der Jahrhundertmaxima und (c) das $1/100$ Quantil Q_{100} aus der Gumbel-Fitfunktion Gleichung (4.15) basierend auf Jahresmaximawerten des Histogramms $P_{365}(m)$. Die Abweichungen des Gumbel-Fits vom Histogramm $P_{365}(m)$ beruhen auf den Gaussverteilten Daten, für die $R = 365$ noch nicht groß genug für das Gumbel-Limit ist. Es findet aber so in der praktischen Hydrologie Anwendung.

4.2.5 Einfluss der Langzeitkorrelationen auf Jahrhundertereignisse

In diesem Abschnitt geht es um die Auswirkung der Langzeitkorrelationen auf die Abschätzung der Größe eines typischen Jahrhundertereignisses, d. h. ein typisches Maximum wie es im Mittel nur einmal in hundert Jahren vorkommt. Aussagen über sogenannte Jahrhundertereignisse sind insbesondere in der hydrologischen Risikoabschätzung sehr großer Überschwemmungen von Bedeutung [Chow 1964, Jenkinson 1969, Raudkivi 1979, Landwehr 1980, Phien 1987, Chow 1988, Hosking 1997, DVWK 1999, BWG 2003]. Leider reichen nur wenige zuverlässig gemessene Datensätze selten über 100 Jahre hinaus, sodass nicht gewährleistet ist, dass überhaupt ein einziges typisches Jahrhundertereignis in einem Datensatz vorkommt. Daher ist es in der Praxis ein äußerst heikles Unterfangen aufgrund der meist sehr schwachen Datenlage verlässliche Aussagen zu treffen. Genauso gut könnte aber auch in einem Datensatz ein 1000-jähriges Ereignis aufgetreten sein. Es muss also erst einmal geklärt werden, was man in der Praxis unter einem "typischen Jahrhundertereignis" versteht.

Wären im Idealfall die Verteilungsdichten $P(x)$ und $P_R(m)$ der untersuchten Zeitreihe vollständig bekannt, d. h. wäre die Zeitreihe unendlich lang, dann gibt es zwei alternative Definitionen, Abbildung 4.16(a) und (b), für typische Jahrhundertereignisse.

Die erste Definition, Abbildung 4.16(a), beruht auf der Verteilungsdichte $P(x)$ (in diesem Beispiel eine Gauss-Verteilung von Tagesdaten) und gibt als Größe für typische Jahrhundertereignisse das Quantil q_{36500} an (rote gestrichelte Linie). Das ist die Schranke, über der nur $1/36500$ aller Tagesdaten x liegen. Sie wird im Mittel also nur alle 100 Jahre (36500 Tage) einmal überschritten. Für diese Definition muss $P(x)$ sehr exakt bekannt sein. Das heißt, eine

Messreihe muss sehr lang sein, so dass bereits einige Jahrhundertereignisse in ihr aufgetreten sind, damit das rechte Ende von $P(x)$ ausreichend bestimmbar ist. Da aber q_{36500} nur auf der Verteilung $P(x)$ beruht, ist diese Größe unbeeinflusst von möglichen Korrelationen und einer möglichen Clusterbildung der m_j . Die Definition berücksichtigt alle Werte, die das Quantil überschreiten, unabhängig davon, ob sie sich innerhalb desselben Jahrhunderts ereignen oder nicht.

Die zweite Definition, Abbildung 4.16(b), basiert auf der Verteilungsdichte $P_{36500}(m)$ (blaues Histogramm) der Jahrhundertereignisse m innerhalb verschiedener Blöcke der Länge $R = 100$ Jahre = 36500 Tage. Ein typisches Jahrhundertereignis wird hier definiert als das mittlere Jahrhundertmaximum m_{36500} (blaue gestrichelte Linie), also das erste Moment der Verteilungsdichte $P_{36500}(m)$. Zwar werden in dieser Definition die Korrelationen berücksichtigt und Cluster von Jahrhundertereignissen, das heißt mehrere Jahrhundertereignisse innerhalb eines Blocks der Länge $R = 100$ Jahre, kann es hier nicht geben⁵. Jedoch verlangt auch diese Definition sehr lange Reihen, da man mehrere Jahrhundertereignisse benötigt, um eine vernünftige Verteilung $P_{36500}(m)$ ermitteln zu können.

In echten Datensätzen beschränkt sich der statistische Umfang meist auf 30 Jahre, 50 Jahre oder vielleicht 100 Jahre Tagesdaten. Das ist nicht genug für eine ausreichende Bestimmung der Verteilungsdichten $P(x)$ bzw. $P_{36500}(m)$, um daraus vernünftige Angaben für die Größen q_{36500} bzw. m_{36500} machen zu können. Es ist a priori auch nicht gegeben, dass Messreihen dieser Längen überhaupt ein Jahrhundertereignis enthalten. Das erschwert es sehr, die typische Größe eines Jahrhundertereignisses abzuschätzen. Gehört die Verteilungsdichte $P(x)$ der Messungen mit großer Wahrscheinlichkeit zu jenen, welche im Einzugsbereich der Gumbel-Verteilung liegen, so ist es sinnvoll die Gumbel-Fitformeln Gl. (4.16) und (4.17) auf $P_R(m)$ mit kleineren R -Werten (z. B. Jahresmaxima mit $R = 365$) anzuwenden, um dann eine Näherung für ein typisches Jahrhundertereignis aus dem Gumbel-Fit zu erhalten. Diese Prozedur ist in Abbildung 4.16(c) illustriert, in der das Histogramm $P_{365}(m)$ und der zugehörige Gumbel-Fit (durchgezogene Linie) zu sehen sind. In der Hydrologie schätzt man nun die Größe eines typischen Jahrhundertereignisses mit dem Quantil Q_{100} (oft auch mit HQ_{100} bezeichnet) ab [Jenkinson 1969, Raudkivi 1979, Landwehr 1980, Phien 1987, Chow 1988, Hosking 1997]. Q_{100} ist der Schwellwert in der Gumbel-Fitkurve, welcher nur von 1% aller Jahresmaxima m überschritten wird und in der Abbildung 4.16(c) als grüne gestrichelte Linie dargestellt ist. Aus der integrierten Gumbel-Verteilung $G_{365}(m)$ in Gl. (4.16) ist der Q_{100} -Wert durch $G_{365}(Q_{100}) = 0.99$ herzuleiten. Man erhält hieraus

$$Q_{100} = u - \alpha \ln(-\ln(0.99)), \quad (4.26)$$

und mit den numerischen Werten für Skalenparameter α und Lokalisationsparameter u

$$Q_{100} = m_{365} - [\ln(-\ln(0.99)) + \gamma_e] \frac{\sqrt{6}}{\pi} \sigma_{365} \simeq m_{365} + 3.137 \sigma_{365}, \quad (4.27)$$

wobei m_{365} das mittlere Maximum aller Jahresmaxima ist und σ_{365} die zugehörige Standardabweichung, welche man auch aus kürzeren Zeitreihen (z. B. nur 30 Jahre) recht einfach und zuverlässig bestimmen kann.

Die drei eben vorgestellten Definitionen für hundertjährige Ereignisse, q_{36500} , m_{36500} und Q_{100} , sind zwar ähnlich, zeigen aber unterschiedliche Reaktionen auf Langzeitkorrelationen. Abbildung 4.17 vergleicht für unkorrelierte (a) und langzeitkorrelierte ($\gamma = 0.4$) Gauss-verteilte

⁵Zwei Jahrhundertereignisse liegen bei dieser Definition mindestens 100 Jahre auseinander.

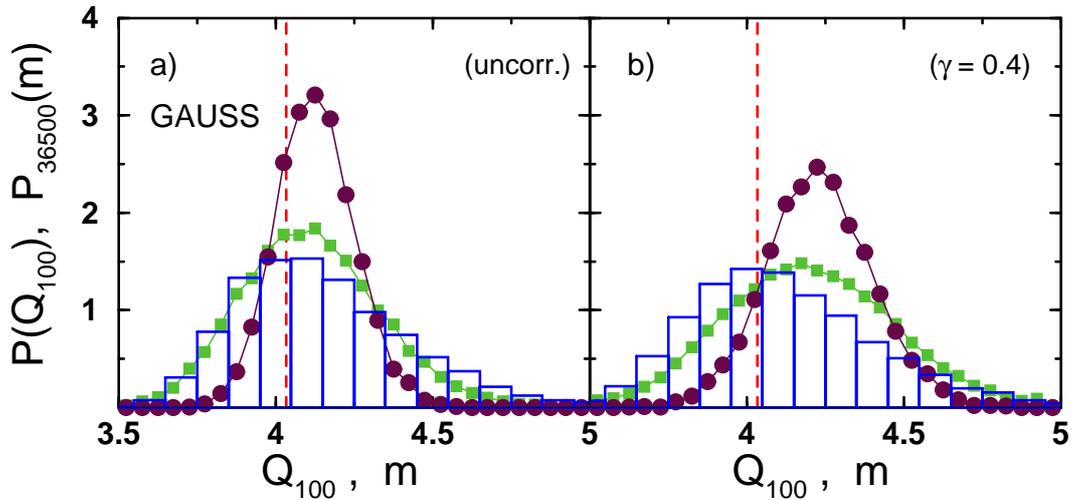


Abbildung 4.17: Vergleich numerischer Ergebnisse für die verschiedenen Jahrhundertereignisse aus Abbildung 4.16 bei unkorrelierten (a) und langzeitkorrelierten ($\gamma = 0.4$) (b) Gauss-verteiltern Daten. Die vertikale gestrichelte Linie repräsentiert das Quantil $q_{36500} = 4.034$, das Histogramm ist die Verteilungsdichte der Jahrhundertmaxima $P_{36500}(m)$. Die Kreise stehen für die Verteilung der Q_{100} -Werte, welche aus unterschiedlichen 100-jährigen Ausschnitten (36500 Tagesdaten) aus 150 Datensätzen der Länge $N = 2^{21}$ bestimmt wurden. Die Quadrate basieren auf 30-jährigen Ausschnitten (10950 Tagesdaten). Abbildung nach [Eichner 2006a].

Daten (b) das Quantil q_{36500} , die Verteilung der Jahrhundertmaxima $P_{36500}(m)$ und zwei Verteilungen von Q_{100} -Werten (für 30-jährige Messreihen und 100-jährige Messreihen). Die Verteilung der Q_{100} -Werte, die aus den Jahresmaxima von 100-jährigen Reihen bestimmt wurden (bordeaux Kreise), haben die gleiche statistische Basis wie die Jahrhundertmaxima aus $P_{36500}(m)$. Da echte Datenreihen aber oft deutlich kürzer als 100 Jahre sind, wird in Abbildung 4.17 auch die Verteilung der Q_{100} -Werte gezeigt, die auf Gumbel-Fits aus 30-jährigen Zeitreihen basieren (grüne Quadrate). Während der Einfluss der Langzeitkorrelationen auf die Q_{100} -Werte aus den Gumbel-Fitkurven der Jahresmaxima recht deutlich ist, zeigt die Verteilungsdichte der Jahrhundertmaxima $P_{36500}(m)$ und das zugehörige m_{36500} nur eine schwache Korrelationsabhängigkeit. Das Quantil q_{36500} als statische Größe bleibt von den Korrelationen vollkommen unbeeinflusst. Man beachte, dass die Verschiebung der Q_{100} -Verteilungen entgegengesetzt zu der leichten Verschiebung von $P_{36500}(m)$ ist. Die leichte Verschiebung von $P_{36500}(m)$ ist auf den Effekt der korrelationsbedingten Linksverbreiterung (siehe Abbildung 4.6) zurückzuführen. Die Rechtsverschiebung der Q_{100} -Histogramme wird vermutlich durch den Einfluss des Skalenparameters α bzw. der Standardabweichung σ_{365} in den Formeln Gln. (4.26) bzw. (4.27) verursacht. Zusätzlich beobachtet man für die Q_{100} auch noch eine Verbreiterung der Verteilungen bei Langzeitkorrelationen, was zu einer ungenaueren Bestimmung von typischen hundertjährigen Ereignissen führt. Aber noch immer ist die Q_{100} -Verteilung deutlich schärfer als $P_{36500}(m)$. Das bedeutet, dass typische Jahrhundertereignisse bei 100-jähriger statistischer Grundlage (bordeaux Kreise) durch Q_{100} zuverlässiger angenähert werden als durch das einzelne Jahrhundertmaximum, welches eine 100-jährige Reihe hergibt. Die Abschätzungen mittels Q_{100} bei 30-jähriger statistischer Grundlage sind

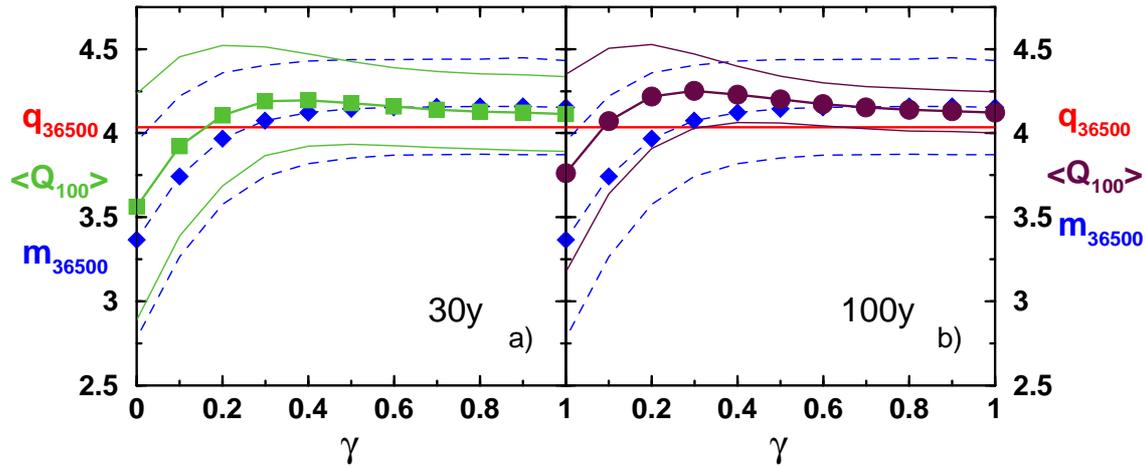


Abbildung 4.18: Hundertjährige Ereignisse von Gauss-verteilten Daten als Funktion der Korrelationsexponenten γ . Der Wert $q_{36500} = 4.034$ (horizontale rote Linie) ist per Definition korrelationsunabhängig, da er nur auf der Verteilung $P(x)$ basiert. Das mittlere Jahrhundertmaximum m_{36500} (\diamond) mit zugehöriger oberer Standardabweichung (gestrichelte blaue Linien) und der mittleren Q_{100} -Wert (\square in (a) und \circ in (b)) mit zugehöriger unterer Standardabweichung (durchgezogene grüne Linien) als Risikoabschätzung zeigen deutliche Korrelationsabhängigkeiten im Bereich $\gamma < 0.2$ (man beachte, dass kleine γ -Werte starke Langzeitkorrelationen bedeuten). In (a) basieren die Werte für Q_{100} auf Ausschnitten der Länge 30 Jahre (10950 Tage), in (b) auf 100 Jahre (36500 Tage). Das Ansteigen der Q_{100} -Kurve gegenüber der m_{36500} -Kurve in (b) ist auf die stärkeren Korrelationen in den Jahresmaxima gegenüber den Jahrhundertmaxima zurückzuführen, welche bei kürzeren Stichproben (wie in (a)) nicht so stark durchschlagen. Abbildung nach [Eichner 2006a].

in ihrer Streuung qualitativ ähnlich wie die Jahrhundertmaxima aus 100-jährigen Reihen.

Um die Korrelationsabhängigkeit der Größenschätzung von Jahrhundertereignissen und ihrer Genauigkeit quantitativ zu untersuchen, wurden Mittelwerte und Standardabweichungen von Histogramme wie in Abbildung 4.17 für künstliche langzeitkorrelierte Daten mit unterschiedlichen Korrelationsexponenten γ untersucht. Abbildung 4.18 zeigt das konstante Quantil q_{36500} , das mittlere Jahrhundertmaximum m_{36500} mit der zugehörigen Standardabweichung s_{36500} und das mittlere geschätzte Jahrhundertereignis $\langle Q_{100} \rangle$ mit der zugehörigen Standardabweichung als Funktionen des Korrelationsexponenten γ . Auch hier wurden die Q_{100} -Werte einmal auf 30-jähriger Basis bestimmt (a) und einmal auf 100-jähriger Basis (b). Das Quantil q_{36500} ist unabhängig von γ , da es nur auf $P(x)$ basiert, und außerdem für die Praxis nicht relevant, da $P(x)$ in der Realität nicht ausreichend gut bekannt ist. Im Gegensatz zu q_{36500} sind m_{36500} und noch viel mehr $\langle Q_{100} \rangle$ durch starke Langzeitkorrelationen ($\gamma < 0.2$) erheblich beeinflusst. Beide Werte werden kleiner mit zunehmenden Korrelationen, d. h. mit kleiner werdendem γ . Aufgrund der starken Langzeitkorrelationen neigen auch die großen Maxima zum Clustern, d. h. es bilden sich Epochen, in denen wesentlich mehr große Maximawerte auftreten (“Berge”) als in schwach korrelierten oder unkorrelierten Daten. Konsequenter Weise fehlen diese Maxima dann an anderen Stellen in der Datenreihe, d. h. es gibt auch Epochen, in denen die Maxima beträchtlich kleiner ausfallen (“Täler”) als in schwach oder unkorrelierten Datenreihen. Mit zunehmenden Korrelationen ($\gamma < 0.2$) bilden sich diese Epochen in Form von Bergen und Tälern in der Zeitreihe immer deutlicher aus und treten auch

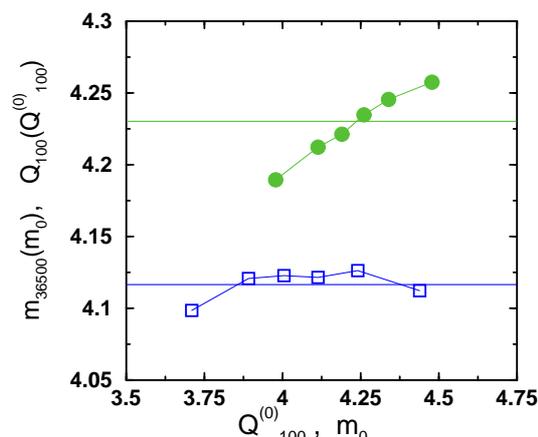


Abbildung 4.19: Bedingte hundertjährige Ereignisse m_{36500} (\square) und Q_{100} (\circ) langzeitkorrelierter ($\gamma = 0.4$) Gauss-verteilter Daten als Funktion von m_0 bzw. $Q_{100}^{(0)}$. Die bedingten $Q_{100}(Q_{100}^{(0)})$, basierend auf Jahresmaxima, zeigen durch den Aufwärtstrend noch einen Korrelationseffekt, während die bedingten Jahrhundertmaxima $m_{36500}(m_0)$ für $\gamma = 0.4$ aufgrund des sehr großen R -Wertes bei beschränkter Reihenlänge (hier $N = 2^{21}$, 150 Runs) keine nennenswerten Korrelationen mehr aufweisen. Die horizontalen Linien sind die zugehörigen unbedingten Werte $Q_{100} = 4.23$ und $m_{36500} = 4.12$. Die Q_{100} -Werte wurden aus 100-jährigen Ausschnitten bestimmt. Abbildung nach [Eichner 2006a].

häufiger auf, wodurch das mittlere Jahrhundertmaximum m_{36500} und auch das geschätzte Jahrhundertereignis Q_{100} , welches auf Jahresmaxima basiert, unter den Wert des Quantils q_{36500} fallen. Die zugehörigen Standardabweichungen (gestrichelte Linien und durchgezogene Linien in Abbildung 4.18), welche die Breiten der Histogramme $P_{36500}(m)$ und $P(Q_{100})$ beschreiben und somit die Genauigkeit der Abschätzungen für kurze Datensätze angeben, wachsen an mit kleiner werdendem γ (d. h. stärker werdenden Langzeitkorrelationen), da die Links-Verbreiterung (siehe Abbildung 4.6) immer beträchtlicher wird. Für $\gamma > 0.5$ sind das mittlere Jahrhundertmaximum m_{36500} und das geschätzte mittlere Jahrhundertereignis $\langle Q_{100} \rangle$ nahezu konstant, also unbeeinträchtigt durch Korrelationen.

Das geschätzte mittlere Jahrhundertereignis $\langle Q_{100} \rangle$ überschätzt für $\gamma < 0.6$ das mittlere Jahrhundertmaximum m_{36500} . Diese systematische Abweichung wird kleiner, wenn $\langle Q_{100} \rangle$ auf 30-jährigen Datensätzen (Abbildung 4.18(a), Quadrate) basiert anstatt auf 100-jährigen (Abbildung 4.18(b), Kreise). Beide Größen unterschätzen q_{36500} für sehr starke Korrelationen ($\gamma < 0.1$), was schon sehr nahe am nichtstationären Verhalten liegt. Die jeweiligen Standardabweichungen von Q_{100} (durchgezogene Linien) fallen für schwach bis mittelstark korrelierte Daten ($\gamma > 0.4$) deutlich kleiner aus als die Standardabweichung σ_{36500} (gestrichelte Linien) der Jahrhundertmaxima m_{36500} . Daraus kann man folgern, dass die Abschätzung eines Jahrhundertereignisses mittels Q_{100} verlässlicher ist als die direkte Bestimmung von m_{36500} . Das heißt, dass die Größe Q_{100} bei kurzen Messreihen eine sehr zuverlässige Vorhersage für typische Jahrhundertereignisse ist. Im Falle sehr starker Langzeitkorrelationen neigt sie jedoch dazu, systematisch größer auszufallen als m_{36500} . In echten hydrologischen Datensätzen, wie z. B. Abflussreihen oder Pegelstände, fallen die Korrelationen selten stärker als $\gamma = 0.3$ aus [Kantelhardt 2003b, Kantelhardt 2006], weshalb die Größe Q_{100} als ein sehr stabiler und sicherer Schätzer für Jahrhundertfluten angenommen werden kann.

Als letztes ist der Effekt der Langzeitkorrelationen auf $\langle Q_{100} \rangle$ analysiert worden. Während es in m_{36500} aufgrund des sehr großen R -Wertes so gut wie keinen Memory-Effekt geben darf (siehe Abbildungen 4.8 und 4.11), sollte Q_{100} immernoch eine Abhängigkeit von der Vergangenheit zeigen, da es auf m_{365} und σ_{365} aus Jahresdaten basiert. Abbildung 4.19 zeigt für langzeitkorrelierte ($\gamma = 0.4$) Gauss-verteilte Daten die bedingte Größe $Q_{100}(Q_{100}^{(0)})$ (Kreise), welche das mittlere Q_{100} all jener Q_{100} -Werte ist, die in der Reihe der Q_{100} unmittelbar auf einen Werte der Größe $Q_{100}^{(0)}$ folgen (völlig analog zur Größe $m_R(m_0)$ aus Abbildung 4.11). Ebenso zeigt Abbildung 4.19 die bedingte Größe $m_{36500}(m_0)$ als Funktion der Bedingung m_0 . Die durchgezogenen horizontalen Linien sind die jeweiligen unbedingten Mittelwerte. Während das bedingte $Q_{100}(Q_{100}^{(0)})$ eine deutliche Abhängigkeit von der Vergangenheit $Q_{100}^{(0)}$ zeigt, ist $m_{36500}(m_0)$ im Prinzip unabhängig von m_0 , da es nahezu horizontal verläuft. Obwohl der Korrelationseffekt auf das bedingte $Q_{100}(Q_{100}^{(0)})$ recht klein ist, bleibt er messbar und kann so ein wenig hilfreich sein, Vorhersagen von Jahrhundertereignissen zu verbessern.

Kapitel 5

Zusammenfassung und Ausblick

Die Untersuchungen zum Einfluss von Langzeitkorrelationen auf die Statistik der Wiederkehrintervalle ergaben, dass die Verteilungen mittlerer bis großer Wiederkehrintervalle r einer gestreckt exponentiellen Funktion folgen. Es konnte im Rahmen dieser Arbeit gezeigt werden, dass dies nicht nur für Gauss-verteilte Daten, sondern auch für exponentiell, Potenzgesetzartig und Log-Normal-verteilte Daten gilt. Außerdem lassen sich die Verteilungsdichtefunktionen $P_q(r)$ zu verschiedenen Quantilwerten q bzw. Wiederkehrperioden R_q mittels der skalierten Auftragung $R_q P_q(r)$ gegen r/R_q zum Kollaps bringen. Das heisst, die Form der Verteilungsdichtefunktionen ist unabhängig vom Quantilwert und nur abhängig vom Korrelationsexponenten γ . Dies hat zur Folge, dass man von Ergebnissen zu kleinen q -Werten auf Ergebnisse zu großen q -Werten, bei denen die Statistik schlecht wird, extrapolieren kann. Weiter wurde gezeigt, dass kleine Wiederkehrzeiten eher einer Potenzgesetzverteilung folgen, die durch einen Exponenten bestimmt wird, der mit dem Korrelationsexponenten γ im Zusammenhang steht.

Es konnte gezeigt werden, dass die Reihen der Wiederkehrintervalle langzeitkorrelierter Daten (auch für große Quantile) ebenfalls langzeitkorreliert sind. Dies hat zur Folge, dass auf große Intervalle eher wieder große folgen und auf kleine Intervalle eher wieder kleine. Aufgrund der Langzeitkorrelationen prägen sich in bestimmten Bereichen regelrechte *Cluster* extremer Ereignisse aus, während sich in anderen Bereichen wiederum lang anhaltende Ruhephasen ausbilden, in denen kaum extreme Ereignisse auftreten.

Dieser *Memory*-Effekt in den Wiederkehrintervallen kann mittels der Betrachtung bedingter Wiederkehrintervalle und bedingter Wahrscheinlichkeiten, d. h. unter Berücksichtigung des unmittelbar vorangegangenen Intervalls r_0 , zur Verbesserung von Erwartungswerten und Vorhersagen ausgenutzt werden. Die bedingte Wiederkehrperiode kann sich erheblich von der unbedingten unterscheiden. Noch drastischer fällt die Situation für die bedingten Latenzzeiten (oder Restwartezeiten) $\tau_q(t|r_0)$ aus, welche die Information berücksichtigen, dass bereits eine Dauer t seit dem letzten Extremereignis verstrichen ist. Die Größe $\tau_q(t|r_0)$ zeigt unter dem Einfluss von Langzeitkorrelationen ein völlig kontraintuitives Verhalten: je länger die Zeit t anhält, umso größer fällt auch $\tau_q(t|r_0)$ aus. Dies steht völlig im Kontrast zum Verhalten bei unkorrelierten Daten, in denen $\tau_q(t|r_0)$ unabhängig von der Vergangenheit r_0 und t ist und trotz anwachsendem t immer $R_q \simeq \tau_q(t|r_0)$ gilt (bis auf Finite-Size Effekte). In langzeitkorrelierten Systemen ist die Aussage, dass ein extremes Ereignis nach langer Ruhephase “überfällig” wird, also falsch.

Die Untersuchungen im Kapitel über den Einfluss der Langzeitkorrelationen auf die Statistik der Maximawerte ergaben, dass die Verteilungsfunktion der Maxima aus Blöcken der Länge R in Gauss- und exponentiell verteilten langzeitkorrelierten Daten mit anwachsendem R langsamer gegen die Gumbel-Verteilung konvergiert, als in unkorrelierten Systemen. Im Gegensatz zur bisherigen Annahme, dass die Reihe der Maxima statistisch unabhängig ist, konnte gezeigt werden, dass für den praxisnahen Fall $R = 365$ ("Jahresmaxima") die Korrelationen nicht verschwinden und die Reihe der Maxima ebenfalls Langzeitkorrelationen mit einem sehr ähnlichen Korrelationsexponenten aufweist. Als Folge bilden sich in den Maxima-Reihen Epochen aus, in denen sich große Extremwerte clustern, aber auch Epochen, in denen nur schwache Extremereignisse auftreten. Dieser Effekt konnte mittels bedingter Verteilungsfunktionen und bedingter Mittelwerte klar gezeigt werden. Sowohl die Verteilung der Maxima als auch das mittlere Maximum werden stark von der Vergangenheit, also etwa vom vorhergehenden Maximum, beeinträchtigt. Auch wenn der Korrelationseffekt absolut gesehen für sehr große Maximawerte aufgrund vorausgesetzter Stationarität immer kleiner wird, so können die bedingten Überschreitungswahrscheinlichkeiten gegenüber den unbedingten (dass heißt, ohne Berücksichtigung der Korrelationen) um ein Vielfaches größer oder kleiner ausfallen und so Prognosen und Risikoabschätzungen verändern.

Außerdem wurde in dieser Arbeit gezeigt, dass Langzeitkorrelationen auch einen schwachen aber messbaren Effekt auf die hydrologische Größe Q_{100} haben, welche zur Abschätzung von Katastrophen wie Jahrhundertfluten verwendet wird. Aber es zeigt sich, dass die Größe Q_{100} auch bei Langzeitkorrelationen eine überraschend verlässliche Näherung für Jahrhundertereignisse ist, insbesondere auch bei kurzen Datensätzen. Nur bei äußerst starken Langzeitkorrelationen ($\gamma < 0.2$) neigt Q_{100} dazu, das tatsächliche Quantil für hundertjährige Ereignisse q_{36500} zu unterschätzen. Korrelationen dieser Stärke sind aber in natürlichen Prozessen, insbesondere in hydrometeorologischen Zeitreihen, eher selten.

Die meisten dieser Ergebnisse, insbesondere das Clustern extremer Ereignisse, konnten in ähnlicher Form bei echten und rekonstruierten Datensätzen trotz geringer Statistik nachgewiesen werden. Aus diesen Ergebnissen lässt sich ableiten, dass zur verbesserten Risikoabschätzung in langzeitkorrelierten Prozessen bedingte Verteilungen, bedingte Mittelwerte und auch bedingte Q_{100} -Werte geeignete Größen sind, als die unbedingten Größen. Der Einfluss vorhandener Langzeitkorrelationen ist besonders auf die Wiederkehrintervalle sehr groß und muss für vernünftige Vorhersagen und Risikoabschätzungen berücksichtigt werden. Mit dieser Arbeit wurde gezeigt, dass allein mittels natürlicher Langzeitkorrelationen bereits ein zeitweilig gehäuftes Auftreten von größeren Klimaphänomenen erklärt werden kann. Der menschliche Beitrag zum Treibhauseffekt ist aber in der heutigen Zeit nicht mehr zu leugnen. Als Konsequenz ist den natürlichen Schwankungen auch ein Trend überlagert, der mit großer Sicherheit das gehäufte Auftreten von Extremereignissen weiter antreibt.

Andere Formen statistischer Abhängigkeiten, die im Rahmen dieser Arbeit nicht betrachtet wurden, wie zum Beispiel Trends oder Multifraktalität (beides findet man u. a. in hydrometeorologischen Messreihen [Kantelhardt 2003b, Kantelhardt 2006, Koscielny 2006]), könnten ebenfalls sehr große Einflüsse auf die Extremwertstatistik haben. Es ist anzunehmen, dass Trends in der Statistik der Wiederkehrintervalle teilweise zu sehr ähnlichen Abweichungen von der Poisson-Statistik führen können wie Langzeitkorrelationen. Multifraktalität nimmt erheblichen Einfluss auf die Quantilauswahl, da große Fluktuationen anderen Skalengesetzen gehorchen als kleine. Da man einem multifraktalen Prozess keinen eindeutigen Skalenexpo-

nennten zuordnen kann, ist ein Kollaps der Verteilungsfunktionen der Wiederkehrintervalle zu verschiedenen Quantilen auszuschließen. In [Altmann 2005] haben die Autoren erste Multifraktalanalysen an den Reihen der Wiederkehrintervalle in langzeitkorrelierten Prozessen durchgeführt. Ähnliche Effekte sind auch in der Statistik der Maxima zu erwarten. Große Maxima können anderen Gesetzmäßigkeiten folgen als mittlere oder kleine Maxima. Erste Ansätze hierzu findet man unter anderem in [L'vov 2001].

Anhang A

Tabelle

Tabelle A.1: Zusammenhang zwischen γ , α und β und die Werte der Normierungsparameter a_γ und b_γ der gestreckt exponentiellen Verteilungsfunktion $P_q(r)$ der langen Wiederkehrintervalle aus Gl. (3.26). Die γ -Abhängigkeit der Normierungsparameter ergibt sich aus den beiden Gleichungen $a_\gamma = \frac{\gamma\Gamma(2/\gamma)}{\Gamma^2(1/\gamma)}$ und $b_\gamma = \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)}$.

γ	$\alpha = 1 - \gamma/2$	$\beta = 1 - \gamma$	a_γ	b_γ
1.00	0.50	0.00	1.00000	1.00000
0.90	0.55	0.10	1.11942	1.15871
0.80	0.60	0.20	1.29445	1.35848
0.70	0.65	0.30	1.56934	1.61683
0.60	0.70	0.40	2.04540	1.96298
0.50	0.75	0.50	3.00000	2.44949
0.40	0.80	0.60	5.43249	3.18149
0.30	0.85	0.70	15.1215	4.40422
0.20	0.90	0.80	126.000	6.85347
0.10	0.95	0.90	92378.0	14.2080

Anhang B

Beweis zur Verteilung der Wiederkehrzeiten

B.1 Zero-Level Crossing nach Newell und Rosenblatt

Bereits 1962 haben G. F. Newell und M. Rosenblatt in [Newell 1962] für das verwandte Problem des *Zero-Level Crossing* zeigen können, dass sich die Verteilungsfunktion der Wiederkehrzeiten bei stationären Gauss-Prozessen mit Potenzgesetz-artig abfallender Autokovarianzfunktion bzw. Autokorrelationsfunktion asymptotisch durch eine gestreckt exponentielle Funktion als obere Schranke abschätzen lässt. Abbildung B.1 erklärt das Prinzip des Zero-Level Crossings. Die von uns gefundene Verteilungsfunktion für das allgemeinere Problem des *q-Level Crossing* ist in der Auftragung r/R_q dieser sehr ähnlich. Es sei an dieser Stelle bereits angemerkt, dass das asymptotische Verhalten für große r/R_q -Werte für die Verteilungsdichte $P_q(r)$ und die dazugehörige Verteilungsfunktion $W_q(r)$

$$W_q(r) = \int_{-\infty}^r P_q(r') dr',$$

dasselbe ist, da P als Ableitung von W für große r/R_q weiterhin durch den gestreckt exponentiellen Verlauf dominiert wird. Der durch die Ableitung hinzugekommene Potenzgesetz-förmige Vorfaktor wird mit anwachsendem r/R_q immer unbedeutender. Der Beweis von Newell und Rosenblatt bezieht sich auf Verweildauern über der Null-Linie. Da sich der Beweis auf Gauss-verteilte Zahlen bezieht, ist das Problem symmetrisch und gilt demnach auch für die von uns definierten Wiederkehrzeiten, d. h. für Verweildauern unterhalb der Null-Linie. Ich stelle den Beweis aus [Newell 1962] hier ausführlich, aber zum besseren Verständnis in leicht abgewandelter Form, dar.

Gegeben sei ein stationärer *Zeit-kontinuierlicher* Gauss-Prozeß $x(t)$ mit Erwartungswert $\bar{x} = 0$ und Varianz $\sigma_x^2 = 1$. Mit *Zeit-kontinuierlich* ist gemeint, dass $x(t)$ zu jedem beliebigen Zeitpunkt $t \in \mathbb{R}^+$ gemessen werden kann, während $x(t_i) \equiv (x_i)$ einen *Zeit-diskreten* Prozeß beschreibt, der nur an bestimmten Zeitpunkten $t_i \in \mathbb{R}^+$ und $i \in \mathbb{N}$ messbar ist. In der Praxis hat man es immer mit (x_i) zu tun, da $x(t)$ nicht beliebig genau bekannt ist.

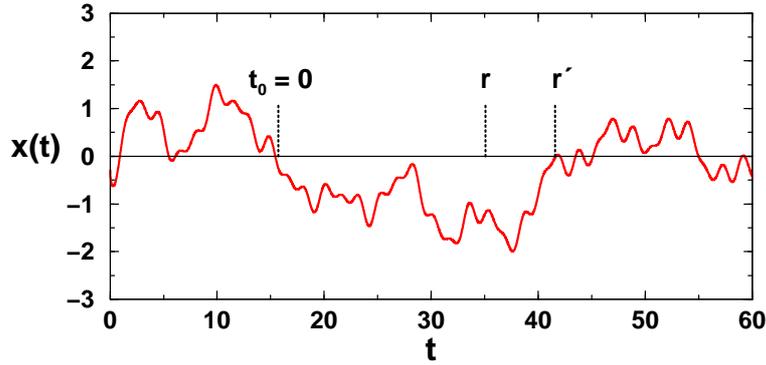


Abbildung B.1: Zero-Level Crossing bei einem kontinuierlichen Gauss-Prozess. Zu einem Zeitpunkt $t = t_0$ durchstößt $x(t)$ die Null-Linie. Von diesem Zeitpunkt an wird die Zeit neu gemessen, das heißt $t_0 := 0$. Die Dauer r' , die der Prozess unterhalb der Nulllinie verweilt, bevor er sie wieder in entgegengesetzter Richtung durchstößt oder auch nur berührt, ist mindestens r .

Für die Kovarianzfunktion¹ $\rho(s) = \langle x(t)x(t+s) \rangle$ des Gauss-Prozesses soll gelten

$$|\rho(s)| < Cs^{-\gamma} \quad \text{mit} \quad 0 < \gamma < 1$$

und einer Konstanten $C > 0$, also $\rho(s) \rightarrow 0$ für $s \rightarrow \infty$, wobei hier $s \in \mathbb{R}^+$ sein darf. Übertragen auf die in dieser Arbeit verwendete Notation der Autokorrelationsfunktion $C_x(s)$ bedeutet dies $C_x(0) = \sigma_x^2 = 1$ und $C_x(1) = C$.

Gesucht ist die Wahrscheinlichkeit $W(r) = P(r' > r)$, wobei r' die Dauer bis zum ersten *Zero-Crossing Zeitpunkt* ist. Das heißt, dass $x(t)$ in einem Intervall der Mindestlänge r nicht nicht größer oder gleich 0 wird,

$$W(r) = P[x(t) < 0 \quad \forall \quad 0 \leq t \leq r]. \quad (\text{B.1})$$

Im folgenden soll gezeigt werden, dass die Wahrscheinlichkeit mit

$$W(r) < \exp(-Kr^\gamma) \quad \text{für ein gewisses} \quad K > 0 \quad (\text{B.2})$$

nach oben abgeschätzt werden kann.

Wenn $x(t) < 0$ für alle $0 \leq t \leq r$ (also sozusagen kontinuierliche t) gilt, dann gilt auch $x(j\Delta) < 0$ für alle $j = 0, 1, \dots, n-1$ für jede beliebige Diskretisierung des Intervalls $[0, r]$ mit Schrittweite Δ und $n = \lceil r/\Delta \rceil$ Schritten, wobei die eckigen Klammern in $\lceil r/\Delta \rceil$ die sogenannten Gauss-Klammern sind und bedeuten, dass n die größte ganze Zahl kleiner oder gleich r/Δ ist. Denn, da jeder Punkt im kontinuierlichen $x(t)$ kleiner 0 ist, muss auch jede Untermenge dieser Punkte kleiner 0 sein, also auch die $x(j\Delta)$. Es folgt

$$W(r) \leq P[x(j\Delta) < 0 \quad \forall \quad j = 0, 1, \dots, n-1]. \quad (\text{B.3})$$

Der Übergang vom kontinuierlichen System ins diskrete hat für die Kovarianzfunktion ρ folgende Konsequenzen: aus $\rho(s)$ wird $\rho(|j-k|\Delta)$ mit $j, k = 0, 1, \dots, n-1$ und $j \neq k$ und die Abschätzung $|\rho(s)| < Cs^{-\gamma}$ geht über in

$$|\rho(|j-k|\Delta)| < C(|j-k|\Delta)^{-\gamma}.$$

¹Da $\sigma_x^2 = 1$ ist die Kovarianzfunktion gleich der Autokorrelationsfunktion.

Der Ausdruck $(\rho_{jk}) = \rho(|j-k|\Delta)$ ist eine *positiv definite* $(n \times n)$ -Kovarianzmatrix, das heißt, (ρ_{jk}) ist symmetrisch und $\sum_{j,k=0}^{n-1} (\rho_{jk}) x_j x_k > 0$, was eine direkte Konsequenz des Gauss-Prozesses ist [Graybill 1983]. Ihre Hauptdiagonalelemente sind $\rho_{jj} := C_x(0) = \sigma_x^2 = 1$. Hieraus folgt die Existenz einer ebenfalls positiv definiten Inversen $(\rho_{jk})^{-1}$ (Theorem 12.2.1. in [Graybill 1983]). Für die Eigenwerte von (ρ_{jk}) gilt $0 < \lambda_1 \leq \dots \leq \lambda_n$ (Theorem 12.2.2 in [Graybill 1983]) und für die Eigenwerte der Inversen $0 < \lambda_n^{-1} \leq \dots \leq \lambda_1^{-1}$, also $\lambda_n^{-1} = 1/\lambda_n$ (Theorem 3.2.7 in [Graybill 1983]). Daraus folgt, dass der Kehrwert des größten Eigenwerts einer invertierbaren Matrix gleichzeitig der kleinste Eigenwert der Inversen Matrix ist. Aus der Invertierbarkeit und der positiven Definitheit von (ρ_{jk}) folgt außerdem, dass die Determinante $|(\rho_{jk})|$ positiv ist.

Die Wahrscheinlichkeit, dass in einem Intervall $[0, r)$ alle $x(j\Delta)$, $j = 0, 1, \dots, n-1$, unter 0 liegen, lässt sich für stationäre Gauss-verteilte Zufallszahlen mit Hilfe der Inversen der Kovarianzmatrix, $(\rho_{jk})^{-1}$, analytisch durch das multivariate Integral angeben [Sachs 1974]

$$W(r) \leq \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \int_{-\infty}^0 dx_0 \dots \int_{-\infty}^0 dx_{n-1} \exp\left(-\frac{1}{2} \sum_{j,k=0}^{n-1} (\rho_{jk})^{-1} x_j x_k\right). \quad (\text{B.4})$$

Dabei deuten die Integrationsgrenzen $-\infty$ bis 0 an, dass alle $x(j\Delta)$ in $[0, r)$ die Nulllinie $q = 0$ nicht überschreiten dürfen. Aus Symmetriegründen dürfen die Integralgrenzen auch von 0 bis ∞ gewählt werden. Dieses multivariate Integral ist analytisch nicht lösbar, da die Inverse der Kovarianzmatrix in ihren Einträgen von n und somit auch von r abhängt und man für jede Diskretisierung mit n Stützstellen eine neue Inverse der Kovarianzmatrix erstellen müsste. Abhilfe liefern die *Extremaleigenschaften der Eigenwerte* (englisch: *extremal properties of eigenvalues*) von positiv definiten Matrizen (Theorem 12.2.14 in [Graybill 1983]):

$$\sum_{j,k=0}^{n-1} (\rho_{jk})^{-1} x_j x_k \geq \lambda_n^{-1} \sum_{j=0}^{n-1} x_j^2 \quad (\text{B.5})$$

Mittels dieser Abschätzungen lässt sich $W(r)$ nach oben beschränken:

$$W(r) \leq \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \int_0^\infty dx_0 \dots \int_0^\infty dx_{n-1} \exp\left(-\frac{1}{2} \lambda_n^{-1} \sum_{j=0}^{n-1} x_j^2\right) \quad (\text{B.6})$$

$$\begin{aligned} &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \int_0^\infty dx_0 \exp\left(-\frac{1}{2} \lambda_n^{-1} x_0^2\right) \dots \int_0^\infty dx_{n-1} \exp\left(-\frac{1}{2} \lambda_n^{-1} x_{n-1}^2\right) \\ &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\int_0^\infty dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) \right)^n \\ &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\frac{\pi}{2\lambda_n^{-1}} \right)^{n/2} \quad (\text{B.7}) \end{aligned}$$

$$\begin{aligned} &= \frac{\lambda_n^{n/2}}{2^n |(\rho_{jk})|^{1/2}} \\ &\stackrel{\text{(i)}}{=} \frac{\lambda_n^{n/2}}{2^n (\prod_{j=1}^n \lambda_j)^{1/2}} \quad (\text{B.8}) \end{aligned}$$

$$\stackrel{\text{(ii)}}{\leq} \left(\frac{\lambda_n}{4\lambda_1} \right)^{n/2} \quad (\text{B.9})$$

Die Gültigkeit von (i) ergibt sich aus dem Theorem 8.6.5 in [Graybill 1983]: eine Determinante setzt sich aus dem Produkt der Eigenwerte zusammen. Der Schritt (ii) beruht auf der monotonen Abschätzung der Eigenwerte ($0 < \lambda_1 \leq \dots \leq \lambda_n$).

Betrachtet man die Messreihe bzw. den zeit-diskreten Prozess $x(j\Delta)_{j=0}^{n-1}$ als Vektor $\vec{x} \in \mathbb{R}^n$, so folgt aus der Eigenwertgleichung $((\rho_{ij}) - \lambda I_n)\vec{x} = 0$ (I_n ist die $(n \times n)$ -Einheitsmatrix, deren Diagonalelemente 1 und alle anderen Elemente 0 sind), dass $((\rho_{ij}) - I_n - (\lambda - 1)I_n)\vec{x} = 0$ gilt und $\lambda - 1$ Eigenwert der Matrix $((\rho_{ij}) - I_n)$ ist.

Mit Hilfe des Theorems 5.6.7 und der Definition 5.6.2 von Matrix-Normen in [Graybill 1983] lässt sich $|\lambda - 1|$ mittels der Zeilensummen abschätzen und es gilt

$$\begin{aligned}
|\lambda - 1| &\leq \max_{1 \leq i \leq n} \sum_{j=1}^n (|\rho_{ij}| - \delta_{ij}) && \text{(B.10)} \\
&= -1 + \max_{1 \leq i \leq n} \sum_{j=1}^n |\rho_{ij}| \\
&= -1 + \max_{1 \leq i \leq n} \sum_{j=1}^n |\rho(|i - j|\Delta)| \\
&= -1 + \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{i-1} |\rho(|i - j|\Delta)| + |\rho(0\Delta)| + \sum_{j=i+1}^n |\rho(|i - j|\Delta)| \right\} \\
&= \max_{1 \leq i \leq n} \left\{ \sum_{l=1}^{i-1} |\rho(l\Delta)| + \sum_{l=1}^{n-i} |\rho(l\Delta)| \right\} \\
&\leq \max_{1 \leq i \leq n} \left\{ \sum_{l=1}^{n-1} |\rho(l\Delta)| + \sum_{l=1}^{n-1} |\rho(l\Delta)| \right\} \\
&= 2 \sum_{l=1}^{n-1} |\rho(l\Delta)| && \text{(B.11)}
\end{aligned}$$

Löst man den Betrag von $|\lambda - 1|$ auf, erhält man für die Abschätzung des Eigenwerts

$$-2 \sum_{l=1}^{n-1} |\rho(l\Delta)| \leq \lambda - 1 \leq 2 \sum_{l=1}^{n-1} |\rho(l\Delta)| \quad \text{(B.12)}$$

$$1 - 2 \sum_{l=1}^{n-1} |\rho(l\Delta)| \leq \lambda \leq 1 + 2 \sum_{l=1}^{n-1} |\rho(l\Delta)| \quad \text{(B.13)}$$

Aus dieser Ungleichung lassen sich die Abschätzung für den größten und kleinsten Eigenwert, λ_n und λ_1 , der Kovarianzmatrix gewinnen

$$\lambda_n \leq 1 + 2 \sum_{l=1}^{n-1} |\rho(l\Delta)| \quad \text{(B.14)}$$

und

$$\lambda_1 \geq 1 - 2 \sum_{l=1}^{n-1} |\rho(l\Delta)|. \quad \text{(B.15)}$$

Aufgrund $|\rho(s)| < Cs^{-\gamma}$ bzw. $|\rho(l\Delta)| < C(l\Delta)^{-\gamma}$, wobei $l = |j - k|$ und $j \neq k$ ist, gilt für die Summen dieser Abschätzungen

$$\sum_{l=1}^{n-1} |\rho(l\Delta)| < C\Delta^{-\gamma} \sum_{l=1}^{n-1} l^{-\gamma} < C\Delta^{-\gamma} \int_{1/2}^{r/\Delta} l^{-\gamma} dl < C\Delta^{-1} r^{1-\gamma}. \quad (\text{B.16})$$

Setzt man das in die Eigenwertabschätzungen (B.14) und (B.15) und diese in die Abschätzung (B.24) für $W(r)$ ein, so erhält man

$$W(r) \leq \left(\frac{\lambda_n}{4\lambda_1} \right)^{n/2} \leq \left(\frac{1 + 2C\Delta^{-1}r^{1-\gamma}}{4(1 - 2C\Delta^{-1}r^{1-\gamma})} \right)^{n/2}. \quad (\text{B.17})$$

Da die Schrittweite Δ der Diskretisierung beliebig ist, darf sie frei gewählt werden. Also kann Δ für das aktuell betrachtete r auch so gewählt werden, dass die r -Abhängigkeit in der Klammer in (B.17) verschwindet und nur noch im Exponent aufgrund von $n = [r/\Delta]$ auftritt. Um das zu erreichen, genügt die Wahl

$$\Delta = K_0 r^{1-\gamma}, \quad (\text{B.18})$$

mit der Einschränkung $K_0 > 2C$, um eine Nullstelle im Nenner von (B.17) zu vermeiden² und einen reellen Ausdruck für $W(r)$ zu gewährleisten. Die freie Schrittweitenwahl Δ wird durch die freie Wahl des Parameters K_0 ersetzt. Setzt man (B.18) in (B.17) ein, erhält man

$$W(r) \leq \left(\frac{1 + 2C/K_0}{4 - 8C/K_0} \right)^{n/2} = \left(\frac{K_0 + 2C}{4(K_0 - 2C)} \right)^{n/2}. \quad (\text{B.19})$$

Da der Ausdruck in der Klammer für alle wählbaren $K_0 > 2C$ größer als $1/4$ bleibt, gilt

$$W(r) \leq \left(\frac{1}{4} \right)^{n/2} = e^{-\ln(4)n/2}. \quad (\text{B.20})$$

Man beachte, dass in n immer noch ein K_0 steckt: aus $n = [r/\Delta]$ folgt $n \geq r/\Delta = r^\gamma/K_0$

$$W(r) \leq \left(\frac{1}{4} \right)^{n/2} \leq e^{-\frac{\ln(4)r^\gamma}{2K_0}} = e^{-Kr^\gamma}. \quad (\text{B.21})$$

□

B.2 Erweiterung auf q -Level Crossing (POT Methode)

Für die Erweiterung des Beweises auf Quantile $q > 0$ setzen wir, dem Beweisschema folgend, in Gl. (B.4) ein und tauschen die oberen Integralgrenzen durch q aus. Dadurch wird der Wert des multivariaten Integrals erheblich vergrößert. Die Abschätzung aus den Extremeigenschaften der Eigenwerte (B.5) ist unabhängig von q und gilt weiterhin. Ebenso hält auch das Argument der Symmetrie des Gauss-Prozesses, sodass die Ungleichung (B.6) stimmt. Aber bei der Berechnung des äquivalenten Ausdrucks von Gl. (B.7) bricht der analytische Beweis

²Anm.: Es gibt auch eine triviale Einschränkung nach oben, nämlich durch $\Delta \ll r$, da man sonst keine brauchbare Zerlegung erhält.

zusammen, da es keinen geschlossenen Ausdruck zur Berechnung von Gauss-Integralen gibt³. Stattdessen ist aber folgende Abschätzung möglich.

$$W(r) \leq \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\int_{-q}^{\infty} dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) \right)^n \quad (\text{B.22})$$

$$\begin{aligned} &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\int_0^{\infty} dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) + \int_{-q}^0 dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) \right)^n \\ &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\int_0^{\infty} dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) + \int_0^q dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) \right)^n \\ &\stackrel{\text{(i)}}{\leq} \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\int_0^{\infty} dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) + \int_0^{\infty} dx \exp\left(-\frac{1}{2} \lambda_n^{-1} x^2\right) \right)^n \\ &= \frac{1}{(2\pi)^{n/2} |(\rho_{jk})|^{1/2}} \left(\frac{\pi}{\lambda_n^{-1}} \right)^{n/2} \quad (\text{B.23}) \end{aligned}$$

$$\leq \left(\frac{\lambda_n}{2\lambda_1} \right)^{n/2} \quad (\text{B.24})$$

Die Abschätzung in (i) ist zwar sehr grob, sie erhält aber die Struktur von $W(r)$. Durch die selbe Wahl der Schrittweite Δ werden alle Korrelationseigenschaften aus der Basis in den Exponenten verbannt und es folgt völlig analog zur obigen Beweisführung

$$W(r) \leq \left(\frac{1}{2} \right)^{n/2} \leq e^{-\frac{\ln(2)r^\gamma}{2K_0}} = e^{-Kr^\gamma}, \quad (\text{B.25})$$

wobei hier $K = \frac{\ln(2)r^\gamma}{2K_0}$, also gegenüber obigem K um einen Faktor 2 kleiner ist. Diese Erweiterung des Beweises auf beliebige q -Werte ist sehr grob, gibt aber das korrekte asymptotische Verhalten der Verteilungsfunktion wieder.

Der Beweis erklärt jedoch nicht den von uns gefundenen Datenkollaps der Krüven für $q > 0$, den man durch die mit R_q skalierte Auftragung erhält (siehe Abbildung 3.7(b)). Ich vermute aber, dass eine Erklärung für das Skalieren in der Wahl der Schrittweite (B.18) liegt.

³Der häufig verwendete Ausdruck der Gauss'schen Fehlerfunktion ist ein tabellierter Ausdruck (siehe [Bronstein 1995]) und ist hier wenig hilfreich.

Literaturverzeichnis

- [Alligood 1996] *Chaos - an introduction to dynamical systems*, K. T. Alligood, T. D. Sauer und J. A. Yorke, Springer, New York (1996)
- [Altmann 2005] E. G. Altmann und H. Kantz, Phys. Rev. E **71**, 056106 (2005)
- [Anderson 2003] *Fixed broadband wireless systems design*, H. R. Anderson, Wiley & Sons, New York (2003)
- [Arneodo 1995] A. Arneodo, E. Bacry, P. V. Graves und J. F. Muzy, Phys. Rev. Lett. **74**, 3293 (1995)
- [Ashkenazy 2000] Y. Ashkenazy, P. Ch. Ivanov, S. Havlin, C.-K. Peng, A. L. Goldberger und H. E. Stanley, Phys. Rev. Lett. **86**, 1900 (2000)
- [Ausloos 2001] M. Ausloos und K. Ivanova, Phys. Rev. E **63**, 47201 (2001)
- [Beran 1994] *Statistics for long memory processes, monograph in statistics and applied probability*, J. Beran, Chapman & Hall, Boca Raton (1994)
- [Berman 1964] S. M. Berman, Ann. Math. Statist. **35**, 502 (1964)
- [Blender 2003] R. Blender und K. Fraedrich, Geophys. Res. Lett. **30**, 1769 (2003)
- [BoM] 72 australische Temperaturzeitreihen vom National Climate Centre, Bureau of Meteorology, Australien
- [Box 1994] *Time series analysis - forecasting and control*, G. E. P. Box, G. M. Jenkins und G. C. Reinsel, Prentice-Hall, Englewood Cliffs (1994)
- [Bramwell 1998] S. T. Bramwell, P. C. W. Holdsworth und J.-F. Pinton, Nature **396**, 552 (1998)
- [Bramwell 2000] S. T. Bramwell, K. Christensen, J.-Y. Fortin, P. C. W. Holdsworth, H. J. Jensen, S. Lise, J. M. Lopez, M. Nicodemi, J.-F. Pinton und M. Sellitto, Phys. Rev. Lett. **84**, 3744 (2000)
- [Briffa 2000] K. R. Briffa, Quaternary Science Reviews **19**, 87 (2002)
- [Brockwell 1996] *Introduction to time series and forecasting*, P. J. Brockwell und R. A. Davis, Springer, New York (1996)

- [Bronstein 1995] *Taschenbuch der Mathematik*, I. N. Bronstein, K. A. Semendjajew, G. Musiol, H. Mühlig, Verlag Harry Deutsch, Frankfurt am Main (1995)
- [Buldyrev 1995] S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsu, C.-K. Peng, M. Simons und H. E. Stanley, *Phys. Rev. E* **51**, 5084 (1995)
- [Bunde 1994] *Fractals in Science*, Hrsg. A. Bunde und S. Havlin, Springer, Berlin (1994)
- [Bunde 2000] A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter und K. Voigt, *Phys. Rev. Lett.* **85**, 3736 (2000)
- [Bunde 2001] A. Bunde und J. W. Kantelhardt, *Physikalische Blätter* **57** Nr. 5, 49 (2001)
- [Bunde 2002a] A. Bunde und S. Havlin, *Physica A* **314**, 15 (2002)
- [Bunde 2002b] *The science of disasters - climate disruptions, heart attacks, and market crashes*, Hrsg. A. Bunde, J. Kropp und H.-J. Schellnhuber, Springer, Berlin (2002)
- [Bunde 2003] A. Bunde, J. F. Eichner, S. Havlin und J. W. Kantelhardt, *Physica A* **330**, 1 (2003)
- [Bunde 2004a] A. Bunde, J. F. Eichner, S. Havlin und J. W. Kantelhardt, *Physica A* **342**, 308 (2004)
- [Bunde 2004b] A. Bunde, J. F. Eichner, S. Havlin, E. Koscielny-Bunde, H.-J. Schellnhuber und D. Vjushin, *Phys. Rev. Lett.* **92**, 039801 (2004)
- [Bunde 2004c] A. Bunde, J. F. Eichner, J. W. Kantelhardt und S. Havlin in *Proceedings of the international school of physics "Enrico Fermi"*, Hrsg. F. Mallamace und H. E. Stanley, IOS Press, Amsterdam, 421 (2004)
- [Bunde 2004d] A. Bunde, J. F. Eichner, S. Havlin, J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, D. Vjushin und I. Zhidkov, *Proceedings of Iranian Phys. Soc.*, Teheran (2004)
- [Bunde 2005] A. Bunde, J. F. Eichner, J. W. Kantelhardt und S. Havlin, *Phys. Rev. Lett.* **94**, 048701 (2005)
- [BWG 2003] *Hochwasserabschätzung in schweizersichen Einzugsgebieten*, Bericht des BWG, Serie Wasser Nr. 4, Bern (2003)
- [Charney 1979] J. G. Charney und J. Devore, *J. Atmos. Sci.* **36**, 1205 (1979)
- [Chatfield 1982] *Analyse von Zeitreihen*, C. Chatfield, Hanser, München (1982)
- [Chow 1964] *Handbook of applied hydrology: a compendium of water-resources technology*, Hrsg. V. T. Chow, McGraw-Hill, New York (1964)
- [Chow 1988] *Applied hydrology*, V. T. Chow, D. R. Maidment und L. W. Mays, McGraw-Hill, New York (1988)
- [Christensen 2002] J. H. Christensen und O. B. Christensen, *Nature* **421**, 805 (2002)
- [Cohn 2005] T. A. Cohn und H. F. Lins, *Geophys. Res. Lett.* **32**, 23402 (2005)

- [Coles 2001] *An introduction to statistical modeling of extreme values*, S. Coles, Springer, London (2001)
- [Corral 2004] A. Corral, Phys. Rev. Lett. **92**, 108501 (2004)
- [Danielsson 2000] J. Danielsson und G. C. de Vries, Annals d'Economie et de Statistique **60**, 239 (2000)
- [Davison 1990] A. C. Davison und R. L. Smith, J. of the Royal Statist. Soc. **52**, 393 (1990)
- [Dean 2001] D. S. Dean und S. N. Majumdar, Phys. Rev. E **64**, 046121 (2001)
- [Diercke 1996] *Diercke Weltatlas*, Westermann Schulbuchverlag, 220 (1996)
- [DVWK 1999] *Statistische Analyse von Hochwasserabflüssen, Merkblatt 251*, Kommissionsvertrieb Wirtschafts- und Verglagsgesellschaft Gas und Wasser mbH, Bonn (1999)
- [Ehrhardt 2004] G. M. C. A. Ehrhardt, S. N. Majumdar und A. J. Bray, Phys. Rev. E **69**, 016106 (2004)
- [Eichner 2002] J. F. Eichner, Diplomarbeit, Giessen, Sommer 2002
- [Eichner 2003] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin und H.-J. Schellnhuber, Phys. Rev. E **68**, 046133 (2003)
- [Eichner 2006a] J. F. Eichner, J. W. Kantelhardt, A. Bunde und S. Havlin, Phys. Rev. E, im Druck (2006)
- [Eichner 2006b] J. F. Eichner, J. W. Kantelhardt, A. Bunde und S. Havlin, Phys. Rev. E, eingereicht (2006)
- [Embrechts 1997] *Modelling extremal events for insurance and finance*, P. Embrechts, C. Klüppelberg und T. Mikosch, Springer, Berlin (1997)
- [Embrechts 1999] P. Embrechts, S. I. Resnik und G. Samorodnitsky, North American Actuarial Journal **3**, 30 (1999)
- [Esper 2002] J. Esper, E. R. Cook und F. H. Schweingruber, Science **295**, 2250 (2002)
- [Feder 1988] *Fractals*, J. Feder, Plenum Press, New York (1988)
- [Feller 1951] W. Feller, Ann. Math. Stat. **22**, 427 (1951)
- [Fisher 1928] R. A. Fisher und H. C. Tippett, Proc. Cambridge Phil. Soc. **24**, 180 (1928)
- [Fließbach 1995] *Statistische Physik*, T. Fließbach, Spektrum, Heidelberg (1995)
- [Fraedrich 2003] K. Fraedrich und R. Blender, Phys. Rev. Lett. **90**, 108501 (2003)
- [Frechet 1927] M. Frechet, Ann. Soc. Math. Polon. **6**, 93 (1927)
- [Galambos 1978] *The asymptotic theory of extreme order statistics*, J. Galambos, Wiley & Sons, New York (1978)

- [Galambos 1994] *Extreme value theory and applications*, Hrsg. J. Galambos, J. Lechner und E. Simin, Kluwer, Dordrecht (1994)
- [Gell-Mann 2003] A. Bunde, J. F. Eichner, R. B. Govindan, S. Havlin, E. Koscielny-Bunde, D. Rybski und D. Vjushin in *Nonextensive entropy - interdisciplinary applications*, Hrsg. M. Gell-Mann und C. Tsallis, Oxford University Press, New York (2003)
- [Glaser 2001] *Klimageschichte Mitteleuropas, 1000 Jahre Wetter, Klima, Katastrophe*, R. Glaser, Wiss. Buchgesellschaft, Darmstadt (2001)
- [Glaser 2003] R. Glaser und H. Stangl, *Natural Hazards and Earth System Science* **3**, 605 (2003)
- [Gnedenko 1943] B. V. Gnedenko, *Ann. Math.* **44**, 423 (1943)
- [Gottman 1981] *Time-series analysis*, J. Gottman, Cambridge University Press, Cambridge (1981)
- [Govindan 2002] R. B. Govindan, D. Vjuschin, A. Bunde, S. Brenner, S. Havlin und H.-J. Schellnhuber, *Phys. Rev. Lett.* **89**, 028501 (2002)
- [Grassel 2005] *Wetterkatastrophen und Klimawandel: sind wir noch zu retten?*, H. Grassel, C. Pfister, M. Latif, S. Rahmstorf, H. J. Schellnhuber, H. v. Storch u.v.a., Edition Wissen, Pg-Verlag (2005)
- [Graybill 1983] *Matrices with applications in statistics*, F. A. Graybill, Wadsworth Publishing Company, Belmont (1983)
- [Greene 1997] *Production and inventory control handbook*, J. H. Greene, McGraw-Hill, New York (1997)
- [Grissino-Meyer 1996] H. D. Grissino-Mayer in *Tree rings, environment, and humanity*, Hrsg. J. S. Dean, D. M. Meko und T. W. Swetnam, Radiocarbon 1996, Department of Geosciences, The University of Arizona, Tucson (1996)
- [Guclu 2004] H. Guclu und G. Korniss, *Phys. Rev. E* **69**, 065104 (2004)
- [Gumbel 1958] *Statistics of extremes*, E. J. Gumbel, Columbia University Press, New York (1958)
- [Haight 1967] *Handbook of the Poisson Distribution*, F. A. Haight, Wiley & Sons, New York (1967)
- [Henze 1997] *Stochastik für Einsteiger*, N. Henze, Vieweg-Verlag, Wiesbaden (1997)
- [Hosking 1997] *Regional frequency analysis - an approach based on L-Moments*, J. M. R. Hosking und J. R. Wallis, Cambridge University Press, Cambridge (1997)
- [Hupf 1998] P. Hupfer in *Witterung und Klima*, Hrsg. W. Kuttler, Teubner, Stuttgart (1998)
- [Hurst 1951] H. E. Hurst, *Trans. Am. Soc. Civ. Eng.* **116**, 770 (1951)
- [Hurst 1965] *Long-Term Storage in Reservoirs: An experimental Study*, H. E. Hurst, R. Black und Y. M. Sinaika, Constable & Co. Ltd., London (1965)

- [Ivanov 1999] P. Ch. Ivanov, A. Bunde, L. A. N. Amaral, J. Fritsch-Yelle, R. M. Baevsky, S. Havlin, H. E. Stanley und A. L. Goldberger, *Europhys. Lett.* **48**, 594 (1999)
- [Jacobeit 2003] J. Jacobeit, R. Glaser, J. Luterbacher und H. Wanner, *Geophys. Res. Lett.* **30**, 1172 (2003)
- [Jenkinson 1969] *Estimation of maximum floods*, A. F. Jenkinson, World Meteorological Institution, Technical Note **98** (1969)
- [Jones 1998] P. D. Jones, K. R. Briffa, T. P. Barnett und S. F. B. Tett, *Holocene* **8**, 455 (1998)
- [Kantelhardt 2001a] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin und A. Bunde, *Physica A* **295**, 441 (2001)
- [Kantelhardt 2001b] J. W. Kantelhardt, Y. Ashkenazy, P. Ch. Ivanov, A. Bunde, S. Havlin, T. Penzel, H. J. Peter und H. E. Stanley, *Phys. Rev. E* **65**, 051908 (2002)
- [Kantelhardt 2002] J. W. Kantelhardt, S. A. Zschiegner, E. K. Koscielny-Bunde, A. Bunde, S. Havlin und H.E. Stanley, *Physica A* **316**, 87 (2002)
- [Kantelhardt 2003a] J. W. Kantelhardt, T. Penzel, S. Rostig, H. F. Becker, S. Havlin und A. Bunde, *Physica A* **319**, 447 (2003)
- [Kantelhardt 2003b] J. W. Kantelhardt, D. Rybski, S. A. Zschiegner, P. Braun, E. Koscielny-Bunde, V. Livina, S. Havlin und A. Bunde, *Physica A* **330** (2003)
- [Kantelhardt 2006] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde und S. Havlin, *J. Geophys. Res. Atmos.* **111**, 1106 (2006)
- [Kendall 1990] *Time series*, M. Kendall und J. Ord, Oxford University Press (1990)
- [Kiraly 2005] A. Kiraly und I. M. Janosi, *Meteorol. Atmos. Phys.* **88**, 119 (2005)
- [Klemes 1974] V. Klemes, *Water Resour. Res.* **10**, 675 (1974)
- [Knuth 1997] *The art of computer programming*, D. Knuth, Vol. **3**, Addison-Wesley, 113 (1997)
- [Koscielny 1996] E. Koscielny-Bunde, A. Bunde, S. Havlin und Y. Goldreich, *Physica A* **231**, 393 (1996)
- [Koscielny 1998a] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich und H.-J. Schellnhuber, *Phys. Rev. Lett.* **81**, 729 (1998)
- [Koscielny 2006] E. Koscielny-Bunde, J. W. Kantelhardt, P. Braun, A. Bunde und S. Havlin, *J. of Hydrology*, im Druck (2006)
- [Lam 1997] *Introduction to nonlinear physics*, L. Lam, Springer (1997)
- [Landwehr 1980] J. M. Landwehr, N. C. Matalas und J. R. Wallis, *Water Resour. Res.* **16**, 547 (1980)

- [Latif 2004] *Hitzerekorde und Jahrhundertflut. Herausforderung Klimawandel*. M. Latif, Heyne Verlag (2003)
- [Leadbetter 1983] *Extremes and related properties of random sequences and processes*, M. R. Leadbetter, G. Lindgren und H. Rootzen, Springer, New York (1983)
- [Liu 1997] Y. H. Liu, P. Cizeau, M. Meyer, C.-K. Peng und H. E. Stanley, *Physica A* **245**, 437 (1997)
- [Liu 1999] Y. H. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng und H. E. Stanley, *Phys. Rev. E* **60**, 1390 (1999)
- [Livina 2005] V. N. Livina, S. Havlin und A. Bunde, *Phys. Rev. Lett.* **95**, 208501 (2005)
- [Lovejoy 1995] S. Lovejoy und D. Schertzer in *New Uncertainty Concepts in Hydrology and Water Resources*, Hrsg. Z. W. Kundzewicz, Cambridge Univ. Press, New York, 61 (1995)
- [L'vov 2001] V. S. L'vov, A. Pomyalov und I. Procaccia, *Phys. Rev. E* **63**, 056118 (2001)
- [Majumdar 2004] S. N. Majumdar und A. Comtet, *Phys. Rev. Lett.* **92**, 225501 (2004)
- [Mandelbrot 1969] B. B. Mandelbrot und J. R. Wallis, *Water Resour. Res.* **5**, 321 (1969)
- [Mandelbrot 1971] B. B. Mandelbrot, *Water Resour. Res.* **7**, 543 (1971)
- [Mandelbrot 1987] *Die fraktale Geometrie der Natur*, B. B. Mandelbrot, Basel, Boston (1987)
- [Mann 1999] M. E. Mann, R. S. Bradley und M. K. Hughes, *Geophys. Res. Lett.* **26**, 759 (1999)
- [Makse 1996] H. A. Makse, S. Havlin, M. Schwartz und H. E. Stanley, *Phys. Rev. E* **53**, 5445 (1996)
- [Meko 2001] D. M. Meko, M. D. Therrell, C. H. Baisan und M. K. Hughes, *J. Amer. Wat. Resour. Assoc.* **37**, 4 (2001)
- [Moberg 2005] A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko und W. Karlén, *Nature* **433**, 613 (2005)
- [Monetti 2003] R. A. Monetti, S. Havlin und A. Bunde, *Physica A* **320**, 581 (2003)
- [Moore 2001] J. J. Moore, K. A. Huguen, G. H. Miller und J. T. Overpeck, *J. of Paleolimnology* **25**, 503 (2001)
- [Mudelsee 2003] M. Mudelsee, M. Börngen, G. Tetzlaff und U. Grünwald, *Nature* **425**, 166 (2003)
- [Mudelsee 2004] M. Mudelsee, M. Börngen, G. Tetzlaff und U. Grünwald, *J. Geophys. Res.* **109**, 23101 (2004)
- [Naess 1998] A. Naess, *J. of Engineering Mechanics*, ASCE **124**, 252 (1998)
- [Newell 1962] G. F. Newell und M. Rosenblatt, *Ann. Math. Statist.* **33**, 1306 (1962)

- [NOAA] Paleoclimatological Data Bank, National Oceanic & Atmospheric Administration (NOAA), <http://www.ngdc.noaa.gov/paleo/recons.html>
- [Noren 2002] A. J. Noren, P. R. Bierman, E. J. Steig, A. Lini und J. Southon *Nature* **419**, 821 (2002)
- [Papoulis 1984] *Probability, random variables and stochastic processes*, A. Papoulis, McGraw-Hill, Singapore (1984)
- [Papula 1994] *Mathematik für Ingenieure und Naturwissenschaftler - Band 3*, L. Papula, Vieweg-Verlag, Wiesbaden (1994)
- [Parker 1994] D. E. Parker, P. D. Jones, C. K. Folland und A. Bevan, *J. Geophys. Res.* **99**, 14373 (1994)
- [Pelletier 1997] J. D. Pelletier und D. L. Turcotte, *J. of Hydrology.* **203**, 198 (1997)
- [Pelletier 1999] J. D. Pelletier und D. L. Turcotte, *Adv. Geophys.* **40**, 91 (1999)
- [Peng 1991] C.-K. Peng, S. Havlin, M. Schwartz und H. E. Stanley, *Phys. Rev. A* **44**, 2239 (1991)
- [Peng 1992] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons und H. E. Stanley, *Nature* **356**, 168 (1992)
- [Peng 1993] C.-K. Peng, J. Mietus, M. Hausdorff, S. Havlin, H. E. Stanley und A. L. Goldberger, *Phys. Rev. Lett.* **70**, 1343 (1993)
- [Peng 1994] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley und A. L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994)
- [Pfister 1999] *Wetternachhersage - 500 Jahre Klimavariationen und Naturkatastrophen*, C. Pfister, Paul Haupt Verlag (1999)
- [Phien 1987] H. N. Phien, *J. of Hydrology* **86**, 251 (1987)
- [PIK] 18 kontinentale und 2 küstennahe Temperaturzeitreihen vom Potsdamer Institut für Klimafolgenforschung, Potsdam
- [Poisson 1837] *Recherches sur la probabilité des jugements eu matière criminelle et eu matière civile*, precedees des regles generales du calcul des probabilités, S.-D. Poisson, Bachelier, Paris (1837)
- [Prakash 1992] S. Prakash, S. Havlin, M. Schwartz und H. E. Stanley, *Phys. Rev. A* **46**, 1724 (1992)
- [Prasad 1988] R. R. Prasad, C. Meneveau und K. R. Sreenivasan, *Phys. Rev. Lett.* **61**, 74 (1988)
- [Press 1997] *Numerical recipes in C: the art of scientific computing*, W. H. Press, B. P. Flannery, S. A. Teukolsky und W. T. Vetterling, Cambridge University Press, Cambridge (1992)

- [Rangarajan 2000] G. Rangarajan und M. Ding, Phys. Rev. E **61**, 4991 (2000)
- [Raudkivi 1979] *Hydrology*, A. J. Raudkivi, Pergamon Press, Oxford, 1979
- [Raychaudhuri 2001] S. Raychaudhuri, M. Cranston, C. Przybyla und Y. Shapir, Phys. Rev. Lett. **87**, 136101 (2001)
- [Reiss 2001] *Statistical analysis of extreme values: with applications to insurance, finance, hydrology, and other fields* R. D. Reiss und M. Thomas, Birkhauser Verlag (2001)
- [Rodriguez-Iturbe 1997] *Fractal river basins - change and self organization*, I. Rodriguez-Iturbe, Cambridge University Press, Cambridge (1997)
- [Rybski 2002] D. Rybski, Diplomarbeit, Giessen, Frühjahr 2002
- [Rybski 2006] D. Rybski, A. Bunde, S. Havlin und H. von Storch Geophys. Res. Lett., eingereicht (2006)
- [Sachs 1974] *Angewandte Statistik: Planung und Auswertung, Methoden und Modelle*, L. Sachs, Springer, Berlin (1974)
- [Saupe 1988] D. Saupe in *The Science of Fractal Images*, Hrsg. H.-O. Peitgen und D. Saupe, Springer, New York (1988)
- [Schäfer 1998] C. Schäfer, M. G. Rosenblum, J. Kurths und H. H. Abel, Nature **392**, 239 (1998)
- [Schlesinger 1987] M. F. Schlesinger, B. J. West und J. Klafter, Phys. Rev. Lett. **58**, 1100 (1987)
- [Schlittgen 1990] *Zeitreihenanalyse*, R. Schlittgen und B. H. J. Streitberg, Oldenburg-Verlag (1990)
- [Schreiber 1996] T. Schreiber und A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996)
- [Schreiber 2000] T. Schreiber und A. Schmitz, Physica D **142**, 346 (2000)
- [Shiau 2003] J. T. Shiau, Stochastic Environmental Research and Risk Assessment **17**, 42 (2003)
- [Smith 1989] R. L. Smith in *Handbook of applicable mathematics*, Hrsg. W. Ledermann, E. Lloyd, S. Vajda und C. Alexander, Wiley & Sons, New York (1989)
- [Smith 1996] R. L. Smith, Statistical Science **4**, 367 (1996)
- [Soon 2003] W. Soon und S. Baliunas, Climate Res. **23**, 89 (2003)
- [Sornette 1997] D. Sornette und L. Knopoff, Bull. Seismol. Soc. Am. **87**, 789 (1997)
- [Stöcker 1995] *Taschenbuch mathematischer Formeln und moderner Verfahren*, Hrsg. H. Stöcker, Verlag Harry Deutsch, Frankfurt am Main (1995)
- [v.Storch 1999] *Statistical analysis in climate research*, H. v. Storch und F. W. Zwiers, Cambridge University Press, Cambridge (1999)

- [Talkner 2000] P. Talkner und R. O. Weber, Phys. Rev. E **62**, 150 (2000)
- [Tessier 1993] Y. Tessier, S. Lovejoy und D. Schertzer, J. Appl. Meteorol. **32**, 223 (1993)
- [Tessier 1996] Y. Tessier, S. Lovejoy, P. Hubert, D. Schertzer und S. Pecknold, J. Geophys. Res. Atmos. **101**, 26427 (1996)
- [Theiler 1992] J. Theiler, B. Galdrikian, E. Eubank, A. Longtin und J. D. Farmer, Physica D **58**, 77 (1992)
- [Tousson 1925] O. Tousson, Mémoire sur l'Histoire du Nil, Mémoire de l'Institute d'Egypte **18**, 366 (1925)
- [Turcotte 1992] *Fractals and chaos in geology and geophysics*, D. L. Turcotte, Cambridge University Press, Cambridge (1992)
- [Vazquez 2006] A. Vazquez, J. G. Oliveira, Z. Dezso, K.-I. Goh, I. Kondor und A.-L. Barabasi, Phys. Rev. E, im Druck (2006)
- [Vjushin 2002] D. Vjushin, R. B. Govindan, S. Brenner, A. Bunde, S. Havlin und H.-J. Schellhuber, J. Phys.: Condens. Matter **14**, 2275 (2002)
- [Vjushin 2004] D. Vjushin, I. Zhidkov, S. Havlin, A. Bunde und S. Brenner, Geophys. Res. Lett. **31**, 10206 (2004)
- [Weber 2001] R. O. Weber und P. Talkner, J. Geophys. Res. Atmos. **106**, 20131 (2001)
- [Whitcher 2002] B. Whitcher, S. D. Byers, P. Guttorp und D. B. Percival, Water Resour. Res. **38**, 1054 (2002)
- [WKT] Wiener-Khintchin-Theorem: "Das Energiedichtespektrum eines (reellen) Energiesignals ist die Fourier-Transformierte der Autokorrelationsfunktion.", siehe auch Ref. [Press 1997]
- [Zurmühl 1963] *Praktische Mathematik für Ingenieure und Physiker*, R. Zurmühl, Springer, Berlin (1963)

Danksagung

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Armin Bunde bedanken, der mich für das Thema begeistern konnte und mir mit hervorragender und fokussierter Betreuung stets zur Seite stand.

Ein ganz besonderer Dank geht an Herrn Prof. Dr. Jan W. Kantelhardt, der mich bei allen Fragestellungen und Problemen hilfreich, geduldig und unermüdlich unterstützte.

Weiterer Dank gebührt Herrn Prof. Dr. Shlomo Havlin aus Israel, der mir bei seinen zahlreichen Besuchen in Giessen in vielen Diskussionen neue Impulse gab.

Weiterer Dank gebührt den stets hilfsbereiten Kolleginnen im Giessener Institut für Theoretische Physik III Frau Dr. Eva Koscielny-Bunde, Frau Dr. Stefanie Ruß und Frau Ilka Sproates, sowie den ehemaligen Kollegen Dr. Yuriy Hlushchuk und Dr. Markus Ulrich.

Ein ganz spezieller Dank geht an meine zuverlässigen Kollegen und “Mitstreiter” Diego Rybski und Stephan Zschiegner. Unsere gemeinsamen nachmittäglichen Kakao-Pausen haben so manchen arbeitsamen Tag wieder aufgefrischt.

Außerdem danke ich meinem Bruder Dr. Gerrit Eichner von ganzem Herzen für große Dienste zum Verständnis mathematischer Beweisführungen.

Mein größter Dank jedoch geht an meine Eltern Lutz und Edith Eichner, ohne die ich niemals da wäre, wo ich bin. Sie haben mir immer in allen Belangen ihr vollstes Vertrauen und großzügige Unterstützung gegeben.

Hiermit versichere ich, dass ich die vorliegende Promotionsarbeit selbständig angefertigt.

Giessen, im Januar 2006