FACHBEREICH PHYSIK
BERGISCHE UNIVERSITÄT
GESAMTHOCHSCHULE WUPPERTAL

WU D 97-14
15. MAI 1997

Charakterisierung und Klassifikation von Zeitreihen

Diplomarbeit (DII)

vorgelegt von

Andreas Schmitz

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Einleitung

> Sartorius betrachtete das Endstück nochmals, mißtrauisch verfinstert, als suchte er den in dieses Liniengezack eingeflossenen Inhalt zu entschlüsseln.
>
> *Solaris* (Stanisław Lem)

Die Zeit prägt in vielerlei Hinsicht das menschliche Leben. Sie dient unter anderem der Orientierung und als Ordnungskriterium. Jedes Ereignis, jedes Handeln und jede Beobachtung ist fest mit einem Zeitpunkt verknüpft. Auch jeder quantitativen Messung kann der Zeitpunkt, zu dem die Messung stattfand, zugeordnet werden und zeitliche Entwicklungen können nur anhand zeitlich aufeinander folgender Messungen, sogenannten *Zeitreihen*, untersucht werden. Die Untersuchung von zeitlich variierenden Systemen ist in nahezu allen Bereichen der Wissenschaft von Interesse. So werden z.B. in der Astronomie die Bahnen von Himmelskörpern, kosmische Strahlungen oder die Anzahl der Sonnenflecken beobachtet. In der Biologie werden Populationen, Wachstum und Lebensraumveränderungen untersucht. Die Wirtschaftswissenschaften beschäftigen sich mit Börsenkursen, Arbeitslosenzahlen und anderen Marktdaten. Und schließlich werden in der Medizin Herzmuskelpotentiale in Form von *Elektrokardiogrammen* (EKG), die elektrische Hirnaktivität in Form von *Elektroencephalogrammen* (EEG), die Atmung oder verschiedene Blutwerte gemessen.

Für einige der aufgezählten Beispiele sind erste Gesetzmäßigkeiten relativ leicht zu erkennen. So sind Planetenbahnen in unserem Sonnensystem im wesentlichen durch Ellipsen beschreibbar. In anderen Fällen kann auf den ersten Blick kein regelmäßiges Verhalten und auch keine Struktur in den Daten entdeckt werden. Diese Zeitreihen zeigen *Irregularitäten*, die nicht durch periodische oder quasiperiodische[1] Prozesse beschrieben werden können. In der empirischen, linearen Zeitreihenanalyse werden solche Daten durch *Zufallsprozesse* (*stochastische* Prozesse) modelliert und beschrieben. Einen zentralen Punkt nimmt dabei das *ARMA-Modell* ein, das in Abschnitt 1.2.2 eingeführt wird.

[1] Unter Quasiperiodizität versteht man eine Überlagerung von periodischen Prozessen, deren Frequenzverhältnis nicht durch rationale Zahlen ausgedrückt werden kann.

Die Erklärung von Irregularitäten kann aber auch anders motiviert werden. Zu Beginn dieses Jahrhunderts beschäftigte sich Poincaré erstmals intensiv mit dem Dreikörperproblem, also z.B. den Bahnen von drei Himmelskörpern. Er stellte fest, daß eine kleine Änderung der Anfangskonstellation im zeitlichen Verlauf auch in deterministischen Systemen zu völlig unterschiedlichen, komplizierten Bahnen führen kann. Dieses Phänomen, der sensitiven Abhängigkeit von den Anfangsbedingungen, wird *chaotisches* Verhalten genannt. Die *Nichtlinearität* der Gleichungen, die das System beschreiben, ist eine notwendige Bedingung für das Auftreten von Chaos. Die Erkenntnis, daß Irregularitäten auch ohne Zuhilfenahme von Zufallselementen möglich sind, fand ihre Anwendung in der *nichtlinearen Zeitreihenanalyse*, welche als Ergänzung zur üblichen, linearen Zeitreihenanalyse angesehen werden kann. Quantitative Größen aus der Theorie der nichtlinearen, dynamischen Systeme können dabei in vielen Fällen auf die nichtlineare Zeitreihenanalyse übertragen werden. Die Anwendung der nichtlinearen Zeitreihenanalyse ist aber nur dann sinnvoll, wenn der zugrundeliegende Prozeß tatsächlich nichtlinear ist. In Kapitel 2 wird ein Test entwickelt, der Nichtlinearität in Zeitreihen aufspüren kann. Dies ist nicht trivial, da gemessene Zeitreihen fast immer eine Mischung aus nichtlinearen und stochastischen Prozessen darstellen.

Mit einem solchen Test wird die große Menge der Zeitreihen in zwei Gruppen unterteilt. Die eine Gruppe enthält Zeitreihen, die aus einem nichtlinearen Prozeß enstanden sind, und die andere Gruppe die Zeitreihen, deren zugrundeliegender Prozeß linear ist. Der Test führt also eine *Klassifikation* der Zeitreihen durch.

Die Fähigkeit, Situationen und Beobachtungen richtig zu beurteilen und einzuordnen, war und ist von äußerster Wichtigkeit. Mensch und Tier können nur überleben, wenn sie einerseits gefährliche Situationen oder Gegner von ungefährlichen unterscheiden können und andererseits giftige Nahrung von nahrhafter. Dabei sollten diese Klassifizierungen und Entscheidungen möglichst schnell durchgeführt werden, um in kürzester Zeit reagieren zu können. Die entsprechenden Klassen, in die Vorgänge und Wahrnehmungen einsortiert werden, sind in der Regel durch einen Lernprozeß gebildet worden. Gelernt wird z.B. durch die Beobachtungen von Tätigkeiten der Eltern oder durch selbstgemachte Erfahrungen.

Aber auch für ein effizientes Arbeiten des menschlichen Gehirns ist die Klassifikation wesentlich. So muß z.B. ein Mensch, der gelernt hat, einem Baum von anderen Objekten zu unterscheiden, bei einem Spaziergang durch einen Wald nicht bei jedem Anblick eines Baumes lange über die Beobachtung nachdenken. Vielmehr geschieht das Einsortieren in die Klasse „Baum" im Unterbewußtsein nahezu simultan zu der Beobachtung und ohne geistige Anstrengung. Die klare Eingrenzung der Klasse „Baum" anhand der typischen Merkmale eines Baumes ist dabei von entscheidender Bedeutung. Stimmt nur ein Merkmal nicht mit der Klasse überein (z.B. ein Baum mit blauen Blättern), beginnt ein bewußter Denkprozeß.

Auch im heutigen Informationszeitalter spielt die Klassifikation eine immer wichtigere Rolle. Der Mensch muß in der Informationsflut die für ihn interessanten und wichtigen Informationen von den uninteressanten trennen können. Nichts anderes macht z.B. eine Presseagentur mit eingehenden Meldungen. Dabei soll immer mehr der Computer den Menschen entlasten, da sich Computer für die Bearbeitung von großen Datenmengen ideal eignen. Das Sortieren der Daten nach bestimmten Kriterien, also die Klassifizierung der Daten, ist z.B. für ein gezieltes, schnelles Suchen nach Informationen sehr nützlich.

Da Meßdaten und Beobachtungen unter anderem in Form von Zeitreihen vorliegen können, ist in vielen Bereichen die Möglichkeit, Zeitreihen zu klassifizieren sehr wichtig. So kann z.B. ein erfahrener Automechaniker bestimmte Fehler oder Mängel am Fahr- oder Motorengeräusch erkennen. Über elektronisch aufgenommene Geräusche oder andere Telemetriedaten ist eine Unterscheidung der einzelnen Fehler auch in der Zeitreihenanalyse durch die Klassifikation von Zeitreihen möglich. Zusätzlich könnte vor unbekannten oder schwer erkennbaren Fehlern durch die Untersuchung von Abweichungen vom „Normalzustand“ gewarnt werden. In der Medizin hat man es mit ähnlichen Problemen zu tun. Zunächst muß ein Arzt erkennen können, ob ein Patient krank oder gesund ist. Ist der Patient nicht gesund, muß genau diagnostiziert werden, welche Krankheit vorliegt, um eine gezielte Behandlung einzuleiten. In einigen Fällen ist die die Analyse von Zeitreihen, wie EKGs und EEGs, sehr hilfreich. Ein Arzt kann z.B. anhand des EKGs viele Herzfehler erkennen. Auch hier findet also eine Klassifikation von Zeitreihen statt. Krankheiten, die sich bisher durch reines Betrachten z.B. des EEGs nicht eindeutig zugeordnen lassen, können vielleicht mit Hilfe der Zeitreihenanalyse korrekt und frühzeitig erkannt werden.

Die Nützlichkeit der Klassifikation von Zeitreihen für verschiedenste Anwendungen steht also außer Frage. In Kapitel 3 dieser Arbeit wird ein Verfahren zur Klassifikation von Zeitreihen entwickelt und an Beispielen untersucht und angewendet. Der Schwerpunkt liegt dabei auf Klassifikationskriterien aus dem Bereich der nichtlinearen Dynamik.

Kapitel 1

Allgemeines zur Zeitreihenanalyse

In diesem Kapitel sollen einige allgemeine Gesichtspunkte der Zeitreihenanalyse in soweit erläutert werden, wie sie für diese Arbeit relevant sind. Dabei werde ich zuerst auf lineare Methoden und Prozesse eingehen. In Kapitel 2 sollen dann Daten daraufhin klassifiziert werden, ob sie durch einen linearen oder nichtlinearen Prozeß erzeugt worden sind. Dafür ist es notwendig, einige Eigenschaften von linearen Prozessen zu kennen.

1.1 Zeitreihen

Allgemein versteht man unter einer Zeitreihe eine diskrete Folge von Messungen, die zu aufeinanderfolgenden Zeitpunkten aufgenommen werden. Eine Zeitreihe ist durch $\{\vec{s}_n\}, n = 1, \ldots, N$ darstellbar, wobei N die Anzahl der Messungen ist. Der zugrundeliegende Prozeß $\vec{x}(t)$ kann natürlich auch für kontinuierliche Zeiten definiert sein. So kann $\vec{x}(t)$ z.B. die Lösung einer Differentialgleichung sein. Ich betrachte hier lediglich skalare Zeitreihen $\{s_n\}$ mit $n = 1, \ldots, N$. Die Messung bildet den im allgemeinen mehrdimensionalen Zustandsvektor $\vec{x}(t)$ zu einem Zeitpunkt t_n auf einen skalaren Wert ab: $s_n = M(\vec{x}(t_n))$, wobei M eine Meßfunktion ist. Sowohl $\vec{x}(t)$ als auch M sind oft unbekannt. Im allgemeinen werden die Werte $\{s_n\}$ in gleichen Zeitabständen Δ_t gemessen, so daß $t_n = t_0 + n\Delta_t$. Die *Datenaufnahmerate* (*sampling rate*) $1/\Delta_t$ ist die Anzahl der aufgenommenen Werte pro Zeiteinheit. Im Limes $\Delta_t \to 0$ wird die Zeitreihe quasi kontinuierlich.

Von eigentlichem Interesse ist der unbekannte, zugrundeliegende Prozeß $x(t)$. Die Ziele der Zeitreihenanalyse sind die Charakterisierung, Modellierung, Vorhersage, Steuerung, usw. dieses Prozesses. Interessierende Parameter können über die Zeitreihe $\{s_n\}$ berechnet werden. Zwei wesentliche Parameter zur Beschreibung einer Zeitreihe sind der Mittelwert

$$\bar{s} = \frac{1}{N}\sum_{n=1}^{N} s_n = \langle s_n \rangle \tag{1.1}$$

und die Varianz

$$\sigma^2 = \frac{1}{N-1}\sum_{n=1}^{N}(s_n - \bar{s})^2 = \langle (s_n - \bar{s})^2 \rangle, \tag{1.2}$$

wobei durch $\langle \cdot \rangle$ das zeitliche Mittel ausgedrückt wird. Sie beschreiben den durchschnittlichen Wert einer Zeitreihe und deren mittlere Abweichung davon. Der Mittelwert ist ein *Moment erster Ordnung* und die Varianz ein Moment zweiter Ordnung. Sie sind aber lediglich statische Parameter. Sie sagen nichts über die Dynamik der Zeitreihe aus, d.h. sie beschreiben nicht die Beziehung der Werte der Zeitreihe untereinander.

In der Statistik werden die Momente über Erwartungswerte für ein *Ensemble*, d.h. mehrere unabhängige Realisierungen eines Prozesses, geschätzt. So ist also z.B. der Ensemblemittelwert $\mu(t)$ eine Funktion der Zeit. Er ist nur für *stationäre* Prozesse konstant und kann als *Schätzer* für den Mittelwert μ des Prozesses dienen. Unter Stationarität versteht man qualitativ, daß Parameter zeitlich konstant sind. Näher wird auf die Stationarität in Kapitel 3 eingegangen.

In der Zeitreihenanalyse steht aber in der Regel nur *eine* Realisierung eines Prozesses zur Verfügung. Im Gegensatz zu Ensemblemitteln werden *empirische* Momente wie (1.1) und (1.2) über zeitliche Mittel (*Stichprobenmittel*) bestimmt. Die *Ergodizität* garantiert, daß im Limes $N \to \infty$ Stichprobenmittel und Ensemblemittel gleich sind. Sowohl stationäre, chaotische Prozesse als auch stationäre, stochastische Prozesse sind ergodisch, wobei im ersten Fall verschiedene Realisierungen über unterschiedliche Anfangsbedingungen entstehen können. Erst aufgrund der Ergodizität sind auch der Mittelwert (1.1) und die Varianz (1.2) der Zeitreihe Schätzer für die Momente des zugrundeliegenden Prozesses. Im folgenden sind Größen, die über zeitliche Mittel definiert sind, als Schätzer zu verstehen, auch wenn dies nicht explizit angegeben wird.

1.2 Lineare Methoden

Lineare Differentialgleichungen haben stets nur exponentielle oder periodische Lösungen. Die Beschränkung auf stationäre und damit beschränkte Zeitreihen schließt diejenigen exponentiellen Lösungen aus, die über alle Grenzen wachsen. Die exponentiellen Lösungen, die auf einen Wert konvergieren, sind ebenfalls uninteressant für die Beschreibung irregulärer Bewegungen.

Eine Möglichkeit, Irregularität in eine lineare Dynamik einzubauen, besteht in der Verbindung mit stochastischen Prozessen[1]. Durch das Zufallselement geht selbstverständlich der Determinismus verloren. Im einfachsten Fall sind die stochastischen Ereignisse (die Werte der Zeitreihe) unabhängig und unterliegen der gleichen Wahrscheinlichkeitsverteilung. Dies ist ein rein stochastischer, sogenannter *iid-Prozeß* (*independent and identically distributed*) und wird auch als *weißes Rauschen* bezeichnet. Die Unabhängigkeit der Zufallsverteilung von der Zeit ist eine bestimmte Form der Stationarität. Die Unabhängigkeit der Ereignisse selbst deutet aber auf ein Fehlen jeglicher Dynamik hin. Kommt eine zusätzliche (lineare) Dynamik ins Spiel, ergeben sich Korrelationen zwischen Werten zu verschiedenen Zeiten.

1.2.1 Lineare Korrelationen

Der Parameter, der die zeitliche, lineare Abhängigkeit der Werte der Zeitreihe beschreibt, ist die *Autokovarianz*

$$\gamma(\tau) = \frac{1}{N-\tau-1} \sum_{n=1}^{N-\tau} (s_n - \bar{s})(s_{n+\tau} - \bar{s}) = \langle (s_n - \bar{s})(s_{n+\tau} - \bar{s}) \rangle . \qquad (1.3)$$

Dabei ist τ die zeitliche Verzögerung oder das *Lag*. Für $\tau = 0$ ergibt sich die Varianz $\gamma(0) = \sigma^2$. Die Autokovarianz ist ein Moment zweiter Ordnung. Für Analysen wird meist die *Autokorrelation* $\varrho(\tau) = \gamma(\tau)/\gamma(0)$ verwendet, da sie für $\tau = 0$ auf 1 normiert ist. Verschwindet die Autokorrelation für alle $\tau \neq 0$ (d.h. es gilt $\varrho(\tau) = \delta_{\tau,0}$), sind die Werte der Zeitreihe unkorreliert. Wird $\gamma(\tau)$ gegen τ aufgetragen, können leicht die Abhängigkeiten von der Vergangenheit der Zeitreihe abgelesen werden. Diese Darstellung wird auch *Korrelogramm* genannt. Ist die Zeitreihe periodisch mit einer Periode T, zeigt auch die Autokorrelation ein periodisches Verhalten mit derselben Periode. Treten langreichweitige Autokorrelationen auf, d.h. nichtverschwindende Autokorrelationen für τ in der Größenordnung von N, ist die Zeitreihe über den betrachteten Zeitraum nicht stationär.

Jede Zeitreihe kann als Linearkombination aus periodischen Funktionen mit verschiedenen Frequenzen und Amplituden dargestellt werden. Dies geschieht mittels der *Fouriertransformation*, die hier u.a. zum Festlegen der Notation kurz erläutert werden soll. Die Fouriertransformation transformiert eine Zeitreihe in den *Frequenzbereich*, d.h. man verläßt den Gesichtspunkt der Zeitabhängigkeit. Die diskrete Fouriertransformation (DFT) für endliche Zeitreihen $\{s_n\}$ ist gegeben durch

$$\tilde{s}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_n e^{2\pi i k n/N}, \quad k = -N/2, \ldots, N/2 . \qquad (1.4)$$

[1] In Systemen mit unendlich vielen Freiheitsgraden (z.B. Hydrodynamik) gibt es noch die Möglichkeit, unendlich viele Moden anzuregen.

Diese Transformation kann symbolisch als $\{\tilde{s}_k\} = \mathcal{F}\{s_n\}$ geschrieben werden. Sie ist linear und invertierbar. Die inverse diskrete Fouriertransformation (iDFT) ist gegeben durch

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{+N/2} \tilde{s}_k e^{-2\pi i k n/N} . \tag{1.5}$$

In dieser Form ist leicht zu erkennen, daß die Zeitreihe nun als Linearkombination periodischer Funktionen dargestellt ist. Während also die Zeitreihe $\{s_n\}$ beschreibt, was zu einer bestimmten Zeit passiert, geben die komplexen Fourierkomponenten $\{\tilde{s}_k\}$ an, wie stark und mit welcher Phase eine bestimmte Frequenz zum Signal beiträgt. Die Amplitude zur Frequenz $f_k = k/N\Delta_t$ ist der Betrag $A_k = |\tilde{s}_k|$, und die Phase ist gegeben durch $\phi_k = \arctan(\mathrm{Im}(\tilde{s}_k)/\mathrm{Re}(\tilde{s}_k))$.

$|\tilde{s}_k|^2 = A_k^2$ wird das *Periodogramm* genannt. Dieses ist ein Schätzer für das *Frequenz-* oder *Power-Spektrum*, das für die kontinuierliche Zeitreihe $s(t)$ definiert ist. Methoden, mit denen das Frequenzspektrum besser geschätzt werden kann, sind in [22] beschrieben. Durch Analyse des Spektrums können viele Eigenschaften der Zeitreihe bestimmt werden. Scharfe Spitzen im Spektrum deuten auf die Dominanz bestimmter Frequenzen hin. Rauschen in der Zeitreihe führt zu einem Untergrund im Spektrum. Weißes Rauschen enthält alle Frequenzen in gleichem Maße und führt somit zu einem konstanten Untergrund. Ein rein stochastischer Prozeß zeigt im Spektrum keine Struktur, d.h. $\tilde{s}_k$ ist unabhängig von k und konstant bis auf statistische Fluktuationen. Dies deutet darauf hin, daß das Frequenzspektrum in einer bestimmten Weise mit der Autokorrelation verknüpft ist. Tatsächlich besagt das *Wiener-Khinchin-Theorem*, daß

$$\mathcal{F}\{\varrho(n)\} = \{|\tilde{s}_k|^2\} . \tag{1.6}$$

Ist die Zeitreihe reellwertig, so gilt $\tilde{s}_k^* = \tilde{s}_{-k}$ und somit ist $A_k = A_{-k}$ und $\phi_k = -\phi_{-k}$. In diesem Fall sind $\tilde{s}_0$ und $\tilde{s}_{N/2}$ reell, und man erhält N unabhängige, reelle Werte, wie im Fall der Zeitreihe im Zeitbereich.

Gleichung (1.5) ist periodisch in n mit Periode N und kann auch für $n > N$ definiert werden. Die Zeitreihe $\{s_n\}$ wird dadurch periodisch fortgesetzt. Unterschiede zwischen Anfang und Ende der Zeitreihe können also unerwünschte Frequenzkomponenten im Periodogramm erzeugen. Weitere Effekte, die bei der diskreten Fouriertransformation endlicher Zeitreihen auftreten können, werden in [22] behandelt.

1.2.2 ARMA-Modelle

Bisher wurde lediglich auf meßbare Eigenschaften von Zeitreihen eingegangen. Welcher Prozeß die Zeitreihe erzeugt, stand dabei im Hintergrund. Um Vorgänge in Zeitreihen zu verstehen, sind Modelle, die den zugrundeliegenden Prozeß beschreiben, sehr nützlich. Ein solches Modell muß sowohl die linearen Abhängigkeiten als auch die stochastische Natur einer Zeitreihe modellieren können. Ein sehr allgemeines Modell dieser Art ist das *ARMA-Modell*:

$$x_n = \sum_{p=1}^{P} a_p x_{n-p} + \sum_{q=0}^{Q} b_q \eta_{n-q} \,. \tag{1.7}$$

a_p und b_q sind reelle Koeffizienten und die $\{\eta_n\}$ sind gaußverteilte, unabhängige Zufallszahlen mit $\langle \eta_i \eta_j \rangle = \delta_{ij}$. Die erste Summe beschreibt den *autoregressiven* (AR) Aspekt der Zeitreihe, d.h. die Abhängigkeit von der eigenen Vergangenheit. Ist $Q = 0$, so handelt es sich um ein autoregressives Modell von der Ordnung P, kurz AR(P). Der zweite Term ist rein stochastisch und für $P = 0$ spricht man von einem *moving–average–Modell* Q-ter Ordnung, MA(Q). Die zweite Summe ist eine mit b_q gewichtete Mittelung der treibenden Zufallsfolge η_n, wobei jeweils nur über ein verschobenes Fenster der Breite $Q + 1$ gemittelt wird. Gleichung (1.7) wird meist mit ARMA(P,Q) abgekürzt. ARMA-Modelle bilden die Basis für nahezu alle Modellanpassungen in der linearen Zeitreihenanalyse. ARMA-Prozesse sind in den meisten Fällen stationär und gaußverteilt. Die Autokorrelationen und damit das Spektrum können über die Koeffizienten des ARMA-Modells bestimmt werden.

Oft ist es erwünscht, den zukünftigen zeitlichen Verlauf einer Zeitreihe vorhersagen zu können. Eine Möglichkeit in der linearen Zeitreihenanalyse eine Vorhersage zu treffen, besteht darin, ein ARMA-Modell an die Daten anzupassen. Um Vorhersagen im einfachsten Fall eines AR-Modells zu erhalten, müssen dann lediglich die so ermittelten Koeffizienten $\{a_p\}$ und die vorangegangen Werte $\{x_{n-p}\}$ in (1.7) eingesetzt und $\eta_n = 0$ gewählt werden.

Eine umfassende Behandlung der linearen Zeitreihenanalyse und der ARMA-Modelle ist z.B. in [27] zu finden.

1.3 Nichtlineare Zeitreihenanalyse

Mit der linearen Zeitreihenanalyse lassen sich viele irreguläre Zeitreihen erfolgreich bearbeiten. Insbesondere kann die wesentliche Dynamik durch Näherungen oft auf eine lineare reduziert werden, d.h. daß lineare Modelle den Prozeß hinreichend beschreiben. Die Beschränkung auf eine rein lineare Dynamik ist aber oft nicht zufriedenstellend. Das Einbinden von Zufallselementen zur Erklärung von Irregularitäten ist in manchen Fällen durch eine Mittelung über viele unbeobachtete Freiheitsgrade motiviert. In anderen Fällen kann aber das Einführen von Zufallsprozessen erzwungen erscheinen.

Nun ist aber länger bekannt, daß auch *deterministisches Chaos* Signale von komplizierter Struktur erzeugt. Der Determinismus besagt, daß bei genauer Kenntnis eines Zustands zu einer bestimmten Zeit und genauer Kenntnis der Dynamik die Zukunft des Systems bestimmt ist. Trotzdem können starke Irregularitäten auftreten. Die Nichtlinearität der Dynamik ist dabei eine wichtige

Voraussetzung. In der Zeitreihenanalyse müssen neue Methoden erdacht und angewendet werden, die Linearität nicht voraussetzen. Ist über die zugrundeliegende Dynamik aber nichts oder wenig bekannt, kann fast nur vom Determinismus in den Prozessen als solchem Gebrauch gemacht werden.

1.3.1 Deterministische Dynamik

Die Beschränkung auf lineare Prozesse muß aufgehoben werden, um u. a. deterministisches Chaos zu erlauben. Im folgenden werden Prozesse betrachtet, deren Dynamik durch eine Differentialgleichung erster Ordnung

$$\frac{\vec{x}(t)}{dt} = \vec{F}(\vec{x}(t)), \tag{1.8}$$

beziehungsweise im zeitlich diskreten Fall durch eine Abbildung

$$\vec{x}_{n+1} = M(\vec{x}_n) \tag{1.9}$$

beschreibbar sei. $\vec{x}$ ist ein Zustandsvektor aus dem *Phasenraum*. Die Menge der $\{\vec{x}(t)\}$ wird *Orbit* oder *Trajektorie* genannt. Differentialgleichungen höherer Ordnung können stets auf die Form (1.8) gebracht werden.

Die kontinuierliche Darstellung kann diskretisiert werden, indem $\vec{x}(t)$ nur zu bestimmten Zeitpunkten t_n aufgenommen wird. Eine Möglichkeit besteht darin, $t_n = t_0 + n\Delta_t$ zu wählen. Aber auch die Zeitpunkte, in denen $\vec{x}(t)$ eine gewählte Ebene schneidet, können für die Diskretisierung herangezogen werden. Dies wird auch *Poincaré-Schnitt* genannt.

Die Lösungen beider Formen können starke Unregelmäßigkeiten zeigen, d.h. sie zeigen chaotisches Verhalten. Eine Voraussetzung dafür ist die Nichtlinearität von F bzw. M. Eine wesentliche Eigenschaft ist die starke Abhängigkeit vom Anfangszustand. Man spricht von einer chaotischen Dynamik, wenn zwei benachbarte Punkte im Phasenraum sich mit der Zeit im Mittel exponentiell auseinander entwickeln. Der entsprechende Exponent heißt *Lyapunovexponent.*

Der Orbit erreicht für große Zeiten oft eine unter F bzw. M invariante Teilmenge des Phasenraums. Die von einer typischen Anfangsbedingung erreichte invariante Menge heißt *Attraktor.* Ist $\mathcal{A}$ der Attraktor, so gilt $\mathcal{A} = \{F(\vec{x})|\vec{x} \in \mathcal{A}\}$. Attraktoren, die unter chaotischer Dynamik entstehen, können interessante Eigenschaften besitzen, nämlich Selbstähnlichkeit und fraktale Dimension. Es müssen aber nicht alle Punkte des Attraktors gleich oft erreicht werden. Der Bruchteil der Zeit, die ein typischer Orbit in einer Teilmenge S des Attraktors verbringt, ist das *natürliche Maß* $\mu(S)$ von S.

Lineare Methoden können bei nichtlinearen Prozessen zu Fehlinterpretationen und Fehlern führen. So zeigt z.B. das in Abbildung 1.1 dargestellte Spektrum eines nichtlinearen Lasers [1, 41] Strukturen, die auch in Spektren linearer Prozesse auftreten können. Ein ähnliches Spektrum könnte auch ein periodischer Prozeß mit Untergrundrauschen aufweisen.

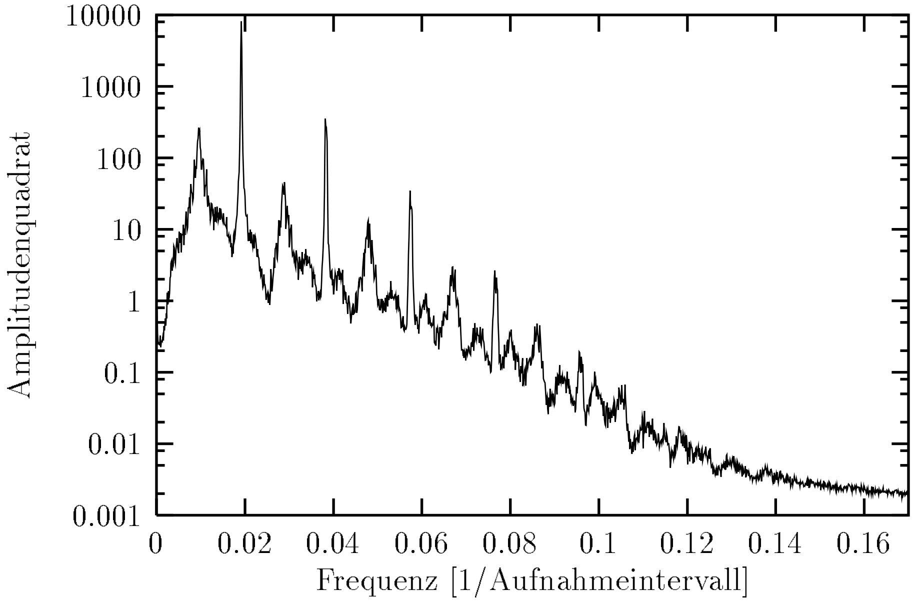

Abbildung 1.1: Spektrum eines nichtlinearen CO_2-Lasers mit Rückkopplung. Die scharfen Spitzen bilden eine Serie von Oberschwingungen.

1.3.2 Einbettung

Zeitreihen sind meistens als skalare Größen gegeben. Sie können aber trotzdem Produkt einer Dynamik sein, die in einem mehrdimensionalen Phasenraum lebt. Die skalare Zeitreihe ist eine irgendwie geartete Projektion der Phasenraumzustände auf die reellen Zahlen. Zur Ermittlung vieler Eigenschaften ist aber eine Repräsentation im Phasenraum unabdingbar. So ist z.B. die eindimensionale Darstellung nicht geeignet, um die Dimension des Attraktors zu bestimmen. Tatsächlich können Phasenraumzustände aus der skalaren Zeitreihe rekonstruiert werden. Die *Zeitverzögerungsrekonstruktion*

$$\vec{s}_n = (s_{n-(m-1)\tau}, \, s_{n-(m-2)\tau}, \, \ldots, \, s_{n-\tau}, \, s_n) \tag{1.10}$$

bettet die Zeitreihe in einen m-dimensionalen Raum ein, wobei τ die zeitliche Verzögerung ist. Der Index n läuft von $(m-1)\tau+1$ bis N, so daß $N-(m-1)\tau$ Einbettungsvektoren konstruiert werden. Es kann unter bestimmten Voraussetzungen gezeigt werden [35, 26], daß für dynamische Systeme der Form (1.8/1.9) eine Repräsentation $\{\vec{s}_n\}$ äquivalent zu dem Originalorbit $\{\vec{x}_n\}$ ist, d.h. eine glatte, invertierbare Abbildung von $\vec{x}_n$ auf $\vec{s}_n$ existiert. Um diese Äquivalenz zu garantieren, stellen die Einbettungstheoreme Bedingungen an die Wahl der Einbettungsdimension m. Offensichtlich muß m größer als die Dimension des Attraktors sein. Es stellt sich heraus, daß m formell auch größer als das zweifache der Boxdimension (s.u.) sein muß. Die Theorie bezieht sich dabei auf unendliche

Zeitreihen, die frei von Rauschen sind. In diesem Grenzfall ist die Zeitverzögerung τ für die Einbettung fast unwesentlich. In der Praxis ist aber auch die Wahl von τ nicht trivial und hängt nicht zuletzt von Δ_t ab. In der Zeitreihe näherungsweise vorhandene Periodizität ist ein guter Orientierungspunkt. Ist man mit einer Zeitreihe konfrontiert, über die keine weiteren Informationen vorliegen, sollten Realisierungen mit verschiedenen m und τ geplottet und inspiziert werden.

1.3.3 Vorhersage

Obwohl für eine nichtlineare Dynamik benachbarte Trajektorien exponentiell auseinanderlaufen können, sind über kurze Zeiten unter Umständen gute Vorhersagen möglich. Auch in der nichtlinearen Zeitreihenanalyse können für die Vorhersage Modelle, die an die Daten gefittet werden, herangezogen werden. Beispiele dafür sind Polynome, Radialbasisfunktionen und neuronale Netze. Diese Modelle verfügen allerdings nicht über einen ähnlichen Stellenwert wie das ARMA-Modell in der linearen Zeitreihenanalyse. Oft können über die zugrundeliegende Dynamik keine Annahmen gemacht werden, was die Entscheidung für ein Modell erschwert. Aber auch ohne Modellanpassung stehen recht gute Vorhersagemethoden zur Verfügung. Diese beruhen allein auf dem angenommenen Determinismus.

Um zu dem Wert s_n der Zeitreihe eine Einschrittvorhersage $F(s_n) \approxeq s_{n+1}$ zu treffen, wird die Zeitreihe wie oben beschrieben, in m Dimensionen eingebettet. Die Punkte im Phasenraum, die in der Nachbarschaft von $\vec{s}_n$ liegen, werden in der ϵ-Umgebung

$$U_n(\epsilon) = \{\vec{s}_i | \, \|\vec{s}_n - \vec{s}_i\| < \epsilon\} \tag{1.11}$$

gesammelt. Als Vorhersage dient der Mittelwert der Folgewerte der Vektoren aus der ϵ-Umgebung:

$$F(s_n) = \frac{1}{\|U_n(\epsilon)\|} \sum_{\vec{s}_i \in U_n(\epsilon)} s_{i+1} \, . \tag{1.12}$$

Dabei ist $\|U_n(\epsilon)\|$ die Anzahl der Vektoren in der ϵ-Umgebung. Man spricht von F als einer lokal konstanten Vorhersage. Im Gegensatz dazu kann eine lokal lineare Vorhersagefunktionen gewählt werden. Als weitere Alternative können die Folgewerte in der Summe (1.12) mit ihrem Abstand zum Punkt $\vec{s}_n$ gewichtet werden. Die Verallgemeinerung auf Mehrschrittvorhersagen ist offensichtlich.

1.3.4 Fraktale Dimension und Korrelationssumme

Eine der wichtigsten Eigenschaften eines Attraktors ist dessen Dimension. Eine mögliche Definition beruht auf der Einteilung des n-dimensionalen Phasenraums in disjunkte, n-dimensionale Würfel der Kantenlänge ϵ. Die Zahl der Würfel, in denen sich eine Teilmenge des Attraktors befindet, sei $N(\epsilon)$. In jedem Würfel wird das invariante Maß μ bestimmt, d.h. der Bruchteil der Zeit, in der sich ein

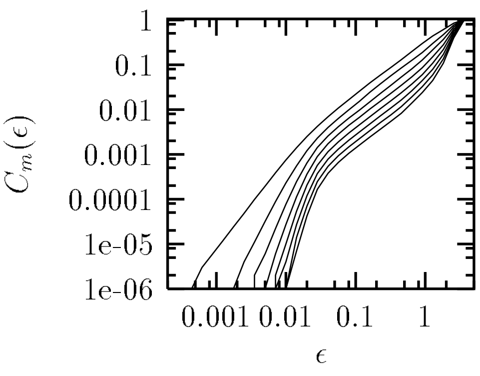

Abbildung 1.2: Korrelationssummen in Abhängigkeit von ϵ für verschiedene Einbettungsdimensionen. Von oben nach unten ist $m = 2, \ldots, 9$.

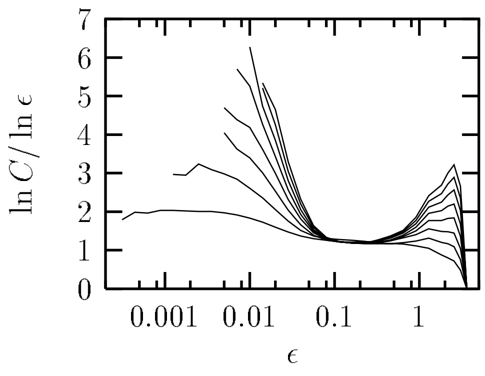

Abbildung 1.3: Die Steigungen der Kurven aus Abbildung 1.2. Deutlich zu sehen ist ein Skalenbereich in dem die Steigung nahezu konstant ist. Von unten nach oben ist $m = 2, \ldots, 9$.

„typischer“ Orbit im Würfel befindet. Das Spektrum der Dimensionen ist dann gegeben durch

$$D_q = \frac{1}{1-q} \lim_{\epsilon \to 0} \frac{\ln \left(\sum_{i=1}^{N(\epsilon)} \mu_i^q \right)}{\ln(1/\epsilon)} . \tag{1.13}$$

Es gilt $D_p \leqslant D_q$ für $p > q$. Im speziellen erhält man für $q = 0$ die *Boxdimension* D_0, die unabhängig vom Maß μ ist, und für $q = 2$ die *Korrelationsdimension* D_2. Im Limes $q \to 1$ erhält man die *Informationsdimension*

$$D_1 = \lim_{\epsilon \to 0} \frac{\sum_{i=1}^{N(\epsilon)} \mu_i \ln \mu_i}{\ln \epsilon} . \tag{1.14}$$

Der Limes zu kleinen ϵ in diesen Dimensionsdefinitionen ist für endliche Zeitreihen nicht bildbar. Die in der Praxis meist angewendete Grassberger-Proccacia-Korrelationsdimension D_2 [8, 9] kann aber alternativ, einfacher und mit weniger Daten über die *Korrelationssumme*

$$C(\epsilon) = \frac{1}{K} \sum_{|i-j| > t_{\min}} \Theta(\epsilon - \|\vec{x}_i - \vec{x}_j\|) \tag{1.15}$$

bestimmt werden. Die Korrelationssumme zählt die Paare von Punkten im Phasenraum, die nicht weiter als ϵ voneinander entfernt sind. Aus praktischen Gründen wird lediglich über Paare summiert, die zeitlich mehr als $t_{\min}$ auseinander liegen. Dies schließt zeitlich korrelierte Paare aus [37]. Die Konstante K wird so gewählt, daß $C(\epsilon)$ für große ϵ eins ergibt.

In der Zeitreihenanalyse werden in der Korrelationssumme (1.15) die Einbettungsvektoren verwendet und $C(\epsilon, m, \tau)$ hängt zusätzlich von der Einbettungsdimension und der Zeitverzögerung ab. Die Korrelationsdimension D_2 kann aus der Steigung von $\ln(C(\epsilon))$ gegen $\ln(\epsilon)$ abgelesen werden. Die Korrelationssummen und deren Steigungen sind in den Abbildungen 1.2 und 1.3 für eine leicht verrauschte Hénon-Abbildung (siehe (2.9)) dargestellt. Der Skalenbereich von ϵ, in dem die Steigungen konstant sind, kann sehr unterschiedlich ausgeprägt sein und ist für verrauschte und kurze Zeitreihen unter Umständen gar nicht vorhanden. Ist ϵ von der Größenordnung der Ausdehnung des Attraktors, geht $C(\epsilon)$ per Konstruktion gegen eins. Bei kleinen ϵ machen sich Rauschen und Fluktuationen bemerkbar. Die Steigung $\ln(C(\epsilon))/\ln(\epsilon)$ strebt dann gegen die Einbettungsdimension, da beim Einbetten Rauschen homogen auf alle Richtungen verteilt wird.

Eine umfassende Einführung in die verschiedenen Formen von Chaos in dynamischen Systemen ist in [21] und [32] zu finden. Die nichtlineare Zeitreihenanalyse wird ausführlich in [14] und [24] behandelt.

Kapitel 2

Tests auf Nichtlinearität

2.1 Einleitung

Zeitlich variierende Prozesse können in nahezu allen Bereichen der Naturwissenschaft beobachtet werden. Um zu einem Verständnis der Vorgänge in solchen Systemen zu gelangen, werden Daten in Form von Zeitreihen aufgenommen. Über den Erzeugungsprozeß der zu analysierenden Daten ist dabei oft nichts oder nur wenig bekannt. Lineare Ansätze sollten in solchen Fällen der erste Schritt sein, denn sie können oft schon zu einem Verständnis der Prozeßnatur führen. Hilfreich ist dabei die Vielzahl an allgemeiner und spezialisierter Literatur und ausgereiften, fertigen Programmpaketen für die lineare Zeitreihenanalyse.

Die Anwendung der Methoden der nichtlinearen Zeitreihenanalyse auf die Daten ist dadurch motiviert, daß die meisten Systeme in der Tat nichtlinear sind. Komplexe Vorgänge sind oft nicht exakt durch lineare Gleichungen beschreibbar. Viele Daten fallen in diese Kategorie. So wird beispielsweise für einen Großteil der physiologischen Daten Nichtlinearität in der Dynamik allgemein angenommen. In solchen Fällen stellt sich allerdings die Frage, ob die Dynamik nicht durch lineare Prozesse angenähert werden kann. Denn dann kann auf das große Gerüst der linearen Zeitreihenanalyse zurückgegriffen werden.

Auf jedem Fall ist es wünschenswert, einen Test zu besitzen, der aussagt, ob sichtbare Nichtlinearitäten in den Daten vorhanden sind. Der Einsatz von nichtlinearen Analysen kann nur dann sinnvoll sein, wenn ein solcher Test erfolgreich Nichtlinearität findet. Gemessene Zeitreihen sind aber selten ein Produkt einer rein linearen oder nichtlinearen Dynamik. Sie sind vielmehr oft eine Mischung aus linearen und nichtlinearen Prozessen. Hinzu kommt fast immer noch mehr oder weniger starkes Rauschen. Ein Auffinden von Nichtlinearität ist also nicht immer einfach. Ziel dieses Kapitels ist es deshalb einen möglichst effektiven und zuverlässigen Test auf Nichtlinearität zu entwickeln.

2.2 Tests in der Zeitreihenanalyse

Tests können sowohl die Verifikation von angenommenen Eigenschaften als auch die Suche nach weiteren, neuen Eigenschaften zum Ziel haben. Die Untersuchung der Eigenschaften kann mit der Formulierung einer Frage beginnen, auf die ein Test eine Antwort geben soll. Typische Fragen sind z.B.:

- Was ist die zugrundeliegende Dynamik?
- Welcher Verteilung unterliegen die Werte der Zeitreihe?
- Sind die Werte korreliert?
- Zeigt die Zeitreihe chaotisches Verhalten?

Insbesondere können *Entscheidungsfragen* formuliert werden, auf die nur mit „ja“ oder „nein“ geantwortet werden kann. Jede Entscheidungsfrage filtert eine bestimmte Eigenschaften der Zeitreihe heraus. Komplexere Fragen können aus mehreren Entscheidungsfragen kombiniert werden.

Zu jeder Entscheidungsfrage wird eine *Nullhypothese* aufgestellt. Die Nullhypothese besagt, daß die Antwort auf die Frage „nein“ ist. Der Test soll diese Hypothese überprüfen. Für das obige dritte Beispiel lautet die Nullhypothese: „Die Werte der Zeitreihe sind unkorreliert.“ Ein Test kann aber nie die Nullhypothese wirklich beweisen oder widerlegen. Der Test kann immer nur mit einer bestimmten Wahrscheinlichkeit die Nullhypothese für die betrachteten Daten verwerfen. Die Nullhypothese muß präzise formuliert werden, damit der Test eine eindeutige Aussage liefert. Außerdem muß sichergestellt werden, daß der Test auch wirklich genau auf die Nullhypothese testet.

Um Entscheidungen über eine bestimmte Eigenschaft treffen zu können, braucht man ein quantitatives Maß für die betreffende Eigenschaft. Dieses Maß ist in der Regel eine *Observable* $T = T(\{s_n\})$, die die Zeitreihe $\{s_n\}$ auf eine reelle Zahl abbildet. Für Tests wird T auch *Teststatistik* genannt. Im einfachsten Fall besteht ein Test aus einer Berechnung von T und einer Kontrolle, ob T in einer für die Nullhypothese charakteristischen Teilmenge von $\mathbb{R}$ liegt. Diese Teilmenge kann z.B. ein bestimmtes Intervall $[T_{\min}, T_{\max}]$ oder ein Punkt T_0 sein.

Soll z.B. auf die Korrelation der Werte der Zeitreihe getestet werden, kann als Teststatistik die Autokorrelation $T = \varrho(\tau = 1)$ gewählt werden, wobei die Nullhypothese für $T = 0$ erfüllt ist. Für $T \neq 0$ wird die Nullhypothese verworfen. Für endliche Zeitreihen unterliegt aber der Wert der Autokorrelation statistischen Fluktuationen, so daß in der Praxis auch für unkorrelierte Daten $\varrho(\tau)$ von 0 abweicht. So liefert z.B. die Varianz des Schätzers der Autokorrelation (1.3) für weißes Rauschen $1/N + \mathcal{O}(1/N^2)$ für alle $\tau \neq 0$ [27]. Die obige Nullhypothese ist also nur im Limes für $N \to 0$ korrekt. Wird sie für endliche Zeitreihen verwendet, liefert der Test falsche Aussagen.

Die Autokorrelation ist eine sehr einfache Teststatistik, für die Erwartungswert und Varianz auch für allgemeine endliche Zeitreihen berechnet werden können. D.h. die Verteilung der Werte T für verschiedene Zeitreihen, die die Nullhypothese erfüllen, ist annähernd bekannt. Ist diese Verteilung z.B. eine Gaußverteilung, so kann einfach anhand des Abstandes vom Mittelwert von $T(\{s_n\})$ die Wahrscheinlichkeit für die Richtigkeit der Nullhypothese angegeben werden. Im allgemeinen kann aber über die Verteilung der T keine Aussage gemacht werden. Nur in wenigen Fällen kann sie unter speziellen Bedingungen analytisch berechnet werden. In Abschnitt 2.2.2 wird gezeigt, wie die Verteilung anhand von vielen Datensätzen, die die Nullhypothese erfüllen, numerisch approximativ ermittelt werden kann.

2.2.1 Maße für die Güte eines Tests: Size und Trennschärfe

Die Leistungsfähigkeit und Qualität von statistischen Tests wird durch zwei wichtige Parameter beschrieben. Diese sind bestimmt durch die zwei Arten von Fehlern, die bei einem Test mit einem zu untersuchenden Datensatz auftreten können.

Zunächst besteht die Möglichkeit, daß der Test die Nullhypothese verwirft, obwohl sie in Wirklichkeit wahr ist. Dies wird „Fehler vom Typ I“ genannt. Die Wahrscheinlichkeit für einen solchen Fehler ist die *Size* α des Tests. Wird ein Signifikanzlevel von $1-p$ vorgegeben, sollte die Size p nicht überschreiten. Für gewöhnlich wird p, die *nominelle* Size[1], vorher festgesetzt und der Test entsprechend realisiert. Es muß gewährleistet werden, daß die wirkliche Size α des Tests mit der nominellen übereinstimmt. Ist die Size zu groß, verliert der Test seine Aussagefähigkeit. Formell ist eine Size kleiner als p zwar zulässig, führt aber zu einer unerwünschten Senkung der Leistungsfähigkeit des Tests. Es sollte also vor jeder Durchführung eines Tests die Size ermittelt werden. Dies geschieht indem möglichst viele Tests mit Daten, die die Nullhypothese erfüllen, durchgeführt werden. Der Quotient aus der Anzahl der Verwerfungen der Nullhypothese und der Anzahl der Tests ist die ermittelte Size des Tests. Stimmt sie nicht mit der nominellen überein, läßt das auf eine unzureichende Konstruktion des Tests schließen.

Ein „Fehler vom Typ II“ ist das Akzeptieren einer Nullhypothese, die in Wirklichkeit falsch ist. Die Wahrscheinlichkeit dafür wird mit β bezeichnet und $1-\beta$ ist die *Trennschärfe* oder *Power* des Tests. Die Trennschärfe gibt an, wie stark bzw. wie wahrscheinlich die Nullhypothese für die zu untersuchende Zeitreihe mit einem bestimmten Test abgelehnt werden kann, d.h. eine möglichst hohe Trennschärfe ist wünschenswert. Verschiedene Realisierungen eines Tests können unterschiedliche Trennschärfen aufweisen. Die Trennschärfe kann als Maß für die Güte der einzelnen Realisierungen betrachtet werden. Ein wichtiger Punkt bei der Durchführung eines Tests ist die gewählte Teststatistik. Wird sie ungünstig

[1] Zur Unterscheidung wird die Size α auch *empirische* Size genannt.

gewählt, kann das zu einer Senkung der Trennschärfe führen. In diesem Kapitel werden verschiedene Teststatistiken auf ihre Trennschärfe hin untersucht. Es ist die Teststatistik zu bevorzugen, die in möglichst vielen Fällen die maximale Trennschärfe liefert.

2.2.2 Tests mit Surrogaten

Monte-Carlo-Methoden werden in der Physik häufig eingesetzt. Mit Ihnen können numerische Ergebnisse erzielt werden, die analytisch nur schwer oder gar nicht zu erhalten sind. Ihnen gemeinsam ist die zufällige, einer wohldefinierten Wahrscheinlichkeitsverteilung unterliegende Auswahl von Konfigurationen oder Prozessen. An diesen Konfigurationen können Observablen numerisch gemessen werden. Bei einer endlichen, aber sehr großen Anzahl von Konfigurationen kann so ein „typisches", kleineres Ensemble herausgegriffen werden und nur anhand dieser die Erwartungswerte der Observablen ermittelt werden.

Die Approximation der Verteilung der Teststatistik T für Datensätze, die die Nullhypothese erfüllen, geschieht ebenfalls unter Zuhilfenahme einer Monte-Carlo-Methode. Es ist dafür notwendig, möglichst viele zufällige Zeitreihen zu erzeugen, die die Nullhypothese erfüllen. Diese „künstlichen" Zeitreihen werden auch *Surrogate* genannt [39]. Die Erzeugung der Surrogate hängt in starkem Maß von der Nullhypothese ab und ist nicht trivial. Die Surrogate dürfen die Eigenschaft, auf die hin der Originaldatensatz untersucht werden soll, nicht aufweisen. Alle anderen Eigenschaften der Originaldaten sollten aber auch in den Surrogaten zu finden sein, was bei der Erzeugung von Surrogaten berücksichtigt werden muß.

Dabei unterscheidet man zwei verschiedene Ansätze. Zum einen kann das Surrogat durch einen Prozeß erzeugt werden, der die Nullhypothese erfüllt und aus der Originalzeitreihe approximierte Parameter verwendet. Dies wird eine *typische Realisierung* genannt, da die Eigenschaften von Surrogaten und Originaldaten nicht exakt übereinstimmen müssen. So haben z.B. zwei ARMA-Prozesse mit exakt gleichen Parametern nicht exakt gleiche Stichproben-Mittelwerte und Stichproben-Varianzen. Im Gegensatz dazu werden bei der *bedingten Realisierung* bestimmte Eigenschaften der Originaldaten exakt auf den Surrogaten reproduziert. So kann z.B. die Stichproben-Mittelwerte oder die Stichproben-Varianzen stets durch Umskalierung in Übereinstimmung gebracht werden. In [40] wird anhand einiger Beispiele sehr deutlich gezeigt, daß bedingte Realisierungen vorzuziehen sind, da sie im Gegensatz zu typischen Realisierungen korrekte Size und höhere Trennschärfe liefern.

In dem obigen Beispiel eines Tests auf Korrelation der Werte einer Zeitreihe ist die Erzeugung von Surrogaten relativ einfach. Ein Surrogat kann hier durch mehrfache, zufällige Vertauschung der Werte der Zeitreihe erzeugt werden. Dies zerstört eine eventuell im Originaldatensatz vorhandene Korrelation. Andere Eigenschaften, z.B. die Verteilung der Werte, bleiben dabei erhalten. Für andere

Tests kann die Surrogaterzeugung wesentlich schwieriger sein.

Der Test (Surrogattest) wird nun folgendermaßen durchgeführt. Es werden zunächst B Surrogate erzeugt. Auf allen Surrogaten wird dann die Teststatistik $T_b, b = 1, \ldots, B$ berechnet. Die Verteilung der $\{T_b\}$ folgt bei korrekter Erzeugung der Surrogate annähernd der Verteilungsfunktion von T. Die gleiche Teststatistik wird auf der Originalzeitreihe $T_0 = T(\{s_n\})$ berechnet. Liegt T_0 nun im oberen oder unteren Bereich der Werte $\{T_b\}$, d.h. im „Schwanz" der empirischen Verteilung, so wird die Nullhypothese verworfen. Solch ein Vorgehen wird *zweiseitiger* Test genannt, da die Nullhypothese sowohl für große als auch für kleine T_0 verworfen wird.

Präziser formuliert, wird die Nullhypothese verworfen, wenn T_0 in der nach Größe sortierten Liste $\{T_b,\ b = 0, \ldots, B\}$ zu den $(B+1)\alpha/2$ kleinsten oder zu den $(B+1)\alpha/2$ größten Werten gehört. Dieses Verfahren ist so konstruiert, daß die Wahrscheinlichkeit für einen „Fehler vom Typ I" wie erwünscht durch die Size α gegeben ist. Minimal werden für einen Test $B = 2/\alpha - 1$ Surrogate benötigt. In diesem Fall wird die Nullhypothese verworfen, wenn T_0 der kleinste oder der größte der $\{T_b\}$ ist. Soll der Test eine kleinere Size α und damit ein höheres Signifikanzniveau aufweisen, werden mehr Surrogate benötigt.

Wird erwartet, daß die Teststatistik für Daten, die die Nullhypothese nicht erfüllen, stets im unteren oder stets im oberen Bereich der $\{T_b\}$ liegt, spricht man von einem *einseitigen* Test. In diesem Fall reichen $B = 1/\alpha - 1$ Surrogate für einen Test aus.

Für gewöhnlich wird $B = 2n/\alpha - 1$ für einen zweiseitigen Test und $B = n/\alpha - 1$ für einen einseitigen Test gewählt, wobei $n \in \mathbb{N}$. Der Grund, mehr Surrogate als minimal nötig für einen Test zu verwenden, ist eine leichte Erhöhung der Trennschärfe für große B [19]. Die Size α hingegen ist unabhängig von der Anzahl der Surrogate.

In einigen Fällen wird bei einem Surrogattest eine etwas andere Vorgehensweise gewählt. So können für die Werte $\{T_b\}$ der Surrogate Mittelwert, Varianz und andere Parameter berechnet werden. Liegt T_0 dann z.B. zwei Standardabweichungen vom Mittelwert entfernt, kann die Nullhypothese mit 95% abgelehnt werden. Dieses Verfahren hat aber zwei Nachteile. Zum einen werden relativ viele Surrogate benötigt, um die Parameter genügend genau abschätzen zu können. Gravierender ist aber die Tatsache, daß die Verteilungsfunktion vorher bekannt sein muß. Die Aussage des obigen Beispiels ist nur korrekt, wenn die Teststatistik für die Nullhypothese gaußverteilt ist. Falsche Annahmen über die Verteilungsfunktion führen sehr wahrscheinlich zu einem falschen Testergebnis.

Für einige der Teststatistiken aus Abschnitt 2.5.1 sah die Wahrscheinlichkeitsverteilung auf den ersten Blick tatsächlich wie eine Gaußverteilung aus. Ein Kolmogorov-Smirnov-Test ergab aber nur eine 10%-Wahrscheinlichkeit für Gaußverteilung der Werte. Ich werde deshalb im folgenden ausschließlich das nichtparametrische Verfahren verwenden.

2.3 Erzeugen von Surrogaten

Bei der Erzeugung von Surrogaten für Tests auf Nichtlinearität ist zu beachten, daß ihre linearen Eigenschaften mit denen des Originaldatensatzes übereinstimmen. Die Autokorrelation der Originaldaten sollte also auch in den Surrogaten zu finden sein. Damit sind die Momente zweiter Ordnung der Surrogate vorgegeben. Eine weitere wichtige Eigenschaft ist die Verteilungsfunktion der Originaldaten. Sie sollte in den Surrogaten möglichst genau nachmodelliert werden. Im Extremfall heißt das, daß das Surrogat die gleichen Werte wie die Originalzeitreihe annimmt und nur die zeitliche Reihenfolge verändert ist.

Können Surrogate mit gleicher Autokorrelation und Verteilungsfunktion wie die zu untersuchende Zeitreihe konstruiert werden, lautet die entsprechende Nullhypothese: „Die Daten $\{s_n\}$ sind das Ergebnis eines Gauß'schen, linearen Prozesses x_n, der mit einer monotonen Meßfunktion $m(x_n) = s_n$ gemessen wurde.“ Ein Verwerfen dieser Nullhypothese ist natürlich kein Beweis für deterministisches Chaos in den Daten.

2.3.1 ARMA-Modell-Anpassung

Eine erste Idee, Surrogate für Nichtlinearitätstests zu erzeugen, ist, ein ARMA-Modell (1.7) an die Daten $\{s_n\}$, $n = 1, \ldots, N$ zu fitten. Die AR-Koeffizienten a_j können dabei durch Lösen des linearen Gleichungssystems

$$\sum_{j=1}^{J} C_{ij} a_j = b_i \quad \text{mit } i = 1, \ldots, J \tag{2.1}$$

ermittelt werden, wobei $C_{ij} = \gamma(i-j) = \langle (s_{n+i} - \bar{s})(s_{n+j} - \bar{s}) \rangle$ die quadratische und symmetrische Kovarianzmatrix und $b_i = \gamma(i)$ die Autokovarianz mit Zeitverzögerung i ist. Beide werden nur bis zu einer maximalen Zeitverzögerung von J berechnet. Einerseits ist es wünschenswert die Autokovarianzen mit möglichst vielen Zeitverzögerungen mitzuberücksichtigen, andererseits kommt es bei Zeitverzögerungen nahe N zu großen statistischen Schwankungen, da nur noch über wenige Werte gemittelt wird. Bei dieser Ermittlung der AR-Koeffizienten wird davon ausgegangen, daß der stochastische Anteil durch Mittelung wegfällt.

Ein AR(J)-Prozeß mit den ensprechenden Koeffizienten a_j erzeugt dann ein Surrogat. Die Surrogate unterscheiden sich durch den Einfluß der gaußverteilten Zufallszahlen η_n. Dies ist ein Beispiel für eine typische Realisierung. Die Autokorrelationen von Originaldaten und den Surrogaten stimmen nicht exakt, sondern nur bis auf Fluktuationen überein. Da die Werte der Surrogate gaußverteilt sind, können die Verteilungsfunktionen von Originaldaten und Surrogaten aber sehr stark voneinander abweichen.

2.3.2 FT-Surrogate

Das Wiener-Khinchin Theorem (1.6) besagt, daß zwei Datensätze mit gleicher Autokorrelation auch gleiches Frequenzspektrum besitzen und umgekehrt. Um Surrogate mit gleicher Autokorrelation wie die Originaldaten zu erzeugen, kann man also Surrogate mit gleichem Frequenzspektrum konstruieren. Folgendes Schema erzeugt solche Surrogate:

1. Die Daten $\{s_n\}$ werden mit einer diskreten Fouriertransformation (1.4) in den Frequenzbereich transformiert. Man erhält Amplituden $\{A_k = |\tilde{s}_k|\}$ und Phasen $\{\phi_k\}$.

2. Die Phasen $\{\phi_k\}$ werden durch gleichverteilte Zufallszahlen zwischen 0 und 2π ersetzt.

3. Eine inverse Fouriertransformation wird durchgeführt und man erhält $\{s'_n\}$.

Per Konstruktion hat das Surrogat $\{s'_n\}$ das gleiche Frequenzspektrum wie die Originaldaten. Es handelt sich also hier um eine bedingte Realisierung im Sinne der Autokorrelation. Die Verteilungsfunktion des Surrogates kann aber auch hier von der der Originaldaten abweichen. Sind die Werte des zu untersuchenden Datensatzes allerdings gaußverteilt, werden für große N auch die Werte des Surrogates der Gaußverteilung unterliegen. Diese Tatsache macht man sich in der folgenden Methode zur Erzeugung von Surrogaten zunutze.

2.3.3 AAFT-Surrogate

Die in der Literatur am häufigsten angewendete Methode zur Erzeugung von Surrogaten ist die *Amplitude Adjusted Fourier Transformation* (AAFT). Hierbei werden die Daten zunächst in die Gaußverteilung überführt, fouriertransformiert und danach wieder auf die Ausgangsverteilung gebracht. Im Detail sieht das wie folgt aus:

1. Es werden N gaußverteilte Zufallszahlen $\{\eta_n\}$ erzeugt.

2. Die Originaldaten $\{s_n\}$ werden jetzt auf die gaußverteilten Werte umskaliert. Dabei wird der kleinste Wert der $\{s_n\}$ durch den kleinsten Wert der $\{\eta_n\}$ ersetzt. An die Stelle des zweitkleinsten der $\{s_n\}$ tritt der zweitkleinste der $\{\eta_n\}$, usw. Bis am Ende der größte Wert der $\{s_n\}$ durch den größten der $\{\eta_n\}$ ersetzt wird. Diese Prozedur heißt *Reskalierung über die Rangordnung.* Die so entstandenen Daten $\{g_n\}$ sind gaußverteilt und folgen der zeitlichen Entwicklung der Originaldaten.

3. Die $\{g_n\}$ werden fouriertransformiert, die resultierenden Phasen durch gleichverteilte Zufallszahlen ersetzt und zurücktransformiert. Man erhält ebenfalls gaußverteilte $\{g'_n\}$.

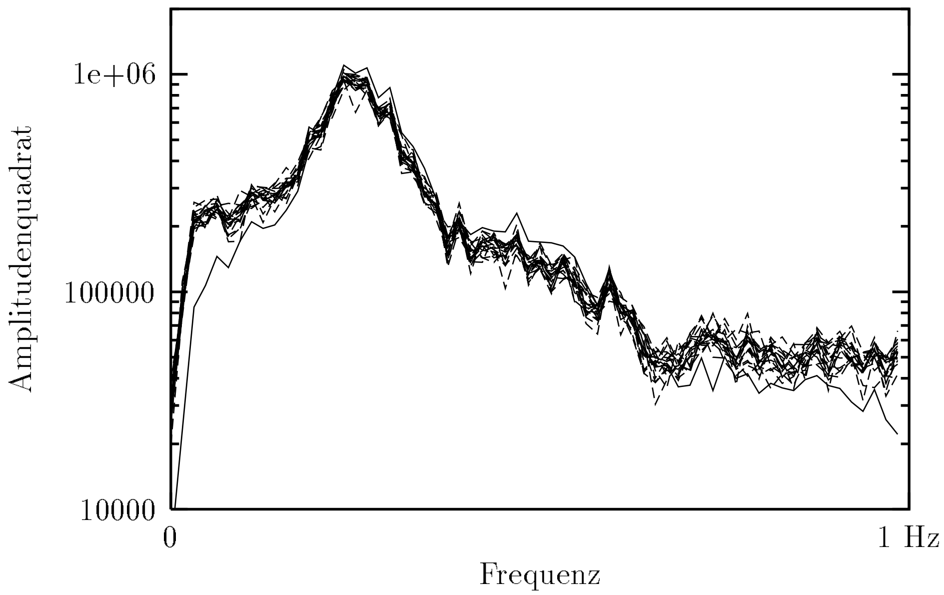

Abbildung 2.1: Abweichung der Spektren für Atmungsdaten einer Person (durchgezogen) und 19 AAFT-Surrogaten (gestrichelt)

4. Die $\{g'_n\}$ werden über die Rangordnung zurückskaliert auf die Verteilung der Originaldaten $\{s_n\}$. Das resultierende Surrogat $\{s'_n\}$ hat per Konstruktion die gleiche Verteilungsfunktion und annähernd das gleiche Frequenzspektrum wie die Originaldaten.

Das Spektrum stimmt nicht exakt überein, da die Reskalierung das Spektrum für gewöhnlich verzerrt. In Abbildung 2.1 wird deutlich, daß es systematische Abweichungen gibt.

2.3.4 Neues verbessertes Verfahren

Im folgenden wird ein neues Verfahren zur Erzeugung von Surrogaten für Nichtlinearitätstests vorgestellt [29]. Es erhält wie die AAFT-Methode die Verteilungsfunktion exakt. Die Diskrepanz zwischen den Spektren der Surrogate und den Originaldaten ist allerdings speziell bei langen Datensätzen wesentlich geringer als beim AAFT-Verfahren.

Der Algorithmus besteht aus einem einfachen Iterationsschema mit zwei abwechselnden Schritten.

1. Zunächst muß eine in ansteigender Reihenfolge sortierte Version $\{\bar{s}_n\}$ der Zeitreihe $\{s_n\}$ abgespeichert werden. Es gilt also $\bar{s}_n \geqslant \bar{s}_m$ für $n > m$. Ebenfalls abgespeichert werden sollten die Frequenzamplituden $\{A_k = |\tilde{s}_k|\}$ der Zeitreihe.

2. Um einen zufälligen Anfangsdatensatz $\{s_n^{(0)}\}$ für die folgende Iteration zu bekommen, werden die Werte der Zeitreihe zufällig permutiert. Dies verändert trivialerweise die Verteilung nicht.

3. Als ersten Schritt der Iteration i wird $\{s_n^{(i)}\}$ fouriertransformiert. Die resultierenden Amplituden werden verworfen und durch das Spektrum $\{A_k\}$ der Originalzeitreihe ersetzt. Die Phasen $\{\phi_k\}$ bleiben unverändert. Nun wird zurücktransformiert und man erhält den Datensatz $\{\hat{s}_n^{(i)}\}$. Dieser hat das gleiche Spektrum wie die Originaldaten, aber die Verteilung ist in der Regel zerstört.

4. Im zweiten Schritt der Iteration wird die Verteilung angepaßt. Dies geschieht wieder über eine Reskalierung über die Rangordnung. Dabei nutzt man die abgespeicherte, sortierte Zeitreihe $\{\bar{s}_n\}$. Für das Ergebnis $\{\hat{s}_n^{(i)}\}$ des ersten Schrittes wird eine Indextabelle $\{I(n)\}, n = 1, \ldots, N$ aufgestellt. Hat $\hat{s}_n^{(i)}$ den Rang m, so ist $I(n) = m$. Der kleinste Wert bekommt also $I(n) = 1$ und der größte $I(n) = N$. Mit dieser Indextabelle läßt sich die Reskalierung mittels $s_n^{(i)} = \bar{s}_{I(n)}$ durchführen. Durch diesen Schritt ist die Verteilung wieder den Originaldaten angepaßt, wobei das Spektrum leicht abgewandelt wurde.

Die beiden Iterationsschritte 3 und 4 werden abwechselnd wiederholt. Das Spektrum des Surrogates paßt sich dabei immer mehr dem Spektrum der Originaldaten an. Für endliche N ist eine Konvergenz gegen das Originalspektrum nicht immer zu erwarten, da die Änderung des Spektrums im ersten Iterationsschritt zu klein sein kann, um im zweiten Schritt eine Umordnung der Werte zu bewirken. Das Iterationsschema konvergiert in so einem Fall nicht exakt gegen das gewünschte Spektrum.

2.4 Eigenschaften des neuen Verfahrens

2.4.1 Konvergenz

Um die Konvergenz dieses Verfahrens untersuchen zu können, wird nach jeder Iteration die Abweichung des Spektrums vom Original ermittelt. Dafür muß die Abweichung quantitativ ausgedrückt werden. Ich verwende hier als Größe für die relative Abweichung des Spektrums

$$\Delta\tilde{S}_i = \frac{\sum_{k=-N/2}^{+N/2} (\tilde{s}'^{(i)}_k - \tilde{s}_k)^2}{\sum_{k=-N/2}^{+N/2} \tilde{s}_k^2} . \tag{2.2}$$

Darin ist $\tilde{s}'^{(i)}_k$ die Fouriertransformierte des Surrogates nach der Iteration i und $\tilde{s}_k$ die Fouriertransformierte der Originaldaten. Um Fluktuation in den einzel-

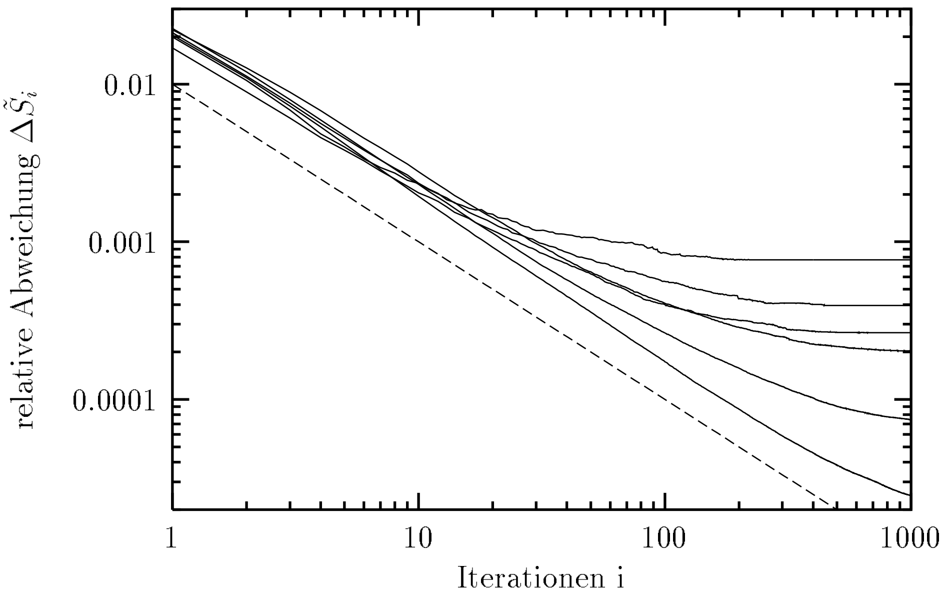

Abbildung 2.2: Konvergenz der Spektren der Surrogate für einen AR(1)-Prozeß. Von oben nach unten wurden Zeitreihen der Länge $N = 1024, 2048, \ldots, 32768$ verwendet. Die gestrichelte Linie ist $\propto 1/i$.

nen Frequenzen des Spektrums zu unterdrücken, wurden beide Spektren zusätzlich geglättet, indem über 21 benachbarte Frequenzen gemittelt wurde, $\tilde{s}_k^2 = \sum_{j=k-10}^{k+10} \tilde{s}_j^2$.

Zur Veranschaulichung zeigt Abbildung 2.2 die Konvergenz des Iterationsschemas für Zeitreihen mit unterschiedlichen Längen N. Verwendet wurden Daten aus einem AR(1)-Prozeß $x_n = 0.7x_{n-1} + \eta_n$ mit der Meßfunktion $s_n = x_n^3$. Für jedes $N = 1024, 2048, \ldots, 32768$ wurden Zeitreihen $\{s_n\}$ und Surrogate $\{s'_n\}$ berechnet. Die größte Änderung des Spektrums findet in der ersten Iteration statt, da in ihr das weiße Spektrum des permutierten Datensatzes auf das gewünschte Spektrum der Originaldaten gebracht wird. Vor der ersten Iteration ist $\Delta\tilde{S}_0 = 0.2 \pm 0.01$, was wesentlich über den Werten für $\Delta\tilde{S}_i$ mit $i \geqslant 1$ in Abbildung 2.2 liegt. Nach weiteren Iterationen gleicht sich das Spektrum der Surrogate weiter dem Spektrum der Originaldaten an, bis schließlich ein von N abhängender Sättigungswert erreicht wird. In Abbildung 2.2 ist zu sehen, daß die relative Abweichung dabei ungefähr proportional zu $1/i$ ist. Ist man ab einer bestimmten Genauigkeit für die Spektren der Surrogate zufrieden, kann das Iterationsschema an der entsprechenden Stelle abgebrochen werden. Der Sättigungswert der Spektrumsabweichung, der nach einer bestimmten Anzahl von Iterationen erreicht wird, ist wie erwartet für lange Zeitreihen kleiner als für kurze Zeitreihen. Grob kann dieses Verhalten für den obigen Prozeß durch eine $1/\sqrt{N}$-Abhängigkeit beschrieben werden, was in Abbildung 2.3 verdeutlicht wird. Im Vergleich dazu ist die Abweichung

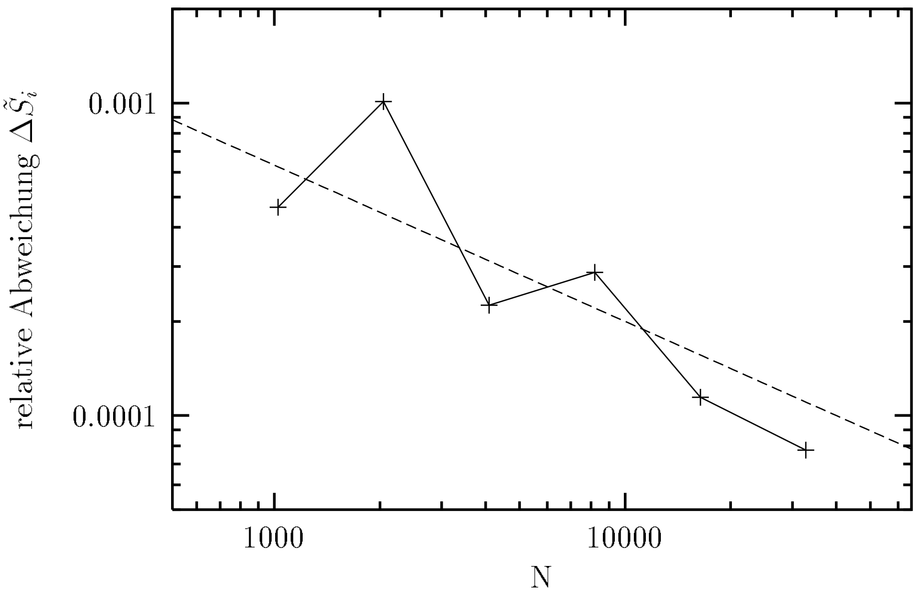

Abbildung 2.3: Die Sättigungswerte der Kurven aus Abbildung 2.2 in Abhängigkeit von N. Die gestrichelte Linie ist $\propto 1/\sqrt{N}$.

der Spektren von AAFT-Surrogaten zum Spektrum der Originaldaten für alle betrachteten N größer als $\Delta\tilde{S} = 0.015$. Ähnliches Verhalten zeigt sich auch für andere linear korrelierte Prozesse. Es kann aber nicht als allgemein gültig angesehen werden. Für Daten, die mit einer diskreten Mackey-Glass-Gleichung [17] erzeugt wurden, ergab sich z.B. eine exponentielle Konvergenz proportional zu $\exp(-0.4i)$. Die entsprechenden Sättigungswerte der Spektrumsabweichung zeigten annähernd eine $1/N^{\frac{3}{2}}$- Abhängigkeit.

Allen betrachteten Fällen gemeinsam ist aber eine schnelle Konvergenz, die mit zunehmender Länge der Zeitreihe kleinere Abweichungen vom Originalspektrum zeigt. Das genaue Verhalten der Konvergenz scheint aber sowohl von der Verteilung der Daten als auch von den linearen Korrelationen abzuhängen und ist nicht pauschal erklärbar.

2.4.2 Size im Vergleich zu AAFT-Surrogaten

Bei der Definition der Size in Abschnitt 2.2.1 wurde bereits kurz erwähnt, daß die aktuelle Size eines Tests bei ungeeigneten Surrogaten von der nominellen abweichen kann. Um die Vorteile des neuen Verfahrens zur Surrogaterzeugung für Nichtlinearitätstests zu zeigen, wähle ich im folgenden ein Beispiel, bei dem ein Test mit AAFT-Surrogaten tatsächlich eine zu große Size liefert und damit unbrauchbar wird. Der verwendete Datensatz basiert wieder auf einem AR(1)-Prozeß $x_n = 0.95x_{n-1} + \eta_n$ und einer monotonen Meßfunktion $s_n = x_n\sqrt{|x_n|}$,

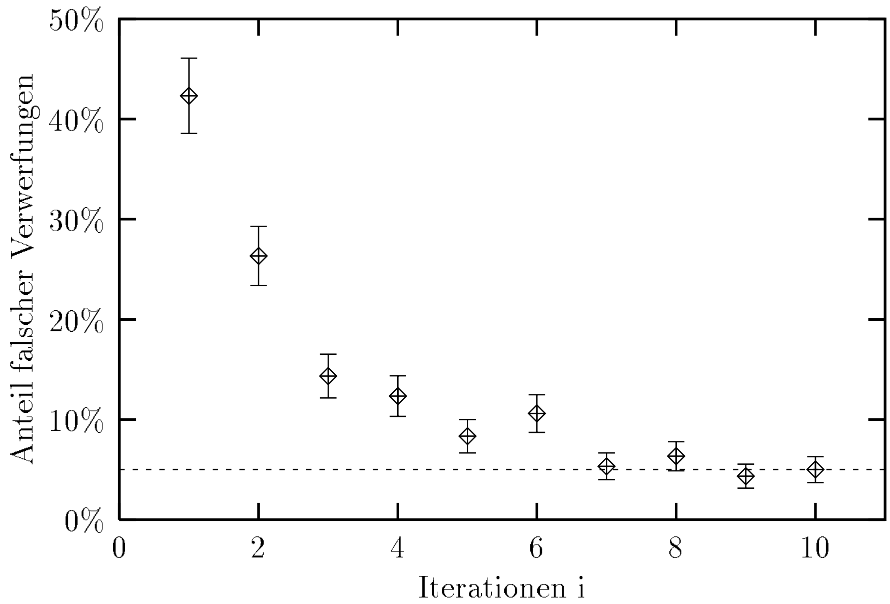

Abbildung 2.4: Anteil der falschen Verwerfungen für 300 Tests in Abhängigkeit von der Anzahl der Iterationen, die bei der Surrogaterzeugung durchgeführt wurden. Die nominelle Size beträgt 5%. Tests mit AAFT-Surrogaten führten zu 66% falschen Verwerfungen.

erfüllt also die Nullhypothese. Es wurden Zeitreihen mit $N = 2048$ erzeugt und in $m = 2$ Dimensionen eingebettet. Für jeden Test wurden jeweils 19 Surrogate erzeugt und als Teststatistik diente ein Vorhersagefehler, welcher im nächsten Abschnitt detailiert beschrieben wird (siehe Gleichung (2.5)). Die Nullhypothese wurde auf einem Signifikanzniveau von 95% verworfen, wenn der Vorhersagefehler für die Originalzeitreihe kleiner war als für alle 19 Surrogate. Um die Size zu ermitteln wurden 300 unabhängige Tests durchgeführt und die Anzahl der falschen Verwerfungen gezählt. Anstelle der erwarteten 5% zeigten die Tests mit AAFT-Surrogaten einen Anteil von $66\% \pm 5\%$ falschen Verwerfungen. Um die Abhängigkeit der Size von der Anzahl der Iterationen im neuen Verfahren darzustellen, habe ich bei der Surrogaterzeugung immer nur eine bestimmte Anzahl von Iterationen durchgeführt. Das Ergebnis ist in Abbildung 2.4 dargestellt. Die richtige Size von 5% wird in diesem Fall für Surrogate erreicht, die mit sieben oder mehr Iterationen erzeugt wurden. Das sind wesentlich weniger Iterationen, als für die Konvergenz des Spektrums nötig sind. Natürlich ist dieses Beispiel nicht allgemeingültig. Die Daten sind besonders stark korreliert und die Meßfunktion ist nur in geringem Maße nichtlinear.

Das neue Verfahren zur Surrogaterzeugung zeigt also entscheidende Vorteile gegenüber der AAFT-Methode. Verbleibende, mögliche Probleme sollen im folgenden diskutiert werden. So wurde z.B. das Spektrum durch das Periodogramm geschätzt, ohne daß dabei verschiedene Fenstermethoden („windowing“)

(siehe z.B. [22]) angewendet wurden. Fluktuationen in den einzelnen Frequenzen des Periodogramms der Originaldaten werden somit auf das Periodogramm der Surrogate übertragen. Eine exakte Übereinstimmung der Periodogramme ist also nicht in jedem Fall erwünscht. Im übrigen fixiert das Periodogramm $N/2+1$ Parameter (die Amplituden der Fouriertransformierten), so daß die „Auswahl“ an möglichen zufälligen Surrogaten eingeschränkt ist. Es kann nicht ausgeschlossen werden, daß diese Einschränkung den Phasen der Fouriertransformierten im Laufe des Iterationsschemas eine in irgendeiner Weise geartete Struktur aufzwingt und die Phasen nicht mehr rein zufällig sind. Solch eine schwer zu erkennende Struktur könnte künstlich Nichtlinearität in den Surrogaten hervorrufen. Dies wiederum könnte dazu führen, daß ein Test in den Surrogaten mehr Nichtlinearität findet als in den Daten selbst. Erfüllt der zu untersuchende Datensatz die Nullhypothese, können so zusätzliche falsche Verwerfungen entstehen, d.h. die Size wird erhöht. Für einseitige Tests besteht diese Gefahr nicht, da sie die Nullhypothese nur dann verwerfen, wenn die Daten mehr Nichtlinearität zeigen als die Surrogate. Die Trennschärfe des Tests wird durch eventuelle Nichtlinearität in den Surrogaten aber in jedem Fall gesenkt.

Es sei noch einmal darauf hingewiesen, daß die Nullhypothese ein Gauß'scher, linearer Prozeß mit einer monotonen Meßfunktion ist. So erfüllt z.B. ein linearer Prozeß mit einer Meßfunktion, die Werte zu unterschiedlichen Zeiten kombiniert, nicht die Nullhypothese. Ein Beispiel für eine solche Meßfunktion ist $s_n = x_n^2 x_{n-1}$. Ebenfalls in der Nullhypothese nicht enthalten sind Daten, die aus den ersten Differenzen eines nichtlinear umskalierten, Gauß'schen, linearen Prozesses gewonnen wurden [23]. Ein Verwerfen der Nullhypothese ist aber ein starker Hinweis auf Nichtlinearität in den Daten.

2.5 Analyse der Trennschärfe

Bisher habe ich mich hauptsächlich mit den Eigenschaften und der Erzeugung von Surrogaten für Nichtlinearitätstests beschäftigt. Dabei war es wichtig, eine möglichst allgemeine Nullhypothese aufzustellen und eine korrekte Size des Tests zu garantieren. In diesem Abschnitt soll die Frage beantwortet werden, wie Nichtlinearität gemessen werden kann, d.h. welche Teststatistiken für Tests auf Nichtlinearität verwendet werden können. Wie bereits in Abschnitt 2.2.1 erwähnt, hängt die Trennschärfe eines Tests erheblich von der gewählten Teststatistik ab. Zur Wahl einer Teststatistik gibt es eine Fülle von Vorschlägen, und der Phantasie sind hier nahezu keine Grenzen gesetzt. Es muß lediglich eine Größe gefunden werden, die in irgendeiner Form die Nichtlinearität der Daten wiedergibt.

Sind bestimmte Eigenschaften der zu untersuchenden Zeitreihe a priori bekannt, können diese Informationen bei der Wahl einer geeigneten Teststatistik nützlich sein. Es wird sich zeigen, daß sich bestimmte Teststatistiken in einigen Fällen hervorragend eignen und in anderen Fällen völlig unbrauchbar sind.

Oft sind allerdings keine Informationen über die Eigenschaften der Zeitreihe verfügbar. Unter diesen Umständen sollte eine Teststatistik gewählt werden, die in möglichst vielen Fällen eine hohe Trennschärfe aufweist. Solch eine robuste Teststatistik zu finden, ist Ziel dieses Kapitels. Die Ergbnisse sind auch in [30] zusammengefaßt.

2.5.1 Teststatistiken für Nichtlinearitätstests

Im folgenden werden die verschiedenen Teststatistiken, die zur Untersuchung herangezogen wurden, aufgeführt. Bei der Auswahl wurde darauf geachtet, daß ein großer Bereich an qualitativ verschiedenen Teststatistiken abgedeckt wird. Die meisten anderen Möglichkeiten sind einer der hier verwendeten Größen sehr ähnlich. Die hier angegeben Teststatistiken können also als Repräsentanten verschiedener, ähnlicher Größen angesehen werden. Natürlicherweise sind viele Größen aus der Theorie für nichtlineare dynamische Systeme entliehen.

Die erste Größe, die als Teststatistik herangezogen werden soll, leitet sich aus der Dimension des Attraktors ab. Aus dem Spektrum D_q der Dimensionen ist die Grassberger-Procaccia- Korrelationsdimension D_2 die meist angewendete und untersuchte in der nichtlinearen Zeitreihenanalyse. Ich verwende hier den von Theiler [38] angegebenen *maximum-likelihood*-Schätzer

$$T^{\mathrm{ML}}(m, \tau, \epsilon) = \frac{C_m(\epsilon)}{\int_0^\epsilon \frac{C_m(\epsilon')}{\epsilon'} d\epsilon'} \tag{2.3}$$

für D_2, welcher auch *Takens' estimator* [36] genannt wird. Dabei ist $C_m(\epsilon)$ die in Kapitel 1 angesprochene Korrelationssumme (1.15). Hier und im folgenden werde ich die Parameter der Teststatistiken angeben, die bei den Untersuchungen verändert und angepaßt wurden. Im Falle von (2.3) sind das die Einbettungsdimension m, die Zeitverzögerung τ und die obere Skalengrenze ϵ. Ein Surrogattest mit einer Dimension als Teststatistik ist einseitig, da für einen Originaldatensatz mit Nichtlinearität eine kleinere Dimension als für die Surrogate erwartet wird. Die Dimension der Surrogate nimmt ungefähr den Wert der Einbettungsdimension m an, da die zufälligen Surrogatwerte durch die Einbettung homogen in alle Raumrichtungen verteilt werden.

Alternative und ähnliche Größen sind die Boxdimension D_0, die Informationsdimension D_1 und die punktweise Dimension d_p. Für die punktweise Dimension wird der Skalenexponent der Nachbarabstände für einen einzelnen Referenzpunkt im Phasenraum berechnet. Als Teststatistik kann dann der Mittelwert oder der Median der Werte für mehrere verschiedene Referenzpunkte dienen [33, 3].

Die nächste Teststatistik ist an die BDS-Statistik von Brock et al. angelehnt und basiert ebenfalls auf der Korrelationssumme $C_m(\epsilon)$. In [4, 5] wird gezeigt, daß für unabhängige Zufallszahlen $C_m(\epsilon) = [C_1(\epsilon)]^m$ gilt, und mit $C_1^2 - C_2$ ein formaler Test auf Unabhängigkeit entwickelt. Ich verwende hier eine leicht abgewandelte

Form:

$$T^{\mathrm{BDS}}(m, \tau, \epsilon) = C_m(\epsilon)/[C_1(\epsilon)]^m\,. \tag{2.4}$$

Auch diese Teststatistik führt zu einen einseitigen Test. Für die Surrogate werden Werte $T^{\mathrm{BDS}} \approxeq 1$ erwartet, wohingegen für Daten, die die Nullhypothese nicht erfüllen, $T^{\mathrm{BDS}} > 1$ gilt.

Auch die dritte Teststatistik kommt aus der Theorie der deterministischen, dynamischen Systeme. Die Standardabweichung einer Vorhersage für eine bestimmte Zeit von dem wirklich angenommen Wert ist der Vorhersagefehler

$$T^{\mathrm{VF}}(m, \tau, \epsilon) = \left(\frac{1}{N-1}\sum_n [s_{n+1} - F(s_n)]^2\right)^{1/2}\,. \tag{2.5}$$

Die Vorhersage $F(s_n)$ kann dabei auf verschiedene Art und Weise realisiert werden. Ich verwende hier die lokal konstante Vorhersage (1.12) aus Kapitel 1. Bei der Bestimmung von $F(s_n)$ verwende ich sowohl vergangene als auch zukünftige Nachbarn in der ϵ-Umgebung, die zeitlich weiter als t_{min} von s_n entfernt sind. Um aussagekräftige, tatsächliche Vorhersagen zu erhalten, wäre ein solches Vorgehen nicht korrekt. Hier dient der Vorhersagefehler aber als Teststatistik, die lediglich mit den Surrogaten verglichen wird und nichts über die Qualität tatsächlicher Vorhersagen aussagt.

Neben der Möglichkeit die Nachbarpunkte verschieden zu wichten, kann auch ein Modell an die Daten gefittet werden. Sind die Parameter des Modells ermittelt, können Vorhersagen durch einfaches Einsetzen der Vergangenheitswerte der Zeitreihe getroffen werden. In [2] wird z.B. ein nichtlineares, diskretes Volterra-Wiener-Korenberg-Modell für die Vorhersage verwendet. Da es auch in dieser Arbeit um das Aufspüren von nichtlinearer Dynamik in kurzen Zeitreihen geht, werde ich meine Ergebnisse mit den dort erzielten vergleichen.

Anstelle der Standardabweichung in (2.5) kann auch die mittlere, absolute Abweichung $\langle |s_{n+1} - F(s_n)| \rangle$ oder ihr Logarithmus verwendet werden.

Auch der maximale Lyapunovexponent, als eine wichtige, charakterisierende Größe der Dynamik, ist als Teststatistik vorstellbar. Wie der maximale Lyapunovexponent aus Zeitreihen geschätzt werden kann, ist in [12] beschrieben.

Die beiden letzten Teststatistiken sind Verallgemeinerungen der linearen Korrelationen bzw. Momente höherer Ordnung. Dies ist zum einen die *3-Punkt-Autokorrelation* oder Autokorrelation dritter Ordnung

$$T^{\mathrm{C3}}(\tau_1, \tau_2) = \frac{\langle (s_n - \bar{s})(s_{n-\tau_1} - \bar{s})(s_{n-\tau_2} - \bar{s}) \rangle}{\langle (s_n - \bar{s})^3 \rangle}\,. \tag{2.6}$$

Analog zum Wiener-Khichnin Theorem (1.6) sind die Autokorrelationen höherer Ordnung mit den *Polyspektren* verknüpft. Die Fouriertransformierte der Autokorrelation dritter Ordnung wird *Bispektrum* genannt [34]. Hier ist wichtig, daß

Autokorrelationen mit Ordnung größer zwei für Gauß'sche, lineare Prozesse verschwinden. Für die Surrogate erwartet man also $T^{\mathrm{C3}} = 0$, während ein zu untersuchender Datensatz mit Nichtlinearität sowohl in negative als auch positive Richtung davon abweichen kann. Ein entsprechender Test ist also zweiseitig zu wählen. Ich werde hier lediglich den Fall $\tau_2 = 2\tau_1$ betrachten und für τ_1 einfach τ schreiben.

Die letzte Größe, die als Teststatistik verwendet werden soll, ist ein Maß für die Zeitumkehrinvarianz

$$T^{\mathrm{REV}}(\tau) = \frac{\langle (s_n - s_{n-\tau})^3 \rangle}{\langle (s_n - s_{n-\tau})^2 \rangle} . \tag{2.7}$$

Werden die s_n durch s_{-n} ersetzt, geht T^{REV} in $-T^{\mathrm{REV}}$ über. Ist die zugrundeliegende Dynamik der Zeitreihe zeitumkehrinvariant, gilt entsprechend $T^{\mathrm{REV}} = 0$. Da lineare Prozesse zeitumkehrinvariant sind, erwarten wir für die Surrogate einen Wert von $T^{\mathrm{REV}} = 0$. Außerdem muß auch mit dieser Teststatistik ein zweiseitiger Test durchgeführt werden, da ein Abweichen von der Nullhypothese wieder in beide Richtungen möglich ist.

2.5.2 Verifikation der Size

Ist eine bestimmte Realisierung eines Tests konstruiert worden, muß zunächst die Size des Tests ermittelt und mit der nominellen Size verglichen werden. Die Konsequenzen von Abweichungen wurden bereits in Abschnitt 2.2.1 beschrieben. Die Size kann über eine Serie von unabhängigen Tests ermittelt werden. Für jede der oben aufgeführten Teststatistiken muß eine solche Serie von Tests durchgeführt werden. Die verwendeten Surrogate werden mit dem neuen Verfahren aus Abschnitt 2.3.4 erzeugt, wobei das Iterationsschema zehnmal durchlaufen wird.

Es wird ein Signifikanzniveau von 90% vorgegeben, so daß für einseitige Tests 9 Surrogate und für zweiseitige Tests 19 Surrogate erzeugt werden müssen. In der Praxis werden für gewöhnlich Tests mit einem Signifikanzniveau von mindestens 95% verwendet. Da ich hier relativ viele Tests mit unterschiedlichen Teststatistiken durchführen will, habe ich ein relativ niedriges Signifikanzniveau gewählt, um den Rechenaufwand zu senken.

Um die Size zu ermitteln, muß mit Daten getestet werden, die die Nullhypothese erfüllen. Die hier verwendeten Datensätze sind einmal ein AR(1)-Prozeß $x_n = 0.9x_{n-1} + \eta_n$, der mit der monotonen Meßfunktion $s_n = x_n^3$ aufgenommen wird, und zweitens gaußverteilte Zufallszahlen. In beiden Fällen bestanden die Zeitreihen aus $N = 2048$ Werten. Mit jeder Teststatistik wurden 200 Tests für beide Datensätze durchgeführt. Die Ergebnisse sind in Tabelle 2.1 aufgeführt. Die Fehler basieren auf der Annahme, daß die Anzahl der Verwerfungen binomialverteilt sind. In fast allen Fällen stimmt die empirische Size α mit der nominellen von 0.1 innerhalb der Fehler überein. Insbesondere gibt es keinen Fall, in dem die

Teststatistik	Size α	
	AR(0)	AR(1)
T^{ML}	0.11 ± 0.02	0.09 ± 0.02
T^{BDS}	0.09 ± 0.02	0.07 ± 0.02
T^{VF}	0.10 ± 0.02	0.10 ± 0.02
T^{C3}	0.09 ± 0.02	0.08 ± 0.02
T^{REV}	0.11 ± 0.02	0.10 ± 0.02

Tabelle 2.1: Der Anteil der Verwerfungen aus 200 Tests für verschiedene Teststatistiken. Als Datensätze wurden gaußverteilte Zufallszahlen und ein AR(1)-Prozeß verwendet.

Size zu groß ist, was den Test unbrauchbar machen würde. Relativ kleine Abweichungen nach unten können höchstens zu einer leichten Senkung der Trennschärfe führen. Für den AR(1)-Prozeß sind die Abweichungen von der nominellen Size etwas größer, was durch die relativ starke Nichtlinearität der Meßfunktion erklärbar ist.

2.5.3 Vergleich der Teststatistiken über die Trennschärfe

Die Trennschärfe eines Tests kann ermittelt werden, indem mehrere unabhängige Tests mit Datensätzen, die die Nullhypothese nicht erfüllen, durchgeführt werden. Ist Q die Anzahl der Tests und Y die Anzahl der erfolgreichen Verwerfungen der Nullhypothese, so ist die Trennschärfe $1 - \beta = Y/Q$, wobei für große Q die Genauigkeit, mit der die Trennschärfe angegeben werden kann, besser wird.

Um die einzelnen Teststatistiken bezüglich ihrer Trennschärfe zu untersuchen und zu vergleichen, müssen also für jede Teststatistik viele Tests durchgeführt werden. Dabei ist es nicht ratsam, als zu untersuchenden Datensatz eine Zeitreihe zu verwenden, die stark von der Nullhypothese abweicht, da dadurch für alle obigen Teststatistiken stets eine Trennschärfe von 1 herauskommt. Ein Vergleich ist dann natürlich nicht möglich. Im Gegenteil sollte die Zeitreihe relativ „nah“ an der Nullhypothese liegen. Es müssen also Daten erzeugt werden, die zwar nah an der Nullhypothese liegen, sie aber doch *nicht* erfüllen. Zeitreihen, die die Nullhypothese nicht erfüllen, sind leicht über nichtlineare, chaotische Abbildungen oder Differentialgleichungen zu erzeugen. Eine Möglichkeit, solche Datensätze näher an die Nullhypothese heranzubringen, besteht darin, einen rein stochastischen Prozeß zu überlagern. Auf einen deterministischen Prozeß $\{x_n\}$ wird also Rauschen $\{\eta_n\}$ addiert. Damit die Autokorrelationen der verrauschten Daten nicht von denen der rauschfreien abweichen, wird Rauschen mit gleichem Frequenzspektrum addiert. Oft wird diese Art Rauschen auch *in-band noise* genannt. Wie Rauschen mit gleichem Spektrum erzeugt werden kann, habe ich bereits in Abschnitt 2.3.2 gezeigt. Die $\{\eta_n\}$ sind nichts anderes als ein FT-Surrogat mit zufälligen Phasen.

Die verrauschte Zeitreihe $\{s_n\}$ wird mittels

$$s_n = \sqrt{\frac{1}{1+a^2}}(x_n + a\eta_n) \tag{2.8}$$

zusammengesetzt, wobei a das *relative Rauschniveau* ist. Für $a = 0$ ist $s_n = x_n$ und im Limes $a \to \infty$ erhält man $s_n = \eta_n$. Der Vorfaktor in (2.8) garantiert, daß die verrauschte Zeitreihe die gleiche Varianz wie der reine Prozeß besitzt. Dadurch können Einflüsse einer eventuellen Sensitivität der Teststatistik auf die Varianz ausgeschlossen werden. Für jede Teststatistik kann eine Folge von Tests mit unterschiedlichem Rauschniveau durchgeführt werden. Interessant ist das Verhalten der Trennschärfe mit steigendem Rauschniveau. Speziell ist zu erwarten, daß die Trennschärfe ab einem bestimmtem Rauschniveau kleiner als 1 wird. Diese Schwelle $a_{\max}$ wird für die einzelnen Teststatistiken unterschiedlich sein.

Bevor ich auf die Ergebnisse im einzelnen eingehe, möchte ich näher die Wahl der freien Parameter in den Teststatistiken diskutieren. Diese Parameter sind in den Teststatistiken (2.3)-(2.7) die Zeitverzögerung τ, die Einbettungsdimension m und der Radius ϵ in der Korrelationssumme bzw. in der Vorhersage. Diese Parameter sollen hier aber nicht so eingestellt werden, daß z.B. die Dimension des Attraktors möglichst genau ermittelt werden kann oder der Vorhersagefehler möglichst klein wird. Vielmehr soll mit der Wahl der Parameter eine möglichst große Trennschärfe erreicht werden. Die Wahl von τ ist in einigen Fällen „natürlich“ vorgegeben. So ist z.B. bei rekursiven Abbildungen oft $\tau = 1$ geeignet. Ein Abweichen von diesem Wert senkt oft die Trennschärfe bereits erheblich. In den anderen Fällen wurden verschiedene Möglichkeiten durch einige Tests mit einem festem Rauschniveau bezüglich ihrer Trennschärfe untersucht und die Zeitverzögerung gewählt, die maximale Trennschärfe lieferte. Die Änderungen von τ ergaben aber in vielen Fällen keine wesentlichen Änderungen der Trennschärfe.

Auch für die Einbettungsdimension m wurden mehrere Möglichkeiten ausprobiert und diejenige mit maximaler Trennschärfe gewählt. Oft ergibt sich dabei tatsächlich für m ungefähr die zweifache Boxdimension, wie in Abschnitt 1.3.2 über die Einbettung von Zeitreihen bereits angedeutet. Für relativ kurze Zeitreihen sind große Einbettungsdimensionen meist ungünstig, da bei festem ϵ für größere m immer weniger Nachbarn im Einbettungsraum gefunden werden können und statistische Fluktuationen zum Tragen kommen. Das Durchspielen verschiedener Möglichkeiten von τ und m kommt dem praktischen Fall sehr nahe, in dem über die Zeitreihe diesbezüglich keine Informationen vorliegen.

Etwas näher möchte ich auf die Wahl von ϵ eingehen, da die Abhängigkeit der Trennschärfe von diesem Parameter wesentlich stärker sein kann. Um die Trennschärfe zu ermitteln, muß zunächst ein Datensatz herangezogen werden, der die Nullhypothese nicht erfüllt. Ich verwende hier die nichtlineare, chaotische

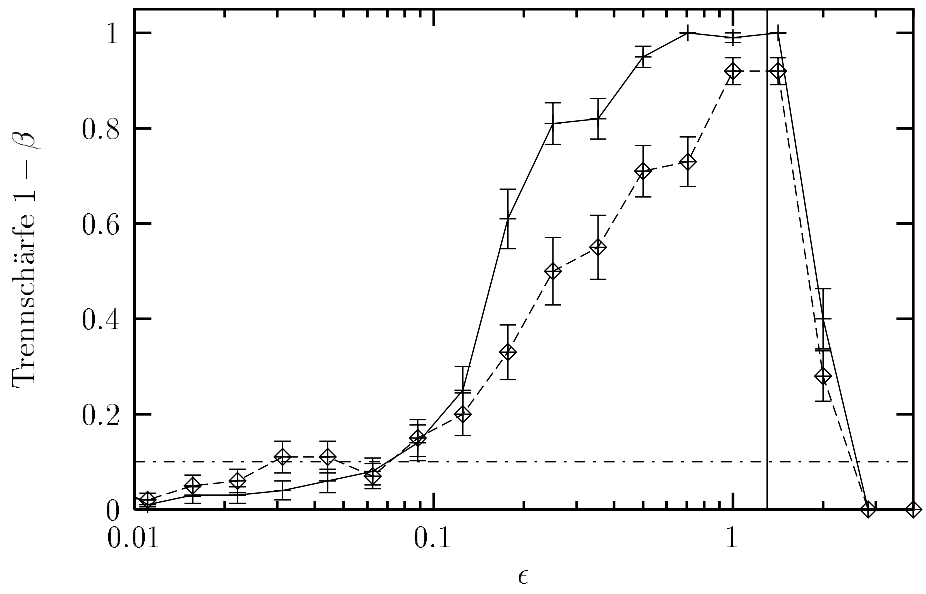

Abbildung 2.5: Abhängigkeit der Trennschärfe von ϵ für die Hénon-Abbildung mit den relativen Rauschniveaus $a = 1.2$ (durchgezogen) und $a = 1.4$ (gestrichelt). Als Teststatistik wurde der Vorhersagefehler verwendet. Für die vertikale Linie gilt $\epsilon = (s_{\max} - s_{\min})/4$.

Abbildung

$$\begin{aligned} x_{n+1} &= A - x_n^2 + By_n \\ y_{n+1} &= x_n \end{aligned} \quad , \tag{2.9}$$

welche unter dem Namen *Hénon-Abbildung* [10] bekannt ist. Die Parameter wurden wie gewöhnlich auf die Werte $A = 1.4$ und $B = 0.3$ gesetzt. Als zu untersuchende Zeitreihe werden $N = 2048$ der x-Werte herangezogen, auf die unter Verwendung von (2.8) noch Rauschen addiert wird. Nun wird für verschiedene ϵ eine Serie von $Q = 100$ Tests durchgeführt, um die Trennschärfe zu ermitteln. Für jeden Test werden weitere 2048 x-Werte ermittelt und jeweils neu Rauschen addiert. Dadurch können die Tests als wirklich unabhängig voneinander betrachtet werden. Die Abhängigkeit der Trennschärfe von ϵ ist in Abbildung 2.5 am Beispiel des Vorhersagefehlers (2.5) mit $m = 3$ und $\tau = 1$ als Teststatistik dargestellt. Es ist deutlich zu sehen, daß sowohl für kleine als auch für große ϵ die Trennschärfe stark absinkt und sogar unter der nominellen Size von 0.1 (9 Surrogate) liegt. In diesem Bereichen wird der Vorhersagefehler als Teststatistik unbrauchbar. Das Verhalten zu großen ϵ ist dadurch zu erklären, daß der Vorhersagefehler (2.5) für große ϵ mit der Standardabweichung der Zeitreihe selbst übereinstimmt, da die Vorhersage F dann über alle Werte der Zeitreihe mittelt. Da die Surrogate aber per Konstruktion dieselbe Varianz wie der Originaldatensatz besitzen, kann es zu keiner Verwerfung der Nullhypothese kommen. Für kleine ϵ können unter Umständen keine Nachbarn im Phasenraum gefunden werden, was eine Vorhersa-

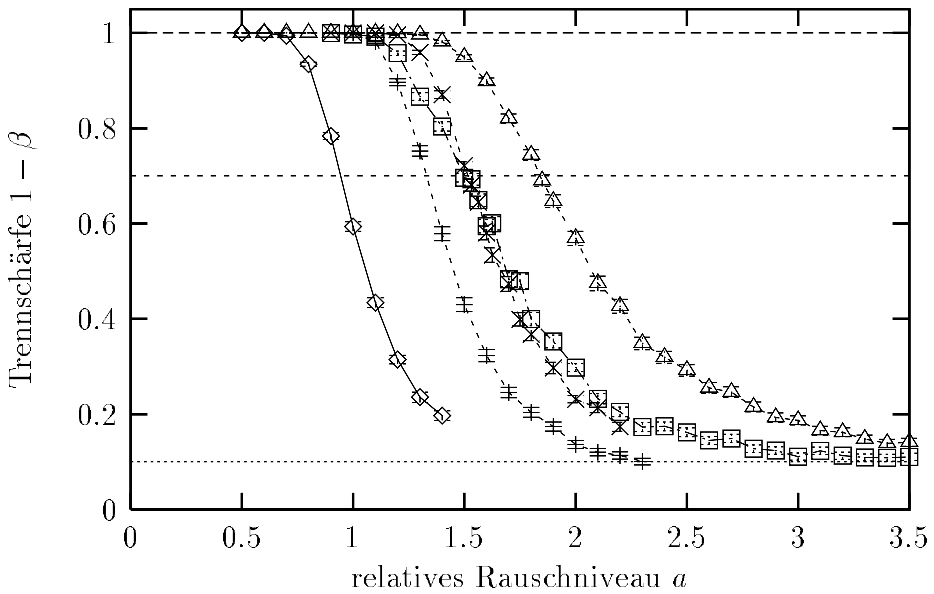

Abbildung 2.6: Vergleich der Trennschärfen für verschiedene Teststatistiken. Verwendet wurden Daten der Hénon-Abbildung mit verschiedenen Rauschniveaus a. Von links nach rechts: T^{ML}, T^{BDS}, T^{VF} (Kreuze), T^{C3} und T^{REV}.

ge mittels (1.12) unmöglich macht. Mein Programm verwendet in diesem Fall den Mittelwert der Zeitreihe als Vorhersagewert, wodurch es zu denselben Effekten kommt, wie für große ϵ.

Weitergehende Versuche mit anderen Datensätzen haben gezeigt, daß für den Vorhersagefehler das Maximum der Trennschärfe ungefähr bei $\epsilon = (s_{\max} - s_{\min})/4$ liegt, wobei $s_{\min}$ der kleinste und $s_{\max}$ der größte Wert der Zeitreihe ist. Der entsprechende Wert ist als vertikale Linie in Abbildung 2.5 eingetragen. Für die Teststatistiken (2.3) und (2.4) ergibt sich die maximale Trennschärfe bei einem Wert von ϵ, der ungefähr bei der Hälfte der Standardabweichung der Zeitreihe liegt. Diese Einstellungen für die ϵ-Werte der einzelnen Teststatistiken werde ich im folgenden verwenden und nicht mehr explizit angeben. Sie wurden auch in Abschnitt 2.5.2 bei der Verifikation der Size verwendet.

Alternativ habe ich versucht, ϵ proportinal zu $(1/N)^{\frac{1}{m}}$ zu wählen, um der Tatsache, daß für große Einbettungsdimensionen weniger Nachbarn gefunden werden können, entgegen zu wirken. Dies lieferte aber insbesondere für höhere Einbettungsdimensionen zu große ϵ-Werte, so daß die Trennschärfe gesenkt wurde.

Zur Untersuchung der Abhängigkeit der Trennschärfe vom relativen Rauschniveau a wurden für jede Teststatistik $Q = 2000$ Tests mit unterschiedlichem Rauschniveau durchgeführt. Als Zeitreihe dienten wieder 2048 x-Werte der Hénon-Abbildung. In Abbildung 2.6 ist zu sehen, wie die Trennschärfe mit steigendem

Teststatistik	Parameter	$a_{\max}$	
		$\beta = 0.05$	$\beta = 0.3$
T^{ML}	$m = 2$	0.7	0.9
T^{BDS}	$m = 3$	1.1	1.3
T^{VF}	$m = 3$	1.2	1.5
T^{C3}	$\tau = 1$	1.1	1.5
T^{REV}	$\tau = 1$	1.4	1.8

Tabelle 2.2: Maximal erlaubte Rauschniveaus, um mit einer Wahrscheinlichkeit von 95% bzw. 70% die Nullhypothese eines Gauß'schen, linearen Prozesses zu verwerfen. Resultate für die Hénon-Abbildung mit 2048 Werten.

Rauschniveau wie erwartet sinkt. Für stark verrauschte Zeitreihen stimmt die Trennschärfe mit der Size α überein, da der Test die Nullhypothese dann nur noch zufällig mit der Wahrscheinlichkeit α verwerfen kann. Das Signifikanzniveau wurde wieder auf 90% festgelegt, um den Rechenaufwand im Rahmen zu halten. Das Verhalten in Abhängigkeit vom Rauschniveau ist für die verschiedenen Teststatistiken qualitativ gleich und unterscheidet sich im wesentlichen nur in dem Rauschniveau, ab dem die Trennschärfe zu sinken beginnt. In Tabelle 2.2 sind die maximalen relativen Rauschniveaus $a_{\max}$ eingetragen, bis zu dem die Trennschärfe über 0.95 bzw. über 0.7 liegt. Tests mit einer Trennschärfe unter 70% besitzen nur eine sehr geringe Aussagekraft und sollten in der Praxis nicht angewendet werden. Ebenfalls in Tabelle 2.2 eingetragen sind die Einstellungen der Parameter, die maximale Trennschärfe für die einzelnen Teststatistiken lieferten. Ein Abweichen von $\tau = 1$ lieferte in allen Fällen wesentlich schlechtere Ergebnisse. Zusammenfassend läßt sich sagen, daß die Größe für die Zeitumkehrinvarianz im Falle der Hénon-Abbildung die am besten geeignete Teststatistik ist. Sie lieferte für alle Rauschniveaus die größte Trennschärfe.

In [2] wurde ein Vorhersagefehler mit Vorhersagen, die auf einem Volterra-Wiener-Modell basieren, als Teststatistik herangezogen. Dort wurde für einen Datensatz mit 1000 Werten der Hénon-Abbildung ein maximales Rauschniveau von 70% ermittelt, um mit 99-prozentiger Wahrscheinlichkeit Nichtlinearität zu erkennen. Ein von mir durchgeführter Test mit 1024 Werten der Hénon-Abbildung und dem Vorhersagefehler (2.5) lieferte noch bei einem Rauschniveau von 100% eine Trennschärfe von 1 (200 unabhängige Tests). Die Zeitumkehr-Teststatistik erreichte sogar bis $a = 1.2$ noch 200 Verwerfungen. Interessant ist hier, daß relativ einfach konstruierte Teststatistiken die beste Wahl für Nichtlinearitätstests sind.

Um die einzelnen Teststatistiken bezüglich ihrer Trennschärfe möglichst umfassend untersuchen zu können, reicht ein Datensatz nicht aus. Es müssen Tests mit Prozessen verschiedener Natur durchgeführt werden. Die große Menge der Prozesse, die die Nullhypothese nicht erfüllen, ist aber nicht einfach abzustecken.

Im folgenden werde ich drei weitere Prozesse zur Erzeugung der zu untersuchenden Datensätze heranziehen.

Dabei ist ein Vergleich der Teststatistiken über einen Plot wie in Abbildung 2.6 sehr rechenaufwendig, da für jeden Punkt eine ganze Serie von Tests durchgeführt werden muß. Eine andere Möglichkeit, die Teststatistiken zu vergleichen, besteht darin das maximale relative Rauschniveau $a_{\max}$ der einzelnen Teststatistiken zu ermitteln, bis zu dem die Trennschärfe oberhalb einer festgelegten Grenze liegt. Aber auch hier müssen relativ viele Tests mit unterschiedlichen Rauschniveaus durchgeführt werden, um möglichst genau diese Grenze zu finden. Ich gehe im folgenden nach einer etwas anderen Methode vor. Ich wähle ein festes Rauschniveau a aus dem Intervall, in dem sich die Trennschärfen der Teststatistiken relativ deutlich voneinander unterscheiden. Dieses Intervall reicht für die Hénon-Abbildung in Abbildung 2.6 grob von $a = 1.3$ bis ungefähr $a = 2.0$. Ein solches Intervall ist relativ breit und ein einzelner Wert aus diesem Intervall kann relativ schnell durch einige Tests gefunden werden. Für diesen Wert des relativen Rauschniveaus werden dann die Trennschärfen der einzelnen Teststatistiken über ausreichend viele Tests miteinander verglichen. Dabei nutze ich die Tatsache, daß die Teststatistiken bei verschiedenen Rauschniveaus sich bezüglich ihrer Trennschärfe qualitativ nicht unterscheiden, was durch Abbildung 2.6 bestätigt wird.

Die Hénon-Abbildung (2.9) ist ein Prozeß, welcher zeitlich diskret ist. Bei der Erzeugung des nächsten Datensatzes, verwende ich eine Dynamik der zeitlich kontinuierlichen Form (1.8). Die drei gekoppelten, gewöhnlichen Differentialgleichungen

$$\begin{aligned} \dot{x} &= -\sigma x + \sigma y \\ \dot{y} &= -xz + rx - y \\ \dot{z} &= xy - bz \end{aligned} \tag{2.10}$$

sind als *Lorenz-System* [16] bekannt, und standardmäßig werden die Parameter zu $\sigma = 10$, $b = 8/3$ und $r = 28$ gewählt. Um aus (2.10) eine Zeitreihe zu generieren, wird das Differentialgleichungssystem mit dem Runge-Kutta-Verfahren vierter Ordnung (siehe z.B. [22]) integriert und der Wert der x-Koordinate im Abstand von 0.08 Zeiteinheiten aufgenommen. Ist die gewünschte Anzahl von Werten aufgenommen, werden x, y und z zwischengespeichert, um weitere Daten von diesem Punkt im Phasenraum ausgehend zu gewinnen. Für die Tests wurden wieder $N = 2048$ Werte pro Zeitreihe verwendet, auf die Rauschen mit einem Niveau von $a = 1.3$ addiert wurde. Welche Trennschärfen die Teststatistiken lieferten, ist in Tabelle 2.3 aufgetragen. Mit weniger Tests wurde verfiziert, daß die angegebenen Parameter die Trennschärfe der verschiedenen Teststatistiken maximierten. Im Vergleich zu den Resultaten der Tests mit der Hénon-Abbildung wird klar, daß es keine pauschal beste Teststatistik gibt. Die für die Hénon-Abbildung noch am besten geeignete Zeitumkehr-Teststatistik versagt bei der Erkennung von Nichtlinearität in den Lorenz-Daten vollkommen. Daran wird deutlich, daß Untersuchungen mit Daten aus verschiedenartigen Prozessen auf jeden Fall sinn-

Teststatistik	Parameter	Trennschärfe $1-\beta$
T^{ML}	$m=3$	0.25 ± 0.02
T^{BDS}	$m=2$	0.24 ± 0.02
T^{VF}	$m=4$	0.66 ± 0.02
T^{C3}	$\tau=3$	0.09 ± 0.01
T^{REV}	$\tau=3$	0.10 ± 0.01

Tabelle 2.3: Anteil der erfolgreichen Verwerfungen aus 1000 Tests mit verrauschten Lorenz-Daten ($a=1.3$). Tests mit den Teststatistiken T^{C3} und T^{REV} lieferten keine signifikanten Verwerfungen.

Teststatistik	Parameter	Trennschärfe $1-\beta$
T^{ML}	$m=3$	0.61 ± 0.03
T^{BDS}	$m=3$	0.86 ± 0.02
T^{VF}	$m=3$	0.79 ± 0.02
T^{C3}	$\tau=3$	0.45 ± 0.03
T^{REV}	$\tau=1$	0.35 ± 0.02

Tabelle 2.4: Anteil der erfolgreichen Verwerfungen aus 600 Tests mit verrauschten NMR-Laser-Daten ($a=0.8$).

voll sind. Der Vorhersagefehler ist für die Lorenz-Daten eindeutig die beste Wahl.

Auch in [2] wurden Lorenz-Daten für Nichtlinearitätstests herangezogen. Wie bereits erwähnt, wurden in dieser Arbeit Vorhersagefehler eines Volterra-Wiener-Modelles als Teststatistik verwendet. Für Zeitreihen mit 1000 Werten, die aus dem Lorenz-System erzeugt wurden, konnte bis zu einem Rauschniveau von 50% noch eine Trennschärfe über 99% beobachtet werden. Mit dem von mir verwendeten Vorhersagefehler konnte in 200 Tests noch bis zu 60% additivem Rauschen in allen Fällen die Nullhypothese verworfen werden.

Die Daten für die nächsten Tests sind einem Experiment entnommen. Der Datensatz besteht aus 60000 Werten aus einem Experiment mit einem NMR (*nuclear magnetic resonance*) Laser [7]. Um mehrere Tests zur Bestimmung der Trennschärfe durchführen zu können, wurde der große Datensatz in 600 Zeitreihen mit jeweils 1000 Werten unterteilt. Da auch hier alle Teststatistiken für die reinen Daten eine Trennschärfe von 1 lieferten, wurde Rauschen mit $a=0.8$ addiert. Die Resultate sind in Tabelle 2.4 aufgetragen. Alle Teststatistiken können signifikant Nichtlinearität in den Zeitreihen erkennen, wobei wieder eine andere Teststatistik zur maximalen Trennschärfe führt. Diese ist hier, wenn auch nicht stark dominierend, die BDS-Statistik.

Als letztes sollen gekoppelte Abbildungen zur Erzeugung der Daten herange-

Teststatistik	Parameter	Trennschärfe $1-\beta$
T^{ML}	$m=2$	0.39 ± 0.02
T^{BDS}	$m=3$	0.83 ± 0.01
T^{VF}	$m=3$	0.71 ± 0.02
T^{C3}	$\tau=1$	0.40 ± 0.02
T^{REV}	$\tau=1$	0.70 ± 0.02

Tabelle 2.5: Trennschärfen der verschiedenen Teststatistiken für 20 gekoppelte, logistische Abbildungen. Rauschen mit $a = 1.3$ wurde addiert.

zogen werden. Dabei betrachte ich zunächst die *logistische* Abbildung

$$x_{n+1} = r x_n(x_n - 1) = \lambda(x_n) \tag{2.11}$$

mit $r = 4$. Gekoppelt werden die Abbildungen auf einem periodischen Gitter mit M Gitterpunkten über

$$x_n^{(m)} = \lambda\left(\frac{\kappa}{2}x_n^{(m-1)} + (1-\kappa)x_n^{(m)} + \frac{\kappa}{2}x_n^{(m+1)}\right), \quad m = 1, \ldots, M, \tag{2.12}$$

wobei $\kappa \in [0,1]$ der Kopplungsfaktor ist. Die untersuchte Zeitreihe besteht aus $N = 2048$ Werten des ersten Gitterpunktes $x_n^{(1)}$ auf einem Gitter mit $M = 20$ und Kopplunksfaktor $\kappa = 0.5$. Zusätzlich wurde Rauschen mit $a = 1.3$ addiert. Die Resultate der Trennschärfen aus jeweils 1000 Tests sind in Tabelle 2.5 mit den Parametern der Teststatistiken aufgetragen. Auch hier kann mit allen verwendeten Teststatistiken Nichtlinearität in den Daten erkannt werden.

Bisher habe ich Rauschen addiert, um die Daten nichtlinearer Prozesse in Richtung Nullhypothese zu verschieben. Dies wurde dadurch motiviert, daß auch gemessene Zeitreihen mehr oder weniger mit Rauschen kontaminiert sind und so die Analysen erschweren. Ein anderes Problem tritt bei Zeitreihen auf, bei deren Messung über Prozesse mit vielen Freiheitsgraden gemittelt wurde. Auch in solchen Systemen kann die Dynamik unter Umständen nur schwer von stochastischen Prozessen unterschieden werden. Eine Mittelung über mehrere Prozesse kann also auch die Erkennbarkeit von Nichtlinearität erschweren. Deshalb betrachte ich im folgenden Zeitreihen, die durch Mittelung von M *ungekoppelten* ($\kappa = 0$) *Zelt*-Abbildungen

$$x_{n+1} = \begin{cases} 2x_n & \text{für } x_n \leqslant 1/2 \\ 2 - 2x_n & \text{für } x_n > 1/2 \end{cases} \tag{2.13}$$

entstanden sind. Wie erwartet ist in Abbildung 2.7 zu sehen, daß die Trennschärfe mit der Anzahl M der Abbildungen, über die gemittelt wird, sinkt. Zur Ermittlung der Trennschärfe wurden 200 unabhängige Tests mit Zeitreihen durchgeführt, auf die kein zusätzliches Rauschen addiert wurde. Im Vergleich mit Abbildung 2.6 fällt auf, daß das Verhalten für verschiedene Teststatistiken sich hier

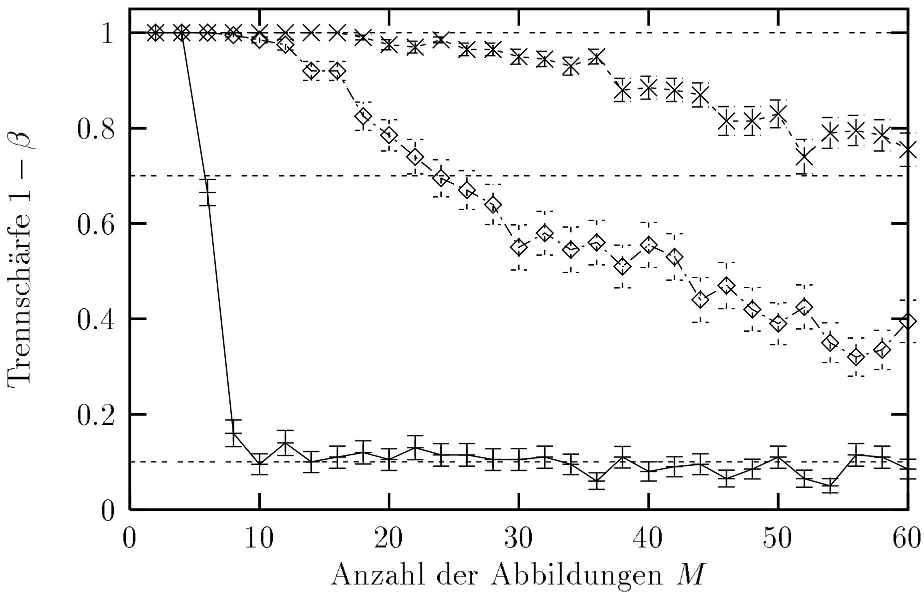

Abbildung 2.7: Trennschärfen in Abhängigkeit von der Anzahl der Zelt-Abbildungen, über die gemittelt wurde. Die verwendeten Teststatistiken von oben nach unten: T^{REV}, T^{VF} und T^{ML}. Es wurden jeweils 200 Tests mit Zeitreihen ohne Rauschen durchgeführt.

stärker unterscheidet. Für den *maximum-likelihood*-Schätzer der Dimension fällt die Trennschärfe bei einer bestimmten Anzahl M abrupt auf die Size ab. Im Gegensatz dazu kann die Zeitumkehr-Teststatistik sogar für $M = 60$ noch in über 70% der Fälle Nichtlinearität in den Daten erkennen.

In Tabelle 2.6 sind die Trennschärfen aller verwendeter Teststatistiken für $M = 16$ ungekoppelte Zelt-Abbildungen aufgetragen. Neben dem Dimensionsschätzer liefern auch die BDS-Statistik und die 3-Punkt-Autokorrelation keine signifikanten Verwerfungen.

Teststatistik	Parameter	Trennschärfe $1-\beta$
T^{ML}	$m = 2$	0.11 ± 0.03
T^{BDS}	$m = 2$	0.10 ± 0.03
T^{VF}	$m = 4$	0.92 ± 0.03
T^{C3}	$\tau = 1$	0.10 ± 0.03
T^{REV}	$\tau = 1$	1.00 ± 0.00

Tabelle 2.6: Anteil der erfolgreichen Verwerfungen aus 100 Tests für einen über 16 ungekoppelte Zelt-Abbildungen gemittelten Datensatz.

2.6 Zusammenfassung

Die Resultate zeigen, daß für Prozesse unterschiedlicher Natur verschiedene Teststatistiken für Tests auf Nichtlinearität geeignet sind. Aus den ausgewählten Teststatistiken kann nur der Vorhersagefehler T^{VF} in allen betrachteten Fällen eine akzeptable Trennschärfe liefern. Die anderen Teststatistiken sind in mindestens einem der Beispiele völlig unbrauchbar. Die Zeitumkehrinvarianz T^{REV} lieferte speziell für die zeitlich diskreten Abbildungen oft die besten Ergebnisse. Im Falle der Lorenz- und der Laser-Daten ist sie aber die am wenigsten geeignete. Asymmetrie unter Zeitumkehr ist zwar ein guter Indikator für Nichtlinearität, aber keine notwendige Bedingung.

Welcher Teststatistik in einer speziellen Situation der Vorzug gegeben werden sollte, hängt stark von der Menge der zur Verfügung stehenden Daten ab. Bei großen Datenmengen können in unabhängigen Tests die Teststatistiken verglichen und entsprechend ausgewählt werden. Typischerweise stehen aber oft nur wenige und kurze Datensätze zur Verfügung, für die teilweise bereits erheblicher zeitlicher und technischer Aufwand betrieben wurde. In solchen Fällen sollte eine robuste, allgemein funktionierende Teststatistik mit möglichst wenigen Parametern gewählt werden, also z.B. der hier untersuchte Vorhersagefehler (2.5). Gibt es aber z.B. per visueller Inspektion Hinweise auf Zeitasymmetrie in den Daten, so liefert wahrscheinlich T^{REV} die besten Resultate.

Es können auch mehrere Tests mit unterschiedlichen Teststatistiken für einen Datensatz durchgeführt werden. Die nominelle Size ist dann durch die Wahrscheinlichkeit gegeben, daß mindestens einer der Tests eine falsche Verwerfung der Nullhypothese liefert. So führen z.B. fünf unabhängige Tests mit einem Signifikanzniveau von 95% zu einer nominellen Size von $p = 1-(0.95)^5 \approxeq 0.23$, wodurch das Resultat des Tests nicht mehr als signifikant angesehen werden kann. In so einem Fall sollte also entweder das Signifikanzniveau der einzelnen Tests erhöht oder weniger Tests durchgeführt werden. In diesen (wenigen) Tests ist die Verwendung von robusten Teststatistiken zu empfehlen.

Die Nullhypothese des monoton gemessenen, Gauß'schen, linearen Prozesses ist bis jetzt die allgemeinste Nullhypothese, die nichtlineare Dynamik ausschließt und für die ein entsprechender Test durchführt werden kann. Das in Abschnitt 2.3.4 vorgestellte Verfahren zur Surrogaterzeugung ist gerade auf diese Nullhypothese zugeschnitten und die hier angewendeten Tests untersuchen *jegliche* Abweichung von dieser Nullhypothese. Wie bereits erwähnt, existieren Prozesse, die diese Nullhypothese nicht erfüllen, an denen man aber auf der Suche nach nichtlinearer, deterministischer Dynamik nicht interessiert ist. In einem solchen Fall sind Teststatistiken wie T^{REV} und T^{C3} nicht zu empfehlen, da sie auch relativ sensitiv für diese Prozesse sein können.

Kapitel 3

Klassifikation von Zeitreihen

3.1 Einleitung

Eine der wesentlichen Aufgaben der Naturwissenschaften und speziell der Physik ist die Charakterisierung und Beschreibung von beobachteten Ereignissen. Ein erster möglicher Ansatz in der Analyse solcher Ereignisse, liegt in der Suche nach Strukturen in der Vielzahl der verschiedenen Beobachtungen. Dabei ist das Zusammenfassen von gleichen Ereignissen in Gruppen oder Klassen eine wichtige Grundvoraussetzung. Die damit verbundene Trennung verschiedenartiger Beobachtungen erlaubt erst die getrennte und detailierte Analyse.

Das Problem der *Klassifikation* läßt sich daher auch in vielen Bereichen der Wissenschaft finden. Hat z.B. in der Elemtarteilchenphysik ein Detektor ein Teilchen aufgespürt, stellt sich sofort die Frage nach der Art des Teilchens. Das Teilchen soll dabei anhand seiner Eigenschaften in den vorhandenen „Teilchenzoo“ einsortiert werden. In der Informatik ist die Mustererkennung ein aktuelles Thema. So sollen z.B. verschiedene handgeschriebene Buchstaben korrekt identifiziert werden. Die Biologen wollen giftige Pilze von eßbaren unterscheiden können, usw.

Auch die Klassifikation von Zeitreihen kann von Interesse sein. So muß z.B. ein Arzt anhand des EKGs eines Patienten bekannte Herzfehler oder -krankheiten erkennen und vorhersagen können. Ein weiteres Beispiel für eine Klassifikation von Zeitreihen ist der in Kapitel 2 entwickelte Test auf Nichtlinearität. Weitere Analysen hängen vom Ergebnis eines solchen Tests ab. In diesem Kapitel soll eine Methode erarbeitet werden, um verschiedenartige Zeitreihen voneinander unterscheiden zu können und ähnliche Zeitreihen in Gruppen zusammenzufassen. Die Zeitreihen sollen also nach ihren Eigenschaften klassifiziert werden. Dabei konzentriere ich mich hier auf die Klassifikation von Zeitreihen unter den Gesichtspunkten der nichtlinearen Zeitreihenanlyse.

Die Frage nach der Stationarität einer Zeitreihe ist eng verknüpft mit der Frage, ob mehrere Zeitreihen gleiche Eigenschaften besitzen. Man braucht sich nur die betreffenden Zeitreihen zu einem großen Datensatz zusammengefaßt vorzustellen und die Stationarität dieses Datensatzes zu untersuchen. Umgekehrt läßt sich jede Zeitreihe unterteilen, um die Eigenschaften der Segmente untereinander zu vergleichen. Klassifizierungsmethoden können also auch für die Stationaritätsanalyse herangezogen werden. Dabei ist es wichtig, daß die zugrundeliegende Dynamik in den einzelnen Segmenten der Zeitreihe näherungsweise konstant ist. Wenn die Parameter der Dynamik sich langsam („adiabatisch") ändern, ist diese Konstanz auf genügend kurzen Segmenten gegeben.

Im Falle von Systemen mit diskreten Zuständen sollen diese Zustände erkannt und zeitlich getrennt werden. Es gibt Fälle, in denen die Anzahl der Zustände und sogar die Zustände selbst vorher bekannt sind. Das Hauptaugenmerk liegt dann auf der zeitlichen Trennung der Zustände, d.h. es sollen die Zeitpunkte der Übergänge gefunden werden. Die Einteilung der Zeitreihe sollte in diesem Fall so vorgenommen werden, daß in jedem Segment möglichst nur ein Zustand vom System angenommen wird. Eine ähnliche Situation ergibt sich, wenn mehrere Zeitreihen gegeben sind, die von unterschiedlichen Prozessen erzeugt wurden. Die Frage ist dann, welche Zeitreihen von welchem Prozesse erzeugt wurden.

Ziel dieses Kapitels ist es also, globale Unterschiede innerhalb einer langen oder zwischen mehreren Zeitreihen aufzuspüren. Diese Unterschiede sollen dann als Kriterium zur Klassifizierung herangezogen werden.

3.2 Stationarität

In diesem Abschnitt soll die Stationaritätsanalyse motiviert werden. Auf die Begriffe und Methoden der linearen Zeitreihenanalyse werde ich ebenfalls eingehen.

Für Nichtstationarität gibt es viele Ursachen. Eine davon ist, daß sich Prozesse auf verschiedenen Zeitskalen kombinieren. So ist z.B. die Lufttemperatur sowohl von der Tageszeit als auch von der Jahreszeit abhängig. Wird sie aber lediglich über einen Zeitraum von zwei Monaten gemessen, erscheint die Jahreszeitenabhängigkeit als globale Veränderung. Es gibt eine langsame Zeitentwicklung, über kurze Zeiträume hingegen bleibt die Dynamik konstant. Ähnlich verhält es sich mit Systemen, die von einem Zustand in einen anderen übergehen. Dies kann sprunghaft oder kontinuierlich vor sich gehen.

Neben dem Interesse für die zeitliche Entwicklung in einer Zeitreihe, gibt es weitere Gründe für ein Interesse an der Stationarität. So setzen fast alle quantitativen Verfahren voraus, daß die Zeitreihe stationär ist. Wird dies vorher nicht getestet, kann das zu Fehlern und Fehlinterpretationen führen. Auch für die Überwachung eines Experimentes kann eine Analyse der Stationarität wichtig sein. So ist es z.B. oft erwünscht, daß bestimmte Parameter während der Messung konstant bleiben. Mit einem Stationaritätstest kann die Konstanz dieser Parameter verifiziert werden. Auch unerwünschte äußere Einflüsse können so untersucht werden.

Im folgenden werde ich eine Definition für die Stationarität angeben und auf die Stationaritätsanalyse von Zeitreihen eingehen. Dabei muß wieder zwischen Prozeß und Zeitreihe unterschieden werden. Gegeben sei ein diskreter Prozeß $\{x_n\}$ mit $n \in T$, wobei T eine Menge $\{1, \ldots, N\}$ oder ganz $\mathbb{Z}$ sein kann. Ein Parameter $P = P(n)$ sei für jeden Zeitpunkt n über ein Ensemblemittel definiert. Der Prozeß $\{x_n\}$ ist stationär bezüglich des Parameters P, wenn

$$P(n) = P_c = \text{konstant} \quad \text{für alle } n \in T \tag{3.1}$$

gilt. Der Parameter P ist also unabhängig von der Zeit. In der linearen Zeitreihenanalyse werden über (3.1) verschiedene Stationaritätsbegriffe definiert. Ein Prozeß ist *mittelwertstationär*, wenn (3.1) für die Ensemble-Mittelwertbildung gilt. *Autokovarianzstationär* sind Prozesse, bei denen die Autokovarianz γ ausschließlich von der Zeitverzögerung τ abhängt. Ein autokovarianzstationärer Prozeß ist stets auch *varianzstationär* ($\tau = 0$). Ist ein Prozeß mittelwertstationär und autokovarianzstationär spricht man von *schwacher Stationarität.*

Bei schwacher Stationarität sind lediglich die Momente erster und zweiter Ordnung zeitunabhängig. Die Stationarität kann aber noch weiter verschärft werden. Dafür betrachtet man eine endliche Teilmenge $\{i\}$ von T und die gemeinsame Ensemble-Verteilungsfunktion $p(\{x_i\})$ der Werte zu den Zeitpunkten $\{i\}$. Der Prozeß wird *streng stationär* genannt, wenn

$$p(\{x_i\}) = p(\{x_{i+t}\}) \tag{3.2}$$

für alle Teilmengen $\{i\}$ und alle t gilt, d.h. wenn die Verteilungsfunktion invariant gegenüber Zeitverschiebung ist. Da bei Vorgabe der Verteilungsfunktion nicht notwendigerweise alle Momente existieren müssen, folgt aus der strengen nicht unbedingt die schwache Stationarität. Existieren aber Mittelwert und Autokovarianz für p, ist diese Schlußfolgerung erlaubt. In der linearen Zeitreihenanalyse ist meist die schwache Stationarität gemeint, wenn von Stationarität die Rede ist.

Für die Überprüfung der Stationarität einer Zeitreihe, die nur eine Realisierung des zugrundeliegenden Prozesses darstellt, müssen wieder zeitliche Mittel anstelle von Ensemblemitteln verwendet werden. Da man aber gerade die zeitliche Abhängigkeit eines Parameters P ermitteln will, darf natürlich nicht über die ganze Zeitreihe gemittelt werden. Vielmehr wird die Zeitreihe in J unter Umständen überlappende Segmente unterteilt und ein Parameter $P_j, j = 1, \ldots, J$ auf jedem Segment berechnet. Ist P_j dann im Rahmen der statistischen Fluktuationen unabhängig von j, gilt die Zeitreihe als stationär bezüglich P. Um Nichtstationarität nachzuweisen, kann man versuchen, die Segmente anhand der P_j zu klassifizieren.

3.3 Klassifikation

Die Möglichkeit, Zeitreihen zu klassifizieren, ist in vielen Bereichen sehr nützlich. Ich möchte deshalb nun präzisieren, was hier unter Klassifikation zu verstehen ist. Klassifikation ist eine Einteilung in Gruppen bzw. Klassen mit gleichen

Eigenschaften. Die erste wichtige Frage ist: Nach welchen Kriterien soll klassifiziert werden? So kann man z.B. die Bevölkerung nach Alter, Größe oder Einkommen gliedern und entsprechend verschiedene Aussagen erhalten. Natürlich können mehrere Kriterien für eine Klassifizierung herangezogen werden, wobei unter Umständen die Kriterien unterschiedlich gewichtet werden müssen. Die Klassifikation kann von dieser Gewichtung abhängen. Hier soll das Hauptaugenmerk auf die nichtlineare Dynamik gelegt werden. Als Kriterien dienen also Größen aus der Theorie des deterministischen Chaos.

Auch die Eigenschaften der Klassen können näher spezifiziert werden. Eine wichtige Entscheidung betrifft die Anzahl der Klassen. Werden zu wenige Klassen gewählt, besteht die Gefahr, innerhalb der Klassen noch stark voneinander abweichende Elemente zu finden. Bei zu vielen Klassen kann die Anzahl der Elemente in einer Klasse so klein werden, daß statistische Fluktuationen zum Tragen kommen. In beiden Fällen kann die Aussagekraft der Klassifikation verloren gehen.

Bei der Klassifizierung sind zwei Arten zu unterscheiden. Es ist nicht ungewöhnlich, daß die Anzahl und Eigenschaften der Klassen vorgegeben sind. EEGs können z.B. in die Klassen Schlaf und Wachzustand unterteilt werden. Man weiß also vorher, daß es zwei Klassen gibt und kann die Klassen genau beschreiben. Ein gegebenes, unbekanntes EEG soll dann in eine der Klassen einsortiert werden. Andererseits ist eine solche „natürliche" Vorgabe nicht selbstverständlich. Die Grenzen zwischen den Klassen sind oft unbekannt, und der Klassifizierungsalgorithmus sollte die Klassenabgrenzung wählen, die möglichst eindeutige Gruppen schafft. Ein Beispiel hierfür ist ein Stationaritätstest für einen Prozeß mit unbekanntem Parameterverlauf.

3.3.1 Realisierung

Wie wird nun eine Klassifizierung realisiert? Da Klassifizierungen in vielen Bereichen eingesetzt werden, sind sehr viele verschiedene Algorithmen entwickelt worden. Sie werden auch als *Clusteralgorithmen* bezeichnet, und oft als *neuronale Netze* realisiert. Ihnen gemeinsam ist die Minimierung einer *Kostenfunktion*, die in irgendeiner Form die Güte der Klassifikation beschreibt.

Neuronale Netze müssen meist in einer Lernphase trainiert werden, wofür in der Regel große Datenmengen notwendig sind. Dabei unterscheidet man zwischen *supervised learning* und *unsupervised learning*. Im ersten Fall erfolgt das Lernen über einen „Lehrer", der dem neuronalen Netz den Erfolg oder Mißerfolg mitteilt. Dafür muß der Lehrer *wissen*, wie die Trainingsdaten zu klassifizieren sind. Da ein solches Wissen nicht vorausgesetzt werden kann, werde ich mich hier auf Klassifizierungsmethoden ohne Lehrer konzentrieren. Eine Verallgemeinerung eines Clusteralgorithmus auf *supervised learning* ist meist einfach.

Clusteralgorithmen arbeiten entweder mit mehreren Eigenschaften der zu klassifizierenden Objekte oder mit Abständen der Objekte untereinander. Die Eigenschaften müssen sich in Werten ausdrücken lassen, was nicht immer eindeutig ist. Abstände zwischen den Objekten zu definieren, stellt sich oft auch als nicht trivial heraus. Insbesondere, wenn die Objekte nicht in einen Parameterraum eingebettet werden können.

Um Zeitreihen (oder Segmente von Zeitreihen) zu klassifizieren, werden oft ein oder mehrere Parameter für jede Zeitreihe berechnet. Eine Zeitreihe wird dadurch auf einen Punkt im m-dimensionalen Raum reduziert, wobei m die Anzahl der herangezogenen Parameter ist. Alle Eigenschaften der Zeitreihe werden auf diesen m-dimensionalen Raum projiziert. Bei dieser Projektion geht auf jeden Fall Information verloren. Bei ungeeigneter Wahl der Parameter können zwei in Wahrheit sehr verschiedene Zeitreihen auf einen Punkt fallen, womit sie ununterscheidbar werden. Praxisnäher ist der Fall, daß mehrere verschiedene Zeitreihen nach der Projektion im m-dimensionalen Raum eine „Wolke" bilden, in der keine Struktur mehr zu erkennen ist. Oft wird sogar nur *ein* bestimmter Parameter P verwendet, so daß entscheidende Informationen unberücksichtigt bleiben können. Die Verteilung des Parameters P muß für eine sinnvolle Einteilung in K verschiedene Klassen K Maxima aufweisen.

Wie erwähnt, können Clusteralgorithmen auch mit Abständen zwischen den zu klassifizierenden Objekten arbeiten. Eine völlig neue Idee ist es, *Abstände zwischen Zeitreihen* zur Klassifizierung der Zeitreihen zu verwenden (siehe auch [31]). Ich werde in Abschnitt 3.4 näher erläutern, was unter Abständen zwischen Zeitreihen zu verstehen ist und wie man sie definiert. Zunächst soll aber der von mir verwendete Clusteralgorithmus beschrieben werden.

3.3.2 Clusterbildung

Gegeben seien nun J Objekte und ihre gegenseitigen Abstandsmaße d_{ij} mit $i, j = 1, \ldots, J$. Mathematisch muß eine Abstandsfunktion folgende Bedingungen erfüllen:

1. $d_{ij} \geqslant 0$
2. $d_{ii} = 0$
3. $d_{ij} = d_{ji}$ (Symmetrie)
4. $d_{ij} \leqslant d_{ih} + d_{hj}$ (Dreiecksungleichung)

Der euklidische Abstand ist ein Beispiel, das diese Bedingungen erfüllt. Für den hier verwendeten Clusteralgorithmus sind aber nicht alle Bedingungen von Nöten. Von diesen Eigenschaften werde ich lediglich $d_{ij} \geqslant 0$ und die Symmetrie voraussetzen. Letztere ist aber kaum eine Beschränkung, da ein Abstandsmaß stets durch $d'_{ij} = \frac{1}{2}(d_{ij} + d_{ji})$ symmetrisiert werden kann. Da die d_{ij} also hier keine Abstände im mathematischen Sinne sind, werde ich sie fortan *Unähnlichkeitsmaße* (*dissimilarity measures*) nennen. Diese Bezeichnung wurde gewählt, weil für ähnliche bzw. gleichartige Objekte d_{ij} kleiner sein soll als für unähnliche bzw. verschiedenartige Objekte.

Die J Objekte sollen nun in K Cluster[1] unterteilt werden. Dafür wird zunächst die *Zugehörigkeit*

$$u_i^{(\nu)} = \begin{cases} 1 & \text{wenn Objekt } i \text{ in Cluster } \nu \\ 0 & \text{sonst} \end{cases} \tag{3.3}$$

definiert. Da jedes Objekt nur genau einem Cluster angehören kann, gilt

$$\sum_{\nu=1}^{K} u_i^{(\nu)} = 1\,. \tag{3.4}$$

Ein Cluster $C^{(\nu)} = \{i | u_i^{(\nu)} = 1\}$ ist vollständig durch seine Elemente definiert, und $|C^{(\nu)}| = \sum_{i=1}^{J} u_i^{(\nu)}$ ist die Anzahl der Elemente in Cluster ν. Ein Cluster soll möglichst gleiche Objekte enthalten, d.h. die Einteilung soll so ausgewählt werden, daß das mittlere Unähnlichkeitsmaß *innerhalb* jedes einzelnen Clusters minimal wird. Das mittlere Unähnlichkeitsmaß des Objektes i zu den Elementen des Clusters ν ist durch

$$D_i^{(\nu)} = \frac{1}{|C^{(\nu)}|} \sum_{j=1}^{J} u_j^{(\nu)} d_{ij} \tag{3.5}$$

gegeben. Damit kann das mittlere Unähnlichkeitsmaß innerhalb des Clusters ν als

$$D^{(\nu)} = \frac{1}{|C^{(\nu)}|} \sum_{i=1}^{J} u_i^{(\nu)} D_i^{(\nu)} \tag{3.6}$$

geschrieben werden. Um nun ein Maß für die mittlere Unähnlichkeit in allen Clustern zu bekommen, muß über die $D^{(\nu)}$ gemittelt werden. Dabei wird mit der Anzahl der Elemente in den Clustern gewichtet, so daß Cluster mit mehr Elementen stärker zur Summierung beitragen. Das totale, mittlere Unähnlichkeitsmaß ist also $D = (1/K) \sum_{\nu=1}^{K} |C^{(\nu)}| D^{(\nu)}$. Damit ist eine Kostenfunktion aufgestellt, die es nun bezüglich der $u_i^{(\nu)}$ zu minimieren gilt, um die optimale Klassifizierung zu finden. In Abhängigkeit der $u_i^{(\nu)}$ ist die Kostenfunktion durch

$$E_K(u_i^{(\nu)}) = \sum_{\nu=1}^{K} \frac{1}{|C^{(\nu)}|} \sum_{i,j=1}^{J} u_i^{(\nu)} u_j^{(\nu)} d_{ij} \tag{3.7}$$

gegeben, wobei der Ausdruck von führenden Konstanten befreit wurde, die für die Minimierung keine Rolle spielen.

[1] Von nun an werde ich den gebräuchlicheren Begriff Cluster für Klassen verwenden.

In der Literatur zur Clusterbildung [15] findet man viele Kostenfunktionen, die Gleichung 3.7 ähnlich sehen. Eine alternative Form wichtet die Summe mit dem Quadrat der Elemente in den Clustern. Es kann gezeigt werden, daß die Klassifizierung in diesem Fall zu Clustern mit gleicher Größe neigt, was oft nicht erwünscht ist. Eine weitere Freiheit besteht darin, das Unähnlichkeitsmaß mit einem Exponenten zu versehen.

Im oben beschriebenen Fall betrachtet man lediglich *harte* Cluster, d.h. ein Element ist genau einem Cluster zugeordnet. Im Gegensatz dazu kann man *fuzzy*[2] Cluster betrachten, bei denen die Zugehörigkeiten $u_i^{(\nu)} \in [0,1]$ kontinuierlich sind. Die Normierung (3.4) muß natürlich gültig bleiben. Das Minimieren der Kostenfunktion (3.7) mit kontinuierlichen Zugehörigkeiten führte aber in allen betrachteten Fällen zu harten Clustern, also $u_i^{(\nu)} = 0, 1$. Dies geschieht nicht, wenn die Zugehörigkeiten mit einem Exponenten größer 1 versehen werden. In [15] wird die Kostenfunktion für fuzzy Cluster mit quadratischen Zugehörigkeiten benutzt. Der Exponent 2 ist aber rein willkürlich und es spricht auch nichts gegen nichtganzzahlige Werte. Meine Untersuchungen mit verschiedenen Kostenfunktionen haben ergeben, daß mit quadrierten Zugehörigkeiten die Cluster „zu fuzzy“ sind, d.h. Elemente, die per Konstruktion in ein Cluster gehören, werden keinem Cluster zugeordnet und bekommen $u_i^{(\nu)} = 1/K$.

Für die optimale Klassifikation wird das globale Minimum der Kostenfunktion E_K gesucht. Dies wird erschwert durch die Tatsache, daß die Kostenfunktion für gewöhnlich mehrere lokale Minima aufweist. Ist die Anzahl der Cluster nicht in irgendeiner Form vorgegeben, ist es natürlich, die Klassifizierung mit verschiedenen K zu wiederholen. Hat man für K Cluster die optimale Klassifikation gefunden und erhöht die Anzahl der Cluster um 1, so kann die Kostenfunktion E_{K+1} einfach weiter minimiert werden, indem ein beliebiges Objekt dem hinzugekommenen Cluster zugeordnet wird. Es gilt also $E_K \geqslant E_{K+1}$, und für die maximale Anzahl $K = J$ der Cluster gilt $E_J = 0$. Bei Klassifizierungen mit verschiedenen K sollten also auf jeden Fall die resultierenden Cluster näher betrachtet und untersucht werden.

3.3.3 Minimierung der Kostenfunktion

Die Kostenfunktion 3.7 hängt nur von diskreten Zugehörigkeiten $u_i^{(\nu)}$ ab. Für sie soll das globale Minimum gefunden werden. Wie bereits erwähnt, gibt es aber viele lokale Minima. Die Anzahl der möglichen, *verschiedenen* Klassifikationen ist $K^J/K!$, da es jeweils $K!$ zu einander analoge Klassifikationen gibt. Das Einsetzen aller Möglichkeiten in die Kostenfunktion zur Ermittlung des globalen Minimums wäre sehr rechenaufwendig, da die Anzahl der Möglichkeiten für viele Objekte J sehr groß ist. Ähnliche Probleme treten in der Physik häufig auf. In diesen Fällen ist die Anwendung von *simulated annealing* Methoden (simu-

[2] Diese werden in der Literatur auch *weiche* Cluster genannt.

liertes Abkühlen) sehr erfolgreich. Dafür wird das Problem als thermodynamisches System betrachtet. Durch langsames Senken der Temperatur T gelangt das System in seinen Grundzustand, d.h. in den Zustand minimaler Energie. Die Kostenfunktion ist also analog zur Energie des Systems. Ein Zustand bzw. eine Konfiguration des System ist durch die Menge der Zugehörigkeiten $\{u_i^{(\nu)}\}$, d.h. durch eine Klassifikation gegeben. Gestartet wird mit einer zufälligen Konfiguration. Eine neue Konfiguration kann dadurch erzeugt werden, daß ein beliebiges Objekt aus einem Cluster entfernt und einem anderen Cluster zugeordnet wird. Die neue Konfiguration wird mit der durch die Boltzmann-Verteilung gegebenen Wahrscheinlichkeit

$$p = \begin{cases} \exp(-\frac{\Delta E}{T}) & \text{für } \Delta E > 0 \\ 1 & \text{für } \Delta E \leqslant 0 \end{cases} \tag{3.8}$$

akzeptiert. Dabei ist ΔE die Energiedifferenz $E_{neu} - E_{alt}$. Hat die neue Konfiguration geringere Energie als die vorherige, wird sie auf jeden Fall akzeptiert. Daß Konfigurationen mit höherer Energie akzeptiert werden, ist um so wahrscheinlicher je größer die Temperatur T ist. Dadurch kann das System aus einem lokalen Minimum herauskommen. Wird das System langsam abgekühlt, wird das Akzeptieren von Konfigurationen mit höherer Energie immer unwahrscheinlicher, bis es schließlich für $T = 0$ unmöglich wird. Ging die Abkühlung langsam genug vonstatten, befindet sich das System am Ende im Grundzustand mit minimaler Energie. Da bei endlicher Abkühlungsrate der Endzustand aber nicht mit dem Grundzustand übereinstimmen muß, wird zusätzlich auch während des Abkühlens der tiefste erreichte Zustand gespeichert. Die Verwendung von Gleichung (3.8) geht auf [20] zurück und wird nach ihrem Erfinder *Metropolis-Algorithmus* genannt. Ein weiteres Anwendungsbeispiel ist z.B. in [22] zu finden.

Die Anfangstemperatur sollte nicht zu hoch gewählt werden, da sonst alle neuen Konfigurationen akzeptiert werden und das System sich lediglich zufällig ändert. Bei einer zu niedrigen Anfangstemperatur kann das System frühzeitig in ein lokales Minimum gelangen. Ein guter Startwert für T kann ermittelt werden, indem einige zufällige Konfigurationen erzeugt werden. Der Wert für die Anfangstemperatur sollte auf jeden Fall über deren größter Energiedifferenz liegen.

Das Abkühlen kann auf verschiedene Arten implementiert werden. Wichtig ist nur, daß die Temperatur wirklich langsam gesenkt wird. Was unter „langsam“ zu verstehen ist, muß im Einzelfall entschieden werden. Eine Möglichkeit besteht darin, nach einer bestimmten Anzahl von Konfigurationsänderungen T durch T/k mit $k > 1$ zu ersetzen.

3.4 Abstandsmaße für Zeitreihen

Um die oben erläuterte Klassifizierungsmethode auf Zeitreihen anwenden zu können, muß ein Abstandsmaß bzw. Unähnlichkeitsmaß zwischen zwei Zeitreihen gefunden werden. Hierfür gibt es viele Möglichkeiten. Ein wichtiger ist Punkt dabei, daß das Abstandsmaß nicht über einen einzelnen, auf beiden Zeitreihen berechneten Parameter definiert wird. Dies führt gerade zu den in Abschnitt 3.3.1 beschriebenen Problemen und Einschränkungen.

Können beide Zeitreihen gut durch Modelle gleicher Art beschrieben werden, kann man versuchen über die Unterschiede in den Parametern der Modelle ein Unähnlichkeitsmaß zu definieren. Für Daten aus linearen Prozessen wäre z.B. ein Vergleich der ARMA-Koeffizienten der beiden Zeitreihen eine Möglichkeit. Es könnten auch die Kovarianzen zwischen den Werten der beiden Zeitreihen herangezogen werden. Ich will hier aber in erster Linie auf Zeitreihen mit nichtlinearer Dynamik eingehen, bei denen eine Modellanpassung mit verschiedenen Schwierigkeiten verbunden sein kann.

Viele der in Abschnitt 2.5.1 vorgestellten Teststatistiken können leicht auf ein Unähnlichkeitsmaß zwischen Zeitreihen verallgemeinert werden. Gegeben seien zwei Zeitreihen $X = \{x_n\}, \quad n = 1, \ldots, N$ und $Y = \{y_m\}, \quad m = 1, \ldots, M$ und ihre eingebetteten Vektoren $\vec{x}_n$ und $\vec{y}_m$. In [13] wird beschrieben, wie die Grassberger-Procaccia-Korrelationssumme [8] zur *Kreuzkorrelationsumme*

$$C_{XY}(\epsilon) = \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} \Theta(\epsilon - \|\vec{x}_n - \vec{y}_m\|) \tag{3.9}$$

verallgemeinert werden kann. Ist $X = Y$ ergibt sich die bekannte Korrelationssumme C_{XX} (1.15) mit dem Unterschied, daß in (1.15) Paare mit $|n - m| < t_{\min}$ nicht berücksichtigt werden. Eine naheliegende Möglichkeit, die Kreuzkorrelationssumme (3.9) zu normieren, ist die C_{XY} durch

$$C'_{XY} = \frac{C_{XY}}{\sqrt{C_{XX} C_{YY}}} \tag{3.10}$$

zu ersetzen, so daß $C'_{XX} = 1$. Um eine korrekte Normierung zu garantieren, sollten in (3.9) auch für $X \neq Y$ Paare mit $|n - m| < t_{\min}$ ausgeschlossen werden. Nun sind C'_{XY} oder $(C'_{XY})^2$ aber für Zeitreihen mit ähnlichen Attraktoren größer als für verschiedenartige. Ein Unähnlichkeitsmaß kann daraus z.B. durch

$$d_{ij} = |1 - C'_{ij}| \tag{3.11}$$

konstruiert werden. Der Absolutbetrag in diesem Ausdruck ist wichtig, da die C'_{ij} mit $i \neq j$ aufgrund von numerischen Fluktuationen durchaus Werte größer 1 annehmen können. Als Unähnlichkeitsmaß sind alternativ auch höhere Potenzen von (3.11) vorstellbar. Da die Kreuzkorrelationssumme symmetrisch in X und Y ist, ist auch das Unähnlichkeitsmaß (3.11) symmetrisch. Die Kreuzkorrelationssumme findet auch in [18] eine Anwendung in der Analyse von Nichtstationarität in Zeitreihen.

Auch Vorhersagefehler können relativ leicht verallgemeinert werden und als Unähnlichkeitsmaß zwischen Zeitreihen dienen, was ausführlich in [28] beschrieben wurde. Auf den beiden Zeitreihen X und Y können zunächst getrennt voneinander zwei Vorhersagefunktionen F_X und F_Y ermittelt werden. Machen die

Vorhersagefunktionen von Parametern Gebrauch, kann im Prinzip über Unterschiede dieser Parameter ein Unähnlichkeitsmaß definiert werden. Ein Zusammenhang zwischen den Parametern und der Dynamik der Vorhersagefunktion ist aber für gewöhnlich nur schwer herstellbar, da nicht selten verschiedene Parameter nahezu gleiches dynamisches Verhalten liefern können. Vorhersagen sind aber, wie bereits in Abschnitt 1.3.3 beschrieben, auch ohne Parameter realisierbar. Eine Verallgemeinerung des Vorhersagefehlers (2.5) ist der *Kreuzvorhersagefehler*

$$\sigma_{XY} = \left(\frac{1}{M-1} \sum_m [y_{m+1} - F_X(y_m)]^2 \right)^{1/2} , \tag{3.12}$$

wobei die Vorhersage F_X auf die Zeitreihe Y angewendet wird. Zur Wahl von F gibt es wieder verschiedene Alternativen, und ich werde wie im vorangegangen Kapitel die lokal konstante Vorhersage verwenden. Im allgemeinen ist der Kreuzvorhersagefehler σ_{XY} nicht symmetrisch und muß für den oben beschriebenen Clusteralgorithmus z.B. durch

$$d_{ij} = \frac{(\sigma_{ij} + \sigma_{ji})^2}{4\sigma_{ii}\sigma_{jj}} \tag{3.13}$$

symmetrisiert werden. Dabei wurde in (3.13) auch gleich auf $d_{ii} = 1$ normiert.

Eine weitere Möglichkeit ein Unähnlichkeitsmaß zu definieren, besteht in einer Verallgemeinerung des Lyapunovexponenten. Man betrachtet in diesem Fall zwei benachbarte Trajektorien aus den beiden betrachteten Zeitreihen und deren Entwicklung mit der Zeit. Die Charakterisierung der zeitlichen Entwicklung des Abstandes der Trajektorien ist aber nicht immer einfach, da die zugrundeliegende Dynamik der Zeitreihen sehr verschieden sein kann.

3.5 Beispiele

Die Verwendung des beschriebenen Clusteralgorithmus soll in diesem Abschnitt an einigen Beispielen illustriert werden. Dafür werden Zeitreihen mit unterschiedlicher Dynamik erzeugt und kontrolliert, ob sie entspechend klassifiziert werden können.

In den ersten Beispielen werden Zeitreihen verwendet, die auf einer generalisierten *Bäcker*-Abbildung der Form

$$\left.\begin{aligned} u_{n+1} &= \beta u_n \\ v_{n+1} &= v_n/\alpha \end{aligned}\right\} \quad \text{für } v_n \leqslant \alpha$$
$$\left.\begin{aligned} u_{n+1} &= 1/2 + \beta u_n \\ v_{n+1} &= \tfrac{v_n - \alpha}{1-\alpha} \end{aligned}\right\} \quad \text{für } v_n > \alpha \tag{3.14}$$

basieren, wobei die Parameter α und β Werte aus dem Intervall $(0,1)$ annehmen können. Diese Abbildung ist stückweise linear und ihre Lyapunovexponenten können analytisch zu

$$\begin{aligned} \lambda_1 &= \alpha \ln \tfrac{1}{\alpha} + (1-\alpha) \ln \tfrac{1}{1-\alpha} \\ \lambda_2 &= \ln \beta \end{aligned} \tag{3.15}$$

berechnet werden [32]. Eine Veränderung von β bei festem α führt also zu Prozessen mit unterschiedlicher Dynamik, aber gleichem maximalem Lyapunovexponenten λ_1.

Um aus der Bäcker-Abbildung (3.14) eine Zeitreihe $\{s_n\}$ zu erzeugen, wird die Summe $u_n + v_n$ aufgenommen, das laufende Mittel subtrahiert und auf eine laufende Varianz von 1 normiert, das heißt, es wird

$$s_n = \frac{w_n - \langle w \rangle_k}{\sqrt{\langle (w_n - \langle w \rangle_k)^2 \rangle_k}}, \quad w_n = u_n + v_n \tag{3.16}$$

gebildet, wobei durch $\langle \cdot \rangle_k$ die Mittelung über die Indizes $n-k$ bis $n+k$ ausgedrückt wird. Diese Konstruktion verhindert, daß die Klassifizierung der Zeitreihen über Unterschiede im Mittelwert oder der Varianz erzielt werden kann.

Als erstes Beispiel habe ich 50 Zeitreihen mit $\beta = 0.6$ und 50 Zeitreihen mit $\beta = 0.8$ erzeugt. Der Parameter α wurde konstant auf einem Wert von $\alpha = 0.4$ gehalten. Jede Zeitreihe bestand aus $N = 400$ Werten, und bei der Berechnung des laufenden Mittels in (3.16) wurde $k = 50$ gewählt. Die Klassifizierung dieser Zeitreihen wird dadurch erschwert, daß viele Parameter kaum variieren. Der Mittelwert und die Varianz sind per Konstruktion bis auf Fluktuationen gleich. Aber auch die Autokorrelationen unterscheiden sich nur in geringem Maße. Da bei der Erzeugung der Zeitreihen α konstant gehalten wurde, sind auch die maximalen Lyapunovexponenten gleich.

Um den Clusteralgorithmus anzuwenden, wurden zunächst die Kreuzvorhersagefehler σ_{ij} zwischen den Zeitreihen berechnet. Dafür wurden die Zeitreihen in $m = 2$ Dimensionen eingebettet und bei der Vorhersage über Nachbarn innerhalb eines Radius von $\epsilon = 0.25$ gemittelt. Die Unähnlichkeitsmaße d_{ij} sind über Gleichung (3.13) aus den Kreuzvorhersagefehlern gebildet worden. Die Kostenfunktion E (3.7) wurde mit dem beschriebenen Metropolis-Algorithmus für zwei Cluster ($K = 2$) minimiert. Mehrere Wiederholungen mit verschiedenen zufälligen Anfangskonfigurationen führten stets zum gleichen Minimum, so daß angenommen werden kann, daß es sich tatsächlich um das globale Minimum handelt. Um die resultierende Klassifikation anschaulich darzustellen, sind in Abbildung 3.1 die mittleren Unähnlichkeitsmaße $D_i^{(\nu)}$ der einzelnen Zeitreihen zu den beiden Clustern gegeneinander aufgetragen. Der Clusteralgorithmus kann erfolgreich die Zeitreihen mit unterschiedlichem β in zwei Cluster einteilen. Da für Zeitreihen, die Cluster $\nu = 1$ zugeordnet wurden, $D_i^{(1)} < D_i^{(2)}$ gilt, sind diese in Abbildung 3.1 in der Gruppe links oben zu finden.

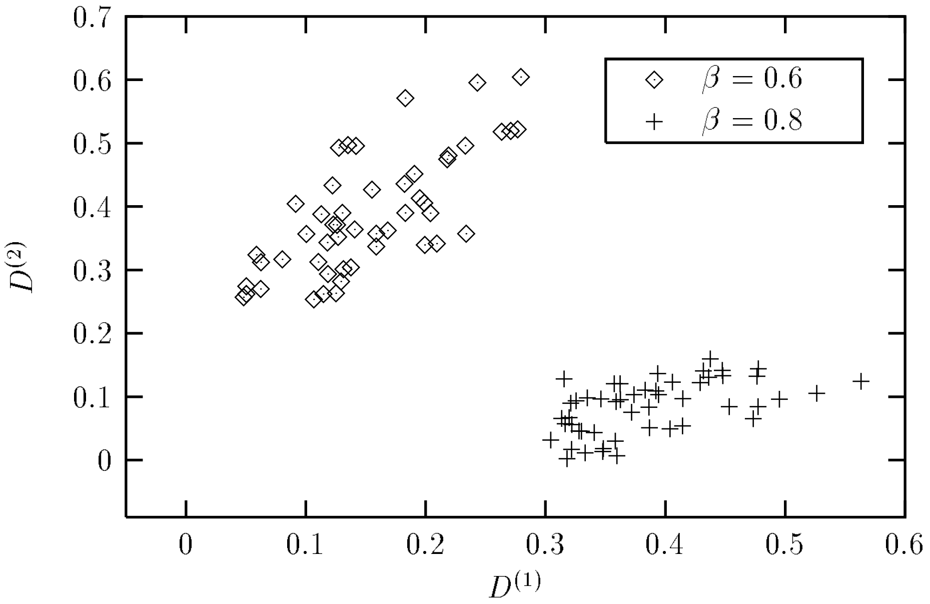

Abbildung 3.1: Mittlere Unähnlichkeitsmaße $D_i^{(\nu)}$ von 100 Zeitreihen gegeneinander aufgetragen. Deutlich zu sehen sind zwei Cluster, die den beiden Werten für β der Bäcker-Abbildung entsprechen.

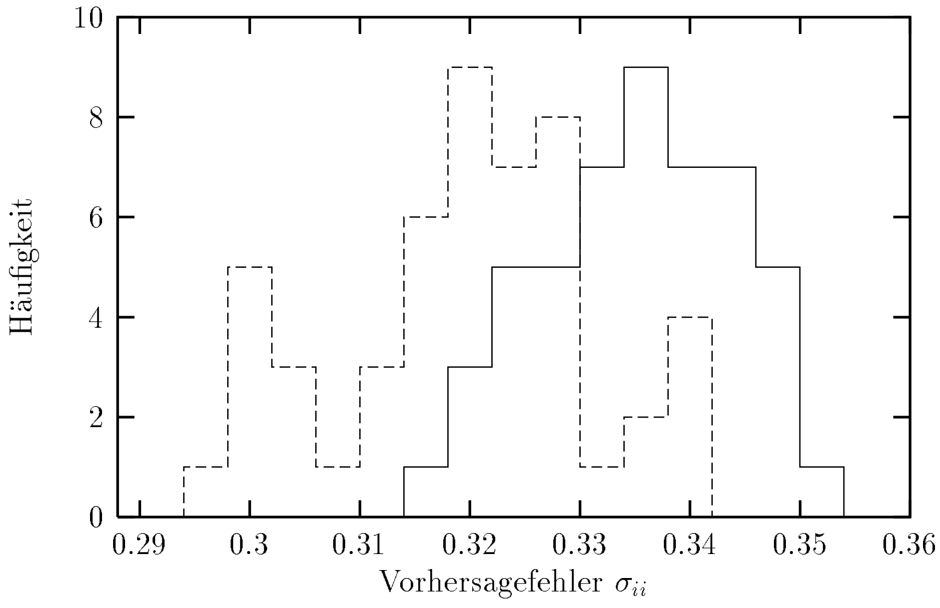

Abbildung 3.2: Verteilung der Vorhersagefehler σ_{ii} für jeweils 50 Zeitreihen, die über die Bäcker-Abbildung mit $\beta = 0.6$ (gestrichelt) und $\beta = 0.8$ (durchgezogen) erzeugt wurden.

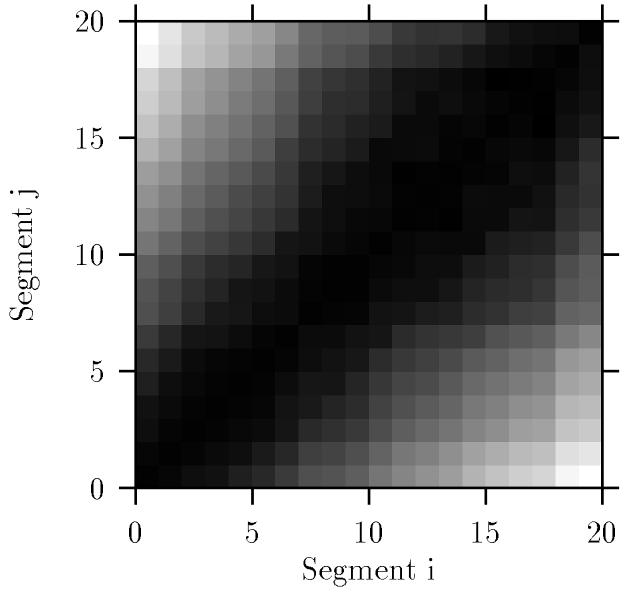

Abbildung 3.3: Normierte Kreuzkorrelationssummen C'_{ij} zwischen Segmenten einer Zeitreihe, die auf einer Bäcker-Abbildung basiert, deren Parameter β kontinuierlich variiert wurde. Die Graustufen geben linear die Werte von $C'_{ij} \approxeq 0.45$ (weiß) bis $C'_{ij} \approxeq 1.03$ (schwarz) wieder.

Bei der Bildung der Unähnlichkeitsmaße wurde auf konstante d_{ii} (3.13) normiert. Diese Wahl wurde getroffen, um zu demonstrieren, daß auch ohne die Information in den Vorhersagefehlern σ_{ii} der einzelnen Zeitreihen eine deutliche Klassifizierung erzielt werden kann. Üblicherweise werden beim Vergleich von Zeitreihen die Werte bestimmter Parameter verglichen, die für jede einzelne Zeitreihe berechnet werden. Um den Vorteil der Verwendung von Unähnlichkeitsmaßen weiter zu verdeutlichen, habe ich versucht, anhand der Vorhersagefehler σ_{ii} die Zeitreihen mit $\beta = 0.6$ von denen mit $\beta = 0.8$ unterscheiden zu können. In Abbildung 3.2 sind die Verteilungen der Vorhersagefehler σ_{ii} für beide β-Werte in Form von Histogrammen aufgetragen. Liegt der Vorhersagefehler einer Zeitreihe im mittleren Bereich der in Abbildung 3.2 aufgetragenen Werte, kann die Zeitreihe nicht eindeutig einem der β-Werte zugeordnet werden. Wird für die Klassifizierung ein bestimmter Vorhersagefehler als Grenze zwischen den Clustern festgelegt, gibt es stets Zeitreihen, die nicht dem richtigen Cluster zugeordnet werden und damit zu einer falschen Klassifikation führen.

Auch im zweiten Beispiel wird die Bäcker-Abbildung zur Erzeugung der Zeitreihen verwendet. Im Unterschied zum ersten Beispiel wird nun eine lange Zeitreihe $\{s_n\}$ mit $n = 1, \ldots, N = 40000$ generiert, wobei der Wert des Parameters β kontinuierlich von 0 auf 1 erhöht wird, also $\beta = n/N$. Der Parameter α wurde wieder auf $\alpha = 0.4$ gesetzt und nicht variiert. Die resultierende Zeitreihe

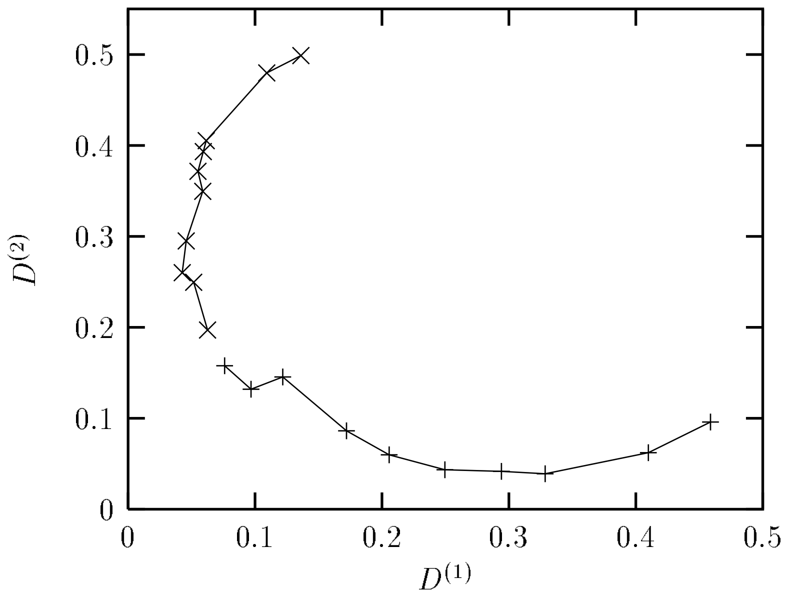

Abbildung 3.4: Mittlere Unähnlichkeitsmaße $D_i^{(\nu)}$ von 20 Zeitreihensegmenten zu zwei Clustern $\nu = 1, 2$. Die Zugehörigkeit der Segmente zu den beiden Clustern wird durch die unterschiedlichen Symbole dargestellt. Der Parameter β der zugrundeliegenden Bäcker-Abbildung änderte sich kontinuierlich von 0 bis 1.

ist bis auf Fluktuationen schwach stationär. Wie im vorangegangenen Beispiel ist sie auch bezüglich anderer Parameter stationär. Werden diese Parameter zur Untersuchung gewählt, kann die Zeitreihe fälschlicherweise für stationär gehalten werden. In [28] wird für die gleiche Zeitreihe gezeigt, wie über Kreuzvorhersagefehler die Nichtstationarität nachgewiesen werden kann. Ich zerlege hier die Zeitreihe in 20 Segmente mit jeweils 2000 Werten und berechne die gegenseitigen Kreuzkorrelationssummen. Dafür wurde die Zeitreihe wieder in $m = 2$ Dimensionen eingebettet und $\epsilon = 0.25$ gewählt. Die Werte der über (3.10) normierten Kreuzkorrelationssummen C'_{ij} sind in Abbildung 3.3 als Graustufen dargestellt. Diese Darstellung kann als eine Verallgemeinerung der *recurrence plots* aus [6] angesehen werden. Ein deutliches Abfallen der C'_{ij} für große Abstände $|i - j|$ ist zu beobachten, wodurch die Annahme der Stationarität widerlegt wird. Dieses visuell gewonnene Resultat kann durch den in dieser Arbeit beschriebenen Clusteralgorithmus quantitativ bestätigt werden. Dafür werden zunächst über (3.11) aus den Kreuzkorrelationssummen Unähnlichkeitsmaße d_{ij} berechnet, die für eine Klassifizierung benötigt werden. Durch Minimierung der Kostenfunktion E mit $K = 2$ werden die Segmente der Zeitreihe in 2 Cluster unterteilt.

Die Klassifikation ist in Abbildung 3.4 wie schon in Abbildung 3.1 durch gegenseitiges Auftragen der mittleren Unähnlichkeitsmaße $D_i^{(\nu)}$ der Zeitreihen zu den beiden Clustern dargestellt. Die Segmente der Zeitreihe bilden in dieser Dar-

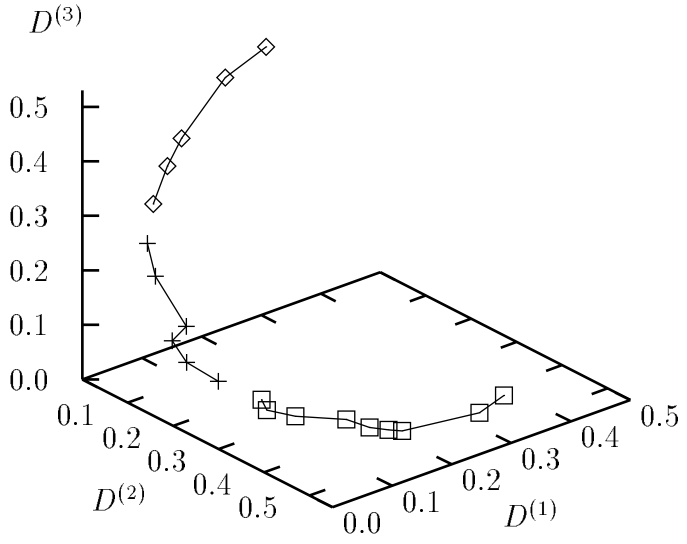

Abbildung 3.5: Die gleiche Darstellung wie in Abbildung 3.4, mit dem Unterschied, daß hier drei Cluster gebildet wurden.

stellung einen „Pfad“, was ein deutliches Zeichen für einen kontinuierlichen Parameterverlauf ist. Die Zeitreihensegmente sind dabei in der zeitlichen Reihenfolge auf dem Pfad angeordnet. Ein ähnliches Bild zeigt sich bei einer Klassifizierung in drei Cluster (siehe Abbildung 3.5). Sowohl in Abbildung 3.4 als auch in Abbildung 3.5 fällt auf, daß die Cluster nicht die gleiche Anzahl von Elementen besitzen, obwohl β linear erhöht wurde. Die Cluster, die kleineren β-Werten entsprechen, enthalten jeweils mehr Elemente. Daran wird deutlich, daß Variationen von β bei kleinen β-Werten über die Kreuzkorrelationssummen einfacher erkannt werden können als Variationen bei großen β-Werten.

Um die Aussagekraft der vorangegangen Beispiele beurteilen zu können, wird im folgenden eine Zeitreihe untersucht, deren Dynamik sich über den betrachteten Zeitraum nicht ändert. Als Zeitreihe dienten dabei $N = 32000$ x-Werte der Hénon-Abbildung (2.9) mit den üblichen Parametern. Zur Berechnung der Kreuzkorrelationssummen wurde die Zeitreihe in 40 Segmente unterteilt und in $m = 2$ Dimensionen eingebettet. Der maximale Abstand ϵ zweier Punkte im Phasenraum, bis zu dem Paare in der Korrelationssumme mitgezählt werden, wurde zu $\epsilon = 0.3$ gewählt. Die Werte der resultierenden, normierten Kreuzkorrelationssummen lagen aufgrund von endlichen Fluktuationen zwischen $C'_{ij} = 0.96$ und $C'_{ij} = 1.015$. Bei der Minimierung der Kostenfunktion E wurde auch hier sehr schnell das globale Minimum gefunden. Bei mehreren Wiederholungen zeigten

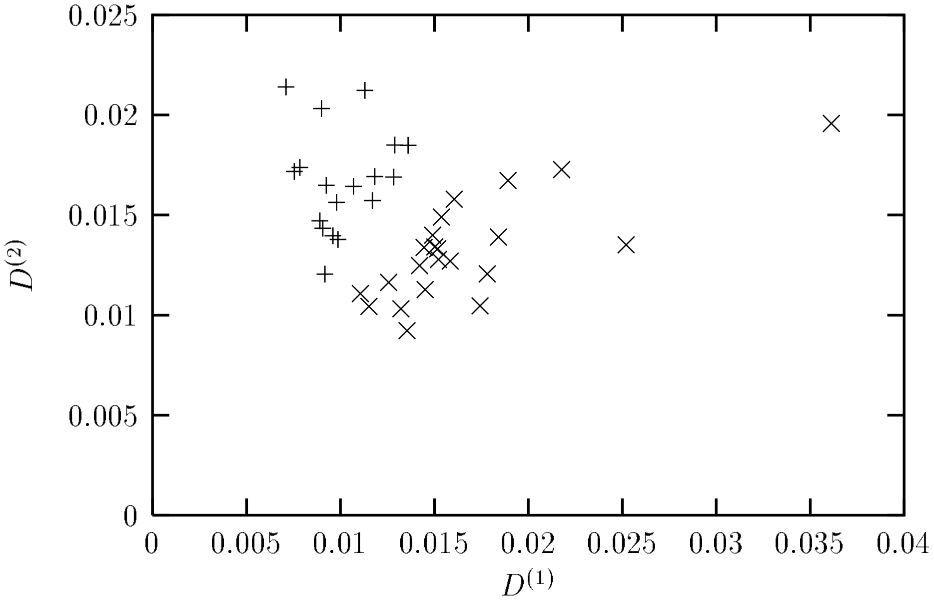

Abbildung 3.6: Mittlere Unähnlichkeitsmaße $D_i^{(\nu)}$ von 40 Segmenten einer Hénon-Zeitreihe gegeneinander aufgetragen. Die unterschiedlichen Symbole geben die Zugehörigkeiten zu den beiden Clustern wieder.

sich allerdings einige Fälle, bei denen sich am Ende Konfigurationen einstellten, die lokale Minima bildeten. Obwohl dieses Problem durch noch langsameres Senken der Temperatur verhindert werden konnte, ist dies schon ein Hinweis auf ein schwächer ausgeprägtes, globales Minimum als in den vorherigen Beispielen. Die optimale Klassifikation ist in Abbildung 3.6 dargestellt. Im Vergleich zu Abbildung 3.1 fällt auf, daß hier die Cluster viel dichter zusammen liegen. Im Falle der Bäcker-Abbildung ist es möglich, die Cluster sowohl durch die Gerade $D^{(1)} \approxeq 0.3$ als auch durch die Gerade $D^{(2)} \approxeq 0.2$ voneinander zu trennen. Im Gegensatz dazu gibt es hier viele Zeitreihensegmente, deren mittleres Unähnlichkeitsmaß zum zugeordneten Cluster größer ist, als das mittlere Unähnlichkeitsmaß einiger Elemente des anderen Clusters. Weiter ist in Abbildung 3.6 zu sehen, daß der „mittlere Abstand" der Cluster kleiner als die Ausdehnung der Cluster selbst ist. Von einer eindeutigen Trennung kann also hier nicht die Rede sein. Ein weiterer, wichtiger Hinweis auf die Stationarität der Zeitreihe, ist die Art und Weise, wie die Zeitreihensegmente den Clustern zugeordnet wurden. Die Zugehörigkeiten der zeitlich sortierten Segmente zu den Clustern $\nu = 1, 2$ ist durch

1,2,2,2,1,1,2,2,2,1,1,1,2,1,2,2,2,2,2,1,2,1,2,1,2,2,1,2,1,1,1,2,1,1,2,1,2,1,2,2

gegeben. Die Zugehörigkeiten sehen zufällig aus und zeigen kein erkennbares System. Wird eine ähnliche Klassifizierung für eine gemessene Zeitreihe gefunden, kann es sich entweder um eine auf kleinen Zeitskalen stark nichtstationäre Zeitreihe oder um eine über den Meßzeitraum stationäre Zeitreihe handeln. Die erste Annahme kann durch weitere Untersuchungen mit kürzeren Segmenten[3] als unwahrscheinlich verworfen werden, wenn sich auch dann keine Struktur in den Zugehörigkeiten zeigt.

Als letztes Beispiel für die Anwendung der Klassifikation von Zeitreihen, werden ungekoppelte Abbildungen betrachtet. Eine Zeitreihe $\{s_n\}$ wird durch Mittelung über M *Ulam*-Abbildungen

$$x_{n+1,m} = 2 - x_{n,m}^2 \,, \quad s_n = \sum_{m=1}^{M} x_{n,m} \tag{3.17}$$

gewonnen. Für jedes $M = 1, \ldots, M_{\max}$ werden 10 Zeitreihen mit $N = 1000$ Werten erzeugt, die zusätzlich auf eine Varianz von 1 umskaliert wurden. Zur Klassifizierung werden zwischen den Zeitreihen Kreuzvorhersagefehler mit einer Einbettungsdimension von $m = 3$ berechnet. Der Clusteralgorithmus war bis zu einem Wert von $M_{\max} = 19$ in der Lage, alle 190 Zeitreihen korrekt den 19 Clustern zuzuordnen. Erst für $M_{\max} = 20$ kam es zu den ersten 15 falschen Zuordnungen der insgesamt 200 Zeitreihen.

Werden 20 Zeitreihen erzeugt, wovon die eine Hälfte durch Mittelung über M Abbildungen erzeugt wurde und die andere Hälfte durch Mittelung über $M - 1$ Abbildungen, konnte der Algorithmus sogar bis $M = 30$ die Zeitreihen korrekt in zwei Cluster unterteilen.

3.6 Anwendungen

Nachdem im vorangegangen Abschnitt die Anwendung und Funktionstüchtigkeit des Clusteralgorithmus anhand von konstruierten Beispielen demonstriert wurde, sollen nun einige praktische Anwendungen beschrieben werden. Sie sollen einen ersten Einblick in mögliche Anwendungsgebiete geben und werden hier nur ansatzweise behandelt.

3.6.1 Neuer Surrogattest

Zunächst will ich zeigen, wie die Klassifikation von Zeitreihen für die in Kapitel 2 ausführlich beschriebenen Surrogattests verwendet werden kann. In diesem Fall wird die Nullhypothese verworfen, wenn es dem Clusteralgorithmus gelingt, den Originaldatensatz von den B Surrogaten zu trennen. Um die korrekte Size α zu garantieren, können die $B + 1$ Zeitreihen aber nicht einfach in zwei Cluster unterteilt werden. Die Wahrscheinlichkeit, daß ein Datensatz, der die Nullhypothese erfüllt, ein einzelnes Cluster bildet, ist geringer als die vorgegebene Size

[3] Damit die Segmente auch dann noch genügend Werte für sinnvolle Analysen enthalten, muß unter Umständen die Datenaufnahmerate erhöht werden.

$\alpha = 1/(B+1)$, da der Datensatz mit einigen Surrogaten zu einem Cluster zusammengefaßt werden könnte. Die Größe der beiden Cluster muß deshalb auf 1 und B festgelegt werden. Die Minimierung der Kostenfunktion E geschieht also unter einer Nebenbedingung. Die Anwendung des Metropolis-Verfahrens ist dabei nicht ratsam, da das Minimum der Kostenfunktion E schneller durch Einsetzen der verbleibenden $B+1$ möglichen Konfigurationen ermittelt werden kann.

Die ersten Tests wurden mit Zeitreihen aus 2048 x-Werten der Hénon-Abbildung (2.9) durchgeführt. Es wurden jeweils 9 Surrogate mit dem in Abschnitt 2.3.4 beschriebenen Verfahren erzeugt, was einem Signifikanzniveau von 90% entspricht. Zur Clusterbildung wurden Kreuzkorrelationssummen mit einer Einbettungsdimension von $m = 3$ zwischen den Originaldaten und den Surrogaten berechnet. Die Nullhypothese konnte bis zu einem relativen Rauschniveau von $a = 1.1$ mit einer Wahrscheinlichkeit größer als 95% verworfen werden. Ein Vergleich mit den Werten aus Tabelle 2.2 zeigt, daß sich in diesem Fall mit der Klassifizierung über die Kreuzkorrelationssumme keine Verbesserung der Trennschärfe des Tests erzielen läßt.

In einer zweiten Testreihe verwendete ich Kreuzvorhersagefehler und die NMR-Laser-Daten [7], um auch hier die Ergebnisse mit denen aus Abschnitt 2.5.3 vergleichen zu können. Die Trennschärfe wurde ermittelt, indem 600 Tests mit jeweils 1000 Werten des langen Datensatzes mit 80% additivem Rauschen durchgeführt wurden. Für die Kreuzvorhersagefehler zwischen den Daten und den 9 Surrogaten wurde eine Einbettungsdimension von $m = 3$ gewählt. Als Unähnlichkeitsmaße dienten die symmetrisierten Kreuzvorhersagefehler $d_{ij} = \sigma_{ij} + \sigma_{ji}$. Es wurde nicht wie in (3.13) auf konstante d_{ii} normiert, um die Information in den Vorhersagen σ_{ii} im Clusteralgorithmus mitzuberücksichtigen. In den 600 Tests kam es zu 434 Verwerfungen, was einer Trennschärfe von 0.72 ± 0.02 entspricht. Dies liegt erstaunlicherweise unter dem in Tabelle 2.4 angegebenen Wert für den Vorhersagefehler, obwohl die Kreuzvorhersagefehler die Vorhersagefehler beinhalten. Tests mit gaußverteilten Zufallszahlen ergaben eine Size von $a = 0.11 \pm 0.01$, so daß eine Senkung der Trennschärfe aufgrund einer Abweichung von der nominellen Size wahrscheinlich nicht in Frage kommt. Die gewählten Parameter m und ϵ für die Kreuzvorhersagen entsprechen den Werten, die für Tests mit dem Vorhersagefehler (2.5) die maximale Trennschärfe lieferten. Dies müssen aber nicht die idealen Parameter für die Kreuzvorhersagefehler sein. In [31] wurden Tests mit anderen Parametern durchgeführt, die mit Vorhersagefehlern lediglich in $61.5\% \pm 2.5\%$ der Fälle die Nullhypothese erfolgreich verwerfen konnten. Tests mit Kreuzvorhersagefehlern und den gleichen Parametern lieferten eine Trennschärfe von 0.69 ± 0.02. Dies zeigt, daß für die Kreuzvorhersagefehler die Abhängigkeit der Trennschärfe von den Parametern erneut untersucht werden muß. Es ist zu erwarten, daß durch Variation speziell von ϵ höhere Trennschärfen zu erreichen sind.

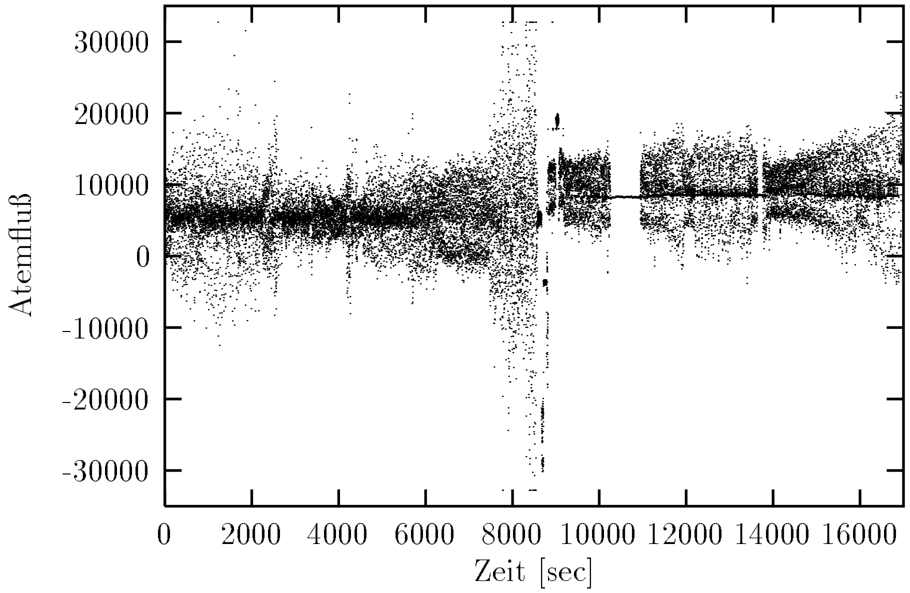

Abbildung 3.7: Die Atemrate einer Versuchsperson während einer Nacht.

3.6.2 Anwendung auf physiologische Daten

Als eine erste Anwendung auf gemessene, physiologische Daten soll die Atemrate einer Testperson untersucht werden. Über einen Zeitraum von ungefähr fünf Stunden wurde an einem Patienten im Schlaf zweimal in der Sekunde der Atemfluß durch die Nase gemessen. Diese Zeitreihe ist Bestandteil von Datensatz B, der vom Santa Fe Institute anläßlich eines Zeitreihenwettbewerbes ausgegeben wurde. Weitere Informationen zu diesen Daten sind in [25] zu finden. Es ist nicht zu erwarten, daß die Atemrate während einer ganzen Nacht stationär bleibt und in Abbildung 3.7 sind auch erhebliche qualitative Unterschiede in verschiedenen Zeitabschnitten zu sehen. Veränderungen des Mittelwertes und der Varianz sind leicht zu erkennen und auch Analysen der Autokorrelationen zeigen starke Schwankungen. Um eventuelle Wechsel in der nichtlinearen Dynamik zu untersuchen, wurden Kreuzvorhersagefehler herangezogen. Dafür wurden zunächst die Werte von 7500–9250 Sekunden und von 10250–11000 Sekunden entfernt, da diese Bereiche sehr wahrscheinlich einzelne Cluster bilden würden, die von den anderern isoliert wären. Im ersten entfernten Bereich stoßen die Meßwerte an die Grenze des Meßbereiches und eine nähere Betrachtung des zweiten Bereichs zeigt relative Nichtstationaritäten auf großen Zeitskalen. Die verbleibenden 29000 Werte wurden in 29 Segmente unterteilt, welche einzeln auf eine Verteilung mit Mittelwert 0 und Varianz 1 umskaliert wurden. In [28] wird gezeigt, daß die Vorhersagefehler auf den einzelnen Segmenten nur leicht variieren und nicht für eine eindeutige Unterscheidung der Abschnitte herangezogen werden können. Um den Clusteral-

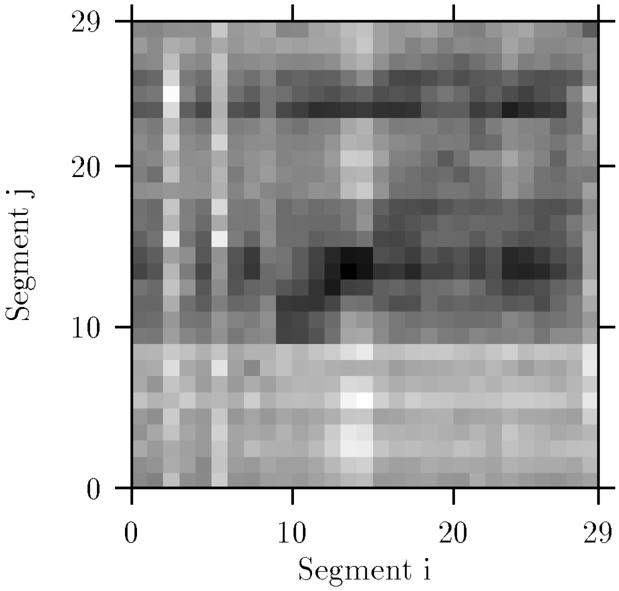

Abbildung 3.8: Kreuzvorhersagefehler σ_{ij} zwischen 29 Segmenten der Atemraten-Zeitreihe. Die Werte liegen zwischen $\sigma_{ij} \approxeq 0.41$ (schwarz) und $\sigma_{ij} \approxeq 1.11$ (weiß).

gorithmus anzuwenden, wurden Kreuzvorhersagefehler mit einer Einbettungsdimension von $m = 2$ und einem Nachbarradius von $\epsilon = 0.25$ berechnet. Sie sind in Abbildung 3.8 als Graustufen dargestellt, wobei ein deutlicher Wechsel zwischen Segment 9 und 10 zu sehen ist. Über Gleichung (3.13) wurden Unähnlichkeitsmaße d_{ij} berechnet und die Kostenfunktion E mit unterschiedlicher Anzahl von Clustern K minimiert. Die resultierenden Cluster sind in Tabelle 3.1 für $K = 2$ und $K = 3$ angegeben. Klassifizierungen mit mehr als drei Clustern führen lediglich zu einer weiteren Unterteilung der für $K = 3$ gebildeten Cluster. Dies läßt vermuten, daß im wesentlichen drei unterschiedliche Formen der Dynamik vorliegen. In Abbildung 3.9 sind die entsprechenden, mittleren Unähnlichkeitsmaße $D_i^{(\nu)}$ für $\nu = 1, 2, 3$ gegeneinander aufgetragen, wobei jeweils auf zwei der $D_i^{(\nu)}$ projiziert wurde. Aber auch in den Projektionen sind bereits drei voneinander getrennte Cluster zu sehen. Wie im Falle der stationären Hénon-Zeitreihe (siehe Abbildung 3.6) ist auch hier der Abstand der Cluster von der Größenordnung der Ausdehnung der Cluster. Der wesentliche Unterschied liegt darin, daß hier die zeitlich sortierten Segmente in sieben „Blöcken“ den Clustern zugeordnet wurden und somit eine deutliche Struktur in den Zugehörigkeiten erkennbar ist.

2 Cluster		3 Cluster	
Segmente	Cluster	Segmente	Cluster
1-11	1	1-9	1
12-18	2	10-12	2
19-23	1	13-15	3
24-26	2	16-23	2
27-29	1	24-25	3
		26-27	2
		28-29	1

Tabelle 3.1: Klassifikation der 29 Segmente der Atemraten-Zeitreihe für $K = 2, 3$ Cluster.

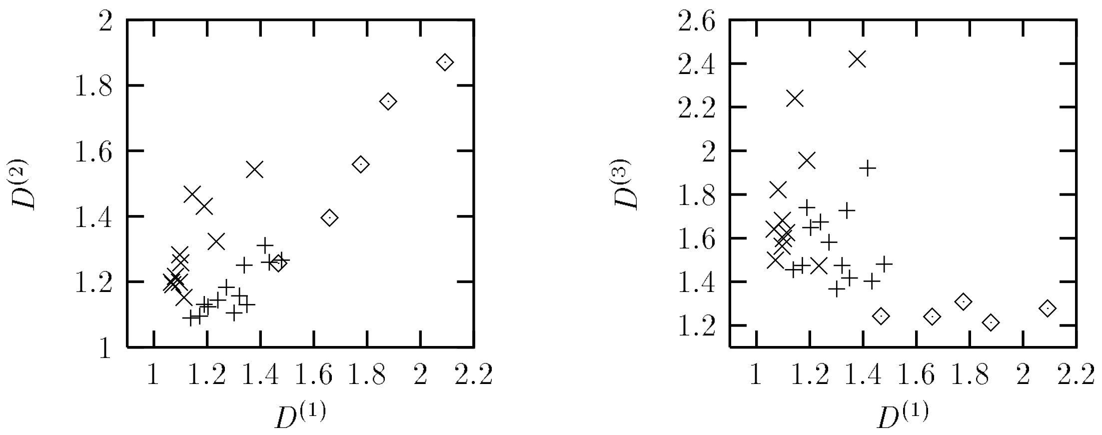

Abbildung 3.9: Mittlere Unähnlichkeitsmaße $D_i^{(\nu)}$ der 29 Segmente der Atemraten-Zeitreihe gegeneinander aufgetragen. Die unterschiedlichen Symbole geben die Zugehörigkeiten zu den drei Clustern an.

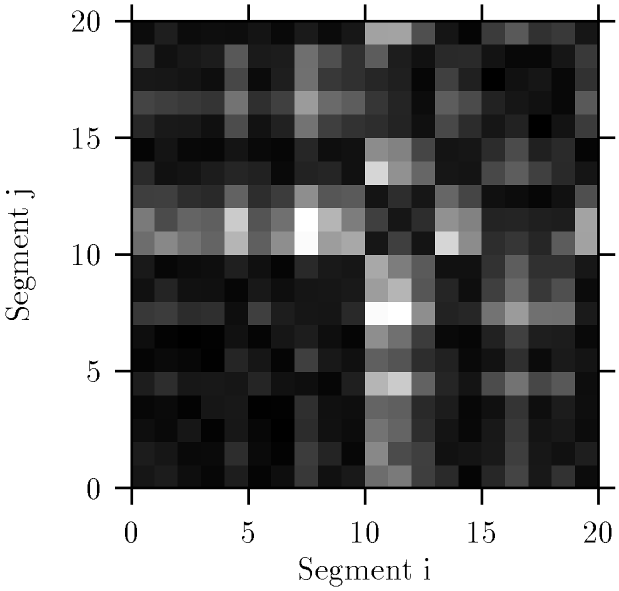

Abbildung 3.10: Normierte Kreuzkorrelationssummen C'_{ij} zwischen Segmenten eines EEGs vor einem epileptischen Anfall. Die Graustufen geben linear die Werte von $C'_{ij} \approxeq 0.89$ (weiß) bis $C'_{ij} \approxeq 1.01$ (schwarz) wieder.

Die letzte, aber wahrscheinlich interessanteste Anwendung, die ich aufführen will, ist die Vorhersage von epileptischen Anfällen. Es wird allgemein angenommen, daß im EEG eines Patienten bereits vor einem epileptischen Anfall Anzeichen dafür erkannt werden können. Patienten, bei denen eine Operation, d.h. in vielen Fällen Entfernen von Gehirnteilen, ein zu großes Risiko bedeuten würde, müssen mit ihren Anfällen leben. Für diese Patienten kann die Vorhersage eines Anfalls im täglichen Leben sehr nützlich sein. So könnte z.B. eine Autofahrt unternommen werden, die im Falle eines Anfalls unterbrochen wird. Dies ist vielen Patienten nicht möglich, da Anfälle plötzlich und ohne vorherige Anzeichen auftreten können. Einige Patienten sind sogar in der Lage bei Kenntnis eines nahenden Anfalls, diesen z.B. durch Wechseln der Tätigkeit zu verhindern. Dadurch motiviert, arbeiten einige Gruppen intensiv an der Vorhersage von epileptischen Anfällen. Mit LENA, Faculté de Medicine Pitié-Salpêtrière in Paris unter der Leitung von F. Varela und der Epileptologischen Klinik, Uni-Klinikum Bonn unter der Leitung von K. Lehnertz und C.E. Elger konnten wir auch interessante Kontakte knüpfen.

Die hier verwendeten Daten wurden von der Gruppe aus Paris zur Verfügung gestellt. Die EEGs von mehreren Patienten wurden vor, während und nach einem Anfall mit vier Elektroden in der Nähe des Fokus aufgenommen. Die im folgenden Beispiel verwendete Zeitreihe besteht aus 40000 Werten des zweiten Kanals unmittelbar vor dem Anfall eines Patienten. Dies entspricht einem Zeitraum von 200

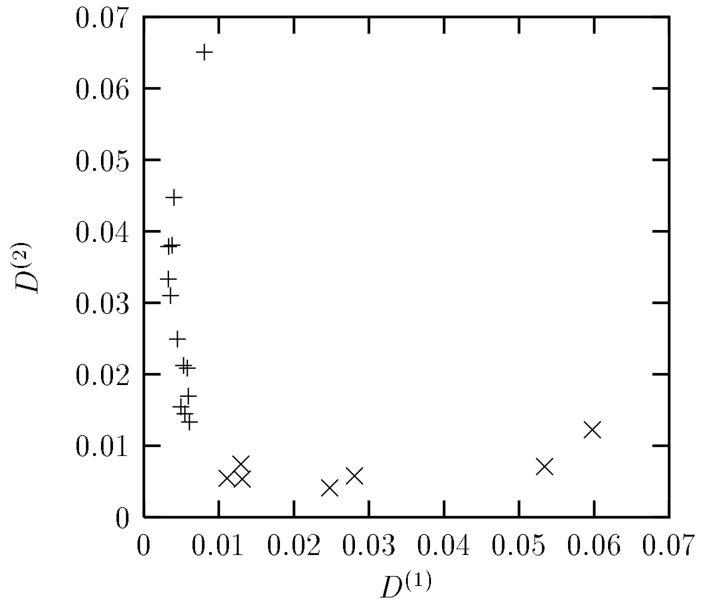

Abbildung 3.11: Mittlere Unänhnlichkeitsmaße $D_i^{(\nu)}$ von 20 Segmenten eines EEGs vor einem epileptischen Anfall gegeneinander aufgetragen.

Sekunden. Zur Untersuchung eventueller Nichtstationarität des EEGs in diesem Zeitraum, wurde die Zeitreihe in 20 Segmente unterteilt und Kreuzkorrelationssummen mit $m = 2$ und einem ϵ, welches ungefähr 2/3 der Standardabweichung entspricht, berechnet. Die Segmente wurden dabei nicht wie in den vorherigen Beispielen auf den gleichen Mittelwert oder die gleiche Varianz umskaliert, da hier Veränderungen jeglicher Art von Interesse sind. Schon anhand der in Abbildung 3.10 dargestellten Kreuzkorrelationssummen ist ungefähr 100 Sekunden vor dem Anfall eine Veränderung zu erkennen. Um den Clusteralgorithmus anzuwenden wurden wieder über (3.11) Unähnlichkeitsmaße d_{ij} berechnet. Bei einer Klassifizierung mit zwei Clustern bildeten die Segmente 1-10, 14, 15 und 20 ein Cluster und die restlichen Segmente das andere Cluster. Die entsprechenden, mittleren Unähnlichkeitsmaße $D_i^{(\nu)}$ sind in Abbildung 3.11 dargestellt, wobei eine deutliche Trennung der Cluster sichtbar ist.

Die relativ deutliche Nichtstationarität in diesem Beispiel ist aber nicht typisch. In den EEGs anderer Patienten sind in einigen Fällen keine Anzeichen für einen Anfall sichtbar. In anderen Fällen tritt eine Änderung der Varianz der Werte des EEGs vor dem Anfall auf. Auch die Häufigkeit der Spitzen, sogenannter *Spikes*, kann einen Hinweis auf einen bevorstehenden Anfall geben. Die hier beschriebenen Unähnlichkeitsmaße, die auf Kreuzvorhersagefehlern und Kreuzkorrelationssummen basieren, können sehr stark von der Anzahl der Spikes abhängen, da für sie nur wenige Nachbarn im Einbettungsraum gefunden werden können. Durch eine Umskalierung auf Gleichverteilung der Werte über die

Rangordnung können diese Effekte vermieden werden, aber auch neue Probleme auftreten.

In der Vorhersage von epileptischen Anfällen kann auch die Verwendung von Klassifikationsmethoden mit Lehrer (*supervised learning*) sinnvoll sein. Dabei werden Zeitreihensegmente, die unmittelbar vor einem Anfall liegen oder zeitlich weiter entfernt sind, dem entsprechenden Cluster zugeordnet, d.h. ihre Zugehörigkeit $u_i^{(\nu)}$ wird bei der Minimierung der Kostenfunktion nicht variiert. Zu welchem der Cluster weitere, neue Zeitreihensegmente gehören, wird durch die Minimierung der Kostenfunktion bezüglich ihrer Zugehörigkeiten ermittelt. Es ist auch möglich, die „festen“ Zugehörigkeiten bei weiteren Anfällen laufend, entsprechend anzupassen. Andere Klassifikationsmethoden mit Lehrer, wie z.B. neuronale Netze, sollten aber als Alternativen in Betracht gezogen werden.

Eine allgemein geeignte Methode zur Vorhersage von epileptischen Anfällen existiert bis jetzt nicht und die Methoden müssen unter Umständen individuell an jeden Patienten angepaßt werden. Die hier durchgeführten Analysen von EEGs zur Vorhersage von epileptischen Anfällen sind also nur als erste Schritte auf diesem Gebiet anzusehen.

Neben den hier beschriebenen Anwendungen ist natürlich eine Vielzahl weiterer Anwendungen denkbar (siehe z.B. [11]). In der Medizin könnte der Clusteralgorithmus sowohl in der Diagnose als auch in der Vorhersage von Krankheiten eingesetzt werden. Aber auch die Vorhersage von Erdbeben in der Geologie, die Untersuchung von Fehlern in Maschinen oder die Meterologie sind als Anwendungsgebiete vorstellbar.

3.7 Zusammenfassung

In diesem Kapitel wurde eine Methode zur Klassifzierung von Zeitreihen entwickelt. Sie basiert auf drei wesentlichen Bestandteilen:

- Ein Abstands- oder Unähnlichkeitsmaß für Zeitreihen. Hier wurde die Kreuzkorrelationssumme (3.9) und der Kreuzvorhersagefehler (3.12) verwendet.
- Eine Kostenfunktion (3.7), die die Güte der Klassifikation wiedergibt.
- Ein Verfahren zur Minimierung der Kostenfunktion. Da harte Cluster (3.3) betrachtet wurden, lag dafür der Metropolis-Algorithmus nahe.

Der hier verwendete Clusteralgorithmus für Zeitreihen zeigte in den Beispielen und in ersten Anwendungen seine Funktionstüchtigkeit. Seine Nützlichkeit für die Stationaritätsanalyse wurde ebenfalls demonstriert.

Wird ein konkreter Fall untersucht, können verschiedene Abwandlungen des Algorithmus ausprobiert werden, wodurch unter Umständen bessere Ergebnisse erzielt werden können. Zunächst kann die Wahl des Unähnlichkeitsmaßes zwischen den Zeitreihen von entscheidender Bedeutung sein. In den Beispielen wurden dabei beliebig entweder Kreuzvorhersagefehler oder Kreuzkorrelationssummen gewählt, da sie im wesentlichen die gleichen Ergebnisse lieferten. Die Untersuchung von möglichst vielen, weiteren Unähnlichkeitsmaßen kann im speziellen Fall aber sehr nützlich sein, wobei auch Unähnlichkeitsmaße, die an Größen der linearen Zeitreihenanalyse angelehnt sind, nicht unberücksichtigt bleiben sollten.

Weitere Alternativen bestehen in der Verwendung einer abgewandelten Kostenfunktion und eines anderen Minimierungsverfahrens. Die Verwendung ähnlicher Kostenfunktionen wurde bereits kurz in Abschnitt 3.3.2 diskutiert. Die Minimierung der Kostenfunktion kann in vielen Fällen der Kostenfunktion angepaßt werden und unter Umständen sogar analytisch möglich sein. Der von mir verwendete Metropolis-Algorithmus ist auf diskrete Zustände, also hier harte Cluster, zugeschnitten und für kontinuierliche Variablen nicht zu empfehlen.

Literaturverzeichnis

[1] F. T. Arecchi, W. Gadomski und R. Meucci, *Generation of chaotics dynamics by feedback on a laser*, Phys. Rev. A **34**, 1617 (1986)

[2] M. Barahona und C.-S. Poon, *Detection of nonlinear dynamics in short, noisy time series*, Nature **381**, 215 (1996)

[3] N. Birbaumer, W. Lutzenberger, H. Rau, G. Mayer-Kress und C. Braun, *Perception of music and dimensional complexity of brain activity*, Int. J. Bifurcations and Chaos **6**, 267 (1996)

[4] W. A. Brock, W. D. Dechert, J. A. Scheinkman und B. LeBaron, *A Test for Independence Based on the Correlation Dimension*, University of Wisconsin Press, Madison (1988)

[5] W. A. Brock, D. A. Hseih und B. LeBaron, *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic evidence*, MIT Press, Cambridge, MA, (1991)

[6] J. P. Eckmann, S. Oliffson Kamphorst und D. Ruelle, *Recurrence plots of dynamical systems*, Europhys. Lett. **4**, 973 (1987)

[7] M. Finardi, L. Flepp, J. Parisi, R. Holzner, R. Badii und E. Brun, *Topological and metric analysis of heteroclinic crises in laser chaos*, Phys. Rev. Lett. **68**, 2989 (1992)

[8] P. Grassberger und I. Procaccia, *Measuring the strangeness of strange attractors*, Physica D **9**, 189 (1983)

[9] P. Grassberger und I. Procaccia, *Characterization of strange attractors*, Phys. Rev. Lett. **50**, 346 (1983)

[10] M. Hénon, *A two dimensional map with a strange attractor*, Commun. Math. Phys. **50**, 69 (1976)

[11] J. L. Hernández, R. Biscay, J. C. Jimenez, P. Valdes und R. Grave de Peralta, *Measuring the dissimilarity between EEG recordings through a non-linear dynamical system approach*, Int. J. Bio-Med. Comp. **38**, 121 (1995)

[12] H. Kantz, *A robust method to estimate the maximal Lyapunov exponent of a time series*, Phys. Lett. A **185**, 77 (1994)

[13] H. Kantz, *Quantifying the closeness of fractal measures*, Phys. Rev. E **49**, 5091 (1994)

[14] H. Kantz und T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press (1997)

[15] L. Kaufman, *Finding groups in data: an introduction to cluster analysis*, Wiley, New York (1990)

[16] E. N. Lorenz, *Deterministic nonperiodic flow*, J. Atmos. Sci. **20**, 130 (1963)

[17] M. C. Mackey und L. Glass, Science **197**, 287 (1977)

[18] R. Manuca und R. Savit, *Stationarity and nonstationarity in time series analysis*, Physica D **99**, 134 (1996)

[19] F. H. C. Marriott, *Barnard's Monte–Carlo tests: how many simulations?*, App. Statist. **28**, 75 (1978)

[20] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller und E. Teller, Journal of Chemical Physics **21**, 1097 (1953)

[21] E. Ott, *Chaos in dynamical systems*, Cambridge University Press (1993)

[22] W. H. Press et al., *Numerical Recipes*, second edition, Cambridge University Press (1995)

[23] D. Prichard, *The correlation dimension of differenced data*, Phys. Lett. A **191**, 245 (1994)

[24] M. B. Priestley, *Non–linear and Non–stationary Time Series Analysis*, Academic Press, London (1988)

[25] D. R. Rigney, A. L. Goldberger, W. Ocasio, Y. Ichimaru, G. B. Moody und R. Mark, *Multi-Channel physiological data: Description and Analysis* in A. S. Weigend und N. A. Gershenfeld, editors, *Time Series Prediction: Forecasting the Future and Understanding the Past*, Santa Fe Institute Studies in the Science of Complexity, Proc. Vol. XV, Addison-Wesley (1993)

[26] T. Sauer, J. Yorke und M. Casdagli, *Embedology*, J. Stat. Phys. **65**, 579 (1991)

[27] R. Schlittgen und H. J. Streitberg, *Zeitreihenanalyse*, 2. Auflage, Oldenbourg Verlag, München Wien (1987)

[28] T. Schreiber, *Detecting and analysing nonstationarity in a time series using nonlinear cross–predictions*, Phys. Rev. Lett. **78**, 843 (1997)

[29] T. Schreiber und A. Schmitz, *Improved surrogate data for nonlinearity tests*, Phys. Rev. Lett. **77**, 635 (1996)

[30] T. Schreiber und A. Schmitz, *On the discrimination power of measures for nonlinearity in a time series*, Phys. Rev. E **55**, 5443 (1997)

[31] T. Schreiber und A. Schmitz, *Classification of time series with nonlinear similarity measures*, (eingereicht zu Phys. Rev.) (1997)

[32] H. G. Schuster, *Deterministic Chaos: An Introduction*, Physik Verlag, Weinheim (1988)

[33] J. E. Skinner, M. Molnar und C. Tomberg, Integrative Physiological and Behavioral Science **29**, 217 (1994)

[34] T. Subba Rao und M. M. Gabr, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics Vol. 24, Springer, New York (1984)

[35] F. Takens, *Detecting strange attractors in turbulence*, Lecture Notes in Math. Vol. 898, Springer, New York (1981)

[36] F. Takens, *On the numerical determination of the dimension of an attractor*, in: B. L. J. Braaksma, H. W. Broer und F. Takens, eds., *Dynamical Systems and Bifurcations*, Lecture Notes in Math. Vol. 1125, Springer, Heidelberg–New York (1985)

[37] J. Theiler, *Spurious dimension from correlation algorithms applied to limited time–series data*, Phys. Lett. A **34**, 2427 (1986)

[38] J. Theiler, *Lacunarity in a best estimator of fractal dimension*, Phys. Lett. A **135**, 195 (1988)

[39] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian und J.D. Farmer, *Testing for nonlinearity in time series: the method of surrogate data*, Physica **D 58**, 77 (1992)

[40] J. Theiler und D. Prichard, *Constrained–realization Monte–Carlo method for hypothesis testing*, Physica D **94**, 221 (1996)

[41] A. Varone, A. Politi und M. Ciofini, *CO_2 laser with feedback*, Phys. Rev. A **52**, 3176 (1995)