BERGISCHE UNIVERSITÄT
GESAMTHOCHSCHULE WUPPERTAL
FACHBEREICH PHYSIK

Erkennung von Nichtlinearitäten und wechselseitigen Abhängigkeiten in Zeitreihen

Dissertation
zur Erlangung des Doktorgrades im Fach Theoretische Physik
an der Bergischen Universität Wuppertal

vorgelegt von

Andreas Schmitz

Dezember 2000
WUB-DIS 2000-11

Abstract

The present work addresses two central questions in the analysis of time series. The first part deals with methods to test for nonlinear structure in measured signals. Only a positive outcome of such a test justifies the application of the advanced techniques of nonlinear time series analysis. We will discuss some limitations and caveats of the popular method of surrogate data for nonlinearity tests. Further we present a new way to generate surrogate data that overcomes these caveats and features an extraordinary flexibility. This flexibility for the first time allows e.g. to test for nonlinearity in unevenly sampled time series. Further examples and applications of the new method are studied.
The second part of the present work deals with the analysis of dependencies between time series. This is motivated by the recently rising interest in synchronisation phenomena of coupled chaotic systems. The central practical question is, wheather one can tell the dominant direction of the coupling between two systems. Standard measures of interdependencies are not optimized to answer this question. To address this problem, two new concepts will be developed and discussed. The first measure is motivated by information theory and is called transfer entropy, while the second is oriented closer to synchronisation itself. We shall demonstrate that we are indeed able to tell the direction of coupling by the asymmetry of the measures. Further investigations on unidirectionally coupled map lattices reveal further interesting properties. Finally we present some first applications to experimental data.

Inhaltsverzeichnis

Abbildungsverzeichnis

Einleitung

Die nichtlineare Zeitreihenanalyse bildet das Bindeglied zwischen der Theorie der nichtlinearen, dynamischen Systeme einerseits und der Analyse von Daten andererseits. Die grundlegende Theorie der chaotischen, dynamischen Systeme wurde dabei im Wesentlichen von Mathematikern, wie z. B. Poincaré und Birkhoff, in der ersten Hälfte des 20. Jahrhunderts entwickelt und später von Kolmogorov, Smale, Sinai, Arnold und Ruelle — um nur einige zu nennen — ausgebaut. Ihr Ziel war die theoretische Beschreibung und die Bestimmung von Eigenschaften von Systemen, deren Dynamik als bekannt vorausgesetzt wurde. Auf der anderen Seite standen Experimentatoren mit Daten, die an Systemen aufgenommen wurden, deren zugrunde liegende Dynamik häufig nicht bekannt war. Beobachtete Irregularitäten in den Daten wurden dabei zunächst als das Ergebnis von externen Störungen, Messrauschen oder einer hohen Komplexität des Systems angesehen. Erst in den Achtzigerjahren gelangte eine weitere alternative Erklärung ins öffentliche Bewusstsein: Die Idee des deterministischen Chaos, welches sich durch eine sensitive Abhängigkeit von den Anfangsbedingungen auszeichnet. Die Tatsache, dass selbst einfachste Systeme, wie z. B. Doppelpendel oder nichtlineare, elektrische Schaltkreise irreguläres, chaotisches Verhalten zeigen können, faszinierte besonders die Physiker. Für sie war und ist stets auch die Einfachheit der zugrunde liegenden Mechanismen ein Kriterium für die Attraktivität eines neuen Modells.

Es zeigte sich schnell, dass die Theorien nicht direkt auf die experimentellen Daten anwendbar waren und erst adäquate Werkzeuge, nämlich die Methoden der nichtlinearen Zeitreihenanalyse, entwickelt werden mussten. Die rasante Geschwindigkeit dieser Entwicklung war nicht zuletzt auf die parallele Entwicklung von immer leistungsfähigeren und preiswerteren Computern zurückzuführen. Durch sie wurde es erstmals möglich, nicht nur große Datenmengen aufzunehmen und statistisch zu analysieren, sondern auch Gleichungen numerisch zu lösen, die analytisch nicht zugänglich waren. In der ersten Euphorie wurde vielerorts überschnell behauptet, in den unterschiedlichsten (nichtphysikalischen) Systemen „Chaos“ gefunden zu haben. Einen entscheidenden Vorsprung im Umgang mit Daten hatten hier die Physiker, welche gewohnt sind, vorurteilsfrei an ein gegebenes Problem heranzugehen und zunächst einfachere Alternativen auszuschließen. Erste gewonnene Ergebnisse wurden kritisch hinterfragt und die statistische Signifikanz angezweifelt. Ein statistischer *Test* für die Existenz von Chaos sollte Klarheit schaffen.

Dabei ist das theoretische Kriterium für das Auftreten von Chaos denkbar einfach: Ist der maximale Lyapunov–Exponent des Systems, welcher ein Maß für die exponentielle Divergenz einer infinitesimalen Abweichung darstellt, *positiv* und *endlich*, ist das System chaotisch. Leider ist der maximale Lyapunov–Exponent nur schwer aus einem Datensatz zu schätzen, was gleichfalls für die Angabe des Fehlers der Schätzung gilt. Zudem weist Rauschen theoretisch einen unendlichen Lyapunov–Exponenten auf, während Schätzer selbstverständlich nur endliche Werte liefern können. Eine Unterscheidung zwischen Rauschen und Chaos wäre also ebenfalls nicht leicht. Auch hier ist also die direkte Umsetzung der Theorie nicht die praktikabelste.

Eine Voraussetzung für das Auftreten von Chaos ist die Nichtlinearität des Systems. Darunter ist die Nichtlinearität der zugrunde liegenden, dynamischen Gleichungen des Systems zu verstehen, welche entweder in der kontinuierlichen Form

$$\frac{\mathrm{d}\vec{x}(t)}{\mathrm{d}t} = \vec{F}\left[\vec{x}(t)\right]$$

als *Differentialgleichungen* oder in der zeitlich diskretisierten Form

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

als *Abbildung* geschrieben werden können. Können in einem Datensatz keine Nichtlinearitäten nachgewiesen werden, kann die Existenz von Chaos zwar nicht ausgeschlossen werden, eine sinnvolle Anwendung der Methoden der nichtlinearen Zeitreihenanalyse lässt sich dann aber nicht mehr rechtfertigen. Andersherum ist aber natürlich der Nachweis von Nichtlinearität kein Beweis für das Auftreten von chaotischem Verhalten, sodass in diesem Fall weitere Untersuchungen notwendig sind. Die *lineare* Form der oben angegebenen Differentialgleichung kann leicht geschlossen gelöst werden und es stellt sich heraus, dass unter der Voraussetzung der Endlichkeit für lange Zeiten lediglich konstante oder periodische Lösungen auftreten. Diese können normalerweise leicht von chaotischen Lösungen unterschieden werden. Anders sieht dies aus, wenn das System hochdimensional ist oder in M stochastische Terme, d. h. Rauschanteile, erlaubt werden.

In Kapitel 1 werden zunächst die etablierten Standardmethoden für Nichtlinearitätstest vorgestellt und diskutiert. Implementationen dieser Methoden lassen sich in frei verfügbaren Programmpaketen wie z. B. TISEAN [HKS99] finden und sind in vielen Fällen ausreichend. Im folgenden Kapitel gehe ich dann auf einige Probleme der Standardmethoden ein und stelle daraufhin einen völlig neuen Ansatz für einen statistischen Test vor. Diese neue Methode löst nicht nur die Probleme der Standardmethoden, sondern zeichnet sich auch durch eine außergewöhnliche Flexibilität aus. Diese Flexibilität erlaubt es z. B., erstmals auch Zeitreihen mit variierenden Zeitabständen auf ihre Nichtlinearität hin zu testen. Weitere Beispiele und Anwendungen auf experimentelle Zeitreihen beschließen diesen ersten Teil meiner Arbeit.

Der zweite Teil dieser Arbeit beschäftigt sich mit der Abhängigkeit *zwischen* Zeitreihen. Das verstärkte Auftreten von *multivariaten* Zeitreihen, also mehreren *gleichzeitig* aufgenommenen Zeitreihen, ist ebenfalls auf die schnelle Entwicklung der Computer zurückzuführen, mit denen immer größere Datenmengen verwaltet werden können. Der Grund für diesen zusätzlichen Messaufwand ist selbstverständlich die Hoffnung der Experimentatoren, zu einem besseren Verständnis des beobachteten Systems zu gelangen. Dabei stellt sich in vielen Fällen als erstes die Frage nach den Zusammenhängen, Gemeinsamkeiten und Abhängigkeiten zwischen den gemessenen Größen. In diese Richtung zielende Bestrebungen in der nichtlinearen Zeitreihenanalyse sind also naheliegend.

Die vielfältigen Möglichkeiten und Schwierigkeiten dieser Aufgabenstellung sollen im Folgenden an einem einfachen, allgemein verständlichen Beispiel illustriert werden. Angenommen wir könnten z. B. per Telemetrie die Geschwindigkeit zweier Fahrzeuge messen. Können wir anhand dieser Daten herausfinden, ob eines der beiden Fahrzeuge das andere abschleppt? Zunächst ist es sicherlich notwendig, einen hinreichend langen Datensatz zu besitzen, der Situationen wie Beschleunigungen, Abbremsungen und auch Stillstände enthält. Sind die Geschwindigkeiten von beiden Fahrzeugen stets gleich, würde dies auf die

Existenz einer Abschleppstange zwischen den Fahrzeugen hindeuten. Was aber, wenn es sich um ein Abschleppseil handelt und nur das vordere Fahrzeug beschleunigt und stets nur das hintere Fahrzeug bremst? Obwohl in beiden Fällen die gleiche Beobachtung gemacht würde, wären die Abhängigkeiten zwischen den Fahrzeugen grundverschieden. Einen Aufschluss darüber, um welchen der beiden Fälle es sich handelt, könnte nur eine zusätzliche Observable, wie z. B. die Stellung der Bremspedale in beiden Fahrzeugen, liefern. Weitere, kompliziertere Situationen sind vorstellbar. So könnten die beiden Fahrzeuge durch eine Feder mit linearer Rückstellkraft oder einem Gummi miteinander verbunden sein. Könnte dies von einer Situation im zähflüssigen Autobahnverkehr unterschieden werden?

Die hier interessierenden Systeme seien zudem chaotisch oder zumindest nichtlinear, sodass naheliegende, lineare Maße keine hinreichende Beschreibung der Abhängigkeiten liefern können. Ein motivierendes Moment war dabei das zunehmende Interesse an *Synchronisationsphänomenen* von gekoppelten, chaotischen Systemen, deren Untersuchung unter Umständen relevante Fragen in vielen wissenschaftlichen Bereichen beantworten kann. So wird z. B. in der Gehirnforschung vermutet, dass die Synchronisation im Gehirn von großer Bedeutung für das „Bindungsproblem“ ist. Weiter wird aber auch vermutet, dass epileptische Anfälle mit einem *erhöhten* Maß an Synchronisation zwischen verschiedenen Bereichen im Gehirn verbunden sind. Diese Anfälle gehen meist von einem bestimmten Bereich, dem Fokus, aus und breiten sich dann über das ganze Gehirn aus. Dabei stellt sich natürlicherweise die Frage, ob bei einer beidseitigen Kopplung, die *dominante Richtung* dieser Kopplung durch Messungen an den einzelnen Systemen bestimmt werden kann. Die in Kapitel 4 vorgestellten Standardmaße zur Abhängigkeitsanalyse sind allerdings nicht auf diesen Fall optimiert und neue Ansätze notwendig.

Zwei völlige neue Methoden zur Bestimmung von Abhängigkeiten zwischen Zeitreihen werden in Kapitel 5 entwickelt. Sie basieren dabei auf zwei unterschiedlichen Grundideen. Während das eine Maß, die Transfer–Entropie, durch die Informationstheorie motiviert wurde, ist die andere Klasse von Abhängigkeitsmaßen eher an der Idee der Synchronisation orientiert. Beide Maße beschreiben die Abhängigkeit zwischen zwei Systemen und weisen eine fundamentale Asymmetrie unter Vertauschung der Systeme auf. Sie sind damit viel versprechende Kandidaten bei der Suche nach der Richtung einer eventuellen Kopplung. In Kapitel 6 werden zunächst die ersten Eigenschaften und die Funktionstüchtigkeit der neuen Methoden an einfachen Modellsystemen verifiziert und deren Leistungsfähigkeit mit den Standardmaßen verglichen. Eine anschließende Anwendung auf gekoppelte Abbildungen auf dem Gitter offenbart weitere Eigenschaften der neuen Abhängigkeitsmaße. Schließlich werden sie auf erste experimentelle Zeitreihen angewendet, wobei diese aus verschiedenen wissenschaftlichen Bereichen, wie Ökonomie, Medizin und Klimaforschung stammen.

Teil I

Tests auf Nichtlinearität

Kapitel 1

Tests auf Nichtlinearität

1.1 Einleitung

In nahezu allen Bereichen der Wissenschaft können zeitlich variierende Prozesse 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 *nicht*linearen Zeitreihenanalyse auf Daten ist meist dadurch motiviert, dass viele Prozesse in der Natur nur schwerlich als linear angenommen werden können. In diese Kategorie fallen nicht nur viele physikalische Systeme, sondern auch die große Klasse der physiologischen Daten. In solchen Fällen stellt sich allerdings stets die Frage, ob die Anwendung der nichtlinearen Zeitreihenanalyse auch durch die *Daten* allein gerechtfertigt werden kann. Dabei spielt natürlich auch die Qualität der Daten bezüglich Aufnahmerate, Länge und Rauschniveau eine entscheidende Rolle.

In jedem Fall ist es wünschenswert, einen objektiven Test zu besitzen, der aussagt, ob *sichtbare* Nichtlinearitäten in den Daten vorhanden sind. Der Einsatz von nichtlinearen Techniken kann nur dann sinnvoll sein, wenn ein solcher Test erfolgreich Nichtlinearität findet. Dabei hat sich die Surrogatmethode von Theiler et al. [TEL+92] als erfolgreicher Ansatz erwiesen, der auch in dieser Arbeit als Grundlage dient und in Abschnitt 1.4 beschrieben wird.

Gemessene Zeitreihen sind aber selten ein Produkt eines rein linearen oder nichtlinearen, niedrigdimensionalen Systems. Sie sind vielmehr oft eine Mischung aus mehreren linearen, nichtlinearen und stochastischen Anteilen. Die in den Daten vorhandenen Nichtlinearitäten können also z. B. durch viele Freiheitsgrade oder starken Rauschanteil verwischt und deshalb nur sehr schwach ausgeprägt sein. Hinzu kommt, dass Nichtlinearitäten in unterschiedlichsten, auch *nicht*dynamischen Formen auftreten können, was nicht immer auf eine nichtlineare, deterministische Dynamik schließen lässt.

In diesem Kapitel werden zunächst die Standardmethoden für Tests auf Nichtlinearität vorgestellt, wobei bereits auf Vor- und Nachteile der einzelnen Methoden eingegangen wird. Im darauf folgenden Kapitel werden bestimmte Probleme der Standardmethoden konkretisiert. Es wird eine verallgemeinerte und flexiblere Methode entwickelt, die einige Probleme löst und den bisher engen Rahmen an möglichen Alternativen, auf die getestet werden kann, sprengt.

1.2 Statistische Tests

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.:

(a) Sind die Werte korreliert?

(b) Welcher Verteilung unterliegen die Werte der Zeitreihe?

(c) Zeigt die Zeitreihe chaotisches Verhalten?

(d) Was ist die zugrunde liegende Dynamik?

Insbesondere können *Entscheidungsfragen* formuliert werden, auf die nur mit „ja“ oder „nein“ geantwortet werden kann. Jede Entscheidungsfrage filtert eine bestimmte Eigenschaft der Zeitreihe heraus. Komplexere Fragen können im Prinzip aus mehreren Entscheidungsfragen kombiniert werden, was aber in solch allgemeinen Fragen wie in Beispiel (d) wenig praktikabel ist.

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 Beispiel (a) 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 bzw. dass die Nullhypothese für einen gegebenen Test korrekt formuliert ist.

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 $\{s_n\}, n = 1 \dots N$, 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 der Schätzer für die *Autokovarianz*

$$\gamma(\tau) = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} (s_n - \bar{s})(s_{n+\tau} - \bar{s}) \tag{1.1}$$

mit einer festen zeitlichen Verzögerung τ dienen.[1] Die Nullhypothese würde dann z. B. für $\gamma(\tau = 1) \neq 0$ verworfen und anderenfalls akzeptiert. Für endliche Zeitreihen unterliegt aber der Wert der Autokovarianz statistischen Fluktuationen, sodass in der Praxis auch für unkorrelierte Daten $\gamma(\tau)$ von 0 abweicht. So liefert z. B. die Varianz des Schätzers der Autokovarianz (1.1) für weißes Rauschen $1/N + \mathcal{O}(1/N^2)$ für alle $\tau \neq 0$ [SS87]. Der obige Test wäre also nur im Limes für $N \to \infty$ korrekt und würde für endliche Zeitreihen falsche Aussagen liefern.

[1] Der Ausdruck $\bar{x}$ kennzeichnet dabei hier und im Folgenden den zeitlichen Mittelwert der Werte $\{x_n\}$.

Die Autokovarianz 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 Abstands 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 1.4 wird gezeigt, wie die Verteilung anhand von vielen Datensätzen, die die Nullhypothese erfüllen, numerisch approximiert werden kann.

1.2.1 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 Signifikanzniveau von $1-p$ vorgegeben, sollte die Size p nicht überschreiten. Für gewöhnlich wird p, die *nominelle* Size, vorher festgesetzt und der Test entsprechend realisiert. Es muss gewährleistet werden, dass die wirkliche, *empirische* 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, kann aber zu einer unerwünschten Senkung der Leistungsfähigkeit des Tests führen. 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 dies 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. Eine möglichst hohe Trennschärfe ist also wünschenswert. Verschiedene Realisierungen eines Tests können unterschiedliche Trennschärfen aufweisen und die Trennschärfe kann als Maß für die Güte der einzelnen Realisierungen betrachtet werden. Ein wichtiger Punkt bei der Konstruktion eines Test ist dabei die Wahl der Teststatistik. Wird sie ungünstig gewählt, kann dies zu einer unerwünschten Senkung der Trennschärfe führen.

1.3 Maße für Nichtlinearität

Für einen Test auf Nichtlinearität ist es nötig eine Teststatistik zu besitzen, die den Grad der vorhandenen Nichtlinearitäten misst. Dazu lassen sich in der Literatur viele unterschiedliche Ansätze finden. Im Folgenden werden die Teststatistiken vorgestellt, die in den weiteren Untersuchungen verwendet werden. Die hier getroffene Auswahl beruht dabei im Wesentlichen auf dem repräsentativen Vergleich in [SS97b] bezüglich ihrer Trennschärfe.

Diese Teststatistiken zeichnen sich durch ihre Sensitivität auch auf schwächste Nichtlinearitäten aus. Weitere Alternativen werden in [SS97b] und den darin enthaltenen Referenzen diskutiert.

Traditionelle Maße für Nichtlinearität basieren auf Verallgemeinerungen der Autokovarianz (1.1) auf höhere Ordnungen. Da alle Momente mit einer Ordnung größer als zwei für lineare Prozesse verschwinden, können sie als Teststatistiken für einen Test auf Nichtlinearität dienen, indem auf einen signifikanten Unterschied von Null getestet wird. Ein Moment dritter Ordnung ist z. B. die *3-Punkt-Autokovarianz*

$$T^{3\mathrm{P}}(\tau_1, \tau_2) = \frac{1}{N - \tau_2 - 1} \sum_{n=1}^{N-\tau_2} (s_n - \bar{s})(s_{n+\tau_1} - \bar{s})(s_{n+\tau_2} - \bar{s}) \,, \tag{1.2}$$

wobei ohne Verletzungen der Allgemeinheit $\tau_1 < \tau_2$ angenommen wurde. Häufig wird $\tau_2 = 2\tau_1$ verwendet. Eine weitere sehr nützliche Größe dritter Ordnung ist

$$T^{\mathrm{rev}}(\tau) = \frac{\sum_{n=1}^{N-\tau} (s_n - s_{n+\tau})^3}{\sum_{n=1}^{N-\tau} (s_n - s_{n+\tau})^2} \,, \tag{1.3}$$

welche als Maß für die Zeitumkehrinvarianz einer Zeitreihe dienen kann. Für die zeitumgekehrte Zeitreihe $\{s_{N-n+1}\}$ ändert sich lediglich das Vorzeichen von T^{rev}, sodass für zeitumkehrinvariante Prozesse $T^{\mathrm{rev}} = 0$ gelten muss. Da lineare Prozesse komplett über ihre Autokovarianzen beschrieben werden können und diese invariant unter Zeitumkehr sind, liefern lineare Prozesse $T^{\mathrm{rev}} = 0$. Prozesse mit nichtlinearer Dynamik zeigen dagegen häufig sehr starke Asymmetrie unter Zeitumkehr, wodurch sich T^{rev} neben seiner Einfachheit und Robustheit sehr gut als Teststatistik eignet [DvHTD95, TEL$^+$92, RG84].

Obwohl nicht verschwindende Momente höherer Ordnung ein starkes Indiz für Nichtlinearität sind, sagen sie nichts über die Art der gefundenen Nichtlinearität aus. Häufig steht hinter einem Test aber die Vermutung auf eine nichtlineare, zugrunde liegende, deterministische Dynamik, sodass als Teststatistiken Größen aus der Theorie nichtlinearer, deterministischer Systeme naheliegen. Um diese Größen auch für skalare Zeitreihen schätzen zu können, müssen die unter Umständen mehrdimensionalen Phasenraumzustände aus der Zeitreihe rekonstruiert werden. Eine sehr nützliche und weitverbreitete Methode, dies zu bewerkstelligen ist die Zeitverzögerungeinbettung

$$\vec{s}_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \ldots, s_{n-\tau}, s_n) \,. \tag{1.4}$$

Dabei ist m die *Einbettungsdimension* und τ die Zeitverzögerung bzw. das *Lag*. Eine skalare Zeitreihe mit N Werten liefert $N_{\mathrm{E}} = N - (m-1)\tau$ Einbettungsvektoren. Der besseren Lesbarkeit wegen, wird im Folgenden auch für N_{E} die Bezeichnung N verwendet. Aus dem jeweiligen Kontext heraus ist die Bedeutung stets eindeutig erkennbar. Auf die Eigenschaften und die Wahl der Parameter der Phasenraumrekonstruktion (1.4) möchte ich hier nicht näher eingehen und auf die existierende Literatur [KS97, Aba96] verweisen.

Eine Klasse sehr robuster Phasenraumobservablen für Nichtlinearitätstests sind Vorhersagefehler. Ein Vorhersagefehler kann z. B. über die Standardabweichung zwischen Vorhersage und dem wirklich angenommen Wert

$$T^{\mathrm{VF}}(m, \tau) = \left(\frac{1}{N-1} \sum_{n=1}^{N-1} [s_{n+1} - F(\vec{s}_n)]^2 \right)^{1/2} \tag{1.5}$$

definiert werden. Dabei kann die nichtlineare Vorhersage $F(\vec{s}_n)$ für den Wert s_{n+1} der Zeitreihe auf verschiedene Art und Weise realisiert werden. Ich werde in dieser Arbeit die lokal konstante Vorhersage verwenden, die im Folgenden beschrieben wird. Zunächst wird für jeden Punkt $\vec{s}_n$ im Einbettungsraum die r–Umgebung

$$U_n(r) = \{\vec{s}_i | \, \|\vec{s}_n - \vec{s}_i\| < r\} \tag{1.6}$$

bestimmt, die aus allen Nachbarpunkten besteht, deren Abstand zum Punkt $\vec{s}_n$ geringer als der Radius r ist und deren zeitlicher Abstand größer als $t_{\min}$ ist. Der Vorhersagewert ist dann nichts anderes als der Mittelwert der Folgewerte der Vektoren aus der r–Umgebung

$$F(\vec{s}_n) = \frac{1}{\|U_n(r)\|} \sum_{\vec{s}_i \in U_n(r)} s_{i+1} \, , \tag{1.7}$$

wobei mit $\|U_n(r)\|$ die Anzahl der Punkte in der r–Umgebung bezeichnet wird. Bei der Bestimmung von $F(\vec{s}_n)$ werden sowohl vergangene als auch zukünftige Nachbarn aus der r–Umgebung verwendet. Um aussagekräftige, tatsächliche Vorhersagen zu erhalten, wäre ein solches Vorgehen nicht korrekt. Hier dient der Vorhersagefehler aber lediglich als Teststatistik und nicht als Qualitätsmaß für Vorhersagen. Zu (1.7) sind natürlich viele Alternativen denkbar. Neben Variationen, wie z. B. verschiedenen Wichtungen und lokal linearen Vorhersagen, sind auch Anpassungen unterschiedlichster Modelle an die Daten möglich, über die dann Vorhersagen erzielt werden können.

Allen in diesem Abschnitt aufgeführten Größen ist gemeinsam, dass ihre Verteilung für ein generisches Ensemble von Zeitreihen nicht bekannt ist. Nur unter sehr speziellen und meist unrealistischen Voraussetzungen an die Zeitreihen und weiteren Vereinfachungen kann ihre Verteilung analytisch bestimmt werden. Im nächsten Abschnitt wird eine Methode zur numerischen Approximation dieser Verteilungen vorgestellt.

1.4 Die Surrogatmethode

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“ Ersatz–Zeitreihen werden auch *Surrogate* genannt [TEL+92]. Die gewünschten Eigenschaften der Surrogate und damit ihre Erzeugung hängen natürlich von der Nullhypothese ab, auf die getestet werden soll. Das Generieren von Surrogaten ist daher oft nicht trivial, nur schwer oder bislang gar nicht zu bewerkstelligen. 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, die im Folgenden beschrieben werden.

1.4.1 Typische Realisierung

Wird die Nullhypothese direkt über eine Klasse von Modellen definiert, liegt es nahe, die Surrogate einfach durch Realisierungen eines Modell zu generieren. Durch unterschiedliche Anfangsbedingungen lässt sich so eine beliebig große Anzahl an Surrogaten erzeugen. Die offenen Parameter des Modells müssen durch Anpassung aus der Originalzeitreihe bestimmt werden. Hier zeigt sich auch schon ein erster Nachteil dieser *typischen Realisierung*. So ist es oft nicht einfach aus einer einzelnen, unter Umständen auch kurzen Zeitreihe die Parameter mit genügend hoher Genauigkeit zu bestimmen. Hinzu kommt häufig das Problem, dass die Ordnung des Modells eine weitere Freiheit darstellt, deren Wahl aber von entscheidender Bedeutung sein kann.

Für das in Abschnitt 1.2 erwähnte Beispiel für einen Test auf Korrelationen in einer Zeitreihe, könnte die typische Realisierung wie folgt aussehen. Die Surrogate werden über eine Folge von Zufallszahlen erzeugt, deren Wahrscheinlichkeitsverteilung mit der angepassten Verteilung der Werte der Originalzeitreihe übereinstimmt. Die Verteilung dieser Zufallszahlen wird dabei aber nicht exakt, sondern nur annähernd übereinstimmen. Werden signifikante Unterschiede in der Autokovarianz (1.1) der Originalzeitreihe und den Surrogaten gefunden, kann die Nullhypothese der Unkorreliertheit verworfen werden.

Um die Methode der typische Realisierung in einem Test auf Nichtlinearität zu verwenden, müsste ein Modell herangezogen werden, das die Klasse der linearen Prozesse abdeckt. In der Tat existiert in der linearen Zeitreihenanalyse ein solches allgemeines Modell, das sogenannte *ARMA–Modell*

$$x_n = \sum_{p=1}^{P} a_p x_{n-p} + \sum_{q=0}^{Q} b_q \eta_{n-q}\,. \qquad (1.8)$$

Dabei sind a_p und b_q reelle Koeffizienten und die $\{\eta_n\}$ gaußverteilte, unabhängige Zufallszahlen mit $\gamma^{(\eta)}(\tau) = 0$. Die erste Summe beschreibt den *autoregressiven* (AR) Aspekt des Prozesses, d. h. die Abhängigkeit von der eigenen Vergangenheit. Ist $Q = 0$, so handelt es sich um ein autoregressives Modell 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.

Zur Anpassung der Parameter des ARMA–Modells wurden unterschiedlichste Ansätze entwickelt, wobei der einfachste auf einer einfachen Korrespondenz zwischen der Autokovarianz (1.1) und den Koeffizienten des AR–Prozesses basiert [SS87]. Entsprechende Surrogate sind dann wieder einfach Realisierungen des angepassten AR-Prozesses.

Bis jetzt sind wir davon ausgegangen, dass die einzigen Freiheiten in dem anzupassenden Modell durch einen endlichen Satz reeller Zahlen beschrieben werden kann. Die Möglichkeit einer unbekannten Funktion im Modell wurde z. B. nicht betrachtet. Eine erwünschte Erweiterung in diese Richtung zeigt die Grenzen der Methode der typischen Realisierung, da dann eine sinnvolle Modellanpassung nur sehr schwierig wäre. In der Praxis tauchen solche Funktionen aber in Form von z. B. Messfunktionen

$$s_n = m(x_n) \qquad (1.9)$$

sehr häufig auf. Die gemessenen Größen in Form von Zeitreihen $\{s_n\}$ zeigen dabei oft auch nichtlineare Abhängigkeit von den interessierenden, intrinsischen Größen $\{x_n\}$ des

Systems, deren Verhalten unter Umständen durch ein bekanntes, lineares Modell beschrieben werden kann.

1.4.2 Bedingte Realisierung

Im Gegensatz zum vorherigen Abschnitt werden bei der *bedingten Realisierung* bestimmte Eigenschaften der Originaldaten *exakt* auf den Surrogaten reproduziert. So können z. B. einfache Eigenschaften wie Mittelwert oder Varianz durch Umskalierung der Surrogate mit der Originalzeitreihe in Übereinstimmung gebracht werden. In [TP96] wird anhand einiger Beispiele sehr deutlich gezeigt, dass die bedingte Realisierung stets vorzuziehen ist, da sie im Gegensatz zur typischen Realisierung korrekte Size und höhere Trennschärfe liefert.

In dem oben erwähnten Beispiel eines Tests auf Korrelation ist auch die Erzeugung von Surrogaten als bedingte Realisierungen relativ einfach. Ein Surrogat kann hier durch mehrfache, zufällige Vertauschung der Werte der Zeitreihe erzeugt werden. Dies zerstört jegliche, eventuell im Originaldatensatz vorhandene Korrelationen. Andere Eigenschaften, wie z. B. die Verteilung der Werte, bleiben dabei aber im Gegensatz zur typischen Realisierung exakt erhalten.

Bei einem Test auf Nichtlinearität ist die entscheidende, zu erhaltende Eigenschaft der Surrogate die Autokovarianz (1.1), die, wie bereits erwähnt, in Beziehung zu den Koeffizienten des ARMA–Modells (1.8) steht. Diese direkte Korrespondenz zwischen Modellklasse und Eigenschaften ist aber eher die Ausnahme und nicht als Regelfall anzusehen. Da interessierende Nullhypothesen aber häufig über Modellklassen formuliert werden, ist das Formulieren entsprechender Eigenschaften oft schwierig oder gar unmöglich. Auch wenn die Eigenschaften bekannt sind, kann ihre gleichzeitige Realisierung problematisch sein. Oft können sie nur näherungsweise in den Surrogaten reproduziert werden. Ziel dieser Arbeit ist es, eine möglichst allgemeine Methode zur Implementierung mehrerer unterschiedlicher Eigenschaften zu entwickeln, welche in Kapitel 2 vorgestellt wird.

1.4.3 Tests mit Surrogaten

Ist die Methode zur Generierung der Surrogate in Einklang mit der zu untersuchenden Nullhypothese, wird der Test folgendermaßen durchgeführt. Zunächst werden B Surrogate erzeugt und auf allen Surrogaten 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. Eine solche mögliche Abweichung in beide Richtungen ist z. B. für 3-Punkt-Autokovarianz (1.2) oder die Zeitumkehrasymmetrie (1.3) als Teststatistik zu erwarten, sodass in diesen Fällen ein zweiseitiger Test nötig wäre.

In einem Surrogattest 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 zweiseitigen Test $B = 2/\alpha - 1$ Surrogate benötigt. In diesem Fall wird die Nullhypothese verworfen, wenn T_0 der kleinste oder der größte der Liste $\{T_b\}$ 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. Ein Beispiel für eine Teststatistik, für die nur ein einseitiger Test durchgeführt werden muss, ist der Vorhersagefehler (1.5), welcher zum Verwerfen der Nullhypothese auf der Originalzeitreihe signifikant kleiner sein sollte als für die Surrogate.

Für gewöhnlich wird die Anzahl der Surrogate zu $B = 2n/\alpha - 1$ für einen zweiseitigen Test und zu $B = n/\alpha - 1$ für einen einseitigen Test gewählt, wobei $n \in \mathbb{N}$ ist. Der Grund, mehr Surrogate als minimal nötig ($n = 1$) für einen Test zu verwenden, ist eine leichte Erhöhung der Trennschärfe für große B [Mar78]. 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 aber sehr wahrscheinlich zu einem falschen Testergebnis. Für einige der Teststatistiken aus Abschnitt 1.3 sieht 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. Um möglichst wenige Annahmen über die Verteilung der Teststatistik zu machen, werde ich im Folgenden ausschließlich das nichtparametrische Verfahren verwenden.

1.5 Methoden im Frequenzraum

Bei der Erzeugung von Surrogaten für Tests auf Nichtlinearität ist zu beachten, dass ihre linearen Eigenschaften mit denen des Originaldatensatzes übereinstimmen. Bei einer bedingten Realisierung sollten die Surrogate also exakt die gleichen Autokovarianzen $\gamma(\tau)$ wie die Originaldaten aufweisen. Da dies aber auf einem Surrogat nicht offensichtlich direkt zu bewerkstelligen ist, wechselt man mittels der *diskreten Fouriertransformation*

$$\tilde{s}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_n e^{2\pi i k n/N}, \quad k = -N/2, \ldots, N/2 \tag{1.10}$$

in der Frequenzraum, welche ich im Folgenden kurz als $\{\tilde{s}_k\} = \mathcal{F}\{s_n\}$ schreibe. Die Normierung in Gleichung (1.10) ist so gewählt, dass die inverse Transformation $\mathcal{F}^{-1}$ durch einen einfachen Vorzeichenwechsel im Exponenten aus $\mathcal{F}$ hervorgeht. Die komplexen $\tilde{s}_k$ lassen sich durch ihre Amplituden $A_k = |\tilde{s}_k|$ und Phasen $\phi_k = \arctan(\mathrm{Im}(\tilde{s}_k)/\mathrm{Re}(\tilde{s}_k))$ darstellen. Für die Generierung von Surrogaten ist wichtig, dass die quadrierten Fourieramplituden A_k^2 einen Schätzer für das *Frequenzspektrum* darstellen, welches über das *Wiener–Khinchin–Theorem*

$$\mathcal{F}\{\varrho(n)\} = \{A_k^2\} \tag{1.11}$$

mit den *Autokorrelationen* $\varrho(\tau) = \gamma(\tau)/\gamma(0) \equiv \gamma(\tau)/\sigma^2$ verknüpft ist. Zwei Zeitreihen mit gleichem Frequenzspektrum besitzen also die gleiche Autokovarianz. Diese Tatsache wird bei der im folgenden Abschnitt beschriebenen Methode zur Erzeugung von Surrogaten ausgenutzt.

1.5.1 FT–Surrogate

Für einen einfachen Nichtlinearitätstests genügt es demzufolge, Surrogate zu generieren, die in den Fourieramplituden mit den Originaldaten übereinstimmen. Die verbleibenden Freiheiten (im Frequenzraum) liegen in der Wahl der Phasen. Das folgende, einfache Schema erzeugt entsprechende Surrogate:

1. Die Zeitreihe $\{s_n\}$ wird mit der diskreten Fouriertransformation (1.10) in den Frequenzbereich transformiert.

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

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

Diese Surrogate besitzen per Konstruktion die gleichen Fourieramplituden wie der Originaldatensatz und werden auch *phasenrandomisierte* Surrogate genannt. Wie bereits erwähnt, sind diese Amplituden aber lediglich ein Schätzer für das Frequenzspektrum, das über Gleichung (1.11) mit der Autokovarianz verknüpft ist. Eine bessere Schätzung des Frequenzspektrums ist also wünschenwert und es existieren dazu eine Reihe weiterer Methoden [PTVF92]. Für die Surrogaterzeugung ist aber auch die Rücktransformation in Schritt 3 von entscheidender Bedeutung.

Ein Test mit den nach obigem Schema konstruierten Surrogaten testet auf die Nullhypothese eines gaußschen, linearen Prozesses. Wurden die Daten allerdings über eine nichtlineare Messfunktion (1.9) gewonnen, sodass die Verteilung der Originalzeitreihe von der Gaußverteilung abweicht, würde der Test die Nullhypothese korrekterweise verwerfen. Will man eine solche instantane, nichtlineare Messfunktion in die Nullhypothese aufnehmen, müssen die entsprechenden Surrogate neben der Autokovarianz auch die Verteilungsfunktion mit den Originaldaten teilen.

1.5.2 AAFT–Surrogate

Im Sinne der bedingten Realisierung bedeutet ein exaktes Reproduzieren der Verteilung, dass die Surrogate dieselben Werte annehmen wie die Originalzeitreihe. Die verbleibende Freiheit liegt in ihrer Reihenfolge. Diese Freiheit wird aber dadurch eingeschränkt, dass auch die Fourieramplituden bzw. die Autokovarianzen übereinstimmen sollen. Wie beide Bedingungen gleichzeitig exakt erfüllt werden können, ist nicht offensichtlich.

Die im Folgenden beschriebene Methode der *Amplitude Adjusted Fourier Transformation* (AAFT) von Theiler et al. [TEL+92] kann die Fourieramplituden aber annähernd in Übereinstimmung bringen, während die Verteilung exakt reproduziert wird. Dabei 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 die ebenfalls gaußverteilten $\{g'_n\}$.

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

Können Surrogate mit gleicher Autokovarianz und Verteilungsfunktion wie die zu untersuchende Zeitreihe konstruiert werden, lautet die entsprechende Nullhypothese: „Die Daten $\{s_n\}$ sind das Ergebnis eines gaußschen, linearen Prozesses x_n, der mit einer monotonen Messfunktion $m(x_n) = s_n$ gemessen wurde.“ Die Monotonie der Messfunktion ist Bestandteil der Nullhypothese, da wir im letzten Schritt des Schemas stillschweigend die Invertierbarkeit der Messfunktion vorausgesetzt haben. Dies ist die einzige bekannte Verallgemeinerung der Nullhypothese für einen Test auf Nichtlinearität in der Sprache der Modellklasse. Weitere Verallgemeinerungen können nur über die Festlegung von weiteren Eigenschaften erreicht werden. Eine sehr flexible und allgemeine Methode, um dies zu erreichen, wird im nächsten Kapitel entwickelt.

Wie erwähnt, wird das Spektrum der Originalzeitreihe nicht exakt auf den Surrogaten reproduziert. In [SS96] wird deutlich gezeigt, dass die Abweichungen des Spektrums speziell für kurze, stark korrelierte Zeitreihen signifikant sind. Das Spektrum der AAFT-Surrogate tendiert dazu, flacher als das Originalspektrum zu sein. Die Reskalierung in den Schritten 2 und 4 ist lediglich eine Approximation der unbekannten Messfunktion, die für endliche Zeitreihen nicht exakt ist. Die Residuen zwischen den durch approximierter und wirklicher Messfunktion gemessenen Werten sind aber zeitlich unabhängig und haben daher ein weißes, flaches Spektrum. Dies führt zu einem additiven, flachen Untergrund und damit einem Abflachen des Spektrums.

1.5.3 Iteratives Verfahren

Die folgende Methode [SS96] erreicht eine wesentlich bessere Annäherung des Spektrums der Surrogate an das Originalspektrum als die AAFT. Der Algorithmus besteht aus einem einfachen Iterationsschema mit zwei abwechselnden Schritten.

1. Zunächst wird eine in ansteigender Reihenfolge sortierte Version $\{\bar{s}_n\}$ der Zeitreihe $\{s_n\}$ erzeugt. Außerdem werden einmalig die Frequenzamplituden $\{A_k = |\tilde{s}_k|\}$ der Zeitreihe berechnet.

2. Die folgende Iteration wird mit einer zufällig permutierten Zeitreihe $\{s_n^{(0)}\}$ gestartet.

3. Im ersten Schritt der Iteration i wird $\{s_n^{(i)}\}$ fouriertransformiert. Die resultierenden Amplituden werden verworfen und durch die Amplituden $\{A_k\}$ der Originalzeitreihe ersetzt. Die Phasen $\{\phi_k\}$ bleiben unverändert. Nach der inversen Transformation erhält man die Sequenz $\{\hat{s}_n^{(i)}\}$.

4. Im zweiten Schritt der Iteration wird die Verteilung angepasst. Dies geschieht wieder über eine Reskalierung über die Rangordnung. Dabei nutzt man die sortierte Zeitreihe $\{\bar{s}_n\}$ und erhält $\{s_n^{(i+1)}\}$.

Die beiden Iterationsschritte 3 und 4 werden abwechselnd wiederholt. Das Spektrum des Surrogates $\{s_n^{(i)}\}$ passt sich dabei immer mehr dem Spektrum der Originaldaten an, während die Verteilung stets exakt übereinstimmt. In [SS96] wird gezeigt, dass die verbleibende Diskrepanz zunächst mit bei steigender Anzahl der Iterationen i sinkt und dann einen konstanten Wert annimmt, der von der Länge der Zeitreihe N abhängt. Dabei ist für größere N die verbleibende Diskrepanz kleiner. Das Erreichen dieses Plateaus in der Spektrumsdiskrepanz könnte als Abbruchkriterium dienen. Nach weiteren Iterationen läuft das Schema allerdings in einen Fixpunkt $\{s_n^{(i+1)}\} = \{s_n^{(i)}\}$, was ein wesentlich einfacheres Abbruchkriterium darstellt. Dieser Fixpunkt wird dann erreicht, wenn die Änderung im ersten Iterationsschritt zu klein ist, um im zweiten Schritt eine Umordnung der Werte zu bewirken. Ist man an Surrogaten interessiert, die eine größtmögliche Übereinstimmung in den Korrelationen zeigen, kann auch nach Schritt 3 abgebrochen und $\{\hat{s}_n^{(i)}\}$ verwendet werden. Eine eventuell starke Abweichung in der Verteilung sollte vor dem Test jedoch ausgeschlossen werden.

Kapitel 2

Flexible Methode zur Surrogaterzeugung

2.1 Probleme der Standardmethoden

Die im vorangegangenen Kapitel besprochenen Standardmethoden zur Erzeugung von Surrogaten für Nichtlinearitätstests weisen einige Nachteile auf. Als erstes ist sicherlich ihre Starrheit zu nennen. Wie ich bereits erwähnt habe, ist bislang keine Erweiterung der Nullhypothese in Bezug auf die Modellklasse bekannt. Aber auch eine Verallgemeinerung unter Einbezug weiterer Eigenschaften ist nicht offensichtlich. In diesem Kapitel werde ich eine Methode entwickeln, mit der beliebige Eigenschaften der Surrogate fixiert werden können. Neben dem Mangel an Flexibilität gibt es aber auch weitere Schwierigkeiten, die bei unzureichender Vorsicht zu falschen Testaussagen führen können. In den folgenden Abschnitten gebe ich zwei Beispiele.

2.1.1 Periodizität

Die bisher diskutierten Methoden zur Surrogaterzeugung basieren alle auf dem Wiener–Khinchin–Theorem (1.11), welches Autokorrelation und Spektrum einer Zeitreihe in Beziehung bringt. Dem entsprechend versuchen diese Methoden die Fourieramplituden der Originaldaten exakt auf den Surrogaten zu reproduzieren, um ihre Autokorrelationen in Übereinstimmung zu bringen. Ein Blick auf Gleichung (1.10) zeigt aber, dass die Fourieramplituden periodisch in k mit der Periode N sind, d. h. $\tilde{s}_{k+jN} = \tilde{s}_k$, $j \in \mathbb{Z}$, gilt. Die gleiche Periodizität zeigt auch das Ergebnis der inversen Fouriertransformationen, nämlich das Surrogat. Demzufolge werden anstelle der erwünschten Autokovarianzen (1.1) die *periodischen* Autokovarianzen

$$\gamma_{\mathrm{p}}(\tau) = \frac{1}{N} \sum_{n=1}^{N} (s_n - \bar{s})(s_{\mathrm{mod}(n+\tau-1,N)+1} - \bar{s}) \tag{2.1}$$

realisiert, wobei die Modulo–Funktion $\mathrm{mod}(a, b)$, der Rest bei Division von a und b, die Periodizität widerspiegelt. Nur in den seltensten Fällen kann man aber tatsächlich annehmen, dass der zu untersuchende Datensatz eine Periode einer unendlich fortgesetzten, periodischen Zeitreihe ist. Zudem ist diese Annahme auch nicht mit der formulierten Nullhypothese zu vereinbaren.

Neben diesen eher „formalen“ Problemen, kann der scheinbar unbedeutende Unterschied zwischen $\gamma(\tau)$ und $\gamma_{\mathrm{p}}(\tau)$ aber zu deutlichen Artefakten in den Surrogaten führen.

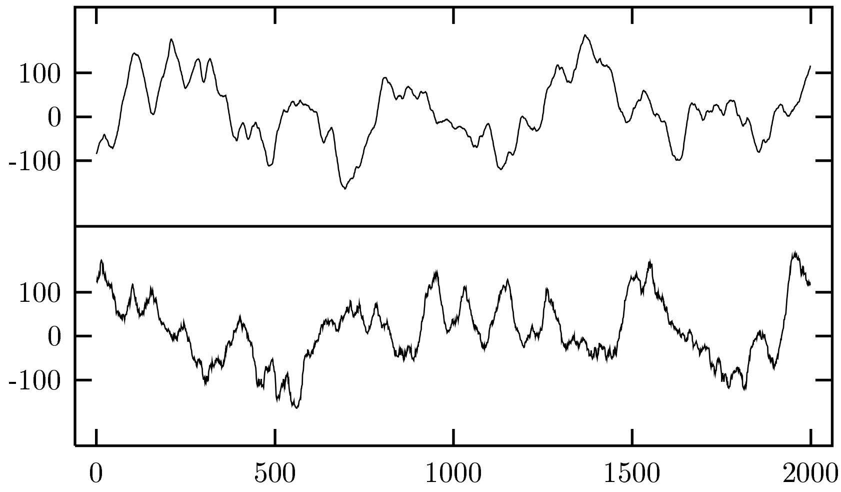

Abbildung 2.1: Auswirkung einer Abweichung zwischen Anfangs– und Endwert einer Zeitreihe auf ein Surrogat. Oben ist eine Realisation des AR(2)–Prozesses mit den Koeffizienten aus Gleichung (2.2) dargestellt. Unten ein Surrogat nach der Methode aus Abschnitt 1.5.3, welches exakt dieselben Fourieramplituden aufweist, aber dennoch deutliche Unterschiede zeigt.

Damit verbunden ist die Gefahr, dass die angewandte Teststatistik empfindlich auf diese Artefakte reagiert und der Test eine falsche Aussage liefert. Insbesondere treten solche Probleme bei einer Abweichung zwischen Anfangs– und Endwert der Zeitreihe auf [TLR93]. Ist man lediglich an einem guten Schätzer für das Spektrum interessiert, können diese Randeffekte z. B. durch Fenstertechniken oder das Anhängen von Nullen minimiert werden. Für die Erzeugung von Surrogaten sind diese Methoden aber nicht anwendbar, da sie nicht invertierbar sind.

Die Auswirkungen einer Abweichung zwischen Anfangs– und Endwert sollen im Folgenden an einem Beispiel demonstriert werden. Dazu betrachte ich eine Realisation des AR(2)–Prozesses

$$x_n = 1.929x_{n-1} - 0.93x_{n-2} + \eta_n \tag{2.2}$$

mit trivialer Messfunktion, d. h. $s_n = x_n$. Die Koeffizienten wurden dabei so gewählt, dass der Prozess nahezu instabil ist und langreichweitige Korrelationen aufweist. In Abbildung 2.1 (oben) ist gut zu erkennen, dass die Zeitreihe relativ glatt ist und ein große Differenz zwischen Anfangs– und Endwert zeigt. Würde die Zeitreihe periodisch fortgesetzt, entspräche diese Differenz einem Sprung, der wesentlich größer ist als typische Differenzen innerhalb der Zeitreihe. Ein solcher Sprung liefert, ähnlich einer Rechteckwelle, in allen Frequenzen Beiträge zum Spektrum. Die entsprechenden Phasen sind dabei so aufeinander abgestimmt, dass der Sprung zeitlich lokalisiert auftritt. In einem Surrogat, das die Amplituden erhält, aber die Phasen durch zufällige ersetzt, werden die Frequenzanteile des Sprungs zeitlich verschmiert. In dem in Abbildung 2.1 unten gezeigten Surrogat ist dies am deutlichsten an dem Spektralanteil in den hohen Frequenzen zu erkennen. Das Surrogat weist nicht die Glattheit der Originalzeitreihe auf. Hingegen zeigt das Surrogat eine nahezu perfekte Übereinstimmung zwischen Anfangs– und Endwert und es existiert kein zeitlich lokaler Sprung mehr.

Die Autokorrelationen $\varrho(\tau) = \gamma(\tau)/\gamma(0)$ und die periodischen Autokorrelationen $\varrho_{\mathrm{p}}(\tau) = \gamma_{\mathrm{p}}(\tau)/\ \gamma_{\mathrm{p}}(0)$ der Originalzeitreihe und eines Surrogates sind in Abbildung 2.2

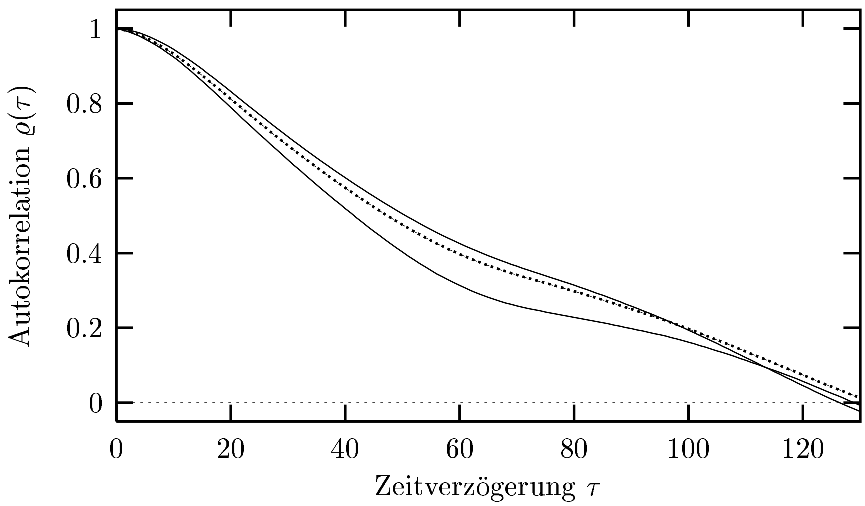

Abbildung 2.2: Diskrepanz zwischen den Autokorrelationen des AR(2)–Prozesses (2.2) (oben) und eines Surrogates (unten). Auch die periodischen Autokorrelationen ϱ_{p} (gestrichelt) weichen von der gewünschten Autokorrelationen ab, stimmen aber zwischen Original und Surrogat überein.

aufgetragen. Die Autokorrelationen $\varrho(\tau)$ des Surrogates weichen dabei zum Teil stark von denen des AR–Prozesses ab, welche für eine bedingte Realisierung aber exakt reproduziert werden sollten. Die *periodischen* Autokorrelationen der Originaldaten und des Surrogates stimmen wie erwartet überein. Verbleibende kleine Abweichungen zwischen ihnen ergeben sich aus dem Iterationschritt, in dem die Verteilung angepasst wird. Würde die Iteration nach der Anpassung der Fourieramplituden abgebrochen, würden die periodischen Autokorrelationen zwar exakt übereinstimmen, aber die gleichen Abweichungen von der gewünschten Autokorrelation $\varrho(\tau)$ zeigen.

Wie fatal diese Abweichungen zwischen den Autokorrelationen sein können, zeigt das Ergebnis des folgenden Tests, in dem ein nichtlinearer Vorhersagefehler als Teststatistik dient. Zunächst werden mithilfe der Iterationsmethode 19 Surrogate für die Zeitreihe aus Abbildung 2.1 erzeugt. Um die Vorhersagen (1.7) zu berechnen, werden die Zeitreihen in $m = 2$ Dimensionen eingebettet. Die Zeitverzögerung der Einbettung wird $\tau = 120$ gewählt, was ungefähr der ersten Nullstelle der Autokorrelation entspricht. Der Nachbarschaftsradius r wird auf ein Viertel der Standardabweichung gesetzt. Es sei betont, dass diese Wahl der Parameter nicht außergewöhnlich ist und gängiger Praxis entspricht. Der sich daraus ergebende Vorhersagefehler (1.5) der Originalzeitreihe ist dabei deutlich kleiner als die Vorhersagefehler für die 19 Surrogate, was einem Verwerfen der Nullhypothese mit einem Signifikanzniveau von 95% entspricht. Der Test „erkennt“, dass der Sprung nur sehr unwahrscheinlich das Resultat eines linearen Prozesses sein kann.

Natürlich hätte in diesem Fall eine visuelle Inspektion der Surrogate gereicht, um eine deutliche Abweichung von der Originalzeitreihe zu erkennen. Es wurde auch bewusst eine Zeitreihe gewählt, deren Differenz zwischen Anfangs– und Endwert so groß ist, dass die zusätzlichen Frequenzanteile im Surrogat leicht zu erkennen sind. Wäre die Abweichung zwischen Anfangs– und Endwert kleiner, könnten aber unter Umständen keine Unterschiede mehr zwischen Daten und Surrogaten zu sehen sein. Eine Teststatistik wie der Vorhersagefehler, der empfindlich auf Abweichungen in den Autokorrelationen reagiert, könnte aber dennoch die Unterschiede aufspüren und ein Verwerfen der Nullhypothese bewirken.

Eine einfache Möglichkeit, Effekte auszuschließen, die aus einer Abweichungen zwischen Anfangs– und Endpunkt resultieren, besteht darin, Werte an den Enden abzuschneiden. Die Anzahl der verworfenen Werte am Anfang und Ende der Zeitreihe sollten dabei so gewählt werden, dass eine möglichst lange Teilsequenz entsteht, deren Endwerte näherungsweise übereinstimmen. Dadurch werden Artefakte in den Surrogaten, die aus einem untypisch großen Sprung in der Originalzeitreihe resultieren, minimiert. Die unter Umständen verbleibende Diskrepanz zwischen den Steigungen am Anfang und Ende der Teilsequenz spielt meist eine wesentlich geringere Rolle. Ist die Zeitreihe aber relativ glatt und der periodische Übergang vom Ende zum Anfang untypisch spitz, sollten auch die Steigungen an den Enden in Übereinstimmung gebracht werden.

Das Anpassen der Enden ist eine sehr einfache, aber wichtige Vorsichtsmaßnahme, die in vielen Fällen ausreicht, falsche Aussagen eines Nichtlinearitätstests zu vermeiden. Nichtsdestotrotz werden auf den Surrogaten aber weiterhin die periodischen Autokorrelationen realisiert, die auch bei einem Übereinstimmen der Enden starke Abweichungen von den gewünschten Autokorrelationen aufweisen können. Eine neue Methode, die tatsächlich die Autokorrelationen $\varrho(\tau)$ auf den Surrogaten realisiert, wird in Abschnitt 2.2 entwickelt. Allerdings stellt diese Methode wesentlich höhere Ansprüche an die Computerleistung und braucht für gewöhnlich mehr Zeit zum Erzeugen einzelner Surrogate. Die Wahl der Methode hängt also vom jeweiligen Fall ab. Sind die Daten mit einem großen zeitlichen oder finanziellen Aufwand aufgenommen worden oder gar nur einmal messbar (wie z. B. in der Geologie), wird der zusätzliche zeitliche Aufwand bei der Auswertung der Daten kaum ins Gewicht fallen.

2.1.2 Nichtstationarität

Ein wesentlicher Bestandteil der Nullhypothese, der bisher nicht erwähnt wurde, ist die Stationarität. Die Surrogate, die auf den Methoden aus Kapitel 1 basieren, sind stets stationär. Dies lässt sich am leichtesten daran erkennen, dass stets nur die Fourieramplituden erhalten bleiben, während die Phasen durch zufällige ersetzt werden. Die Fourieramplituden sind aber ein Bestandteil der Repräsentation der Daten im Frequenzraum und enthalten allein keine zeitliche Information. Nach der inversen Fouriertransformation erhalten wir eine Linearkombination aus unabhängigen, periodischen und stationären Funktionen, die damit ebenfalls stationär ist.

Die Tatsache, dass die Nullhypothese also die Stationarität beinhaltet, lässt vermuten, dass *jegliche* Form von Nichtstationarität in der zu untersuchenden Zeitreihe zu einem Verwerfen der Nullhypothese führt. Dies trifft aber nicht immer zu und hängt in starkem Maße von der verwendeten Teststatistik ab. Die Größen, die als Teststatistik für Nichtlinearitätstests dienen können, setzen in nahezu allen Fällen Stationarität voraus und liefern für nichtstationäre Zeitreihen meist keine sinnvollen Werte. So können z. B. für ein dynamisches System, dessen Parameter sich zeitlich stark ändern, keine sinnvollen Vorhersagen über die lokal konstante Vorhersage (1.7) getroffen werden. Die im Phasenraum zu unterschiedlichen Zeiten wiederkehrende Trajektorie würde zu jedem Zeitpunkt einer Dynamik mit anderen Parametern folgen und sinnvolle Vorhersagen unmöglich machen. Die Nichtstationarität sorgt also in einem solchen Fall für größere Vorhersagefehler und senkt damit die Wahrscheinlichkeit für ein Verwerfen der Nullhypothese. In [SS99a] haben wir das Beispiel eines AR(1)–Prozesses betrachtet, dessen Parameter linear mit der Zeit variieren. In 1000 Test mit unterschiedlichen Realisierungen dieses Prozesses war in 350 Fällen der Vorhersagefehler (1.5) größer als die Vorhersagefehler der jeweils neun Surrogate. Nur in 2.3% der Tests wurde die Nullhypothese verworfen. Diese Senkung der Verwerfwahrscheinlichkeit könnte allerdings auch bei einem nichtstationären, *nichtlinearen* Prozess dazu führen, dass die Nullhypothese fälschlicherweise akzeptiert wird.

Anders sieht es bei der Verwendung von Schätzern für die Dimension als Teststatistik aus. Für gaußsche, lineare Prozesse erwartet man im Einbettungsraum eine Punktwolke, die sich in alle Raumrichtungen ausdehnt. Die Dimensionsschätzer liefern für die Surrogate also schlicht die Einbettungsdimension. Die Einbettungsdimension ist aber natürlich auch die obere Schranke für den Wert, den ein Schätzer für die Dimension liefern kann. Dimensionsschätzer können also im Gegensatz zum Vorhersagefehler für die Originalzeitreihe nicht größer sein als für die Surrogate. Verschiedene Beispiele mit nichtstationären, linearen Prozessen zeigten, dass bei Verwendung von Dimensionsschätzern als Teststatistik, die Nullhypothese mit hoher Wahrscheinlichkeit verworfen wurde [Tim98, SS99a].

Diese Beispiele zeigen, wie schwierig der Umgang mit nichtstationären Zeitreihen speziell im Hinblick auf Nichtlinearitätstests ist. Es wäre wünschenswert, eventuelle Nichtstationarität möglichst vor einem Test auf Nichtlinearität ausschließen zu können. Dies ist aber aufgrund der vielfältigen Quellen und Formen von Nichtstationaritäten nahezu unmöglich. Grundlegende Vorsichtsmaßnahmen sollten aber in jedem Falle getroffen werden. Dabei erweist sich als erster Ansatz oft schon die visuelle Inspektion der Zeitreihe und ihrer Repräsentationen in höhere Dimensionen als sehr hilfreich. Zur quantitativen Analyse von Nichtstationarität existieren zahlreiche Methoden wie z. B. Rekurrenzplots [EKR87], [MS96], usw. Ist die Zeitreihe hinreichend lang, kann sie unterteilt werden und die charakterisierenden Größen auf den einzelnen Segmenten gemessen und verglichen werden. Starke Schwankungen zwischen den Segmenten lassen dann Nichtstationarität vermuten. Gibt es Anhaltspunkte dafür, dass die Dynamik sich innerhalb einer langen Zeitreihe oder zwischen verschiedenen Datensatz nicht wesentlich ändert, kann der Test auf Nichtlinearität auch lediglich auf *einem* Segment durchgeführt werden, welches annähernd stationär ist. Ein Verwerfen der Nullhypothese auf diesem Segment würde den Einsatz von nichtlinearen Methoden auf dem gesamten Datensatz rechtfertigen.

Steht allerdings nur eine einzelne kurze Zeitreihe zur Verfügung, für die Nichtstationarität nicht ausgeschlossen werden kann, ist die Herausnahme eines Segmentes aber nicht mehr praktikabel. Eine sinnvolle Berechnung der Teststatistik auf einem einzelnen kurzen Segment ist dann nicht mehr möglich. In diesen Fällen kann sich aber unter Umständen die Anwendung der Methode aus [SS99a] lohnen, mit der nichtstationäre Surrogate erzeugt werden können. Dabei wird die zu untersuchende Zeitreihe in Segmente unterteilt und für jedes Segment individuell ein Surrogat nach den Standardmethoden erzeugt. Die Teilsurrogate werden anschließend durch einfaches Aneinanderreihen zu einem langen Surrogat zusammengefügt. Da das resultierende Surrogate auf der Zeitskala der Segmentlänge die gleiche Nichtstationarität wie die Originalzeitreihe zeigt, kann die Teststatistik auf der ganzen Länge der Zeitreihe berechnet werden. Eventuelle Abweichungen der Teststatistik aufgrund von Nichtstationaritäten wären dann auf der Originalzeitreihe und den Surrogaten gleich. Auch bei dieser Methode sollte aber die Segmentlänge vorsichtig gewählt werden. Sie sollte einerseits kürzer als die typische Zeitskala der Nichtstationarität sein. Andererseits gehen zwischen den Segmenten jegliche Korrelationen verloren, sodass bei einer Einteilung in viele kurze Segmente die Korrelationen des zusammengefügten Surrogates stark von denen der Originalzeitreihe abweichen können. Dies könnte sich wiederrum in den Werten für die Teststatistik widerspiegeln. In [SS99a] gebe ich einige Beispiele, in denen diese Methode bei geeigneter Wahl der Segmentlänge zum Erfolg führt.

Mit der im nächsten Abschnitt vorgestellten Methode können beliebige Eigenschaften der Originalzeitreihe auch auf den Surrogaten reproduziert werden. Sind die Formen der Nichtstationaritäten des zu untersuchenden Datensatz bekannt, können dadurch auch ähnlich nichtstationäre Surrogate erzeugt werden. Obwohl die sinnvolle Berechnung von Teststatistiken auf nichtstationären Zeitreihen fraglich ist, bleibt der Test so immerhin ausgewogen und besitzt die korrekte nominelle Size. Die Nullhypothese muss dann natürlich im Sinne der implementierten Nichtstationarität modifiziert werden.

2.2 Simulated Annealing Methode

Die Standardmethoden des vorangegangen Kapitels sind in den meisten Fällen, in denen Daten auf Nichtlinearität getestet werden sollen, ausreichend. Allerdings weisen sie einige Probleme auf, von denen einige bereits diskutiert wurden. Der größte Nachteil ist aber sicherlich die stark eingeschränkte Klasse an möglichen Nullhypothesen, für die ein Test durchgeführt werden kann. Die Fourieramplituden und eine eventuell vorhandene, monotone Messfunktion sind sie einzigen bekannten, interessanten Eigenschaften, für die entsprechende inverse Transformationen existieren. Nur dadurch ist es möglich, diese Eigenschaften auf ansonsten zufälligen Surrogaten zu realisieren. Eine allgemeine Methode, die nicht auf der Invertierbarkeit der Eigenschaften beruht, wäre also wünschenswert.

Für einen Surrogattest auf eine einzelne interessierende Eigenschaft müssten idealerweise Surrogate erzeugt werden, in denen lediglich diese eine zu untersuchende Eigenschaft zerstört ist. Alle anderen Eigenschaften sollten möglichst erhalten bleiben. Diese Herangehensweise ist aber nur für äußerst einfache Eigenschaften durchführbar und kaum für abstrakte Eigenschaften wie z. B. Nichtlinearität oder Stationarität zu realisieren. Einfacher ist es, auf einem zunächst zufälligen Surrogat Schritt für Schritt Eigenschaften zu implementieren, die bereits bekannt oder für die Analyse uninteressant sind. Dabei sollte mit den einfachsten Eigenschaften begonnen werden, um einfache Modelle auszuschließen, bevor komplizierte Alternativen in Erwägung gezogen werden. Diese Vorgehensweise ist auch als *Occam's Razor* bekannt. Eine Methode [Sch98], mit der dies möglich ist, wird im Folgenden beschrieben.

Die interessierenden Eigenschaften müssen zunächst quantifiziert werden und als endliche Menge von Observablen $\{O_j\}$, die auf einer Zeitreihe gemessen werden können, formuliert werden. Diese Größen lassen sich meist leicht als Bedingungen der Art

$$B_j(\{s'_n\}) = 0, \quad j = 1, \dots, J \tag{2.3}$$

an das Surrogat $\{s'_n\}$ umschreiben. In vielen Fällen können diese Bedingungen einfach über die Differenz der Größen zwischen Surrogat und Originalzeitreihe $\{s_n\}$

$$B_j(\{s'_n\}) = O_j(\{s'_n\}) - O_j(\{s_n\}) \tag{2.4}$$

definiert werden. Wenn die Observable O_j eine zu untersuchende Eigenschaft charakterisiert, dann stimmt diese Eigenschaft auf Surrogat und Originalzeitreihe überein, wenn B_j verschwindet. Die Erfüllung von $B_j = 0$ für alle j ist also eine direkte Umsetzung der bedingten Realisierung, die in Abschnitt 1.4.2 beschrieben wurde. Diese meist große Anzahl von Bedingungen kann stets in die Form einer einzelnen, reellen *Kostenfunktion*

$$E(\{s'_n\}) = \left(\sum_{j=1}^{J} |w_j B_j(\{s'_n\})|^q \right)^{1/q} \tag{2.5}$$

gebracht werden, welche ein globales Minimum annimmt, wenn *alle* Bedingungen erfüllt sind. Dies ist unabhängig von der Wahl der Gewichte $w_i \neq 0$ und der Ordnung q der Mittelwertbildung bzw. der Normierung des Abstandsvektors $\vec{B}$. Durch die Einstellung von q kann die verwendete Norm gewählt werden. So erhält man für $q = 2$ den Euklidischen Abstand oder die L^2–Norm, für $q = 1$ die L^1– bzw. Manhattan–Norm und im Limes $q \to \infty$ trägt nur noch die größte Komponente von $\vec{B}$ bei und man erhält die Maximumsnorm. Formell ist auch das geometrische Mittel im Limes $q \to 0$ möglich. Über die Wichtung lassen sich die relativen Größenverhältnisse der Komponenten B_j ausgleichen oder Prioritäten auf bestimmte Bedingungen setzen, die unter Umständen schwerer zu realisieren sind. Eine pauschal geeignete Wahl dieser Parameter lässt sich aber nicht angeben und muss von Fall zu Fall entschieden werden. Eines der wichtigsten Kriterien wird dabei der benötigte Rechenaufwand sein.

Sollen zum Beispiel die Autokovarianzen (1.1) zwischen Surrogat und Originalzeitreihe übereinstimmen, wie z. B. in einem Test auf Nichtlinearität, könnte die Kostenfunktion

$$E(\{s'_n\}) = \sum_{\tau=0}^{\tau_{\max}} |\gamma'(\tau) - \gamma(\tau)| \tag{2.6}$$

lauten. Dabei wurden hier die Gewichte $w_\tau = 1$ gleichverteilt und $q = 1$ gewählt. Außerdem ist man diesem Fall nur an einer Realisierung der Autokovarianzen bis zu einer Zeitverzögerung von $\tau = \tau_{\max}$ interessiert. Dies könnte z. B. durch vernachlässigbare, langreichweitige Korrelationen begründet sein.

Das gesuchte, globale Minimum der Kostenfunktion E in Gleichung (2.5) muss im Prinzip durch eine Minimierung über *alle* möglichen Surrogate $\{s'_n\}$ gefunden werden. Eine Suche in diesem unendlichen und hochdimensionalen Raum ist selbst mit geschicktesten Minimierungsmethoden nahezu hoffnungslos. Eine zusätzliche Bedingung kann dieses Problem aber lösen. In vielen Fällen sollen die Surrogate der gleichen Verteilung folgen wie die Originalzeitreihe, d. h. sie sollen beide gleichwahrscheinlich das Resultat einer unbekannten, zugrunde liegenden Wahrscheinlichkeitsverteilung sein. Dies gilt, wie im vorangegangen Kapitel erwähnt, auch im Falle von Tests auf Nichtlinearität. Dort waren somit im Sinne der bedingten Realisierung alle Surrogate schlicht Permutationen der Originalzeitreihe. Es wurde also lediglich die zeitliche Reihenfolge der N Zeitreihenwerte vertauscht. Die Anzahl der möglichen Permutation $N!$ ist zwar immer noch sehr groß, aber endlich. Dieser Wechsel von einem unendlichen zu einem endlichen Konfigurationsraum erleichtert die Minimierung erheblich.

Durch die Einführung einer Kostenfunktion wurde das Problem (2.3), die Nullstellen einer nichtlinearen, hochdimensionalen Funktion zu finden, in ein Minimierungsproblem (2.5) umgewandelt. Im Allgemeinen ist dieser Weg für die Nullstellensuche von Funktionen mit kontinuierlichen Argumenten nicht empfehlenswert. Die resultierende Kostenfunktion besitzt viele lokale Minima, an denen man häufig nicht interessiert ist und die die Suche nach dem globalen Minimum erschweren. Für diskrete und beschränkte Argumente, wie im vorliegenden Fall, existieren aber adäquate Algorithmen, mit denen sich Minimierungsprobleme, auch mit einer großen Anzahl von lokalen Minima, behandeln lassen.

2.2.1 Minimierung der Kostenfunktion

Zur Minimierung von Funktionen existiert eine Fülle von Methoden, welche recht umfassend in [PTVF92] behandelt werden. Die Wahl der Methode hängt stark von dem individuellen Problem ab, wobei Kriterien wie z. B. die Verfügbarkeit der ersten Ableitung oder die Dimension des Argumentenraums die Entscheidung maßgeblich beeinflussen. Einige Methoden basieren auf einem vorgegebenem Startpunkt, von dem ausgehend sie schnellstmöglich das nächstliegende, lokale Minimum finden. Eine solche Vorgehensweise ist im vorliegenden Fall natürlich nicht anwendbar, da wir das globale Minimum einer Funktion suchen, die viele lokale Minima aufweist. Zwei bekannte Methoden, mit denen sich solche Funktionen behandeln lassen, sind die *genetischen Algorithmen* und die Methode des *Simulated Annealing* [KJV83, Vid93]. Während die erste auf einem Modell für die biologische Evolution mit Mutationen und Reproduktionen basiert, hat das Simulated Annealing einen langsamen Abkühlungprozess zur Grundlage. Im Labor können über einen solchen Prozess reine Einkristalle erzeugt werden, deren Ausdehnung übliche Kristalle um Größenordnungen übersteigt. Solche Einkristalle stellen einen energetisch sehr niedrigen Zustand dar, woraus sich der Name dieser Methode ableitet.

Um die Methode des Simulated Annealing anzuwenden, wird das Problem als thermodynamisches System betrachtet, wobei die Kostenfunktion die Rolle der Energie übernimmt. Durch langsames Senken der Temperatur T gelangt das System in seinen Grundzustand, d. h. in den Zustand minimaler Energie. Der aktuelle Zustand des Systems entspricht dem Argument der Kostenfunktion (2.5), was im vorliegenden Fall ein Surrogat ist. Das gesuchte Surrogat, das die gewünschten Bedingungen erfüllt, entspricht dem Grundzustand. Die Minimierungsmethode besteht nun darin, dass man von einem Zustand ausgehend neue Konfigurationen des Systems erzeugt und deren Energie mit der des Ausgangszustandes vergleicht. Im Falle der Zeitreihen können neue Surrogate einfach dadurch erzeugt werden, indem ein oder mehrere zufällige Wertepaare vertauscht werden. Ein solcher Wertetausch ändert den Wert der Kostenfunktion um ΔE, während die Verteilung des Surrogates davon unberührt bleibt. Die Wahrscheinlichkeit p dafür, dass die neue Konfiguration akzeptiert wird und als Ausgangspunkt für weitere Änderungen dient, hängt natürlicherweise von der Energieänderung ΔE und der Temperatur T ab. Befindet sich ein System im thermodynamischen Gleichgewicht, werden die Zustände mit der Wahrscheinlichkeit besucht, die durch die Boltzmann–Verteilung $e^{-E/T}$ gegeben ist. Dies kann in einer Simulation dadurch erreicht werden, dass Änderungen mit der Wahrscheinlichkeit

$$p(\Delta E, T) = \begin{cases} e^{-\Delta E/T} & \text{für } \Delta E > 0 \\ 1 & \text{für } \Delta E \leq 0 \end{cases} \tag{2.7}$$

akzeptiert werden. Diese Idee geht auf Metropolis et al. [MRR+53] zurück und wird häufig auch *Metropolis–Algorithmus* genannt. Dabei ist von entscheidender Bedeutung, dass auch Konfigurationen mit höherer Energie ($\Delta E > 0$) akzeptiert werden können. Nur dadurch ist es möglich aus einem lokalen Minimum herauszukommen. Eine weitere wichtige Eigenschaft des Algorithmus ist die Ergodizität, die unabhängig von der Temperatur T garantiert, dass jeder Zustand von jedem anderen Zustand aus erreicht werden kann. Wird die Temperatur nun genügend langsam gesenkt, sodass sich das System immer nahe dem thermodynamischen Gleichgewicht befindet, endet das System für $T = 0$ im Grundzustand bzw. im globalen Minimum der Kostenfunktion.

Zusammenfassend lassen sich Surrogate mit dieser Methode wie folgt erzeugen. Man beginnt zunächst mit einer zufälligen Permutation der Originalzeitreihe, was einer „heißen“ Konfiguration mit hoher Temperatur T und Energie E entspricht. Die Kostenfunktion einer neuen Konfigurationen, die über den Tausch von Werten des Surrogates erreicht werden, werden mit dem jeweils alten Wert der Kostenfunktion verglichen, d. h. ΔE berechnet. Die neue Konfiguration wird anschließend in Abhängigkeit von ΔE mit der

Wahrscheinlichkeit p (2.7) akzeptiert. Die Temperatur wird langsam gesenkt und die Kostenfunktion des Surrogates nimmt kontinuierlich kleinere Werte an. Der Vorgang wird abgebrochen, wenn die Kostenfunktion E unter eine vorher festgelegte Schwelle gesunken ist. Die genaue Vorgehensweise wird im Folgenden erläutert.

2.2.2 Wahl der Parameter

Die im vorangegengen Abschnitt beschriebene Methode des Simulated Annealing zur Minimierung beinhaltet einige Parameter und Freiheiten, die hier erläutert werden sollen. Dazu zählen z. B. die Wahl der Anfangstemperatur und das Schema zur Senkung der Temperatur. Aber auch auf eine geeignete Wahl der Updates, d. h. der Erzeugung von neuen Konfigurationen, wurde bis jetzt nicht näher eingegangen. Zu diesen offenen Fragen existieren keine allgemein gültigen Antworten. Ich werde deshalb im Folgenden speziell auf das Problem der Surrogaterzeugung eingehen und die aus Erfahrung gewonnenen Ergebnisse diskutieren.

Relativ einfach lässt sich ein guter Richtwert für die Anfangstemperatur T_0 bestimmen. Wird sie zu niedrig gewählt, läuft das System schnell in ein lokales Minima. Sie sollte also zu Beginn so gewählt werden, dass sich das System leicht über den ganzen Konfigurationsraum bewegen kann. Um entsprechend hohe Akzeptanzwahrscheinlichkeiten p zu erreichen, sollte T_0 nach Gleichung (2.7) größer sein als die maximal mögliche Energiedifferenz ΔE. Diese Energiedifferenz kann bestimmt werden, indem mehrere zufällige, unabhängige Konfigurationen erzeugt und deren Energiedifferenzen berechnet werden. Als Richtwert für T_0 kann dann der maximale Wert dieser Energiedifferenzen oder deren Mittelwert dienen. Die spätere Anwendung dieser Idee zeigte, dass die durch diese Methode bestimmte Anfangstemperatur in allen Fällen zu hoch war und eine erste wesentliche Änderung der Kostenfunktion erst nach starkem Absenken der Temperatur zu erkennen war. Da für gewöhnlich mehrere Surrogate erzeugt werden sollen, kann man die Temperatur, ab der die Kostenfunktion sinkt, bei der Erzeugung des ersten Surrogates ablesen und für die weiteren Surrogate verwenden. Dadurch lassen sich unnötige Annealingschritte vermeiden.

Ein weitaus schwieriger Punkt ist die Wahl des genauen Verlaufs der Abkühlung. Wird die Temperatur zu schnell gesenkt, kann das System in einem lokalen Minimum hängenbleiben. Je langsamer das Abkühlen vonstatten geht, um so niedriger ist für gewöhnlich der am Ende erreichte Wert der Kostenfunktion. Es ist sinnvoll, eine Schranke E_{stop} für die Kostenfunktion vorzugeben, bei deren Unterschreiten der Prozess beendet wird. Eine solche Schranke ist im Fall der Surrogate durch die Genauigkeit gegeben, mit der die gewünschten Eigenschaften der Originalzeitreihe auf den Surrogaten reproduziert werden sollen. Das Abkühlen sollte dann so eingestellt werden, dass E_{stop} in möglichst wenigen Annealingschritten erreicht wird. Ich verwende im Folgenden das in [PTVF92] vorgeschlagene, exponentielle Schema zur Abkühlung. Dabei wird die Temperatur entweder nach S_{total} Metropolisschritten oder nach S_{ak} akzeptierten Updates um einen Faktor $\alpha < 1$ auf $T' = \alpha T$ gesenkt. Während α im Wesentlichen den Abfall der Temperatur bestimmt und frei gewählt werden kann, sollten S_{total} und S_{ak} proportional zur Systemgröße N und zur Anzahl der Bedingungen J gewählt werden. Während ein großes S_{total} dafür sorgt, dass das System für jede Temperatur in ein thermodynamisches Gleichgewicht laufen kann, wird durch ein $S_{\text{ak}} < S_{\text{total}}$ ein unnötiger Aufenthalt bei hohen Temperaturen verhindert. Die genaue Wahl dieser Parameter werde ich im jeweiligen Fall angeben. Im weiteren dient der Wert von S_{ak} auch als Abbruchkriterium. Werden keine Updates mehr akzeptiert ($S_{\text{ak}} = 0$), hängt das System in einem Minimum. Ist der entsprechende Wert der Kostenfunktion aber noch zu hoch, muss das System wieder „geschmolzen“ werden oder die Prozedur mit einer langsameren Abkühlung wiederholt werden.

2.2.3 Optimierung des Verfahrens

Wird das im vorangegangen beschriebene Verfahren zur Surrogaterzeugung naiv implementiert, bemerkt man bald seinen wesentlich Nachteil. Die Methode benötigt erheblich mehr Rechenzeit als die Standardmethoden aus Kapitel 1. Es sind aber einige Änderungen möglich, die die Rechenzeit zum Teil um Größenordnungen reduzieren.

Wie bereits erwähnt, können neue Konfigurationen einfach durch den Austausch zweier zufälliger Werte der Zeitreihe erzeugt werden. Um zu entscheiden, ob die neue Konfiguration akzeptiert werden kann, muss die Energiedifferenz ΔE bestimmt werden. Dafür ist es aber in vielen Fällen nicht nötig die Kostenfunktion sowohl vor als auch nach dem Update zu berechnen und miteinander zu vergleichen. Vielmehr kann die Kostenfunktion für die neue Konfiguration oft aus dem alten Wert ermittelt werden. Dies gilt zum Beispiel für die Autokovarianz $\gamma(\tau)$. Werden beim Update die Werte mit den Zeitindizes i und j getauscht, ändern sich in der Summe (1.1) der Autokovarianz für ein bestimmtes τ nur vier Terme, nämlich die für $n = i, j, i - \tau$ und $j - \tau$. Anstelle der $N - \tau$ Terme müssen also nur vier Terme neu berechnet werden, wodurch speziell für lange Zeitreihen eine immense Geschwindigkeitssteigerung erzielt werden kann. Bei der Implementierung dieser Direktberechnung der Kostenfunktion sollte man allerdings Vorsicht walten lassen. Zum einen müssen Fälle wie $i, j \leq \tau$ und $|i - j| = \tau$ gesondert behandelt werden. Zum anderen könnten sich Rundungsfehler in der Berechnung der Energiedifferenzen über mehrere Updates hinweg fortpflanzen. Um dies auszuschließen, kann die Kostenfunktion von Zeit zu Zeit, z. B. nach S_{total} Schritten, komplett neu berechnet werden.

Eine weitere Rechenzeitersparnis kann durch eine weitere Überlegung erzielt werden, die in vielen Fällen die Berechnung von einigen Summanden in der Kostenfunktion (2.5) überflüssig macht. Dabei nutzt man aus, dass für die Entscheidung über das Akzeptieren eines Updates eine gleichverteilte Zufallszahl $R \in [0, 1]$ gewürfelt wird, die selbstverständlich auch vor der Berechnung von ΔE ermittelt werden kann. Eine neue Konfiguration wird nach Gleichung (2.7) akzeptiert, wenn $\exp(-\Delta E/T) \geq R$ gilt. Setzt man $\Delta E = E_{\text{neu}} - E_{\text{alt}}$ ein, löst nach E_{neu} auf und potenziert mit q, ergibt sich:

$$E_{\text{neu}}^q \leq (E_{\text{alt}} - T \ln R)^q \,. \tag{2.8}$$

Ein Vergleich mit der Gleichung der Kostenfunktion (2.5) zeigt, dass der linke Term in (2.8) nun lediglich eine Summe über positive Summanden ist. Vergleicht man während des Aufsummierens den jeweiligen Wert der Summe mit der rechten Seite von Gleichung (2.8), kann die Summation abgebrochen werden, sobald die Ungleichung verletzt wird. In diesem Fall wird die neue Konfiguration frühzeitig verworfen.

Bisher bestand ein Update aus einem Tausch zweier Werte der Zeitreihe. Ein solches Update resultiert zu Beginn der Abkühlung bei hoher Temperatur in einer kleinen relativen Änderung der Kostenfunktion und wird mit hoher Wahrscheinlichkeit akzeptiert. Am Ende der Prozedur bedeutet ein Tausch zweier zufälliger Werte aber eine relative große Änderung der Kostenfunktion und wird bei kleinen Temperaturen häufig abgelehnt. Die *Akzeptanzrate* A, der Quotient aus der Anzahl akzeptierter und vorgeschlagener Updates, sinkt also mit der Temperatur T im Laufe des Abkühlvorgangs. Wünschenswert ist aber eine konstante Akzeptanzrate von $A \approx 50\%$ [Bha88], was einen Kompromiss zwischen zu kleinen und zu großen Änderungen darstellt. Ist die aktuelle Akzeptanzrate zu hoch, kann sie gesenkt werden, indem zur Erzeugung neuer Konfigurationen *mehrere* Wertepaare der Zeitreihe getauscht werden. Auch im entgegengesetzten Fall einer zu kleinen Akzeptanzrate gibt es eine Ausgleichsmöglichkeit. Dafür werden die beiden Werte der Zeitreihe, die getauscht werden sollen, mit einer Wahrscheinlichkeit gewürfelt, die von ihrer Größendifferenz bzw. von ihrem Unterschied im Rang abhängt. Der Rang $r(s_n)$ eines Wertes s_n der Zeitreihe ist seine Position bzw. sein Index in der entsprechenden, sortierten Zeitreihe. Der kleinste Wert hat also Rang $r = 1$ und der größte Rang $r = N$. Der Tausch zweier Werte mit großem Rangunterschied resultiert im Allgemeinen in einer größeren Änderung der Kostenfunktion als der Tausch zweier Werte mit annähernd gleichem Rang. Dies gilt z. B. auch für die Autokovarianz, die in nahezu allen folgenden Anwendungen Bestandteil der Kostenfunktion sein wird. Um also die Akzeptanzrate zu erhöhen, können Wertepaare $\{s_i, s_j\}(i \neq j)$ mit einer Wahrscheinlichkeit $p_{ij}(d, \mu)$ gewählt werden, wobei

$$d = |r(s_i) - r(s_j)| - 1 \tag{2.9}$$

der Rangunterschied ist. Die Wahrscheinlichkeitsverteilung p_{ij} sollte dabei so gewählt werden, dass sie für $d = 0$ ein Maximum aufweist und mit steigendem d abfällt. Der Parameter μ charakterisiert die „Breite“ dieser Verteilung und sollte proportional zu N/T variiert werden. Die Implementierung dieser von d abhängenden Verteilung p_{ij} brachte in allen betrachteten Fällen eine erhebliche Geschwindigkeitssteigerung, sodass tiefere Minima mit gleichem Rechenaufwand erreicht werden können. Allerdings scheint die exakte Form der Wahrscheinlichkeitsverteilung p_{ij} dabei keine entscheidenende Rolle zu spielen. So habe ich z. B. keinen Geschwindigkeitsunterschied zwischen einer exponentiellen und einer Gaußverteilung feststellen können.

Schließlich ist auch bei der Durchführung des Tests selbst, eine relativ einfache Rechenzeitersparnis möglich. Benötigt man für einen Test mit einem bestimmten Signifikanzniveau z. B. B Surrogate, so erhöht sich auch der Rechenaufwand um das B–fache. Die Rechenzeit zur Berechnung der Teststatistik fällt normalerweise im Vergleich zur Erzeugung der Surrogate kaum ins Gewicht. Im einseitigen Test wird die Nullhypothese verworfen, wenn der Wert der Teststatistik für alle Surrogate größer (oder kleiner) ist als der der Originalzeitreihe. Liegt aber schon für das erste Surrogat des Wert der Teststatistik unter (über) dem der Originalzeitreihe, kann die Nullhypothese nicht verworfen werden. Die zeitaufwendige Erzeugung weiterer Surrogate ist dann natürlich nicht mehr nötig. Ähnlich verhält es sich bei einem zweiseitigen Test, bei dem unter Unterständen schon zwei Surrogate ausreichen, um die Nullhypothese zu akzeptieren.

2.2.4 Einfaches Beispiel

Die Funktionsweise der Simulated Annealing Methode und die Wahl der Parameter soll im Folgenden an einem Beispiel illustriert werden. Als Datensatz dient der monatlich bestimmte *Southern Oscillation Index* (SOI) von 1876 bis 1999, was 1483 Werten entspricht. Dieser Index berechnet sich aus Luftdruckdifferenzen über dem Pazifik und spielt bei der Analyse des Klimaphänomens *El Niño* eine wichtige Rolle [GMB87a, GMB87b]. Diese Zeitreihe ist ein Beispiel für einen Datensatz, dessen zugrunde liegender Prozess zwar als nichtlinear angenommen werden kann, statistische Tests auf Nichtlinearität dafür aber keine Bestätigung finden [SS00]. Eine solche eine skalare Zeitreihe, die über ein System mit vielen Freiheitsgraden mittelt, ist in der Analyse unter Umständen nicht von einem linearen, stochastischen Prozess unterscheidbar. In Abbildung 2.3 ist zu sehen, dass der

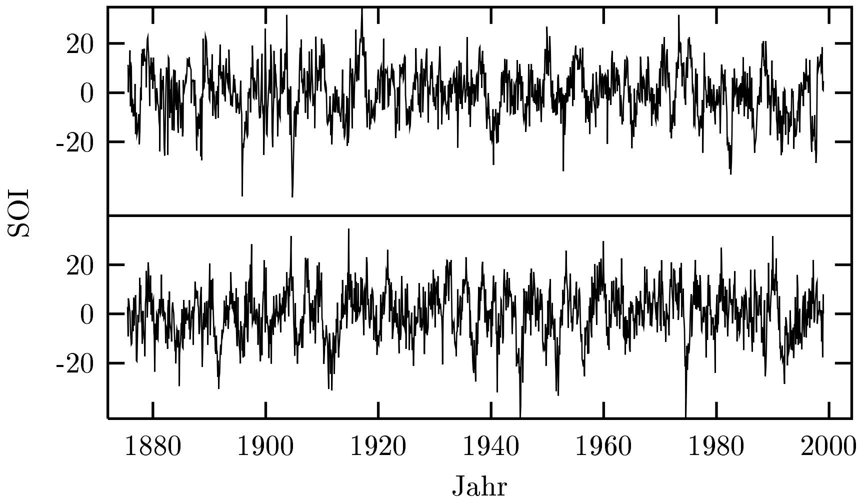

Abbildung 2.3: Zeitreihe des Southern Oscillation Index (SOI) und ein Surrogat (unten), das die gleichen Autokorrelationen besitzt.

SOI keine außergewöhnlichen Merkmale zeigt und nichts dagegen spricht einen Surrogattest mit den Standardmethoden durchzuführen. Der Datensatz dient hier lediglich als einfaches Beispiel, um erste Eigenschaften der neuen Methode zu illustrieren.

Um Surrogate für einen Test auf Nichtlinearität zu erzeugen, müssen lediglich die Autokovarianzen $\gamma(\tau)$ reproduziert werden, sodass als Kostenfunktion die Gleichung (2.6) gewählt werden kann. Wie bereits erwähnt, entspricht dies der Wahl $w_\tau = 1$ und $q = 1$. Die Autokorrelation der Zeitreihe fällt relativ rasch mit τ ab und hat bereits bei ungefähr $\tau_0 = 12$ seine erste Nullstelle. In diesem Fall hätte man die Gewichte also auch proportional zu $w_\tau \sim 1/\tau$ oder $w_\tau \sim e^{-\tau}$ wählen können, um den kurzreichweitigen Autokorrelation mehr Gewicht zu verleihen. Dies kann aber auch durch eine gezielte Wahl von $\tau_{\max}$ erreichen werden, indem man z. B. $\tau_{\max} = \tau_0$ setzt. Um bei einem Test die größtmögliche Sicherheit auf die Korrektheit der Testaussage zu bekommen, sollten aber *alle* Autokorrelationen berücksichtigt und $\tau_{\max} = N - 1$ gewählt werden. Allerdings sind Autokorrelationen mit einer Zeitverzögerung τ im Größenordnungsbereich von N nur schwer zu schätzen und zeigen große statistische Fluktuationen. Im folgenden Beispiel habe ich deshalb eine obere Grenze von $\tau_{\max} = 500$ gewählt, was ungefähr $N/3$ entspricht.

Die nächste Überlegung betrifft die gewünschte Genauigkeit, mit der die Eigenschaften zwischen Originaldaten und Surrogaten übereinstimmen sollen. Dementsprechend sollte die Schranke E_{stop} gewählt werden, bei der der Abkühlungsprozess abbricht. Ich wähle hier $E_{\text{stop}} = 25.0$, sodass die Autokorrelationen $\varrho(\tau) \in [-1, 1]$ im Mittel eine Abweichung von 0.0005 zeigen. Um die Abweichungen nicht nur im Mittel, sondern absolut zu beschränken, kann die Maximumsnorm ($q \to \infty$) verwendet werden. Für die Abkühlungsprozedur wurde der Temperaturfaktor auf $\alpha = 0.98$ und die Schranken für die Metropolisschritte auf $S_{\text{ak}} = N$ und $S_{\text{total}} = 10N$ gesetzt. Sie wurden so gewählt, dass ein Erreichen der vorgegebenen Grenze E_{stop} sichergestellt ist.

In Abbildung 2.4 ist zu sehen, wie sich der Wert der Kostenfunktion mit sinkender Temperatur (von rechts nach links) verändert. Zunächst zeigt sich bei hohen Temperaturen ($T > 10$) ein Plateau mit $E \approx 2600$. Da mit einer zufälligen Permutation gestartet wird, sind die Autokovarianzen des Surrogates $\gamma'_\tau \approx 0$ und die Kostenfunktion nimmt den

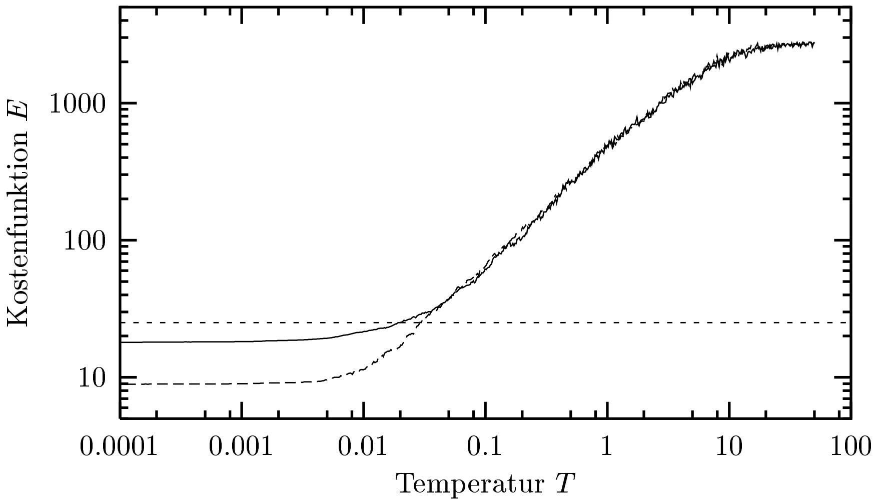

Abbildung 2.4: Die Kostenfunktion E sinkt während des Simulated Annealings mit der Temperatur. Während für die durchgezogene Linie beliebige Paare getauscht werden, werden bei der gestrichelten Linie Paare mit ähnlicher Größenordnung bevorzugt. Dadurch wird das Ziel ($E_{\mathrm{stop}} = 25.0$) fünfmal so schnell erreicht.

Wert der summierten Beträge der Autokovarianzen der Originalzeitreihe an. Das Plateau ist ein Zeichen dafür, dass mit einer zu hohen Temperatur gestartet wurde. Weitere Surrogate könnten mit einer Anfangstemperatur von $T_0 \approx 20$ erzeugt werden. Für weiter sinkende Temperaturen folgt ein starkes Absinken der Kostenfunktion, wobei in diesem Fall annähernd $E \sim T^{0.8}$ gilt. Bei tiefen Temperaturen läuft das System in ein Minimum und der Wert der Kostenfunktion wird konstant ($E \approx 17.8$). Die gewünschte Genauigkeit ($E = 25.0$) wird für $T \approx 0.02$ erreicht. Die benötigte Rechenzeit auf einem Standard–PC mit 450 MHz beträgt dabei ca. 500 Sekunden. Werden aber zusätzlich die Tauschpaare mit der in Abschnitt 2.2.3 beschriebenen Wahrscheinlichkeit p_{ij} gewählt, reduziert sich die Rechenzeit auf ca. 100 Sekunden. Als Wahrscheinlichkeitsverteilung wurde dabei eine Gaußverteilung gewählt, deren Standardabweichung proportional zu T verkleinert wurde. Ein weiterer Vorteil dieser Methode ist — wie ebenfalls in Abbildung 2.4 zu sehen — das Erreichen eines deutlich tieferen Minimums $E \approx 8.9$, wodurch die Genauigkeit ungefähr verdoppelt wurde. Weitere Versuche bei der Implementierung dieser Methode zeigten einen Geschwindigkeitsfaktor im Bereich 2–10.

Wie bereits erwähnt, resultiert der Geschwindigkeitsgewinn aus dem Versuch die Akzeptanzrate A während des Abkühlens auf einem konstanten Niveau von 50% zu halten. In Abbildung 2.5 ist aufgetragen, wie sich die Akzeptanzrate in Abhängigkeit von der Temperatur ändert. Werden die Tauschpaare für die Updates beliebig gewählt, sinkt A rasch ab und schon ab $T \approx 0.4$ wird nur noch jedes zehnte Update akzeptiert. Dies resultiert in unnötig häufigen Berechnungen der Kostenfunktion und ist deshalb sehr langsam. Werden aber die Tauschpaare $\{i, j\}$ mit der Wahrscheinlichkeit p_{ij} gewählt, deren Breite entsprechend angepasst wird, sinkt die Akzeptanzrate nicht unter 40% bis das System sich in einem Minimum niederlässt. Erst wenn die Breite der Verteilung $\mu = 1$ erreicht ist, d. h. nur noch Werte getauscht werden, die Nachbarn in der sortierten Zeitreihe sind, sinkt auch hier die Akzeptanzrate.

An dieser Stelle sei darauf hingewiesen, dass diese Methode bei diskreten Zeitreihenwerten, wie z. B. bei digital gemessenen Daten, zu Problemen führen kann. In solchen

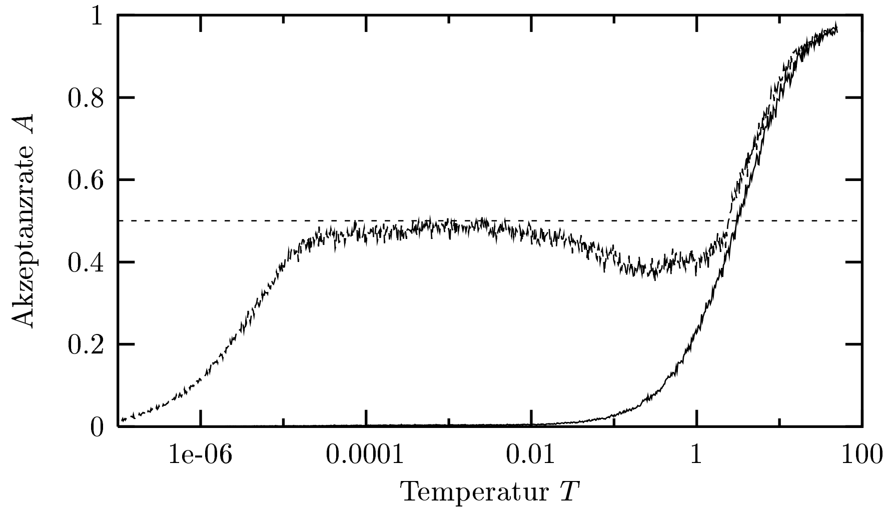

Abbildung 2.5: Die Akzeptanzrate A fällt stark mit der Temperatur T ab, falls jeweils beliebige Paare getauscht werden (untere Kurve). Durch Auswahl der Tauschpaare über eine angepasste Wahrscheinlichkeitsverteilung kann die Akzeptanzrate auf einem Niveau über 40% gehalten werden.

Zeitreihen tauchen meist viele gleiche Werte auf, deren Tausch die Kostenfunktion nicht ändert. Da in diesem Fall $\Delta E = 0$ gilt, wird ein solcher Tausch auf jeden Fall akzeptiert. Werden, wie hier, hauptsächlich Paare mit ähnlichem Rang getauscht, werden viele Updates akzeptiert, ohne dass die Kostenfunktion sinkt. Die Bedingung $S_{\text{ak}} = 0$ kann in einem solchen Fall nicht mehr als Abbruchkriterium dienen. Durch die Addition von schwachem Rauschen, dessen Amplitude unter dem Diskretisationsabstand der Zeitreihenwerte liegt, kann dieses Problem gelöst werden.

2.3 Eigenschaften

Bevor im nächsten Abschnitt die Flexibilität und die Möglichkeiten der neuen Methode an einigen Anwendungen demonstriert werden, sollen zunächst einige wesentliche Eigenschaften des Verfahrens untersucht werden. Dazu wird es mit den Standard–Fouriermethoden verglichen und die Konvergenz– und Genauigkeitseigenschaften analysiert.

2.3.1 Periodizität

In Abschnitt 2.1.1 wurde gezeigt, dass die Standardmethoden zur Surrogaterzeugung für Zeitreihen mit großer Differenz zwischen Anfangs– und Endwert falsche Testaussagen liefern können. Dies liegt daran, dass sie aufgrund der Periodizitätsannahme der Fouriertransformation die periodischen Autokovarianzen (2.1) auf den Surrogaten reproduzieren, und nicht wie erwünscht die Autokovarianzen $\gamma(\tau)$ aus Gleichung (1.1). Da diese aber mit der neuen Methode über die Kostenfunktion *direkt* auf den Surrogaten implementiert werden können, sollten die in Abbildung 2.1 zu sehenden Artefakte mit dem neuen Verfahren nicht mehr auftreten. Für die folgende Untersuchung wird die gleiche Zeitreihe wie in Abschnitt 2.1.1 verwendet, welche das Ergebnis eines stark korrelierten AR(2)–Prozesses war.

Zunächst muss wieder eine Kostenfunktion E aufgestellt werden, für die die Surrogate einen möglichst kleinen Wert liefern sollen. Diesmal wähle ich zur Quantifizierung der Diskrepanz zwischen den Autokorrelationen der Originaldaten und der Surrogate die Maximumsnorm ($q = \infty$) der Differenzen:

$$E(\{s'_n\}) = \max_{\tau=0..\tau_{\max}} |\varrho'(\tau) - \varrho(\tau)| \,. \tag{2.10}$$

Dadurch ist garantiert, dass die Autokorrelationen für alle τ gleich schnell konvergieren und keine Ausreißer existieren, deren starke Abweichung an anderer Stelle ausgeglichen werden. Die starken langreichweitigen Korrelationen machen die Surrogaterzeugung über die Minimierung der Kostenfunktion besonders schwierig. Werden alle $\tau_{\max} = N-1$ Autokorrelationen in die Kostenfunktion mit einbezogen, konvergiert die Minimierungsprozedur nur sehr langsam. So lieferte ein Programmlauf auf einem Rechner mit 450 MHz selbst nach vier Tagen Rechenzeit kein zufriedenstellendes Endergebnis und es blieben sichtbare hochfrequente Anteile. Eine wesentlich günstigere Methode die langreichweitigen Korrelationen zu fixieren, besteht in dem Ausschluss des Anfangs– und Endpunktes aus dem Permutationsschema. Durch diese Fixierung der Endpunkte wird die Autokovarianz mit der größtmöglichen Zeitverzögerung, $\gamma(\tau = N-1) = s_1 s_N$, exakt auf den Surrogaten reproduziert. Weiter wird dadurch die unerwünschte Übereinstimmung zwischen Anfangs– und Endwert in den Surrogaten aufgehoben, welche bei den Fourier–basierten Methoden stets gegeben war. Es sei betont, dass es sich hierbei nicht um einen „faulen Trick" handelt. Es wird schlichtweg die Flexibilität der neuen Methode ausgenutzt. Im Übrigen stellt die Fixierung der Endpunkte keine große Einschränkung des Spielraumes der möglichen Permutationen dar und eine große Vielfalt an Surrogaten ist nach wie vor möglich.

Da durch die Fixierung der Endpunkte indirekt für die Implementierung der langreichweitigen Korrelationen gesorgt ist, kann nun ein wesentlich kleineres $\tau_{\max}$ gewählt werden. Im Folgenden verwende ich $\tau_{\max} = 130$, was ungefähr der ersten Nullstelle der Autokorrelationsfunktion entspricht. Für die Minimierung der Kostenfunktion werden die Parameter $\alpha = 0.9995$, $S_{\text{ak}} = 10N$ und $S_{\text{total}} = 100N$ gewählt. Durch diese langsame Abkühlung ist es möglich sehr tiefe Minima zu erreichen. In Abbildung 2.6 sind einige Stationen eines Surrogates während der Minimierungsprozedur dargestellt. Die beim Start der Prozedur erzeugte zufällige Permutation ist völlig unkorreliert und weist bis auf Fluktuationen verschwindende Autokorrelationen auf. Die maximale Abweichung tritt dabei für kleine τ auf und die Kostenfunktion (2.10) liefert $E \approx 1$. Da durch die Maximumsnorm in der Kostenfunktion bei der Minimierung die größten Abweichungen zuerst abgebaut werden, finden die stärksten Änderungen in der Autokorrelationsfunktion zunächst bei kleinen Zeitverzögerungen τ statt. Dies ist deutlich in den Autokorrelationen der dritten Zeitreihe von oben in Abbildung 2.6 zu sehen. Das nächste Surrogat liefert mit $E \approx 0.11$ den gleichen Wert für die Kostenfunktion, den auch das in Abbildung 2.1 dargestellte Surrogat liefern würde. Da hier die größte Abweichung aber für kleine τ auftritt, und nicht wie in Abbildung 2.2 bei $\tau \approx 60$, zeigen die beiden Surrogate deutliche Unterschiede. Im weiteren Verlauf der Abkühlung passen sich die Autokorrelationen der Surrogate immer mehr den Autokorrelationen der Originalzeitreihe an, bis schließlich keine Abweichungen mehr zu erkennen sind. Nach circa neun Stunden Rechenzeit läuft das System in ein Minimum und das abschließende Surrogat wird ausgegeben, welches in Abbildung 2.6 unten gezeigt ist. Es hat per Konstruktion die gleichen Endpunkte wie die Originalzeitreihe und weist im Gegensatz zu Fourier–basierten Surrogaten auch die gleiche Glattheit auf. Der zeitliche Verlauf ist aber zufällig und völlig anders als in den Originaldaten.

Der Preis für die Vermeidung von Periodizitätseffekten ist hier der relativ hohe Rechenaufwand. Dies liegt in diesem Beispiel an den starken Korrelationen, die im Laufe der Minimierung im Surrogat aufgebaut werden müssen. So war das vorletzte Surrogat

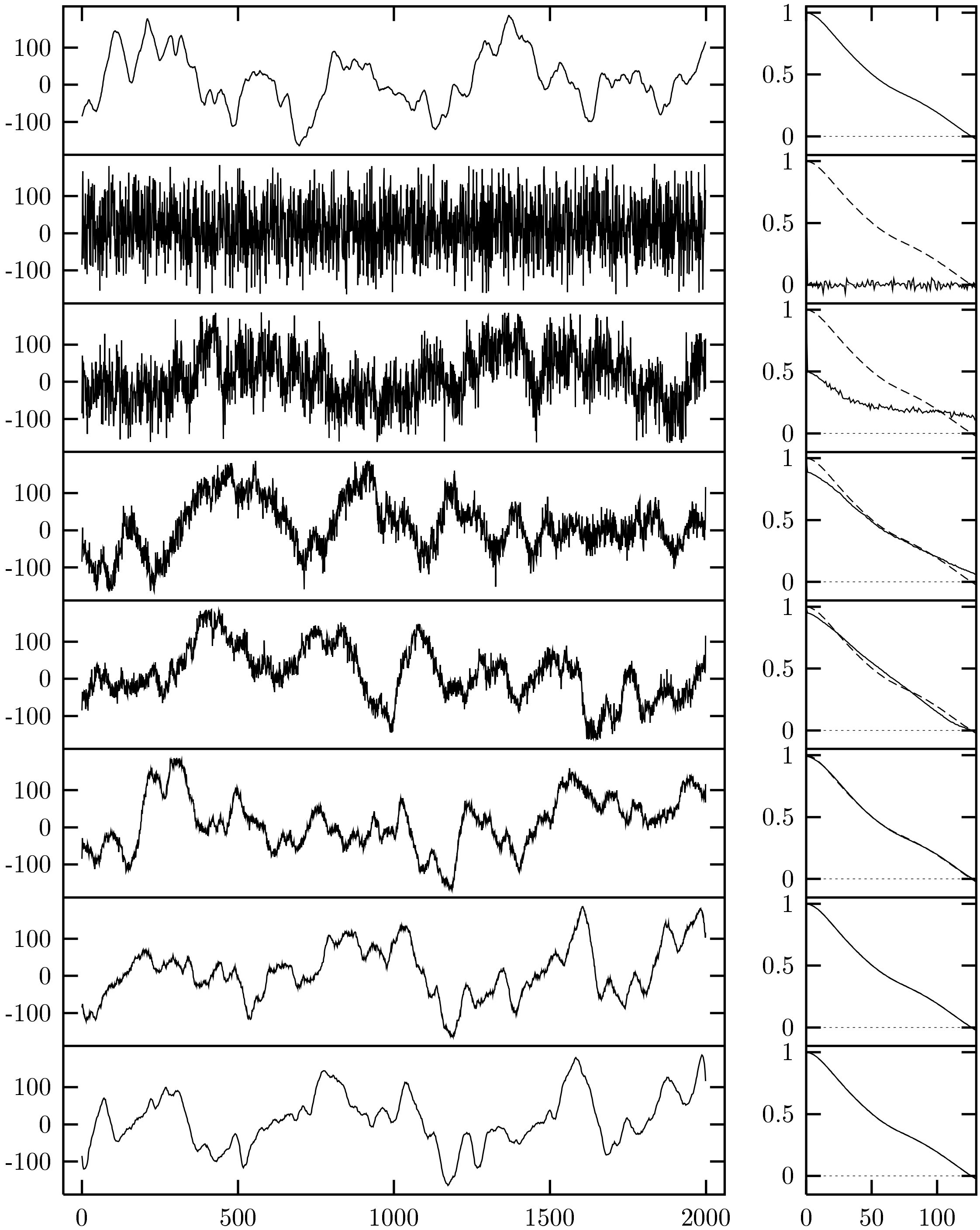

Abbildung 2.6: Interimssurrogate (links) und deren Autokorrelationen (rechts) während der Minimierung der Kostenfunktion (2.10). Die Werte der Kostenfunktion E von oben nach unten sind $E = 0.0$ (Originaldaten), $E = 1.0$ (zufällige Permutation), $E = 0.5$, $E = 0.11$ (vergleiche Surrogat aus Abbildung 2.1) und $E = 0.05, 0.01, 0.001, 0.0000036$.

in Abbildung 2.6 schon nach der Hälfte der Rechenzeit erreicht und erst nach weiterer, rechenintensiver Minimierung der Kostenfunktion wurde ein Surrogat mit der gewünschten Glattheit gefunden. Zur Reproduktion der Glattheit müssen die kurzreichweitigen Autokorrelationen also sehr genau übereinstimmen.

Für einen Nichtlinearitätstest mit einem Signifikanzniveau von 95% wären im Prinzip 19 Surrogate nötig, was im vorliegenden Fall den unangemessen hohen Rechenaufwand von ungefähr einer Woche bedeuten würde. Im Einzelfall muss, wie in Abschnitt 2.1.1 erläutert, abgewogen werden, ob die neue, rechenintensive Methode einem Abschneiden von Daten an den Enden vorzuziehen ist, um Periodizitätseffekte zu vermeiden. Hier kann der Test allerdings schon nach einem Surrogat mit einem Ergebnis abgebrochen werden. Schon das erste Surrogat liefert nämlich einen kleineren Vorhersagefehler[1] als die Originalzeitreihe, sodass auch mit weiteren Surrogaten die Nullhypothese eines gaußschen, linearen Prozesses nicht verworfen werden kann. Die Auflösung der Endpunktartefakte in den Surrogaten liefert also das erwartete Testergebnis.

2.3.2 Genauigkeit

Im Prinzip kann durch das Simulated Annealing ein beliebig tiefes Minimum der Kostenfunktion erreicht werden, solange die Temperatur genügend langsam gesenkt wird. Die einzige praktische Begrenzung liegt in der zur Verfügung stehenden Rechenzeit. Im Gegensatz zu üblichen Minimierungsproblemen sind wir hier aber nicht am exakten, globalen Minimum mit $E = 0$ interessiert. Die Surrogate, die dies erfüllen, sind mit großer Wahrscheinlichkeit die Originalzeitreihe selbst oder einfache Transformationen davon. Im Falle der Autokovarianz als Bedingung könnte dies z. B. die zeitumgekehrte Zeitreihe sein. Aber auch zyklische Permutationen der Originalzeitreihe haben, bis auf Endpunkteffekte, nahezu gleiche Autokovarianzen und erfüllen $E \approx 0$ recht gut. Allerdings ist es unwahrscheinlich, dass eine dieser wenigen Transformationen aus der großen Menge der $N!$ Permutationen jemals erreicht wird, solange N groß ist. Ist die Zeitreihe aber sehr kurz und die Anzahl der Bedingungen I so groß, dass das Problem überspezifiziert ist, können diese unerwünschten Surrogate auftreten. Eine sinnvolle Analyse von Zeitreihen mit derart wenigen Datenpunkten ist allerdings ohnehin fraglich. Es sei darauf hingewiesen, dass ähnliche Probleme selbstverständlich auch bei der Verwendung der Standardmethoden auftreten können. Die Flexibilität der neuen Methode erlaubt es aber, bestimmte unerwünschte Surrogate zu vermeiden. So könnten zusätzliche Terme in die Kostenfunktion eingebaut werden, die Ähnlichkeiten mit bekannten Transformationen unterdrücken. Im Übrigen zerstören Ähnlichkeiten zwischen der Originalzeitreihe und den Surrogaten nicht die Gültigkeit des Tests, da die Size dadurch schlimmstenfalls erhöht wird. Allerdings senkt eine solche Ähnlichkeit die Trennschärfe und damit die Leistungsfähigkeit des Tests.

Die Tatsache, dass die Surrogate möglichst kleine Werte der Kostenfunktion liefern sollen, ohne $E = 0$ exakt zu erfüllen, wirft die Frage auf, ob genügend viele lokale Minima mit $E \approx 0$ existieren. Eine allgemeine Antwort auf diese Frage lässt sich nur schwer geben, da sie in starkem Maße von der verwendeten Kostenfunktion abhängt. Allerdings lassen sich plausible Argumente finden, dass für große N auch für sehr kleine Werte der Kostenfunktion noch sehr viele Konfigurationen existieren. Sei $p(E)$ die Verteilung der Werte, die die Kostenfunktion E unter den $N!$ möglichen Permutationen annehmen kann. Diese hat für gewöhnlich ein Maximum bei einem Wert der Kostenfunktion, der typischerweise von zufälligen Permutationen angenommen wird. In den vorangegangen Beispielen hatte dieses

[1] Dabei wurden die gleichen Parameter wie in Abschnitt 2.1.1 verwendet.

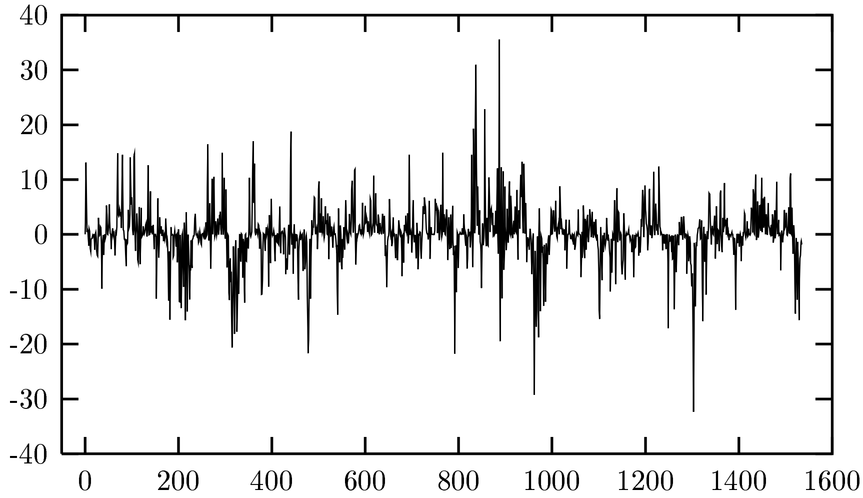

Abbildung 2.7: Stark korrelierter AR(3)–Prozess (2.11), der mit $s_n = x_n|x_n|$ gemessen wurde.

Maximum einen Abstand von ungefähr 40 (SOI) bzw. 76 (AR2) Standardabweichungen von $E = 0$. Dabei wurde im ersten Fall $q = 1$ und im zweiten Fall die Maximumsnorm ($q = \infty$) verwendet. Weiter sei $P(\bar{E}) = \int_0^{\bar{E}} p(E)\mathrm{d}E$ die kumulative Wahrscheinlichkeitsverteilung, die angibt wie wahrscheinlich Konfigurationen mit einer Kostenfunktion $E \leq \bar{E}$ sind. Natürlich nimmt $P(\bar{E})$ für $\bar{E}$ nahe bei Null sehr kleine Werte an, ist aber für positive $\bar{E}$ stets von Null verschieden. Dies gilt für einen Großteil an Kostenfunktionen und ist im Beispiel der Autokovarianzen offensichtlich gegeben. Die Anzahl der möglichen Konfigurationen mit $E < \bar{E}$ ist proportional zu $P(\bar{E})N!$. Die Fakultät $N! \sim N^N$ wächst dabei schneller als exponentiell an. Und obwohl in den erwähnten Beispielen $P(\bar{E})$ extrem klein ist und Größenordnungen von 10^{-350} bzw. 10^{-1250} annimmt, ergeben sich aufgrund der immensen Anzahl an Konfigurationen noch 10^{3712} bzw. 10^{4480} Konfigurationen mit $E < \bar{E}$. Dabei wurde für $p(E)$ die Gaußverteilung angenommen, welche mit $\exp(-E^2)$ abfällt. Aber auch schneller abfallende Verteilungen können die Anzahl der möglichen Konfigurationen nur dann unterdrücken, wenn N klein ist.

Das Konzept der bedingten Realisierung verlangt im Grunde genommen ein exaktes Übereinstimmen der Eigenschaften zwischen Surrogaten und Originalzeitreihe. Für endliche Zeitreihen wird diese Bedingung aber nur von einigen trivialen Transformationen exakt erfüllt, an denen man als Surrogat nicht interessiert ist. In der Praxis kann man nur versuchen, die Eigenschaften so genau wie möglich auf den Surrogaten zu reproduzieren. Diesem Ziel kommt man von der AAFT über das iterative Verfahren bis nur neuen Simulated Annealing Methode immer näher. Eine Senkung der Diskrepanz zwischen Surrogaten und Originaldaten verkleinert eine mögliche Abweichung der empirischen von der nominellen Size. Da viele der verwendeten Teststatistiken für Nichtlinearität, wie z. B. der nichtlineare Vorhersagefehler, auch von den linearen Eigenschaften der Zeitreihe abhängen, senkt sich natürlich auch deren Varianz. Dadurch erhöht sich die Trennschärfe des Tests, wodurch das Auffinden von kleinsten Anzeichen von Nichtlinearität möglich wird. Andererseits kann diese Senkung der Varianz der Teststatistik entgegen den Erwartungen auch zu falschen Verwerfungen der Nullhypothese führen. Solche Fälle wurden von Kugiumtzis [Kug99] aufgezeigt, wobei die Surrogate über das iterative Verfahren erzeugt wurden. Dieses Phänomen soll im Folgenden an einem Beispiel demonstriert werden.

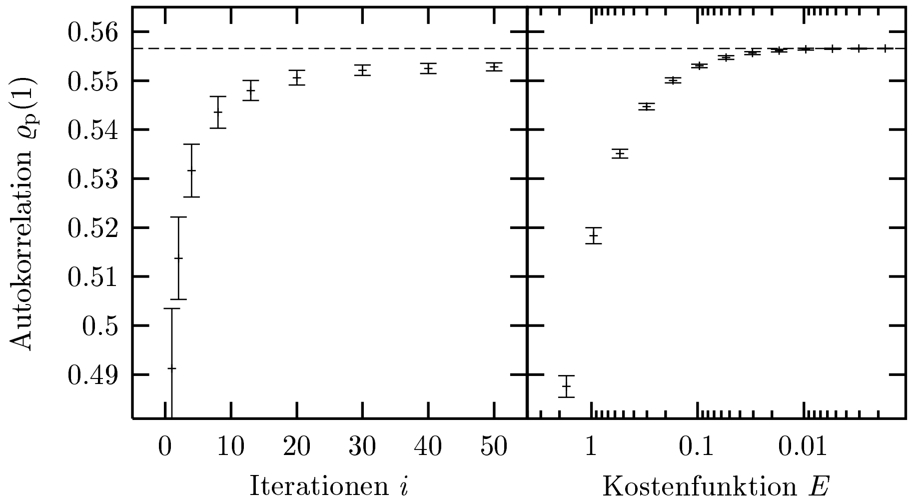

Abbildung 2.8: Mittelwert und Standardabweichung der periodischen Autokorrelation $\varrho_p(1)$ für 200 Surrogate in Abhängigkeit von den Iterationschritten (links) und der Kostenfunktion (rechts). Die horizontale Linie entspricht dem Wert des Original–AR(3)–Prozesses. Während das iterative Verfahren den Sollwert nicht erreicht, kommt man mit der neuen Methode diesem Wert beliebig nahe.

Als Zeitreihe dient diesmal ein stark korrelierter AR(3)–Prozess

$$x_n = 0.99x_{n-1} - 0.8x_{n-2} + 0.65x_{n-2} + \eta_n \,, \tag{2.11}$$

der zusätzlich mit der nichtlinearen Messfunktion

$$s_n = x_n|x_n| \tag{2.12}$$

gemessen wird. Eine Realisierung dieses Prozesses, die im Folgenden als Grundlage dient, ist in Abbildung 2.7 dargestellt. Die Kombination aus der starker Korreliertheit und der relativ starken Nichtlinearität in der Messfunktion macht es für diesen Datensatz sehr schwierig Surrogate zu erzeugen. Die iterative Methode konvergiert nur sehr langsam und benötigt ungefähr 130 Iterationen, um den Fixpunkt zu erreichen. In Abbildung 2.8 ist im Vergleich zwischen iterativem und Simulated Annealing Verfahren dargestellt, wie sich eine *lineare* Teststatistik mit steigender Genauigkeit verhält. Als Observable dient dabei die periodische Autokorrelation $\varrho_p(\tau = 1)$, welche auf jeweils 200 Surrogaten gemessen wurde. Für die iterative Methode nähert sich der Mittelwert der Autokorrelation mit steigender Anzahl der Iterationschritte einem konstanten Wert. Aber obwohl Fourier–basierte Surrogate die periodischen Autokorrelationen zu implementieren versuchen, erreichen sie nicht den Wert der Originalzeitreihe. Der Abstand beträgt am Ende ungefähr drei Standardabweichungen. Offensichtlich gelingt es mit dem iterativen Verfahren nicht, sowohl Verteilung als auch Spektrum gleichzeitig zu realisieren. In einem solchen Fall könnten nichtlineare Teststatistiken, die auch von den linearen Korrelationen abhängen, zu falschen Verwerfungen der Nullhypothese führen. Für die rechte Seite von Abbildung 2.8 wurden Surrogate erzeugt, die die Kostenfunktion $E = \max_{\tau=1}^{100} |\gamma'_p(\tau) - \gamma_p(\tau)|$ minimieren. Da in diesem Fall die Korrelationen direkt implementiert werden, ist es nicht verwunderlich, dass der Mittelwert sich im Laufe der Minimierung dem Sollwert der Originalzeitreihe nähert. Für $E < 0.01$ ist der Abstand kleiner als eine Standardabweichung und ab $E \approx 0.001$ verteilen sich die Werte der Surrogate symmetrisch um den Sollwert. Die Standardabweichung

σ wird dabei mit der Kostenfunktion immer kleiner und ist annähernd proportional zu E. Im Gegensatz zur iterativen Methode existieren also in diesem Fall keine verbleibenden Diskrepanzen zwischen Surrogaten und Originalzeitreihe.

Um beschriebene Probleme für iterative Surrogate zu vermeiden, schlägt Kugiumtzis [Kug99] vor, die Verwendbarkeit der Surrogate durch einen Test mit einer linearen Teststatistik zu verifizieren. Im oben genannten Fall wäre dadurch die Diskrepanz zwischen Surrogaten und Originalzeitreihe sofort aufgefallen. Für Simulated Annealing Surrogate habe ich eine solche Diskrepanz bisher in keinem Fall beobachtet. Es ist aber nicht auszuschließen, dass z. B. bei ungeschickter Wahl der Kostenfunktion unerwünschte *systematische* Abweichungen auftreten können. Die Flexibilität der neuen Methode erlaubt im Prinzip auch die Vermeidung dieses Problems. So ist es z. B. möglich, eine Kostenfunktion

$$E(S'_1, \ldots, S'_M) = \left[\sum_{j=1}^{J} \left| w_j \left(\sum_{m=1}^{M} B_j(S'_m) \right) \right|^q \right]^{1/q} \tag{2.13}$$

aufzustellen, die alle M Surrogate berücksichtigt und explizit deren summierte Abweichung minimiert. Die Surrogate $S'_1, \ldots, S'_M$ werden dabei in *einem* Minimierungsprozess *gleichzeitig* erzeugt. Durch die innere Summe über die Surrogate gehen systematische Abweichungen dabei stärker in die Kostenfunktion ein als statistische und werden somit bei der Minimierung stärker unterdrückt. Dieses Vorgehen bedeutet allerdings einen äußerst großen Rechenaufwand und der Sinn einer Umsetzung dieser Idee ist daher mehr als fraglich.

2.4 Anwendungen

Bisher war stets die Autokovarianz die einzige Eigenschaft, die auf den Surrogaten reproduziert werden sollte. Das Motiv dafür war, dass in einem Test auf Nichtlinearität die linearen Eigenschaften der Originalzeitreihe in den Surrogaten erhalten werden sollten. In vielen Fällen reicht dafür aber die Verwendung der Fourier–basierten Standardmethoden aus. Nur wenn Periodizitätsprobleme zu erwarten sind oder das iterative Verfahren nur unzulänglich konvergiert, sollte in diesem einfachen Fall auf Simulated Annealing Surrogate zurückgegriffen werden.

Im Folgenden soll gezeigt werden, wie durch die Flexibilität der neuen Methode die einfache Nullhypothese eines linearen Prozesses mit einer eventuellen nichtlinearen Messfunktion sinnvoll erweitert werden kann. Bekannte, bereits verstandene oder uninteressante Eigenschaften können über die Kostenfunktion in die Surrogate eingebaut werden. Darüber hinaus lassen sich Surrogate für Datensätze erzeugen, deren Behandlung sich den Standardmethoden entzieht.

2.4.1 Nichtstationarität

In Abschnitt 2.1.2 wurde auf die Problematik von Nichtstationarität in der zu untersuchenden Zeitreihe eingegangen. Die Tatsache, dass die Stationarität impliziter Bestandteil der Nullhypothese ist, führt dazu, dass jegliche Form von Nichtstationarität in den Daten eine (korrekte) Verwerfung der Nullhypothese nach sich ziehen kann. In einigen Situationen kann eine unerwünschte, aber unvermeidbare Nichtstationarität in den Daten bereits *vor* der nichtlinearen Analyse bekannt sein. Dies ist z. B. der Fall, wenn die Messung durch eine externe Störung oder durch den Drift eines Parameters des Systems beeinflusst

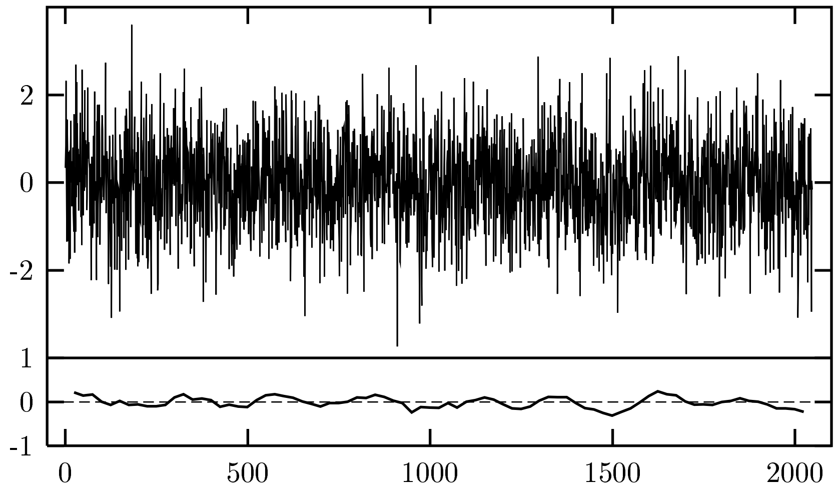

Abbildung 2.9: Unabhängige, gaußverteilte Zufallszahlen mit einer addierten periodischen Schwankung (oben) und deren laufender Mittelwert (unten).

wird. Eine solche nichtstationäre Zeitreihe sollte nicht direkt in einem Test verwendet werden, da bei einem eventuellen Verwerfen der Nullhypothese die Ursache dafür nicht bestimmt werden kann. Eine Möglichkeit, die Auswirkungen der bekannten Nichtstationaritäten zu unterbinden, ist die „Subtraktion" bzw. ein *De–Trending* der vermuteten Nichtstationarität. Dafür muss die Form bzw. der Verlauf der Nichtstationarität entweder durch Anpassung an die Daten modelliert oder durch zusätzliche, parallele und unabhängige Messungen gewonnen werden. Dies ist aber in vielen Fällen sehr schwierig und nicht praktikabel. Die Alternative ist die Implementierung der Nichtstationarität in die Surrogate.

Um mit einem einfachen Beispiel zu beginnen, betrachten wir die Zeitreihe aus Abbildung 2.9, welche aus unabhängigen, gaußverteilten Zufallszahlen besteht. Zusätzlich wurde eine Sinuswelle $A\sin(2\pi t/T)$ mit einer Amplitude $A = 0.2$ und einer Periode $T = 256$ addiert, welche in diesem Fall den nichtstationären Untergrund darstellen soll. Soll der Datensatz auf mögliche Korrelationen hin untersucht werden, lässt sich ein Test mit Surrogaten durchführen, die lediglich zufällige Permutationen der Originalzeitreihe sind. Als Teststatistik kann die Autokorrelation $\varrho(1)$ dienen. Führt man einen solchen Test mit der oben genannten Zeitreihe und 39 Surrogaten durch, so wird die Nullhypothese der Unkorreliertheit der Werte mit einem Signifikanzniveau von 95% verworfen. Dies ist dadurch zu erklären, dass der langsam variierende Untergrund natürlich selbst Korrelationen aufweist. Ist aber die Existenz und Ursache des nichtstationären Untergrundes bekannt und für die Analyse uninteressant, müssen für einen Test entsprechende Surrogate erzeugt werden. Im Falle eines additiven Untergrunds, der lediglich den zeitlich lokalen Mittelwert verändert, könnte eine der zu fixierenden Eigenschaften der Surrogate der laufende Mittelwert

$$m_b(t) = \frac{1}{b}\sum_{n=t}^{t+b-1} s_n \tag{2.14}$$

sein. Dabei ist t der Beginn und b die Breite des Fensters, über das die Werte der Zeitreihe

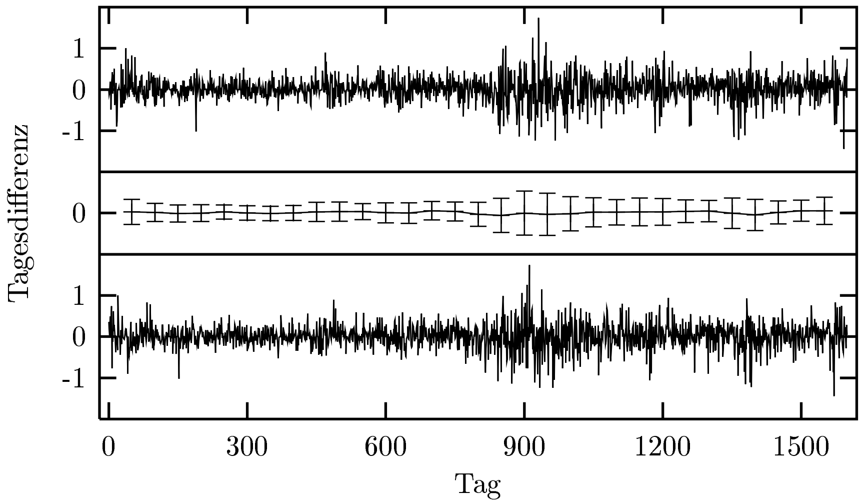

Abbildung 2.10: Tagesdifferenzen der BUND Future Notierung (oben) und deren laufende Mittelwerte und Standardabweichungen. Das Surrogat (unten) hat die gleichen Autokorrelationen und die gleichen laufenden Mittelwerte und Standardabweichungen.

$\{s_n\}$ gemittelt werden. Daraus könnte eine Kostenfunktion der Art

$$E = \sum_{j=0}^{J-1} \left[m'_b(jN/J) - m_b(jN/J) \right]^2 \ , \tag{2.15}$$

aufgestellt werden, welche die Abweichungen des laufenden Mittelwertes zwischen Originalzeitreihe und Surrogat in J um N/J versetzten Fenstern misst. In Abbildung 2.9 sind unten die laufenden Mittelwerte über Fenster der Breite $b = 100$, welche jeweils um 25 Punkte versetzt sind, aufgetragen. Die periodische Struktur des addierten Untergrundes ist schwach zu erkennen. Weiter wurden 39 Surrogate erzeugt, die die Kostenfunktion aus Gleichung (2.15) minimieren, und wieder deren Autokorrelation mit den zu untersuchenden Daten verglichen. Die erweiterte Nullhypothese einer unkorrelierten Zeitreihe mit langsam variierendem Mittelwert konnte diesmal korrekterweise nicht verworfen werden.

Das vorangegangene Beispiel war sowohl in Hinblick auf die Nullhypothese als auch auf die auftretende Nichtstationarität besonders einfach. Im Folgenden soll eine Finanzzeitreihe auf Nichtlinearität hin untersucht werden. Dabei handelt es sich um den täglich notierten BUND Future, welcher freundlicherweise von Thomas Schürmann von der WGZ–Bank in Düsseldorf zur Verfügung gestellt wurde. In Abbildung 2.10 (oben) sind 1600 Tagesdifferenzen von Dezember 1990 bis Februar 1997 darstellt. Es ist deutlich zu sehen, dass die Varianz und in kleinerem Maße auch der Mittelwert auf einer größeren Zeitskala variieren als die Änderungen der Zeitreihe selbst. Die Nichtstationarität ist wesentlich stärker ausgeprägt als im vorangegangen, konstruierten Beispiel und die laufende Standardabweichung variiert bis zu einem Faktor drei. Diese Nichtstationaritäten könnten leicht zu einem Verwerfen der Nullhypothese eines linearen Prozesses führen, was formell zwar richtig, in vielen Fällen aber unerwünscht ist. Im Folgenden sollen Surrogate für einen Nichtlinearitätstest erzeugt werden, die die gleichen laufenden Mittelwerte und Standardabweichungen besitzen wie der Originaldatensatz. Dafür werden diese Größen in Fenstern von 100 Tagen bestimmt, die jeweils um 50 Tage überlappen. Die entsprechenden Werte sind in der Mitte von Abbildung 2.10 aufgetragen. Weiterhin müssen wir für einen Test auf Nichtlinearität auch die Autokovarianzen $\gamma(\tau)$ der Surrogate in Übereinstimmung mit der Originalzeitreihe bringen. Da die Werte nur schwach korreliert sind, berücksichtige ich in der Kostenfunktion nur Autokovarianzen bis zu einer Zeitverzögerung von $\tau_{\max} = 5$. In die Kostenfunktion fließen damit drei unterschiedliche Eigenschaften (Mittelwerte, Standardabweichungen und Autokovarianzen) ein, welche sich in ihrer Größenordnung unterscheiden. In einem solchen Fall können die Größenverhältnisse durch geeignete Wahl der Gewichte w_j in der Kostenfunktion (2.5) aufeinander angepasst werden. Hier skaliere ich die drei Eigenschaften so, dass sie im Intervall $[-1, 1]$ liegen. Als Norm wähle ich die Euklidische Norm mit $q = 2$. Durch Minimierung der resultierenden Kostenfunktion erhält man Surrogate, von denen eines unten in Abbildung 2.10 dargestellt ist. Wird das Maß für die Zeitasymmetrie (1.3) für die Originalzeitreihe und 19 Surrogate berechnet und verglichen, kann die Nullhypothese eines linearen Prozesses mit zeitabhängigen Mittelwerten und Varianzen nicht verworfen werden.

Neben den in diesem Abschnitt implementierten Erweiterungen der Nullhypothese sind beliebige, weitere Formen von Nichtstationaritäten vorstellbar. Sie müssen lediglich in Form einer Observablen formulierbar sein und in die Kostenfunktion eingebaut werden. Eine mögliche Erweiterung für Nichtlinearitätstests mit nichtstationären Zeitreihen wäre die Implementierung von laufenden Autokovarianzen. Aber auch die lokale Wahrscheinlichkeitsverteilung der Zeitreihenwerte könnte zeitlich angepasst werden, indem das Permutationsschema geändert wird und sich die Werte s_n nur um maximal $\Delta n_{\max}$ verschieben lassen. Durch Kombination dieser beiden Ansätze könnten Surrogate für einen Nichtlinearitätstest erzeugt werden, dessen Nullhypothese ein ARMA–Prozess mit langsam variierenden Parametern und veränderbarer Messfunktion ist.

2.4.2 Multivariate Daten

Häufig werden von einem zu untersuchenden System mehrere Zeitreihen gleichzeitig aufgenommen. Dabei sind z. B. an verschiedenen Orten gemessene, gleiche Größen oder auch physikalisch unterschiedliche Größen vorstellbar. Obwohl multivariate Daten sehr häufig auftreten, lassen sich in der Literatur relativ wenig nichtlineare Methoden finden, die ihre Analyse erlauben. Für einen Test auf Nichtlinearität ist es im Prinzip legitim die Zeitreihen einzeln zu untersuchen. Durch ein Auffinden von Nichtlinearität in einem Anteil der Zeitreihen, der signifikant höher ist als die Size des Tests, lässt sich die Anwendung von nichtlinearen Methoden rechtfertigen. Dabei werden die Zeitreihen allerdings als unabhängig von einander angenommen. Eine natürliche Erweiterung der Nullhypothese des ARMA–Prozesses ist ein *multivariater*, gaußscher, linearer Prozess. Neben den Autokovarianzen und Verteilungsfunktionen der individuellen Zeitreihen müssen in diesem Fall auch die gegenseitigen *Kreuzkovarianzen*

$$\Gamma_{XY}(\tau) = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} (x_n - \bar{x})(y_{n+\tau} - \bar{y}) \quad (\text{für} \quad \tau \geq 0)^2 \tag{2.16}$$

zwischen den jeweiligen Zeitreihen $\{x_n\}$ und $\{y_n\}$ fixiert werden. Das *Korrelationstheorem*, eine Verallgemeinerung des Wiener–Khinchin–Theorems (1.11), erlaubt auch hier eine Realisierung im Frequenzraum. Eine entsprechende Verallgemeinerung der iterativen Methode aus Abschnitt 1.5.3 wird in [SS00] beschrieben, indem die Einzelspektren und

[2] Für $\tau < 0$ müssen die Grenzen der Summe und die Normierung entsprechend angepaßt werden.

das Kreuzspektrum angepasst werden. Dabei werden sowohl eventuell vorhandene Nichtlinearitäten *in* den Zeitreihen als auch nichtlineare Abhängigkeiten *zwischen* den Zeitreihen in den Surrogaten zerstört. In einigen Fällen kann dies aber unerwünscht sein.

In [Sch98] und [SS00] wird das Beispiel einer gleichzeitigen Aufnahme von Atmung und Herzschlagrate einer Versuchsperson im Schlaf betrachtet. Diese Zeitreihen sind Bestandteil des Datensatzes B, der vom Santa Fe Institute anlässlich eines Zeitreihenwettbewerbs veröffentlicht wurde [WG93]. Werden Atem– und Herzrate einzeln betrachtet, lassen sich in beiden leicht Nichtlinearitäten finden. Dies macht sich unter anderem durch eine starke Asymmetrie unter Zeitumkehr bemerkbar. Da erfahrungsgemäß eine Abhängigkeit zwischen Atem– und Herzrate besteht, stellt sich die interessante Frage, ob die Nichtlinearität einer der beiden Komponenten aus der Abhängigkeit von der anderen erklären lässt. Bei einem entsprechenden Test wird eine Zeitreihe fixiert, während für die andere ein Surrogat erzeugt wird. Sollen lediglich lineare Abhängigkeiten zwischen den Zeitreihen untersucht werden, reicht auch hier wieder die Fixierung der Kreuzkorrelationen aus. Werden aber sowohl *alle* Kreuzkorrelationen zu einer *unveränderlichen* Zeitreihe als auch alle Autokorrelationen fixiert, wie es bei Methoden im Frequenzraum der Fall ist, bleiben keine Freiheiten für ein zufälliges Surrogat. Im Gegensatz dazu erlaubt die Simulated Annealing Methode die Implementierung von Kovarianzen auf bestimmten, interessierenden Teil–Intervallen des τ–Bereichs. So gewonnene Surrogate wurden in [Sch98] für die Herzrate erzeugt. Es zeigte sich, dass speziell die Zeitumkehrasymmetrie der Herzrate nicht allein mit einer linearen Abhängigkeit von der Atemrate erklärt werden kann.

Ich möchte hier die grundlegende Idee dieser Vorgehensweise an einem konstruierten Modell illustrieren. Dazu betrachte ich als erste Zeitreihe die x–Komponente der chaotischen Hénon–Abbildung

$$\begin{array}{rcl} x_n & = & 1.4 - x_{n-1}^2 + 0.3 y_{n-1} \\ y_n & = & x_{n-1} \end{array} \quad . \tag{2.17}$$

Diese sei in der folgenden Art mit einem AR(1)–Prozess gekoppelt

$$z_n = 0.9 z_{n-1} + C x_{n-1} + \eta_n \, , \quad n = 1, \ldots, N = 1000 \, , \tag{2.18}$$

wobei über C die Stärke der Koppelung variiert werden kann. Für $C = 0$ ist die Zeitreihe $\{z_n\}$ ein linearer AR(1)–Prozess, aber mit steigendem C wird der Prozess über die Koppelung mit der Hénon–Abbildung nichtlinear. Wird für z. B. $C = 1$ ein Nichtlinearitätstest mit 39 Fourier–basierten Surrogaten durchgeführt, wird die Nullhypothese eines gaußschen, linearen Prozesses mit einer Wahrscheinlichkeit von 95% verworfen. Werden Mittelwert und Standardabweichung der Zeitumkehrstatistik (1.3) mit $\tau = 1$ für 500 Surrogate berechnet, so zeigt sich, dass der Wert für die Originalzeitreihe ungefähr drei Standardabweichungen vom Mittelwert ($T^{\text{rev}}(1) \approx 0$) entfernt ist. Um zu untersuchen, ob die gefundene Nichtlinearität ihre Ursache lediglich in der linearen Koppelung mit der nichtlinearen Zeitreihe $\{x_n\}$ hat, müssen Surrogate erzeugt werden, die die gleichen Kreuzkorrelationen aufweisen. Wie in Abbildung 2.11 zu sehen, fallen die Kreuzkorrelationen mit steigendem $|\tau|$ rasch ab und ich beziehe im Folgenden nur die Kreuzkorrelationen im Bereich $\tau = -100, \ldots, 100$ in die Kostenfunktion ein. Weiter werden in der Kostenfunktion die Autokovarianzen $\gamma(\tau)$ bis $\tau_{\max} = 200$ berücksichtigt. Da beide Zeitreihen auf gleiche Varianz normiert wurden, sind Kreuz– und Autokovarianzen von gleicher Größenordnung und die Gewichte in der Kostenfunktion (2.5) können konstant $w_j = 1$ gewählt werden. Die resultierende Kostenfunktion mit Maximumsnorm ($q = \infty$) wird minimiert, indem die

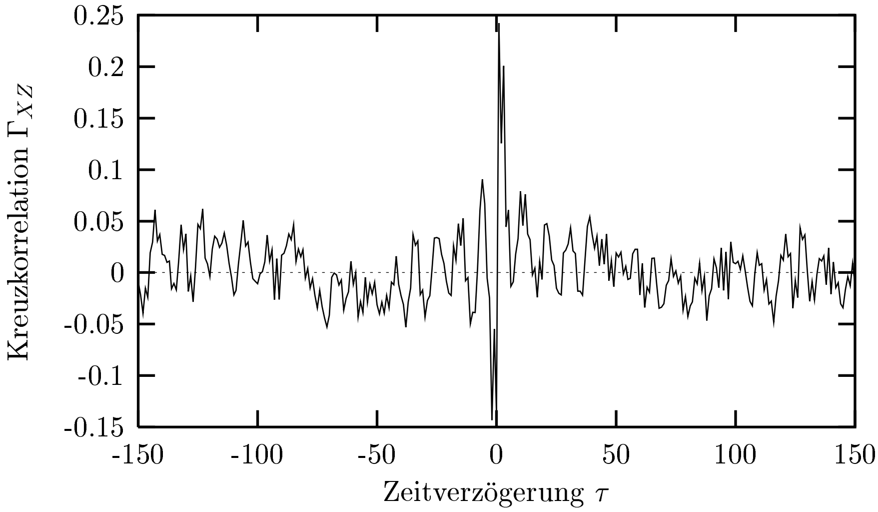

Abbildung 2.11: Kreuzkorrelationen zwischen der x–Komponente der Hénon–Abbildung und einem damit gekoppelten AR(1)–Prozess (2.18).

Temperatur nach $S_{\mathrm{ak}} = 5N$ bzw. $S_{\mathrm{total}} = 50N$ Updates um den Faktor $\alpha = 0.995$ gesenkt wird. Bei einem erneuten Test mit der Zeitumkehrstatistik und 39 Annealing–Surrogaten liegt der Wert der Originalzeitreihe nun nicht außerhalb des durch die Surrogate definierten Bereichs, sondern ungefähr in der Mitte. Die Nullhypothese eines linearen Prozesses, der linear an ein nichtlineares System gekoppelt ist, kann also korrekterweise nicht verworfen werden. Ein Verwerfen dieser Nullhypothese könnte verschiedene Ursachen haben. Zum einen könnte die untersuchte Zeitreihe tatsächlich das Produkt eines nichtlinearen zugrunde liegenden Prozesses sein. Aber auch die Koppelung könnte nichtlinearer Natur sein und damit nicht ausreichend durch die Kreuzkorrelationen beschrieben werden.

Die Analyse von multivariaten Daten ist damit bei weitem nicht erschöpft und der zweite Teil dieser Arbeit beschäftigt sich eingehend mit der Analyse von Abhängigkeiten zwischen Zeitreihen.

2.4.3 Zeitreihen mit unterschiedlichen Zeitintervallen

Die große Mehrheit der Arbeiten in der Zeitreihenanalyse beschäftigen sich mit Zeitreihen, deren Werte in festen Zeitabständen aufgenommen wurden. Unregelmäßig aufgenommene Zeitreihen werden meist nicht berücksichtigt, obwohl sie in vielen Gebieten der Wissenschaft auftreten, wenn Messungen durch praktische Bedingungen beschränkt sind. Beispiele dafür sind astronomische Beobachtungen, die oft nur nachts und bei klarem Wetter stattfinden können, oder Finanzmarktdaten, die an Wochenenden und Feiertagen Lücken aufweisen. Der Ausschluss solcher Zeitreihen ist dabei hauptsächlich durch technische Aspekte begründet und damit zu erklären, dass viele Methoden nicht auf den Fall der unregelmäßig aufgenommenen Zeitreihen verallgemeinert werden können. Dies gilt sowohl für einfachste, lineare Observablen wie die Autokovarianz als auch für die Methoden der nichtlinearen Zeitreihenanalyse.

Eine erste Idee Zeitreihen mit variierenden Zeitintervallen zu behandeln besteht darin, die Werte der Zeitreihe auf Zeitpunkte mit gleichem Abstand zu interpolieren. Dies ist aber für einen Test auf Nichtlinearität nicht zu empfehlen, da die Interpolation unerwünschte Nichtlinearitäten nach sich ziehen könnte, die nachher nicht von wirklicher nichtlinearer Struktur in den Daten unterschieden werden kann. In dem speziellen Fall, in dem die Daten zwar in regelmäßigen Zeitabständen aufgenommen wurden, aber an einigen Stellen Lücken aufweisen, ist es möglich diese Lücken mit Nullen aufzufüllen. Das Permutationsschema für die Surrogate sollte diese Lücken unberücksichtigt lassen, sodass sie an der gleichen Stelle bleiben. Der resultierende Fehler im Schätzer der Autokovarianz ist dann für Originalzeitreihe und Surrogate gleich und der Test behält seine Gültigkeit. Ist die Anzahl der Lücken aber sehr hoch oder sind die Lücken sehr lang, ist dieser Ansatz aber nicht mehr praktikabel.

Im Falle von Zeitreihen, die keine Regelmäßigkeit in den Zeitabständen aufweisen, benötigen wir andere Größen bzw. Schätzer, die die linearen Eigenschaften der Daten beschreiben. Sobald wir ein solches Maß für die linearen Korrelationen einer unregelmäßig aufgenommenen Zeitreihe gefunden haben, können wir sie in die Form einer Kostenfunktion bringen und entsprechend Surrogate erzeugen. Auch hier sind — wie im Falle von herkömmlichen Zeitreihen — zwei Ansätze möglich, wobei der eine im Frequenzbereich und der andere im Zeitbereich angesiedelt ist.

Zunächst ist es möglich, auch das Frequenzspektrum einer Zeitreihe $\{s_n\}, n = 1, \ldots, N$, mit beliebigen, aber chronologisch sortierten Zeitpunkten $\{t_n\}$ zu schätzen. Ein solcher Schätzer ist das *Lomb*–Periodogramm [Lom76, Sca82]

$$P(\omega) = \frac{1}{2\sigma^2} \left\{ \frac{\left[\sum_{n=1}^{N} (s_n - \bar{s}) \sin \omega (t_n - \tilde{t}\,)\right]^2}{\sum_{n=1}^{N} \sin^2 \omega (t_n - \tilde{t}\,)} + \frac{\left[\sum_{n=1}^{N} (s_n - \bar{s}) \cos \omega (t_n - \tilde{t}\,)\right]^2}{\sum_{n=1}^{N} \cos^2 \omega (t_n - \tilde{t}\,)} \right\}, \tag{2.19}$$

wobei $\tilde{t}$ über

$$\tan(2\omega\tilde{t}\,) = \frac{\sum_{n=1}^{N} \sin 2\omega t_n}{\sum_{n=1}^{N} \cos 2\omega t_n} \tag{2.20}$$

definiert ist. Diese Gleichung ist das Resultat einer *least–squares* Anpassung des Modells $s(t) = a \cos \omega t + b \sin \omega t$ an die Daten für eine feste Frequenz ω. Aus diesem Grund wird das Lomb–Periodogramm auch häufig *least–squares*– oder LS–Periodogramm[3] genannt. Für Zeitreihen mit konstanten Zeitintervallen $\Delta_t = t_n - t_{n-1}$ (für alle n) reproduziert das Lomb–Periodogramm $P(2\pi n/N\Delta_t)$ die quadrierten Amplituden der Fouriertransformation. Abgesehen von diesem speziellen Fall, existiert keine inverse Transformation für das Lomb–Periodogramm. Daher ist es nicht möglich, den Lomb–Schätzer für das Spektrum in einer der Standardmethoden zur Surrogaterzeugung aus Abschnitt 1.5 zu verwenden, da diese alle auch eine Rücktransformation benötigen.

Das Einbinden in die neue Methode ist im Gegensatz dazu leicht möglich, da hier die inverse Transformation nicht verwendet wird. Das Lomb–Periodogramm kann wie üblich in eine Kostenfunktion der Art

$$E = \sum_{j=1}^{N_f} |P(\omega_j) - P'(\omega_j)| \tag{2.21}$$

eingebaut werden [SS99b]. Dabei wurden konstante Gewichte und $q = 1$ gewählt. Schwieriger gestaltet sich die Wahl der Frequenzen ω_j, an denen das Spektrum von Surrogat

[3] Die Abkürzung LS kann dabei sowohl für *least–squares* als auch für Lomb–Scargle, den Entwicklern dieser Methode, stehen.

und Originalzeitreihe in Übereinstimmung gebracht werden sollen. Während bei der Fouriertransformation sowohl die Anzahl der Frequenzen N_f als auch die Frequenzen selbst natürlicherweise festgelegt sind, gibt es für unregelmäßig aufgenommene Zeitreihen keine kanonische Wahl. Es ist in jedem Fall sinnvoll, sich das Lomb–Periodogramm der zu untersuchenden Zeitreihe anzuschauen, um Anhaltspunkte dafür zu bekommen, welche Frequenzen vernünftigerweise gewählt werden sollten. Dabei sollten z. B. Spitzen im Spektrum auf jeden Fall berücksichtigt werden. Naheliegend und vorurteilsfrei ist die Wahl von $N_f = N/2$ äquidistanten Frequenzen $\omega_j = j\omega_1$, wobei ω_1 so gewählt werden sollte, dass der interessierende Frequenzbereich abgedeckt wird. Orientieren kann man sich auch an den Werten der kleinsten und größten möglichen Frequenz, welche zum einen über der Gesamtzeitraum $\Delta_{\max} = t_N - t_1$ der Zeitreihe und zum anderen über den kleinsten Zeitabstand $\Delta_{\min} = \min_n(t_n - t_{n-1})$ gewonnen werden können.

Bei der Implementierung des Lomb–Periodogramms im Annealing–Prozess sind wie im Falle der Autokovarianz einige Optimierungen möglich, die unnötige verbrauchte Rechenzeit vermeiden. Zunächst ist die benötigte Rechenzeit für die Berechnung des kompletten Lomb–Periodogramms von der Ordnung $O(NN_f)$. Da aber lediglich die Zeitreihenwerte $\{s_n\}$ permutiert werden, während die Zeitpunkte $\{t_n\}$ unverändert bleiben, müssen nicht bei jedem Wertetausch alle Summen in Gleichung (2.19) komplett neu berechnet werden. So können die Werte für $\tilde{t}$ und die Summen im Nenner bereits zu Beginn des Annealings für jede Frequenz ω_j berechnet und abgespeichert werden. In den Summen im Zähler ändern sich bei einem Wertetausch auch nur zwei Terme, die neu berechnet werden müssen. Dadurch reduziert sich die Ordnung der benötigten Rechenzeit für ein Update auf $O(N_f)$.

Alternativ können die linearen Eigenschaften einer Zeitreihe mit unterschiedlichen Zeitintervallen auch direkt im Zeitbereich beschrieben werden. Da der Schätzer (1.1) der Autokovarianz aber unmittelbar auf dem mehrfachen Vorkommen von Zeitreihenwerten mit Zeitabstand τ basiert, ist dieser nicht direkt umsetzbar. Um die nötige Statistik für den Schätzer zu erhalten, ist es vielmehr nötig, den Bereich der möglichen Zeitabstände τ in kleine Intervalle der Länge δ_τ zu unterteilen und die Wertepaare $\{s_i, s_j\}$, deren Zeitabstand in diese Intervalle fallen, zu zählen. Konkret sei

$$B_{ij}(\tau, \delta_\tau) = \{(i,j) : \tau < t_i - t_j \leq \tau + \delta_\tau\} \tag{2.22}$$

die Menge der Indexpaare (i,j), deren zugehörige Zeitabstände im Intervall $[\tau, \tau + \delta_\tau]$ liegen. Damit lässt sich ein Schätzer für die Autokovarianz der Form

$$\gamma_{\delta_\tau}(\tau + \frac{1}{2}\delta_\tau) = \frac{1}{||B_{ij}(\tau, \delta_\tau)||} \sum_{B_{ij}(\tau,\delta_\tau)} (s_i - \bar{s})(s_j - \bar{s}) \tag{2.23}$$

konstruieren, welcher die Autokovarianz mit der Zeitverzögerung $\tau + \delta_\tau/2$ in der Mitte des Intervalls $[\tau, \tau + \delta_\tau]$ schätzt. Der Ausdruck $||B_{ij}(\cdot)||$ steht dabei für die Anzahl der Elemente der Menge $B_{ij}(\cdot)$. Für Zeitreihen mit konstantem Zeitabstand Δ_t stimmt der Schätzer $\gamma_{\Delta_t}(\tau)$ mit den üblichen Autokovarianzen $\gamma(\tau)$ überein.

Die Wahl des Parameters δ_τ muss hier allerdings mit Vorsicht getroffen werden. Wird δ_τ zu klein gewählt, kann die Einteilung zu fein sein, sodass bestimmte Bereiche keine Wertepaare enthalten ($||B_{ij}(\cdot)|| = 0$) und damit der Schätzer (2.23) der Autokovarianz für die entsprechende Zeitverzögerung nicht definiert ist. Für eine vernünftige Schätzung der Autokovarianz sollte jedes Intervall genügend Elemente enthalten, um numerische Fluktuationen herauszumitteln. Dies gilt insbesondere für die vorliegende Anwendung, in der

Surrogate erzeugt werden sollen, die in ihren Autokovarianzen mit der Originalzeitreihe übereinstimmen sollen. Wird die Autokovarianz nur über einige wenige Wertepaare geschätzt, und versucht, dies exakt auf den Surrogaten zu reproduzieren, kann der Spielraum der möglichen Permutationen stark eingeengt sein. Das Problem ist also bei zu feiner Einteilung überspezifiziert und es werden nicht die gewünschten, zufälligen Surrogate produziert. Im Gegensatz dazu wird bei zu grober Einteilung die Autokorrelationsfunktion geglättet, sodass unter Umständen wesentliche Struktur verloren gehen kann.

Auch hier können wieder Optimierungen genutzt werden, die die benötigte Rechenzeit für ein Update reduzieren. Zunächst ist Berechnung aller Autokovarianzen von der Ordnung $O(N^2)$, da die Zugehörigkeit aller Paare (i, j) in die Zeitintervalle ermittelt werden muss. Da diese Zugehörigkeiten aber nur von den unveränderlichen Zeitpunkten t_n abhängen, bleiben diese von der Permutation unberührt. Beim Tausch eines Wertepaares müssen also wieder nur wenige Terme in den Summen neu berechnet werden.

Welche der beiden vorgestellten Alternativen in einem Nichtlinearitätstest mit einer unregelmäßig aufgenommenen Zeitreihe verwendet werden sollte, hängt vom individuellen Fall ab. Sind Periodizitäten in der Zeitreihe vorhanden, ist sicherlich das Lomb–Periodogramm die natürlichere Wahl. In anderen Fällen ist unter Umständen die Implementierung der Autokovarianzen vorzuziehen. Es ist in jedem Fall ratsam, vor einer Entscheidung sowohl das Frequenzspektrum als auch die Autokovarianzfunktion zu inspizieren und aus der gewonnenen Information eine adäquate Kostenfunktion aufzustellen. Die folgenden Resultate wurden alle über die Reproduktion des Lomb–Periodogramms in den Surrogaten gewonnen. Ein Beispiel mit einer Umsetzung der Autokovarianzen aus Gleichung (2.23) lässt sich in [SS00] finden.

Um die Funktionstüchtigkeit der Methode für Zeitreihen mit unterschiedlichen Zeitabständen zu demonstrieren, soll zunächst ein einfaches Beispiel untersucht werden. Bevor aber ein Surrogattest auf Nichtlinearität tatsächlich durchgeführt werden kann, benötigen wir noch eine Teststatistik, mit der die Nichtlinearitäten in unregelmäßig aufgenommenen Zeitreihen gemessen werden können. Die für herkömmliche Zeitreihen bewährten Teststatistiken [SS97b] sind allerdings meist nicht ohne weiteres auf den vorliegenden Fall verallgemeinerbar. Die naheliegende Idee der Interpolation der Zeitreihe auf Zeitpunkte mit gleichen Abstand, um so die üblichen Größen berechnen zu können, war für die Realisierung der linearen Eigenschaften in den Surrogaten, wie bereits argumentiert, nicht empfehlenswert. Im Gegensatz dazu ist dieser Ansatz für die Berechnung von Teststatistiken durchaus erlaubt, da die Teststatistik lediglich als Vergleichsgröße dient. Eine gleiche systematische Abweichung der Teststatistik für Surrogate und Originaldatensatz kann die Gültigkeit des Tests nicht verletzen.

Im Folgenden verwende ich aber nicht den „Umweg“ über die Interpolation, sondern eine Verallgemeinerung des Maßes (1.3) für die Zeitumkehrinvarianz. Dieses ist aufgrund seiner einfachen Struktur leicht auf den Fall der variierenden Zeitabstände erweiterbar. Eine mögliche Verallgemeinerung dieser Teststatistik ist z. B.

$$T^{\text{rev}} = \frac{1}{(\sigma^2)^{3/2}(N-1)} \sum_{n=1}^{N-1} \left(\frac{s_{n+1} - s_n}{t_{n+1} - t_n} \right)^3 , \tag{2.24}$$

was einfach den Mittelwert der Steigungen zur dritten Potenz misst. Für Zeitreihen mit konstantem Zeitabstand $t_{n+1} - t_n \equiv 1$ stimmt dies, abgesehen von der unterschiedlichen Normierung, mit dem Maß aus Gleichung (1.3) mit $\tau = 1$ überein. Es sei darauf hingewiesen, dass eine Asymmetrie unter Zeitumkehr zwar ein guter Indikator für Nichtlinearität

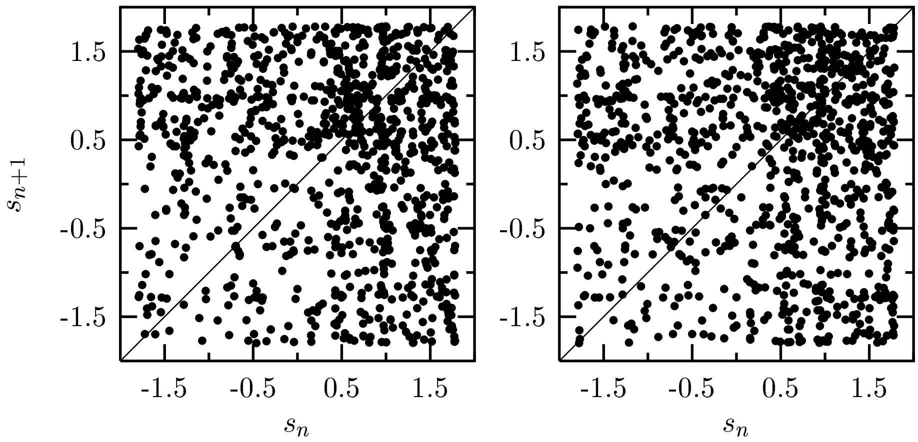

Abbildung 2.12: Zeitlich aufeinander folgende Werte einer unregelmäßig aufgenommenen Hénon–Zeitreihe (links) und eines Surrogates (rechts). Der Test findet signifikante Unterschiede in der Zeitasymmetrie zwischen Originalzeitreihe und Surrogaten.

ist, aber nicht viel Ausschluss darüber gibt, welcher Natur diese Nichtlinearität ist. Hier wurde diese Teststatistik lediglich aufgrund ihrer Einfachheit und Robustheit gewählt.

Im ersten Beispiel soll die Funktionstüchtigkeit eines Tests mit Surrogaten, deren Lomb–Periodogramm mit der Originalzeitreihe übereinstimmt, verifiziert werden. Um eine nichtlineare Zeitreihe mit unterschiedlichen Zeitabständen zu bekommen, werden zunächst 10000 x–Werte der Hénon–Abbildung aus Gleichung (2.17) erzeugt. Aus diesen Daten werden anschließend $N = 1000$ Werte zufällig herausgenommen und nach ihren Zeitindizes sortiert. Die Zeitpunkte t_n sind dabei einfach durch den Index der Werte in der langen Ursprungszeitreihe gegeben. Die resultierenden Zeitabstände $t_n - t_{n-1}$ liegen im Bereich von eins bis 80 und sind exponentialverteilt. In Abbildung 2.12 sind links die zeitlich aufeinanderfolgenden Werte dieser Zeitreihe gegeneinander aufgetragen. Dabei ist ansatzweise die bekannte Form des Hénon–Attraktors zu erahnen, welcher aus Wertepaaren mit Zeitabstand eins aufgebaut ist.

Für den Test auf Nichtlinearität werden Surrogate erzeugt, die das Lomb–Periodogramm der Hénon–Zeitreihe reproduzieren. Die zu minimierende Kostenfunktion 2.21 berücksichtigt dabei $N_f = N/2 = 500$ gleichverteilte Frequenzen. Als Maximalfrequenz wurde $\nu_{\max} = \omega_{\max}/2\pi = 1/2$ gewählt, was der Nyquist–Frequenz einer Zeitreihe mit Zeitabstand eins entspricht und die maximal auflösbare Frequenz darstellt. Durch Minimierung der Kostenfunktion werden 19 Surrogate erzeugt, was einem Signifikanzniveau von 90% entspricht, da für die verwendete Teststatistik (2.24) ein zweiseitiger Test durchgeführt werden muss. Die anschließende Berechnung der Teststatistik ergibt für die Surrogate den Bereich $-0.30 < T^{\text{rev}} < 0.18$ und für die Hénon–Zeitreihe $T^{\text{rev}} = -0.58$. Da der Wert der Originalzeitreihe außerhalb des durch die Surrogate aufgespannten Intervalls liegt, wird die Nullhypothese eines linearen Prozesses verworfen. In diesem einfachen Fall liefert der Test also in der Tat das korrekte Ergebnis.

Als weitere Verifikation für die korrekte Funktionsweise soll im Folgenden ein Test mit einer linearen Zeitreihe durchgeführt werden. Als Basis diente dabei ein AR(1)–Prozess $x_n = 0.95x_{n-1} + \eta_n$, welcher zusätzlich mit der nichtlinearen Messfunktion $s_n = m(x_n) = x_n\sqrt{|x_n|}$ aufgenommen wurde. Aus der so gewonnenen Zeitreihe mit 10000 Werten wurden wieder 1000 zufällig herausgenommen, um eine Zeitreihe mit unterschiedlichen Zeitabstän-

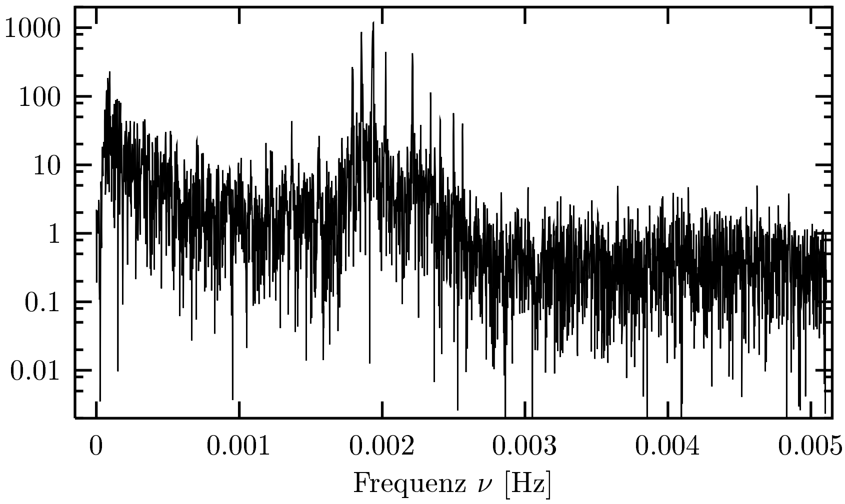

Abbildung 2.13: Das Lomb–Periodogramm von Datensatz E zeigt sowohl um $\nu \approx 0.002$ als auch bei kleinen Frequenzen einen hohen spektralen Anteil.

den zu konstruieren. Der anschließende Test mit 19 Surrogaten war diesmal korrekterweise nicht in der Lage die Nullhypothese zu verwerfen. Für eine zuverlässige Überprüfung der Brauchbarkeit der Lomb–Surrogate müsste die Size des Tests möglichst genau ermittelt werden. Dafür wäre allerdings die Durchführung von vielen unabhängigen Tests mit linearen Zeitreihen nötig. Der damit verbundene Rechenaufwand wäre aber immens und ließe sich kaum mit der Relevanz der Frage rechtfertigen.

Als letztes Beispiel soll der Surrogattest auf experimentelle Daten aus der Astronomie angewendet werden. Als Basis dient dabei Datensatz E, welcher vom Santa Fe Institute anlässlich eines Zeitreihenwettbewerbs veröffentlicht wurde [WG93]. Er beinhaltet mehrere Messreihen der (über kurze Zeiten integrierten) Lichtintensität eines variablen, weißen Zwerges, welche vom *Whole Earth Telescope* [NWC$^+$90] aufgenommen wurden. Die Zeitreihe besteht aus 17 Teilen unterschiedlicher Größe und umspannt insgesamt ungefähr 3.5 Tage. Während einige der Teile sich zeitlich überlappen, existieren zwischen anderen Teile relativ große Zeitabstände, welche eine sinnvolle Interpolation unmöglich machen. Innerhalb der Blöcke sind die Zeitabstände der Messwerte mit $\Delta_t = 10$s konstant. Ein besonderes Interesse an Langzeitentwicklungen und damit niedrigfrequenten Anteilen macht die Untersuchung der Daten als Ganzes wünschenswert. Das in Abbildung 2.13 dargestellte Lomb–Periodogramm der kompletten Zeitreihe zeigt auch in der Tat einen hohen spektralen Anteil bei niedrigen Frequenzen.

Um den zeitlichen Aufwand des Surrogattests im Rahmen zu halten, habe ich die Daten weiter reduziert und über jeweils zwölf aufeinander folgende Messwerte integriert. Die resultierende Zeitreihe umfasst $N = 2260$ Werte und ist in Abbildung 2.14 oben dargestellt. Innerhalb der teilweise überlappenden Blöcke weisen die Daten nun einen Zeitabstand von $\Delta_t = 120\,$s auf, während dazwischen Lücken von bis zu 10000 s auftreten. Bei der Erzeugung der Surrogate fließt in die Kostenfunktion das Lomb–Periodogramm an $N_f = 1130$ gleichverteilten Frequenzen bis $\nu_{\max} = 1/240$ Hz ein. Dies entspricht wieder der Nyquist–Frequenz einer Zeitreihe mit konstantem Zeitabstand $\Delta_t = 120\,$s, sodass alle hochfrequenten Effekte, die auf dem Überlappen der Teile beruhen, unberücksichtigt bleiben. Ein entsprechendes, durch Minimierung dieser Kostenfunktion gewonnenes Surrogat ist eben-

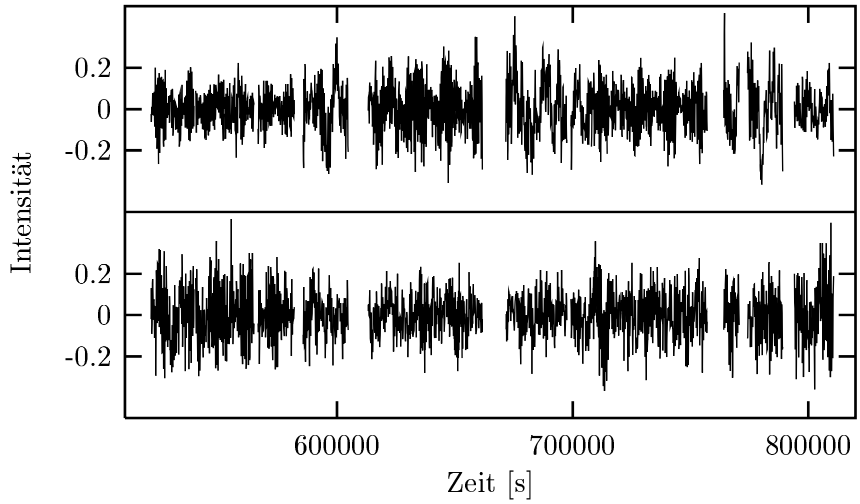

Abbildung 2.14: Der über zwölf Messwerte integrierte Datensatz E (oben) weist Lücken von bis zu 10000 s auf, welche eine sinnvolle Interpolation unmöglich machen. Das Surrogat (unten) stimmt in seinem Lomb–Periodogramm und seiner Verteilung mit dem Originaldatensatz überein und besitzt damit gleiche lineare Eigenschaften.

falls in Abbildung 2.14 dargestellt. Wird anschließend die Zeitumkehr–Teststatistik auf 19 Surrogaten berechnet, so liegt diese im Bereich $-13 \times 10^{-7}\mathrm{s}^{-3} < T^{\mathrm{rev}} < 29 \times 10^{-7}\mathrm{s}^{-3}$. Im Vergleich dazu liefern die Originaldaten den Wert $T^{\mathrm{rev}} = -0.6 \times 10^{-7}\mathrm{s}^{-3}$, welcher ziemlich genau in der Mitte des Intervalls liegt. Die Nullhypothese eines linearen, zeitumkehrinvarianten Prozesses kann also nicht verworfen werden.

Weitere Untersuchungen des Datensatzes zeigen, dass die Autokorrelationen in den einzelnen Blöcken teilweise erheblich voneinander abweichen. Die Zeitreihe kann damit bezüglich ihrer linearen Eigenschaften nicht als stationär angesehen werden. Passende Surrogate für den kompletten Datensatz müssten also über den in Abschnitt 2.1.2 beschriebenen Ansatz erzeugt werden. Da in diesem speziellen Fall die Zeitreihe aus Blöcken mit konstanter Aufnahmerate besteht, ist es hier möglich, auf den individuellen Blöcken Nichtlinearitätstests mit den Standardmethoden durchzuführen. Auch diese finden aber keine signifikanten Nichtlinearitäten und können die Nullhypothese nicht verwerfen.

Die in diesem Abschnitt beschriebene Methode macht es erstmals möglich, Zeitreihen mit unterschiedlichen Zeitintervallen auf Nichtlinearitäten hin zu testen. Dies wurde durch die neue, flexible Methode zur Erzeugung von Surrogaten ermöglicht, worauf hier auch der Schwerpunkt gelegt wurde. Den Aspekt der Verallgemeinerung der Teststatistiken auf diesen Fall habe ich hier nur kurz angesprochen. Weitere Untersuchungen auf diesem Gebiet sind aber sicherlich erforderlich, um unregelmäßig aufgenommene Zeitreihen zuverlässig auf Nichtlinearitäten hin testen zu können.

Kapitel 3

Zusammenfassung und Ausblick

Im ersten Teil dieser Arbeit wurde eine neue Methode entwickelt, um Zeitreihen auf eventuelle nichtlineare Strukturen hin zu testen. Mit einem solchen Nichtlinearitätstest soll entschieden werden, ob die Irregularitäten in einer Zeitreihe stochastischer Natur sind oder zum Teil auf nichtlinearen, zugrunde liegenden Mechanismen beruhen. Neben dieser wesentlichen Charakterisierung der Zeitreihen, sagt ein solcher Test auch aus, ob sich die Anwendung der Methoden der nichtlinearen Zeitreihenanalyse auf die zu untersuchenden Daten lohnt. Als eines der leistungsfähigsten Konzepte hat sich dabei ein Monte–Carlo–Test mit zufälligen Ersatz–Zeitreihen, sogenannten Surrogaten, durchgesetzt.

Die gängigen Standardmethoden zur Erzeugung von Surrogaten implementieren die notwendigen linearen Eigenschaften auf den Surrogaten über eine möglichst exakte Realisierung des Frequenzspektrums. Damit verbunden sind einige Probleme, wie z. B. die Periodizitätsannahme der Fourier–Transformation und die schwierige Behandlung von nichtstationären Zeitreihen. Außerdem wäre es wünschenswert, die linearen Eigenschaften direkt in der Form von Autokorrelationen auf den Surrogaten zu realisieren. All diese Probleme existieren nicht für die hier vorgestellte, neue Methode. Sie implementiert die Autokorrelationen über eine Kostenfunktion, die zur Erzeugung von Surrogaten minimiert werden muss. Dafür wird hier das Verfahren des Simulated Annealing verwendet, mit dem es im Prinzip möglich ist, beliebig tiefe Minima der Kostenfunktion und damit beliebige Genauigkeit für die Eigenschaften der Surrogate zu erreichen.

Der wesentliche Vorteil der neuen Methode gegenüber den Standardmethoden ist aber ihre Flexibilität. So lassen sich nahezu beliebige Eigenschaften auf den Surrogaten realisieren. Dadurch ist es erstmals möglich, auch Zeitreihen mit variierenden Zeitintervallen auf Nichtlinearität hin zu testen. Weiter ist die Behandlung von nichtstationären Zeitreihen möglich, indem bekannte Nichtstationaritäten der Originalzeitreihe auf den Surrogaten reproduziert werden. Schließlich wurde in dieser Arbeit auch ein Beispiel gezeigt, das die Flexibilität der neuen Methode ausnutzte, um eine feste Beziehung der Surrogate zu einer zweiten Zeitreihe zu fixieren. Die demonstrierte Flexibilität hat aber auch ihren Preis: Die neue Methode benötigt wesentlich mehr Rechenzeit als die Standardmethoden, welche auf die Fast–Fourier–Transformation zurückgreifen können. Dieser Nachteil ist aber natürlich nur bei großen Datenmengen und besonders langen Zeitreihen von entscheidender Bedeutung.

Auch wenn die neue Methode ursprünglich einen genaueren und flexibleren Nichtlinearitätstest zum Ziel hatte, sind andere Tests z. B. auf Stationarität oder auf die Existenz höherer Momente vorstellbar. Die einfache Implementierung von beliebigen Eigenschaften auf den Surrogaten sorgt für eine Vielfalt an Möglichkeiten, denen keine Grenzen gesetzt sind. Weitere Anwendungen der neuen Methode sind also zu erwarten. Eine sinnvolle, zukünftige Aufgabe wäre sicherlich noch die Optimierung des Minimierungsverfahrens, welche den wesentlichen Teil der Rechenzeit benötigt. Durch eine solche Entwicklung könnte der verbleibende Nachteil gegenüber den Standardmethoden immer mehr an Bedeutung verlieren.

Teil II

Abhängigkeiten zwischen Zeitreihen

Kapitel 4

Maße für Abhängigkeiten

4.1 Einleitung

Die Frage nach Abhängigkeiten zwischen gemessenen Signalen stellt sich in vielen Bereichen. Dies gilt im Speziellen für *gleichzeitig* gemessene Zeitreihen. Die Aufnahme von mehreren Zeitreihen in einem Experiment geht immer mit der Hoffnung auf ein besseres Verständnis des untersuchten Systems einher. Dabei stellt sich nach der Messung als erstes die Frage, welche *zusätzliche* Information über das System durch den zusätzlichen Messaufwand gewonnen werden konnte. Die Analyse von Abhängigkeiten zwischen den Zeitreihen kann bei der Beantwortung dieser Frage sehr hilfreich sein, wobei der Nachweis einer Abhängigkeit oft von entscheidender Bedeutung ist. Werden beispielsweise an einem System physikalisch unterschiedliche Größen gemessen, kann eine Abhängigkeit zwischen ihnen auf einen physikalischen Zusammenhang dieser Größen hindeuten.

Aber auch die Abhängigkeiten von physikalisch gleichen Größen, die an *einem* System gemessen wurden, können von Interesse sein. Dies ist z. B. der Fall, wenn das System aus Teilsystemen aufgebaut ist, deren Beziehung zueinander untersucht werden soll. Die Analyse von Abhängigkeiten zwischen diesen Teilsystemen kann offenbaren, ob sie isoliert sind und damit getrennt betrachtet werden können oder eine Kopplung zwischen ihnen existiert. Weiter könnte dann die Stärke, Richtung und zeitliche Variation einer solchen Kopplung mit geeigneten Abhängigkeitsmaßen untersucht werden. Ein System, für das diese Fragen im Moment auf großes Interesse stoßen, ist das menschliche Gehirn. Bei der Messung eines Elektroenzephalogramms (EEG) messen die einzelnen Elektroden die Aktivität von verschiedenen Teilbereichen des Gehirns. Die Untersuchung deren Beziehung untereinander könnte weitere Aufschlüsse über die Funktionsweise des Gehirns geben. Ein Phänomen, welches in diesem Zusammenhang häufig zur Sprache kommt, ist die *Synchronisation*, welche genauer in Abschnitt 5.2.1 behandelt wird. Dabei verhalten sich anfangs autonome Teilsysteme bei genügend starker Kopplung in fester Beziehung zueinander.

Die naive Vorstellung von einer Abhängigkeit zwischen zwei Systemen ist meist durch eine sehr starke Asymmetrie geprägt. Auf der einen Seite existiert ein Aktor oder Treiber, der das aktive der beiden Systeme darstellt. Auf der anderen Seite steht das passive Antwortsystem, welches lediglich auf den Aktor reagiert. Dieser Fall der *unidirektionalen* Kopplung ist aber nur eine spezielle Form von möglichen Abhängigkeiten. In der Natur wirkt eine bestehende Kopplung häufig in beide Richtungen, sodass eine *gegenseitige* Abhängigkeit besteht. Einfache, mechanische Beispiele dafür sind das Doppelpendel (ein Pendel an einem Pendel) oder zwei durch eine Feder verbundene Pendel. In beiden Fällen existiert eine beidseitige Abhängigkeit zwischen den Systemen, wobei bei den über eine Feder gekoppelten Pendeln diese Abhängigkeit in beide Richtungen gleich und damit symmetrisch ist.

Weitere Szenarien sind vorstellbar. So könnten sich z. B. beide Systeme annähernd autonom bewegen und lediglich ein System das andere über eine kleine Kopplung stören. Ein Beispiel dafür wären zwei elektrische Schwingkreise, von denen einer durch Abstrahlung den anderen leicht beeinflusst. Die Analyse von Abhängigkeiten zwischen gemessenen Zeitreihen könnte in einem solchen Fall unter Umständen die Richtung der Störung bestimmen. Weiter wäre es auch möglich, dass beide Systeme mit einem unbekannten und unsichtbaren, dritten System in Verbindung stehen. Dieses dritte System könnte dabei sowohl als „Vermittler“ für die Kopplung als auch als autonomer Treiber für beide Systeme fungieren. Im letzteren Fall würden beide Systeme vom gleichen, dritten System abhängen und so unter Umständen Gemeinsamkeiten aufweisen. Es wäre wünschenswert ein Maß zu besitzen, mit dem zwischen Gemeinsamkeiten einerseits und Abhängigkeiten andererseits unterschieden werden könnte. Zwei nahezu identische oder durch einfache Transformation auseinander hervorgehende Zeitreihen weisen große Gemeinsamkeiten auf. Über die Abhängigkeit zwischen ihnen lässt sich hingegen nur schwer eine Aussage machen.

Schließlich soll noch erwähnt werden, dass die Analyse von Gemeinsamkeiten und Abhängigkeiten auch rein praktischer Natur sein kann. Dies ist z. B. der Fall, wenn es darum geht, große Datenmengen zu reduzieren und so die Analyse redundanter Informationen zu verhindern.

4.1.1 Eigenschaften von Abhängigkeitsmaßen

So vielfältig wie die verschiedenen Formen an möglichen Abhängigkeiten zwischen Systemen sind auch die Methoden und Maße, mit denen sie quantifiziert und gemessen werden können. Hier sollen einige grundlegende Eigenschaften und Klassen dieser Maße diskutiert werden. Sei $A(X,Y)$ ein solches skalares Maß, welches die Abhängigkeit zwischen den Systemen X und Y beschreibe. Eine erste wichtige Klassifikation teilt die möglichen Maße danach ein, ob sie unter der Vertauschung von X und Y symmetrisch sind, d. h. ob $A(X,Y) = A(Y,X)$ gilt. Will man an A die Richtung einer Kopplung ablesen, ist dies nur bei einem asymmetrischen Maß möglich, indem man $A(X,Y)$ und $A(Y,X)$ vergleicht. Allerdings existieren auch Maße mit trivialen Asymmetrien, wie z. B. $A(X,Y) = -A(Y,X)$, welche häufig ebenfalls nicht viel Aufschluss über die Richtung der Kopplung geben.

Weiter lassen sich die Abhängigkeitsmaße in „statische“ und „dynamische“ unterteilen. Unter statischen Maßen will ich dabei Größen verstehen, die nicht die Dynamik der einzelnen Systeme oder der Kopplung mit einbeziehen und allein auf der (zeitlich gemittelten) Verteilung der Zustandsvektoren beruhen. Viele Maße verwenden natürlich über Gleichung (1.4) rekonstruierte Zustandsvektoren, welche Zeitreihenwerte zu unterschiedlichen Zeiten beinhalten. Diese Zeitverzögerungsvektoren beinhalten damit bereits für sich gesehen Information über die Dynamik, sodass in diesem Fall eine Unterscheidung fragwürdig ist. Während über statische Maße bereits Stärke und Richtung einer Kopplung ermittelt werden können, haben dynamische Maße den Vorteil unter Umständen auch Aufschluss über die Dynamik der Kopplung geben zu können.

In den nächsten Abschnitten werden zunächst einige gängige Standardmaße für Abhängigkeiten zwischen Zeitreihen beschrieben und diskutiert. Im nächsten Kapitel werden dann zwei neue, viel versprechende Ansätze entwickelt, die als ein entscheidendes Merkmal die Asymmetrie aufweisen. Dadurch kann es, wie bereits erwähnt, unter Umständen möglich sein, auch die Richtung einer Kopplung zwischen zwei Systemen aus den Daten zu extrahieren.

Die in den folgenden Abschnitten diskutierten Größen, beschreiben stets die Beziehung zwischen *zwei* Zeitreihen. Diese werden durchgängig mit $X = \{x_n\}$ und $Y = \{y_n\}$ bezeichnet und ihre Werte x_n und y_n sollen zu gleichen Zeitpunkten $t_n, n = 1, \ldots, N$ aufgenommen worden sein.

4.2 Lineare Abhängigkeitsmaße

4.2.1 Lineare Korrelationen

Eine erste Größe zur Quantifizierung von Abhängigkeiten ist uns bereits in Abschnitt 2.4.2 des ersten Teils dieser Arbeit begegnet: Die *Kreuzkovarianz*

$$\Gamma_{XY}(\tau) = \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} (x_n - \bar{x})(y_{n+\tau} - \bar{y}) \quad (\text{für} \quad \tau \geq 0) \tag{4.1}$$

beschreibt die Korrelationen zwischen den zwei Zeitreihen X und Y und damit die *lineare* Abhängigkeit zwischen den zugrunde liegenden Systemen. Für $\tau < 0$ müssen die Grenzen der Summe auf $n = |\tau|, \ldots, N$ angepasst werden und bei der Normierung durch $(N - |\tau|)$ dividiert werden. Es kann leicht gezeigt werden, dass

$$\Gamma_{XY}(-\tau) = \Gamma_{YX}(\tau) \tag{4.2}$$

gilt. An Gleichung (4.2) können zwei wesentliche Eigenschaften der Kreuzkovarianz abgelesen werden. Zum einen ist die *Auto*kovarianz $\gamma_X(\tau) = \Gamma_{XX}(\tau)$ (im Gegensatz zur Kreuzkovarianz) eine gerade Funktion mit $\gamma_X(-\tau) = \gamma_X(\tau)$. Aus diesem Grund genügt es, lediglich die Autokovarianzen mit positiven Zeitverzögerungen τ zu betrachten. Zweitens ist die Kreuzkovarianz mit $\tau = 0$ symmetrisch unter Vertauschung der Zeitreihen X und Y, d. h. $\Gamma_{XY}(0) = \Gamma_{YX}(0)$. Diese Symmetrie macht es unmöglich, z. B. die Richtung einer Kopplung zwischen zwei Systemen an der Kreuzkovarianz abzulesen.

Für die Analyse von Abhängigkeiten zwischen zwei Zeitreihen ist zu beachten, dass aus einer signifikant nicht verschwindenden Kreuzkovarianz $\Gamma_{XY} \neq 0$ lediglich die Existenz einer *linearen* Abhängigkeit gefolgert werden kann. Da hier unter Abhängigkeit aber stets *jegliche* Abhängigkeit, also inklusive nichtlinearer Abhängigkeit, zu verstehen ist, ist der Umkehrschluss nicht zulässig. D. h. aus einer Abhängigkeit folgt nicht unbedingt eine von Null verschiedene Kreuzkovarianz.

Weitere Größen, die im Zusammenhang mit der Untersuchung von Abhängigkeiten in der linearen Zeitreihenanalyse auftauchen, sollen im Folgenden kurz skizziert werden. Zum einen findet man häufig die *Kreuzkorrelation*

$$\rho_{XY}(\tau) = \frac{\Gamma_{XY}(\tau)}{\sqrt{\Gamma_{XX}(0)\Gamma_{YY}(0)}}, \tag{4.3}$$

welche lediglich eine auf die Einzelvarianzen $\Gamma_{XX}(0) = \sigma_X^2$ bzw. $\Gamma_{YY}(0) = \sigma_Y^2$ normierte Kreuzkovarianz darstellt. Sie hat den Vorteil, dass ihr Wert auf das Intervall $[-1, 1]$ beschränkt ist und so leicht als Maß für die Stärke der Korrelation dienen kann. Während unabhängige Systeme $\rho_{XY}(\tau) \approx 0$ zeigen, weisen identische Systeme einen Wert von $\rho_{XX}(\tau) = 1$ auf. Es ist zu beachten, dass diese Normierung im Gegensatz zur Autokorrelation für die Kreuzkorrelation ρ_{XY} mit $\tau = 0$ natürlich nicht zwangsläufig eins liefert.

Weiterhin ist auch hier wieder eine Analyse im Frequenzbereich möglich. Analog zum Wiener–Khinchin–Theorem (1.11), welches Autokorrelation und Spektrum verknüpft, existiert das *Korrelationtheorem.* Es setzt die Kreuzkorrelation $\rho_{XY}(\tau)$ mit dem *Kreuzspektrum* $\tilde{s}_{XY}(k) = \mathcal{F}\{\rho_{XY}(n)\}$ in Beziehung. Die auf die Einzelspektren normierte Variante des Kreuzspektrums ist die *Kohärenz*

$$\tilde{C}_{XY}(k) = \frac{|\tilde{s}_{XY}(k)|^2}{\sqrt{|\tilde{s}_X(k)|^2|\tilde{s}_Y(k)|^2}}\,, \tag{4.4}$$

welche sehr häufig in der Literatur zu finden ist. Während die Kreuzkorrelation also die linearen Abhängigkeiten zur Zeitverzögerung τ misst, gibt die Kohärenz die linearen Abhängigkeiten der beiden Zeitreihen zur Frequenz f_k an.

4.2.2 Nichtlineare Verallgemeinerung

In Abschnitt 1.3 wurden höhere Ordnungen der Autokorrelation herangezogen, um als Maße für Nichtlinearität in Zeitreihen zu dienen. Eine entsprechende Verallgemeinerung der Kreuzkorrelation würde es ermöglichen, dass auch nichtlineare Abhängigkeiten gemessen werden können. So könnten z. B. durch

$$\Gamma^{(rs)}_{XY}(\tau) = \frac{1}{N-\tau}\sum_{n=1}^{N-\tau}(x_n-\bar{x})^r(y_{n+\tau}-\bar{y})^s\,, \quad r,s=1,2,\ldots \tag{4.5}$$

und Variation von r uns s beliebige, statistische Abhängigkeiten zwischen den Zeitreihen X und Y ermittelt werden. Als statistisch unabhängig würden die Zeitreihen nur gelten, wenn $\Gamma^{(rs)}_{XY}(\tau)$ für alle r, s und τ verschwindet. Die große Anzahl der zu untersuchenden Möglichkeiten der freien Parameter macht dieses Maß allerdings sowohl sehr rechenintensiv als auch sehr unhandlich. Praktischere Größen zur Untersuchung von nichtlinearen Abhängigkeiten werden im folgenden Abschnitt beschrieben.

4.3 Nichtlineare Abhängigkeitsmaße

Lineare Korrelationen sind nur eine mögliche Form, in der Abhängigkeiten zwischen Systemen auftreten können. Wurden in den zu untersuchenden Zeitreihen mit einem statistischen Test bereits Nichtlinearitäten gefunden, ist die Analyse von nichtlinearen Abhängigkeitsmaßen naheliegend. Die Betrachtung von ausschließlich linearen Korrelationen wäre in diesen Fällen unbefriedigend und unter Umständen auch nicht ausreichend. Sie ist aber in jedem Fall ratsam, da lineare Korrelationen selbstverständlich auch zwischen nichtlinearen Systemen auftreten können.

Die Vielfalt an Methoden, nichtlineare Abhängigkeiten zwischen System zu messen, ist natürlicherweise wesentlich größer als die Anzahl an möglichen Ansätzen in der linearen Zeitreihenanalyse. Diesbezüglich unterscheidet sich dieses Gebiet nicht von anderen Bereichen der Zeitreihenanalyse, wie z. B. der Modellierung. Im Folgenden sollen einige gängige Standardmethoden zur Analyse von nichtlinearen Abhängigkeiten diskutiert werden. Sie dienen zum Teil als Basis aber auch als Vergleichswerte für die im nächsten Kapitel vorgestellten, neuen Methoden.

4.3.1 Kreuzkorrelationssumme

Eines der interessantesten Phänomene, die man bei nichtlinearen Systemen beobachten kann, ist das Auftreten von *seltsamen Attraktoren.* Ein Attraktor ist die Menge der vom System angenommenen Punkte im Phasenraum, die sich unter der Dynamik auf sich selbst abbildet. Im Falle von nichtlinearen, dynamischen Systemen können diese von nicht ganzzahliger (d. h. fraktaler) Dimension sein. Eines der wichtigsten Werkzeuge zur numerischen Ermittlung der Dimension aus einer Zeitreihe ist die *Korrelationsumme*

$$C(r) = \frac{2}{(N - t_{\min})(N - t_{\min} - 1)} \sum_{n=1}^{N} \sum_{m=n+t_{\min}}^{N} \Theta(r - \|\vec{x}_n - \vec{x}_m\|)\,, \tag{4.6}$$

welche den Anteil an Punktepaaren $(\vec{x}_n, \vec{x}_m)$ zählt, die im Phasenraum nicht weiter als r voneinander entfernt sind. Im Falle von skalaren Zeitreihen muss diese zunächst nach Gleichung (1.4) in einen genügend hochdimensionalen Raum eingebettet werden, um die Vektoren $\vec{x}_n$ zu erhalten. $\Theta(\cdot)$ ist die *Heaviside*–Stufenfunktion mit $\Theta(z > 0) = 1$ und $\Theta(z \leq 0) = 0$. Bei einer Wahl $t_{\min} > 1$ werden zeitlich benachbarte Paare von der Summation ausgeschlossen und so der Einfluss von zeitlichen Korrelationen unterbunden [The86]. Der eigentliche Aufwand bei der Berechnung von $C(r)$ besteht in der Suche von Nachbarn im Einbettungsraum, wofür eine Reihe schneller Algorithmen, wie z. B. [Gra90a], existieren. Aus der Korrelationssumme $C(r)$ lässt sich über den Grenzwert

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{4.7}$$

zu kleinen Radien r die *Korrelationsdimension* D_2 bestimmen. Im Falle von Zeitreihen ist dieser Grenzwert aber aufgrund der endlichen Länge und eventuellem Rauschanteil nicht sinnvoll bildbar. Ist aber bereits für endliche Abstände r ein Bereich sichtbar, in dem $\log C(r)$ und $\log r$ proportional zueinander sind, kann aus der Steigung $\mathrm{d} \log C(r)/\mathrm{d} \log r$ die Korrelationsdimension D_2 geschätzt werden.

Eine Verallgemeinerung der Korrelationssumme $C(r)$ auf zwei Zeitreihen X und Y ist die *Kreuzkorrelationssumme* [Kan94]

$$C_{XY}(r) = \frac{1}{N^2} \sum_{n,m=1}^{N} \Theta(r - \|\vec{x}_n - \vec{y}_m\|)\,, \tag{4.8}$$

welche für identische Zeitreihen $X = Y$ wie die univariate Korrelationsumme $C(r)$ skaliert. Sie ist symmetrisch unter der Vertauschung von X und Y. Im Grenzwert zu unendlichen Zeitreihen gilt für hinreichend kleine Abstände r die Ungleichung $C_{XY}(r) \leq \frac{1}{2}(C_{XX}(r) + C_{YY}(r))$. Ist die Kreuzkorrelationssumme C_{XY} identisch mit den individuellen Korrelationssummen C_{XX} und C_{YY} kann auf identische Verteilungen bzw. fraktale Maße geschlossen werden [Kan94]. Die Kreuzkorrelationssumme kann also als *Ähnlichkeitsmaß* zwischen Attraktoren angesehen werden. Damit fand sie erfolgreiche Anwendung bei der Untersuchung von Symmetrien von Attraktoren [SG97] und der Klassifikation von Zeitreihen [SS97a]. Weiter kann sie auch bei der Frage nach der Nichtstationarität einer Zeitreihe hilfreich sein [MS96]. Sind die zu untersuchenden Zeitreihen X und Y aber von unterschiedlicher Natur, d. h. z. B. an unterschiedlichen Orten oder Systemen gemessen worden, ist eine geeignete Interpretation der Kreuzkorrelationssumme schwierig und damit ihre sinnvolle Anwendung als Abhängigkeitsmaß fraglich.

4.3.2 Informationstheoretische Maße

Die Informationstheorie geht im Wesentlichen auf die Arbeit von Shannon zur mathematischen Theorie der Kommunikation zurück [SW49]. In ihr werden Zufallsvariablen und Zufallsprozesse betrachtet, deren Output die Symbole eines bestimmten *Alphabets* $\mathcal{A} = \{A_1, A_2, \ldots\}$ annehmen kann. Wird jedem Element des Alphabets eine Wahrscheinlichkeit $p(i) \equiv p(A_i)$ mit $\sum_i p(i) = 1$ zugeordnet, kann daraus die *Shannon–Entropie*

$$I(X, \mathcal{A}) = -\sum_i p(i) \log_2 p(i) \tag{4.9}$$

bestimmt werden, welche eine der zentralen Größen der Informationstheorie darstellt. Sie gibt die mittlere *Information* an, die aus der Beobachtung eines Output–Symbols der Zufallsvariable X gewonnen werden kann. In der Sprache der Kodierungstheorie gibt sie die mittlere Anzahl der benötigten Bits an, um unabhängige Outputs der Zufallsvariablen optimal zu kodieren. Die Basis des Logarithmus bestimmt dabei lediglich, in welcher Einheit die Information I gemessen wird, wobei der Logarithmus zur Basis zwei einem Informationsmaß in Bits entspricht.

Im Falle von skalaren Zeitreihen besteht das Alphabet aus einem Teilintervall der reellen Zahlen bzw. bei einer m–dimensionalen Einbettung (1.4) eine Teilmenge des $\mathbb{R}^m$. Eine sinnvolle Quantifizierung der Entropie ist in diesem Fall nur durch eine Vergröberung des Zustandsraumes und anschließender Schätzung der Wahrscheinlichkeitsverteilung $p(i)$ möglich. Eine Möglichkeit, dies zu bewerkstelligen, besteht in der Einführung einer Unterteilung bzw. *Partition* des Intervalls der angenommenen Zeitreihenwerte in B gleichgroße Boxen. Die entsprechenden Box–Wahrscheinlichkeiten $p(i)$ können z. B. einfach über $p(i) \approx N_i/N$ geschätzt werden, wobei N_i die Anzahl der Zeitreihenwerte ist, die in Box $i = 1, \ldots, B$ fallen. Im Falle von eingebetteten Zeitreihen entsprechen die Boxen natürlich m–dimensionalen Würfeln, die den Phasenraum disjunkt abdecken. Ein freier Parameter bei der Wahl der Partition ist die Boxgröße δ und es stellt sich heraus, dass die Entropie mit der Boxgröße variiert und für kleine δ einem Potenzgesetz folgt. Der Grenzwert

$$D_1 = \lim_{\delta \to 0} \frac{-I(X, \delta)}{\log \delta} \tag{4.10}$$

ist die *Informationsdimension* des Systems, welche eine wichtige Größe im Spektrum der verallgemeinerten Dimensionen D_q darstellt. Ihre Beziehung zu den verallgemeinerten Korrelationssummen und Entropien wird am Ende dieses Abschnitts diskutiert, wenn es darum geht, schnellere Schätzer für die Entropien zu finden.

Bisher wurde die Entropie lediglich als univariate Größe, die auf einer Zeitreihe berechnet wird, behandelt, die selbstverständlich nicht zur Analyse von Abhängigkeiten zweier Zeitreihen dienen können. Eine einfache Verallgemeinerung der Entropie (4.9) auf zwei Zufallsvariablen ist die *zusammengesetze* Entropie

$$I(X, Y) = -\sum_{i=1}^{B_x} \sum_{j=1}^{B_y} p(i,j) \log p(i,j)\,. \tag{4.11}$$

Dabei ist $p(i,j)$ die Wahrscheinlichkeit, dass sich zu einem festen Zeitpunkt das System X in Box i und das System Y in Box j aufhält. Entsprechend gelten für die zusammengesetzten Wahrscheinlichkeiten $p(i,j)$ bei Summierung über einen Index $\sum_i p(i,j) = p_y(j)$ bzw. $\sum_j p(i,j) = p_x(i)$. Die Verallgemeinerung von Gleichung (4.11) auf mehrere Zeitreihen ist offensichtlich.

Zur Ableitung eines Maßes für die Abhängigkeit von X und Y ist entscheidend, dass die zusammengesetzte Wahrscheinlichkeit faktorisiert, $p(i,j) = p_x(i)p_y(j)$, wenn die Systeme völlig unabhängig sind. Setzt man dies in Gleichung (4.11) ein, kann die zusammengesetzte Entropie in die zwei Einzelentropien, $I(X,Y) = I(X)+I(Y)$, aufgespaltet werden. Daraus lässt sich leicht ein Maß für die Abhängigkeit von X und Y bilden. Ein solche Größe ist z. B. die *gegenseitige Information*

$$M(X,Y) = I(X) + I(Y) - I(X,Y)\,, \tag{4.12}$$

welche ein Maß für die Information ist, die man über einen Wert der Zeitreihe Y gewinnt, wenn man den entsprechenden Wert von X (innerhalb der Genauigkeit der Boxgröße δ) kennt. Für unabhängige Zeitreihen liefert Gleichung (4.12) den Wert Null, und falls Y komplett von X abhängt, ist $M(X,Y) = I(Y)$. Im Falle von identischen Systemen ($Y = X$) liefert die zusammengesetzte Entropie $I(X,X) = I(X)$ und ist damit in diesem Fall identisch mit der gegenseitigen Information $M(X,X) = I(X)$. Außerdem sind sowohl $I(X,Y)$ als auch $M(X,Y)$ symmetrisch unter Vertauschung von X und Y. Sie können also nicht dazu dienen z. B. die Richtung einer Kopplung zwischen zwei Systemen zu untersuchen. Die Verallgemeinerung von M auf mehrere Systeme Z_d mit $d = 1,\ldots,D$

$$M(Z_1,\ldots,Z_D) = \sum_{d=1}^{D} I(Z_d) - I(Z_1,\ldots,Z_D) \tag{4.13}$$

wird *Redundanz* genannt.

Das Skalenverhalten der gegenseitigen Information $M(X,Y)$ für kleine Boxgrößen δ ist etwas komplizierter als das der Shannon–Entropie. Für *rauschfreie*, deterministische Systeme mit deterministischer Kopplung hängt M wie die Shannon–Entropie von δ ab und es ist analog zu Gleichung (4.10) möglich, eine „gegenseitige Informationsdimension“ zu definieren. Im Gegensatz dazu konvergiert M für Systeme *mit* Rauschanteil im Limes $\delta \to 0$ auf einen konstanten, endlichen Wert und ist unabhängig von der Partition.

Eine alternative Herleitung der gegenseitigen Information basiert auf der *Kullback–Leibler–Entropie*, welche für ein neues Maß aus Kapitel 5 eine entscheidende Rolle spielt und deshalb hier diskutiert wird. Ist die Wahrscheinlichkeitsverteilung $p(i)$ einer Zufallsvariablen X bekannt, kann aus ihr die optimale Kodierung konstruiert werden, die im Mittel die durch die Entropie $I(X)$ bestimmte Anzahl Bits benötigt. Wird statt dessen eine andere Wahrscheinlichkeitsverteilung $q(i)$ zur Kodierung herangezogen, ist die *überschüssige* Information gerade die Kullback–Leibler–Entropie [Kul59]

$$K = \sum_i p(i) \log \frac{p(i)}{q(i)}\,, \tag{4.14}$$

welche in der Literatur auch unter den Namen *Divergenz* oder *relative* Entropie zu finden ist. Es kann gezeigt werden, dass $K \geq 0$ gilt (siehe z. B. [Gra90b]), wobei $K = 0$ genau dann gilt, wenn die Wahrscheinlichkeitsverteilungen $q(i)$ und $p(i)$ identisch sind. Auch die gegenseitige Information $M(X,Y)$ kann als eine Kullback–Leibler–Entropie verstanden werden. Sie gibt die überschüssige Information wieder, die sich ergibt, wenn die zwei Zufallsvariablen X und Y unter der Annahme der Unabhängigkeit durch die Wahrscheinlichkeitsverteilung $q(i,j) = p_x(i)p_y(j)$ anstelle der zusammengesetzten Wahrscheinlichkeitsverteilung $p(i,j)$ kodiert wurden. Die entsprechende Kullback–Leibler–Entropie ist

$$M(X,Y) = \sum_{i=1}^{B_x} \sum_{j=1}^{B_y} p(i,j) \log \frac{p(i,j)}{p_x(i)p_y(j)}\,, \tag{4.15}$$

was nach Umformen des Logarithmus mit dem Ausdruck (4.12) für die gegenseitige Information übereinstimmt. Damit gilt auch $M(X,Y) \geq 0$, wobei in diesem Fall die Identität nur bei einer Unabhängigkeit von X und Y gilt.

Der naheliegendste und einfachste Weg, die in den Entropien auftauchenden Wahrscheinlichkeiten zu schätzen, besteht in dem bereits angesprochenen Box–Verfahren. Dabei wurde der Phasen– bzw. Einbettungsraum in gleichgroße Boxen der Größe δ unterteilt und dann die Aufenthaltswahrscheinlichkeit durch einfaches Auszählen der Punkte in den Boxen ermittelt. Ist N die Anzahl der Punkte insgesamt und N_i die Anzahl der Punkte in der i–ten Box, dann ist $p(i) \approx N_i/N$. Dabei können die Wahrscheinlichkeiten in Phasenraumgebieten mit niedriger, lokaler Dichte (also kleinem N_i) nicht so gut geschätzt werden wie in Phasenraumgebieten mit hoher Dichte. Entsprechend wurden verfeinerte Box–Verfahren mit variierenden Boxgrößen, die von der lokalen Dichte abhängen, entwickelt [FS86, PAD93]. Weiterhin können für kleine N_i Korrekturterme für die Entropie verwendet werden [Gra88].

Alternativ kann die Entropie auch über die Korrelationssumme berechnet werden. Anstelle der Aufenthaltswahrscheinlichkeiten in Boxen werden dabei Aufenthaltswahrscheinlichkeiten in den Umgebungen von Punkten mit Radius r verwendet. Dadurch können die *verallgemeinerten* Entropien mit den *verallgemeinerten* Korrelationssummen in Beziehung gebracht werden. Für Details der Herleitung dieser Beziehungen sei auf die Referenzen [PS87, GSS91] und [PT95] verwiesen. Ich fasse hier nur die wichtigsten Resultate zusammen, die für die praktischen Berechnungen benötigt werden und für das nächste Kapitel von Bedeutung sind.

Eine Verallgemeinerung der Shannon–Entropie (4.9) ist das Spektrum der *Renyi*–Entropien [Ren70]

$$I_q(\delta) = \frac{1}{1-q} \log \sum_{i=1}^{B} [p(i)]^q \,, \quad q \in \mathbb{R} \,, \tag{4.16}$$

welches im Limes $q \to 1$ die Shannon–Entropie liefert (d. h. $I_1 \equiv I$). Diese stehen mit den verallgemeinerten Korrelationssummen $C_q(r)$ [PS87] gemäß

$$I_q(\delta) \approx -\log C_q(r = \delta/2) \tag{4.17}$$

in Beziehung. Die in Abschnitt 4.3.1 angegebene Korrelationssumme (4.6) entspricht der verallgemeinerten Korrelationssumme mit $q = 2$. In Verbindung mit der Shannon–Entropie I_1 steht die Korrelationssumme C_1, welche durch den Ausdruck

$$\log C_1(r) = \frac{1}{N} \sum_{n=1}^{N} \log \left[\frac{1}{N} \sum_{m=1}^{N} \Theta(r - \|\vec{x}_n - \vec{x}_m\|) \right] \tag{4.18}$$

gegeben ist. Ähnlich wie bei der Korrelationssumme C_2 (Gleichung 4.6) sollten auch hier bei der Berechnung von C_1 die Paare mit kleinem, zeitlichen Abstand von der Summe ausgeschlossen werden, um den Einfluss von zeitlichen Korrelationen zu vermeiden. Schließlich sei noch auf die Verallgemeinerung der Gleichungen (4.7) und (4.10) hingewiesen. Die Korrelationsdimension D_2 und die Informationsdimension D_1 sind zwei Elemente des Dimensionsspektrums D_q, welches über

$$D_q = \lim_{r \to 0} \frac{\log C_q(r)}{\log r} = \lim_{\delta \to 0} \frac{-I_q(\delta)}{\log \delta} \tag{4.19}$$

mit den verallgemeinerten Korrelationssummen und Entropien verknüpft ist.

Welchen Vorteil hat es nun, dass wir alle Entropien durch Korrelationssummen ausdrücken können? Dazu beachte man, dass die Summen in den Korrelationssummen einen Kern–Schätzer [Sil86] darstellen, wobei in der angegebenen Form der einfache Kern $K(\|\vec{x}_n - \vec{x}_m\|/r) = \Theta(r - \|\vec{x}_n - \vec{x}_m\|)$ verwendet wurde. Kernschätzer sind im Allgemeinen den naiven Schätzern von mehrdimensionalen Histogrammen, wie z. B. dem Box–Verfahren, in ihrer Genauigkeit überlegen. Für höhere Einbettungsdimensionen lässt sich zudem auch ein Geschwindigkeitsvorteil erwarten.

Die natürliche Wahl für den Parameters q bei der Bestimmung von Entropien ist $q = 1$. Zum einen liefert die Verwendung von C_1 die ursprüngliche Form der Shannon–Entropie I_1. Zum anderen ist I_1 die einzige der verallgemeinerten Entropien I_q, welche Additivität ausweist. Aber auch die Wahl von $q = 2$ und damit C_2 hat ihre Vorteile. Aufgrund der Symmetrie von C_2 (Gleichung 4.6) kann sie von allen Korrelationssummen C_q mit Abstand am schnellsten berechnet werden. Weiter ist sie die einzige der verallgemeinerten Korrelationsummen, welche keine *finite–sample*–Korrektur benötigt [Gra88] und damit über einen größeren r–Bereich sinnvolle Resultate liefern kann.

Im Falle von endlichen und eventuell verrauschten Zeitreihen ist der Limes zu kleinen Skalen, also kleinen Boxgrößen δ bzw. Nachbarschaftsradien r, nicht sinnvoll bildbar. Um Größen, wie z. B. Entropien, aber trotzdem zu charakteristischen und vergleichbaren Maßen zu machen, muss der Grenzwert geeignet ersetzt werden. Zunächst ist es empfehlenswert sich die Skalenabhängigkeit der betreffenden, zu untersuchenden Größen anzuschauen. Daraus lässt sich meist schon ein sinnvoller Skalenbereich ablesen und unter Umständen ein geeignetes δ oder r finden. Um mehrere informationstheoretische Größen miteinander vergleichen zu können, sollte aber in allen Maßen die gleiche, einmal gewählte Auflösung verwendet werden.

Kapitel 5

Zwei neue Ansätze

5.1 Transfer–Entropie

5.1.1 Zeitverzögerte, gegenseitige Information

Die gegenseitige Information $M(X,Y)$ (4.12) ist eines der attraktivsten Konzepte bei der Untersuchung von Abhängigkeiten zwischen Zeitreihen. Dies liegt zum einen an dem relativ geringen Aufwand, mit dem sie implementiert werden kann, und der es ermöglicht, relativ schnell erste Ergebnisse zu gewinnen. Zum zweiten erlauben die gewonnenen Resultate eine sehr anschauliche Interpretation, da $M(X,Y)$ die Information (in Bits) angibt, die die Systeme über das jeweils andere System beinhalten. Leider können aus der *gegenseitigen* Information — wie der Name schon sagt — keine Aussagen über eine eventuelle Richtung der Kopplung der Systeme gemacht werden. Sie ist ein Maß für eine Größe, die sich beide Systeme „teilen" und damit symmetrisch unter Vertauschung von X und Y. Im nächsten Abschnitt soll eine neue informationstheoretische Größe motiviert und hergeleitet werden, die auf [Sch00b] zurückgeht und eine entsprechende Asymmetrie aufweist, und die es erlaubt, die Richtung der Informationsübertragung zwischen Systemen zu untersuchen.

Eine erste Idee die gegenseitige Information zu desymmetrisieren, besteht in der Einführung einer Zeitverschiebung τ zwischen den beiden Zeitreihen. Es wird also die *zeitverzögerte*, gegenseitige Information

$$M_{XY}(\tau) = \sum_{i=1}^{B_x} \sum_{j=1}^{B_y} p(i_n, j_{n+\tau}) \log \frac{p(i_n, j_{n+\tau})}{p(i_n)p(j_n)} \tag{5.1}$$

gebildet, welche für $\tau = 0$ mit der gegenseitigen Information aus Gleichung (4.15) übereinstimmt. Die Zeitindizes n bzw. $n+\tau$ sollen dabei die *relative* Zeitverzögerung zwischen den Zeitreihen verdeutlichen. Zur besseren Lesbarkeit wurden außerdem die Systemindizes weggelassen, also z. B. $p_x(i)$ durch $p(i)$ ersetzt. Hier und im Folgenden wird der Index i stets zum System X und der Index j zum System Y gehören. Bei Vertauschung von X und Y erfüllt die zeitverzögerte, gegenseitige Information die gleiche Beziehung (Gleichung (4.2)) wie die Kreuzkovarianz: $M_{XY}(-\tau) = M_{YX}(\tau)$. Diese Eigenschaft ist allen Größen gemein, deren Asymmetrie *allein* aus einer zeitlichen Verzögerung zwischen den beiden Systemen herrührt.

Die Betrachtung von zeitverzögerten Größen bei der Untersuchung von Abhängigkeiten zwischen zwei Zeitreihen kann sich aus zwei Gründen lohnen. Zum einen ist es möglich, dass die vermutete Kopplung zwischen den Systemen nicht augenblicklich erfolgt, sondern eine gewisse Zeit braucht. Die Information „fließt" also mit einer endlichen Geschwindigkeit von einem System zum anderen. Damit ein solcher Effekt beobachtet werden kann, muss das Zeitintervall zwischen aufeinander folgenden Zeitreihenwerten natürlich kleiner sein als die Zeitskala, auf der die Kopplung stattfindet. Aber auch eine instantane Kopplung zwischen zwei Systemen ist unter Umständen an zeitverzögerten Größen erkennbar. Dies ist möglich, wenn die durch die Kopplung übertragene Information durch die zeitliche Dynamik der Systeme nicht verloren geht, sondern wenigstens zum Teil weitergetragen wird. Die instantane Kopplung wäre damit indirekt auch über ein zeitverzögertes Maß untersuchbar. Es ist natürlich wünschenswert, die durch die Kopplung übertragene Information von der durch die (chaotische) Dynamik erzeugten Information unterscheiden zu können.

In der nichtlinearen Zeitreihenanalyse findet die zeitverzögerte, gegenseitige Information z. B. bei der Suche nach passenden Einbettungsparametern (1.4) ihre Anwendung. Eine sinnvolle Wahl für die Zeitverzögerung bei der Einbettung ist das τ, bei dem $M_{XX}(\tau)$ sein erstes Minimum zeigt [FS86]. Dadurch können redundante Informationen in einem Einbettungsvektor vermieden werden. In dieser Arbeit wird $M_{XY}(\tau)$ als Vergleichwert für die im nächsten Abschnitt vorgestellten, informationstheoretischen Größen dienen.

5.1.2 Bedingte Entropie

Die Informationstheorie hat ihren Ursprung in der Entwicklung einer Theorie für die *Kommunikation* zwischen zwei Systemen. Da es sich dabei um eine Informationsübermittlung von einem Sender zu einem Empfänger handelt, ist diese per Definition asymmetrisch. Eine wichtige Rolle bei der verrauschten bzw. gestörten Kommunikation spielt die *bedingte* Wahrscheinlichkeit

$$p(i|j) = \frac{p(i,j)}{p(j)}\,. \tag{5.2}$$

Sie gibt die Wahrscheinlichkeit dafür an, dass das Symbol X_i empfangen wird, wenn das Symbol Y_j gesendet wurde. Von der Interpretation innerhalb der Kommunikationstheorie losgelöst, ist $p(i|j)$ die Wahrscheinlichkeit im System X den Zustand i zu beobachten, wenn sich System Y im Zustand j befindet. Entsprechend lässt sich die *bedingte* Entropie

$$I(X|Y) = -\sum_{i=1}^{B_x}\sum_{j=1}^{B_y} p(i,j)\log p(i|j) \tag{5.3}$$

bilden, welche über $I(X|Y) = I(X,Y) - I(Y)$ mit der zusammengesetzten Entropie (4.11) in Beziehung steht. Die bedingte Entropie $I(X|Y)$ gibt also den mittleren *Mangel* an Information über das System X wieder, wenn wir den Zustand von System Y kennen. Folglich steht sie mit der gegenseitigen Information über

$$M(X,Y) = I(X) - I(X|Y) \tag{5.4}$$

in Beziehung. Obwohl die bedingte Entropie eine Asymmetrie unter Vertauschung der Systeme X und Y aufweist, ist sie bei der Untersuchung von gerichteten Abhängigkeiten zwischen Zeitreihen nicht hilfreich. Ihre Asymmetrie $I(X|Y) - I(Y|X) = I(X) - I(Y)$ beruht nämlich einzig auf dem Unterschied in den individuellen Entropien $I(X)$ und $I(Y)$ der Systeme. Würden zwei gekoppelte Systeme mit gleichen individuellen Entropien betrachtet, wäre die bedingte Entropie in beide Richtungen gleich. Sie kann also nicht herangezogen werden, um eine Aussage über die Richtung einer Kopplung zu machen, außer — und dies ist in einem sehr begrenzten Umfang der Fall — die Ungleichheit von $I(X)$ und $I(Y)$ erlaubt bereits eine solche Aussage.

Analog zu Gleichung (4.14) kann auch für bedingte Wahrscheinlichkeiten die Kullback–Leibler–Entropie gebildet werden. Da die im nächsten Abschnitt beschriebene Größe eine Kullback–Entropie dieser Art ist, soll hier kurz darauf eingegangen werden. Für einen bekannten Zustand j des Systems Y ist die Kullback–Entropie

$$K(X|j) = \sum_{i=1}^{B_x} p(i|j) \log \frac{p(i|j)}{q(i|j)} . \tag{5.5}$$

Sie gibt die überschüssige Information an, die kodiert wird, wenn anstelle der wahren bedingten Wahrscheinlichkeiten $p(i|j)$ die Wahrscheinlichkeiten $q(i|j)$ für die Zustände i angenommen werden. Nach einer Mittelung über alle Zustände j mit entsprechenden Wahrscheinlichkeiten $p(j)$ erhält man die mittlere Kullback–Entropie

$$K(X|Y) = \sum_{i=1}^{B_x} \sum_{j=1}^{B_y} p(i,j) \log \frac{p(i|j)}{q(i|j)} . \tag{5.6}$$

5.1.3 Transfer–Entropie

Innerhalb eines Systems können die bedingten Wahrscheinlichkeiten $p(i_{n+1}|i_n)$ von einem Zeitpunkt n zum Zeitpunkt $n+1$ auch als Übergangswahrscheinlichkeiten von einem Zustand i_n zum Zustand $i' = i_{n+1}$ angesehen werden. Die Dynamik eines Prozesses kann also approximativ durch ein Modell beschrieben werden, welches ausschließlich die Übergangswahrscheinlichkeiten einbezieht. Ein solches Modell ist ein *Markov–Prozess* k–ter Ordnung. Er geht davon aus, dass ein Zustand nur von den k vorangegangenen Zuständen des Systems abhängt. Diese Eigenschaft lässt sich als

$$p(i_{n+1}|i_n, \ldots, i_{n-k+1}) = p(i_{n+1}|i_n, \ldots, i_{n-k+1}, i_{n-k}, \ldots) \tag{5.7}$$

schreiben und macht deutlich, dass die Wahrscheinlichkeit das System zum Zeitpunkt $n+1$ im Zustand i_{n+1} zu finden nicht von den Zuständen zu Zeitpunkten kleiner $n-k+1$ abhängt. Im Falle von Zeitreihen können diese Übergangswahrscheinlichkeiten wieder näherungsweise durch das Box–Verfahren oder über Korrelationssummen bestimmt werden. Die so geschätzten Übergangswahrscheinlichkeiten können dann wiederum z. B. für Vorhersagen innerhalb der Genauigkeit der Boxgröße dienen.

Sind die k vorangegangenen Zustände bekannt, wird zur Kodierung des folgenden Zustands im Mittel die Information benötigt, die durch die *Entropie–Rate*

$$\begin{aligned} G^{(k)}(X) &= -\sum p(i_{n+1}, i_n, \ldots, i_{n-k+1}) \log p(i_{n+1}|i_n, \ldots, i_{n-k+1}) \\ &= I(X_{n+1}|X_n, \ldots, X_{n-k+1}) \\ &= I(X_{n+1}, \ldots, X_{n-k+1}) - I(X_n, \ldots, X_{n-k+1}) \end{aligned} \tag{5.8}$$

gegeben ist. Dabei wird in der Summe der ersten Zeile über alle Zustände von X zu den Zeitpunkten $n+1, \ldots, n-k+1$ summiert. In der Form der letzten Zeile wird deutlich, dass die Entropierate G lediglich eine Differenz zwischen zusammengesetzten Entropien darstellt, welche sich nur durch die Einbeziehung eines weiteren Zustands unterscheiden. Im Falle von eingebetteten Zeitreihen (1.4) ist dies die Differenz aus den Entropien der

$m = k+1$ und $m = k$ dimensionalen Zeitverzögerungsvektoren.[1] Das Verhalten der Entropierate G bei kleinen Skalen r ist gerade umgekehrt zum Skalenverhalten der gegenseitigen Information. Während für verrauschte Systeme die Entropierate im Limes $r \to 0$ divergiert, existiert der Grenzwert für deterministische Systeme und ist die *Kolmogorov–Sinai*–Entropie des Systems [Kol93].

Bislang wurde die Entropierate $G(X)$ lediglich als univariate Größe diskutiert. Sie spiegelt im Gegensatz zur Shannon–Entropie die Dynamik des Systems wider. Eine sinnvolle Verallgemeinerung der Entropierate auf *mehrere* Systeme kann also unter Umständen die Dynamik der Abhängigkeiten zwischen den Systemen widerspiegeln. Eine naheliegende Verallgemeinerung auf zwei Systeme wäre die zusammengesetzte Entropierate

$$G^{(k)}(X,Y) = I(X_{n+1}, Y_{n+1} | X_n, \ldots, X_{n-k+1}, Y_n, \ldots, Y_{n-k+1}) \,, \tag{5.9}$$

welche analog zur zusammengesetzten Entropie $I(X,Y)$ gebildet wurde. Daraus könnte ähnlich zu Gleichung (4.12) eine „gegenseitige Informationsrate“ konstruiert werden. Da aber auch $G(X,Y)$ symmetrisch unter Vertauschung von X und Y ist, wäre nicht viel gegenüber der gegenseitigen Information $M(X,Y)$ gewonnen.

Eine andere Möglichkeit besteht darin, die Markov–Eigenschaft (5.7) zu verallgemeinern. Auf der rechten Seite von Gleichung (5.7) können beliebige Zustände, von denen der Zustand i_{n+1} unabhängig ist, eingefügt werden, ohne die Gültigkeit der Gleichung zu verletzen. D. h., dass im Falle der Unabhängigkeit von X und Y auch Zustände j des Systems Y eingefügt werden können, und man

$$p(i_{n+1} | i_n, \ldots, i_{n-k+1}) = p(i_{n+1} | i_n, \ldots, i_{n-k+1}, j_n, \ldots, j_{n-l+1}) \tag{5.10}$$

erhält. Ist hingegen der Zustand i_{n+1} von einem oder mehreren der l Zustände $j_n, \ldots, j_{n-l+1}$ abhängig, verliert Gleichung (5.10) ihre Gültigkeit. Sie gilt also nur unter der Annahme der Unabhängigkeit. Der überschüssige Informationsgehalt infolge der Ungültigkeit dieser Annahme lässt sich als Kullback–Leibler–Entropie (5.6) der Form

$$T(Y \to X) = \sum p(i_{n+1}, \ldots, i_{n-k+1}, j_n, \ldots, j_{n-l+1}) \log \frac{p(i_{n+1} | i_n, \ldots, i_{n-k+1}, j_n, \ldots, j_{n-l+1})}{p(i_{n+1} | i_n, \ldots, i_{n-k+1})} \tag{5.11}$$

ausdrücken und wurde in [Sch00b] *Transfer–Entropie* genannt. Da diese den Informationsübertrag von einem System in das andere misst, ist sie per Konstruktion asymmetrisch unter Vertauschung von X und Y. Sie stellt damit einen viel versprechenden Kandidaten für die Analyse der Richtung von Kopplungen dar. Es sei betont, dass l nicht unbedingt der Ordnung des Markov–Prozesses entsprechen muss, durch den das System Y modelliert werden kann. Vielmehr kann über l bestimmt werden, wie weit in der Vergangenheit von Y nach Einflüssen auf X gesucht werden soll. Bei ähnlichen Systemen ist $l = k$ eine naheliegende Wahl, während aus Rücksicht auf die Rechenzeit auch $l = 1$ gewählt werden kann.

Werden die bedingten Wahrscheinlichkeiten in Gleichung (5.11) über die Beziehung (5.2) in Quotienten aus zusammengesetzten Wahrscheinlichkeiten umgeformt, lässt sich die Transfer–Entropie auch als Summe aus gegenseitigen Informationen (4.12) und Redundanzen (4.13) schreiben,

$$\begin{aligned} T(Y \to X) &= I(X_{n+1} | X_n) - I(X_{n+1} | X_n, Y_n) && (5.12) \\ &= M(X_{n+1}, X_n, Y_n) - M(X_{n+1}, X_n) - M(X_n, Y_n) \,, && (5.13) \end{aligned}$$

[1] Der Einfachheit halber wurde hier eine Zeitverzögerung von $\tau = 1$ angenommen.

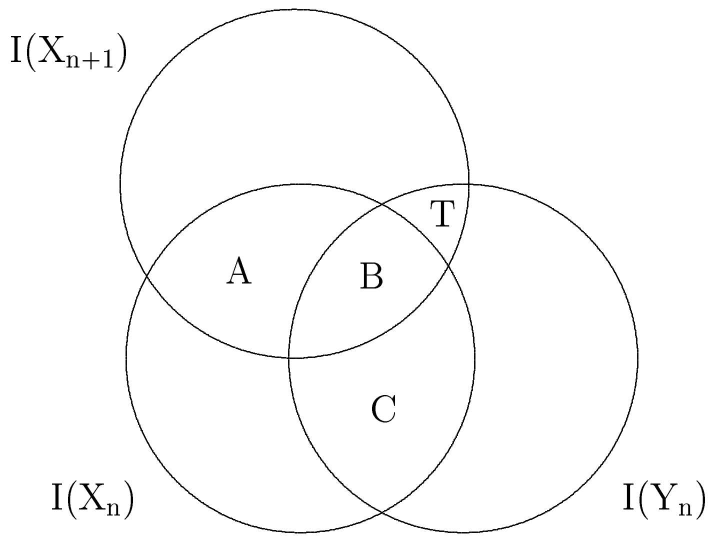

Abbildung 5.1: Anschauliche Darstellung der Beziehungen zwischen den einzelnen Entropien (Kreise) und den entsprechenden gegenseitigen Informationen (Schnittmengen). Die Transfer–Entropie (5.11) entspricht der Menge, die durch den Buchstaben T gekennzeichnet ist.

wobei hier der Einfachheit halber $k = l = 1$ angenommen wurde. Sind die Systeme X und Y unabhängig, verschwindet der letzte Term in Gleichung (5.13) und die beiden ersten Terme heben sich gegeneinander auf. Der wesentliche Unterschied zwischen $T(Y \to X)$ und der zeitverzögerten, gegenseitigen Information $M_{YX}(\tau = 1)$ kann anschaulich anhand von Abbildung 5.1 verstanden werden. In ihr sollen die Kreise die Entropien der Systeme X_n,Y_n und des zeitverschobenen Systems X_{n+1} symbolisieren. Die gegenseitigen Informationen entsprechen den Schnittmengen der Kreise, sodass z. B. $M(X, Y)$ der Vereinigungsmenge B$\cup$C entspricht. Die Transfer–Entropie $T(Y \to X)$ ist im Diagramm mit dem Buchstaben T gekennzeichnet und die zeitverzögerte, gegenseitige Information $M_{YX}(1)$ lässt sich mit der Menge B$\cup$T identifizieren. Während also $M_{YX}(1)$ zum Teil auf den Abhängigkeiten zwischen X und Y zu gleichen Zeiten n beruht, misst die Transfer–Entropie den reinen Abhängigkeitsanteil zwischen X_{n+1} und Y_n. Man beachte, dass die Analogie des Diagramms aber nicht so weit geht, dass daraus Größenverhältnisse abgelesen werden können. So gilt im Allgemeinen *nicht* $M_{YX}(1) \geq T(Y \to X)$, da B nicht zwangsläufig positiv sein muss.

Eine weitere Größe, die für die Untersuchung der Abhängigkeit zwischen X und Y vorstellbar wäre, ist

$$\begin{aligned} U(Y \to X) &= M(X_{n+1}, X_n, Y_n) - M(X_{n+1}, X_n) \qquad (5.14) \\ &= \sum p(i_{n+1}, \dots, i_{n-k+1}, j_n, \dots, j_{n-l+1}) \log \frac{p(i_{n+1}, \dots, i_{n-k+1}, j_n, \dots, j_{n-l+1})}{p(i_{n+1}, \dots, i_{n-k+1}) p(j_n, \dots, j_{n-l+1})} \\ &\hat{=} \mathrm{C} \cup \mathrm{B} \cup \mathrm{T}, \end{aligned}$$

welche sich von der Transfer–Entropie (5.13) nur durch den ohnehin symmetrischen Term $M(X_n, Y_n)$ unterscheidet. In der Form der zweiten Zeile ist erkennbar, dass es sich hierbei wieder um eine Kullback–Leibler–Entropie handelt. Sie misst *jegliche* gegenseitige Information zwischen dem System Y_n einerseits und den Systemen X_n und X_{n+1} andererseits und stellt damit eine weitere mögliche Verallgemeinerung der gegenseitigen Information dar. $U(Y \to X)$ weist unter Vertauschung von X und Y die gleiche Differenz wie die

Transfer–Entropie auf, $U(Y \to X) - U(X \to Y) = T(Y \to X) - T(X \to Y)$, ist aber vom Betrag stets größer: $U(Y \to X) \geq T(Y \to X)$. Ähnlich wie die zeitverzögerte, gegenseitige Information beruht U zum Teil auf den Abhängigkeiten zwischen den Systemen X und Y zu gleichen Zeiten n. In diesem Sinne weist die Transfer–Entropie von allen drei Größen die „stärkste" Asymmetrie auf.

5.2 Interdependenzen

5.2.1 Motivation: Synchronisation

Eines der ersten beobachteten Synchronisationsphänomene beschrieb Huygens in seinem Buch *Horologium Oscillatorium* [Huy73]. Er beobachtete, dass die Pendel zweier Pendeluhren, die an entgegengesetzten Seiten einer Wand hingen, sich nach einer gewissen Zeit synchronisierten und mit festem Phasenunterschied bewegten. Die beiden periodischen Systeme wurden dabei über die Wand, die die Uhren trennte, schwach gekoppelt. Später wurde Synchronisation auch bei gekoppelten, *chaotischen* Systemen gefunden [FY83, PC90], wobei zunächst lediglich die *identische* Synchronisation betrachtet wurde und erst später eine Verallgemeinerung formuliert wurde.

Seien X und Y zwei gekoppelte, deterministische Systeme, deren zeitliche Entwicklung durch die Gleichungen

$$\vec{x}_{n+1} = F(\vec{x}_n) \qquad (5.15)$$

$$\vec{y}_{n+1} = G(\vec{y}_n, \vec{x}_n) \qquad (5.16)$$

beschrieben werden kann. Dabei ist das System X ein autonomer Treiber und Y das Antwortsystem, das von X abhängt. Sind die Systeme abgesehen von der Kopplung identisch, $G(\cdot, 0) = F(\cdot)$ ist es möglich, dass nach einer transienten Dynamik die beiden Systeme sich unabhängig von den Anfangszuständen exakt synchron bewegen und $\vec{y}_n = \vec{x}_n$ erfüllen. In diesem Fall spricht man von identischer Synchronisation, welche natürlicherweise relativ leicht nachzuweisen ist. Vor einigen Jahren wurde der Begriff der Synchronisation von Rulkov et al. [RSTA95] auf Systeme mit unterschiedlicher Dynamik erweitert. Man spricht von *verallgemeinerter* Synchronisation, wenn

$$\vec{y}_n = \psi(\vec{x}_n) \qquad (5.17)$$

erfüllt wird, also der Zustandsvektor $\vec{y}_n$ des Antwortsystems allein von dem Zustandsvektor $\vec{x}_n$ des treibenden Systems abhängt. Eine weitere Fallunterscheidung wurde von Pyragas [Pyr96] eingeführt. Er nannte eine Synchronisation mit glatter Transformation ψ *stark* und sprach im anderen Fall von *schwacher* Synchronisation. Letztere tritt bei schwächeren Kopplungen auf, ist aber natürlicherweise schwerer nachzuweisen.

Es ist zu beachten, dass die umgekehrte Beziehung zu Gleichung (5.17), $\vec{x}_n = \phi(\vec{y}_n)$, *nicht* als Kriterium für Synchronisation herangezogen werden kann. Vielmehr kann gezeigt werden, dass sie für Systeme der Art (5.15) unter sehr allgemeinen Annahmen stets erfüllt ist. Dies ist unabhängig von einer eventuellen Synchronisation der Systeme. Weiter ist zu beachten, dass ein Nachweis von Gleichung (5.17) keine Aussage über die *Richtung* der Kopplung erlaubt. Dies ist offensichtlich, wenn ψ bijektiv ist und eine eindeutige Umkehrfunktion $\phi = \psi^{-1}$ existiert. Trotzdem wurde die Möglichkeit eine Treiber/Antwortsystem–Beziehung aus Daten abzuleiten in den letzten Jahren kontrovers in den Veröffentlichungen zu diesem Thema diskutiert. Während die Autoren von [QMAV99] und [SSC$^+$96] behaupteten, dass dies in der Tat möglich sei, wurde in [AGLE99, QAG00] argumentiert, dass lediglich das „aktivere“ der beiden Systeme, d. h. das System mit mehr angeregten Freiheitsgraden, ermittelt werden kann. Auf diese Fragestellung werde ich im Laufe dieser Arbeit noch zurückkommen.

Neben der Bestimmung der Richtung der Kopplung ist auch der Nachweis von verallgemeinerter Synchronisation im Allgemeinen nicht einfach. In besonders einfachen Fällen ist es unter Umständen möglich, die Funktion ψ direkt annähernd anzupassen. Da aber diese Anpassung niemals exakt sein kann, ist es schwer zu entscheiden, ob kleine Abweichungen von ihr auf ein Fehlen jeglicher Synchronisation hindeuten. Ist die Dynamik der Systeme bekannt, sodass beliebig lange Realisierungen mit unterschiedlichen Anfangsbedingungen betrachtet werden können, kann verallgemeinerte Synchronisation auf zwei verschiedene Arten nachgewiesen werden. Die erste Methode beruht auf der Schätzung der $d + r$ Lyapunov–Exponenten des Gesamtsystems. Dabei entsprechen die ersten d Lyapunov–Exponenten denen des autonomen, treibenden Systems X. Die restlichen r Lyapunov–Exponenten sind die des Antwortsystems Y, welches als nicht–autonomes System mit einem äußeren Treiber betrachtet wird. Verallgemeinerte Synchronisation liegt dann und genau dann vor, wenn alle r Lyapunov–Exponenten des Antwortsystems negativ sind. Infinitesimale Abweichungen in den Anfangsbedingungen werden also exponentiell unterdrückt. Darüber hinaus verliert das Antwortsystem Y im Falle der Synchronisation sogar *jegliche* Abhängigkeit vom eigenen Anfangszustand und hängt nur noch vom Treiber X ab. Dies war ja gerade die Aussage von Gleichung (5.17). Auf diesem Kriterium basiert die zweite Methode zum Nachweis von Synchronisation. Dabei betrachtet man *zwei* identische Antwortsysteme mit unterschiedlichen Anfangsbedingungen, die von demselben Treiber X angetrieben werden. Liegt verallgemeinerte Synchronisation zwischen X und Y vor, dann synchronisieren die beiden Antwortsysteme identisch, was wieder relativ leicht z. B. durch Auftragen der Differenzen nachgewiesen werden kann.

Letztere Methode ist natürlich im Falle von einzelnen, gemessenen Zeitreihen nicht realisierbar. Aber auch die Methode der Lyapunov–Exponenten ist nicht empfehlenswert, da diese im Allgemeinen aus Zeitreihen nur schwer geschätzt werden können. Praxisorientiertere Methoden werden im nächsten Abschnitt diskutiert, wobei ein viel versprechendes, neues Maß im Detail beschrieben wird.

Zum Schluss möchte ich noch kurz den Stellenwert der Analyse von Synchronisation hervorheben. Sie tritt nicht nur in einer Reihe von physikalischen Systemen, wie z. B. Lasern und elektrischen Schaltkreisen auf, sondern auch in vielen anderen Bereichen der Wissenschaft. Speziell in der Biologie besitzen Synchronisationsphänomene eine herausragende Bedeutung. Von einfachen Beispielen, wie z. B. der Koordination von Armen und Beinen beim Gehen, bis zu komplizierteren Abhängigkeiten zwischen z. B. Herzschlag und Atmung spielt die Synchronisation der einzelnen Vorgänge eine wesentliche Rolle. Eines der interessantesten Phänomene ist dabei auf dem Gebiet der Hirnforschung zu finden. Dort wird vermutet, dass epileptische Anfälle mit einer übermäßig starken Synchronisation von Neuronen oder Gruppen von Neuronen im Gehirn einher geht. Die eventuelle Möglichkeit, aus der Analyse von Gehirn–Signalen den Ursprungsort (Fokus) des Anfalls zu bestimmen oder gar den Zeitpunkt von epileptischen Anfällen vorherzusagen, macht dies zu einem Gebiet, in dem im Moment intensiv geforscht wird. In diesem Zusammenhang ist es nicht verwunderlich, dass viele Papiere zum Thema Synchronisation auch die Anwendung auf EEG–Signale behandeln [AGLE99, QMAV99, SSC$^+$96].

5.2.2 Interdependenzmaße S und H

Im vorangegangenen Abschnitt wurde die Analyse von verallgemeinerter Synchronisation unter dem theoretischen Gesichtspunkt betrachtet. Es wurden Nachweismöglichkeiten für den Fall erörtert, in dem die exakte Dynamik der beteiligten Systeme bekannt ist. Im Gegensatz dazu hat man es in der Praxis meist mit nur einer einzelnen Realisierung eines unbekannten, zugrunde liegenden Prozesses zu tun. Dazu kommt immer noch ein mehr oder weniger starker Rauschanteil, dessen Größenordnung meist auch nicht bekannt ist. Ein direkter Nachweis der Synchronisation über Gleichung (5.17) ist also nahezu ausgeschlossen. Unabhängig von einer eventuellen Synchronisation, wäre es aber auch wünschenswert zwischen dem Treiber und Antwortsystem unterscheiden zu können, falls eine solche Beziehung zwischen den beiden beobachteten Systemen tatsächlich existiert. Im Folgenden wird diskutiert, welche Ansätze bei Zeitreihen diesbezüglich zum Erfolg führen können.

Die wesentliche Aussage des funktionalen Zusammenhangs in Gleichung (5.17) besagt, dass jedem Zustand im System X genau ein Zustand im System Y zugeordnet ist. Wenn also das System X nach einer Zeit t zu einem Zustand $\vec{x}_n$ *exakt* zurückkehrt, dann kehrt auch der Zustand von Y exakt zum gleichen Zustand $\vec{y}_n$ zurück. Diese exakte Wiederkehr hat für chaotische Systeme mit zufälliger Anfangsbedingung allerdings verschwindende Wahrscheinlichkeit. Noch weniger ist sie für endliche, verrauscht gemessene Zeitreihen zu erwarten. Allerdings kann als abgeschwächtes Kriterium dienen, dass zu zwei ähnlichen Zuständen im System X gleichzeitig zwei ebenfalls ähnliche Zustände im System Y gehören. Auf diesem Kriterium basierte bereits die Originalarbeit zur verallgemeinerten Synchronisation von Rulkov et al. [RSTA95], in der ein Maß verwendet wurde, welches auf gegenseitigen, falschen nächsten Nachbarn beruht. In [SSC$^+$96, QMAV99] wurde diese Idee in Form von Vorhersagefehlern implementiert, wobei ähnlich zu Gleichung (1.7) nächste Nachbarn in X zur Vorhersage verwendet wurden. Dies wird dann mit dem Vorhersagefehler verglichen, bei dem in der Vorhersage die zeitgleichen Werte aus dem System Y verwendet werden. Schließlich wurde von Arnhold et al. [AGLE99] ein Verfahren vorgeschlagen, welches nicht den „Umweg“ über die Vorhersagen geht, sondern ausschließlich auf Verhältnissen von Nachbarschaftsradien beruht. Dieses Verfahren wird im Folgenden im Detail beschrieben.

Die Zeitreihen X und Y seien als Zustandsvektoren $\{\vec{x}_n\}$ bzw. $\{\vec{y}_n\}$ mit $n = 1, \ldots, N$ gegeben, welche aus skalaren Zeitreihen wie üblich durch Zeitverzögerungseinbettung (1.4) gewonnen werden können. Für jeden Punkt $\vec{x}_n$ werden zunächst die k nächsten Nachbarn gesucht, d. h. die Punkte, die zu $\vec{x}_n$ die kleinsten Abstände aufweisen. Als Abstandsnorm kann dabei z. B. der Euklidische Abstand oder die Maximumsnorm verwendet werden. Die Indizes dieser k Nachbarspunkte werden mit $r_n(i)$, $i = 1, \ldots, k$, bezeichnet. Der mittlere quadrierte Abstand dieser Nachbarn zu dem Punkt $\vec{x}_n$ ist durch

$$R_n^{(k)}(X) = \frac{1}{k} \sum_{i=1}^{k} (\vec{x}_n - \vec{x}_{r_n(i)})^2 \tag{5.18}$$

gegeben. Für $k \ll N$ ist $R_n^{(k)}(X)$ wesentlich kleiner als der mittlere quadrierte Radius $R(X) = \frac{1}{N} \sum_n R_n^{(N-1)}(X)$ der gesamten Punktmenge $\{\vec{x}_n\}$. Darüber hinaus erwartet man die Proportionalität

$$\frac{R_n^{(k)}(X)}{R(X)} \propto \left(\frac{k}{N} \right)^{2/D_1} , \tag{5.19}$$

wobei D_1 die Informationsdimension (4.10) des Wahrscheinlichkeitsmaßes von X ist. Die entsprechende Methode D_1 zu schätzen, heißt *Fixed–Mass*–Verfahren [BP85]. Die Zwei im Exponent der rechten Seite von Gleichung (5.19) trägt der Tatsache Rechnung, dass es sich bei den Größen auf der linken Seite um *quadrierte* Radien handelt.

Analog zu Gleichung (5.18) lässt sich auch für die Zeitreihe Y der mittlere quadrierte Abstand von einem Punkt $\vec{y}_n$ zu seinen k nächsten Nachbarn

$$R_n^{(k)}(Y) = \frac{1}{k}\sum_{i=1}^{k}(\vec{y}_n - \vec{y}_{s_n(i)})^2 \tag{5.20}$$

definieren, wobei hier die Indizes der k Nachbarn mit $s_n(i)$ bezeichnet wurden. Eine erste Größe, die die beiden Systeme in Beziehung stellt, ist der *bedingte* Abstand, welcher die nächsten Nachbarn des jeweils anderen Systems berücksichtigt. Der auf X bedingte mittlere quadrierte Abstand vom Punkt $\vec{y}_n$ ist durch

$$R_n^{(k)}(Y|X) = \frac{1}{k}\sum_{i=1}^{k}(\vec{y}_n - \vec{y}_{r_n(i)})^2\,, \tag{5.21}$$

definiert, wobei im Gegensatz zu Gleichung (5.20) die Indizes der k nächsten Nachbarn von $\vec{x}_n$ verwendet wurden. Hängt das System Y in starkem Maße von X ab, dann impliziert eine Nachbarschaft in X auch eine Nachbarschaft in Y. Demzufolge ist $R_n^{(k)}(Y|X)$ in diesem Fall von der gleichen Größenordnung wie $R_n^{(k)}(Y)$ und der Quotient

$$S_n^{(k)}(Y|X) = \frac{R_n^{(k)}(Y)}{R_n^{(k)}(Y|X)} \tag{5.22}$$

ist von der Ordnung $\mathcal{O}(1)$. Sind die Systeme hingegen völlig unabhängig, ist $R_n^{(k)}(Y|X)$ eher von der Größenordnung von $R_n^{(N-1)}(Y)$ und damit wesentlich größer als $R_n^{(k)}(Y)$, sodass $S_n^{(k)}(Y|X) \ll 1$ ist. Der Ausdruck (5.22) kann also als *lokales* Abhängigkeitsmaß angesehen werden. Durch Mittelung über alle n Zustände erhält man das *globale* Abhängigkeitsmaß

$$S^{(k)}(Y|X) = \frac{1}{N}\sum_{n=1}^{N}\frac{R_n^{(k)}(Y)}{R_n^{(k)}(Y|X)}\,. \tag{5.23}$$

Da für die mittleren Abstände per Konstruktion $R_n^{(k)}(Y|X) \geq R_n^{(k)}(Y) > 0$ gilt, liegt $S^{(k)}(Y|X)$ im Intervall $(0,1]$. Für unabhängige Systeme nimmt $S^{(k)}(Y|X)$ einen kleinen, aber endlichen Wert an, der über Gleichung (5.19) mit der Dimension von Y verknüpft ist. Für abhängige Systeme steigt $S^{(k)}(Y|X)$ mit steigendem Abhängigkeitsgrad der Systeme und für identische Systeme ($X = Y$), also z. B. bei identischer Synchronisation, gilt $S^{(k)}(Y|X) = S^{(k)}(Y|Y) = 1$. Es sei betont, dass (5.23) lediglich ein Abhängigkeitsmaß ist und keine Aussage über die *kausalen* Zusammenhänge zwischen den Systemen X und Y erlaubt.

Die Größen $R_n^{(k)}(X|Y)$ und $S^{(k)}(X|Y)$, welche auf Abständen in X basieren, lassen sich völlig analog zu Gleichungen (5.21) und (5.23) bilden und spiegeln die Abhängigkeit des Systems X von dem System Y wider. $S^{(k)}(X|Y)$ ist damit im Allgemeinen asymmetrisch unter Vertauschung von X und Y, $S^{(k)}(X|Y) \neq S^{(k)}(Y|X)$, und könnte somit Informationen über die Richtung einer Kopplung liefern. Hängt z. B. das System X stärker von Y ab als umgekehrt, dann erwartet man die Beziehung $S^{(k)}(X|Y) > S^{(k)}(Y|X)$. Eine kausale Beziehung zwischen den Systemen lässt sich daraus aber nicht ableiten. In [AGLE99, QAG00] wird argumentiert, dass in einem solchen Fall Y lediglich als das „aktivere“ der beiden Systeme angesehen werden kann. Dabei steht hohe Aktivität mit einer größeren Zahl an angeregten Freiheitsgraden und damit größerer Dimension in Verbindung.

In [AGLE99] werden viele mögliche Alternativen zu Gleichung (5.23) diskutiert, die gegenüber $S(X|Y)$ bezüglich ihrer Abhängigkeitssensitivität eine Steigerung bringen könnten. Eine dieser Alternativen, welche auch in [QAG00] untersucht wurde, wird im Folgenden beschrieben. Der Quotient (5.22) stellt den bedingten Abstand $R_n^{(k)}(Y|X)$ mit dem Abstand zu den nächsten Nachbarn $R_n^{(k)}(Y)$ in Relation. Eine andere Möglichkeit besteht darin, den bedingten Abstand mit dem Abstand zu *zufälligen* oder *allen* anderen Punkten zu vergleichen, also den Quotienten $R_n^{(N-1)}(Y)/R_n^{(k)}(Y|X)$ zu bilden. Durch Mitteln kann daraus wieder ein globales Abhängigkeitsmaß der Art

$$H^{(k)}(Y|X) = \frac{1}{N}\sum_{n=1}^{N} \log \frac{R_n^{(N-1)}(Y)}{R_n^{(k)}(Y|X)} \tag{5.24}$$

gewonnen werden. Die Verwendung des geometrischen Mittels ist dabei durch die Ergodentheorie motiviert, in der geometrische Mittel meist robuster und einfacher zu interpretieren sind. Da $H(Y|X)$ lokale und globale Abstände vergleicht und damit vom Radius der gesamten Punktmenge $\{\vec{y}_n\}$ abhängt, ist es im Gegensatz zu $S(Y|X)$ nicht normiert und nach oben beschränkt. Für unabhängige Systeme sind die Radien in Gleichung (5.24) annähernd gleich und $H(Y|X)$ verschwindet. Dies ist ein weiterer Unterschied zu S, welches für unabhängige Systeme einen endlichen Wert (5.19) in Abhängigkeit von k liefert. Weiter ist zu beachten, dass $H(Y|X)$ im Gegensatz zu S auch negative Werte liefern kann. Dies gilt für den unwahrscheinlichen Fall, dass aus einer Nachbarschaft in X tendenziell große Abstände in Y folgen, und damit der Quotient in Gleichung (5.24) kleiner als 1 ist. Dementsprechend wird $H(Y|X) = 0$ nicht nur für unabhängige Systeme angenommen, sondern könnte auch das unwahrscheinliche Resultat eines exakten gegenseitigen Aufhebens von positiven und negativen Termen sein. Diese Tatsache ist neben der Asymmetrie unter Vertauschung von X und Y einer der wesentlichen Unterschiede zur gegenseitigen Information.

Kapitel 6

Eigenschaften und Anwendungen

Im vorangegangenen Kapitel wurden zwei neue Ansätze für Abhängigkeitsmaße vorgestellt. Im Folgenden sollen zunächst deren grundlegende Eigenschaften anhand einiger einfacher Beispiele illustriert werden. Dabei wird ihre Leistungsfähigkeit sowohl untereinander als auch mit den Standardmethoden aus Kapitel 4 verglichen. Nachdem die Funktionstüchtigkeit und die Vorteile gegenüber den Standardmethoden herausgearbeitet sind, wende ich mich komplizierteren Systemen zu. Diese sind zum einen so konstruiert, dass aus ihnen weitere wichtige Eigenschaften der neuen Maße extrahieren werden können. Zum anderen können diese Systeme die Grenzen des Möglichen aufzeigen und damit als lehrreiche Beispiele für die Untersuchung von gemessenen Zeitreihen angesehen werden. Schließlich sollen einige erste interessante Anwendungen auf Daten aus den Gebieten Ökonomie, Medizin und Klimaforschung vorgestellt werden.

6.1 Einfache, gekoppelte Systeme

6.1.1 Identische Synchronisation

In diesem Abschnitt sollen die Eigenschaften der verschiedenen Abhängigkeitsmaße aus den vorangegangenen Kapiteln an jeweils zwei gekoppelten, chaotischen Systemen illustriert werden. Dazu werden zunächst zwei *identische* Systeme betrachtet, sodass identische Synchronisation auftreten kann. Bei den betrachteten Systemen handelt es sich dabei um Hénon–Abbildungen, welche bereits in Abschnitt 2.4.2 verwendet wurden. Das treibende System X sei völlig autonom und gehorche der Dynamik

$$\begin{aligned} {x_1}' &= 1.4 - x_1^2 + b_x x_2 \\ {x_2}' &= x_1 \,, \end{aligned} \tag{6.1}$$

was bei der üblichen Wahl von $b_x = 0.3$ mit Gleichung (2.17) übereinstimmt. Die ungestrichenen Größen x_i entsprechen dabei den Werten zur Zeit n und die gestrichenen Größen ${x_i}'$ den Werten zur Zeit $n+1$. Daran gekoppelt sei das Antwortsystem Y mit der Dynamik

$$\begin{aligned} {y_1}' &= 1.4 - ((1-K)y_1 + Kx_1)y_1 + b_y y_2 \\ {y_2}' &= y_1 \,, \end{aligned} \tag{6.2}$$

wobei über den Parameter K die Stärke der Kopplung variiert werden kann. Für $K = 0$ sind die Systeme unabhängig und auch das Antwortsystem Y zeigt eine autonome Dynamik, welche für $b_y = 0.3$ mit der von System X identisch ist. In diesem Fall würden bei

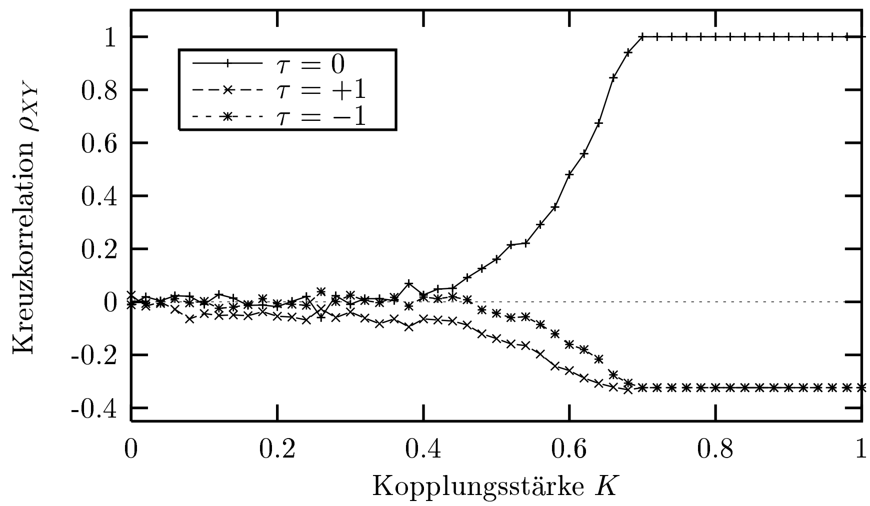

Abbildung 6.1: Kreuzkorrelationen zwischen zwei gekoppelten Hénon–Abbildungen (6.1) und (6.2) in Abhängigkeit von K für verschiedene τ. Die zeitverzögerten Kreuzkorrelationen ($\tau \pm 1$) weichen systematisch voneinander ab.

exakt übereinstimmenden Anfangsbedingungen beide Systeme dieselbe Trajektorie aufweisen. Bei unterschiedlichen Anfangsbedingungen hingegen bewegen sich beide Systeme völlig unabhängig voneinander. Das Synchronisationsverhalten dieser gekoppelten Systeme für $K > 0$ wurde in [QAG00] anhand des maximalen Lyapunov–Exponenten des Antwortsystems untersucht. Dieser ist für $K > 0.7$ negativ und die Systeme synchronisieren identisch. Zusätzlich tritt im Bereich $0.47 < K < 0.52$ schwache, verallgemeinerte Synchronisation auf. Im Folgenden sollen anhand dieses Systems die Eigenschaften der Abhängigkeitsmaße verglichen werden.

Die zur Untersuchung herangezogenen Zeitreihen bestehen jeweils aus den ersten Komponenten x_1 bzw. y_1 der Systeme X und Y. Die Anfangsbedingungen werden in beiden Systemen stets unterschiedlich gewählt. Um den Einfluss von transientem Verhalten auszuschließen, werden die ersten 10^6 Iterationen verworfen, bevor die Zeitreihen mit $N = 10000$ Werten aufgenommen werden. Außerdem wurde in jedem Fall verifiziert, dass die Ergebnisse nicht von den Anfangsbedingungen abhängen.

Zunächst sollen die linearen Korrelationen zwischen den Systemen X und Y in Abhängigkeit von der Kopplungsstärke untersucht werden. Dazu wurden die Kreuzkorrelationen aus Gleichung (4.3) für $\tau = -1, 0, 1$ berechnet und in Abbildung 6.1 aufgetragen. Die Kreuzkorrelation $\rho_{XY}(\tau = 0)$ ist für Kopplungsstärken bis ungefähr $K \approx 0.4$ mit Null verträglich und steigt dann monoton bis $K \approx 0.7$ an. Darüber synchronisieren die Systeme identisch ($Y = X$) und die Kreuzkorrelation nimmt den Wert der Autokorrelation des Treibersystems X an, welche für $\tau = 0$ aufgrund der Normierung 1 ist. An der Kreuzkorrelation $\rho_{XY}(0)$ ist also deutlich das Einsetzen der identischen Synchronisation zu erkennen. Aber auch für Kopplungsstärken im Bereich $0.4 < K < 0.7$ ist eine lineare Abhängigkeit zwischen den Systemen vorhanden. Weiterhin sind in Abbildung 6.1 die zeitverzögerten Kreuzkorrelationen mit $\tau = \pm 1$ aufgetragen. Auch an ihnen ist das Eintreten der identischen Synchronisation zu erkennen, da auch sie für $K > 0.7$ beide exakt mit der Autokorrelation $\varrho_X(|\tau| = 1) \approx -0.29$ der Hénon–Abbildung übereinstimmen. Wesentlich interessanter ist aber die Tatsache, dass die Kreuzkorrelation $\rho_{XY}(1)$ für Kopp-

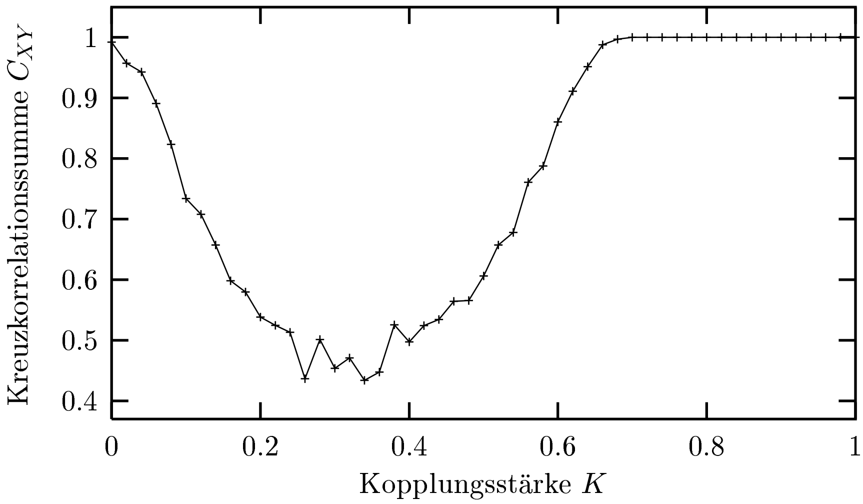

Abbildung 6.2: Kreuzkorrelationssumme von zwei gekoppelten Hénon–Abbildungen in Abhängigkeit von der Kopplungstärke K. Sowohl für unabhängige als auch für identisch synchronisierte Systeme wird der Maximalwert erreicht.

lungsstärken $K < 0.7$ systematisch kleiner als die Kreuzkorrelation $\rho_{XY}(-1) = \rho_{YX}(1)$ ist. Vom Betrag her ist also die Korrelation zwischen x und y' größer als die Korrelation zwischen y und x', welche für $K < 0.45$ sogar annähernd verschwindet. Betrachtet man die Gleichungen (6.1) und (6.2), ist dieses Verhalten aber nicht verwunderlich. Während ${y_1}'$ explizit von x_1 abhängt, ist das System X autonom und völlig unabhängig von Y. Erst bei Einsetzen der schwachen, verallgemeinerten Synchronisation bei $K > 0.47$ ist auch deren Korrelation von Null verschieden. Abgesehen von dieser Tatsache, sind aber keine weiteren Anzeichen für die schwache Synchronisation in den Kreuzkorrelationen zu erkennen. Sie ist im Gegensatz zur identischen Synchronisation offensichtlich wesentlich schwieriger nachzuweisen. Es sei bemerkt, dass eine Betrachtung der Kreuzkorrelationen mit $|\tau| > 1$ keine zusätzlichen Erkenntnisse an den Tag bringt. Für $\tau = 2$ ist die gleiche Asymmetrie wie für $\tau = 1$ beobachtbar. Diese verschwindet aber für $|\tau| \geq 3$, was damit zusammenhängt, dass auch die Autokorrelation der Hénon–Abbildung mit steigender Zeitverzögerung τ rasch abfällt. Zusammenfassend lässt sich sagen, dass in diesem einfachen Fall über die Kreuzkorrelation bereits wesentliche Erkenntnisse über die Abhängigkeit der Systeme gewonnen werden können.

Als erste Größe, die die nichtlinearen Beziehungen zwischen den Systemen beschreibt, soll die Kreuzkorrelationssumme aus Abschnitt 4.3.1 betrachtet werden. Dafür müssen die Zeitreihen zunächst in einen genügend hochdimensionalen Raum über Gleichung (1.4) eingebettet werden. Wie an Gleichung (6.1) zu sehen, reicht bei der Hénon–Abbildung eine Einbettungsdimension von $m = 2$ aus, um den Attraktor eindeutig abzubilden. Eine Dimensionsanalyse des Antwortsystems Y zeigt aber, dass die Dimension dessen Attraktors größer als zwei sein kann. Im Folgenden habe ich daher stets Einbettungen mit $m = 3$ verwendet. Weiter muss eine Wahl für den Nachbarschaftsradius r gewählt werden. Da es sich hier um rauschfreie und relativ lange Zeitreihen handelt, existiert ein relativ großer r–Bereich, in dem $\log r$ und $\log C_{XY}(r)$ proportional zueinander sind. Im Folgenden habe ich $r = 0.1$ gewählt, welches in diesem Skalenbereich liegt. Eine andere Wahl würde qualitativ aber dieselben Resultate liefern. Anschließend wurde in Abhängigkeit von der Kopplungsstärke K die normierten Kreuzkorrelationssummen $C'_{XY} = C_{XY}/\sqrt{C_{XX}C_{YY}}$

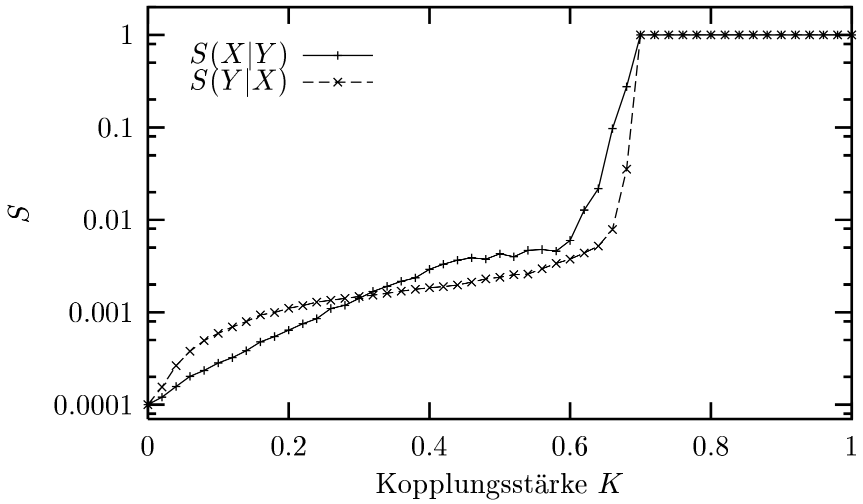

Abbildung 6.3: $S(X|Y)$ und $S(Y|X)$ für zwei identische, gekoppelte Hénon–Abbildungen in Abhängigkeit von K. Der Anstieg bei $K \approx 0.7$ geht über zwei Zehnerpotenzen.

zwischen den Systemen X und Y berechnet. Das Resultat ist in Abbildung 6.2 aufgetragen. Der Maximalwert von $C'_{XY} = 1$ wird dabei sowohl für $K > 0.7$ als auch für $K = 0$ angenommen. In diesen beiden Fällen weisen die Systeme X und Y exakt die gleichen Attraktoren auf. Es ist aber anhand der Kreuzkorrelationssumme nicht möglich, zu entscheiden, ob die Trajektorien der beiden Systeme unabhängig (wie bei $K = 0$) oder völlig identisch (wie bei $K > 0.7$) sind. Dies bestätigt die Aussage aus Abschnitt 4.3.1, dass die Kreuzkorrelationssumme ein Maß für die Ähnlichkeit von Attraktoren ist und nicht unbedingt eine Aussage über mögliche Abhängigkeiten macht. Ein weiterer Nachteil der Kreuzkorrelationssumme ist deren Symmetrie unter Vertauschung von X und Y. Es ist hier auch nicht möglich, eine Asymmetrie durch das Einführen einer Zeitverzögerung zwischen den Systemen herbeizuführen, wie dies für die Kreuzkovarianz oder die gegenseitige Information der Fall ist. Da in Gleichung (4.8) für beide Systeme über die Zeit summiert wird, würde eine Ersetzung von z. B. $\vec{y}_m$ durch $\vec{y}_{m+\tau}$ keinen anderen Wert für C_{XY} liefern.

Als nächstes sollen die Eigenschaften der in Abschnitt 5.2.2 vorgestellten Abhängigkeitsmaße S und H untersucht werden. Wie im Fall der Kreuzkorrelationssumme wurde für beide Systeme die Einbettungsdimension $m = 3$ gewählt. Weiter muss eine sinnvolle Wahl für die Anzahl der nächsten Nachbarn k getroffen werden. Um die mittleren Radien zu den nächsten Nachbarn sinnvoll schätzen zu können, sollte k nicht zu klein gewählt werden. Andererseits sollte k wesentlich kleiner als die Länge N der Zeitreihe sein. Ich habe in Übereinstimmung mit [QAG00] festgestellt, dass die Wahl von k keinen wesentlichen Einfluss auf die qualitativen Resultate hat. Im Folgenden habe ich jeweils $k = 20$ verwendet. Bei der Bestimmung von Abständen im Einbettungsraum wurde die Maximumsnorm verwendet.

Das Verhalten der Abhängigkeitsmaße $S(X|Y)$ und $S(Y|X)$ aus Gleichung (5.23) ist in Abhängigkeit von der Kopplungsstärke K in Abbildung 6.3 dargestellt. Die S–Achse wurde dabei logarithmisch gewählt, um das Verhalten bei schwacher Kopplung besser hervorzuheben. Auch hier ist wieder ein deutlicher Anstieg bei Eintreten der identischen Synchronisation bei $K = 0.7$ sichtbar. Allerdings ist dieser Anstieg wesentlich stärker als z. B. bei der Kreuzkorrelation ausgeprägt und beträgt mehr als zwei Zehnerpotenzen.

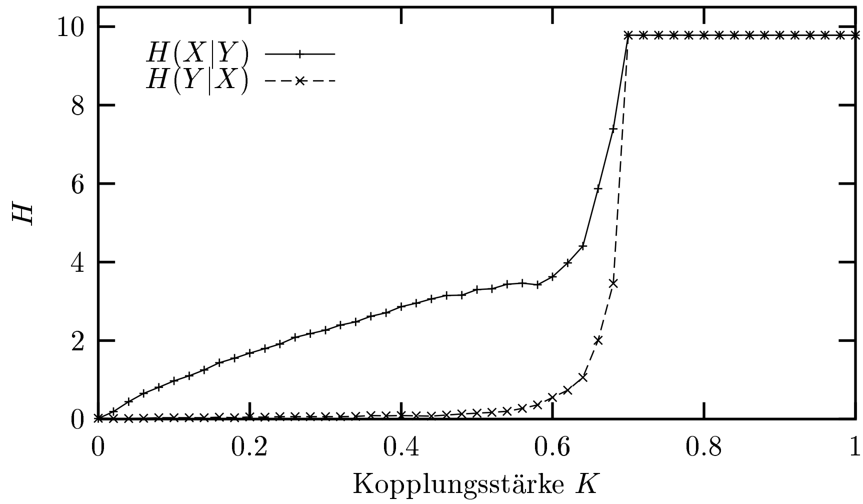

Abbildung 6.4: $H(X|Y)$ und $H(Y|X)$ für zwei identische, gekoppelte Hénon–Abbildungen in Abhängigkeit von K. Unabhängig von K wird $H(X|Y) \geq H(Y|X)$ erfüllt, wobei die Identität nur für die völlige Unabhängigkeit und die identische Synchronisation gilt.

Dass es sich um identische Synchronisation handelt, ist an der Tatsache abzulesen, dass $S(X|Y)$ und $S(Y|X)$ identisch sind und den Wert $S = 1$ liefern. Das Verhalten für $K < 0.7$ muss differenzierter diskutiert werden. Für unabhängige Systeme liefert $S(X|Y)$ nach Gleichung (5.19) einen endlichen Wert, der proportional zu k^{2/D_X} ist. Hat also das System Y eine Dimension $D_Y > D_X$, so erhalten wir bereits für unabhängige Systeme die Asymmetrie $S(Y|X) > S(X|Y)$. Da hier die Systeme identisch sind, findet man für $K = 0$ die Identität $S(Y|X) = S(X|Y)$. Werden die Systeme schwach gekoppelt, steigt die Dimension des Antwortsystems Y, während D_X konstant bleibt. Entsprechend zeigt Abbildung 6.3 für $K < 0.3$ die Beziehung $S(Y|X) > S(X|Y)$. Für Kopplungsstärken $K > 0.3$ kehrt sich diese Ungleichung um, woraus abzulesen ist, dass das System X stärker von Y abhängt als umgekehrt. Dieses Resultat wurde übereinstimmend in [QAG00] und [SSC+96] gefunden, wobei in letzterem gezeigt wurde, dass der Treiber besser durch das Antwortsystem vorhergesagt werden kann als umgekehrt. Dies stimmt auch mit dem Bild überein, dass das Antwortsystem eine größere Anzahl an „aktiven“ Freiheitsgraden besitzt. Kurz vor dem Einsetzen der identischen Synchronisation ist die Asymmetrie $S(X|Y) > S(Y|X)$ am stärksten ausgeprägt. Während aber $S(X|Y)$ im Bereich $0.6 < K < 0.7$ kontinuierlich ansteigt, lässt sich das Verhalten von $S(Y|X)$ eher als ein sprunghafter Anstieg beschreiben. Das Verhalten von S hängt also insgesamt nicht–trivial von der Kopplungstärke K ab und es existiert selbst in diesem einfachen Fall keine allgemein gültige Beziehung zwischen $S(X|Y)$ und $S(Y|X)$.

Anders sieht das im Fall des Abhängigkeitsmaßes H aus, dessen Verhalten in Abbildung 6.4 dargestellt ist. Unabhängig von der Kopplungsstärke gilt vor dem Eintreten der identischen Synchronisation stets

$$H(X|Y) > H(Y|X)\,. \tag{6.3}$$

Während $H(X|Y)$ schon bei kleinen K kontinuierlich mit der Kopplungsstärke ansteigt, ist $H(Y|X)$ nahezu unabhängig von K und liefert sehr kleine Werte. Für unabhängige Systeme ($K = 0$) gilt, wie erwartet, $H(X|Y) = H(Y|X) = 0$. Im Gegensatz zu S

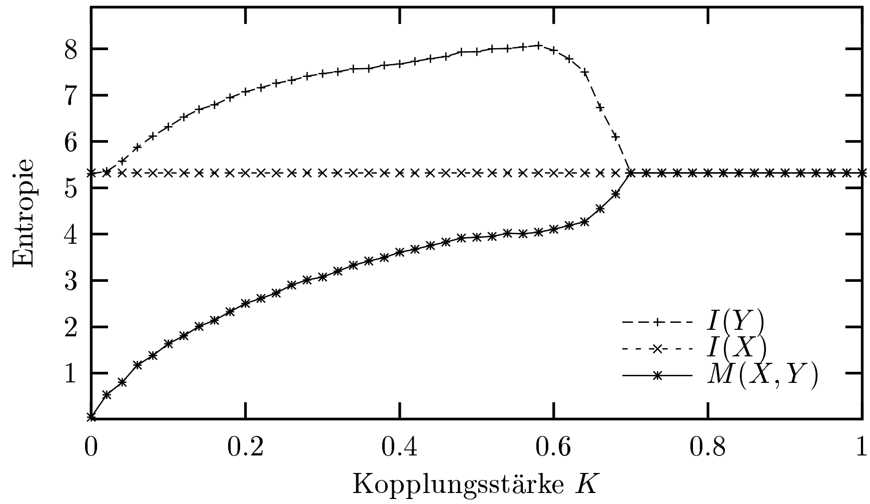

Abbildung 6.5: Entropie des treibenden Systems $I(X)$ und des Antwortsystems $I(Y)$ in Abhängigkeit von der Kopplungsstärke K. Die gegenseitige Information $M(X,Y)$ steigt monoton mit K.

lässt sich bereits für schwache Kopplungen an der Asymmetrie (6.3) ablesen, dass X stärker von Y abhängt als umgekehrt. Erst ab ungefähr $K \approx 0.5$ wird auch $H(Y|X)$ signifikant von Null verschieden. Dies stimmt in etwa mit dem Punkt überein, an dem die schwache, verallgemeinerte Synchronisation auftritt. Bis $K = 0.7$ steigen dann sowohl $H(X|Y)$ als auch $H(Y|X)$ auf den Maximalwert an. Die Tatsache, dass darüber beide Größen exakt übereinstimmen, ist wieder ein Indiz für die identische Synchronisation. Zusammenfassend lässt sich sagen, dass in diesem Fall das Abhängigkeitsmaß H aufgrund seiner eindeutigen Asymmetrie dem Maß S vorzuziehen ist. Die Vorteile gegenüber der zeitverzögerten Kreuzkorrelation sind zum einen die Berücksichtigung von nichtlinearen Abhängigkeiten und zum anderen die stärkere Ausprägung der Asymmetrie.

Zuletzt sollen die informationstheoretischen Größen, welche in den Abschnitten 4.3.2 und 5.1 eingeführt wurden, auf die gekoppelten Hénon–Abbildungen (6.1) angewendet werden. Sie alle basieren auf Entropien, die im Folgenden nach Gleichung (4.17) über die Korrelationsumme $C_1(r)$ geschätzt werden. Auch hier ist es zunächst empfehlenswert, das Skalenverhalten, d. h. die Abhängigkeit vom Nachbarschaftsradius r, zu betrachten. Wie schon bei der Kreuzkorrelationssumme C_{XY} sind auch hier die Skalenbereiche aufgrund der reinen Dynamik und Länge der verwendeten Zeitreihen relativ groß. Im Folgenden wurde der Nachbarschaftsradius auf $r = 0.05$ fixiert, um alle Entropien sinnvoll vergleichen zu können.

Als erste informationstheoretische Größe soll die gegenseitige Information (4.12) zwischen dem treibenden und dem Antwortsystem bestimmt werden. Auch hier werden die Zeitreihen zunächst wieder in einen $m = 3$ dimensionalen Raum eingebettet und anschließend die Korrelationssummen berechnet. In Abbildung 6.5 sind die Einzelentropien und die gegenseitige Information $M(X,Y)$ in Abhängigkeit von der Kopplungsstärke aufgetragen. Da das treibende System X autonom ist und nicht von K abhängt, ist auch seine Entropie $I(X)$ konstant und unabhängig von der Kopplungstärke. Anders sieht das für die Entropie des Antwortsystems Y aus. Sie ist für $K = 0$ mit der Entropie $I(X)$ identisch und steigt dann mit stärker werdender Kopplung an. Sobald ab $K \approx 0.7$ die identische Synchronisation einsetzt, gilt selbstverständlich wieder $I(Y) = I(X)$. An der

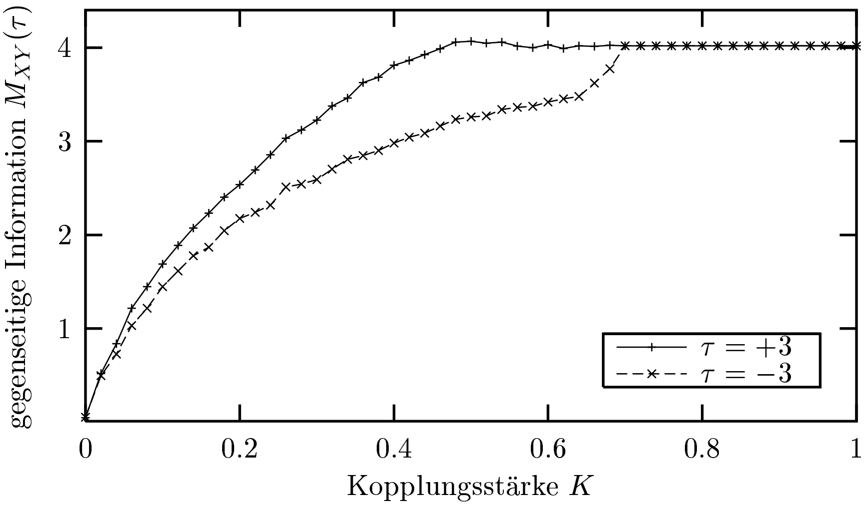

Abbildung 6.6: Zeitverzögerte, gegenseitige Information mit $\tau = \pm 3$ zwischen den gekoppelten Hénon–Abbildungen (6.1) und (6.2).

Tatsache $I(Y) \geq I(X)$ wird deutlich, dass das Antwortsystem zusätzliche Information trägt, die es über die Kopplung aus dem treibenden System gewinnt. Dies lässt sich auch am Verlauf der gegenseitigen Information $M(X, Y)$ in Abhängigkeit von K ablesen. Diese steigt monoton mit der Kopplungsstärke an, wobei der Anstieg vor dem Eintreten der identischen Synchronisation deutlich steiler wird. Für $K > 0.7$ ist die gegenseitige Information $M(X, Y = X)$ identisch mit den Einzelentropien $M(X, X) = I(X) = I(Y)$. Die K–Abhängigkeit von $M(X, Y)$ ist mit der von $H(X|Y)$ in Abbildung 6.4 vergleichbar.

Da die gegenseitige Information symmetrisch unter Vertauschung von X und Y ist, kann sie keine Aussage über die unterschiedlichen Rollen der beiden Systeme treffen. Wie in Abschnitt 5.1.1 beschrieben, kann eine Asymmetrie aber über eine Zeitverzögerung erreicht werden. Abbildung 6.6 zeigt die zeitverzögerten, gegenseitigen Information $M_{XY}(\tau)$ mit $\tau = \pm 3$. Andere Werte für $|\tau|$ liefern qualitativ das gleiche Verhalten. Wie an Gleichung (5.1) zu sehen, bildet $M_{XY}(\tau)$ die gegenseitige Information des Systems X zum Zeitpunkt n und dem System Y zum Zeitpunkt $n + \tau$. Da zukünftige Werte von Y von vergangenen Werten von X und Y abhängen, erwartet man positives $M_{XY}(\tau > 0)$ sobald $K > 0$ ist. Dies wird in Abbildung 6.6 auch bestätigt. Allerdings ist aber auch $M_{XY}(\tau < 0)$ größer als Null, obwohl X nicht von Y abhängt. Dies hängt damit zusammen, dass aufgrund der Dynamik des Systems X natürlich x_1'' (über x_1') von x_1 abhängt. Andererseits hängt aber auch y_1' von x_1 ab. Da also sowohl y_1' als auch x_1'' von x_1 abhängen, tragen beide gemeinsame Information, die von $M_{XY}(\tau = -1)$ gemessen wird. Für die hier verwendeten Einbettungsvektoren mit $m = 3$ ist dieser Effekt natürlich stärker ausgeprägt als bei Verwendung von skalaren Zeitreihen. Er ist aber auch bei $m = 1$ vorhanden. Diese „indirekte“ Abhängigkeit über einen gemeinsamen Vergangenheitswert ist selbstverständlich nicht so stark ausgeprägt wie der Zusammenhang zwischen y_1' von x_1, was zu der Asymmetrie $M_{XY}(\tau > 0) > M_{XY}(\tau < 0)$ führt, die in Abbildung 6.6 zu sehen ist. Für identische Systeme ($K > 0.7$) ist diese Asymmetrie selbstverständlich wieder aufgehoben.

Zu guter letzt soll die Transfer–Entropie (5.11) auf das System der gekoppelten Hénon–Abbildungen angewendet werden. Dazu muss zunächst eine Wahl für den Parameter k, die

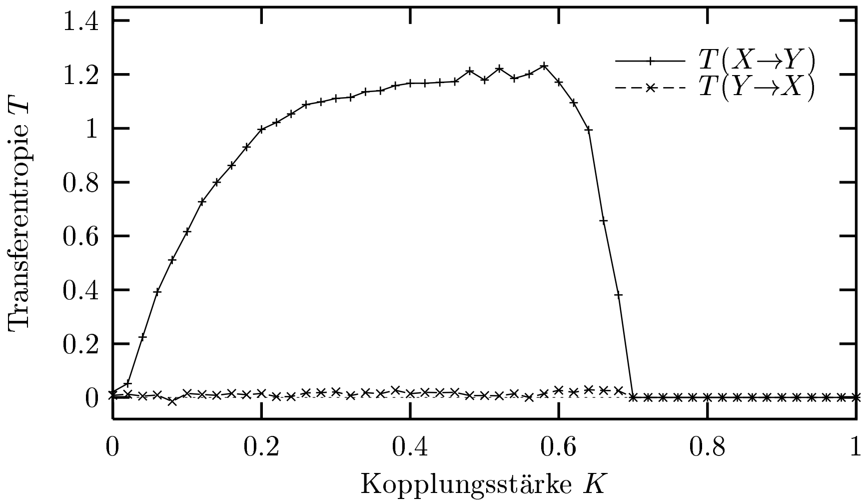

Abbildung 6.7: Transfer–Entropien zwischen Treiber X und Antwortsystem Y der gekoppelten Hénon–Abbildungen. Während $T(Y \to X) \approx 0$ ist, zeigt $T(X \to Y)$ einen starken Anstieg für $K > 0$.

Ordnung des angenommenen Markov–Modells in dem einen System, und den Parameter l, die Größe des Vergangenheitsfensters im anderen System, getroffen werden. Ich habe hier $k = 2$ gewählt, was durch den zweidimensionalen Phasenraum des Treibers (6.1) motiviert ist, und $l = 1$ gesetzt. Auch hier gilt, dass diese Wahl nicht entscheidend für die qualitative Aussage ist und eine andere Wahl lediglich kleine Unterschiede bewirken würde. Die entsprechenden Transfer–Entropien $T(Y \to X)$ und $T(X \to Y)$ sind in Abbildung 6.7 aufgetragen. Es fällt sofort die starke Asymmetrie ins Auge, wobei $T(Y \to X)$ unabhängig von K mit Null verträgliche Werte liefert. Dies stimmt mit der Vorstellung überein, dass das Antwortsystem Y keinen Einfluss auf das treibende System X ausübt. Auf der anderen Seite steigt $T(X \to Y)$ kontinuierlich mit der Kopplungsstärke bis ungefähr $K \approx 0.58$ an, um dann bei Einsetzen der identischen Synchronisation auf Null abzufallen. Für $Y = X$ liefert die Transfer–Entropie $T(X \to X) = 0$, da dann beide Systeme gleichzeitig die gesamte Information über das System enthalten und somit kein Informationsübertrag mehr stattfindet. Die Tatsache, dass $T(Y \to X) = 0$ sowohl für unabhängige als auch für identisch synchronisierte Systeme gilt, ist nicht weiter beunruhigend, da diese Fälle ohnehin sehr leicht über andere Maße unterschieden werden können. Alternativ könnte $U(Y \to X) = T(Y \to X) + M(X,Y)$ aus Gleichung (5.14) verwendet werden, welches für $Y = X$ die Entropie $U(X \to X) = M(X,X) = I(X)$ liefert.

Zusammenfassend lässt sich sagen, dass die Richtung der Kopplung von mehreren der untersuchten Abhängigkeitsmaße korrekt erkannt wurde. Dazu gehören die zeitverzögerten Kreuzkorrelationen und gegenseitigen Informationen einerseits und die Abhängigkeitsmaße $H(X|Y)$ und $T(X \to Y)$ andererseits. Dabei wird in diesem Fall die Unidirektionalität der Kopplung am eindrucksvollsten durch die Asymmetrie der Transfer–Entropie widergespiegelt. Das Eintreten der Synchronisation ist dagegen am leichtesten durch den starken Anstieg von H zu erkennen. In den folgenden Abschnitten werden weitere Systeme betrachtet, um weitere Eigenschaften der verschiedenen Maße vergleichen zu können.

6.1.2 Einfluss von Rauschen

Der hier betrachtete Fall der identischen Systeme (6.1) und (6.2) ist besonders einfach. Dies bedeutet aber nicht, dass er keinen Bezug zu physikalisch existierenden Systemen besitzt. In vielen Fällen können Teilsysteme als wenigstens annähernd identisch angesehen werden. Ein wesentlicher Unterschied zur Praxis ist allerdings die Reinheit der Zeitreihen, die hier betrachtet wurden, wohingegen bei gemessenen Zeitreihen immer ein mehr oder wenig starker Rauschanteil existiert. Im Folgenden soll deshalb der Einfluss von Messrauschen auf die untersuchten Abhängigkeitsmaße diskutiert werden. Dabei sollen aber nicht alle Abbildungen mit verrauschten Zeitreihen erneut wiedergegeben, sondern vielmehr die qualitativen Unterschiede zum rauschfreien Fall erörtert werden.

In [QAG00] wurde bereits der Einfluss von Messrauschen auf die Abhängigkeitsmaße S und H untersucht. Dabei wurden die beiden Fälle betrachtet, dass entweder das treibende *oder* das Antwortsystem verrauscht gemessen wurde. Ich will hier den etwas praxisorientierteren Fall diskutieren, in dem beide Zeitreihen X und Y gleichermaßen durch Rauschen verunreinigt sind. Der addierte Rauschanteil besteht dabei aus gaußverteilten Zufallszahlen mit einer Standardabweichung von $\sigma = 0.1$, was ungefähr 10% der Standardabweichung der betrachteten Zeitreihen entspricht.

Die geringsten Auswirkungen des addierten Rauschens machen sich bei den linearen Kreuzkorrelationen bemerkbar, welche in Abbildung 6.1 dargestellt sind. Dies spricht für die Robustheit dieser Observablen. Der einzige sichtbare Unterschied liegt in dem Verhalten für $K > 0.7$, wo die zeitverzögerten Kreuzkorrelationen $\rho_{XY}(\tau = +1)$ und $\rho_{XY}(\tau = -1)$ nicht mehr exakt übereinstimmen, sondern kleine Abweichungen voneinander zeigen. Diese offensichtliche Auswirkung des Rauschens ist aber selbstverständlich in allen besprochenen Abhängigkeitsmaßen zu finden, da die Zeitreihen X und Y für $K > 0.7$ nun nicht mehr exakt übereinstimmen.

Um das Verhalten der Abhängigkeitsmaße S und H zu verstehen, muss der Einfluss des Rauschens auf die einzelnen verwendeten Radien diskutiert werden. Die über alle Punkte gemittelten Abstände $R^{(N-1)}(\cdot)$ im Zähler von H (5.24) werden nur leicht größer, solange, wie hier der Fall, das Rauschniveau klein im Vergleich zur Standardabweichung der Zeitreihen ist. Sind die Systeme X und Y nur schwach abhängig voneinander, sind $R^{(k)}(Y|X)$ und $R^{(k)}(X|Y)$ von der gleichen Größenordnung wie die entsprechenden $R^{(N-1)}(\cdot)$, sodass sich auch ihr Wert nur leicht erhöht. Für Systeme mit starker Abhängigkeit hingegen, steigen die Abstände $R^{(k)}(\cdot|\cdot)$ bei Hinzufügen von Rauschen stark an, da dadurch die Abhängigkeit zerstört wird. $H(X|Y)$ und $H(Y|X)$ liefern also bei verrauschten Zeitreihen durchgehend kleinere Werte, wobei dieser Effekt mit steigendem Abhängigkeitsgrad zunimmt. Im vorliegenden Beispiel nimmt H für $K > 0.7$ ungefähr den Wert $H \approx 4.2$ an, was mit $H \approx 9.8$ in Abbildung 6.4 verglichen werden muss. Abgesehen von einer allgemeinen Senkung zeigt der Verlauf von H aber keine wesentlichen Unterschiede und auch die Asymmetrie $H(X|Y) > H(Y|X)$ bleibt für alle K erhalten. Wird allerdings, wie in [QAG00] gezeigt, nur eines der Systeme mit Rauschen verunreinigt, kann diese Asymmetrie aber durchaus invertiert werden. Dies ist aber nicht weiter verwunderlich, da dort auch das Rauschen sehr asymmetrisch addiert wurde.

Für das Verhalten von S ist die Auswirkung von Rauschen auf $R^{(k)}(\cdot)$ entscheidend. Da das Rauschen unkorreliert ist und sich damit bei der Einbettung in alle m Richtungen gleich verteilt, tendiert die Dimension für kleine Skalen gegen $D = m$. Sind beide Zeitreihen gleich eingebettet, bedeutet dies, dass nach Gleichung (5.19) $R^{(k)}(X)$ und $R^{(k)}(Y)$ gleich stark ansteigen. Da sich für unabhängige Systeme, wie oben erwähnt, $R^{(k)}(\cdot|\cdot)$

kaum ändert, ist damit auch ein Anstieg von $S^{(k)}(\cdot|\cdot)$ verbunden. Im vorliegenden Beispiel wurde für $K = 0$ der Wert $S(X|Y) \approx S(Y|X) \approx 0.002$ gefunden, was deutlich über dem Wert aus Abbildung 6.3 liegt. Interessant ist das Verhalten mit steigender Kopplungstärke. Schon für kleine K steigt $S(X|Y)$ schneller als $S(Y|X)$, sodass unabhängig von K die erwartete Asymmetrie $S(X|Y) > S(Y|X)$ gilt. Der „Dimensionseffekt" bei schwacher Kopplung wird also offensichtlich durch das Rauschen unterdrückt. Bei starker Abhängigkeit der Systeme steigt $R^{(k)}(\cdot|\cdot)$ unter dem Einfluss von Rauschen stärker an als $R^{(k)}(\cdot)$, was zu einer Senkung von S führt. Im vorliegenden Fall liegt das Plateau ab $K = 0.7$ bei ungefähr $S \approx 0.1$. Beide Effekte zusammengenommen bewirken eine wesentliche Abschwächung des Anstiegs bei Eintreten der identischen Synchronisation.

Der Einfluss von Rauschen auf die informationstheoretischen Größen hängt stark von dem Verhältnis der Rauschamplitude und der Größe des Nachbarschaftsradius r ab. Da im vorliegenden Beispiel das Rauschen eine Standardabweichung von $\sigma = 0.1$ aufweist, liegt der oben verwendete Nachbarschaftsradius $r = 0.05$ nicht mehr im Skalenbereich. Demzufolge messen die Einzelentropien $I(X)$ und $I(Y)$ die Shannon–Entropie des Rauschens, welche deutlich über der der reinen Zeitreihe liegt. Da ich hier nur an den qualitativen Auswirkungen und nicht den Beträgen der Größen interessiert bin, wurde im Folgenden $r = 0.15$ gesetzt, um eine sinnvolle Betrachtung zu ermöglichen. Außerdem wurden der Einfachheit halber die Zeitreihen nicht eingebettet und auch für die Transfer–Entropie $k = l = 1$ gewählt. Die so gewonnene, gegenseitige Information zeigt die gleiche K–Abhängigkeit wie in Abbildung 6.5 und auch die zeitverzögerten, gegenseitigen Informationen weisen die in Abbildung 6.6 dargestellte Asymmetrie auf. Diese ist nun für kleine Zeitverzögerungen τ am stärksten ausgeprägt, da die durch die Dynamik resultierenden Abhängigkeiten durch das Rauschen stark gesenkt werden. Auch die Asymmetrie $T(X \to Y) \geq T(X \to Y)$ der Transfer-Entropie bleibt für verrauschte Zeitreihen erhalten. Der wesentliche Unterschied besteht darin, dass im Fall der identischen Synchronisation bei $K > 0.7$ ein kleiner, aber endlicher Wert erreicht wird. Da der Determinismus durch das Rauschen zerstört wird, „verschmieren" die Übergangswahrscheinlichkeiten $p(i_{n+1}|i_n, \dots, i_{n-k+1})$ im Nenner von Gleichung (5.11) und die bedingte Entropie $I(X_{n+1}|X_n)$ in Gleichung (5.12) wird größer. Das gleiche gilt auch für die bedingte Entropie $I(X_{n+1}|X_n, Y_n)$. Allerdings geschieht dies in geringerem Maße, da $I(X_{n+1}|X_n, Y_n)$ die Zustände beider Systeme zum Zeitpunkt n einbezieht, welche sich nur durch einen geringen Rauschanteil unterscheiden. Diese Redundanz bewirkt, dass $I(X_{n+1}|X_n, Y_n) < I(X_{n+1}|X_n)$ gilt und die Transfer–Entropie T größer als Null wird.

6.1.3 Schwache Synchronisation

Das im vorangegangenen Abschnitt diskutierte System bestand aus zwei identischen Teilsystemen, die bei genügend starker Kopplung identisch synchronisieren. Die Tatsache, dass die Systeme bereits bei $K \approx 0.5$ schwach synchronisierten, ließ sich an keinem der vorgestellten Abhängigkeitsmaße eindeutig ablesen. Im Folgenden soll deshalb ein Beispiel studiert werden, das lediglich schwache, verallgemeinerte Synchronisation aufweist. Es soll untersucht werden, ob auch in diesem Fall das Eintreten der Synchronisation und die Richtung der Kopplung von den Maßen erkannt werden kann.

Das im Folgenden betrachtete System besteht aus zwei unidirektional gekoppelten, unterschiedlichen System und wurde bezüglich der Größen S und H bereits ausführlich in [QAG00] analysiert. Der autonome Treiber ist ein Rössler–System, das den Differential-

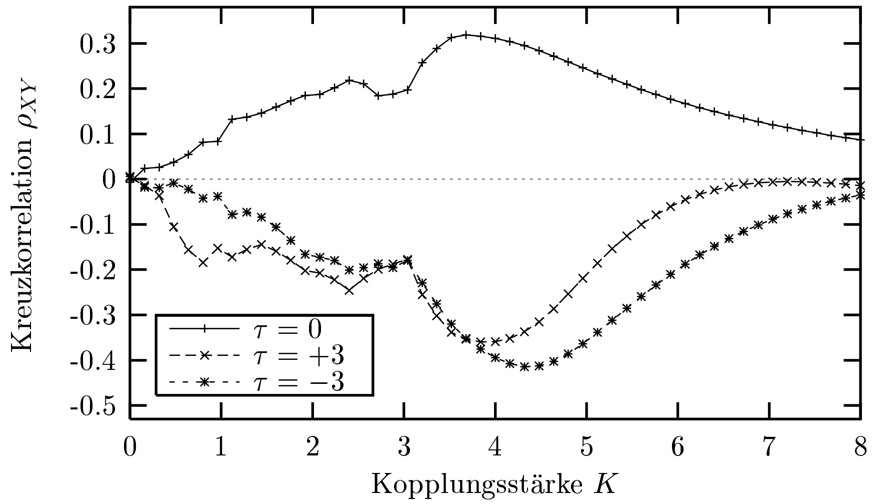

Abbildung 6.8: Kreuzkorrelationen zwischen dem Rössler–System (6.4) und dem daran gekoppelten Lorenz–System (6.5) in Abhängigkeit von K. Ein Vergleich mit Abbildung 6.1 zeigt wesentliche Unterschiede.

gleichungen

$$\begin{aligned} \dot{x}_1 &= -\alpha\,[x_2 + x_3] \\ \dot{x}_2 &= \alpha\,[x_1 + 0.2x_2] \\ \dot{x}_3 &= \alpha\,[0.2 + x_3(x_1 - 5.7)] \end{aligned} \tag{6.4}$$

gehorcht. Dieses sei über seine x_2–Komponente an ein Lorenz–System mit den Gleichungen

$$\begin{aligned} \dot{y}_1 &= \sigma(y_2 - y_1) \\ \dot{y}_2 &= r\,y_1 - y_2 - y_1 y_3 + K x_2^2 \\ \dot{y}_3 &= y_1 y_2 - b\,y_3 \end{aligned} \tag{6.5}$$

gekoppelt. Dabei wurde $\dot{y}_2$ um einen Term proportional zu x_2^2 erweitert, dessen Einfluss wieder über den Parameter K variiert werden kann. Für das Lorenz–System werden die üblichen Parameterwerte $\sigma = 10$, $r = 28$, $b = 8/3$ gewählt, für die sich das System chaotisch verhält. Über den Parameter α kann das Verhältnis der mittleren Perioden der beiden Systeme variiert werden. Ich verwende im Folgenden $\alpha = 10$, sodass beide Perioden ungefähr übereinstimmen. Die resultierenden Systeme synchronisieren sowohl im Bereich $2.1 < K < 2.7$ als auch für Kopplungsstärken $K > 2.9$. Der maximale Lyapunov–Exponent des Antwortsystems ist dabei stets größer als der kleinste Lyapunov–Exponent des Treibers, sodass es sich in beiden Fällen um schwache, verallgemeinerte Synchronisation handelt.

Die zu untersuchenden Zeitreihen wurden gewonnen, indem die Differentialgleichungen (6.4) und (6.5) mit einem Runge–Kutta–Verfahren vierter Ordnung integriert wurden. Die Integrationsschrittweite wurde dabei $\Delta t = 0.003$ gewählt. Alle 100 Schritte werden die Komponenten x_1 und y_1 der Systeme rausgeschrieben, wodurch ihr zeitlicher Abstand ungefähr einem Viertel der mittleren Periode der Systeme beträgt. Das transiente Verhalten wurde abgeschnitten und die Unabhängigkeit der Ergebnisse von den Anfangsbedingungen verifiziert. Die so gewonnenen Zeitreihen umfassen jeweils $N = 10000$ Werte.

Wie im vorangegangenen Abschnitt soll nun das Verhalten der einzelnen Abhängigkeitsmaße für den vorliegenden Fall untersucht werden. Dabei betrachte ich zunächst

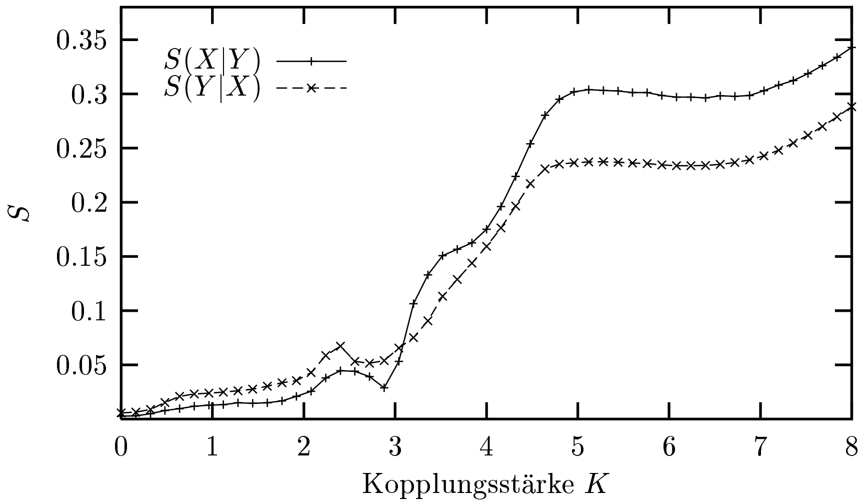

Abbildung 6.9: Abhängigkeitsmaße $S(X|Y)$ und $S(Y|X)$ für ein Lorenz–System Y, welches an ein autonomes Rössler–System X gekoppelt ist.

wieder die linearen Korrelationen in Abhängigkeit von der Kopplungsstärke. Die Kreuzkorrelationen (4.3) sind in Abbildung 6.8 für Zeitverzögerungen $\tau = -3, 0, 3$ aufgetragen. Eine erste interessante Beobachtung ist, dass $\rho_{XY}(\tau = 0)$ nicht wie in Abbildung 6.1 monoton mit der Kopplungsstärke steigt, sondern für $K \approx 3.7$ ein Maximum erreicht und danach wieder abfällt. Die Kreuzkorrelation kann in diesem Fall also nicht als Maß für die Stärke der Kopplung herangezogen werden. Ein weiteres lokales Maximum existiert bei $K \approx 2.4$, also in dem Fenster, in dem die Systeme zum ersten Mal synchronisieren. Der zweite auffällige Unterschied zu Abbildung 6.1 liegt in der Tatsache, dass die zeitverzögerten Kreuzkorrelationen $\rho_{XY}(\tau = \pm 3)$ hier keine von K unabhängige Asymmetrie aufweisen. Ein ähnliches Verhalten zeigen die Kreuzkorrelationen mit anderen τ–Werten. Die Richtung der Kopplung kann also nicht eindeutig an den zeitverzögerten Kreuzkorrelationen abgelesen werden. Im vorliegenden Fall scheint also offensichtlich eine Analyse der linearen Korrelationen allein nicht auszureichen, um die Natur der Kopplung zwischen den Systemen zu verstehen.

Im Gegensatz zu den linearen Korrelationen stimmt das Verhalten der Abhängigkeitsmaße S und H im vorliegenden Beispiel weitgehend mit dem aus Abschnitt 6.1.1 überein. Zur Berechnung wurden die Zeitreihen in einen $m = 5$ dimensionalen Raum eingebettet und die Abstände zu den jeweils $k = 20$ nächsten Nachbar gemessen. Wie in Abbildung 6.9 zu sehen, muss auch hier das Verhalten von S für kleine und große Kopplungsstärken differenziert betrachtet werden. Für schwache Kopplung gilt aufgrund der höheren Dimension des Antwortsystems $S(Y|X) > S(X|Y)$. Da die beiden beteiligten Systeme nun aber im Gegensatz zu den identischen Hénon–Abbildungen, bereits für $K = 0$ leicht unterschiedliche Dimension besitzen, gilt diese Ungleichung auch bei völliger Unabhängigkeit. Diese Tatsache macht die Interpretation der Asymmetrie von S als Hinweis auf die Richtung der Kopplung mehr als fraglich. Das erwartete Verhältnis zwischen $S(Y|X)$ und $S(X|Y)$ wird auch in diesem Beispiel nur bei starken Kopplungen ab ungefähr $K \approx 3$ angenommen. Das Eintreten der Synchronisation hingegen wird in beiden Fällen ($K \approx 2.4$ und $K > 2.9$) deutlich durch erhöhte Werte von S in beide Richtungen angezeigt.

Wesentlich einheitlicher als das Verhalten von S sieht das Verhalten von H in Ab-

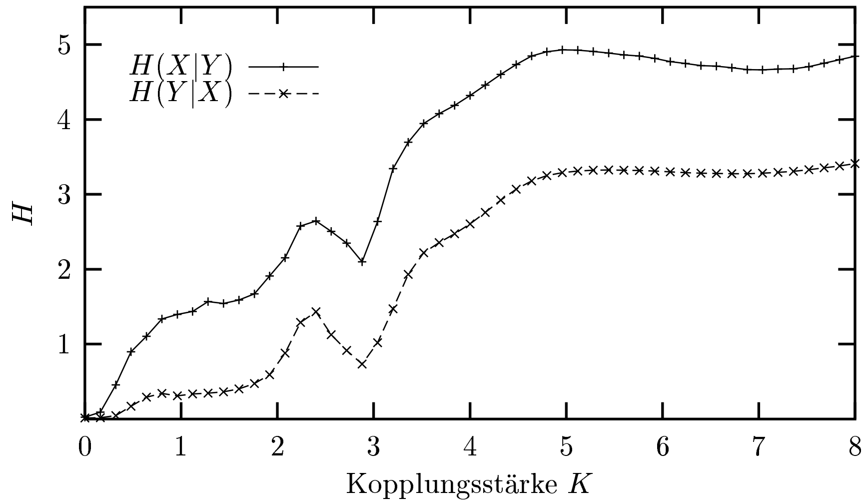

Abbildung 6.10: Abhängigkeitsmaße $H(X|Y)$ und $H(Y|X)$ für ein Lorenz–System Y, welches an ein autonomes Rössler–System X gekoppelt ist. Unabhängig von der Kopplungsstärke K gilt $H(X|Y) > H(Y|X)$.

bildung 6.10 aus. Während für unabhängige Systeme $H(X|Y) \approx H(Y|X) \approx 0$ gilt, ist für alle $K > 0$ stets die Asymmetrie aus Gleichung (6.3) zu beobachten. Daneben ist auch hier wieder deutlich das Eintreten der Synchronisation zu erkennen. Im ersten Bereich $2.1 < K < 2.7$, in dem schwache Synchronisation auftritt, haben sowohl $H(X|Y)$ als auch $H(Y|X)$ ein lokales Maximum und fallen danach deutlich ab. Ab $K \approx 2.9$ steigen dann wieder beide an, was das erneute Eintreten der Synchronisation anzeigt. An dem Abhängigkeitsmaß H lässt sich in den bisher untersuchten Beispielen also sowohl die Richtung der Kopplung als auch der Grad der Abhängigkeit ablesen.

Zur Anwendung der informationstheoretischen Abhängigkeitsmaße auf das vorliegende Beispiel werden die Zeitreihen nicht eingebettet, d. h. $m = 1$ gesetzt. Die Tatsache, dass dann die Werte x_1 und y_1 die Zustände nicht eindeutig festlegen, ist für das Folgende nicht von Bedeutung. Der Nachbarschaftsradius wird zu $r = 0.5$ gewählt, was ungefähr 7% der Standardabweichung der Zeitreihen entspricht. Die so gewonnenen gegenseitigen Informationen $M_{XY}(\tau)$ sind für $\tau = 0, \pm 2$ in Abbildung 6.11 aufgetragen. Alle drei Kurven zeigen ungefähr den gleichen Verlauf und lassen den Bereich der schwachen Synchronisation um $K \approx 2.4$ erkennen. Außerdem spiegelt die Asymmetrie der zeitverzögerten, gegenseitigen Informationen mit $\tau = \pm 2$ die Richtung der Kopplung wider. Die Ausprägung der Asymmetrie wäre für $\tau = \pm 1$ allerdings nicht so stark und die Kurven würden im Bereich $3 < K < 4$ sogar aufeinander fallen. Die Entscheidung über die Richtung der Kopplung anhand der zeitverzögerten, gegenseitigen Informationen ist hier also nicht mehr so einfach wie im Beispiel der gekoppelten Hénon–Abbildungen, wo die Asymmetrie unabhängig von τ war.

In Abbildung 6.12 sind die Transfer–Entropien zwischen dem Lorenz– und dem Rössler–System aufgetragen, wobei die Größen der Vergangenheitsfenster $k = l = 1$ gewählt wurden. Das Verhalten ist dabei dem von H in Abbildung 6.10 sehr ähnlich. Die Richtung der Kopplung kann auch hier an der deutlichen Asymmetrie $T(X \to Y) > T(Y \to X)$ abgelesen werden. Sie zeigt allerdings eine stärke Ausprägung als bei der zeitverzögerten, gegenseitigen Information in Abbildung 6.11. Weiter ist der Verlauf der Kurven im Bereich

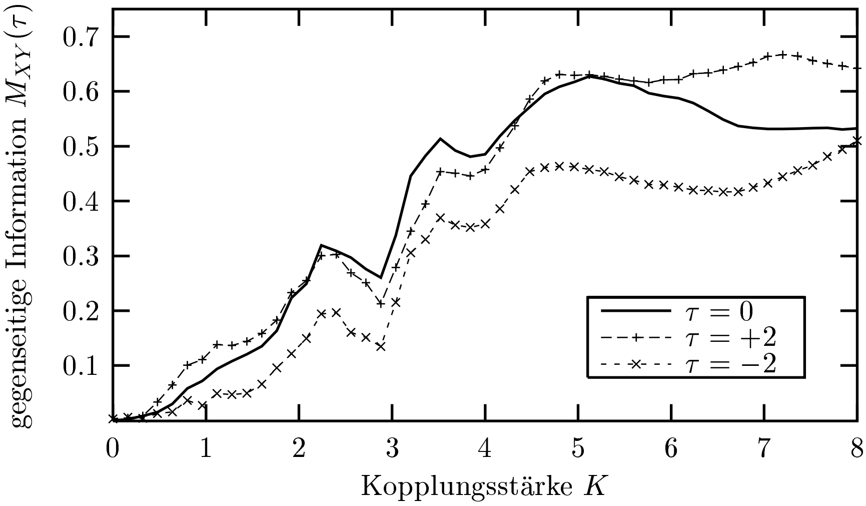

Abbildung 6.11: Gegenseitige Informationen zwischen einem autonomen Rössler–System X und einem daran gekoppeltem Lorenz–System Y für Zeitverzögerungen $\tau = 0, \pm 2$. Unabhängig von der Kopplungsstärke K gilt $M_{XY}(+2) > M_{XY}(-2) = M_{YX}(+2)$.

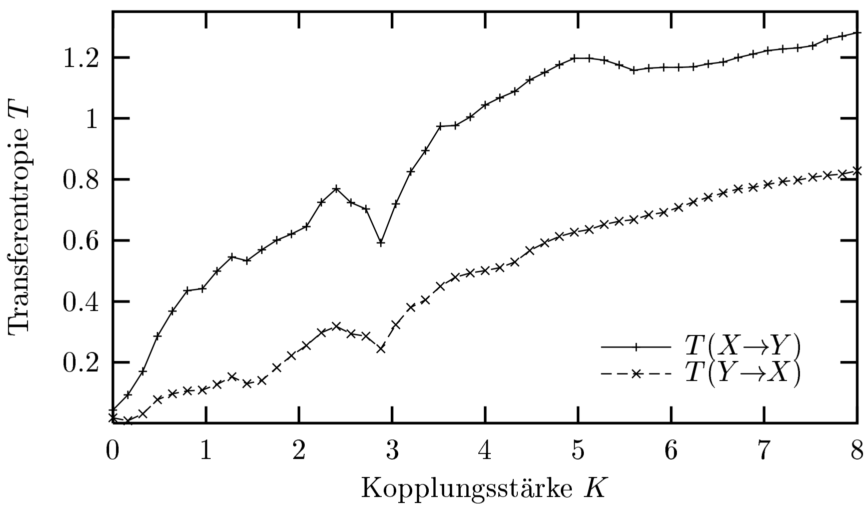

Abbildung 6.12: Transfer–Entropien $T(X \to Y)$ und $T(Y \to X)$ zwischen einem autonomen Rössler–System X und einem daran gekoppeltem Lorenz–System Y. Unabhängig von der Kopplungsstärke K gilt $T(X \to Y) > T(Y \to X)$.

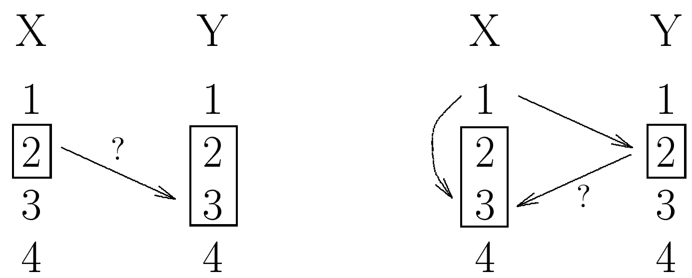

Abbildung 6.13: Veranschaulichung der Transfer–Entropien $T(X{\to}Y)$ (links) und $T(Y{\to}X)$ (rechts) zwischen Treiber X und Antwortsystem Y. Die umrandeten Zahlen kennzeichnen die bei $k = l = 1$ einbezogenen Zustände.

$2 < K < 3$ dem Verhalten der anderen Abhängigkeitsmaße sehr ähnlich. Das Eintreten der schwachen Synchronisation kann also auch über das Verhalten von Transfer–Entropie erkannt werden.

Ein wesentlicher Unterschied zu dem Beispiel der gekoppelten Hénon–Abbildungen ist die Tatsache, dass die Transfer–Entropie $T(Y{\to}X)$ vom Antwortsystem zum treibenden System hier nicht verschwindet. Dies hängt wieder mit der gemeinsamen Vergangenheit der Systeme X und Y zusammen, was bereits im Zusammenhang mit der zeitverzögerten, gegenseitigen Information in Abschnitt 6.1.1 diskutiert wurde. Der Sachverhalt ist in Abbildung 6.13 illustriert, in der die Zahlen die zeitlich aufeinander folgenden Zustände der beiden Systeme widerspiegeln. Links ist der Fall $T(X{\to}Y)$ und rechts $T(Y{\to}X)$ dargestellt, wobei die jeweils einbezogenen Zustände umrandet sind. Der durch die Transfer–Entropie gemessene Einfluss ist durch einen mit einem Fragezeichen gekennzeichneten Pfeil verdeutlicht. Auf der linken Seite stimmt dieser Pfeil mit der Richtung der Kopplung überein, sodass die Transfer–Entropie $T(X{\to}Y)$ groß ist. Die Transfer–Entropie $T(Y{\to}X)$ auf der rechten Seite misst den Einfluss des Zustandes Y_2 auf X_3. Direkt ist ein solcher Einfluss nicht vorhanden, da die Kopplung nicht in diese Richtung wirkt. Allerdings hängen sowohl X_3 (über die Dynamik) als auch Y_2 (über die Kopplung) vom Zustand X_1 ab. Da der Einfluss von X_1 auf X_3 von der Transfer–Entropie mit $k = 1$ nicht berücksichtigt wird, führt der indirekte Einfluss über Y_2 zu einem nicht verschwindenden $T(Y{\to}X) > 0$. Bei den gekoppelten Hénon–Abbildungen verschwand dieser Effekt, wenn dort ein Vergangenheitsfenster der Größe $k = 2$ verwendet wurde. Dies lag an der kurzen Autokorrelationslänge der Hénon–Abbildung. Im aktuellen Beispiel reichen die Korrelationen aber wesentlich weiter und durch eine Wahl von $k > 1$ kann der Effekt zwar abgeschwächt, aber nicht völlig abgeschaltet werden. Die sinnvolle Wahl von k ist ohnehin nach oben beschränkt. Dies liegt daran, dass bei einer Länge k des Vergangenheitsfensters eine $m = k+1$ dimensionale Einbettung der Zeitreihe X nötig ist. Da aber in höheren Dimensionen aufgrund der endlichen Länge der Zeitreihe die Dichte sehr gering wird, können Wahrscheinlichkeiten nur sehr schwer (über große Nachbarschaftsradien r) geschätzt werden. Ist man allerdings lediglich an der Asymmetrie der Transfer–Entropie interessiert, genügt in vielen Fällen $k = 1$. Neben einer Erhöhung der Robustheit hat diese Wahl auch zur Folge, dass die benötigte Rechenzeit gesenkt wird.

Werden die Resultate der beiden diskutierten Beispiele zusammengefasst, lässt sich sagen, dass das Abhängigkeitsmaß H, die Transfer–Entropie T und mit Abstrichen $M_{XY}(\tau)$ die gewünschte Information über die Kopplung liefern. An ihnen kann in beiden Fällen sowohl die Stärke als auch die Richtung der Kopplung korrekt abgelesen werden. In [QAG00] werden weitere Beispiele mit gekoppelten Systemen untersucht, die ebenfalls die

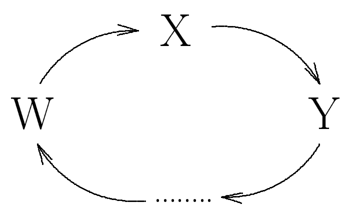

Abbildung 6.14: Geschlossener Ring aus einseitig gekoppelten Teilsystemen. Sind alle Teilsysteme identisch, besitzen sie trotz asymmetrischer Kopplung gleiche individuelle Eigenschaften.

Asymmetrie von H aus Gleichung (6.3) bestätigen.

6.2 Gekoppelte Abbildungen auf dem Gitter

Die vorangegangenen Abschnitte haben gezeigt, dass die Kapitel 5 eingeführten neue Maße zur Abhängigkeitsanalyse sehr nützliche Informationen über die Natur der Kopplung zwischen Systemen geben können. Speziell erkannten sie die in der unidirektionalen Kopplung innewohnende Asymmetrie des Gesamtsystems. Es wurde aber auch betont, dass das Antwortsystem bei eingeschalteter Kopplung als das „aktivere“ der beiden Systeme angesehen werden kann. Nach [AGLE99, QAG00] ist diese höhere Aktivität auf eine größere Anzahl angeregter Freiheitsgrade und damit größere Dimension zurückzuführen. Treiber und Antwortsystem unterscheiden sich also auch in ihren *individuellen* Eigenschaften. Dies gilt auch für die in Abschnitt 6.1.1 untersuchten, *identischen* Systeme, die nach Einschalten der Kopplung eben nicht mehr gleiche Dimension aufweisen. In den Beispielen wurde bereits deutlich, dass das Abhängigkeitsmaß S bei schwacher Kopplung auf einen solchen Unterschied in der Dimension sensitiv reagiert. Ein ähnlicher Einfluss von individuellen Eigenschaften der Systeme kann auch für die anderen Abhängigkeitsmaße nicht ausgeschlossen werden. Es stellt sich also die Frage, welche Aussagekraft der Asymmetrie $H(X|Y) > H(Y|X)$ von H zukommt, wenn gleichzeitig auch stets z. B. $I(Y) > I(X)$ für die Einzelentropien gilt.

Um den Einfluss von individuellen Eigenschaften ausschließen zu können, müssten identische Systeme betrachtet werden, die auch bei Einschalten der Kopplung identisch bleiben. Die Kopplung sollte aber nach wie vor asymmetrisch sein. Um dies zu bewerkstelligen, muss zunächst die Autonomie des Treibers aufgehoben werden und auch dieser an ein zusätzliches System W gekoppelt werden. Dies allein macht aber die Systeme X und Y noch nicht identisch, da beide von unterschiedlichen Systemen, nämlich W und X, getrieben werden. Die völlige Identität der Systeme ist erst gewährleistet, wenn zusätzlich das System Y das System W treibt. In diesem Fall haben wir es mit einem „Ring“ von Systemen zu tun, die in eine Richtung miteinander gekoppelt sind. Dieser Sachverhalt ist schematisch in Abbildung 6.14 dargestellt. Während X immer noch der Treiber von Y ist, weisen nun beide Systeme die gleichen individuellen Eigenschaften auf. Die Asymmetrie zwischen ihnen ist in einem solchen Ring aber nicht mehr perfekt, da Y über das System W auch auf X einen Einfluss ausübt. Dieser Einfluss kann allerdings gesenkt werden, indem ein Ring aus mehr als drei Systemen betrachtet wird. Im Folgenden soll das Verhalten der verschiedenen Abhängigkeitsmaße an einem solchen Ring–System untersucht werden.

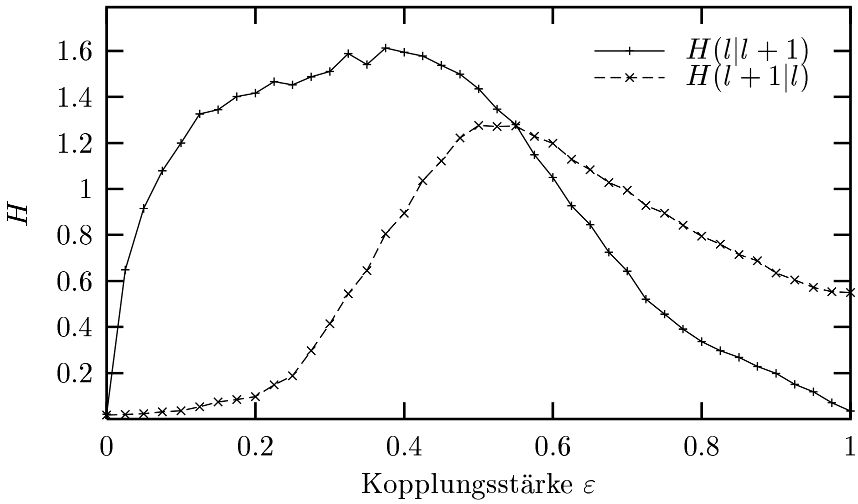

Abbildung 6.15: Abhängigkeitsmaß H zwischen zwei benachbarten Abbildungen l und $l+1$ auf dem Gitter. (Diskussion im Text)

6.2.1 Einseitig gekoppelte Abbildungen

Eine der einfachsten Formen den beschriebenen Ring aus Systemen zu realisieren, sind einseitig gekoppelte Abbildungen auf einem Gitter mit *periodischen* Randbedingungen. Ist L die Anzahl der Abbildungen und x_n^l der Zustand der Abbildung $l = 1, \ldots, L$ zum Zeitpunkt n, nimmt die einseitige Kopplung die Form

$$x_{n+1}^l = (1-\varepsilon)f(x_n^l) + \varepsilon f(x_n^{l-1}) \tag{6.6}$$

an. Dabei spiegelt $f(\cdot)$ die Dynamik der einzelnen Systeme wider, was im einfachsten Fall eindimensionale Abbildungen sind. Dies wird im Folgenden die Zelt–Abbildung

$$f(x) = 1 - |\,2x - 1| \tag{6.7}$$

auf dem Intervall $[0,1]$ sein. Über den Parameter ε kann, ähnlich dem Parameter K in den obigen Beispielen, die Stärke der Kopplung variiert werden. Die im Folgenden betrachteten Zeitreihen umfassen jeweils $N = 10000$ aufeinander folgende Werte von zwei benachbarten Gitterplätzen l und $l+1$. Die Größe des Gitters wurde $L = 50$ gewählt, sodass der indirekte Einfluss von Abbildung $l+1$ auf Abbildung l vernachlässigbar klein ist.

Zunächst soll wie in [Sch00a] das Verhalten des Abhängigkeitsmaßes H auf diesem System betrachtet werden. Dazu wurden die Zeitreihen im $m = 2$ Dimensionen eingebettet und die Abstände zu den jeweils $k = 20$ nächsten Nachbarn bestimmt. Die Abhängigkeitsmaße $H(l|l{+}1)$ und $H(l{+}1|l)$ zwischen zwei Gitterplätzen l und $l{+}1$ sind in Abbildung 6.15 in Abhängigkeit von der Kopplungsstärke ε aufgetragen. Drei interessante Eigenschaften sind zu beobachten:

1. Es gilt $H(l|l+1) \neq H(l+1|l)$, sodass die gefundene Asymmetrie (6.3) von H *nicht allein* auf Unterschieden in den individuellen Eigenschaften der beteiligten Systeme beruhen kann. Das Maß H vergleicht also nicht einfach die Eigenschaften der Systeme, welche hier ja gleich sind, sondern misst in der Tat die Beziehung *zwischen* den Systemen.

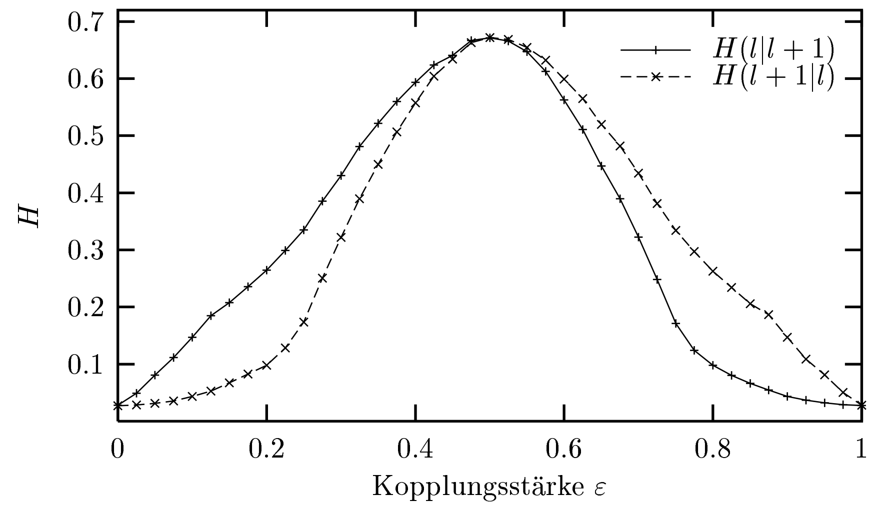

Abbildung 6.16: Abhängigkeitsmaß H zwischen zwei benachbarten Abbildungen l und $l+1$ auf dem Gitter. Im Gegensatz zu Abbildung 6.15 wurden die Zeitreihen diesmal nicht eingebettet.

2. Für $\varepsilon = 1$ ist $H(l{+}1|l) > 0$, obwohl die Abbildungen in diesem Fall völlig unabhängig voneinander sind. Wie in Gleichung (6.6) zu sehen, werden bei $\varepsilon = 1$ die Abbildungen lediglich iteriert und um einen Gitterplatz verschoben, ohne sich gegenseitig zu beeinflussen.

3. Die erwartete Asymmetrie $H(l|l+1) > H(l+1|l)$ gilt nur bis ungefähr $\varepsilon \approx 0.55$ und wird danach invertiert. Dies könnte eine direkte Folge von Punkt 2 sein.

Das in Punkt 2 beschriebene Verhalten bei $\varepsilon = 1$ ist dadurch zu verstehen, dass bei der Berechnung von H zwei–dimensionale Einbettungsvektoren ($m = 2$) verwendet wurden. Es wurde also die Beziehung zwischen den Vektoren

$$\begin{pmatrix} x_n^l \\ x_{n+1}^l \end{pmatrix} \searrow \begin{pmatrix} x_n^{l+1} \\ x_{n+1}^{l+1} \end{pmatrix} \tag{6.8}$$

untersucht. Für $\varepsilon = 1$ besteht die einzige Beziehung zwischen ihnen aus der durch den Pfeil gekennzeichneten funktionalen Abhängigkeit $x_{n+1}^{l+1} = f(x_n^l)$. Dies beschreibt gerade die angesprochene Iteration der Abbildung und die Verschiebung um einen Gitterplatz. Die anderen Komponenten der Vektoren sind unabhängig voneinander, sodass das Verhalten von H ausschließlich auf der Anwendung der Zelt–Abbildung (6.7) beruhen muss. Aufgrund der Stetigkeit der Abbildung $x^{l+1} = f(x^l)$, werden benachbarte Werte in x^l auf benachbarte Werte in x^{l+1} abgebildet. Die entsprechenden bedingten Abstände $R_n^{(k)}(l{+}1|l)$ sind also klein und damit ist $H(l{+}1|l)$ groß. Da die Zelt–Abbildung aber nicht eindeutig invertierbar ist, kann für $H(l|l+1)$ nicht die gleiche Argumentation angewandt werden. Zu jedem x^{l+1}–Wert existieren zwei x^l–Werte, deren Abstand von der gleichen Größenordnung wie die Ausdehnung des Gesamtsystems ist. Da also der Quotient aus $R_n^{(k)}(l|l{+}1)$ und $R_n^{(N-1)}(l)$ von der Ordnung eins ist, verschwindet $H(l|l{+}1)$. Dies erklärt das Verhalten von H in Abbildung 6.15 bei $\varepsilon = 1$. Dieser Effekt sollte verschwinden, wenn H direkt auf den skalaren Zeitreihen, d. h. ohne Einbettung, berechnet wird. In Abbildung 6.16 sind $H(l|l+1)$ und $H(l+1|l)$ mit $m = 1$ in Abhängigkeit von der Kopplungsstärke aufgetragen. Im Gegensatz zu Abbildung 6.15 wurde hier zusätzlich über die

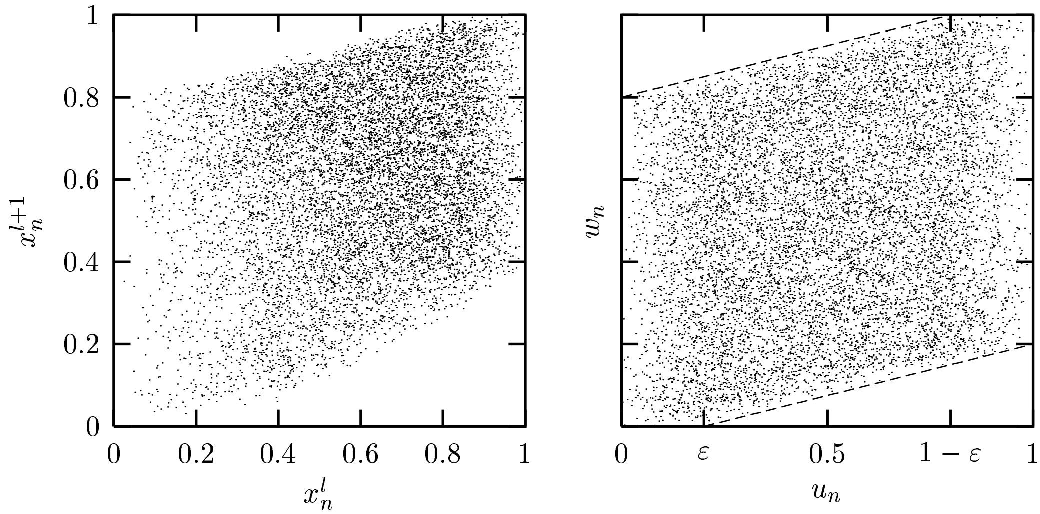

Abbildung 6.17: Realisierungen der einseitig gekoppelten Abbildungen aus Gleichung (6.6) (links) und den nach Gleichung (6.10) gemischten Zufallszahlen (rechts). In beiden Fällen wurde $\varepsilon = 0.2$ gewählt. Die Dichte der Punkte ist ein Maß für die zusammengesetzte Wahrscheinlichkeit $p(x^l, x^{l+1})$ bzw. $p(u, w)$.

$L = 50$ Gitterplätze gemittelt. Wie erwartet verschwindet bei $\varepsilon = 1$ nun sowohl $H(l|l+1)$ als auch $H(l+1|l)$.

In Übereinstimmung mit dem Fall mit $m = 2$ ist allerdings weiterhin die erwartete Asymmetrie $H(l|l+1) > H(l+1|l)$ verletzt. Für $\varepsilon > 0.5$ wird diese offensichtlich invertiert. Darüber hinaus legt die Form der Kurven in Abbildung 6.16 den Verdacht nahe, dass nun eine Beziehung der Art

$$H_\varepsilon(l|l+1) = H_{1-\varepsilon}(l+1|l) \tag{6.9}$$

gilt. Um dies zu verstehen, muss man sich vor Augen halten, dass sowohl x_{n+1}^l als auch x_{n+1}^{l+1} von dem Wert x_n^l und einem weiteren benachbarten Gitterplatz beeinflusst werden. Der Parameter ε bestimmt dabei die Stärke des Einflusses und spielt damit die Rolle eines Wichtungsfaktors für den Einfluss x_n^l. Wird die Wirkung der Abbildung $f(\cdot)$ vernachlässigt, kann dieser Sachverhalt durch eine einfache Mischung der Form

$$\begin{aligned} u_n &= \quad \varepsilon \; z_n^{(1)} + (1-\varepsilon) \; z_n^{(2)} \\ w_n &= (1-\varepsilon) \; z_n^{(3)} + \quad \varepsilon \; z_n^{(2)} \end{aligned} \tag{6.10}$$

modelliert werden. Dabei seien $\{z_n^{(l)}\}$, $l = 1, 2, 3$, unabhängige Zufallszahlen, die der gleichen Wahrscheinlichkeitsverteilung unterliegen. Wird für diese die Gleichverteilung $p(z) = 1$ im Intervall $[0, 1]$ gewählt, zeigt H für dieses einfache Modell in der Tat das gleiche ε–Verhalten wie für die gekoppelten Abbildungen. Würden also $H(U|W)$ und $H(W|U)$ in Abhängigkeit von ε aufgetragen, würde sich das gleiche Bild wie für $H(l|l+1)$ und $H(l+1|l)$ in Abbildung 6.16 ergeben. Auch die in Abbildung 6.17 dargestellten Realisierungen der gekoppelten Abbildungen (x_n^l, x_n^{l+1}) und der gemischten Zufallszahlen (u_n, w_n) sehen sich ähnlich. In beiden Fällen existieren Bereiche, in denen die zusammengesetzte Wahrscheinlichkeit, d.h. die Dichte der Punkte, Null ist.

Um die ε–Abhängigkeit von H zu verstehen, soll im Folgenden das Verhalten der einzelnen beteiligten Abstände diskutiert werden. Der Einfachheit halber konzentriert sich

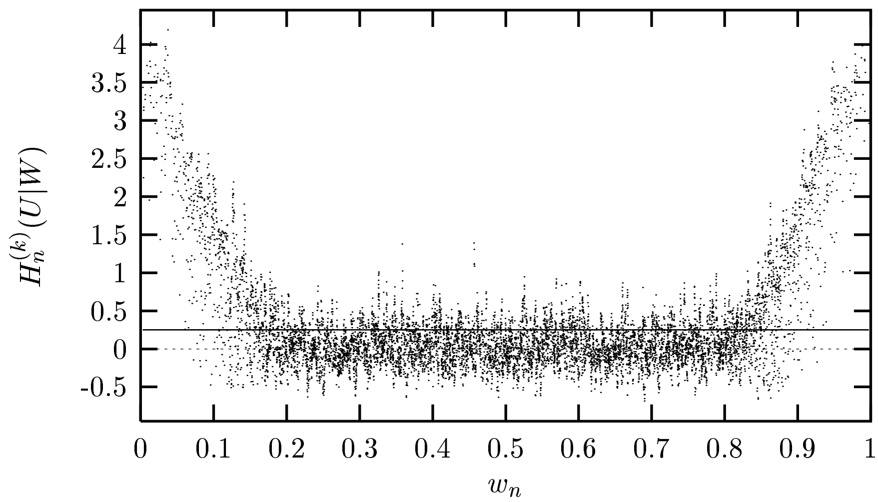

Abbildung 6.18: Der logarithmierte Quotient $H_n^{(k)}(U|W) = \log[R_n^{(N-1)}(U)/R_n^{(k)}(U|W)]$ in Abhängigkeit von w_n. Für $w_n < \varepsilon$ und $w_n > 1-\varepsilon$ sind die Werte deutlich größer als Null. Die durchgezogene, horizontale Linie ist der Mittelwert $H^{(k)}(U|W)$.

diese Diskussion auf die in Abbildung 6.17 rechts dargestellte Realisierung der gemischten Zufallszahlen. Die Argumentation lässt sich aber analog auch für die einseitig gekoppelten Abbildungen führen. In $H(U|W)$ fließen die mittleren quadrierten Abstände $R_n^{(N-1)}(U)$ und $R_n^{(k)}(U|W)$ ein. Sie messen Abstände vom Punkt (u_n, w_n) in der *horizontalen*, d. h. der u–Richtung. Während aber $R_n^{(N-1)}(U)$ den Mittelwert über *alle* Punkte bildet, mittelt $R_n^{(k)}(U|W)$ lediglich über k Punkte, die auf der w–Achse die nächsten Nachbarn von w_n sind und damit in einem dünnen horizontalen Band um w_n liegen. Liegt w_n im Intervall $[\varepsilon, 1-\varepsilon]$ überspannt dieses Band das gesamte u–Intervall $[0,1]$, sodass $R_n^{(k)}(U|W)$ und $R_n^{(N-1)}(U)$ annähernd gleich sind. Anders sieht dies aus, wenn $w_n < \varepsilon$ oder $w_n > 1-\varepsilon$ gilt. In diesem Fall sind die Punkte im horizontalen Band auf ein kleineres u–Intervall beschränkt und der bedingte Abstand $R_n^{(k)}(U|W)$ ist kleiner als $R_n^{(N-1)}(U)$. Da der in $H(U|W)$ verwendete Quotient damit zu Werten größer als Eins tendiert, erhält man im Mittel ein $H(U|W) > 0$. Zur Verdeutlichung dieses Sachverhalts ist in Abbildung 6.18 der logarithmierte Quotient $H_n^{(k)}(U|W)$ gegen w_n aufgetragen. Während dieser im Intervall $[\varepsilon, 1-\varepsilon]$ um Null streut, weist er außerhalb dieses Intervall deutlich größere Werte auf.

Völlig analog kann das Verhalten von $H(W|U)$ diskutiert werden. Die darin enthaltenen Größen $R_n^{(N-1)}(W)$ und $R_n^{(k)}(W|U)$ messen allerdings die Abstände vom Punkt (u_n, w_n) in der *vertikalen*, d. h. der w–Richtung. Der mittlere quadrierte Abstand $R_n^{(N-1)}(W)$ zu *allen* Punkten ist dabei im Mittel von der gleichen Größenordnung wie $R_n^{(N-1)}(U)$. Dies ist leicht einzusehen, wenn man sich vor Augen hält, dass beide Größen lediglich von der jeweiligen Wahrscheinlichkeitsverteilung $p(w)$ bzw. $p(u)$ abhängen. Aufgrund der symmetrischen Form der Mischung in Gleichung (6.10) sind diese aber offensichtlich identisch, d. h. es gilt $p(w) = p(u)$. Das Gleiche gilt selbstverständlich auch für die Wahrscheinlichkeitsverteilungen $p(x^l)$ und $p(x^{l+1})$ der gekoppelten Abbildungen, da diese sich aufgrund der periodischen Randbedingungen nicht unterscheiden.

Der bedingte Abstand $R_n^{(k)}(W|U)$, welcher über k Punkte in einem dünnen vertikalen Band um u_n mittelt, ist wie oben von der Größenordnung von $R_n^{(N-1)}(W)$ solange u_n

im Intervall $[\varepsilon, 1-\varepsilon]$ liegt. Wie in Abbildung 6.17 rechts zu sehen, wird $R_n^{(k)}(W|U)$ außerhalb dieses Intervall auch hier wieder kleiner, wobei dieser Effekt im Vergleich zu $R_n^{(k)}(U|W)$ wesentlich kleiner ist. Während $R_n^{(k)}(U|W)$ von $R_n^{(N-1)}(U)$ für kleine ε in führender Ordnung um $\mathcal{O}(\varepsilon)$ abweicht, ist die analoge Abweichung von $R_n^{(k)}(W|U)$ und $R_n^{(N-1)}(W)$ von der Ordnung $\mathcal{O}(\varepsilon^2)$. Daher gilt für kleine ε die Asymmetrie $H(U|W) > H(W|U)$ bzw. $H(l|l+1) > H(l+1|l)$, welche in Abbildung 6.16 zu sehen war.

Die obige Argumentation basiert auf der in Abbildung 6.17 dargestellten Realisierung mit $\varepsilon = 0.2$. Sie gilt also nur für kleine ε und verliert für $\varepsilon \geq 0.5$ ihre Gültigkeit. Die beobachtete Beziehung (6.9) und damit die Umkehrung der Asymmetrie ist aber leicht über die Mischungsgleichung (6.10) zu verstehen. Da die Zufallszahlen $z^{(1)}$ und $z^{(3)}$ der gleichen Wahrscheinlichkeitsverteilung unterliegen, entspricht dort eine Ersetzung von ε durch $1-\varepsilon$ einer Vertauschung von u und w. Die Beziehung zwischen $H(U|W)$ und $H(W|U)$ ist also für ε nahe bei Eins gerade umgekehrt zum Fall kleiner ε.

Das betrachtete Beispiel der einseitig gekoppelten Abbildungen hat einige neue Eigenschaften des Abhängigkeitsmaßes H offenbart. Zum einen wurde zufriedenstellenderweise gefunden, dass seine Asymmetrie (6.3) nicht allein auf einem Unterschied in den individuellen Eigenschaften der untersuchten Systeme beruht. Es beschreibt damit tatsächlich die Asymmetrie der wechselseitigen Abhängigkeiten zwischen den Systemen. Die Schlussfolgerung, dass es sich damit um eine Treiber–Antwortsystem–Beziehung zwischen den Systemen handelt, erlaubt diese Asymmetrie aber natürlich nicht. Im Beispiel der gemischten Zufallszahlen (6.10) war die gefundene Asymmetrie von $H(U|W)$ lediglich auf den unterschiedlichen Gehalt der Variablen $z^{(2)}$ in u und w zurückzuführen.

Weiter ist zu beachten, dass der Fall $m = 1$ hier nur der Einfachheit halber diskutiert wurde. Das System der gekoppelten Abbildungen weist dagegen eine sehr hohe Dimension auf. Das System wird also in einem niedrigdimensionalen Einbettungsraum nicht hinreichend beschrieben. Wie bei vielen anderen Größen in der nichtlinearen Zeitreihenanalyse, sollte auch bei H das Verhalten unter Variation der Einbettungsdimension m betrachtet werden, um aussagekräftige Resultate zu erhalten. Bei *schwacher* Kopplung kann im vorliegenden Beispiel allerdings bereits bei kleinem m die korrekte Richtung der Kopplung an der Asymmetrie von H abgelesen werden.

Nachdem das Verhalten des Abhängigkeitsmaßes H im Falle der einseitig gekoppelten Abbildungen ausgiebig studiert wurde, sollen im Folgenden die Eigenschaften der informationstheoretischen Größen an diesem Beispiel untersucht werden. Das Verhalten für *kleine* ε wurde für diesen Fall bereits in [Sch00b] analysiert. Bei einer groben Partition von nur zwei Boxen, konnte dort sowohl analytisch als auch numerisch eine ε^2–Abhängigkeit der Transfer–Entropie $T(l \to l+1)$ von Gitterplatz l zu Gitterplatz $l+1$ gefunden werden. Die entgegengesetzte Transfer–Entropie $T(l+1 \to l)$ und die zeitverzögerten, gegenseitigen Informationen in beide Richtungen verschwanden in diesem Fall. In Abbildung 6.19 ist das Verhalten der informationstheoretischen Größen für das gesamte ε–Intervall $[0,1]$ aufgetragen. Dabei wurde für die Transfer–Entropien $k = l = 1$ und bei den gegenseitigen Informationen $m = 1$ gewählt. Als Nachbarschaftsradius in der Korrelationssumme wurde $r = 0.05$ verwendet. Sowohl die Transfer–Entropie als auch die zeitverzögerten, gegenseitigen Informationen zeigen eine starke Asymmetrie. Während die Transfer–Entropie in der Richtung $l+1 \to l$ und die gegenseitige Information $M(\tau = -1)$ mit Null verträglich sind, zeigen $T(l \to l+1)$ und $M(\tau = +1)$ einen monotonen Anstieg mit der Kopplungsstärke ε. Dabei steigt die Transfer–Entropie für kleine ε deutlich stärker an als $M(\tau = +1)$. Für $\varepsilon = 1$ ergibt sich wieder die in Gleichung (6.8) dargestellte Situa-

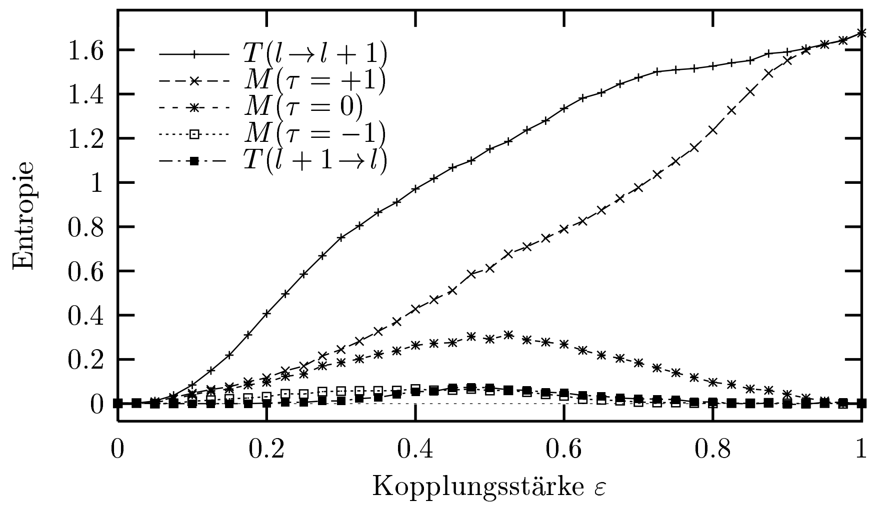

Abbildung 6.19: Transfer–Entropien T und gegenseitige Informationen M mit $\tau = 0, \pm 1$ zwischen zwei benachbarten Gitterplätzen l und $l+1$ der einseitig gekoppelten Abbildungen (6.6).

tion, in der $X_n^{(l+1)}$ sowohl von $X_n^{(l)}$ als auch von $X_{n+1}^{(l+1)}$ unabhängig ist und damit deren gegenseitige Informationen verschwinden. Die Transfer–Entropie (5.13) liefert damit $T(l \to l+1) = M(X_{n+1}^{(l+1)}, X_n^{(l+1)}, X_n^{(l)}) = M(X_{n+1}^{(l+1)}, X_n^{(l)})$, sodass sie, wie in Abbildung 6.19 zu sehen, für $\varepsilon = 1$ mit der gegenseitigen Information $M(\tau = +1)$ übereinstimmt. Die gegenseitige Information $M(\tau = 0)$ ohne Zeitverzögerung ist symmetrisch um $\varepsilon = 0.5$ und verschwindet für $\varepsilon = 0$ und $\varepsilon = 1$, wo die benachbarten Gitterplätze unabhängig voneinander sind. Ihr Verlauf sieht damit dem Verlauf von H mit $m = 1$ in Abbildung 6.16 sehr ähnlich.

Die Analogie zu dem System der gemischten Zufallszahlen (6.10) hält im Gegensatz zum Abhängigkeitsmaß H nicht für die hier verwendeten informationstheoretischen Größen. Dies liegt daran, dass sie im Sinne von Abschnitt 4.1.1 *dynamische* Maße sind, welche im Gegensatz zu statischen Maßen nicht allein von der zusammengesetzten Wahrscheinlichkeit $p(x_n, y_n)$ zu *gleichen* Zeiten n abhängen. Da die Zufallszahlen $z_n^{(l)}$ und damit die Variabeln u_n bzw. w_n unabhängig sind und damit keine Dynamik aufweisen, verschwinden dynamische Maße wie die Transfer–Entropie und die zeitverzögerte, gegenseitige Information für das System (6.10). Statische Maße wie H oder die gegenseitige Information *ohne* Zeitverzögerung liefern dagegen positive Werte. In der Tat stimmt der Verlauf der gegenseitigen Information $M_{UW}(\tau = 0)$ mit dem in Abbildung 6.19 gezeigten überein.

Natürlich lässt sich auch bei dem Abhängigkeitsmaß H eine Zeitverzögerung τ zwischen den beiden Zeitreihen einführen. Dadurch gewinnt man im Gegensatz zu der gegenseitigen Information eine *zusätzliche* Asymmetrie, welche im Folgenden mit *Zeit*asymmetrie bezeichnet wird. Zu einem festen $|\tau|$ existieren damit vier verschiedene Maße. Für das Beispiel der einseitig gekoppelten Abbildungen sind diese in Abbildung 6.20 in Abhängigkeit von ε aufgetragen, wobei $m = 1$ verwendet wurde. Sowohl $H(l|l+1)$ als auch $H(l+1|l)$ zeigen die erwartete Zeitasymmetrie und liefern unabhängig von ε für $\tau = 1$ größere Werte als für $\tau = -1$. Dies stimmt mit der Asymmetrie der informationstheoretischen Größen in Abbildung 6.19 überein. Die H–Achse wurde logarithmisch gewählt, da $H(l+1|l)$ mit $\tau = 1$ für $\varepsilon \to 1$ sehr große Werte liefert. Dies liegt wie im Fall der zweidimensionalen

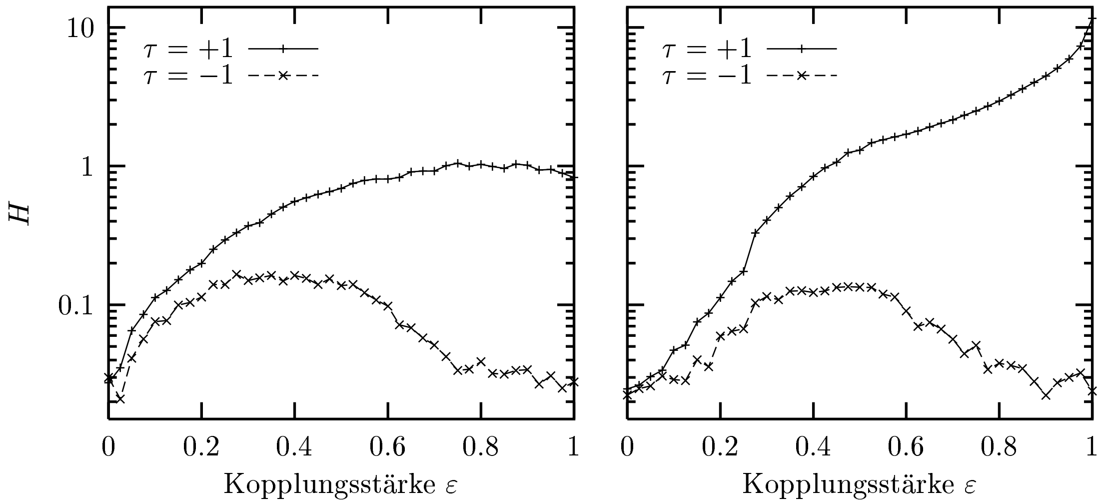

Abbildung 6.20: Zeitverzögerte Abhängigkeitsmaße $H(l|l+1)$ (links) und $H(l+1|l)$ (rechts) mit $\tau = \pm 1$ zwischen benachbarten Abbildungen auf dem Gitter. Die Zeitasymmetrie stimmt in beiden Fällen mit der der informationstheoretischen Größen in Abbildung 6.19 überein.

Einbettung (6.8) an der Beziehung $x^{l+1} = f(x^l)$ für $\varepsilon = 1$ und der Nichtinvertierbarkeit der Abbildung $f(\cdot)$. Ein gemeinsames Auftragen von $H(l|l+1)$ und $H(l+1|l)$ bei $\tau = +1$ würde dementsprechend eine Karikatur der Situation in Abbildung 6.15 liefern, in welcher $m = 2$ verwendet wurde.

6.3 Anwendung auf gemessene Zeitreihen

Nachdem in den vorangegangenen Abschnitten die Funktionstüchtigkeit und Leistungsfähigkeit der neuen Abhängigkeitsmaße an unterschiedlichsten Modellsystemen untersucht wurde, sollen sie im Folgenden auf *gemessene* Zeitreihen angewendet werden. Der wesentliche Unterschied besteht dabei darin, dass bei gemessenen Daten *a priori* häufig keine Information über die zugrunde liegende Dynamik oder die Art der Kopplung vorhanden ist. Weiter ist es in vielen Fällen auch nicht möglich, das beobachtete System kontrolliert zu beeinflussen. Es fehlt also ein Kontrollparameter, dessen Variation Aufschluss über die Natur des zugrunde liegenden Prozesses geben könnte. Die Herangehensweise an einen Messdatensatz muss diesen Bedingungen angepasst sein und wird in vielen Fällen neue Probleme aufwerfen. Bei den folgenden Analysen von gemessenen Zeitreihen werden einige dieser Probleme aufgezeigt und diskutiert. Die untersuchten Daten sind dabei unterschiedlichsten, wissenschaftlichen Bereichen wie Ökonomie, Medizin und Klimaforschung entliehen. Eine Anwendung des Abhängigkeitsmaßes S (5.23) auf EEG–Daten lässt sich in [AGLE99] finden.

6.3.1 Finanzdaten

Die Betrachtung von Daten aus der Ökonomie zeichnet sich in der Zeitreihenanalyse durch anhaltende Popularität aus und bekam mit dem Aufkommen der nichtlinearen Zeitreihenanalyse neuen Auftrieb. Dies liegt daran, dass neue Erkenntnisse auf diesem Gebiet unter Umständen gewinnbringend ausgenutzt werden könnten. Dies gilt im Speziellen für die

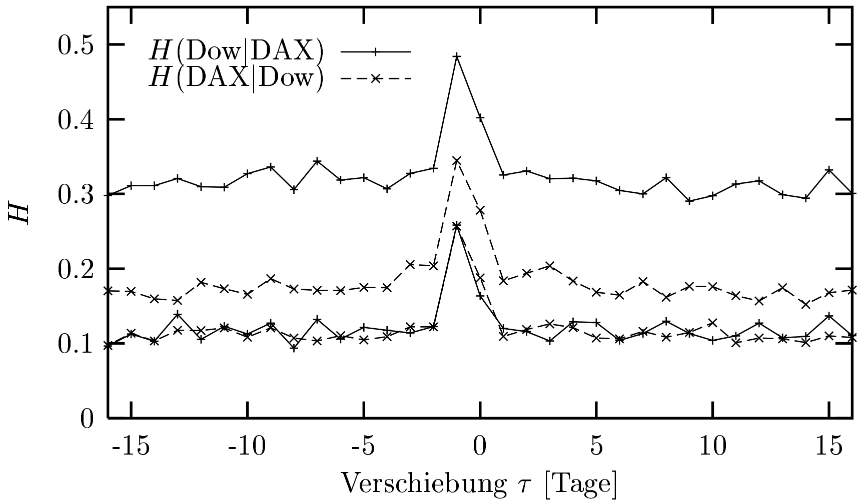

Abbildung 6.21: Zeitverzögertes Abhängigkeitsmaß H zwischen Dow Jones und DAX–Zeitreihe. Während für die beiden oberen Kurven die gesamten Zeitreihen verwendet wurden, wurden für die unteren Kurven Ausnahmeereignisse wie Crashs und Rallyes aus den Zeitreihen entfernt.

Entwicklungen auf dem Wertpapiermarkt. Neben dem offensichtlichen Nutzen von Vorhersagen von Kursen kann das Auffinden jeglicher Struktur einen Vorteil im Handel bringen. Dazu zählen natürlicherweise auch Abhängigkeiten zwischen verschiedenen Kursen.

Die wohl wichtigsten Indikatoren für die Entwicklung an den verschiedenen Börsenplätzen sind Aktienindizes, in welchen die Werte von mehreren Aktien zusammengefasst sind. Für die USA ist dies der *Dow Jones* Index und für Deutschland der *Deutsche Aktien Index* DAX. Im Folgenden soll die Abhängigkeit zwischen ihnen analysiert werden. Die verwendeten Zeitreihen bestehen dabei aus täglichen Notierungen der Indizes von Mai 1985 bis Mai 1998 und enthalten damit etwas mehr als 3000 Werte. Die Daten wurden freundlicherweise von Thomas Schürmann von der WGZ–Bank in Düsseldorf zur Verfügung gestellt. Da das Verhalten beider Indizes auf langen Zeitskalen als exponentieller Anstieg beschrieben werden kann, werden für gewöhnlich die Tagesdifferenzen der *logarithmierten* Notierungen betrachtet. Aus den resultierenden Zeitreihen wurden anschließend alle Feiertage herausgenommen. Diese sind sehr einfach mit den Nullen in den Tagesdifferenzen zu identifizieren und könnten eine Abhängigkeitsanalyse verfälschen, da viele Feiertage in den USA und Deutschland aufeinanderfallen.

Die einfache Berechnung von $H(\mathrm{DAX}|\mathrm{Dow})$ und $H(\mathrm{Dow}|\mathrm{DAX})$ allein würde für die Untersuchung der Abhängigkeit zwischen den beiden Indizes wenig Aufschluss geben. Die Signifikanz der Differenz und Größe der beiden gewonnenen Werte bliebe unklar. Dasselbe würde selbstverständlich auch bei einem Modellsystem gelten, bei dem z. B. H nur für *eine* feste Kopplungsstärke gemessen würde. Erst die *Variation* der Kopplungsstärke erlaubte einen Vergleich und eine Interpretation der Resultate. Im vorliegenden Fall existiert eine solche Freiheit nicht. Bei den Indizes handelt es sich um „Messungen" an Systemen, auf die wir keinen Einfluss haben. Zusätzlich ist hier auch keine Wiederholung der Messung möglich, sodass auch darüber keine Vergleichsmöglichkeiten bestehen.

Als Alternative können in einem solchen Fall die Parameter des Abhängigkeitsmaßes selbst variiert werden. Sinnvoll ist z. B. die Betrachtung der zeitverzögerten Versionen der Maße in Abhängigkeit von der relativen Zeitverschiebung τ. In den meisten Fällen erwartet man für große τ eine Unabhängigkeit zwischen den Systemen, sodass der Wert des Abhängigkeitsmaßes bei großen τ als Vergleichwert dienen kann. Dementsprechend sind in Abbildung 6.21 das zeitverzögerte Maß H zwischen DAX und Dow Jones in Abhängigkeit von der Zeitverschiebung τ aufgetragen. Es ist deutlich zu sehen, dass H für große $|\tau|$ einen (abgesehen von Fluktuationen) konstanten Wert annimmt. Bei der Berechnung von H wurden dabei die Abstände zu den jeweils $k = 15$ nächsten Nachbarn bestimmt. Die qualitativen Ergebnisse hingen dabei nicht von der Wahl von k oder der Einbettungsdimension m ab und in Abbildung 6.21 wurde $m = 1$ verwendet. Betrachtet man zunächst nur die beiden oberen Kurven, kann zweierlei beobachtet werden. Zum einen heben sich nur die Werte für $\tau = 0$ und $\tau = -1$ vom konstanten Untergrund ab und zum anderen gilt unabhängig von τ die Ungleichung $H(\text{Dow}|\text{DAX}) > H(\text{DAX}|\text{Dow})$. Letzteres ist die Beziehung, welche man erwarten würde, wenn der Dow Jones der Treiber des DAX wäre. Eine solche Interpretation wäre aber voreilig. So gilt die gefundene Ungleichung auch für große $|\tau|$, obwohl in diesem Bereich die Zeitreihen unabhängig voneinander sein sollten und damit auch keine Asymmetrie mehr aufweisen sollten. Die systematische Abweichung deutet auf einen grundlegenden Unterschied zwischen den beiden Zeitreihen hin. Tatsächlich fallen bei der Betrachtung der Tagesdifferenzen in beiden Zeitreihen starke Ausreißer auf. Diese entsprechen den Tagen, an denen es zu Börsen–Crashs oder besonders starken Anstiegen kam. Weiter fällt auf, dass diese Ausreißer im Dow Jones teilweise wesentlich stärker ausgeprägt sind und einmal sogar über 20 Standardabweichungen vom Mittelwert abweichen. Da in das Maß H (5.24) der mittlere quadrierte Abstand zu *allen* Punkten einfließt, reagiert es sensitiv auf solche Ausreißer und liefert größere Werte. Um diese Abhängigkeit zu verifizieren wurden in beiden Zeitreihen die Tage entfernt, an denen mindestens in einem Index die Tagesdifferenzen mehr als fünf Standardabweichungen vom Mittelwert abweichen. Aufgrund dieses Kriteriums wurden insgesamt lediglich 13 von ungefähr 3000 Tagen herausgenommen. Die anschließend berechneten $H(\text{Dow}|\text{DAX})$ und $H(\text{DAX}|\text{Dow})$ sind die beiden unteren Kurven in Abbildung 6.21. Sie weisen nun keine von der Zeitverschiebung τ unabhängige Asymmetrie mehr auf und liefern für große τ wie erwartet vergleichbare Werte. Es sei betont, dass dieses Resultat *nicht* bedeutet, dass die DAX–Crashs von den Crashs des Dow Jones abhängen. Die mit Crashs gefundene Asymmetrie war lediglich eine Auswirkung der unterschiedlichen Ausprägung der Crashs in den Zeitreihen. Dies zeigt, dass die Stationarität der betrachteten Zeitreihen eine wichtige Voraussetzung für die Anwendung des Abhängigkeitsmaßes H ist.

Nachdem also H in diesem Fall keine Aussage über die Asymmetrie einer eventuellen Kopplung machen kann, bleibt als einzige signifikante Beobachtung das Verhalten bei $\tau = 0$ und $\tau = -1$. Dabei spiegelt der erhöhte Wert bei $\tau = 0$ natürlich nur die Tatsache wider, dass die beiden Aktienmärkte im Laufe eines Tages sich gegenseitig beobachten und damit auch voneinander abhängen. Die Differenz der beiden Werte bei $\tau = 0$ kann kaum als eine signifikante Abweichung betrachtet werden, da sie von der gleichen Größenordnung wie Differenzen bei großen τ ist. Auch das Verhalten bei $\tau = -1$ kann relativ leicht verstanden werden. Aufgrund der Zeitverschiebung zwischen Deutschland und den USA, läuft die Börse in New York noch einige Stunden weiter, wenn sie in Frankfurt am Abend schließt. Der erhöhte Wert bei $\tau = -1$ sagt aus, dass die Entwicklung des Dow Jones am folgenden Tag vom DAX nachvollzogen wird und insofern eine Abhängigkeit existiert. Die Betrachtung des Abhängigkeitsmaßes H bestätigt also lediglich die erwarteten Beziehungen, die z. B. auch an den linearen Kreuzkorrelationen beobachtet werden können, und es konnten keine grundlegend neuen Erkenntnisse gewonnen werden.

Abschließend soll die Anwendung der Transfer–Entropie auf dieses Beispiel studiert

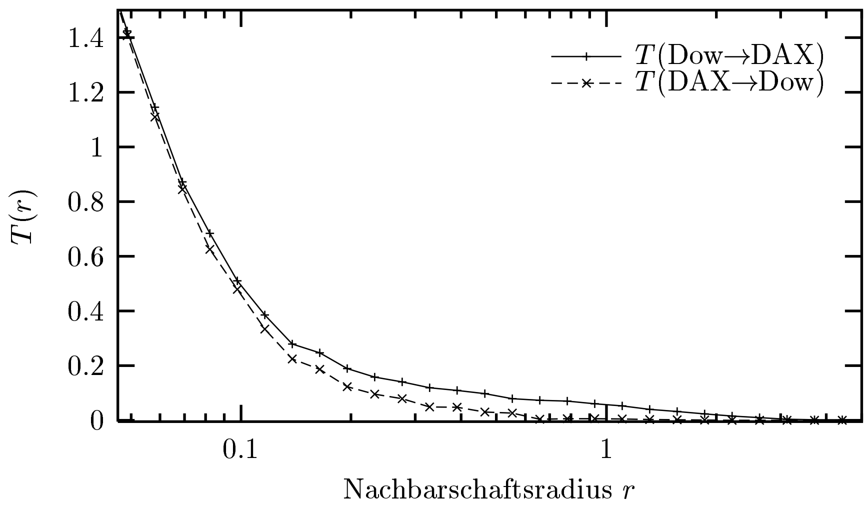

Abbildung 6.22: Transfer–Entropie T zwischen Dow Jones und DAX–Zeitreihe in Abhängigkeit vom Nachbarschaftsradius r. Im gezeigten Bereich gilt $T(\mathrm{Dow}{\rightarrow}\mathrm{DAX}) > T(\mathrm{DAX}{\rightarrow}\mathrm{Dow})$ unabhängig von r. Außerhalb dieses Bereiches zeigen sie keine signifikanten Abweichungen voneinander.

werden. Wie im Fall des Abhängigkeitmaßes H muss auch hier wieder ein Parameter variiert werden, um einen Anhaltspunkt für die Signifikanz der Resultate zu erhalten. Eine naheliegende Wahl ist hier die Betrachtung unter Variation des Nachbarschaftsradius r in der Korrelationssumme, während $k = l = 1$ fixiert werden. In Abbildung 6.22 sind die Transfer–Entropien $T(\mathrm{DAX}{\rightarrow}\mathrm{Dow})$ und $T(\mathrm{Dow}{\rightarrow}\mathrm{DAX})$ in Abhängigkeit von r aufgetragen. Beide Zeitreihen wurden vorher auf eine Standardabweichung von $\sigma = 1$ umskaliert, sodass das Skalenverhalten der beiden Transfer–Entropien direkt miteinander verglichen werden kann. Im gezeigten Bereich zeigen sie eine systematische, von r unabhängige Abweichung. Dies deutet darauf hin, dass der Informationsfluss vom Dow Jones zum DAX stärker ist als umgekehrt. Dies stimmt mit dem beobachteten Verhalten von H überein, wo die Abhängigkeit bei $\tau = -1$ größer war als bei $\tau = 1$. Die in Abbildung 6.22 gefundene Asymmetrie allein bringt also keine neue Information über das Abhängigkeitsverhältnis der Indizes. Weitere Untersuchungen mit größerem Vergangenheitsbereich k und l in der Transfer–Entropie oder relative Zeitverschiebungen zwischen den Zeitreihen zeigten keine eindeutigen Asymmetrien und brachten damit ebenfalls keine neuen Erkenntnisse.

Zusammenfassend lässt sich sagen, dass sowohl die Transfer–Entropie als auch das Abhängigkeitsmaß H lediglich die erwarteten Resultate bestätigt haben. Darüber hinaus konnten keine neuen Erkenntnisse gewonnen werden. Dies ist nicht weiter verwunderlich, da es sich bei den betrachteten Indizes um eine extreme Reduktion der komplizierten Entwicklungen an den Finanzmärkten auf einzelne Zahlen handelt. Da selbst Autokorrelationen innerhalb der Zeitreihen nicht signifikant nachgewiesen werden können, ist es nicht erstaunlich, dass auf Zeitverzögerung basierende Abhängigkeitsmaße keine entscheidenden Resultate liefern.

6.3.2 Physiologische Daten

Auch in der medizinischen Forschung werden in zunehmendem Maße Methoden der nichtlinearen Zeitreihenanalyse angewendet. Damit ist die Hoffnung verbunden, die komplexen Vorgänge, die z. B. im menschlichen Körper stattfinden, besser verstehen zu können. Im

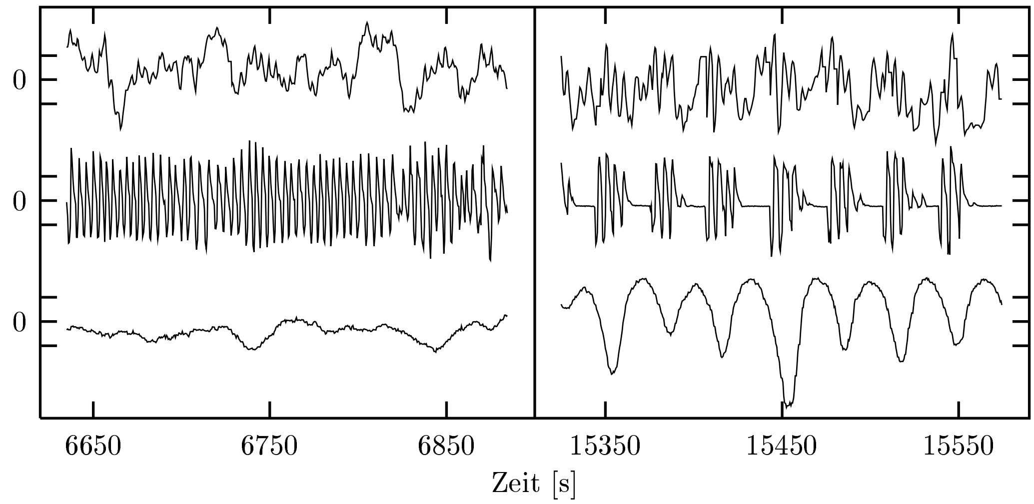

Abbildung 6.23: Im Schlaf aufgenommene Herzrate (oben), Atmung (Mitte) und Blut–Sauerstoff–Konzentration (unten) einer Versuchsperson zu zwei unterschiedlichen Zeitpunkten. Während die Person im rechts dargestellten Zeitraum wach ist, befindet sie sich im linken Zeitraum in Schlafphase zwei.

Folgenden sollen physiologische Daten betrachtet werden, die als Datensatz B im Rahmen eines Zeitreihenwettbewerbs des Santa Fe Institutes veröffentlicht wurden [WG93]. Es handelt sich dabei um gleichzeitig aufgenommene Herz–, Atem– und Blutwert–Daten einer Versuchsperson im Schlaf und umspannt insgesamt einen Zeitraum von fast fünf Stunden. Bereits ein erster Blick auf die Zeitreihen offenbart, dass sie nicht als stationär angenommen werden können. Dies liegt zum einen an den experimentellen Bedingungen. Durch zwischenzeitliches Lösen der Sensoren und langsamen Drifts in den Zeitreihen mussten in regelmäßigen Abständen Rekalibrierung durchgeführt werden. Zum anderen zeigen die Zeitreihen zu unterschiedlichen Zeiten völlig unterschiedliche Strukturen, die schon mit dem Auge unterschieden werden können. Zur folgenden Untersuchung wurden ausschließlich Teilstücke herangezogen, die visuell als annähernd stationär angesehen werden können und keine Drifts oder Sprünge aufweisen. Zusätzlich wurde darauf geachtet, dass im Verlauf der Teilstücke kein Schlafphasenwechsel stattfindet. Die jeweiligen Schlafphasen wurden anhand eines gleichzeitig aufgenommenen EEGs von einem Neurologen bestimmt und notiert. Es wurden dabei die vier Zustände wach, Schlafphase eins, Schlafphase zwei und REM–Schlaf unterschieden. Die REM–Schlafphase ist nur relativ kurz und die entsprechenden Signale nicht stationär, sodass dieser Teil im Folgenden nicht betrachtet wurde. Die resultierenden acht Zeitreihen umspannen zwischen 1000 und 3000 Werte.

In Abbildung 6.23 sind zwei, 250 Sekunden lange Teilstücke der Daten zu unterschiedlichen Zeitpunkten aufgetragen. Bei der oberen Kurve handelt es sich um die Herzrate, welche über Inversion der Abstände zwischen den Herzschlägen und anschließender Interpolation auf gleichmäßige Zeitabstände von 0.5 Sekunden gewonnen wurde. Die mittlere Zeitreihe spiegelt das Lungenvolumen und damit die Atmung der Versuchsperson wider und wurde mit 2 Hz aufgenommen. Schließlich gibt die untere Kurve die Sauerstoff–Konzentration im Blut des Patienten wieder. Da die einzelnen Zeitreihen von unterschiedlichster Natur sind und mit verschieden Methoden aufgenommen wurden, sind sie zur besseren Vergleichbarkeit für die Abbildung umskaliert worden.

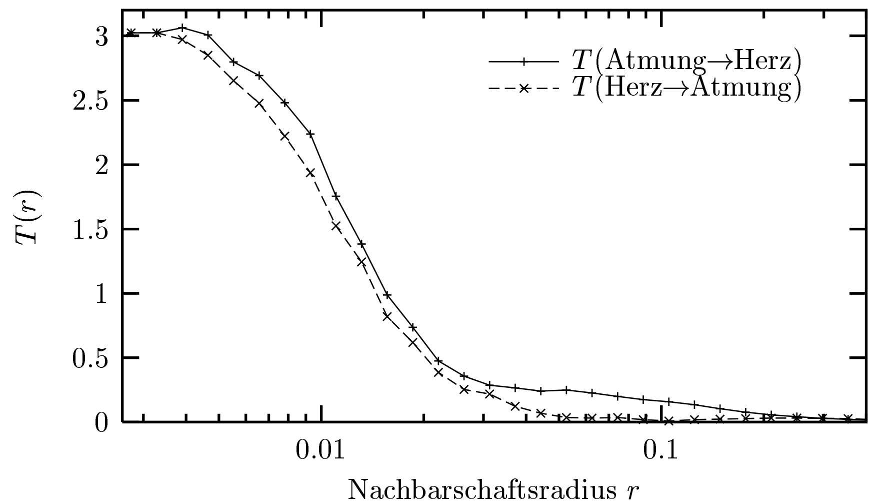

Abbildung 6.24: Transfer–Entropie T zwischen Herzrate und Atmung eines Schlafapnoikers in Abhängigkeit vom Nachbarschaftsradius r. Die systematische Abweichung deutet auf einen stärkeren Informationsfluss von der Atmungszeitreihe zur Herzrate als umgekehrt hin.

Wie in Abbildung 6.23 rechts sehr deutlich zu erkennen ist, leidet die Versuchsperson unter Schlafapnoe. Diese Krankheit zeichnet sich durch häufige Atemstillstände von bis zu 45 Sekunden aus. Diese Schlafstörung kann unterschiedlichste, ernsthafte Erkrankungen nach sich ziehen und sogar zum Tode führen. Die Tatsache, dass ungefähr 5% der Erwachsenen unter dieser Schlafstörung leiden und die Ursache bis jetzt noch unklar ist, unterstreicht die Relevanz der Forschung auf diesem Gebiet.

Zunächst sollen die fünf Zeitreihen betrachtet werden, in denen es zu eindeutigen Atemstillständen wie in Abbildung 6.23 rechts kommt, und die Beziehung zwischen Herzrate und Atmung untersucht werden. Während dieser Zeiträume ist die Versuchsperson unbewusst wach. Dies liegt daran, dass bei Ausbleiben der Atmung das Gehirn eine Überlebensreaktion auslöst, der Schlafapnoiker erwacht, mit einem lauten Schnarchen Luft holt und wieder zu atmen beginnt. Ein gesunder, erholsamer Schlaf ist unter diesen Bedingungen selbstverständlich nicht möglich, woraus sich eine Reihe von Folgeerscheinungen, wie z. B. Tagesmüdigkeit erklären lassen.

Aufgrund der unterschiedlichen Strukturen in Herzrate und Atmungszeitreihe wurden sie vor der Anwendung der Abhängigkeitsmaße umskaliert. Dabei wurden jeweils zwei verschiedene Praktiken angewendet und die Resultate verglichen. Bei der ersten Methode handelt es sich um eine einfache lineare Umskalierung auf Mittelwert Null und Varianz $\sigma^2 = 1$. Dies hat keinen Einfluss auf die Verteilung der Zeitreihen und Ausreißer bleiben Ausreißer. Um deren mögliche Auswirkung zu untersuchen, werden bei der zweiten Umskalierungsmethode beide Zeitreihen auf eine Gleichverteilung im Intervall $[0, 1]$ gebracht.

Unabhängig von der Wahl der Umskalierung sind die Resultate der Transfer–Entropie für die fünf Zeitreihen mit Atemstillständen. Ein Beispiel dafür ist in Abbildung 6.24 dargestellt, wobei hier die Parameter $k = l = 1$ und die Umskalierung auf Gleichverteilung verwendet wurde. In allen Fällen wird über einen großen Skalenbereich r hinweg die Beziehung $T(\text{Atmung} \to \text{Herz}) > T(\text{Herz} \to \text{Atmung})$ gefunden, was auf einen stärkeren Informationsfluss von der Atmung zur Herzrate als in die andere Richtung hindeutet. Dies stimmt mit der Erwartung überein, wenn der Atemstillstand durch eine Fehlfunktion

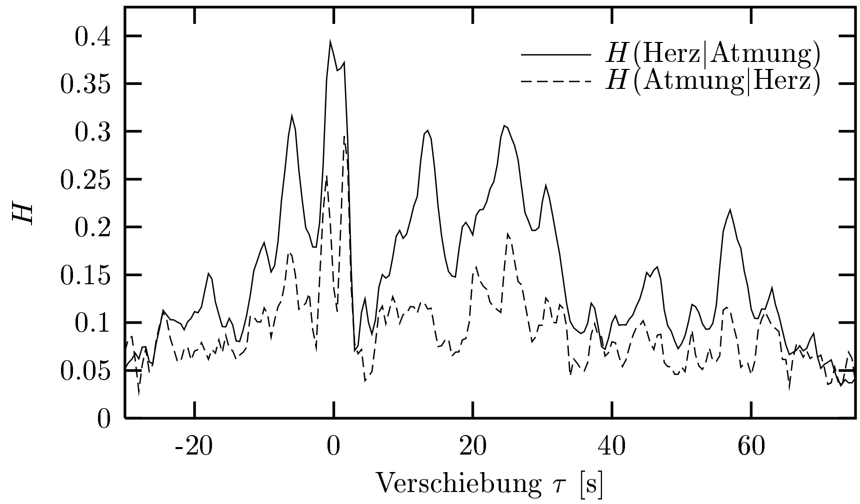

Abbildung 6.25: Abhängigkeitsmaß H zwischen Herzrate und Atmung in Abhängigkeit von der Zeitverschiebung τ zwischen den Zeitreihen. Die gefundene Asymmetrie weist darauf hin, dass die Herzrate stärker von der Atmung abhängt als umgekehrt.

des Gehirns oder durch einen Verschluss im Schlundbereich erklärt werden kann und sich daraus resultierend die Herzrate ändert. Ansatzweise lässt sich dies auch in Abbildung 6.23 rechts erkennen, wo zu Zeiten mit Atemstillstand die Herzrate tendenziell niedriger ist.

Weiter soll die Beziehung zwischen Atmung und Herzrate anhand des Abhängigkeitsmaßes H untersucht werden. Im Gegensatz zu der Transfer–Entropie zeigt sich nun aber eine stärke Abhängigkeit der Resultate von der Wahl des Umskalierungsschemas. Werden die Zeitreihen lediglich auf gleichen Mittelwert und Varianz skaliert, können keine signifikanten, eindeutigen Abweichungen zwischen $H(\text{Herz}|\text{Atmung})$ und $H(\text{Atmung}|\text{Herz})$ beobachtet werden. Dies ändert sich aber, wenn beide Zeitreihen jeweils auf Gleichverteilung gebracht werden. Die resultierenden Werte von H sind in Abbildung 6.25 wieder in Abhängigkeit von der zeitlichen Verschiebung τ aufgetragen. Während außerhalb des gezeigten τ–Bereiches die Werte von $H(\text{Herz}|\text{Atmung})$ und $H(\text{Atmung}|\text{Herz})$ annähernd übereinstimmen, zeigen sie für kleine τ eine deutliche Asymmetrie. Dabei lässt sich deren Richtung so deuten, dass die Herzrate stärker von der Atmung abhängt als umgekehrt. Das Abhängigkeitsmaß H bestätigt also die über die Transfer–Entropie gefundenen Resultate. Allerdings zeigt H hier, wie im Beispiel der Aktienindizes, eine sensitive Abhängigkeit von der Verteilung der Zeitreihen.

In zwei der acht stationären Teilstücke der Gesamtzeitreihe zeigt sich das in Abbildung 6.23 links dargestellte Verhalten, das sich durch eine regelmäßige Atmung auszeichnet. Diese beiden Zeiträume liegen zeitlich vor den oben betrachteten Daten, in denen es zu zwischenzeitlichen Atemstillständen kam. Dabei befindet sich die Versuchsperson in dem einen Zeitraum im Schlafstadium zwei und in dem anderen im Schlafstadium eins. Werden für diese Zeitreihen die Transfer–Entropien berechnet, zeigt sich nun keine von r unabhängige Abweichung mehr zwischen den beiden Richtungen. Vielmehr wird die oben gefundene Asymmetrie lediglich bei großen Skalen beobachtet und zu kleinen Nachbarschaftsradien r gerade umgekehrt. Dieser „Umkehrpunkt“ liegt bei den auf $\sigma = 1$ umskalierten Zeitreihen bei $r \approx 0.1$ und bei den auf Gleichverteilung gebrachten Daten bei $r \approx 0.03$. Da dieser Übergang damit ungefähr in der Mitte des relevanten Skalenbereiches liegt, fällt es in diesem Fall schwer, allein anhand der Transfer–Entropien ein eindeutiges Urteil zu fällen. Allerdings zeigt in diesem Fall auch das Abhängigkeitsmaß H kein eindeutiges Verhalten. Für die eine Zeitreihe (Schlafphase zwei) gilt unabhängig von der Wahl der Umskalierung im Bereich $-5\,\mathrm{s} < \tau < 5\,\mathrm{s}$ die oben gefundene Asymmetrie $H(\mathrm{Herz}|\mathrm{Atmung}) > H(\mathrm{Atmung}|\mathrm{Herz})$, während in der anderen Zeitreihe keine eindeutigen Abweichungen der beiden Richtungen voneinander beobachtet werden können. Von einer eindeutig stärkeren Abhängigkeit in eine Richtung kann also in dem hier betrachteten Fall der regelmäßigen Atmung nicht gesprochen werden. Zur Klärung dieses Sachverhalts wären also weitere Daten nötig, die idealerweise an einem gesunden Menschen aufgenommen worden sind.

Die verbleibende Teilzeitreihe (Schlafstadium zwei) unterscheidet sich in ihrer Struktur noch einmal wesentlich von den anderen, bereits diskutierten Zeitreihen. In ihr wechseln sich Phasen mit relativ gleichmäßiger Atmung und Phasen mit schwacher, unregelmäßiger Atmung ab, wobei die Periode ungefähr 40 Sekunden beträgt. Ein halb so langer, zehn Minuten umfassender Teil dieser Zeitreihe wurde bereits in [Sch00b] auf seine Transfer–Entropien hin untersucht. Das dort gefundene Resultat, $T(\mathrm{Herz}\to\mathrm{Atmung}) > T(\mathrm{Atmung}\to\mathrm{Herz})$, kann ich hier unabhängig von der Wahl der Umskalierung bestätigen. Man beachte, dass diese Beziehung von der Richtung abweicht, die für die fünf Zeitreihen mit Atemstillständen gefunden wurde. Im Gegensatz dazu findet man mit dem Abhängigkeitsmaß H in Übereinstimmung mit den anderen Fällen, dass die Herzrate stärker von der Atmung abhängt als umgekehrt. Die Resultate der beiden Maße scheinen sich also zu widersprechen und auch in diesem Fall wären weitere Daten wünschenswert. Darüber hinaus wäre es interessant zu erfahren, welche Vorgänge nach Meinung von Medizinern dem Atmungsverhalten in dieser Phase zugrunde liegen.

Schließlich soll noch die Rolle der Sauerstoffkonzentration im Blut untersucht werden. Während die Beziehung zwischen Atmung und Herzrate lediglich im Fall der Apnoen–Zeitreihen eindeutige Resultate aufweist, zeigt sich hier ein anderes Bild. Für alle acht Zeitreihen werden die Beziehungen $T(\mathrm{O_2}\to\mathrm{Herz}) > T(\mathrm{Herz}\to\mathrm{O_2})$ und $T(\mathrm{O_2}\to\mathrm{Atmung}) > T(\mathrm{Atmung}\to\mathrm{O_2})$ unabhängig vom Nachbarschaftsradius r und dem Umskalierungsschema gefunden. Mit der Interpretation dieser Ergebnisse sollte man allerdings vorsichtig sein. Die Sauerstoffkonzentration variiert auf wesentlich längeren Zeitskalen als die beiden anderen Observablen und ist deshalb nicht so leicht vergleichbar. Dies schlägt sich z. B. auch in den individuellen Entropien nieder. Während Atmung und Herzrate noch annähernd gleiche Entropien aufweisen, zeigt die Sauerstoff–Zeitreihe in allen Fällen eine deutlich niedrigere Entropie. Diese Unterschiede in den individuellen Entropien können, wie bereits in [Sch00b] erwähnt, zu einer Asymmetrie der Transfer–Entropie führen.

Im Falle der Beziehung zwischen Atmung und Sauerstoffgehalt des Blutes ist eine wechselseitige Beeinflussung offensichtlich. Einerseits wird die Atmung forciert, wenn die Sauerstoffkonzentration zu niedrig ist, und andererseits erhöht eine schnellere Atmung den Sauerstoffgehalt des Blutes. Weiter ist es offensichtlich, dass in beiden Vorgängen eine zeitliche Verzögerung auftritt. Die gefundene Asymmetrie, $T(\mathrm{O_2}\to\mathrm{Atmung}) > T(\mathrm{Atmung}\to\mathrm{O_2})$, würde darauf hindeuten, dass in einem kurzen Zeitrahmen von 0.5 Sekunden der Informationsfluss in der Richtung Sauerstoffkonzentration zu Atmung größer ist als umgekehrt. Auch das Abhängigkeitsmaß H liefert für alle acht Zeitreihen übereinstimmend ein Ergebnis. In einem endlichen τ–Bereich um $\tau = 0$ gilt stets die Beziehung $H(\mathrm{O_2}|\mathrm{Atmung}) > H(\mathrm{Atmung}|\mathrm{O_2})$. In Abbildung 6.26 ist H für eine Zeitreihe mit Atemstillständen aufgetragen, von der ein Ausschnitt bereits in Abbildung 6.23 rechts zu sehen war. Die

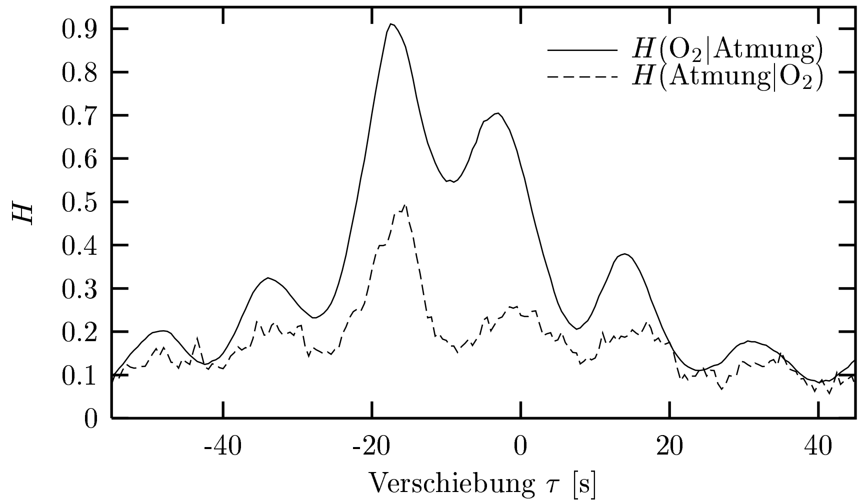

Abbildung 6.26: Abhängigkeitsmaß H zwischen Atmung und Sauerstoffkonzentration im Blut. Die gefundene Asymmetrie weist darauf hin, dass die Sauerstoffkonzentration stärker von der Atmung abhängt als umgekehrt, wobei dieser Effekt in beide Richtungen am stärksten bei $\tau \approx -18\,\mathrm{s}$ ausgeprägt ist.

mögliche Interpretation, dass es sich bei Sauerstoffkonzentration und Atmung um Observablen zweier gekoppelter Systeme handelt, wobei die Atmung das getriebene System darstellt, ist auch hier wieder mit äußerster Vorsicht zu genießen. Naheliegender ist die Interpretation, dass H in der Atmung lediglich das aktivere der beiden Systeme erkennt, was bereits durch einen Blick auf die einzelnen Zeitreihen zu erkennen ist. Die größten Werte von H sind in Abbildung 6.26 in beide Richtungen für $\tau \approx -18\,\mathrm{s}$ zu finden, wobei $H(\mathrm{Atmung}|\mathrm{O}_2)$ an dieser Stelle das einzige deutliche Maximum zeigt. Bei dieser Zeitverschiebung zwischen den Zeitreihen treffen die Minima der Sauerstoffkonzentration mit dem Wiedereinsetzen der Atmung zusammen. Es ist gut möglich, dass es sich dabei um eine verstärkte Wechselwirkung beim Auslösen der oben erwähnten Überlebensreaktion des Körpers handelt.

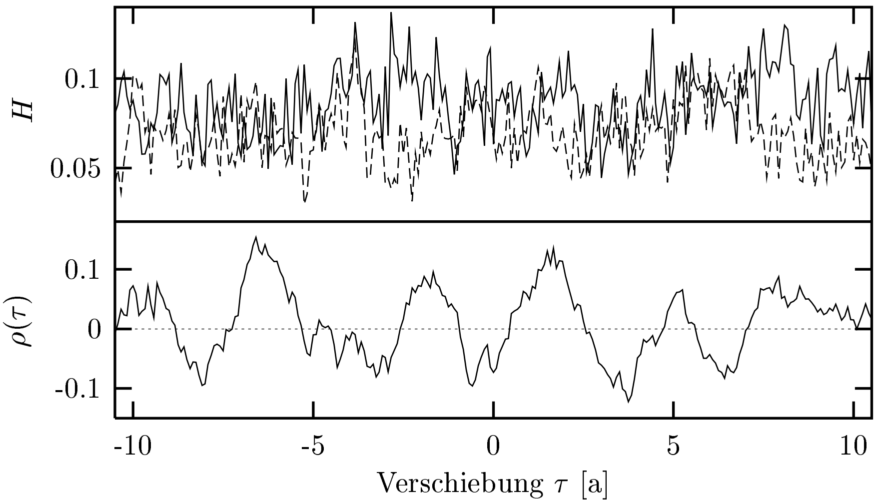

Abbildung 6.27: Abhängigkeitsmaße $H(\mathrm{TADI}|\mathrm{SOI})$ (oben, durchgezogen), $H(\mathrm{SOI}|\mathrm{TADI})$ (oben, gestrichelt) und Kreuzkorrelationen (unten) zwischen SOI und TADI in Abhängigkeit von der Zeitverschiebung τ in Jahren.

6.3.3 Klimadaten

Als abschließendes Beispiel sollen die Abhängigkeiten zwischen zwei bekannten Klimaindizes untersucht werden. Bei dem ersten handelt es sich um den *Southern Oscillation Index* (SOI), der bereits in Abschnitt 2.2.4 als Beispiel herangezogen wurde und in Abbildung 2.3 oben gezeigt ist. Die zweite Zeitreihe ist der *Tropical Atlantic Sea Surface Temperature Dipole Index* (kurz TADI), welcher aus Temperaturdifferenzen des Oberflächenwassers im tropischen Atlantik abgeleitet wird. Die Daten wurden freundlicherweise von Norbert Marwan von der Universität Potsdam zur Verfügung gestellt. Beide Indizes werden monatlich berechnet und umspannen den Zeitraum von 1884 bis 1990, was 1284 Werten entspricht. Beide Zeitreihen zeigen annähernd eine Gaußverteilung und keine außergewöhnlichen Ausreißer, sodass sie für die folgenden Untersuchungen lediglich auf die gleiche Varianz umskaliert wurden.

In Abbildung 6.27 ist das Abhängigkeitsmaß H zwischen den beiden Indizes in Abhängigkeit von der Zeitverschiebung τ dargestellt, wobei die Parameter $m = 1$ und $k = 15$ verwendet wurden. Versuche mit anderen Einbettungsdimensionen m und unterschiedlicher Anzahl von nächsten Nachbarn k zeigten keine nennenswerten Abweichungen von dem hier gezeigten Ergebnis. Sowohl $H(\mathrm{TADI}|\mathrm{SOI})$ als auch $H(\mathrm{SOI}|\mathrm{TADI})$ sind auf niedrigem Niveau nahezu unabhängig von τ und liefern bis auf Fluktuationen gleiche Werte. Eine eindeutige Abhängigkeit zwischen den beiden Zeitreihen kann also über das Maß H nicht gefunden werden. Die ansatzweise realisierte Asymmetrie in den Bereichen um $\tau = \pm 2$ Jahre, an denen die Kreuzkorrelation Maxima aufweist, kann ebenfalls kaum als signifikant bezeichnet werden. Ähnlich sieht das Bild bei den Transfer–Entropien aus, welche in Abbildung 6.28 in Abhängigkeit vom Nachbarschaftsradius r aufgetragen ist. Dabei liefern die Transfer–Entropien in beide Richtungen annähernd gleiche Werte und zeigen keine signifikanten Abweichungen voneinander. Man vergleiche dieses Resultat mit den in Abbildung 6.22 und Abbildung 6.24 gezeigten Ergebnissen.

Zusammenfassend lässt sich sagen, dass in diesem Fall weder das Abhängigkeitsmaß H noch die Transfer–Entropie eine eindeutige Abhängigkeit zwischen den Zeitreihen nachwei-

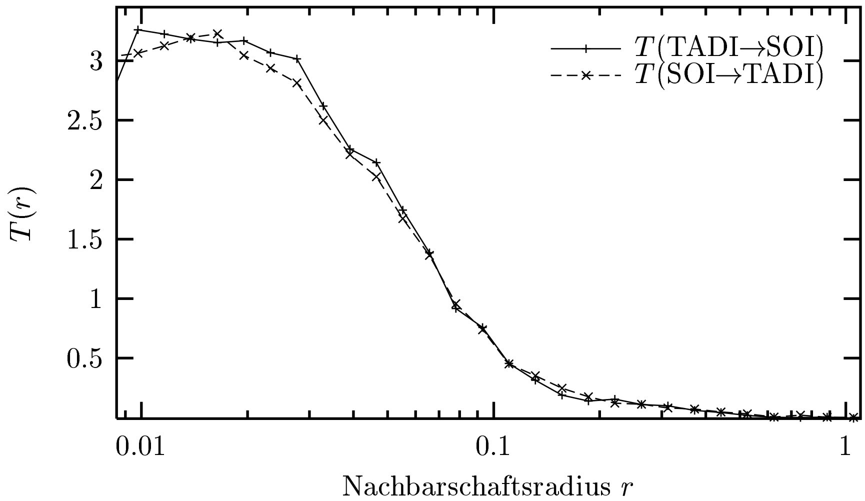

Abbildung 6.28: Die Transfer–Entropien zwischen SOI und TADI zeigen keine eindeutige, von r unabhängige Asymmetrie und können bis auf Fluktuationen als gleich angesehen werden.

sen können. Ähnlich wie im Fall der Aktienindizes haben wir es hier mit einer extremen Reduktion der komplizierten, klimatischen Vorgänge auf einzelne Zahlen zu tun, sodass eventuell vorhandene Beziehungen nur schwer in den Zeitreihen gefunden werden können.

Kapitel 7

Zusammenfassung und Ausblick

Der zweite Teil der vorliegenden Arbeit beschäftigte sich mit der Analyse von Abhängigkeiten zwischen Zeitreihen. Dabei wurden im Hinblick auf gekoppelte Systeme Methoden angestrebt, über die sich sowohl der Grad als auch die dominierende Richtung der Abhängigkeit bestimmen lassen. Es zeigte sich, dass die Frage nach der Stärke der Abhängigkeit in vielen Fällen bereits mit gängigen Standardmaßen beantwortet werden kann. Auf die Bestimmung einer eventuell dominierenden Kopplungsrichtung zwischen den Systemen sind diese Maße aber nicht optimiert. Dadurch motiviert wurden zwei neue, von ihren Grundideen völlig unterschiedliche Methoden entwickelt und diskutiert.

Die erste Größe, die Transfer–Entropie, war dabei durch die Informationstheorie motiviert, welche einen Rahmen zur Beschreibung von sowohl stochastischen als auch nichtlinearen Systemen liefert. Eines der attraktivsten Konzepte bei der Untersuchung von Abhängigkeiten ist dabei die gegenseitige Information, welche den *gemeinsamen* Informationsgehalt zweier Systeme angibt und damit per Konstruktion symmetrisch ist. Im Gegensatz dazu, kann über die Transfer–Entropie der direkte Informationsfluss zwischen zwei Systemen bestimmt und darüber die Richtung einer Kopplung ermittelt werden. Die zweite neue Klasse von Abhängigkeitsmaßen basiert ursprünglich auf der Analyse von Synchronisationsphänomenen von einseitig gekoppelten, chaotischen Systemen. Aus der dabei auftretenden festen Beziehung zwischen den Systemen folgt unter anderem, dass ähnliche Zustände in dem einen System zu gleichen Zeiten ähnliche Zustände im anderen System implizieren. Über die Idee des bedingten Abstands werden so zwei neue Abhängigkeitsmaße S und H konstruiert, welche ebenfalls asymmetrisch sind.

Zur Untersuchung der ersten Eigenschaften der neuen Maße und dem Vergleich zu den Standardmethoden wurden zunächst einfache gekoppelte Modellsysteme betrachtet. Dabei zeichneten sich sowohl die Transfer–Entropie als auch das Abhängigkeitsmaß H in allen betrachteten Fällen durch eine deutliche Asymmetrie aus, an der bereits bei schwacher Abhängigkeit die Richtung der Kopplung abgelesen werden konnte. Die gleichzeitige Beobachtung von Unterschieden in den *individuellen* Eigenschaften der beteiligten Systeme motivierte die Betrachtung eines Ringes aus mehreren Systemen. So weisen Abbildungen auf einem Gitter mit periodischen Randbedingungen trotz der einseitigen Kopplung identische individuelle Eigenschaften auf. Die Anwendung der neuen Maße auf ein solches Ring–System zeigte, dass sie auch in diesem Fall die gewünschte Asymmetrie aufweisen und die Richtung der Kopplung korrekt wiedergeben.

Von der Funktionstüchtigkeit und Leistungsfähigkeit der neuen Abhängigkeitsmaße überzeugt, wurden sie schließlich auf erste gemessene Zeitreihen aus unterschiedlichen wissenschaftlichen Bereichen angewendet. Die betrachteten Daten deckten dabei drei verschiedene Fälle ab. Zum einen konnte im Fall der Klimadaten, wo die Beziehung von zwei globalen Indizes untersucht wurde, keine signifikante Abhängigkeit nachgewiesen werden. Im Gegensatz dazu konnte im zweiten Fall der Aktienindizes mit beiden Maßen eine eindeutige Abhängigkeit zwischen DAX und Dow Jones gefunden werden. Sie bestätigen damit bekannte Resultate, die bereits über lineare Korrelationen gewonnen werden können, gehen über diese Erkenntnisse aber nicht hinaus. Schließlich wurden bei den physiologischen Daten Beziehungen gefunden, die die Standardmaße bisher nicht offenbarten. So wurde eine Abhängigkeit zwischen Atmung und Herzrate eines Schlafapnoikers nachgewiesen, deren dominante Richtung in Übereinstimmung sowohl von der Transfer–Entropie als auch von dem Abhängigkeitsmaß H gefunden wurde.

Weitere, zukünftige Anwendungen der neuen Abhängigkeitsmaße auf experimentelle Daten werden zeigen, inwieweit durch sie neue, über die Standardmaße hinausreichende Ergebnisse gefunden werden können. Parallel sollten die Methoden auf Modellsystemen optimiert werden. Im Falle der Abhängigkeitsmaße S und H wurde deutlich, dass relativ kleine konzeptionelle Änderungen teilweise erhebliche Auswirkungen auf die Leistungsfähigkeit der Methode haben können. Eine systematische Untersuchung von alternativen Implementationen lässt also durchaus eine weitere Leistungssteigerung erwarten. Auch bei der Transfer–Entropie ist die Entwicklung noch nicht abgeschlossen. Der wünschenswerte Übergang zu größeren Vergangenheitsfenstern ist bislang stark durch die Länge der betroffenen Zeitreihen eingeschränkt. Eine adäquate Lösung dieses Problems könnte die Asymmetrie der Transfer–Entropie auch bei relativ kurzen Zeitreihen deutlicher ausprägen.

Literaturverzeichnis

[Aba96] H. D. I. Abarbanel. *Analysis of Observed Chaotic Data.* Springer-Verlag, New York Berlin Heidelberg, 1996.

[AGLE99] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger. A robust method for detecting interdependences: Application to intracranially recorded EEG. *Physica D*, 134(4):419–430, 1999.

[Bha88] Gyan Bhanot. The metropolis algorithm. *Rept. Prog. Phys.*, 51:429, 1988.

[BP85] R. Badii and A. Politi. Statistical description of chaotic attractors: the dimension function. *J. Stat. Phys.*, 40:725–750, 1985.

[DvHTD95] C. Diks, J. C. van Houwelingen, F. Takens, and J. DeGoede. Reversibility as a criterion for discriminating time series. *Phys. Lett. A*, 210(2-3):221–228, 1995.

[EKR87] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhys. Lett.*, 4:973–977, 1987.

[FS86] A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33(2):1134–1140, 1986.

[FY83] H. Fujisaka and T. Yamada. Stability theory of synchronized motion in coupled-oscillator systems. *Prog. Theor. Phys.*, 69:32, 1983.

[GMB87a] N. E. Graham, J. Michaelsen, and T. P. Barnett. An investigation of the El Nino-Southern Oscillation cycle with statistical models - 1. predictor field characteristics. *J. Geophys. Res.*, 92:1425–1427, 1987.

[GMB87b] N. E. Graham, J. Michaelsen, and T. P. Barnett. An investigation of the El Nino-Southern Oscillation cycle with statistical models - 2. model results. *J. Geophys. Res.*, 92:1427–1428, 1987.

[Gra88] P. Grassberger. Finite sample corrections to entropy and dimension estimates. *Phys. Lett. A*, 128:369–373, 1988.

[Gra90a] P. Grassberger. An optimized box-assisted algorithm for fractal dimensions. *Phys. Lett. A.*, 148(1/2):63–68, 1990.

[Gra90b] R. M. Gray. *Entropy and Information Theory.* Springer-Verlag, New York, 1990.

[GSS91] P. Grassberger, T. Schreiber, and C. Schaffrath. Nonlinear time sequence analysis. *Int. J. of Bifurcation and Chaos*, 1(3):521–547, 1991.

[HKS99] R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, 9(2):413–435, 1999.

[Huy73] C. Huygens. *Horologium Oscillatorium*. Muguet, Paris, 1673.

[Kan94] H. Kantz. Quantifying the closeness of fractal measures. *Phys. Rev. E*, 49(6):5091–5097, 1994.

[KJV83] S. Kirkpatrik, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671, 1983.

[Kol93] A. N. Kolmogorov. *Information Theory and the Theory of Algorithms*, volume III of *Selected Works*. Kluwer Academic, Dordrecht, 1993.

[KS97] H. Kantz and Th. Schreiber. *Nonlinear Time Series Analysis*. Cambridge Univ. Press, Cambridge, UK, 1997.

[Kug99] D. Kugiumtzis. Test your surrogate data before you test for nonlinearity. *Phys. Lett. E*, 60:2808, 1999.

[Kul59] S. Kullback. *Information Theory and Statistics*. Dover, New York, 1959.

[Lom76] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophys. and Space Sci.*, 39:447–462, 1976.

[Mar78] F. H. C. Marriott. Barnard's monte–carlo tests: how many simulations? *App. Statist.*, 28:75, 1978.

[MRR+53] N. Metropolis, A. Rosenblut, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machine. *J. Chem. Phys.*, 21:1097, 1953.

[MS96] R. Manuca and R. Savit. Stationarity and nonstationarity in time series analysis. *Physica D*, 99:134–161, 1996.

[NWC+90] R. E. Nather, D. E. Winget, J. C. Clemens, C. J. Hansen, and B. P. Hine. The whole earth telescope: a new astronomical instrument. *Astrophys. J.*, 361:309–317, 1990.

[PAD93] M. Paluš, V. Albrecht, and I. Dvořák. Information-theoretic tests for nonlinearity in time series. *Phys. Lett. A*, 175:203–209, 1993.

[PC90] L. M. Pecora and T. L. Carroll. Synchronization in chaotic systems. *Phys. Rev. Lett.*, 64:821–824, 1990.

[PS87] K. Pawelzik and H. G. Schuster. Generalized dimensions and entropies from a measured time series. *Phys. Rev. A*, 35:481, 1987.

[PT95] D. Prichard and J. Theiler. Generalized redundancies for time series analysis. *Physica D*, 84:476–493, 1995.

[PTVF92] W. H. Press, S. T. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: the art of scientific computing.* Cambridge University Press, Cambridge, England, 1992.

[Pyr96] K. Pyragas. Weak and strong synchronization of chaos. *Phys. Rev. E*, 54(5):4508–4511, 1996.

[QAG00] R. Quian Quiroga, J. Arnhold, and P. Grassberger. Learning driver-response relationships from synchronization patterns. *Phys. Rev. E*, 61(5):5142–5148, 2000.

[QMAV99] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela. Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Physica D*, 127:250–266, 1999.

[Ren70] A. Renyi. *Probability Theory.* North–Holland, Amsterdam, 1970.

[RG84] T. Subba Rao and M. Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, volume 24 of *Lecture Notes in Statistics.* Springer-Verlag, Berlin, 1984.

[RSTA95] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E*, 51(2):980–994, 1995.

[Sca82] J. D. Scargle. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263:835–853, 1982.

[Sch98] Th. Schreiber. Constrained randomization of time series data. *Phys. Rev. Lett.*, 90(10):2105–2108, 1998.

[Sch00a] A. Schmitz. Measuring statistical dependence and coupling of subsystems. *Phys. Rev. E*, 62(5):7508–7511, 2000.

[Sch00b] Th. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461, 2000.

[SG97] P. Schneider and P. Grassberger. Studying attractor symmetries by means of cross-correlation sums. *Nonlinearity*, 10:749–762, 1997.

[Sil86] B. W. Silverman. Kernel density estimation technique for statistics and data analysis. In *Monographs on statistics and applied probability*, volume 26. Chapman and Hall, London, 1986.

[SS87] R. Schlittgen and H. J. Streitberg. *Zeitreihenanalyse.* Oldenbourg Verlag, München Wien, 2. edition, 1987.

[SS96] Th. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 77(4):635–638, 1996.

[SS97a] Th. Schreiber and A. Schmitz. Classification of time series data with nonlinear similarity measures. *Phys. Rev. Lett.*, 79(8):1475–1478, 1997.

[SS97b] Th. Schreiber and A. Schmitz. Discrimination power of measures for nonlinearity in a time series. *Phys. Rev. E*, 55(5):5443–5447, 1997.

[SS99a] A. Schmitz and T. Schreiber. Surrogate data for non–stationary signals. In *Chaos in Brain? Interdisc. Workshop, 10-12 March 1999, Bonn, Germany*, pages 222–225, Bonn, Germany, 1999. Clinic of Epileptology, Rhein.-Friedr.-Wilhelms-Univ., Bonn, Germany.

[SS99b] A. Schmitz and T. Schreiber. Testing for nonlinearity in unevenly sampled time series. *Phys. Rev. E*, 59(4):4044–4047, 1999.

[SS00] Th. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.

[SSC+96] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E*, 54(6):6708–6724, 1996.

[SW49] C. E. Shannon and W. Weaver. *The mathematical theory of information.* University Press, Urbana Ill., 1949.

[TEL+92] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.

[The86] J. Theiler. Spurious dimensions from correlation algorithms applied to limited time-series data. *Phys. Rev. A*, 34:2427–2432, 1986.

[Tim98] J. Timmer. Power of surrogate data testing with respect to nonstationarity. *Phys. Rev. E*, 58(4):5153–5156, 1998.

[TLR93] J. Theiler, P. S. Linsay, and D. M. Rubin. Detecting nonlinearity in data with long coherence times. In A. S. Weigend and N. A. Gershenfeld, editors, *Time series prediction: Forecasting the future and understanding the past*, pages 429–455. Addison Wesley, Reading Mass., 1993.

[TP96] J. Theiler and D. Prichard. Constrained-realization monte-carlo method for hypothesis testing. *Physica D*, 94:221, 1996.

[Vid93] R. V. V. Vidal. Applied simulated annealing. In *Lect. Notes in Econom. and Math.Syst.*, volume 396. Springer Verlag, Berlin, 1993.

[WG93] A. S. Weigend and N. A. Gershenfeld. *Time series prediction: Forecasting the future and unerstanding the past*, volume XV of *Santa Fe Institute Studies in the Science of Complexity, Proc.* Addison -Wesley, Reading, 1993.

Danksagung

Mein Dank gilt in erster Linie Herrn Prof. Dr. P. Grassberger und Priv. Doz. Dr. Thomas Schreiber, die mir die Durchführung sowohl meiner Diplom– als auch meiner Doktorarbeit ermöglicht haben und aus deren Wissens–Pool ich im Laufe meiner Arbeit schöpfen durfte. Sie haben mich stets unterstützt und — meiner Meinung nach — die richtige Mischung aus motivierendem Druck und kreativer Freiheit für mich gefunden. Darüber hinaus haben sie mir die Teilnahme an internationalen Konferenzen und Workshops ermöglicht, über die ich Kontakte zu vielen interessanten Wissenschaftlern knüpfen konnte.

Des Weiteren danke ich allen Kollegen/Freunden in den Arbeitsgruppen, mit denen ich Kontakt hatte und zusammengearbeitet habe. Sie haben dazu beigetragen, dass ich nie den Spaß an der wissenschaftlichen Arbeit verloren habe. Aus der Wuppertaler Gruppe möchte ich dabei speziell Thomas, Helge, Peter und Hans dankend hervorheben. Weiter hatte ich mit Rodrigo, Jochen und Johannes im Forschungszentrum Jülich viele interessante Diskussionen und auch bei Besuchen in Dresden in der Gruppe um Holger Kantz fühlte ich mich immer gut aufgehoben.

Ein Glücksfall war auch meine „provisorische“ Unterbringung in Büro F.10.08, wo ich bis heute meinen Arbeitsplatz habe. Mit meinen — chronologisch sortierten — Büro–Kollegen Matthias, Christoph, Martin R., Martin S. (kann man mitzählen), Pedro, Jens, Bram, Thomas, Erwin (ganz kurz), Boris und Guido hatte ich immer viel Spaß. Ihnen und den anderen Kollegen aus der theoretischen Physik möchte ich für die schöne Zeit danken.

Natürlich danke ich auch meinen Eltern für die jahrzehntelange Unterstützung, wobei ich meine Mutter allerdings enttäuschen muss: Zu einem Nobelpreis hat es bis heute nicht gereicht.

Zu guter Letzt möchte ich der für mich wichtigsten Person danken: Meiner Frau — und Freundin — Bärbel. Sie hat mich während meines ganzen Studiums durch alle Höhen und Tiefen begleitet. Ohne ihren Rückhalt, ihre Unterstützung, ihr Verständnis und die unzähligen schönen Stunden, die wir zusammen verbracht haben, wäre diese Arbeit nicht möglich gewesen. Ihr soll diese Arbeit gewidmet sein.