Zeitreihenanalytische Verfahren für die Analyse nichtlinearer Systeme

Unterlagen zur Summer School 2013 von

PD Dr. Dr. Guido Strunk
guido.strunk@complexity-research.com
www.complexity-research.com
Salisstr. 5-15/6/26 – A-1140 Wien
Weitere Infos unter:
www.complexity-research.com/summerschool.htm

1. Grundsteine – Einführung 2
1.1 Grundproblem. Wie kann man die Komplexität einer Dynamik bestimmen? 2
1.2 Abgrenzung und Messung von Komplexität 3
2. Grundlegende Vorgehensweisen 4
2.1 Stationarität? 4
2.2 Surrogatdatenverfahren 5
3. Phasenraumeinbettung 6
4. Grundprinzipien gängiger Verfahren 7
4.1 Fraktale Geometrie – D2 / PD2 7
4.2 Schmetterlingseffekt – Lyapunov-Exponent LLE 9
4.3 Beinahe gleich – Recurrence Plots 10
4.4 Entropie-Maße: Permutationsentropie / Zentraltendenztransformator 12
4.5 Dynamische Komplexität (Schiepek) 14
4.5.1 Fluktuation – F 14
4.5.2 Verteilung – D 15
5. Software 15
6. Literatur 16

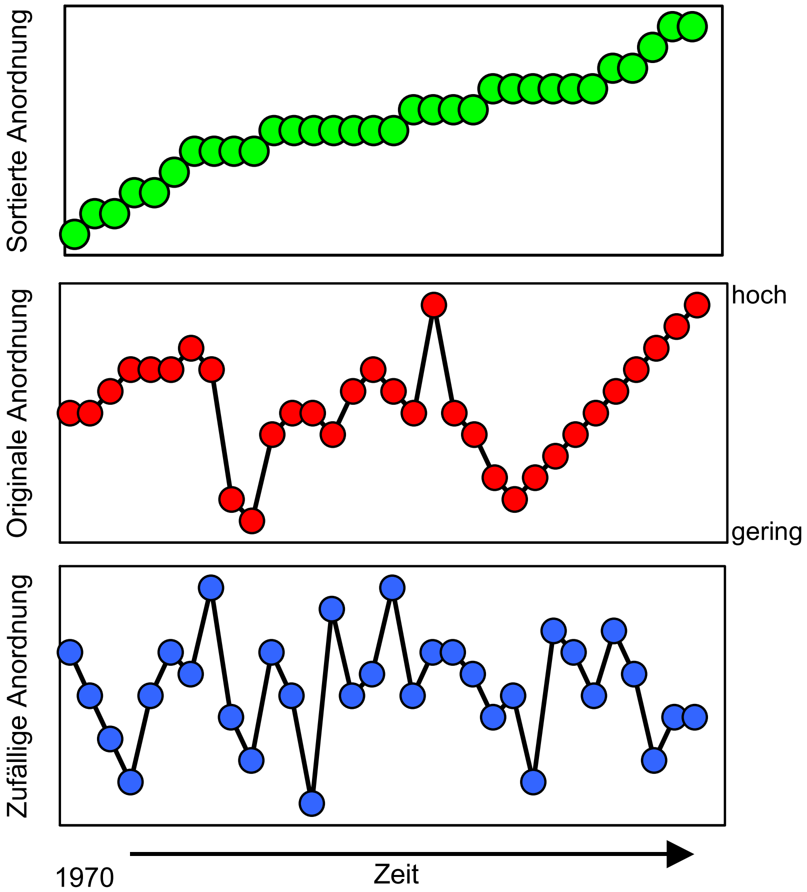
1. Grundsteine – Einführung

- **Grundproblem.** Wie kann man die Komplexität einer Dynamik bestimmen?
- **Grundthema.** Was sind die Unterscheide zwischen Ordnung, Komplexität und Zufall?
- **Grundlegende Vorgehensweisen.** Wie sieht typischer Weise ein Untersuchungsdesign aus?
- **Grundprinzipien der Datenaufbereitung.** Wie kann eine Dynamik „übersichtlich" aufbereitet werden?
- **Grundprinzipien gängiger Verfahren.** Welche Verfahren messen wie die Ordnung, die Komplexität oder den Zufall?
- **Software.** GChaos – complexity-research.com

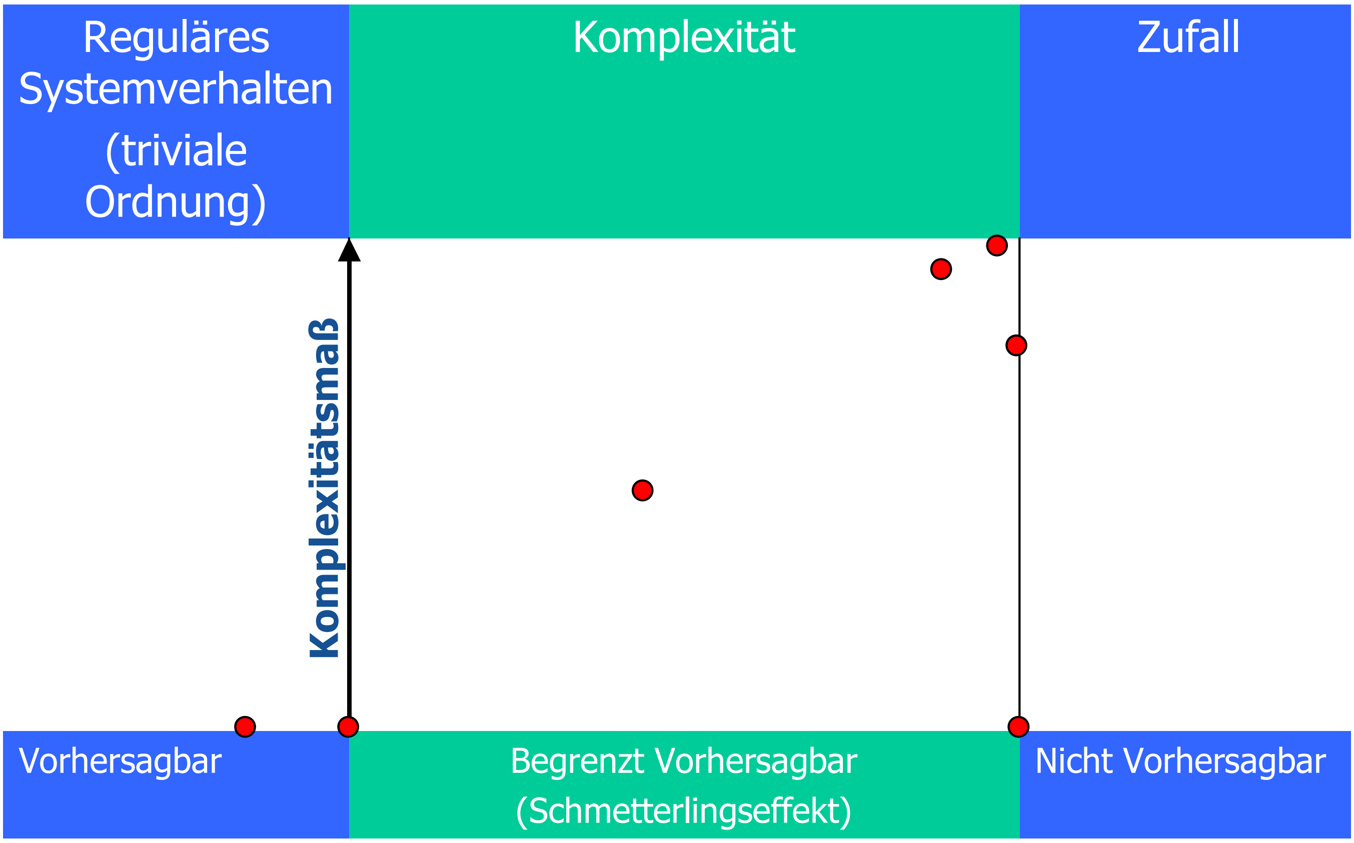
1.1 Grundproblem. Wie kann man die Komplexität einer Dynamik bestimmen?

Ein Beispiel

Zufriedenheit mit der beruflichen Entwicklung



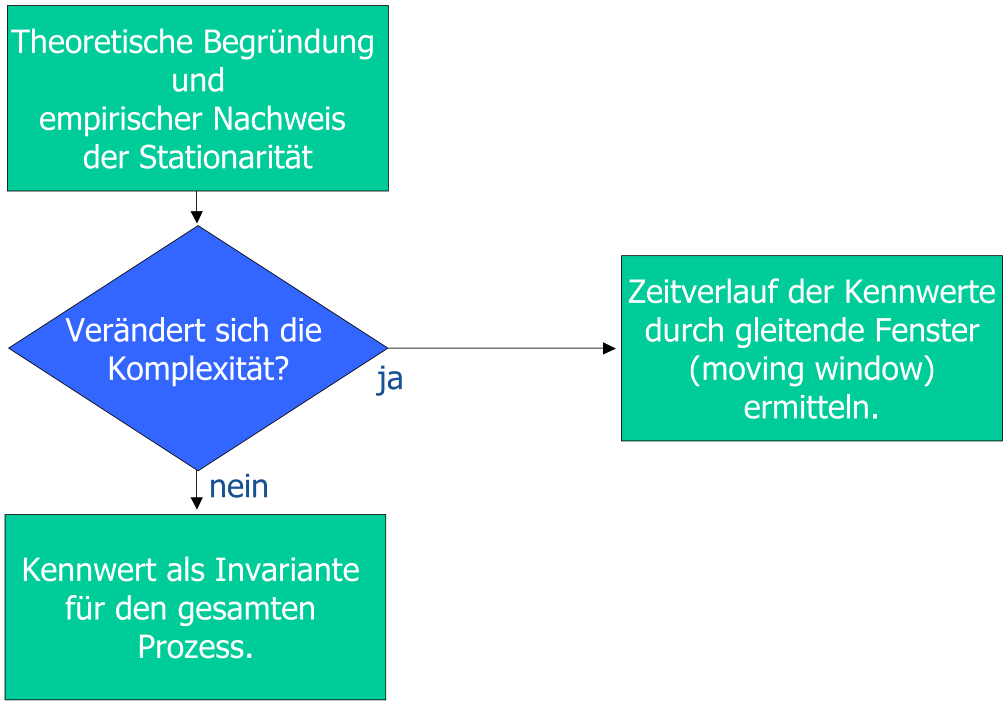
1.2 Abgrenzung und Messung von Komplexität



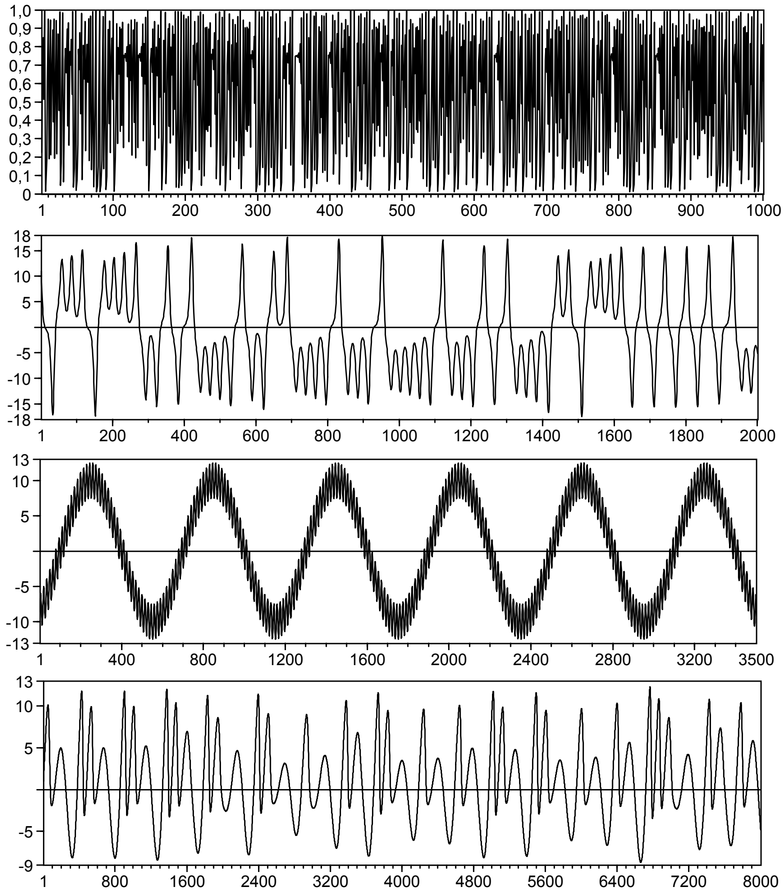
- **Ordnung feststellen.** Z.B. wie oft wiederholt das System sein Verhalten in fast gleicher Weise (Recurrence Plots)? Gemessen wird der Anteil „einfacher" regulärer Muster an der Gesamtdynamik.
- **Zufall feststellen.** Kann die Dynamik von Zufall unterschieden werden? Z.B. Permutationsentropie. Diese besitzt ein Maximum, welches maximalem Zufall entspricht.
- **Komplexität konkret messen.** Z.B. den Schmetterlingseffekt identifizieren und quantifizieren. Die Berechnung des Lyapunov-Exponenten schlägt bei Zufall fehl.

2. Grundlegende Vorgehensweisen

2.1 Stationarität?



Stationäre Analyse – Beispiele



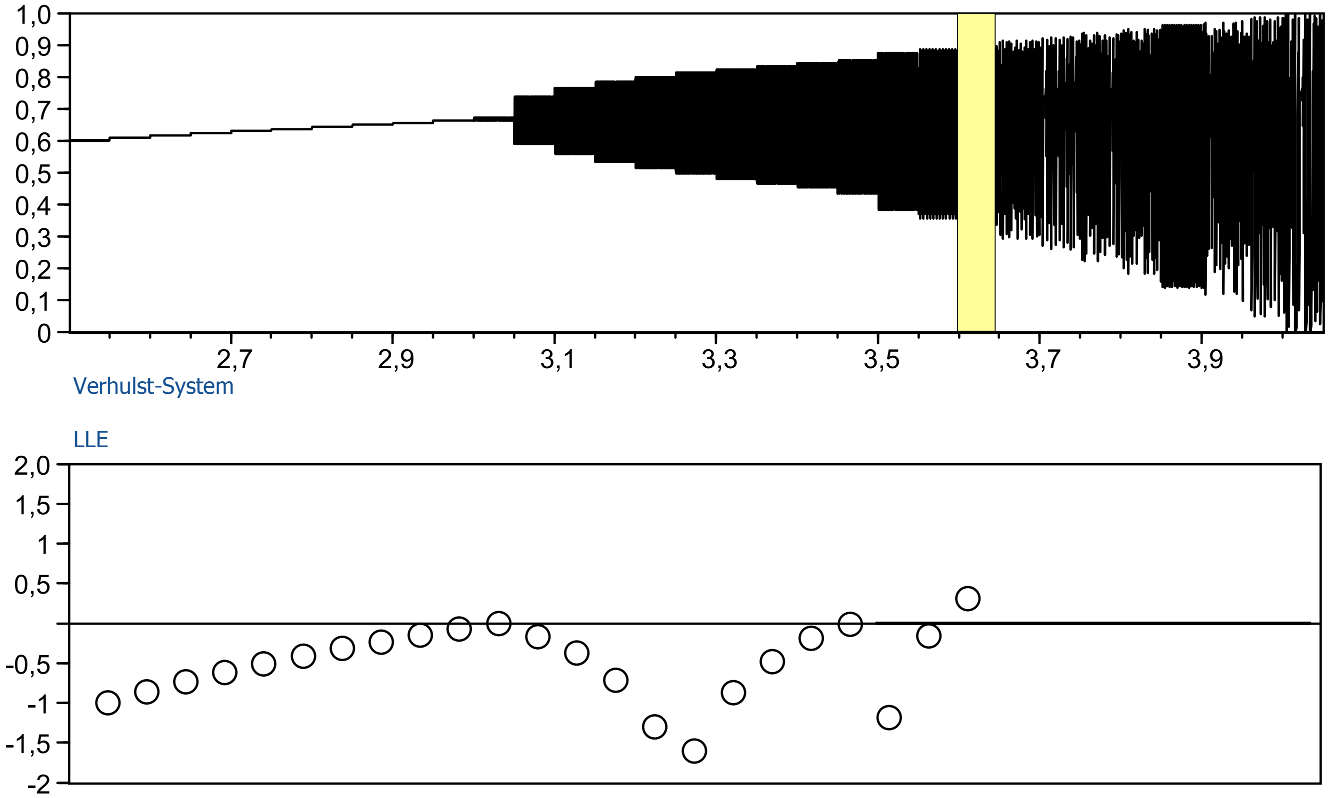
Verhulst-System (r= 4)
D2 = 1,0
LLE = 0,693 (Basis e)

Lorenz-System (x; sigma=10, r=28, b=8/3)
D2 = 2,068
LLE = 0,906 (Basis e)

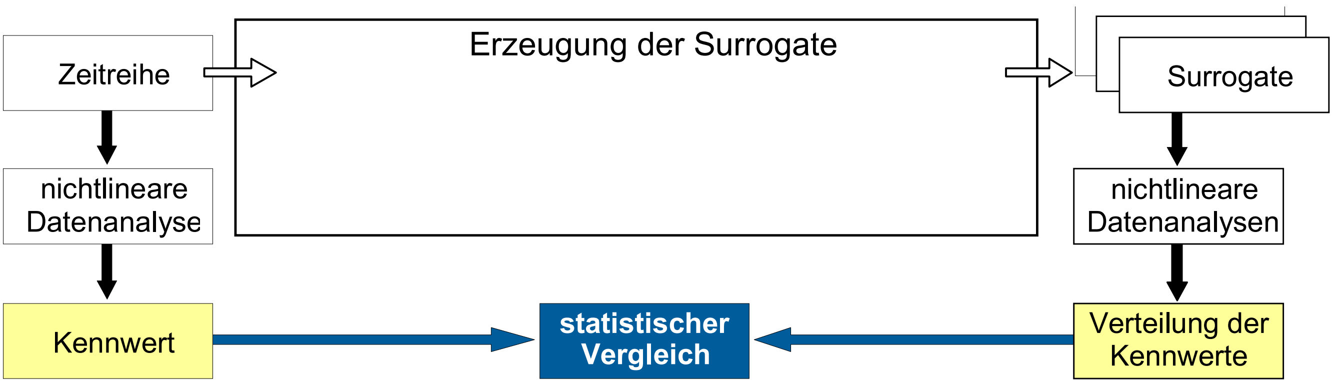
Sinus plus Sinus
D2 = 1,0
LLE = 0,000

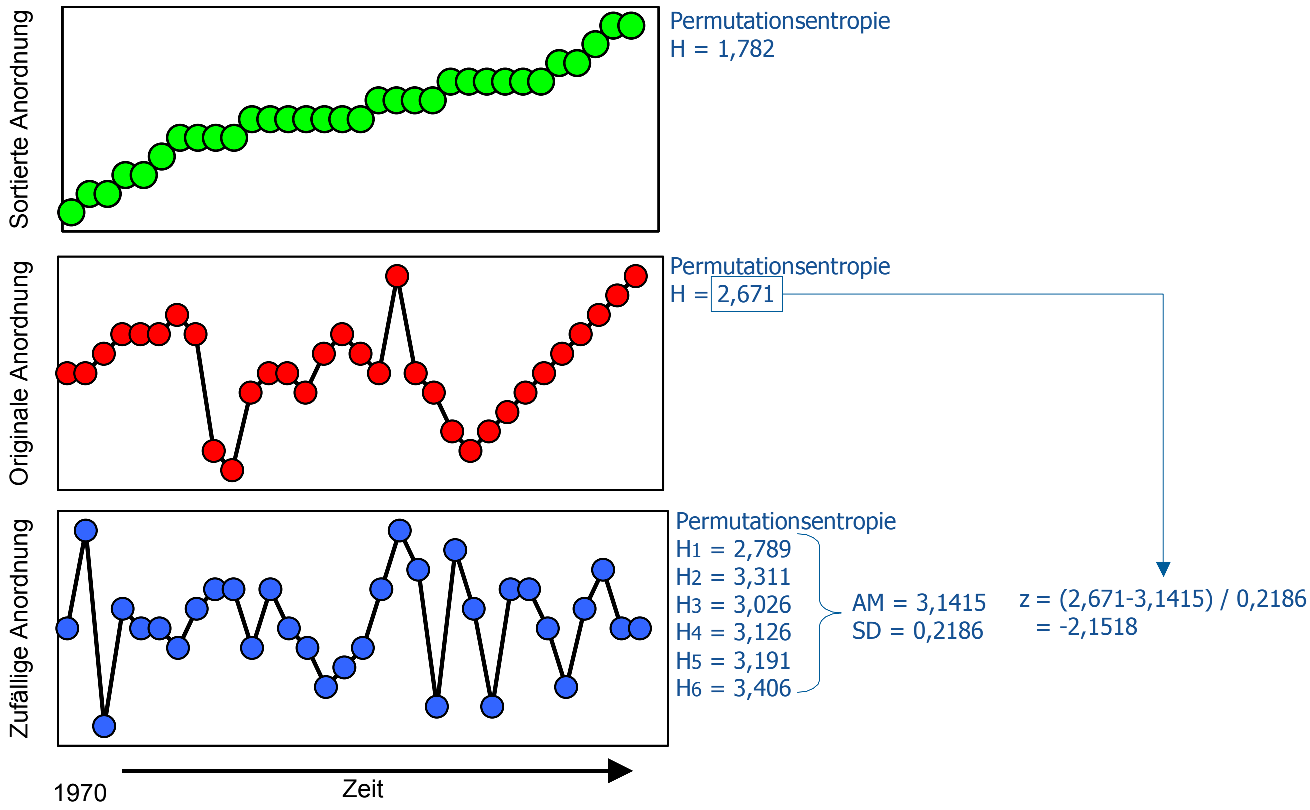
Rössler (x; a=b=0,2, c=5,7)
D2 = 1,991
LLE = 0,071 (Basis e)

Moving Window

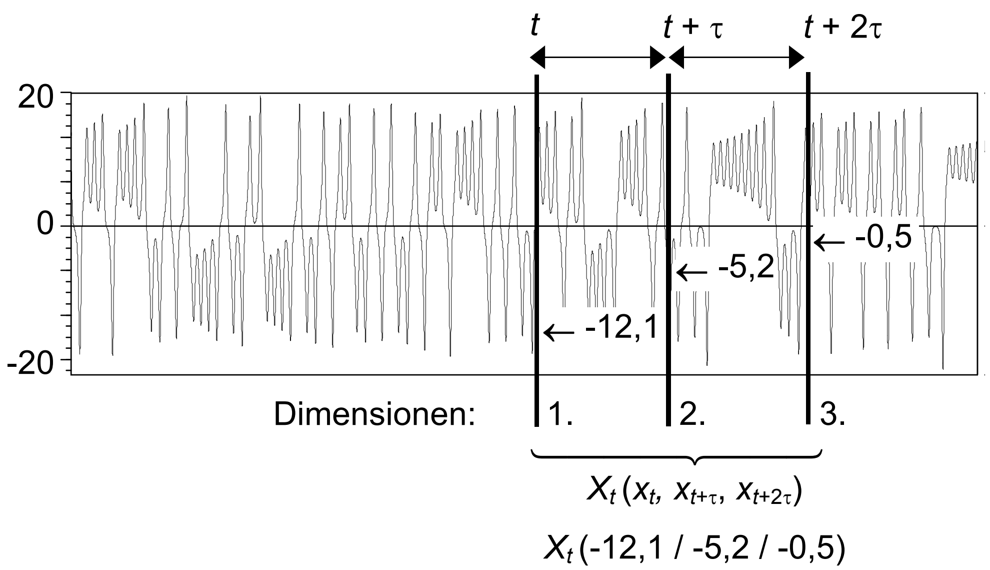


2.2 Surrogatdatenverfahren



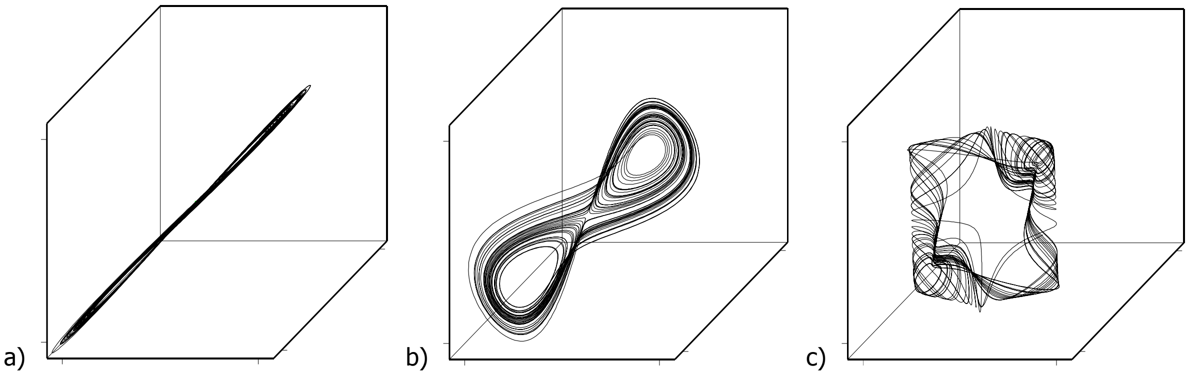


3. Phasenraumeinbettung



Schematische Darstellung der Generierung von Zeitverzögerungskoordinaten

Mehrdimensionale Phasenraumdarstellungen lassen sich auf der Grundlage nur einer Zeitreihe (hier ist es eine Zeitreihe des Lorenz-Systems) erzeugen. Für eine 3-dimensionale Einbettung wird nun ein Messwert x bei t, einer bei $t + \tau$ und einer bei $t + 2\tau$ abgelesen und in dieser Reihenfolge den drei Dimensionen zugeordnet. Dieses Vorgehen wird nacheinander für jeden Messzeitpunkt, also für alle t wiederholt. Das Verfahren ist nicht auf drei Dimensionen begrenzt. Die Komponente für eine vierte Dimension ist nach dem gleichen Prinzip gegeben durch $t + 3\tau$ etc. Für eine „realgetreue" Einbettung des Systems über dieses Verfahren gilt es zunächst, eine passende Zeitverzögerung τ zu bestimmen und festzulegen, wie viele Dimensionen insgesamt betrachtet werden sollen.



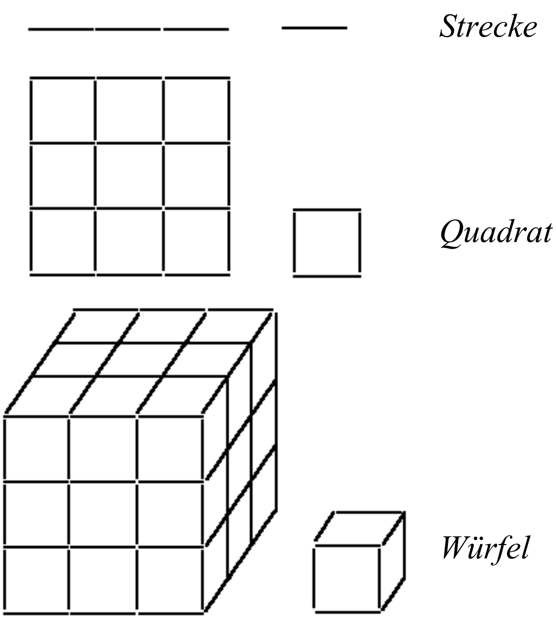
Phasenraumeinbettung des Lorenz-Attraktors für verschiedene Time-Lags

Das Time-Lag in a) ist mit einem Wert von eins eindeutig zu klein gewählt. Der Attraktor kann sich nicht entfalten und gruppiert sich im Wesentlichen entlang der Raumdiagonalen. Das Time-Lag in b) ist optimal gewählt. Es hat einen Wert von sechs. Das Time-Lag in c) ist mit einem Wert von 18 eindeutig zu hoch gewählt. Der Attraktor „kollabiert".

4. Grundprinzipien gängiger Verfahren

4.1 Fraktale Geometrie – D2 / PD2

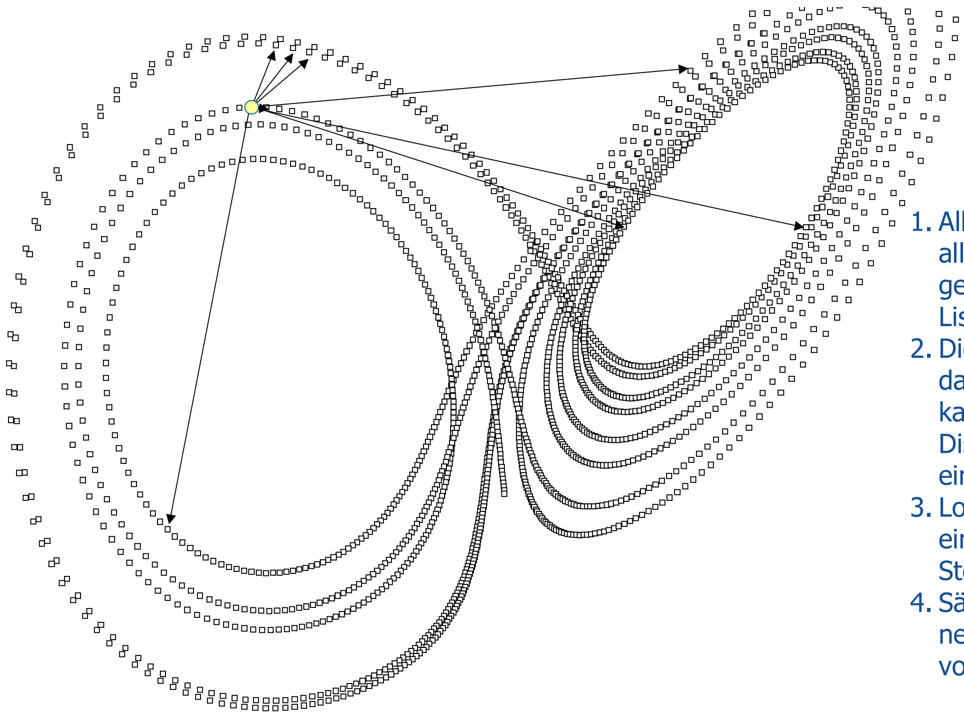
Dimensionalität geometrischer Objekte



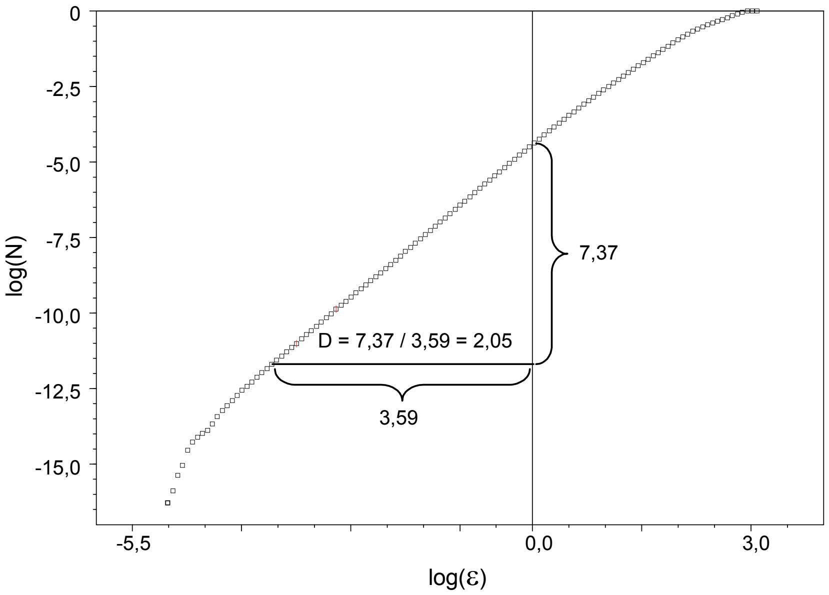
$$N = \varepsilon^{D}$$

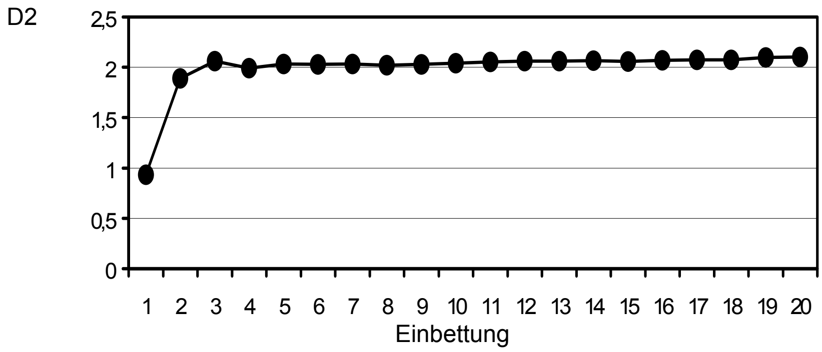
ε: Verkleinerungsfaktor des Vergleichskörpers
D: Dimensionalität
N: Anzahl der Vergleichskörper

Korrelationsdimension – Beispiel: Lorenz-System

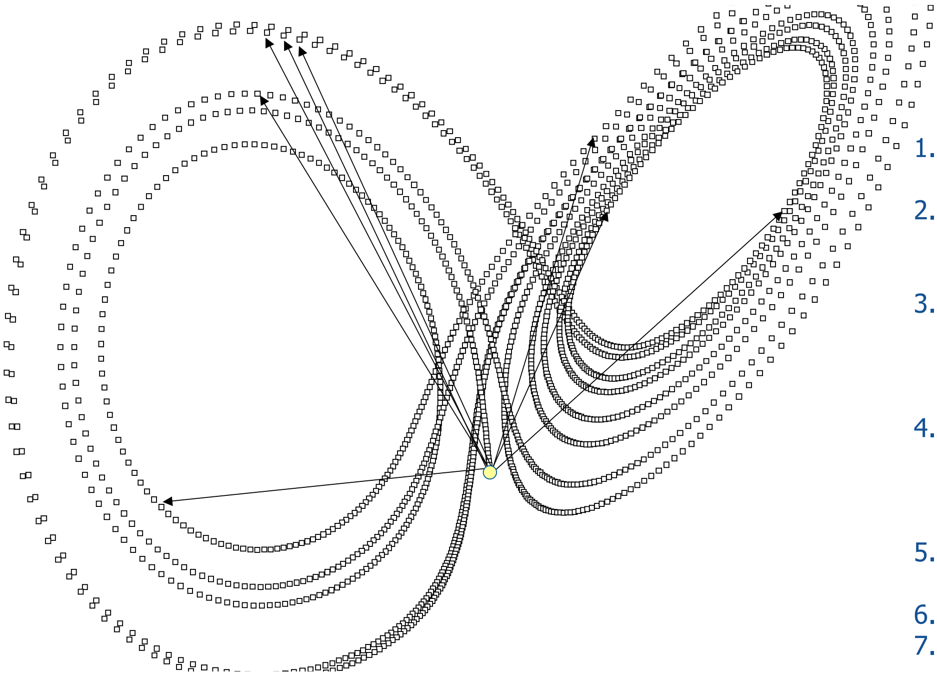


1. Alle Abstände zwischen allen Punkten werden gemessen und in einer Liste notiert.
2. Die Liste wird sortiert, so dass man leicht zählen kann (N) wie viele Distanzen kleiner sind als eine Vergleichsdistanz (ε).
3. Log(N) und Log(ε) bilden eine Grafik deren Steigung D2 heißt.
4. Sättigung für verschiedene Einbettungen muss vorliegen.





- Das Time-Lag muss passen. (Variieren und Testen)
- Es müssen genügend Datenpunkte vorhanden sein, damit das Verfahren sinnvoll angewendet werden kann. (200 bei kleinem D2 eventuell ausreichend; 1000 durchaus üblich als Minimum; bei hohem D2 sollte N>10(D2/2) sein.
- Surrogatdatentests sind erforderlich.



1. Jeder Datenpunkt ist einmal Fokuspunkt.
2. Die Abstände zu den anderen Punkten werden gemessen und in einer Liste notiert.
3. Die Liste wird sortiert, so dass man leicht zählen kann (N) wie viele Distanzen kleiner sind als eine Vergleichsdistanz (ε).
4. Log(N) und Log(ε) bilden eine Grafik deren Steigung PD2 heißt und lokal für den Fokuspunkt gilt.
5. Sättigung für verschiedene Einbettungen muss vorliegen.
6. PD2 für Fokuspunkt speichern.
7. Nächster Fokuspunkt.

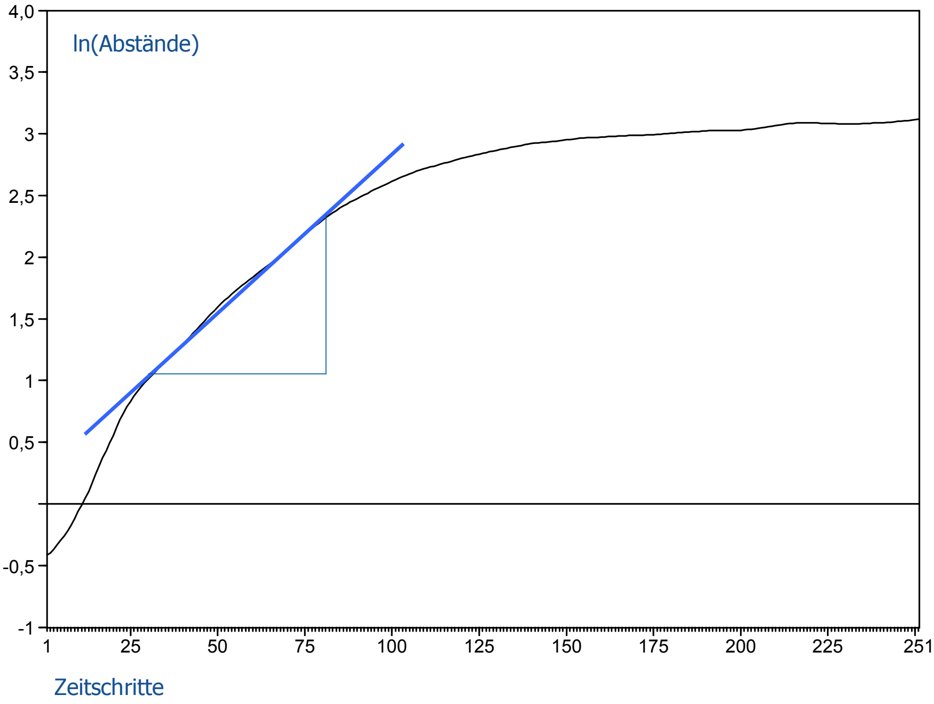
4.2 Schmetterlingseffekt – Lyapunov-Exponent LLE

- Chaos ist weder einfache Ordnung noch Zufall. Chaos ist also der Prototyp für Komplexität. Das einzige Verfahren zum direkten Nachweis von Chaos ist der Lyapunov Exponent (LE).
- Ist der größte LE (LLE) positiv, liegt Chaos vor.

LLE – Rosenstein-Algorithmus

1. Jeder Datenpunkt ist einmal Fokuspunkt
2. Ausschluss von Vorgängern und Nachfolgern um den Fokus
3. Suche nächsten Nachbarn
4. Nachbar wird für Zukunft ausgeschlossen
5. Verfolge beide Punkte eine Zeit lang und registriere die Abstände
6. Wähle Fokuspunkt direkt nach dem letzten Fokuspunkt und beginne erneut mit 2
7. Wenn alle Punkte bearbeitet sind, mittel die Abstände

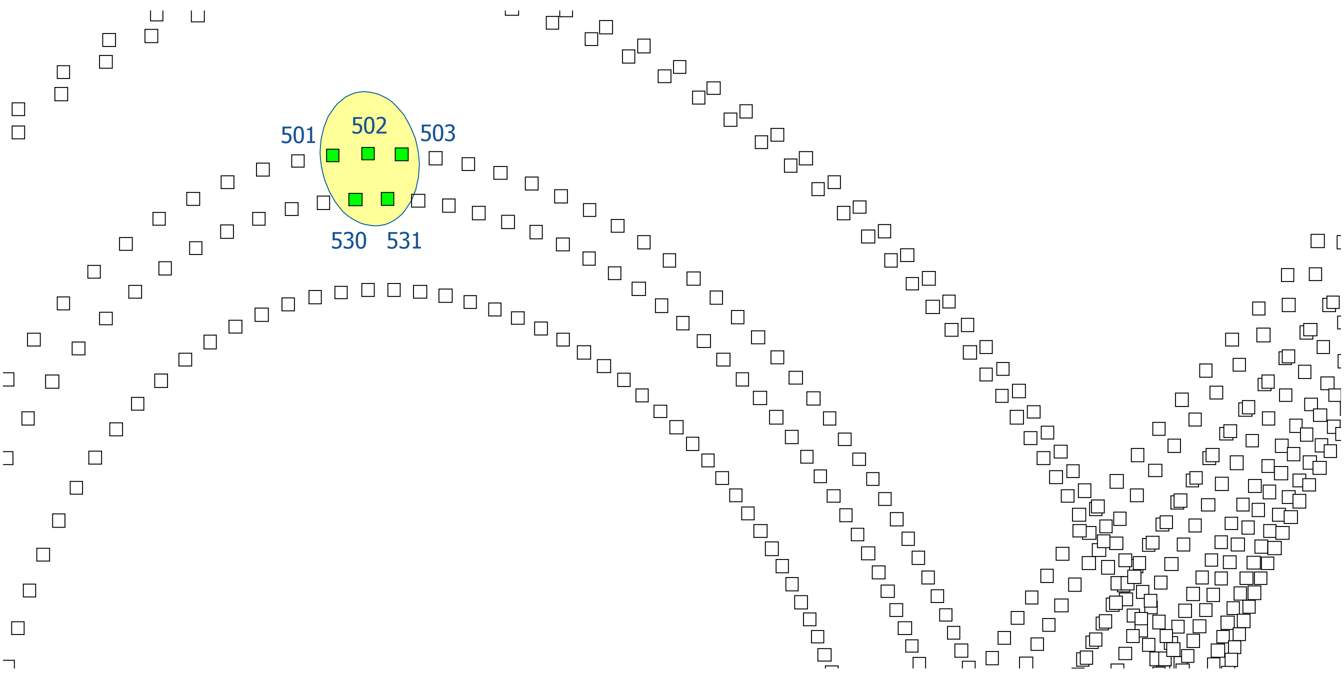
1. Jeder Datenpunkt ist einmal Fokuspunkt
2. Ausschluss von Vorgängern und Nachfolgern um den Fokus
3. Suche nächsten Nachbarn
4. Nachbar wird für Zukunft ausgeschlossen
5. Verfolge beide Punkte eine Zeit lang und registriere die Abstände
6. Wähle Fokuspunkt direkt nach dem letzten Fokuspunkt und beginne erneut mit 2
7. Wenn alle Punkte bearbeitet sind, mittel die Abstände
8. Logarithmiere die Abstände
9. Identifiziere linearen Abschnitt und bestimme die Steigung. Diese ist der LLE

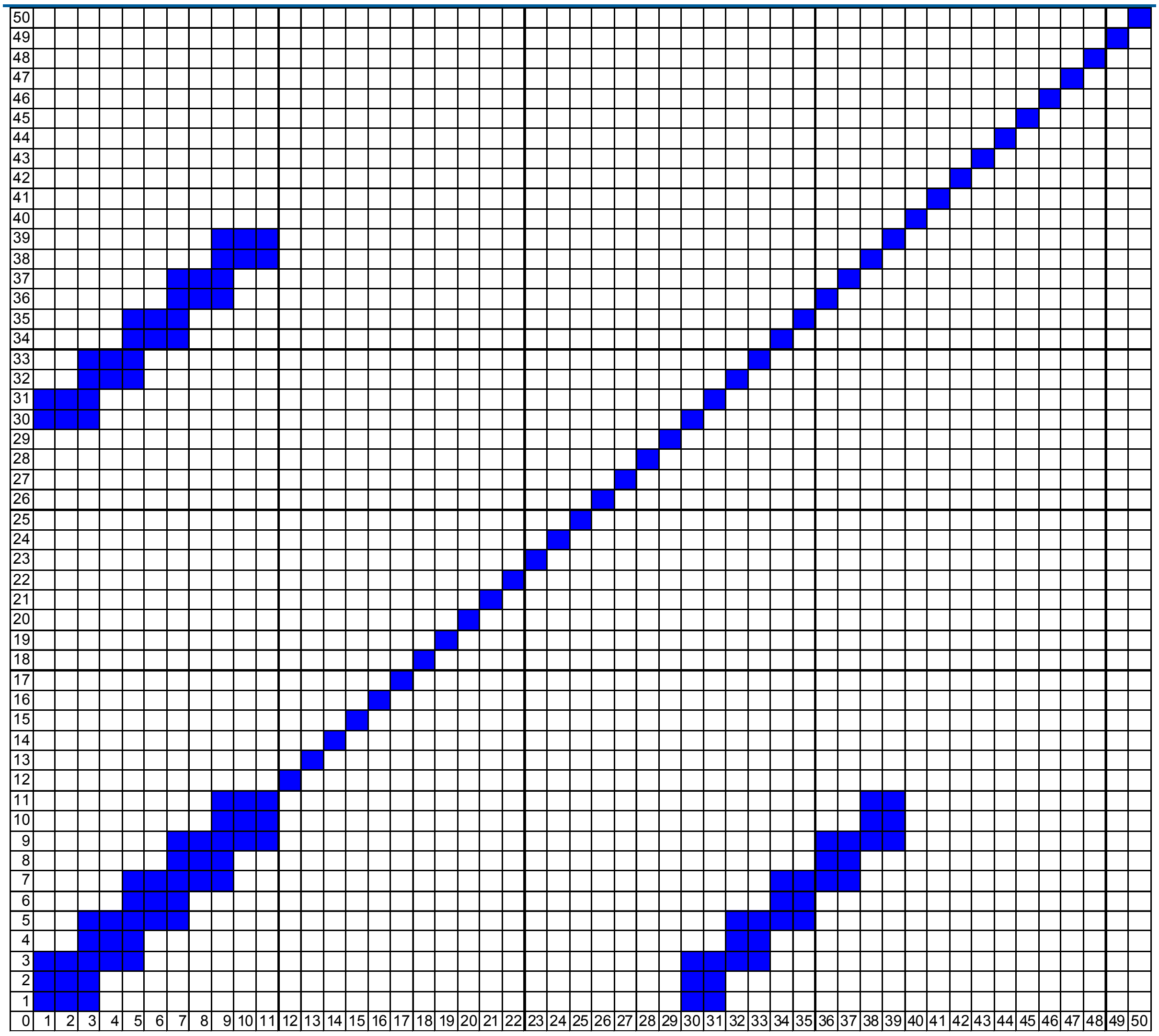


- Das Time-Lag muss passen. (Variieren und Testen)
- 100 bis 500 als Minimum. Ansonsten N>10D.
- Surrogatdatentests sind erforderlich.
- Gleitendes Fenster für die Erfassung nichtstationärer Entwicklungen.

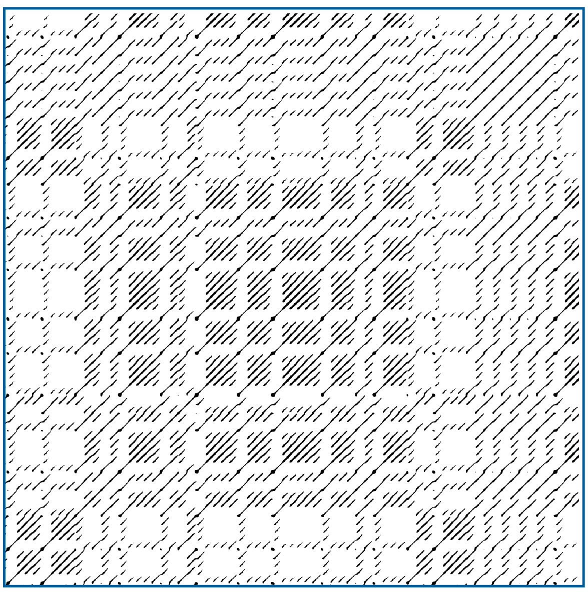
4.3 Beinahe gleich – Recurrence Plots

Sind zwei Datenpunkte einander sehr ähnlich, so wiederholt sich das System in seinem Verhalten. Solche Widerholungen werden gesucht, grafisch dargestellt und interpretiert.

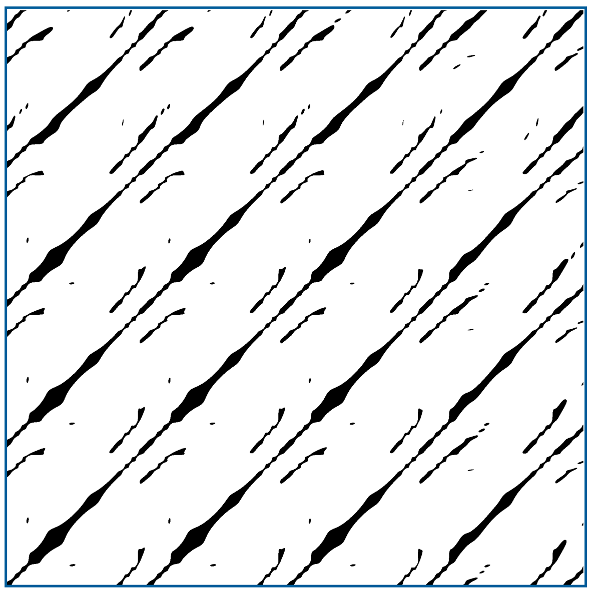




Lorenz-System

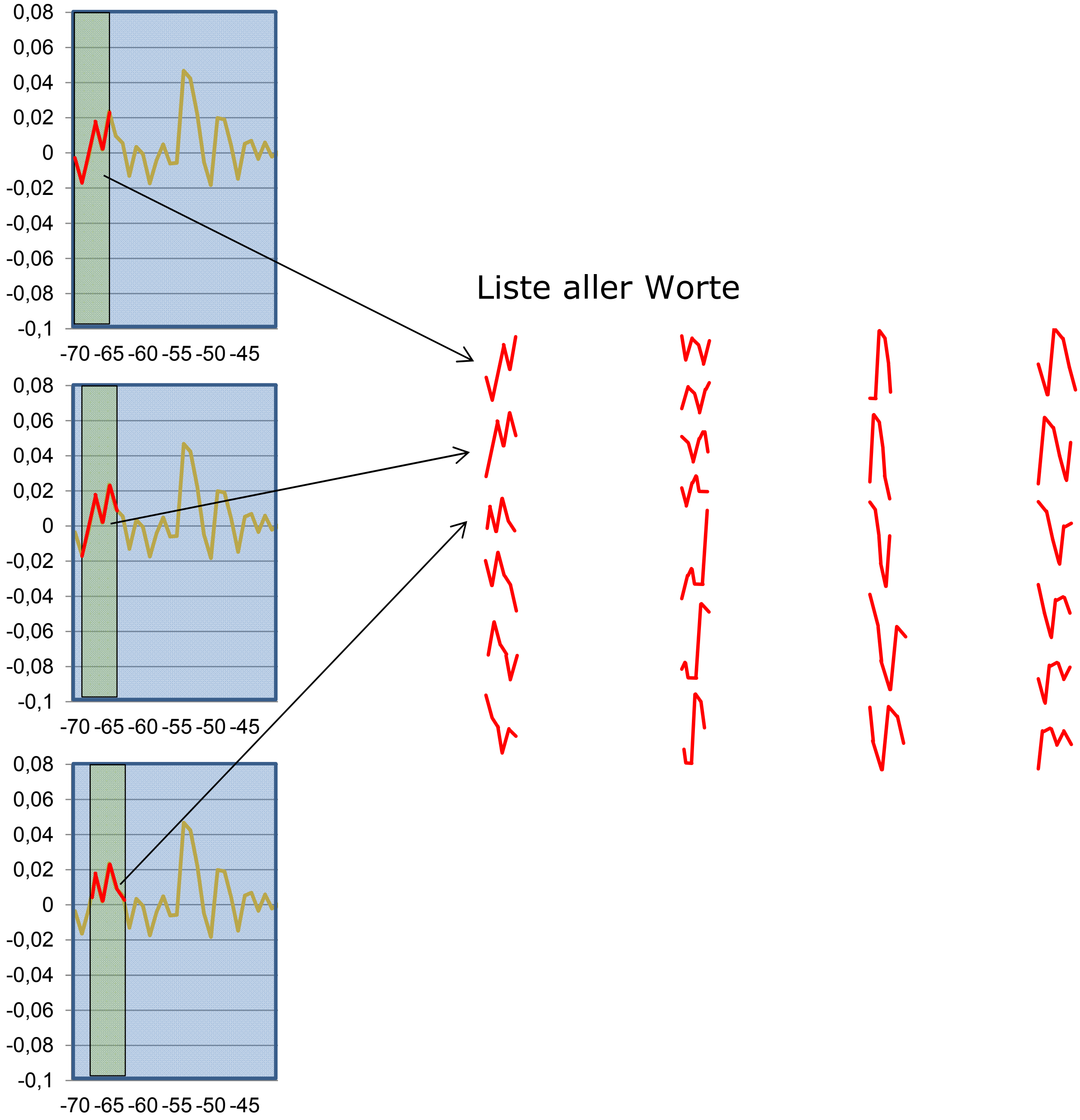


Rössler-System



- **Recurrence Rate.** Ein System, das sich häufig wiederholt ist geordnet: Die Recurrence Rate setzt die Anzahl der Recurrence Punkte in ein Verhältnis zur Anzahl möglicher Punkte.
- **Determinismus.** Längere Phasen der Musterwiederholung sind ein Zeichen für Ordnung: Da diagonale Linien auf solche deterministische Anteile hinweisen, ist der Determinismus definiert als Prozentsatz derjenigen Recurrence Punkte, die diagonale Linien bilden.
- **Divergenz.** Sobald der Schmetterlingseffekt die Trajektorien auseinandertreibt bricht die Diagonale ab. Der Kehrwert der längsten Diagonalen wird als Divergenz bezeichnet und steht in einem Zusammenhang mit dem LLE, ist aber nicht mit diesem identisch.
- Das Time-Lag muss passen. (Variieren und Testen)
- N: 30-50 als Minimum.
- Surrogatdatentests sind hilfreich.
- Gleitendes Fenster für die Erfassung nichtstationärer Entwicklungen.

4.4 Entropie-Maße: Permutationsentropie / Zentraltendenz-transformator



Die Kodierung der Worte kann auf verschiedene Weise erfolgen

Symbolic Dynamics (Rohwertcodierung)

Permutationsentropie (Rangdatencodierung)

GEntropie I (Runs-Test-Codierung)

GEntropie II (Mittelwerts-Codierung)

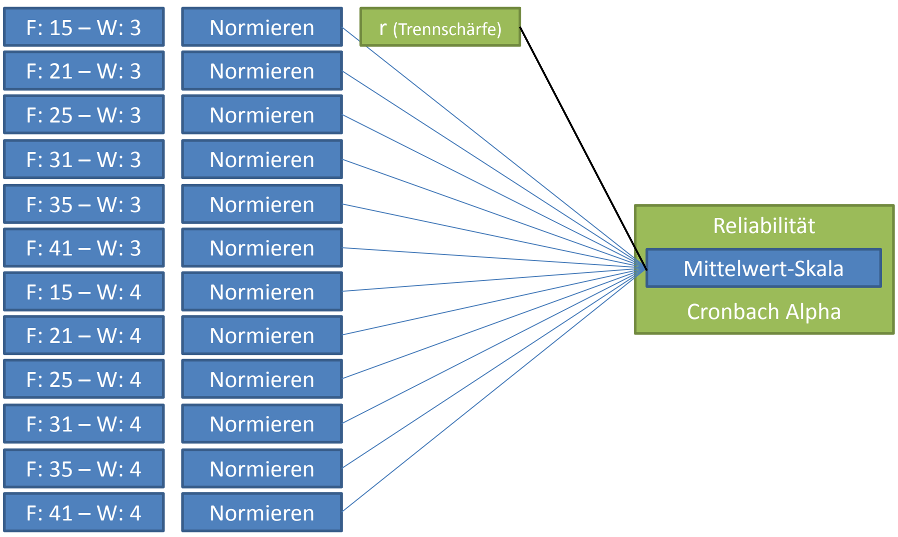
- Eine Gleichverteilung der codierten Worte würde auf Zufall hinweisen. Dieser wird in anderen Zusammenhängen auch als maximale Entropie bezeichnet.
- Möglichkeiten zur Quantifizierung einer so verstandenen Entropie wurden von Claude Shannon 1948 vorgeschlagen.
- Die Entropie ist die mit der Auftretenswahrscheinlichkeit P gewichtete Summe der Informationsgehalte I der einzelnen Worte.

$$I_s = -\sum_{i=1}^{N} P(s_i) \log P(s_i).$$

Der Informationsgehalt ist gegeben durch:

$I(w_i) = -\log_2 P(w_i)$

Beispiele:
$P \to 0$ dann $I \to \infty$
$P \to 1$ dann $I \to 0$



- Kein Time-Lag erforderlich aber Fensterbreite und Wortlänge müssen passen.
- Fensterbreite bzw. N: 20-30 als Minimum.
- Surrogatdatentests sind hilfreich. Bootstrapping ist möglich.
- Gleitendes Fenster für die Erfassung nichtstationärer Entwicklungen.

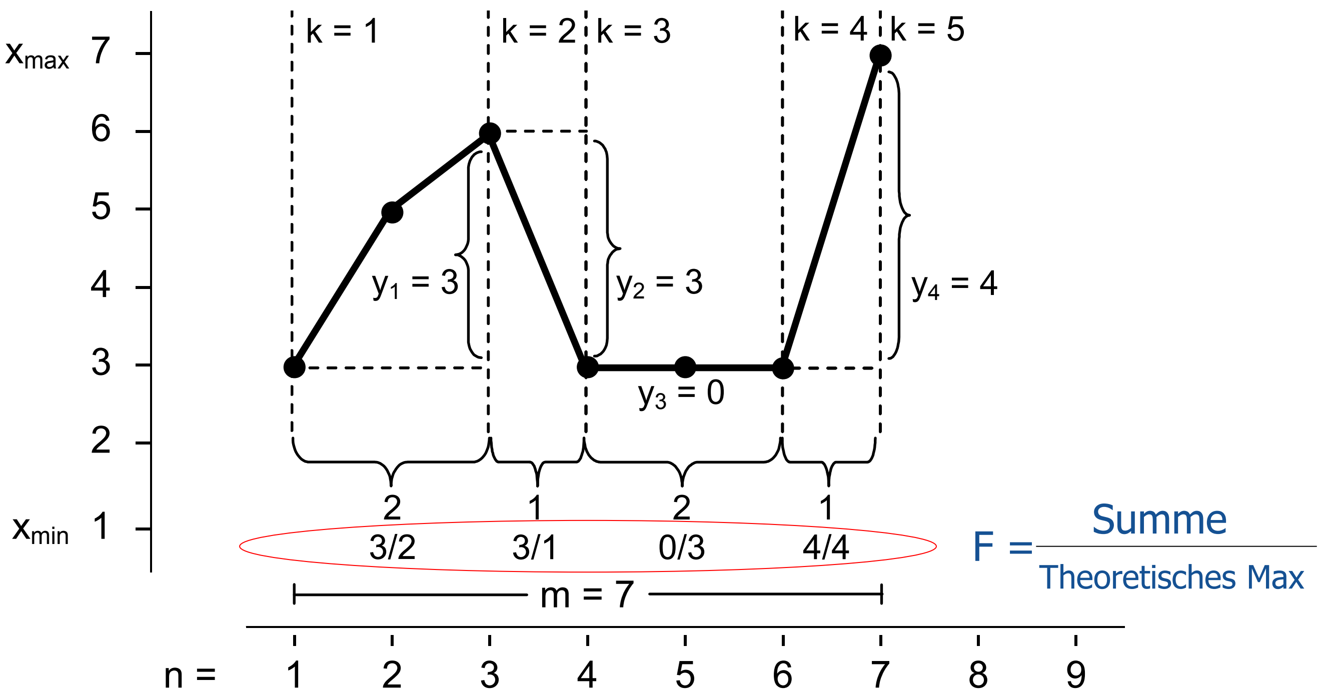
Anmerkungen zur Datenmenge

- Allgemein gilt: Wenn eine triviale Ordnung vorliegt ist die häufig auch mit kleinen Datensätzen nachweisbar.
    - Wenn der Nachweis gelingt, dann ist Ordnung da.
    - Kann die Ordnung auch durch Zufall entstanden sein? Bei weniger Daten ist das auch durch Zufall möglich.

- Wenn bei gegebener Datenmenge keine Ordnung gefunden werden kann, heißt das nicht, dass keine vorliegt.
    - Jedes Verfahren findet nur die Ordnung, nach der es sucht. Liegt eine andere vor, kann diese nicht entdeckt werden.
    - Eine triviale periodische Abfolge von Messwerten kann dann nicht entdeckt werden, wenn weniger Messwerte vorliegen als die Periode dauert.

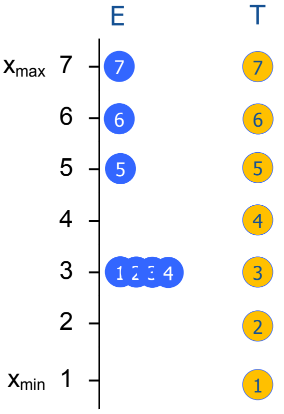
4.5 Dynamische Komplexität (Schiepek)

4.5.1 Fluktuation – F



F wird groß bei vielen hohen Sprüngen in kurzer Zeit.

4.5.2 Verteilung – D



Abstände zwischen theoretisch verteilten Datenpaaren:
$\delta(2\text{-}1) = T2 - T1 = 2 - 1 = 1$
Allgemein:
$\delta(a\text{-}b) = Ta - Tb$

Abstände zwischen empirisch verteilten Datenpaaren:
$\delta'(2\text{-}1) = E2 - E1 = 3 - 3 = 0$
Allgemein:
$\delta'(a\text{-}b) = Ea - Eb$

Abweichung:
$\Delta(a\text{-}b) = \delta(a\text{-}b) - \delta'(a\text{-}b)$

Abweichung:
$\Delta(a\text{-}b) = \delta(a\text{-}b) - \delta'(a\text{-}b)$

Normierte Abweichung:
$A(a\text{-}b) = \Delta(a\text{-}b) / \delta(a\text{-}b)$

Kombinatorik von a, b und Summe aller positiven A:

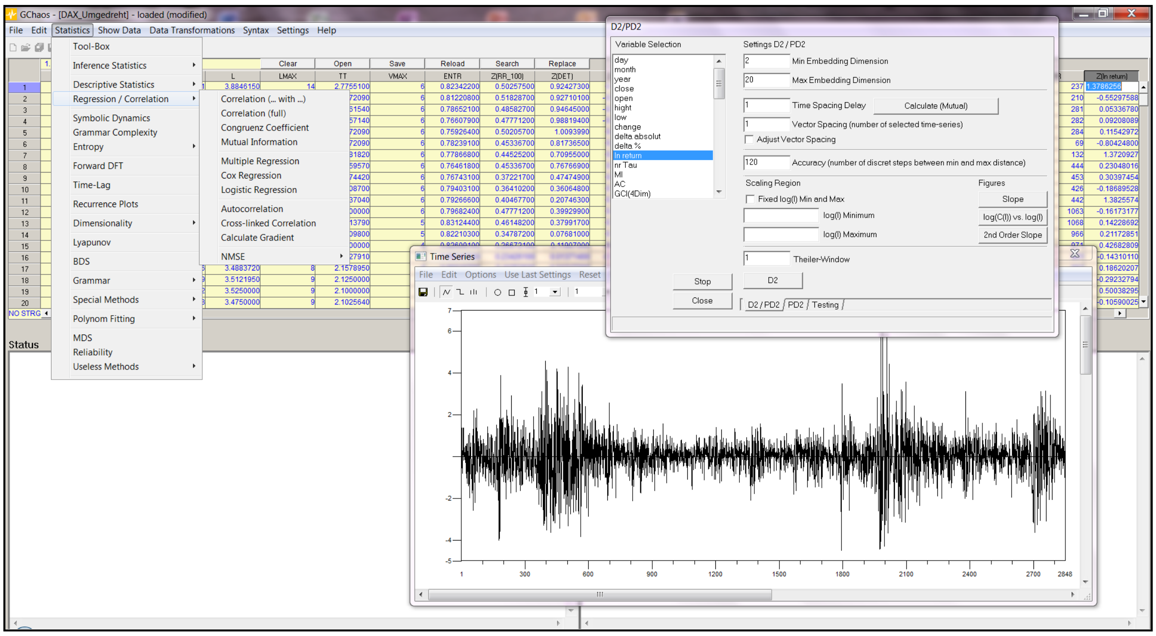
$$D = 1 - \sum A^{+}(a\text{-}b)$$

D wird groß wenn die Skala in voller Breite gleichmäßig ausgenutzt wird. Die Reihenfolge der Werte im Zeitverlauf spielt keine Rolle.

- F – Fluktuation: wird groß bei vielen hohen Sprüngen in kurzer Zeit.
- D – Verteilung: wird groß wenn die Skala in voller Breite gleichmäßig ausgenutzt wird.
- V = F * D
- Selbst für kleine Fenster (N = 7) klinisch relevante Hinweise auf Musterveränderungen (Ordnungsübergänge, die sich durch kritische Fluktuationen auszeichnen).
- Validitätsstudie in BC (Schiepek & Strunk 2010) zeigt die Validität für künstliche Daten. Gruppenstatistik.
- Kein Time-Lag erforderlich.
- Fensterbreite bzw. N: 7 als Minimum.
- Surrogatdatentests sind hilfreich.
- Gleitendes Fenster für die Erfassung nichtstationärer Entwicklungen.
- Misst **Unordnung** bei gegebener Skalenbreite und Fensterbreite.

5. Software

GChaos – http://www.complexity-research.com/DownSoft.htm



6. Literatur

Grundproblem. Wie kann man die Komplexität einer Dynamik bestimmen?
Abgrenzung und Messung von Komplexität

Strunk G (2009) *Die Komplexitätshypothese der Karriereforschung*. Peter Lang, Frankfurt

Strunk G (2009) Operationalizing Career Complexity. *Management Revue,* 20 (3), S. 294-311

Stationarität?

Strunk G, Haken H & Schiepek G (2006) *Ordnung und Ordnungswandel in der therapeutischen Kommunikation*. In: Haken H & Schiepek G (Hrsg) *Synergetik in der Psychologie. Selbstorganisation verstehen und gestalten*. Hogrefe, Göttingen, S. 462-517

Surrogatdatenverfahren

Strunk G & Schiepek G (2006) *Systemische Psychologie*. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens. Spektrum Akademischer Verlag, München

Phasenraumeinbettung

Strunk G & Schiepek G (2006) *Systemische Psychologie*. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens. Spektrum Akademischer Verlag, München

Fraktale Geometrie – D2 / PD2
Schmetterlingseffekt – Lyapunov-Exponent LLE

Strunk G, Haken H & Schiepek G (2006) Ordnung und Ordnungswandel in der therapeutischen Kommunikation. In: Haken H & Schiepek G (Hrsg) *Synergetik in der Psychologie. Selbstorganisation verstehen und gestalten*. Hogrefe, Göttingen, S. 462-517

Strunk G & Schiepek G (2006) Systemische Psychologie. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens. Spektrum Akademischer Verlag, München

Beinahe gleich – Recurrence Plots
Entropie-Maße: Shannon / Permutationsentropie / Zentraltendenztransformator

Strunk G (2009) *Die Komplexitätshypothese der Karriereforschung*. Peter Lang, Frankfurt

Dynamische Komplexität

Schiepek, G. & Strunk, G. (2010) The identification of critical fluctuations and phase transitions in short term and coarse-grained time series-a method for the real-time monitoring of human change processes. *Biological Cybernetics*, 102(3), 197-207

Software und detaillierte Beschreibung der Algorithmen

Strunk, G. (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft. Grundlagen, Methoden, Software und Beispiele (*derzeit noch* unveröffentlichte Habilitationsschrift). Dortmund: TU Dortmund