

Vorlesungen

„Einführung in die Datenvisualisierung“

„Angewandte Datenvisualisierung für
Medizinphysiker“

Kapitel I Visualisierung Zeit-abhängiger Daten

Dr. Frank Weichert
Informatik VII
Technische Universität Dortmund
<http://ls7-www.cs.uni-dortmund.de>

Nur zum persönlichen Gebrauch durch Mitglieder der Universität Dortmund

I.1. Zeitreihen

- I.1.1. Einleitung
- I.1.2. Beispiele
- I.1.3. Einfache lineare Regression
- I.1.4. Ziele der Analyse

I.2. Prognosetechniken und Prognosefehler

- I.2.1. Einleitung
- I.2.2. Qualitative Verfahren
- I.2.3. Quantitative Verfahren
- I.2.4. Szenarioanalyse
- I.2.5. Güte der Prognose

I.3. Komponentenmodelle

- I.3.1. Einführung
- I.3.2. Glatte Komponente
- I.3.3. Saisonkomponente
- I.3.4. Analyse der Residuen
- I.3.5. Vorhersage

I.4. Interpolation und Approximation

- I.4.1. Einleitung
- I.4.2. Interpolation
- I.4.3. Approximation

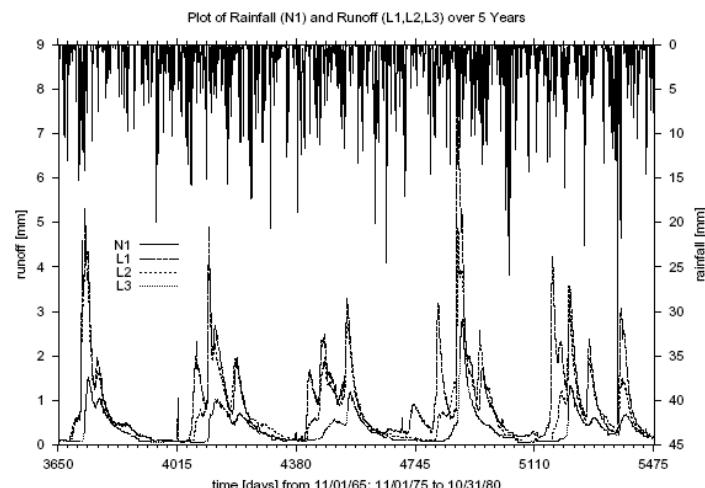
I.5. Funktionale Analyse

- I.5.1. Einleitung
- I.5.2. Histogramme
- I.5.3. Statistische Merkmale
- I.5.4. (Diskrete) Fourier-Transformation
- I.5.5. Wavelet-Transformation
- I.5.6. Phasenraumdarstellung und –einbettung

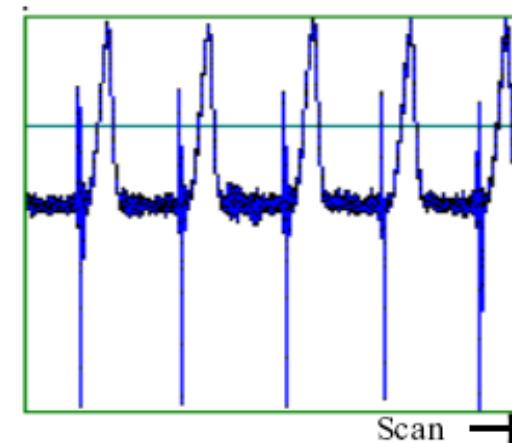
Zeitreihen: (zeitlich) geordnete Folge von Daten (Beobachtungen). Zentral ist, dass die Reihenfolge der Messungen relevant ist und sukzessive Folgenglieder im Allgemeinen nicht als unabhängig betrachtet werden dürfen.

Analyseansätze

- **Zeitbereich:** Betrachtung der zeitlichen Abfolge der Werte über die Zeit hinweg.
Insbesondere Regressions- und Zustandsraummodelle
- **Frequenzbereich:** Betrachtung der Zeitreihe als Überlagerung von Schwingungen mit verschiedenen Frequenzen (z.B. Fourieranalyse)

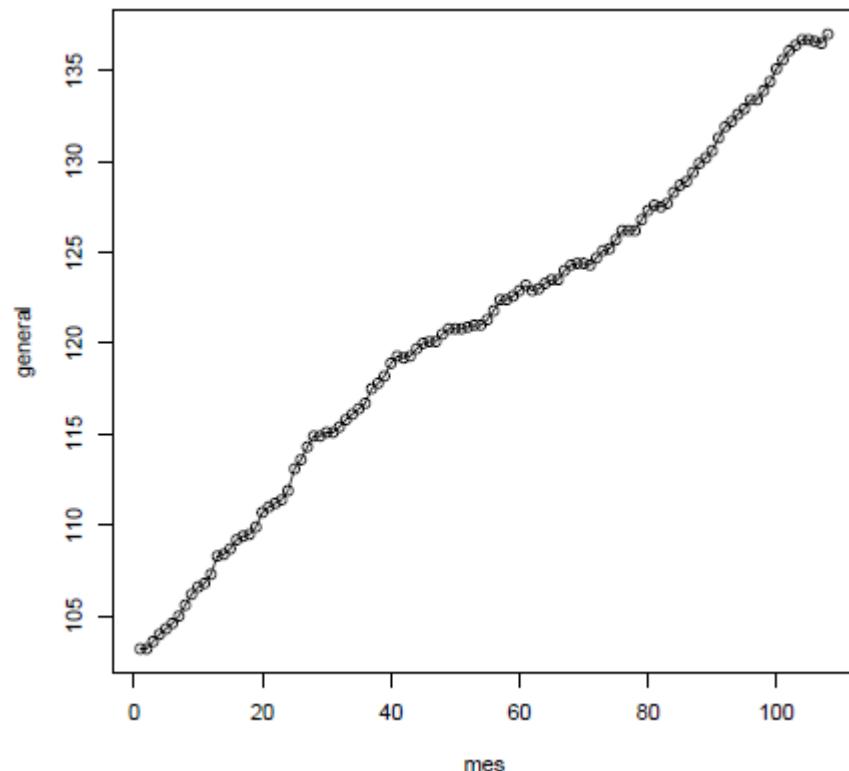


Daten eines ökologischen Systems

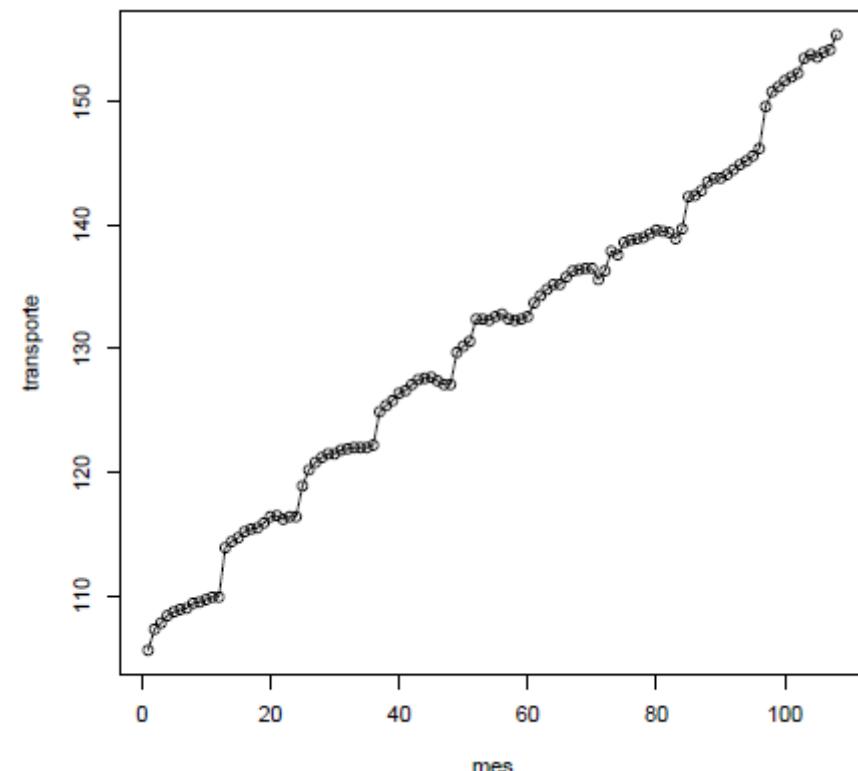


Elektrokardiogramm (EKG)

Beispiele: Preisindizes in Spanien, 1/1993–12/2001 (www.ine.es)



grob gesehen linear
(Globalindex)

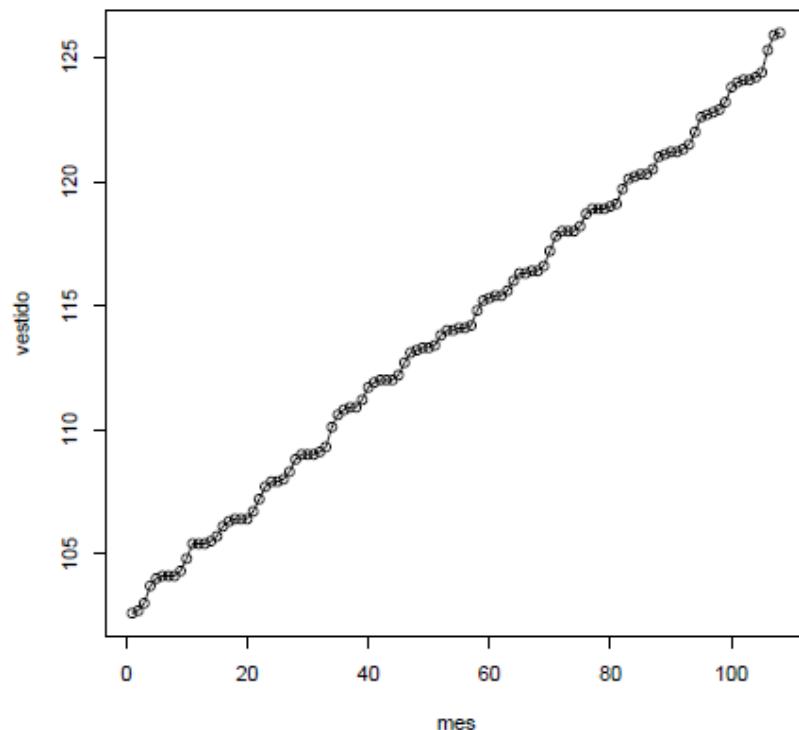


Anstieg mit Preissprüngen
(Transport)

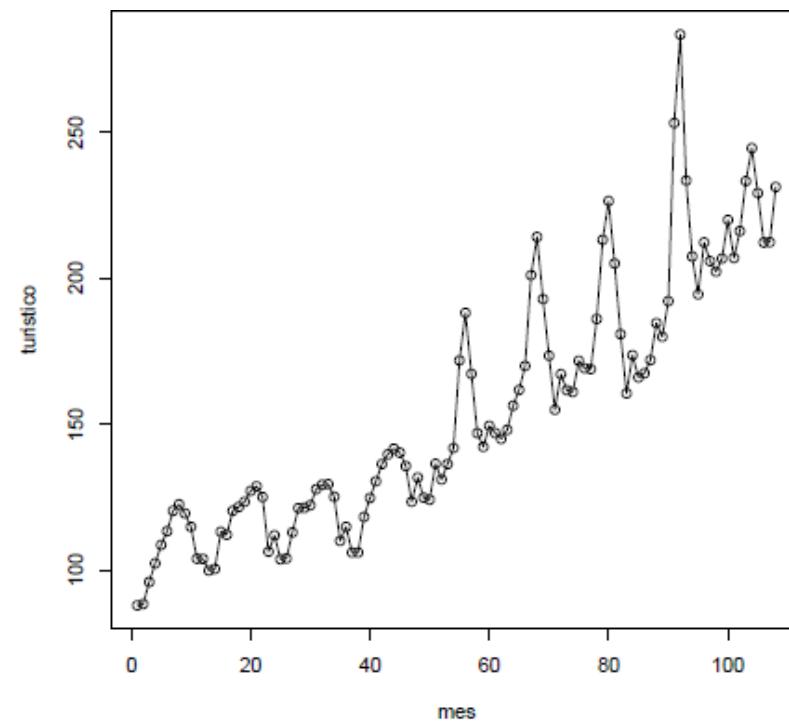
I.1. Zeitreihen

I.1.2. Beispiele

Beispiele: Preisindizes in Spanien, 1/1993–12/2001 (www.ine.es)



linearer Anstieg mit Saisonfigur
(Kleidung)

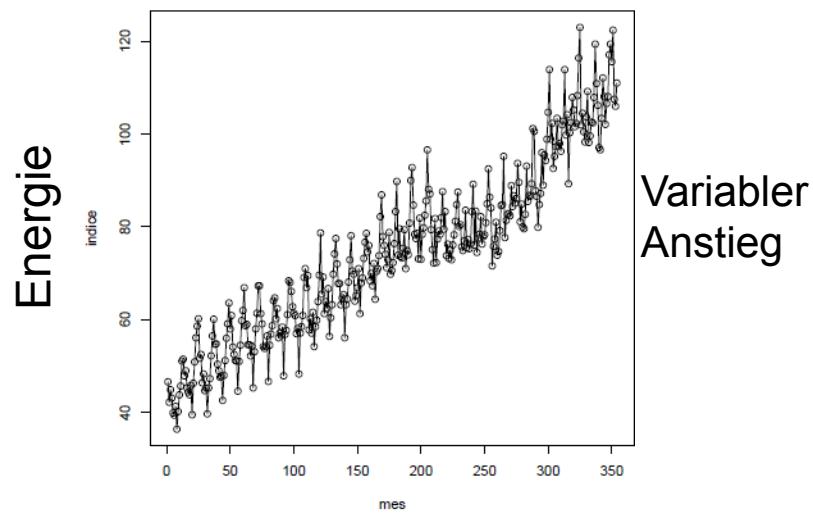


Anstieg, variable Saisonfigur
(Tourismus)

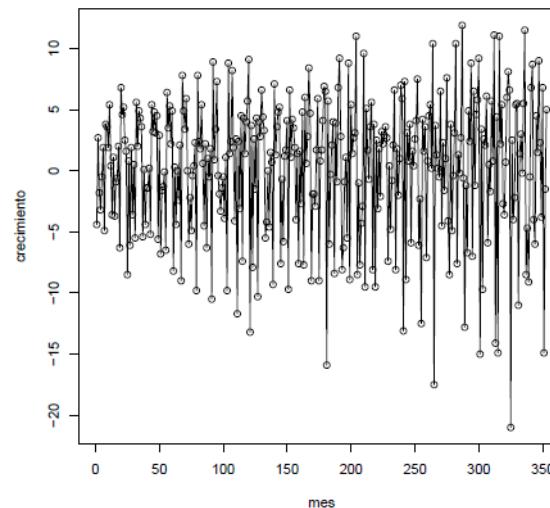
I.1. Zeitreihen

I.1.2. Beispiele

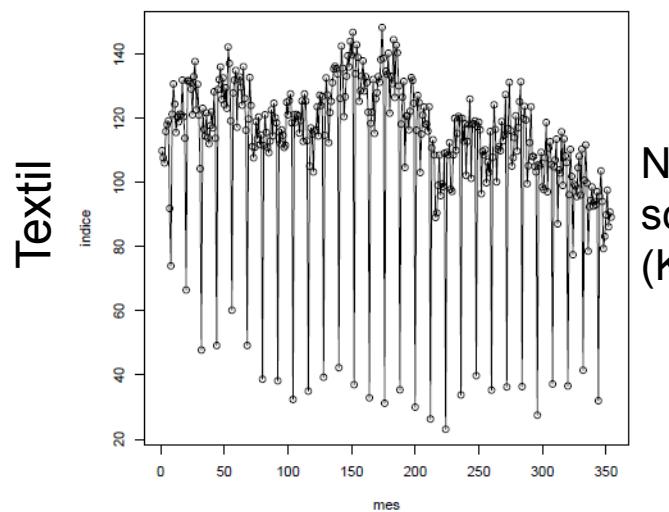
Beispiele: Produktionsindizes (links) und Wachstumsraten (Differenzen, rechts)



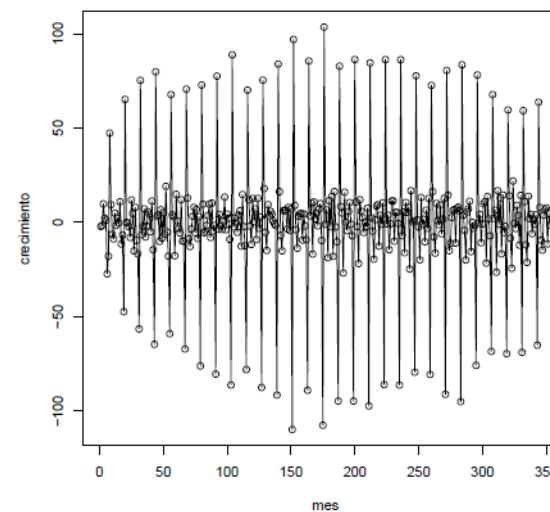
Variabler
Anstieg



Stabiles
Niveau



Niveau
schwankend
(Kalendereffekt)

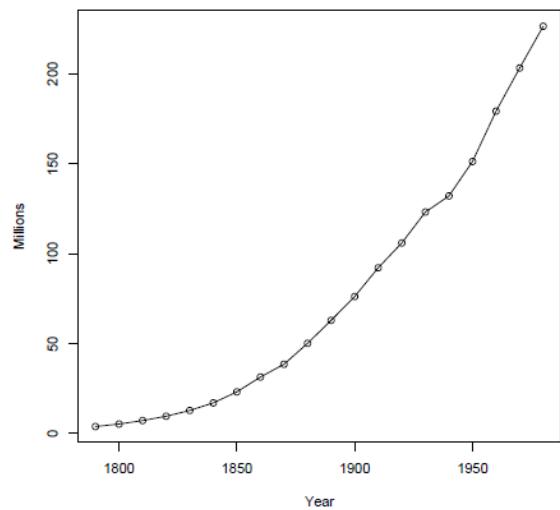


Stabiles
Niveau
&
Kalendereffekt

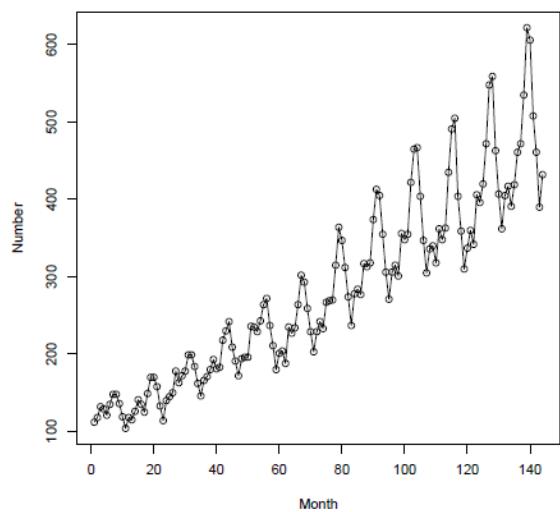
I.1. Zeitreihen

I.1.2. Beispiele

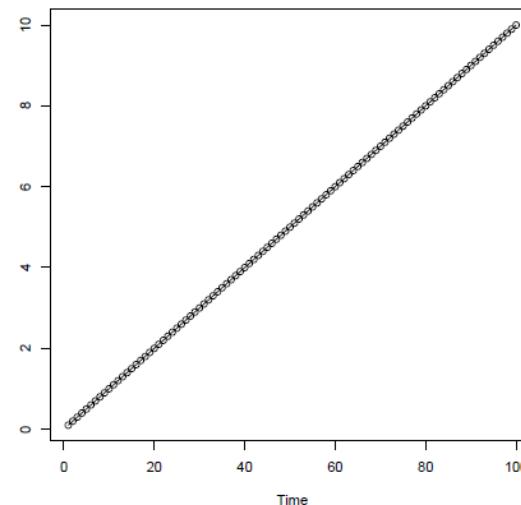
Beispiele: Deterministischer Trend und Saison



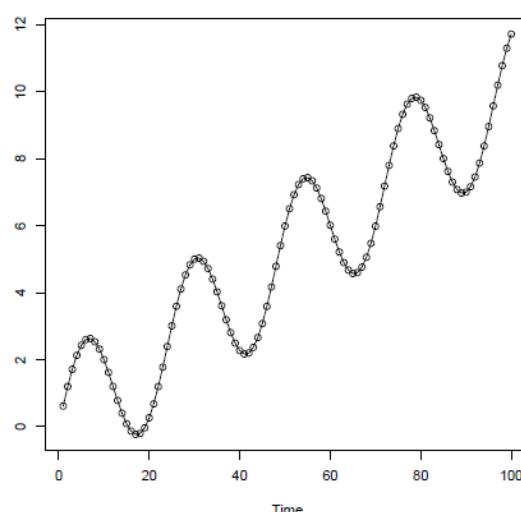
Bevölkerung
der USA



Flugpassagiere
(weltweit)



Linearer
Trend



Linearer
Trend
&
sinusförmige
Saisonfigur

Die Regressionsanalyse beschäftigt sich mit der Schätzung funktionaler Beziehungen zwischen zwei oder mehr metrisch skalierten Merkmalen.

Im Folgenden: zweidimensionale Datensätze → „**Einfache Regression**“

Es wird unterstellt, dass eine metrische Merkmal (die unabhängige Variable, im Folgenden immer mit x bezeichnet) beeinflusst das andere metrische Merkmal (die abhängige Variable, im Folgenden immer mit y bezeichnet).

Gesucht ist also die Funktion $y=f(x)$ durch welche die gegebenen n Wertepaare (x_i, y_i) generiert werden.

Bei der **linearen Regression** wird angenommen, die gesuchte Funktion sei linear von der Form

$$y = a + b \cdot x$$

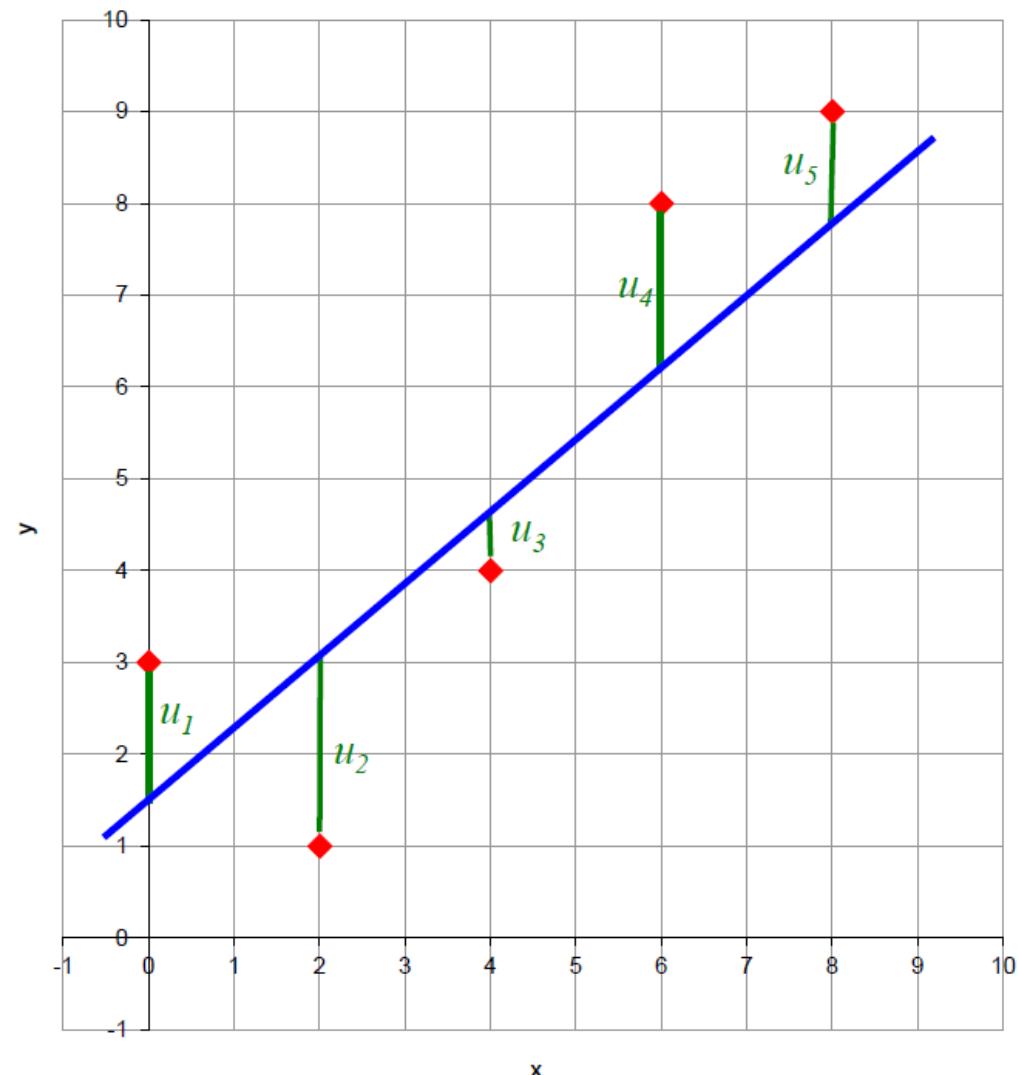
Die vorliegenden Wertepaare erfüllen diese Beziehung in der Regel nicht exakt, d. h. es gibt Abweichungen

$$u_i = y_i - (a + b \cdot x_i), \quad i = 1, \dots, n$$

Die **lineare Einfachregression** läuft also darauf hinaus, die durch den Datensatz gegebene **Punktewolke** im Streudiagramm durch eine **Gerade** so anzunähern, dass die Abweichungen u_i möglichst gering sind.

Beispiel

i	x_i	y_i
1	0	3
2	2	1
3	4	4
4	6	8
5	8	9



Methode der kleinsten Quadrate nach Gauß

Die Koeffizienten a und b der Regressionsgeraden $y = a + b \cdot x$ sind so zu wählen, dass die Summe der quadratischen Abweichungen

$$Q(a, b) = \sum_{i=1}^n u_i^2 = \sum_{i=1}^n [y_i - (a + b \cdot x_i)]^2$$

minimiert wird.

Unter der Voraussetzung $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$

ist die Lösung dieser Minimierungsaufgabe eindeutig und lautet:

$$b = \left(\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})] \right) \Bigg/ \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) = \frac{s_{xy}}{s_x^2}$$

$$a = \bar{y} - b \cdot \bar{x}$$

oder alternativ:

$$b = \left(n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i \right) \Bigg/ \left(n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right)$$

$$a = \bar{y} - b \cdot \bar{x}$$

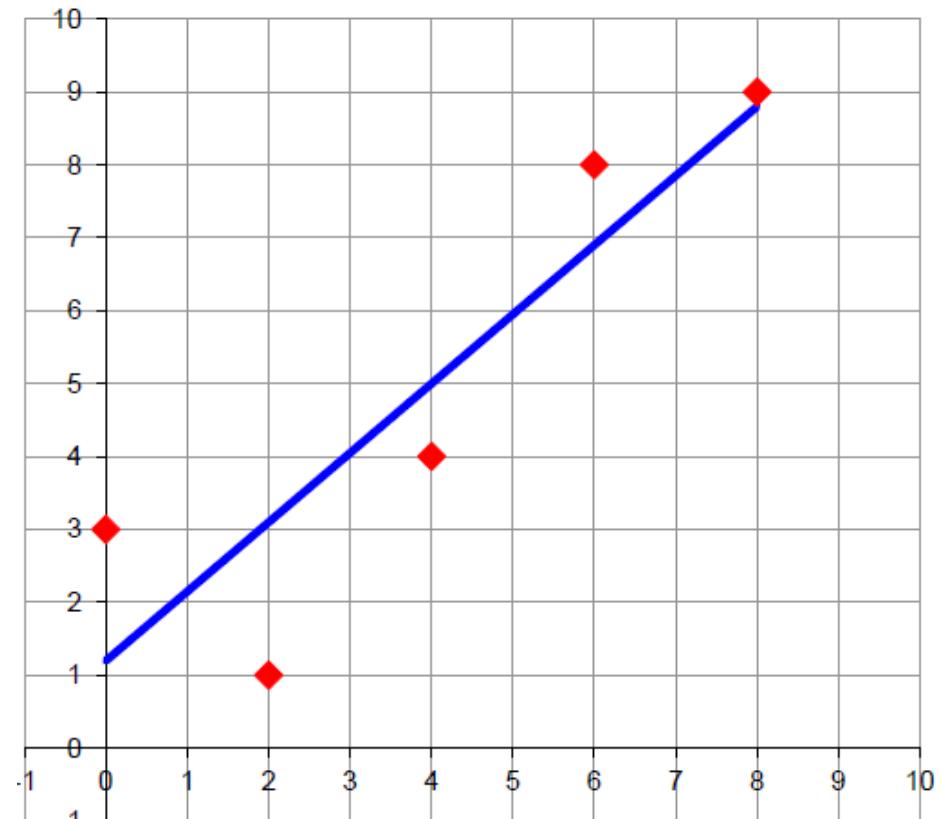
Beispiel

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	0	3	-4	-2	8	16
2	2	1	-2	-4	8	4
3	4	4	0	-1	0	0
4	6	8	2	3	6	4
5	8	9	4	4	16	16
Σ	20	25			38	40
AM	4	5			7,6	8

$$b = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{38}{40} = 0,95$$

$$a = \bar{y} - b \cdot \bar{x} = 5 - 0,95 \cdot 4 = 1,2$$

\Rightarrow Regressionsgerade $y = 1,2 + 0,95 \cdot x$



Bestimmtheits-/Gütemaß

- Zwischen dem Korrelationskoeffizienten r von Bravais-Pearson und dem Regressionskoeffizienten b gibt es folgende Beziehung:

$$b = r \cdot \frac{s_y}{s_x} \quad \text{mit } s_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \text{ und } s_y = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2}$$

- Die sich durch Anwendung der Regressionsgleichung aus den Beobachtungswerten Ergebenden Werte

$$\hat{y}_i = a + b \cdot x_i$$

heißen **theoretische y -Werte** oder durch die Regression erklärten Werte.

- Die Abweichungen

$$\hat{u}_i = y_i - \hat{y}_i$$

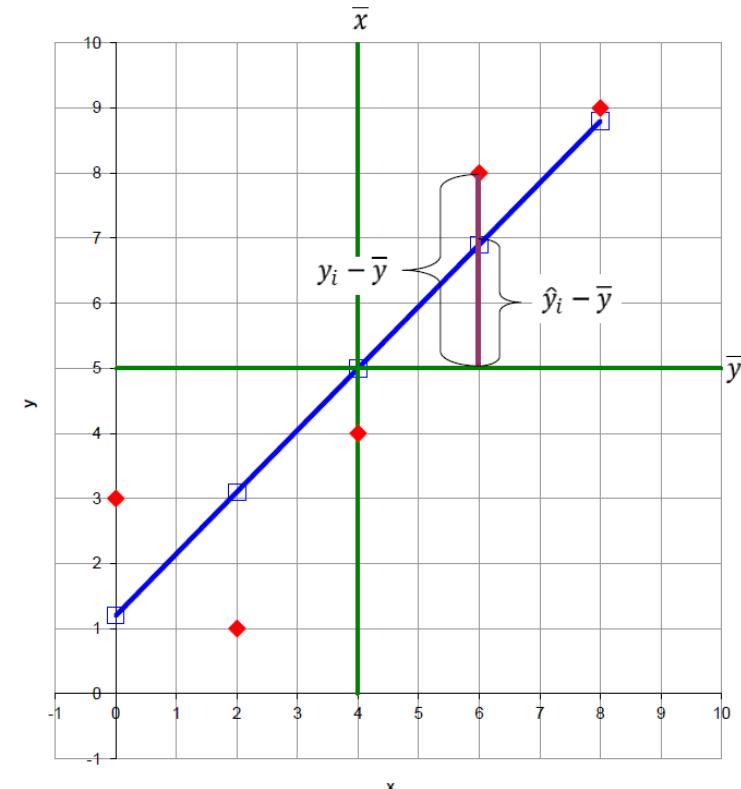
zwischen den beobachteten y -Werten und den theoretischen y -Werten heißen **KQ-Residuen**.

Bestimmtheits-/Gütemaß

Zur Einschätzung der Güte der Regression wird der **Determinationskoeffizient** d berechnet. Diesen wird auch als **Bestimmtheitsmaß** bezeichnet:

$$d = \left(\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \right) / \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \quad \text{mit } \sum_{i=1}^n (y_i - \bar{y})^2 \neq 0$$

Der Determinationskoeffizient misst, welcher Anteil der Varianz (quadratische Abweichung der Beobachtungswerte vom arithmetischen Mittel $(y_i - \bar{y})^2$) durch die Regression erklärt wird.



Bestimmtheits-/Gütemaß

Für den Determinationskoeffizienten gelten folgende Aussagen:

- Der Determinationskoeffizient gibt an, wie groß der Anteil der durch die Regressionsgeraden erklärten quadratischen Abweichungen des abhängigen Merkmals y vom seinem Mittelwert \bar{y} an der Gesamtsumme der quadratischen Abweichungen ist.
- $$d = 1 - \frac{\sum \hat{u}_i}{\sum (y_i - \bar{y})^2}$$
- $0 \leq d \leq 1$
- $d = r^2$
- $d = 1$ genau dann, wenn alle Punkte (x_i, y_i) auf der Regressionsgeraden liegen.
- $d = 0$ genau dann, wenn beide Merkmale unkorreliert sind. Dies ist unter den Annahmen $s_x \neq 0$ und $s_y \neq 0$ genau dann der Fall, wenn $b = 0$ und $a = \bar{y}$.
- Die Anpassung durch die Regressionsgerade ist umso besser, je größer d ist.

Ziele der Zeitreihenanalyse

1. **Beschreibung der Charakteristika einer Zeitreihe und ihrer Entwicklung:** Schätzung der zeitvariablen Lage oder Variabilität (Glättung, Entrauschen). Zerlegung in Trend, Saison, andere zyklische Komponenten, Rauschen.
 2. **Interventionsanalyse / Ausreißererkennung:** Messung besonderer Effekte zu bekanntem oder unbekanntem Zeitpunkt. Resultate irreführend bei Vernachlässigen von Abhängigkeiten.
 3. **Modellierung:** Formulierung und Anpassung stochastischer Modelle, z.B. von ARIMA-Modellen.
 4. **Vorhersage zukünftiger Werte:** Annahme: die zukünftigen Bedingungen sind wie die vorangegangenen.
 5. **Dynamische Regression** einer Zeitreihe auf erklärende Variablen. Aufdeckung von Zusammenhängen, bessere Vorhersagen.
 6. **Multivariate Modelle** zur Beschreibung der dynamischen Beziehungen zwischen verschiedenen Variablen und ihrer Vorhersage.
 7. **Kontrolle**, d.h. optimale Regelung der Einflussvariablen eines Prozesses.
-

I.1. Zeitreihen

- I.1.1. Einleitung
- I.1.2. Beispiele
- I.1.3. Einfache lineare Regression
- I.1.4. Ziele der Analyse

I.2. Prognosetechniken und Prognosefehler

- I.2.1. Einleitung
- I.2.2. Qualitative Verfahren
- I.2.3. Quantitative Verfahren
- I.2.4. Szenarioanalyse
- I.2.5. Güte der Prognose

I.3. Komponentenmodelle

- I.3.1. Einführung
- I.3.2. Glatte Komponente
- I.3.3. Saisonkomponente
- I.3.4. Analyse der Residuen
- I.3.5. Vorhersage

I.4. Interpolation und Approximation

- I.4.1. Einleitung
- I.4.2. Interpolation
- I.4.3. Approximation

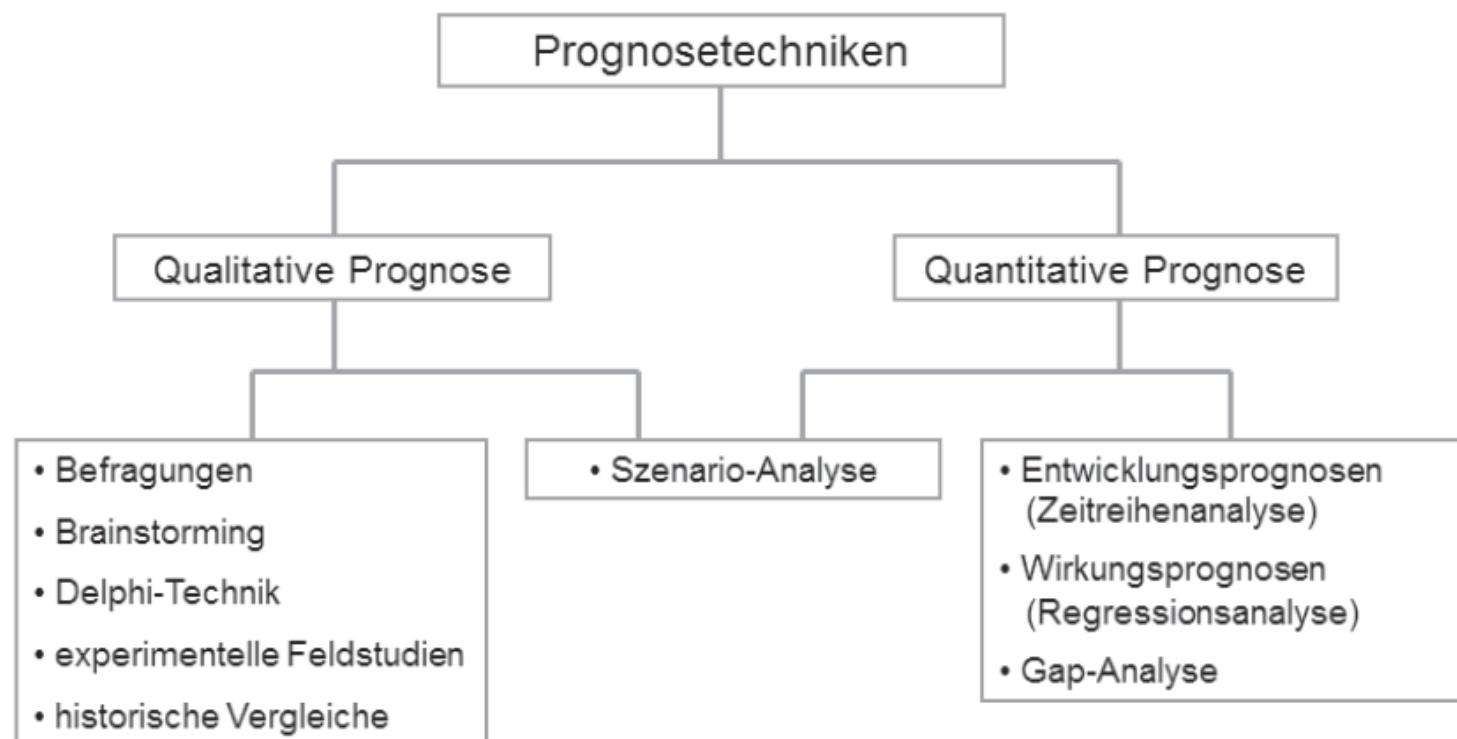
I.5. Funktionale Analyse

- I.5.1. Einleitung
 - I.5.2. Histogramme
 - I.5.3. Statistische Merkmale
 - I.5.4. (Diskrete) Fourier-Transformation
 - I.5.5. Wavelet-Transformation
 - I.5.6. Phasenraumdarstellung und –einbettung
-

Prognose

Unter einer Prognose wird die Vorhersage eines zukünftigen Zustandes oder der zukünftigen Entwicklung eines Systems anhand des gegenwärtigen Zustandes und der Kenntnis der vergangenen Entwicklung dieses Systems verstanden.

Systematisierung der Prognosetechniken



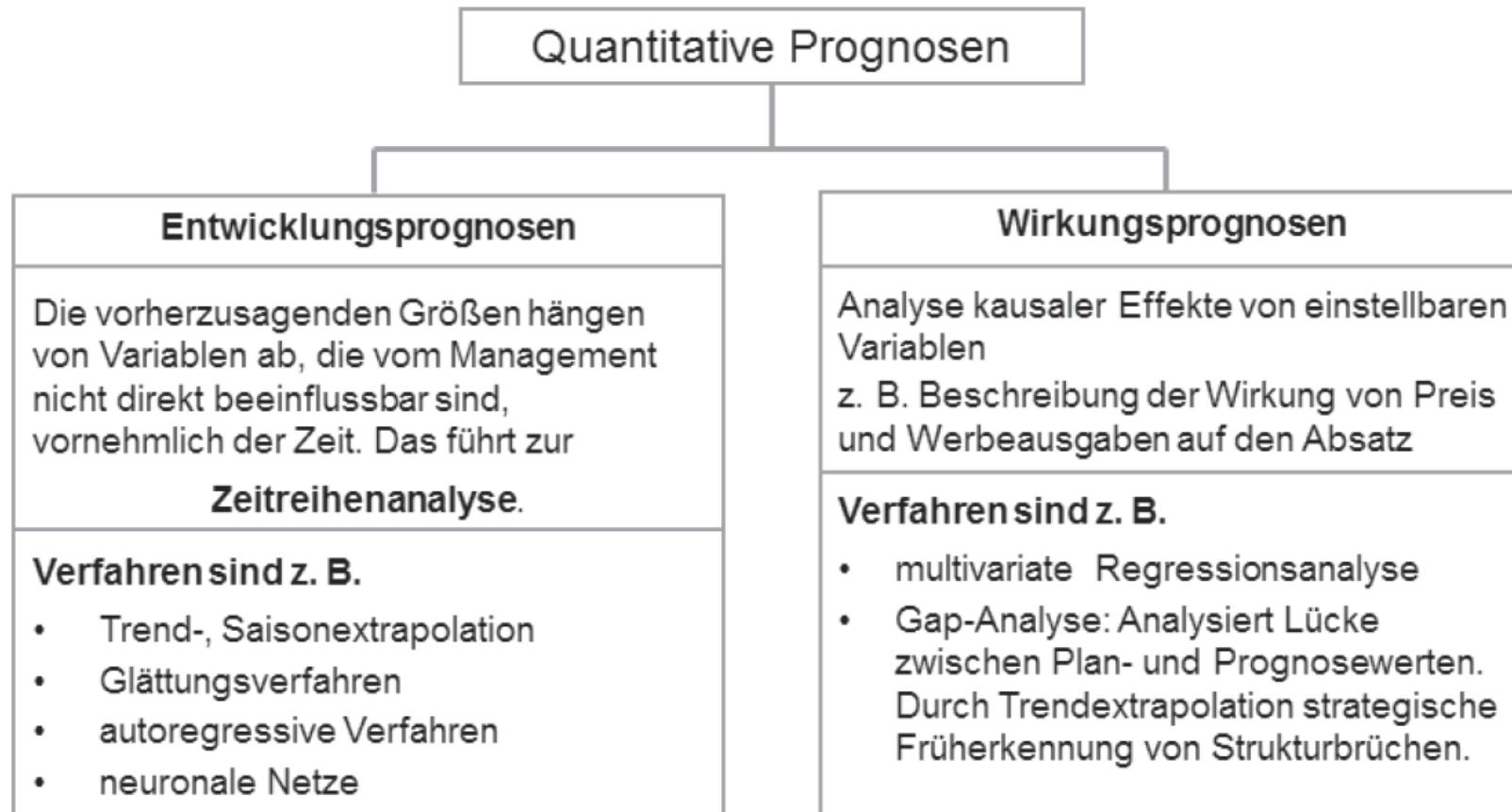
Qualitative Prognosen sind in erster Linie subjektive, nicht notwendig quantifizierbare Beurteilungen einer künftigen Entwicklung.

Exemplarische qualitative Verfahren

- **Visionen und Wahrsagungen**
- **Sammeln von Expertenmeinungen**
 - Brainstorming: Zwei-Phasenprinzip, Gruppensitzung von Experten zum Sammeln von Ideen (1. Phase), Sortierung und Bewertung der Ideen (2. Phase)
 - Delphi-Methode: mehrstufige „Fragebogen-Bewertung“
 - Panel-Konsensus-Verfahren: offene Gruppendiskussion mehrerer Experten
- **Experimentelle Feldversuche**
- **Historische Analogien**

Quantitative Verfahren sind zumeist Rechenverfahren zur Auswertung empirischer Daten zur Vorhersage (quantifizierte Beurteilung) künftiger Zustände oder Wirkungen von Einflussfaktoren.

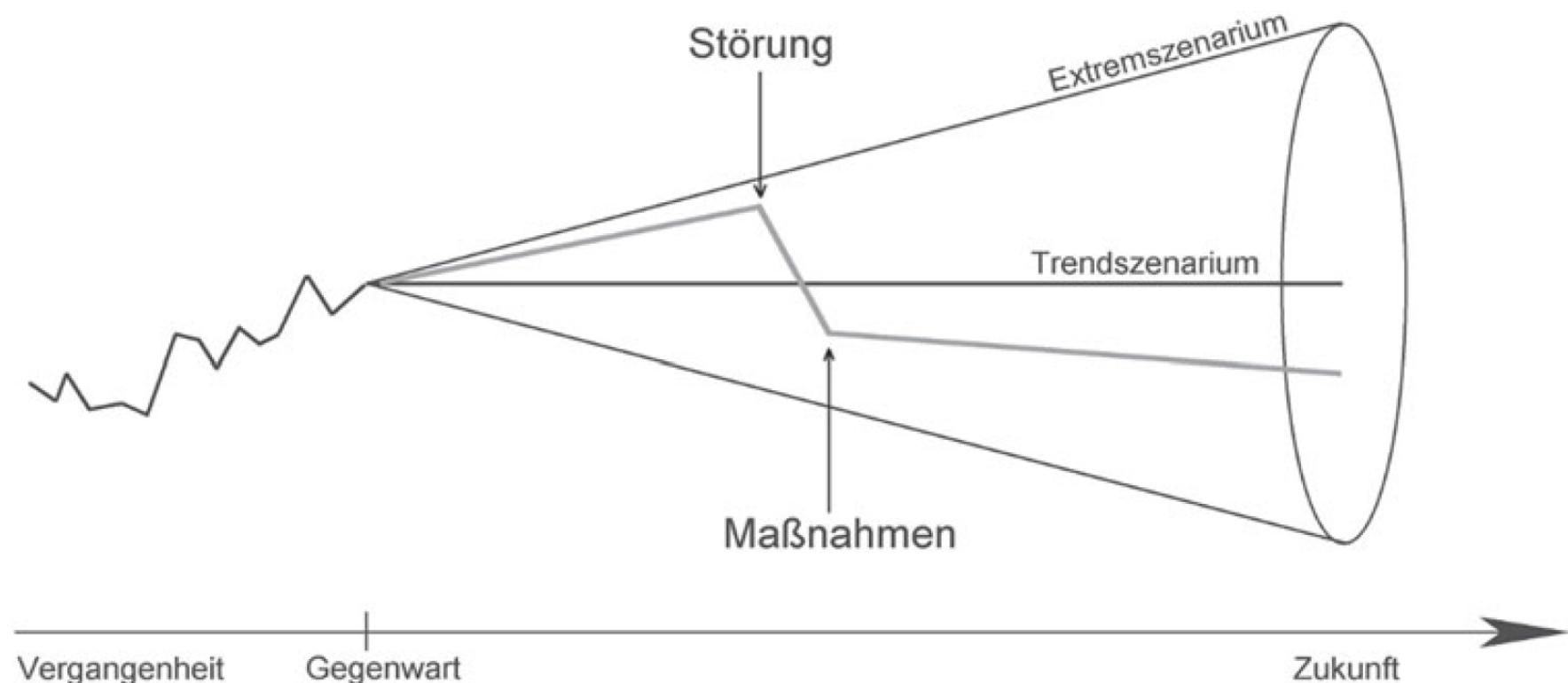
Quantitative Prognoseverfahren



Unter einem **Szenarium** wird eine hypothetische Folge von Ereignissen verstanden, die konstruiert wird, um kausale Zusammenhänge zwischen Entscheidungen und ihren Auswirkungen zu analysieren.

Phasen einer Szenarioanalyse

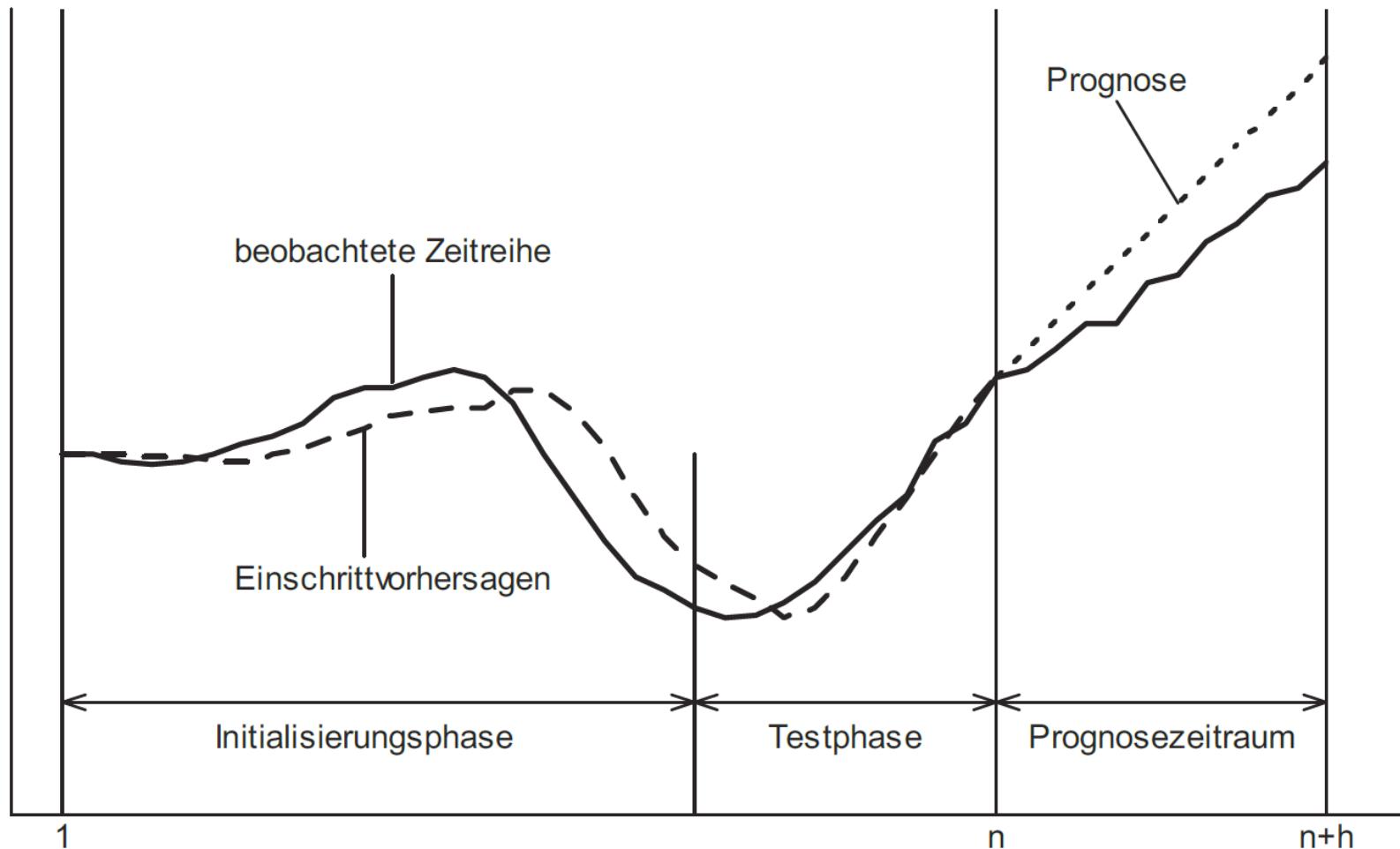
Phase	Gegenstand	Verfahren
1. Analyse	...der Ausgangssituation und der Systemzusammenhänge	...mittels qualitativer Verfahren wie Befragungen und Delphi-Methode
2. Prognose	...der absehbaren Entwicklung, überraschender Ereignisse. Erarbeitung von Entscheidungskriterien	...mittels Brainstorming, quantitativer Verfahren wie vektorieller Zeitreihenanalyse.
3. Synthese	...über Computer-Simulationen und Klassifikationsverfahren	...mittels Computer-Simulationen und Klassifikationsverfahren.

Szenariotrichter: Grundschema der Szenarioanalyse

→ Möglichkeitsspektrum öffnet sich mit fortschreitendem Prognosehorizont

Zur **Güte einer Prognose** werden verschiedene **Fehlermaße** in Abhängigkeit der zeitlichen Betrachtungsphase definiert.

Zeitschema zur Fehlerbeurteilung



Exemplarische Fehlermaße

- mittlerer quadratischer Fehler: $MSE = \frac{1}{h} \sum_{t=n+1}^{n+h} (\hat{y}_t - y_t)^2$
- mittlerer absoluter Fehler: $MAE = \frac{1}{h} \sum_{t=n+1}^{n+h} |\hat{y}_t - y_t|$
- mittlerer absoluter prozentualer Fehler: $MAPE = \frac{100\%}{h} \sum_{t=n+1}^{n+h} \frac{\hat{y}_t - y_t}{y_t}$

mit

Beobachtungswerten y_1, y_2, \dots, y_n

Prognosewerten $\hat{y}_{n+1}, \hat{y}_{n+2}, \dots, \hat{y}_{n+h}$

tatsächlichen (Prognose-) Werten $y_{n+1}, y_{n+2}, \dots, y_{n+h}$

Bemerkungen

- Mit wachsender Anzahl von Parametern, die in ein Prognose-Modell aufgenommen werden, kann das Fehlermaß im Prinzip beliebig klein werden:
→ Gefahr der **Überanpassung**.
- Einsatz von **Vergleichskriterien** zur Beurteilung der Güte unterschiedlicher Prognose-Modelle, z.B. **Ungleichheitskoeffizient U von Theil**

Bewertet den Unterschied eines Prognose-Modells zu einer naiven Prognose, die als Vorhersage $\hat{y}_{n+1} = y_n$ den zuletzt beobachteten Wert verwendet:

$$U = \sqrt{\frac{\sum_{t=n+1}^{n+h} (\hat{y}_t - y_t)^2}{\sum_{t=n+1}^{n+h} (y_{t-1} - y_t)^2}}$$

Fallunterscheidung:

- $U = 0$: Prognose-Modell vollkommen exakt
 - $U < 1$: Prognose-Modell besser als die naive Prognose
 - $U = 1$: Prognose-Modell und naive Prognose gleichwertig
 - $U > 1$: Prognose-Modell schlechter als die naive Prognose
-

I.1. Zeitreihen

- I.1.1. Einleitung
- I.1.2. Beispiele
- I.1.3. Einfache lineare Regression
- I.1.4. Ziele der Analyse

I.2. Prognosetechniken und Prognosefehler

- I.2.1. Einleitung
- I.2.2. Qualitative Verfahren
- I.2.3. Quantitative Verfahren
- I.2.4. Szenarioanalyse
- I.2.5. Güte der Prognose

I.3. Komponentenmodelle

- I.3.1. Einführung
- I.3.2. Glatte Komponente
- I.3.3. Saisonkomponente
- I.3.4. Analyse der Residuen
- I.3.5. Vorhersage

I.4. Interpolation und Approximation

- I.4.1. Einleitung
- I.4.2. Interpolation
- I.4.3. Approximation

I.5. Funktionale Analyse

- I.5.1. Einleitung
 - I.5.2. Histogramme
 - I.5.3. Statistische Merkmale
 - I.5.4. (Diskrete) Fourier-Transformation
 - I.5.5. Wavelet-Transformation
 - I.5.6. Phasenraumdarstellung und –einbettung
-

Komponenten einer Zeitreihe

- a) **Trend** t_t : Grundrichtung, langfristige Entwicklung
- b) **Andere zyklische Komponenten** z_t : mittelfristige Entwicklung, z.B. Konjunkturzyklus, Populationsrhythmen
- c) **Saisonkomponente** s_t : kurzfristige Entwicklung innerhalb der einzelnen Jahre durch saisonbedingte Schwankungen
- d) **Irreguläre Komponente oder Rest** ε_t : einmalige oder seltene Einflüsse und Zufallsschwankungen.

Bemerkungen

- t_t kann positiv (Wachstum)/negativ (Schrumpfung) oder linear/nichtlinear sein.
- z_t und s_t sind nichtlineare zyklische Schwankungen, die im Zeitraum von mehreren Jahren (Konjunktur) oder einem Jahr (Saison) periodisch wiederkehren.
- t_t , z_t und s_t werden als *systematische Komponente* bezeichnet und ε_t als zufällige Komponente

Zeitreihenzerlegungen

- Additive Zeitreihenzerlegung:

$$y_t = \underbrace{t_t + z_t}_{\text{glatte Komponente } g_t} + s_t + \varepsilon_t$$

zyklische Komponente

- Multiplikative Zeitreihenzerlegung: $y_t = t_t \cdot z_t \cdot s_t \cdot \varepsilon_t$

Bemerkungen

- Ein multiplikatives Modell kann durch Logarithmieren auf ein additives Modell zurückgeführt werden: $\log(y_t) = \log(t_t \cdot z_t \cdot s_t \cdot \varepsilon_t)$
 $= \log(t_t) + \log(z_t) + \log(s_t) + \log(\varepsilon_t)$
- Aufteilung in die vier Komponenten für „sehr lange“ Zeitreihen sinnvoll
- Bei „kürzeren“ Zeitreihen ist eine Trennung zwischen Trend und Konjunkturkomponente nicht mehr sinnvoll. Es bleibt eine Aufteilung in drei Komponenten – daher vielfach nur eine glatte Komponente g_t :

$$y_t = g_t + s_t + \varepsilon_t$$

- Bestimmung der **glatten Komponente** ist wichtig zur Prognose einer Zeitreihe.
- Zur Vereinfachung erfolgt im Folgenden (OBdA) eine Beschränkung auf glatter Komponente und Rest: $y_t = g_t + \varepsilon_t$ mit $t = 1, 2, \dots, n$

Trendmodelle (Prognosemodelle) zur Anpassung von glatten Kurven

- Annahme: glatte Komponenten lassen sich durch eine reellwertige Funktion mit „glattem“ Kurvenverlauf, z.B. lineare Funktion, Polynom, Exponentialfunktion, Sinuskurve, beschreiben. Der genaue Verlauf hängt von unbekannten Parametern ab.
- Parameterschätzung: Parameter werden dann beispielsweise mit der Methode der gewöhnlichen kleinsten Quadrate (*ordinary least squares, OLS*) geschätzt.

Ermittlung der **glatten Komponente** bei einem angenommen **linearen Trend**

1. Verwendung des additiven Modells und einer Modell-Geraden als glatte Komponente: $g_t = a + b \cdot t$
2. Bestimmung der optimalen Koeffizienten a und b durch Minimierung der Zielfunktion gemäß OLS:

$$Q(a, b) := \sum_{t=1}^n (y_t - a - b \cdot t)^2$$

Prinzip der „Linearen Einfachregression“ für Beobachtungspaare (x_t, y_t) :

$$\hat{b} = \frac{\sum_{t=1}^n x_t \cdot y_t - n \cdot \bar{x} \cdot \bar{y}}{\sum_{t=1}^n x_t^2 - n \cdot \bar{x}^2}$$

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x}$$

3. Berechnung der Regressionskoeffizienten

$$\hat{b} = \frac{12 \sum_{t=1}^n t \cdot y_t - 6 \cdot n \cdot (n + 1) \cdot \bar{y}}{(n - 1) \cdot n \cdot (n + 1)}$$

$$\hat{a} = \bar{y} - \hat{b} \cdot \frac{n + 1}{2}$$

Unter Beachtung, dass die x_t -Werte den Zeitpunkten t entsprechen und unter Verwendung der bekannten Summenformeln:

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t = \frac{1}{n} \sum_{t=1}^n t = \frac{n + 1}{2}$$

und

$$\sum_{t=1}^n x_t^2 = \sum_{t=1}^n t^2 = \frac{n \cdot (n + 1) \cdot (2n + 1)}{6}$$

Beispiel

Gegeben: Zeitreihe 100, 150, 75, 150, 175, 100, 175, 125, 200, 150, 175, 225, 150, 200, 250, 200

Bestimmung des linearen Trends:

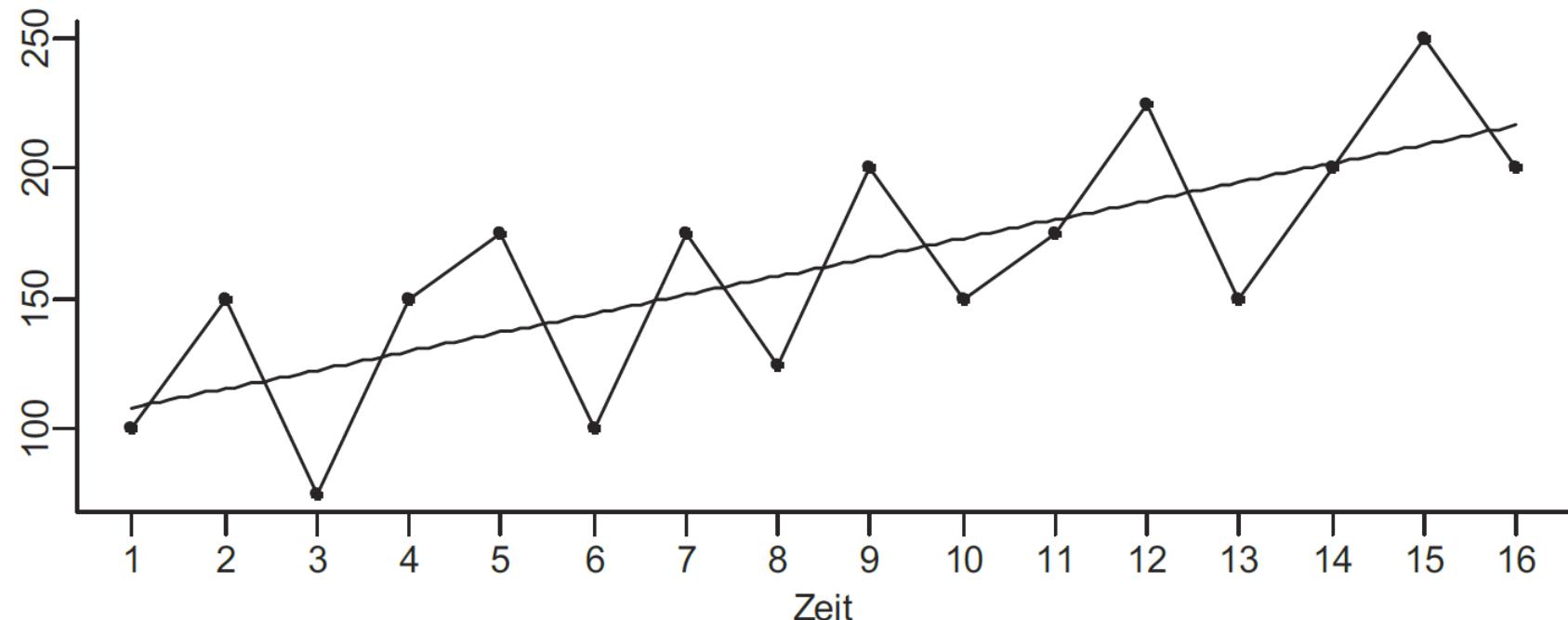
- Berechnung des arithmetischen Mittels $\bar{y} = 162,5$ und der Summe $1 \cdot y_1 + \dots + 16 \cdot y_{16} = 24550$ für die $n = 16$ Beobachtungen.
- Bestimmung der optimalen Regressionskoeffizienten

$$\hat{b} = \frac{12 \cdot 24550 - 6 \cdot 16 \cdot 17 \cdot 162,5}{15 \cdot 16 \cdot 17} = \frac{245}{34} \approx 7,206$$

$$\hat{a} = 162,5 - \frac{245}{34} \cdot \frac{17}{2} = 101,25$$

Beispiel

- geschätzte Trendgerade: $\hat{g}_t = 101,25 + 7,206 \cdot t$



Ermittlung der **glatten Komponente** bei einem angenommen **polynomialen Trend**

1. Verwendung des additiven Modells und eines Polynoms vom Grade k als glatte Komponente: $g_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \dots + \beta_k \cdot t^k$

Bemerkung: für viele Anwendungen ist $k = 2$ oder $k = 3$ (bei einem glatten Kurvenverlauf) ausreichend

2. Bestimmung der optimalen Koeffizienten $\beta_0, \beta_1, \dots, \beta_k$ durch Minimierung der Zielfunktion gemäß OLS:

$$Q(\beta_0, \dots, \beta_k) := \sum_{t=1}^n (y_t - \beta_0 - \beta_1 \cdot t - \dots - \beta_k \cdot t^k)^2$$

3. Partielle Ableitungen von Q nach $\beta_0, \beta_1, \dots, \beta_k$ berechnen und gleich null setzen:

$$\frac{\partial Q(\beta_0, \dots, \beta_k)}{\partial \beta_0} = \sum_{t=1}^n 2(y_t - \beta_0 - \beta_1 \cdot t - \dots - \beta_k \cdot t^k) \cdot (-1) = 0$$

$$\frac{\partial Q(\beta_0, \dots, \beta_k)}{\partial \beta_1} = \sum_{t=1}^n 2(y_t - \beta_0 - \beta_1 \cdot t - \dots - \beta_k \cdot t^k) \cdot (-t) = 0$$

⋮

⋮

$$\frac{\partial Q(\beta_0, \dots, \beta_k)}{\partial \beta_k} = \sum_{t=1}^n 2(y_t - \beta_0 - \beta_1 \cdot t - \dots - \beta_k \cdot t^k) \cdot (-t^k) = 0$$

4. Bestimmung des Normalengleichungssystems:

$$\beta_0 \cdot n + \beta_1 \cdot \sum_{t=1}^n t + \cdots + \beta_k \cdot \sum_{t=1}^n t^k = \sum_{t=1}^n y_t$$

$$\beta_0 \cdot \sum_{t=1}^n t + \beta_1 \cdot \sum_{t=1}^n t^2 + \cdots + \beta_k \cdot \sum_{t=1}^n t^{k+1} = \sum_{t=1}^n t \cdot y_t$$

⋮

⋮

$$\beta_0 \cdot \sum_{t=1}^n t^k + \beta_1 \cdot \sum_{t=1}^n t^{k+1} + \cdots + \beta_k \cdot \sum_{t=1}^n t^{2k} = \sum_{t=1}^n t^2 \cdot y_t$$

Das lineare Gleichungssystem hat genau eine Lösung, welche die gesuchten Schätzwerte $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ liefert.

Beispiel

Gegeben: Zeitreihe 100, 140, 175, 170, 190, 200, 180, 170, 180, 140, 130, 100 und Modell einer quadratischen Parabel $g_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2$

Bestimmung des polynomialen Trends:

- Aufstellen des allgemeinen Normalengleichungssystems in Matrixschreibweise für $k = 2$:

$$\begin{pmatrix} \sum 1 & \sum t & \sum t^2 \\ \sum t & \sum t^2 & \sum t^3 \\ \sum t^2 & \sum t^3 & \sum t^4 \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum t \cdot y_t \\ \sum t^2 \cdot y_t \end{pmatrix}$$

Beispiel (Fortsetzung)

- Aufstellen des Normalengleichungssystems in Matrixschreibweise für die gegebene Zeitreihe mit $n = 12$:

$$\begin{pmatrix} 12 & 78 & 650 \\ 78 & 650 & 6084 \\ 650 & 6084 & 60710 \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1875 \\ 12005 \\ 95315 \end{pmatrix}$$

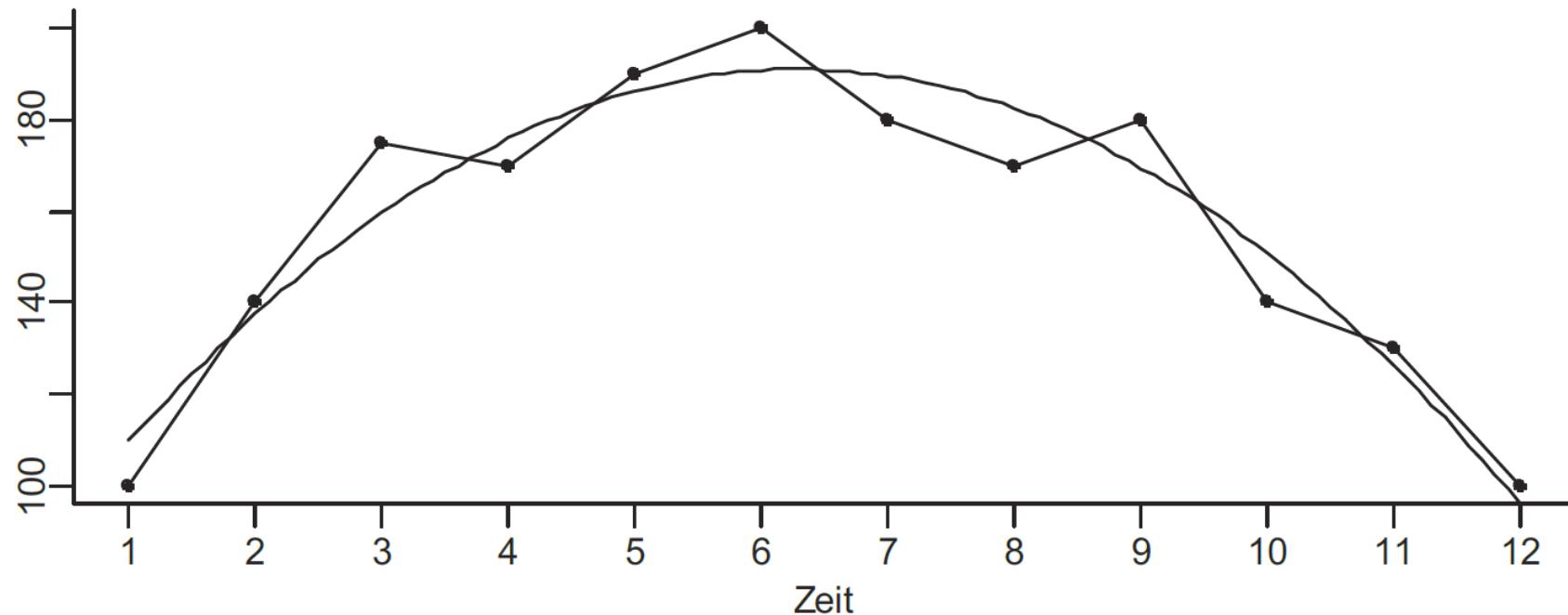
- Lösung des linearen Gleichungssystems (gerundet):

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 76,48 \\ 36,47 \\ -2,90 \end{pmatrix}$$

- Aufstellen der Gleichung für die gesuchte Trendparabel:

$$\hat{g}_t = 76,48 + 36,47 \cdot t - 2,9 \cdot t^2$$

Beispiel (Fortsetzung)



- Prognosen auf Basis eines Parabel-förmigen Modells nur für kurze Prognosehorizonte geeignet, da sie außerhalb des Beobachtungszeitraums schnell „abdriften“

I.3.2. Glatte Komponente

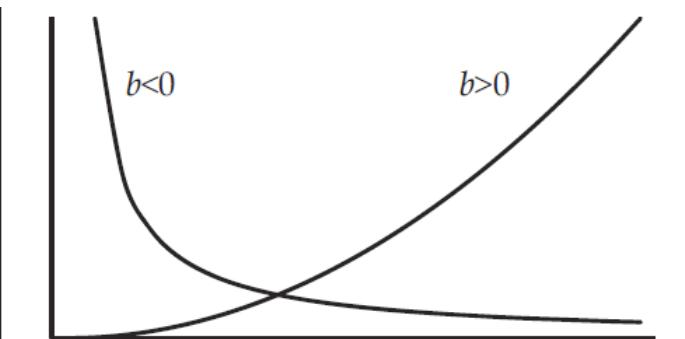
I.3.2.3. Übersicht zu Linearisierbaren Trendansätzen

Ansatz

$$g_t = a \cdot t^b$$

Potenzfunktion

Grafische Darstellung

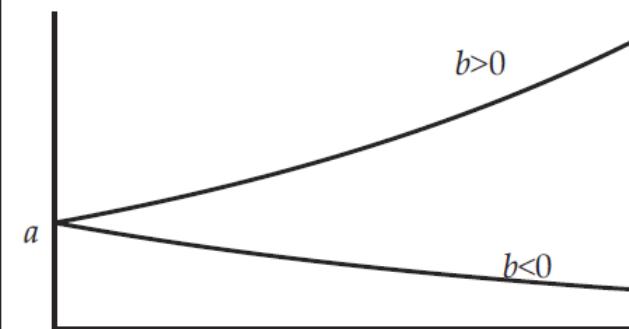


Transformierte Funktion

$$\log y_t = \log a + b \cdot \log t$$

$$g_t = a \cdot e^{b \cdot t}$$

Exponentialfunktion

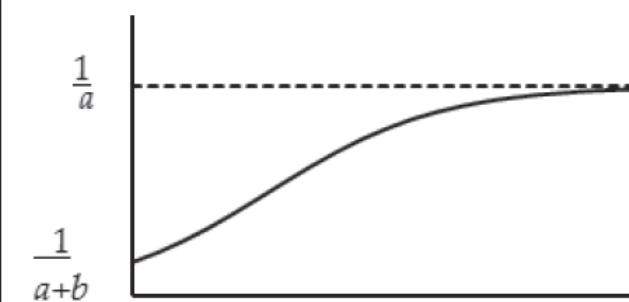


$$\ln y_t = \ln a + b \cdot t$$

$$g_t = \frac{1}{a + b \cdot e^{-t}}$$

logistische Funktion

$$a > 0, b > 0$$



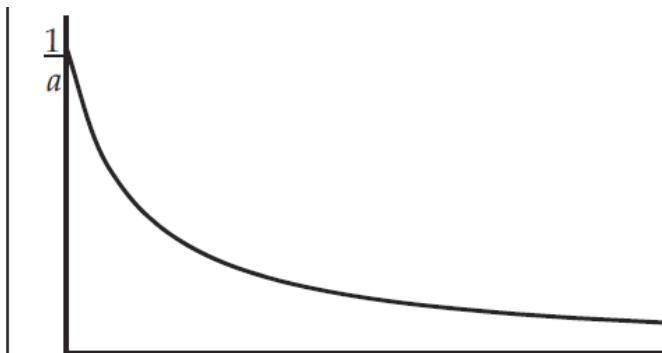
$$\frac{1}{y_t} = a + b \cdot e^{-t}$$

Ansatz

$$g_t = \frac{1}{a + b \cdot t}$$

Hyperbelfunktion

$$a > 0, b > 0$$

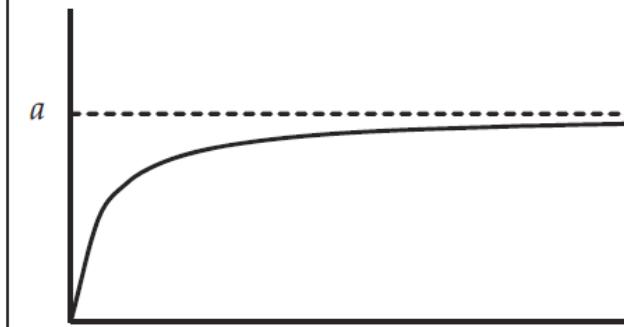
Grafische Darstellung**Transformierte Funktion**

$$\frac{1}{y_t} = a + b \cdot t$$

$$g_t = \frac{a \cdot t}{b + t}$$

Törnquist-Funktion

$$a > 0, b > 0$$



$$\frac{1}{y_t} = \frac{1}{a} + \frac{b}{a} \cdot \frac{1}{t}$$

Exponentielles Glätten

- Überführung einer gegebenen Zeitreihe (y_t) über einen rekursiven Algorithmus in eine neue Zeitreihe, welche aus einer glatten Komponente (\hat{g}_t) besteht.
- Die glatte Komponente ist dabei als ein gewogenes arithmetisches Mittel aus der gesamten Vergangenheit mit exponentiell abklingenden Gewichten definiert:

$$\hat{g}_t = \alpha \cdot \sum_{j=0}^{\infty} (1 - \alpha)^j \cdot y_{t-j}$$

Die Zahl $\alpha \in [0 \dots 1]$ wird als **Glättungsfaktor** bezeichnet.

- Da eine unendliche Vergangenheit nicht zur Verfügung steht, rekursive Berechnung. Jedes \hat{g}_t für $t = 2, 3, \dots, n$ lässt sich rekursiv aus dem aktuellen Beobachtungswert y_t und seinem Vorgänger berechnen:

$$\hat{g}_t = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{g}_{t-1}$$

- Startwert (vielfach) zur Durchführung der Rekursion: $\hat{g}_1 = y_1$

Beweis:

$$\begin{aligned}
 \alpha \cdot y_t + (1 - \alpha) \cdot \hat{g}_{t-1} &= \alpha \cdot y_t + (1 - \alpha) \cdot \alpha \cdot \sum_{j=0}^{\infty} (1 - \alpha)^j \cdot y_{t-1-j} \\
 \hat{g}_t &= \alpha \cdot \sum_{j=0}^{\infty} (1 - \alpha)^j \cdot y_{t-j} \\
 &= \alpha \cdot y_t + \alpha \cdot \sum_{j=0}^{\infty} (1 - \alpha)^{j+1} \cdot y_{t-1-j} \\
 &= \alpha \cdot ((1 - \alpha)^0 \cdot y_t + \sum_{j=1}^{\infty} (1 - \alpha)^j \cdot y_{t-j}) \\
 &= \alpha \cdot \sum_{j=0}^{\infty} (1 - \alpha)^j \cdot y_{t-j} = \hat{g}_t
 \end{aligned}$$

Beispiel

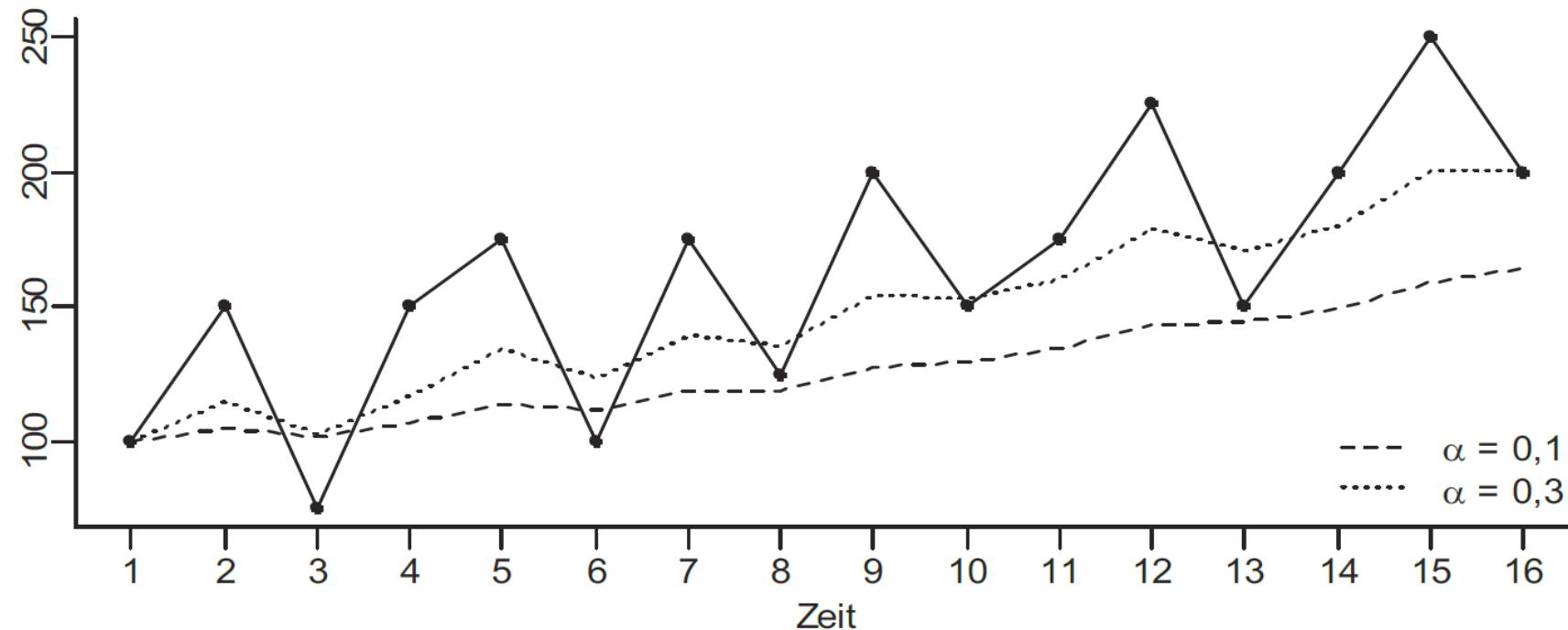
Gegeben: Zeitreihe 100, 150, 75, 150, 175, 100, 175, 125, 200, 150, 175, 225, 150, 200, 250, 200 und Glättungsfaktoren $\alpha = 0.1$ und $\alpha = 0.3$.

Bestimmung der exponentiellen Glättung:

Ausgehend vom Startwert $\hat{g}_1 = y_1 = 100$ berechnen der glatten Komponenten gemäß der Rekursionsvorschrift: $\hat{g}_t = 0.1 \cdot y_t + 0.9 \cdot \hat{g}_{t-1}$; $\hat{g}_t = 0.3 \cdot y_t + 0.7 \cdot \hat{g}_{t-1}$

t	y_t	\hat{g}_t	
		$\alpha = 0,1$	$\alpha = 0,3$
1	100	100,00	100,00
2	150	105,00	115,00
3	75	102,00	103,00
4	150	106,80	117,10
⋮	⋮	⋮	⋮
16	200	163,57	200,39

Beispiel (Fortsetzung)



Saisonkomponente

Eine Zeitreihe (s_t), die regelmäßigen periodischen Schwankungen unterliegt. Typische Periodenlängen l sind $l = 12$ oder $l = 4$, je nachdem, ob es sich bei den Zeitpunkten t um Monate oder Quartale handelt.

Ermittlung:

- Ansätze zur Berechnung der Saisonkomponente zu einer gegebenen Zeitreihe (y_t) unter Beachtung einer festen Saisonfigur, d.h. $s_{t+l} = s_t$ für alle $t = 1, \dots, n - l$:

$$y_t = g_t + s_t + \varepsilon_t \quad \text{oder} \quad y_t = g_t \cdot s_t \cdot \varepsilon_t$$

- Befreiung der Zeitreihe vom Trend durch Schätzung (\hat{g}_t) für die glatte Komponente (g_t) und durch Eliminierung aus (y_t):

$$\hat{r}_t := y_t - \hat{g}_t \quad \text{bzw.} \quad \hat{r}_t := \frac{y_t}{\hat{g}_t}$$

Ermittlung (Fortsetzung):

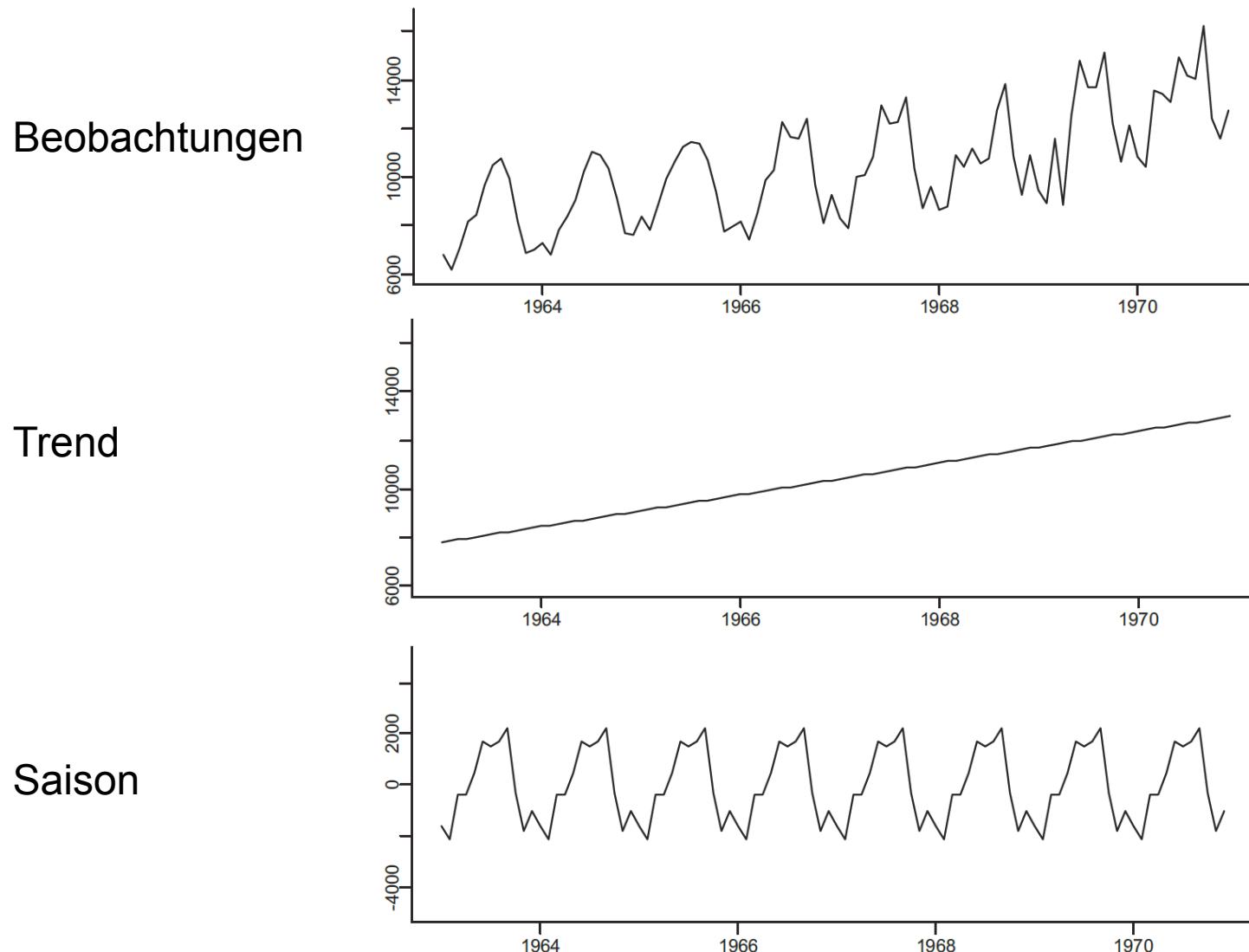
- Zur Vereinfachung sei angenommen, dass $(y_t)_{t=1,\dots,n}$ für m volle Jahre beobachtet wird, d.h. $n = m \cdot l$.
- Saisonfigur kann für $t = 1, 2, \dots, l$ durch den Mittelwert der trendbefreiten Zeitreihenwerte über alle vergleichbaren Zeitpunkte bestimmt werden.
- Dabei ist im additiven Fall das arithmetische und im multiplikativen Fall das geometrische Mittel zu nehmen:

$$\hat{s}_t := \frac{1}{m} (\hat{r}_t + \hat{r}_{t+l} + \dots + \hat{r}_{t+(m-1)\cdot l})$$

bzw.

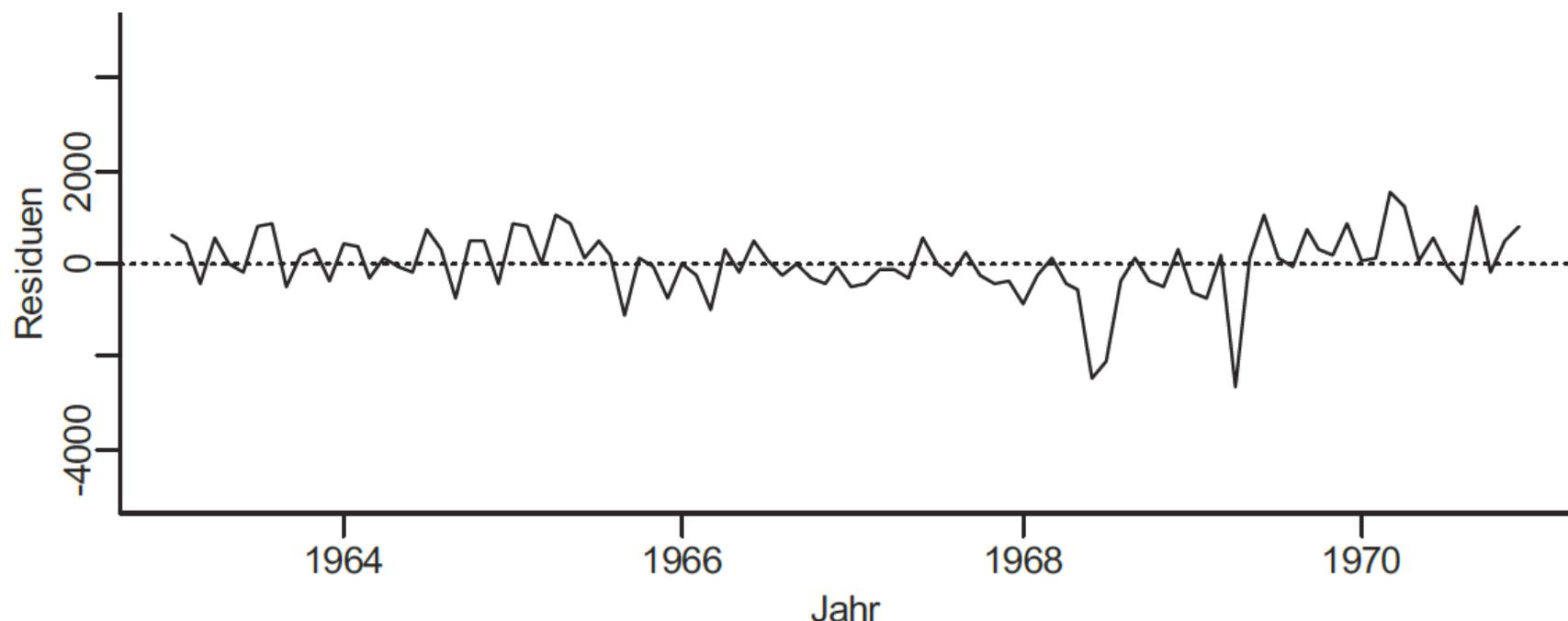
$$\hat{s}_t := (\hat{r}_t \cdot \hat{r}_{t+l} \cdot \dots \cdot \hat{r}_{t+(m-1)\cdot l})^{\frac{1}{m}}$$

Beispiel: monatliche Flugstrecken britischer Linienflugzeuge



- Sofern alle Komponenten zu einer Zeitreihe ermittelt wurden und entsprechende Schätzungen dafür berechnet sind, können diese unter Beachtung des additiven oder multiplikativen Ansatzes aus der gegebenen Zeitreihe eliminiert werden.
- Verbleibend ist das **Residuum** $\hat{\varepsilon}_t = y_t - \hat{g}_t - \hat{s}_t$ bzw. $\hat{\varepsilon}_t = \frac{y_t}{\hat{g}_t \cdot \hat{s}_t}$
- Ableitung einer guten Modellwahl anhand der Residuen, die
 - (1) um den konstanten Mittelwert 0 bzw. 1 schwanken,
 - (2) eine konstante Varianz haben,
 - (3) voneinander unabhängig und
 - (4) normalverteilt sind.

Beispiel: Residuen zu den monatlichen Flugstrecken nach Zerlegung in Trendgerade und Saisonkomponente



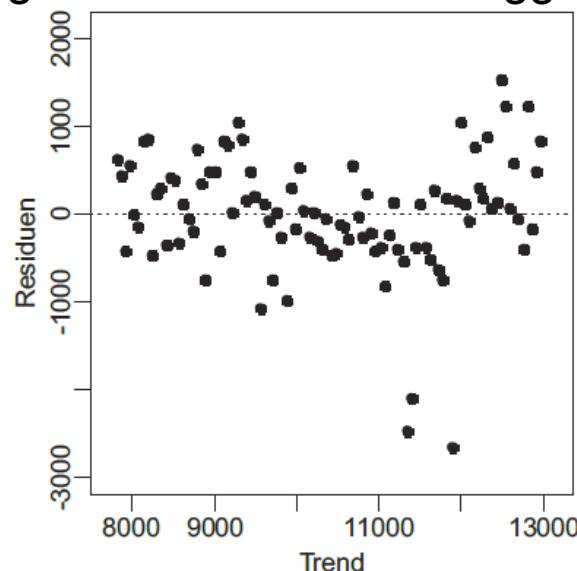
Bemerkung:

Wenn alle Informationen der Zeitreihe durch die gefundenen Komponenten „erklärt“ werden, sollte in der Residualkomponente keine Information mehr enthalten sein.

Trend Residuen Plot: Beurteilung der Varianzfunktion auf Konstanz

- Eintragen der n Punkte $(\hat{g}_t, \hat{\varepsilon}_t)$ in ein rechtwinkliges Koordinatensystem.
- Wenn die Punktewolke einem sich nach rechts oder links öffnenden Trichter gleicht, wächst bzw. sinkt die Varianz mit fortschreitender Zeit.

Beispiel: Monatliche Flugstrecken britischer Fluggesellschaften



→ Erkennbar ist ein sich nach rechts öffnender Trichter, weshalb diese Residuen bezüglich konstanter Varianz nicht gut einzustufen sind.

Ausgangssituation

- Beobachtet wurde eine Zeitreihe y_1, y_2, \dots, y_n , die in Komponenten zerlegt worden ist, wobei die Residuen als hinreichend gut bewertet werden können.
- Anhand der Komponenten sollen Schätzungen $\hat{y}_{n+1}, \hat{y}_{n+2}, \dots, \hat{y}_{n+h}$ für zukünftige Zeitreihenwerte $y_{n+1}, y_{n+2}, \dots, y_{n+h}$ ermittelt werden.

Vorhersage durch Kurvenanpassung

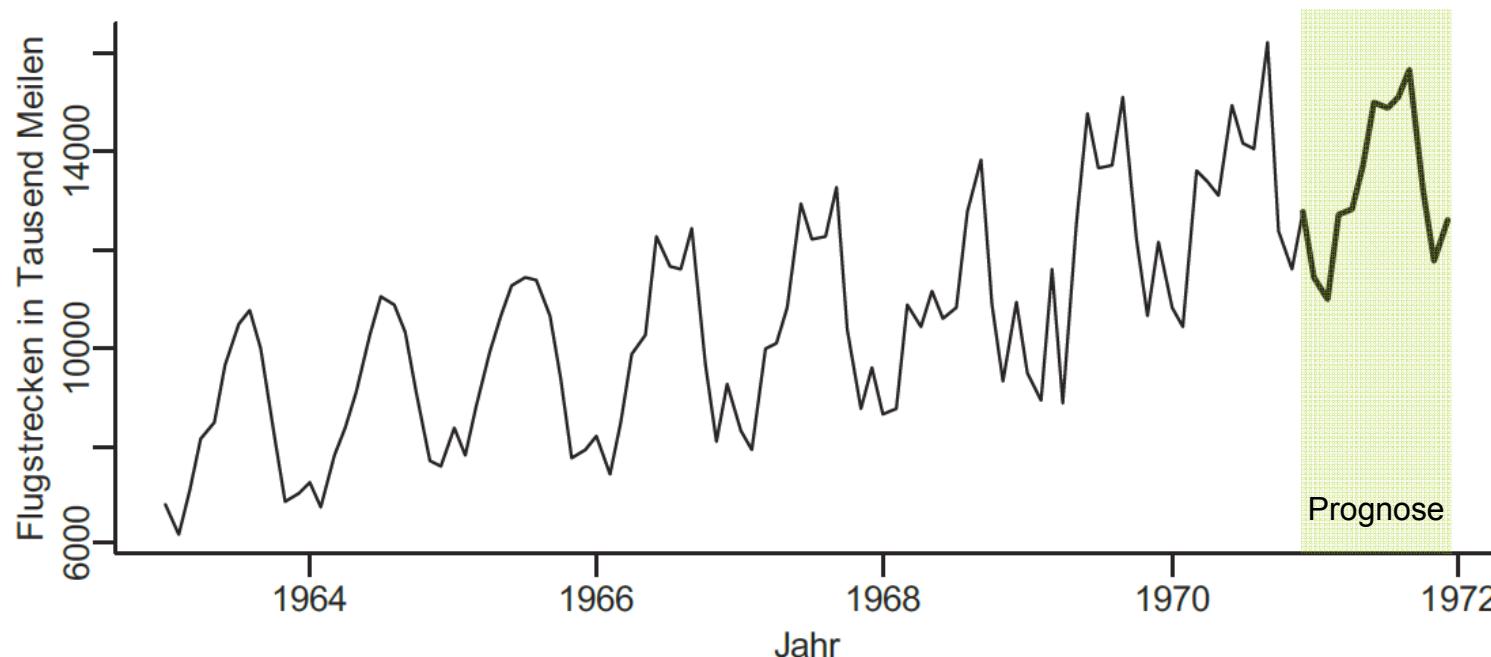
1. An die Zeitreihe y_1, y_2, \dots, y_n eine glatte Funktion \hat{g}_t der Zeit anpassen.
2. Die Saisonkomponente ist, falls vorhanden, durch eine feste Saisonfigur \hat{s}_t geschätzt worden.
3. Prognosen durch Fortsetzen der geschätzten Komponenten:

$$\hat{y}_{n+t} = \hat{g}_{n+t} + \hat{s}_{n+t} \quad \text{bzw.} \quad \hat{y}_{n+t} = \hat{g}_{n+t} \cdot \hat{s}_{n+t} \quad \text{für } t = 1, \dots, h$$

Bemerkungen:

- Residuen werden nicht berücksichtigt, da sie nicht vorhersagbar sind
 - Prognosehorizont h sollte nicht zu groß sein, empfohlen wird $h \leq 0.5 \cdot n$.
-

Beispiel: Prognose über 12 Monate mithilfe der Komponentenzerlegung der Flugstrecken-Zeitreihe



Beobachtung:

- Die Saisonfigur zeigt die typische Doppelspitze in den Sommerferien der letzten fünf Jahre in abgeschwächter Form.
- Das liegt daran, dass in die Prognose auch die drei weiter zurückliegenden Jahre gleichberechtigt eingegangen sind, die im Sommer nur ein einzelnes lokales Maximum aufwiesen.

I.1. Zeitreihen

- I.1.1. Einleitung
- I.1.2. Beispiele
- I.1.3. Einfache lineare Regression
- I.1.4. Ziele der Analyse

I.2. Prognosetechniken und Prognosefehler

- I.2.1. Einleitung
- I.2.2. Qualitative Verfahren
- I.2.3. Quantitative Verfahren
- I.2.4. Szenarioanalyse
- I.2.5. Güte der Prognose

I.3. Komponentenmodelle

- I.3.1. Einführung
- I.3.2. Glatte Komponente
- I.3.3. Saisonkomponente
- I.3.4. Analyse der Residuen
- I.3.5. Vorhersage

I.4. Interpolation und Approximation

- I.4.1. Einleitung
- I.4.2. Interpolation
- I.4.3. Approximation

I.5. Funktionale Analyse

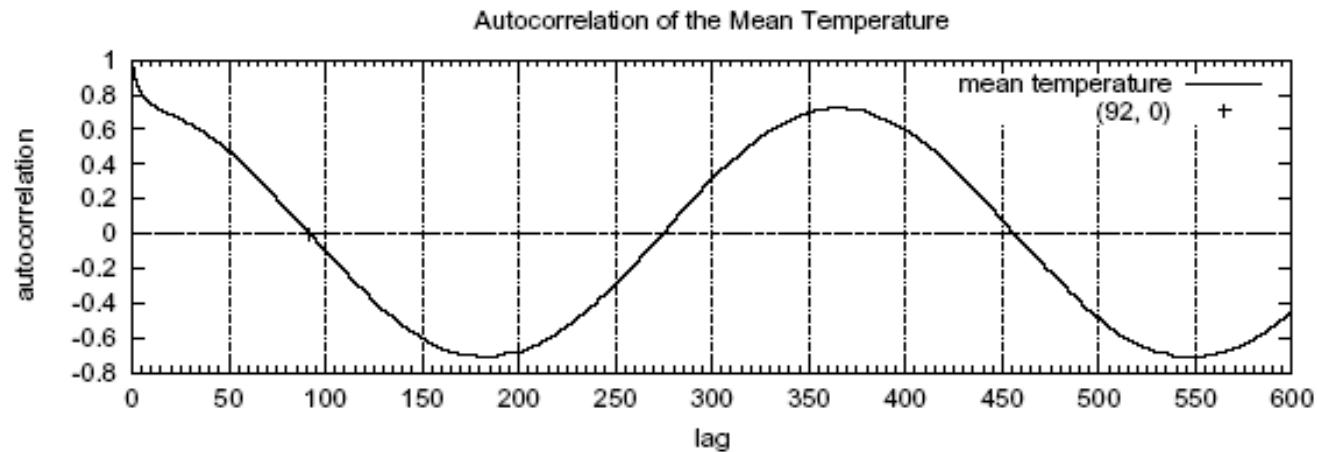
- I.5.1. Einleitung
 - I.5.2. Histogramme
 - I.5.3. Statistische Merkmale
 - I.5.4. (Diskrete) Fourier-Transformation
 - I.5.5. Wavelet-Transformation
 - I.5.6. Phasenraumdarstellung und –einbettung
-

Funktionsgraph

für eindimensionale skalare Funktionen: $f : I \rightarrow \mathbb{R}$, $I \subseteq \mathbb{R}$

Die Funktion wird durch eine Kurve in einem Koordinatensystem repräsentiert

Beispiel:



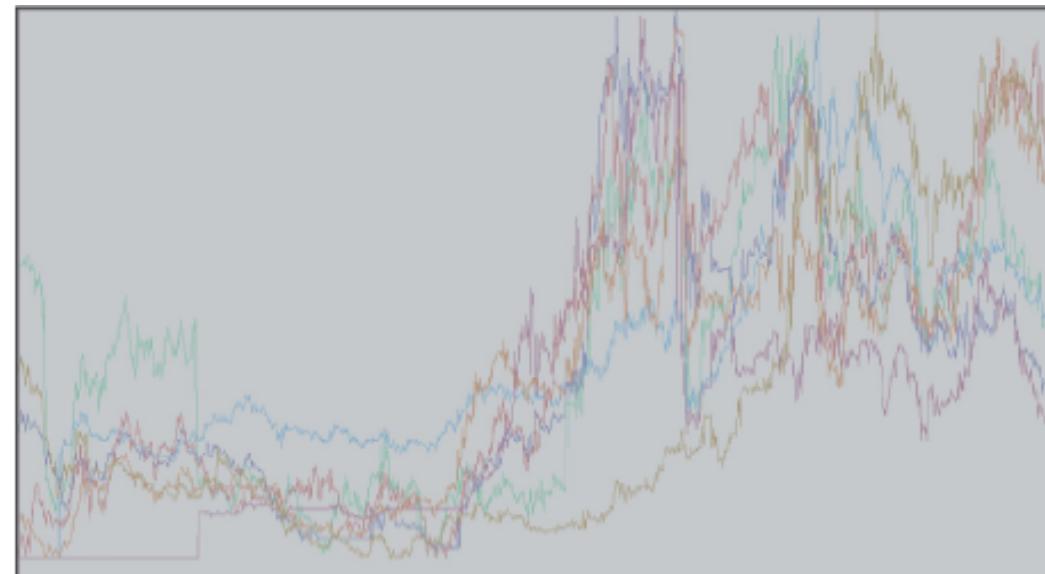
Funktionsgraph

für eindimensionale mehrwertige Funktionen $\mathbf{f} : I \rightarrow \mathbb{R}^d$, $I \subseteq \mathbb{R}$

je ein eindimensionaler Funktionsgraph für jede Komponentenfunktion f_i ,
 $\mathbf{f} := (f_1, f_2, \dots, f_d)$, unterscheidbar gemacht durch Farbe, Strichmuster,
Dicke, etc. („graphische Variable“ nach Bertin)

Beispiel:

Verlauf von acht Aktienkursen
über 20 Jahre, jeder
Aktienkurs durch einen
Funktionsgraphen
repräsentiert.



Interpolation von Stützstellen

findet zu einer Folge von Werten, die eine stetige Funktion repräsentieren, eine Funktion aus einer vorgegebenen Funktionenklasse, die diese Werte annimmt.

Approximation von Stützstellen

findet zu einer Folge von Werten, die eine stetige Funktion repräsentieren, eine Funktion aus einer vorgegebenen Funktionenklasse, die diesen Werten möglichst nahe kommt.

Anwendungen

- Schätzung und Visualisierung der (unbekannten) Funktion, von der die gegebenen Werte stammen.
- Ermöglichung der vergrößerten Darstellung der Funktion (Zoom).

Interpolation von Stützstellen (allgemein)

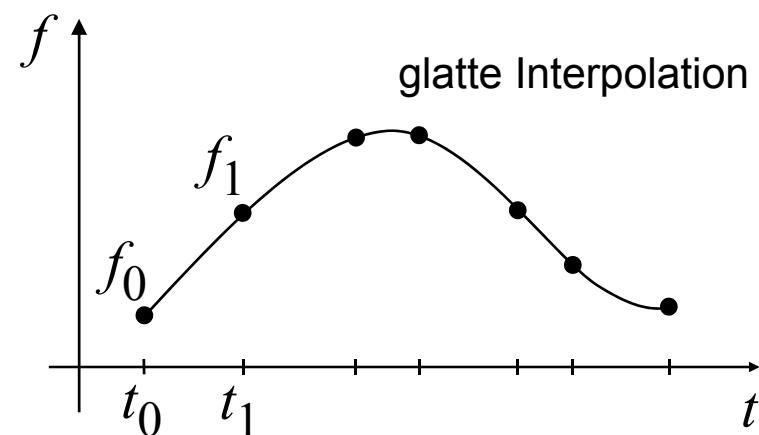
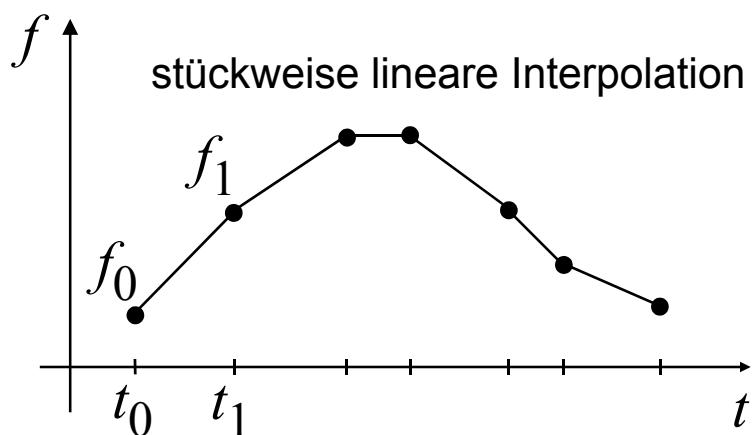
Gegeben: Stützstellen (t_i, \mathbf{f}_i) , $\mathbf{f}_i \in \mathbb{R}^d$, $i = 0, \dots, m - 1$, $t_i < t_{i+1}$.

Gesucht: eine Interpolationsfunktion $\mathbf{f} : [t_0, t_{m-1}] \rightarrow \mathbb{R}^d$ mit $\mathbf{f}(t_i) = \mathbf{f}_i$

Beispiel: stückweise lineare Interpolation:

$$\mathbf{f}(t) = \mathbf{f}_i + \frac{t - t_i}{t_{i+1} - t_i} \cdot (\mathbf{f}_{i+1} - \mathbf{f}_i), \quad t \in [t_i, t_{i+1}], \quad i = 0, \dots, m - 1.$$

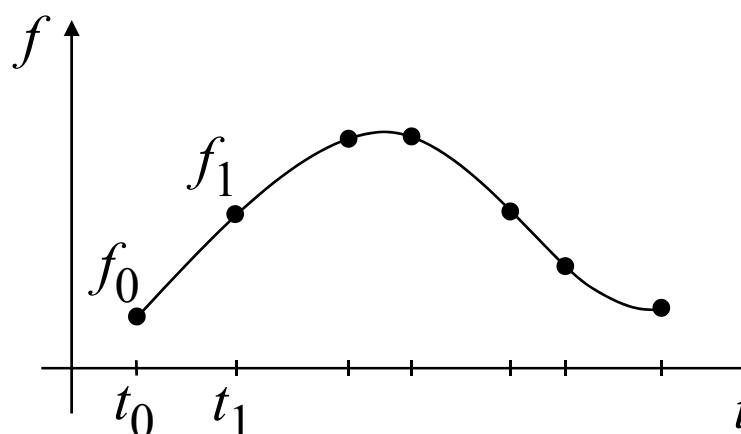
stetig, aber an den Stützpunkten (t_i, \mathbf{f}_i) i.a. nicht glatt.



Interpolation mit Spline-Funktionen

Vorteile:

- Vermeiden des welligen Verhaltens von Interpolationspolynomen durch stückweise polynomiale Interpolation, wobei glatte Übergänge zwischen den verschiedenen Polynomen gefordert werden.
- Verminderung von Fehlern bei der numerischen Auswertung im Falle vieler Stützstellen.



Spline

Funktion s auf a, b heißt Spline vom Grad k , falls folgende Bedingungen gelten:

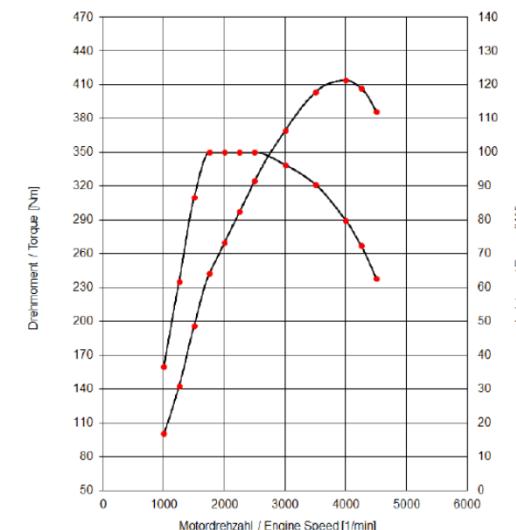
- s ist $k-1$ mal stetig differenzierbar auf $[a, b]$
- Es gibt Stützstellen $a = x_0 < x_1 < \dots < x_n = b$, sodass s in jedem Intervall x_i, x_{i+1} ein Polynom vom Grad k ist.
- Ein interpolierender Spline erfüllt die Bedingungen $s(x_i) = f(x_i)$.
- Raum dieser Splines wird mit $S_k(x_0, \dots, x_n)$ bezeichnet.

Bemerkung:

Splines sind im Wesentlichen stückweise Polynome mit gewissen Stetigkeits- und Differenzierbarkeitsanforderungen an den Intervallgrenzen.

Beispiel:

Spline für Drehmoment und Leistung eines 2.0l Turbo Diesel. Der Spline muss bei sogenannten Pressekurven einerseits die Messdaten glätten darf aber andererseits nicht überschwingen, speziell beim Drehmoment



Kubische interpolierende Splines:

- Auf jedem Teilintervall $[x_{i-1}, x_i]$ ist s ein Polynom vom Grad 3:

$$s(x) = p_i(x) = \alpha_i + \beta_i(x - x_{i-1}) + \gamma_i(x - x_{i-1})^2 + \delta_i(x - x_{i-1})^3$$

- s ist zweimal stetig differenzierbar. Es gilt für $i=1, \dots, n-1$:

$$p_i(x_i) = p_{i+1}(x_i), \quad p'_i(x_i) = p'_{i+1}(x_i), \quad p''_i(x_i) = p''_{i+1}(x_i)$$

- s erfüllt die Interpolationsbedingungen $s(x_i) = f_i, i=0, \dots, n$

→ Insgesamt $3n-1+n+1=4n-2$ Bedingungen für die $4n$ Koeffizienten.

Definition der zwei (Rand-) Zusatzbedingungen:

- Natürlicher Spline: $s''(x_0) = s''(x_n) = 0$
- Hermitescher Spline: $s'(x_0) = f'_0, s'(x_n) = f'_n$
- Periodischer Spline: $s'(x_0) = s'(x_n), s''(x_0) = s''(x_n)$

Darstellung der Koeffizienten des Splines

- Momente $\mu_i = s''(x_i)$
- Funktionswerte f_i mit

$$\alpha_i = f_{i-1}, \quad \beta_i = \frac{f_i - f_{i-1}}{h_i} - \frac{h_i}{6}(\mu_i + 2\mu_{i-1}), \quad \gamma_i = \frac{1}{2}\mu_{i-1}, \quad \delta_i = \frac{\mu_i - \mu_{i-1}}{6h_i}$$

Zudem erfüllen die $n+1$ Momente die $n-1$ Gleichungen:

$$\frac{h_i}{6}\mu_{i-1} + \frac{h_i + h_{i+1}}{3}\mu_i + \frac{h_i + 1}{6}\mu_{i+1} = \frac{f_{i+1} - f_i}{h_{i+1}} - \frac{f_i - f_{i-1}}{h_i}$$

Sowie zwei Zusatzgleichungen je nach Randbedingung

- natürliche Splines: $\mu_0 = \mu_n = 0$
- hermitesche Splines: $\frac{h_1}{3}\mu_0 + \frac{h_1}{6}\mu_1 = \frac{f_1 - f_0}{h_1} - f'_0$ und $\frac{h_n}{6}\mu_{n-1} + \frac{h_n}{3}\mu_n = f'_n - \frac{f_n - f_{n-1}}{h_n}$

Berechnungsprinzip:

1. Verwende $p_i(x_{i-1}) = f_{i-1}$, $p_i''(x_{i-1}) = \mu_{i-1}$, $p_i''(x_i) = p_{i+1}''(x_i)$ und $p_i(x_i) = p_{i+1}(x_i)$.
2. Dies ergibt die Darstellung der 4 Koeffizienten.
3. Um das Gleichungssystem herzuleiten bzw. die Momente zu berechnen die Gleichung $p'_i(x_i) = p'_{i+1}(x_i)$ ausnutzen.
4. Zwischenschritt $\beta_i + 2\gamma_i h_i + 3\delta_i h_i^2 = \beta_{i+1}$.
5. Einsetzen der zuvor hergeleiteten Beziehungen liefert das Ergebnis.

Bemerkung:

Für jede Wahl der Knoten $a = x_0 < x_1 < \dots < x_n = b$ und eine der genannten Zusatzbedingungen gibt es einen eindeutigen kubischen Spline mit $s(x_i) = f(x_i)$.

Beispiel: Berechnung einer natürlichen Spline

Es gilt $\mu_0 = \mu_3 = 0$, $h_i = 1$ sowie für $i = 1, 2$:

$$\frac{h_i}{6} \mu_{i-1} + \frac{h_i+h_{i+1}}{3} \mu_i + \frac{h_{i+1}}{6} \mu_{i+1} = \frac{f_{i+1}-f_i}{h_{i+1}} - \frac{f_i-f_{i-1}}{h_i}$$

i	0	1	2	3
x_i	-1	0	1	2
f_i	5	-2	9	-4

$$\frac{1}{6}0 + \frac{2}{3}\mu_1 + \frac{1}{6}\mu_2 = \frac{9 - (-2)}{1} - \frac{-2 - 5}{1} = 18$$

$$\frac{1}{6}\mu_1 + \frac{2}{3}\mu_2 + \frac{1}{6}0 = \frac{-4 - 9}{1} - \frac{9 - (-2)}{1} = -24$$

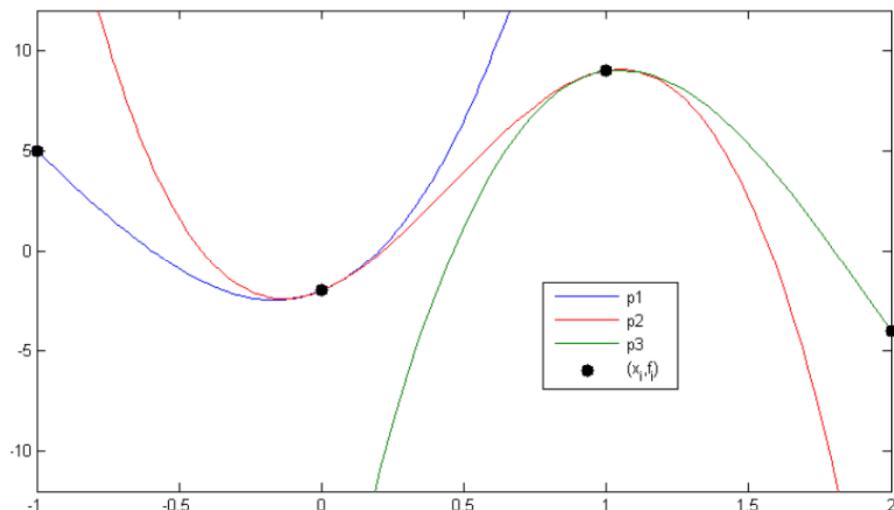
} Lösung $\mu_1=38.4$ und $\mu_2=-45.6$

Koeffizienten der Polynome

$$p_i(x) = \alpha_i + \beta_i(x-x_{i-1}) + \gamma_i(x-x_{i-1})^2 + \delta_i(x-x_{i-1})^3, i=1,2,3$$

i	1	2	3
α	5	-2	9
β	-13.4	5.8	2.2
γ	0	19.2	-22.8
δ	6.4	-14	7.6

Beispiel: Berechnung einer natürlichen Spline

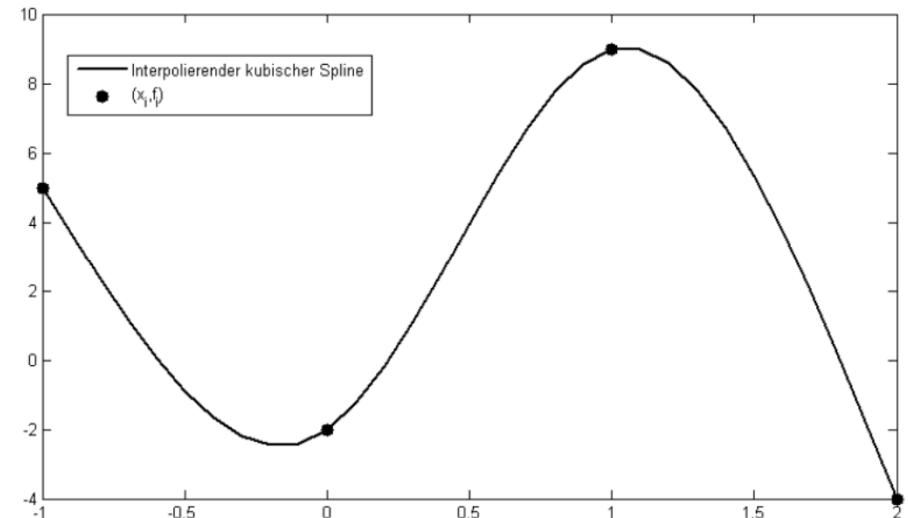


Polynome p_1, p_2, p_3

— $p_1(x) = 6.4x^3 + 19.2x^2 + 5.8x - 2$

— $p_2(x) = -14x^3 + 19.2x^2 + 5.8x - 2$

— $p_3(x) = 7.6x^3 - 45.6x^2 + 70.6x - 23.6$



Resultierende Spline

Interpolation

Auffinden einer Funktion, die gegebene Stützstellen (t_i, \mathbf{f}_i) , $\mathbf{f}_i \in \mathbb{R}^d$, $i = 0, \dots, n$, $t_i < t_{i+1}$, durchläuft und die einer gegebenen Klasse von Funktionen

$$\Psi = \{\mathbf{f}(t, c_1, \dots, c_p) : c_i \in \mathbb{R}, \mathbf{f} : \mathbb{R}^{p+1} \rightarrow \mathbb{R}^d\}$$

entstammt.

Die Freiheitsgrade c_1, \dots, c_p der Funktionen in Ψ werden durch die Interpolationsbedingung $\mathbf{f}(t_i, c_1, \dots, c_p) = \mathbf{f}_i$ gebunden.

Approximation

Auffinden einer Funktion, die gegebene Stützstellen (t_i, \mathbf{f}_i) möglichst gut annähert, jedoch nicht unbedingt durchläuft, und die einer gegebenen Klasse von Funktionen entstammt.

Anwendung

- Behandlung verrauschter Daten, d.h. Daten, die von den korrekten Werten abweichen, z.B. wegen Mess- oder Rundungsfehlern
- Die Interpolation solcher Daten würde eine Funktionenklasse mit einer großen Anzahl von Freiheitsgraden erfordern, um den schwankenden Funktionswerten zu folgen.
- Durch Vorgabe einer Funktionenklasse mit relativ wenigen Freiheitsgraden ist die Funktion möglicherweise dazu nicht in der Lage, wodurch sich im günstigen Fall ein Glättung („Tiefpassfilterung“) ergeben kann.

Funktionsapproximation von Stützstellen

Gegeben:

Stützstellen (t_i, \mathbf{f}_i) , $i = 0, \dots, n$, $t_i < t_{i+1}$, eine Klasse von Funktionen

$$\Psi = \{\mathbf{f} := \mathbf{f}(t, c_1, \dots, c_p) : c_i \in \mathbb{R}, t \in P \subseteq \mathbb{R}, \mathbf{f} : \mathbb{R}^{p+1} \rightarrow \mathbb{R}^d\}$$

sowie Normen $\|\cdot\|_a$, $\|\cdot\|_b$ im \mathbb{R}^d bzw. \mathbb{R}^{n+1} .

Gesucht:

$$c_1, \dots, c_n \text{ mit } \|(\|\mathbf{f}(t_i, c_1, \dots, c_p) - \mathbf{f}_i\|_a)_{i=0, \dots, n}\|_b \text{ minimal.}$$

Bemerkung:

Durch Einsetzen verschiedener Normen erhält man unterschiedliche Approximationen. Als besonders günstig zur Bestimmung einer approximierenden Funktion hat sich die Euklidsche Norm für $\|\cdot\|_a$ und $\|\cdot\|_b$ erwiesen:

Gaußsche Funktionsapproximation:

Funktionsapproximation $\mathbf{f}(t, c_1, \dots, c_p)$ mit $\|\cdot\|_a$, $\|\cdot\|_b$ Euklidsche Norm,

$$\text{d.h. } \sum_{j=1}^d \sum_{i=0}^n (k_j(t_i, c_1, \dots, c_p) - p_{i,j})^2 \text{ minimal,}$$

mit $\mathbf{f} = (k_1, k_2, \dots, k_d)$, $\mathbf{f}_i = (p_{i,1}, p_{i,2}, \dots, p_{i,d})$, $i = 0, \dots, n$.

Zusätzliche Spezialisierung der Funktionsmenge Ψ auf komponentenweise linear kombinierbare Funktionen.

Komponentenweise linear kombinierbare Funktionen:

Eine Menge Ψ von Funktionen $f : P \rightarrow \mathbb{R}^d$, $P \subseteq \mathbb{R}$, heißt linear kombinierbar, falls sich alle $f \in \Psi$ in der Form

$$f(t) = \sum_{k=0}^p c_k e_k(t), \quad c_i \in \mathbb{R}^d,$$

bezüglich einer festen Basis $e_i(t)$, $e_i : P \rightarrow \mathbb{R}$, $i = 0, \dots, p$, linear unabhängiger Funktionen darstellen lassen.

Beispiele für Basen:

Monome x^i \rightsquigarrow Polynomapproximation

Bernstein-Polynome $B_i^n(x)$ \rightsquigarrow Bézier-Approximation

Satz (Gaußsche Approximation mit linear kombinierbaren Funktionen)

Notwendig dafür, dass eine komponentenweise linear kombinierbare Funktion $f(t) = \sum_{k=0}^p c_k e_k(t)$ eine Gaußsche Funktionsapproximation gegebener Stützstellen $(t_0, f_0), \dots, (t_n, f_n)$ liefert, ist, dass die c_k dem Gleichungssystem

$$\sum_{k=0}^p \sum_{i=0}^n e_k(t_i) e_l(t_i) c_k = \sum_{i=0}^n f_i e_l(t_i), \quad l = 0, \dots, p,$$

genügen.

Beweis: Gesucht sind Minima von

$$d := \sum_{i=0}^n \left(\sum_{k=0}^p \mathbf{c}_k e_k(t_i) - \mathbf{f}_i \right)^* \left(\sum_{k=0}^p \mathbf{c}_k e_k(t_i) - \mathbf{f}_i \right).$$

Eine notwendige Bedingung für ein lokales Minimum ist, dass die Ableitung 0 ist:

$$\begin{aligned} \frac{\delta d}{\delta \mathbf{c}_l} &= \sum_{i=0}^n \left(\begin{pmatrix} e_l(t_i) & \dots & e_l(t_i) \end{pmatrix} \left(\sum_{k=0}^p \mathbf{c}_k e_k(t_i) - \mathbf{f}_i \right) \right. \\ &\quad \left. + \left(\sum_{k=0}^p \mathbf{c}_k e_k(t_i) - \mathbf{f}_i \right)^* \begin{pmatrix} e_l(t_i) \\ \vdots \\ e_l(t_i) \end{pmatrix} \right) \\ &= 0. \end{aligned}$$

Durch Umformen ergibt sich daraus

$$\sum_{k=0}^p \sum_{i=0}^n e_k(t_i) e_l(t_i) \mathbf{c}_k = \sum_{i=0}^n \mathbf{f}_i e_l(t_i). \square$$

Bemerkung: Falls die Funktionen e_i wie hier angenommen eine Basis bilden, dann ist dieses Gleichungssystem eindeutig lösbar.

I.1. Zeitreihen

- I.1.1. Einleitung
- I.1.2. Beispiele
- I.1.3. Einfache lineare Regression
- I.1.4. Ziele der Analyse

I.2. Prognosetechniken und Prognosefehler

- I.2.1. Einleitung
- I.2.2. Qualitative Verfahren
- I.2.3. Quantitative Verfahren
- I.2.4. Szenarioanalyse
- I.2.5. Güte der Prognose

I.3. Komponentenmodelle

- I.3.1. Einführung
- I.3.2. Glatte Komponente
- I.3.3. Saisonkomponente
- I.3.4. Analyse der Residuen
- I.3.5. Vorhersage

I.4. Interpolation und Approximation

- I.4.1. Einleitung
- I.4.2. Interpolation
- I.4.3. Approximation

I.5. Funktionale Analyse

- I.5.1. Einleitung
 - I.5.2. Histogramme
 - I.5.3. Statistische Merkmale
 - I.5.4. (Diskrete) Fourier-Transformation
 - I.5.5. Wavelet-Transformation
 - I.5.6. Phasenraumdarstellung und –einbettung
-

Analyse eindimensionaler Funktionen

Gegeben: eine Folge von Stützstellen

$$(t_i, \mathbf{f}_i), i = 0, \dots, n - 1, t_i < t_{i+1}, \mathbf{f}_i \in I\!\!R^d.$$

einer d-dimensionalen Funktion \mathbf{f} , d.h. $\mathbf{f}(t_i) = \mathbf{f}_i$.

Gesucht:

- Beschreibung von Merkmalen der Funktion, z.B. Mittelwert, Trend, Schwingungsverhalten, Konvergenzverhalten
- Modellierung der Funktion: Angabe eines Verfahrens, das die Funktion erzeugt

Anwendung bei der Visualisierung:

- Hervorheben wesentlicher Eigenschaften der Funktion
 - skalierbare Darstellung der Funktion
-

Aspekte:

- **Wahl der Stützstellen** (t_i, f_i):

gestreut: Die Differenz zwischen t_i und t_{i+1} ist variabel.

äquidistant: Die Differenz zwischen t_i und t_{i+1} ist für alle i gleich, d.h.
 $t_{i+1} - t_i = \Delta t$ für eine Konstante Δt .

- **Auflösung der Stützstellen:**

Beachtung des Abtasttheorems zur Vermeidung von Aliasen (siehe später)

- **Art der Parameter (Zeit)-Abhängigkeit:**

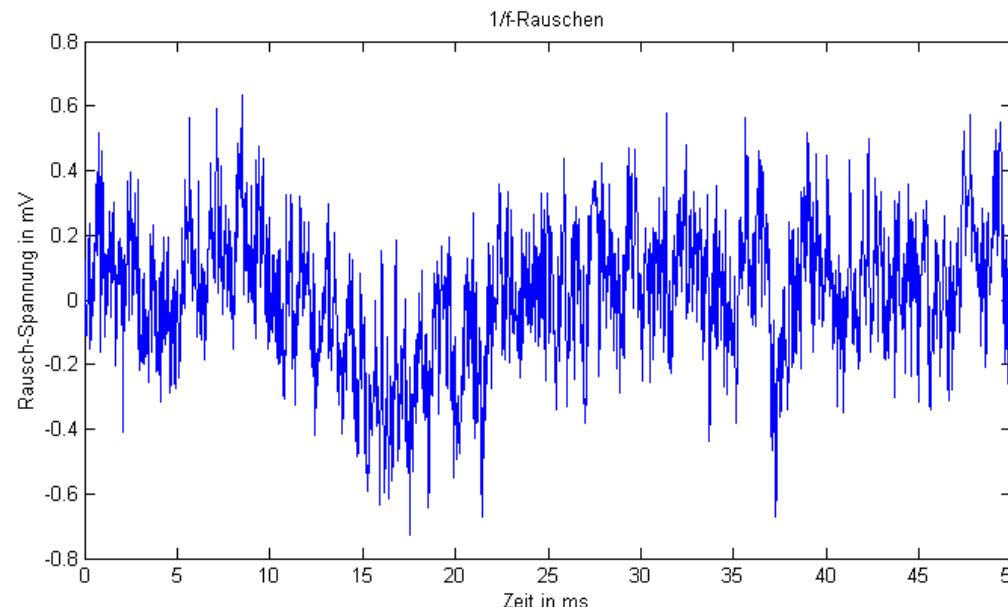
stationär: Die Funktion verhält sich an jeder Stelle des Parameterbereichs nach gegebenen statistischen Merkmalen lokal gleich.

Rauschen:

Die gegebene Funktion \hat{f} hat die Form $\hat{f} := f + r$, wobei r eine Zufallsfunktion und f eine gewöhnliche Funktion, die Originalfunktion, ist.

Möglichkeiten zur Minderung der Rauschstörung:

- subjektive Verbesserung durch Filter
- modellbasierte Restauration der Originalfunktion



Im Folgenden vorgestellte Analyse- und Visualisierungsverfahren:

- Histogramme
- statistische Merkmale
- Frequenzzerlegung (Fourier- und Wavelet-Zerlegung)
- Phasenraum
- Phasenraumeinbettung, Clustering, Dimensionsreduktion

Literatur/Quellen

- M. Sickert, Time Series and Data Analysis, Skript, Fachbereich Mathematik und Informatik, Universität Osnabrück
- A.M. Kirchberger, I. Gröller, Analysis and Visualization of Nonlinear Time Sequences, 7th Eurographics Workshop on Visualization in Scientific Computing, Springer-Verlag, 1996, 101-110

Histogramm

repräsentiert die Anzahl auftretender Werte einer eindimensionalen (diskretisierten) skalaren Funktion $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$

Aufbau:

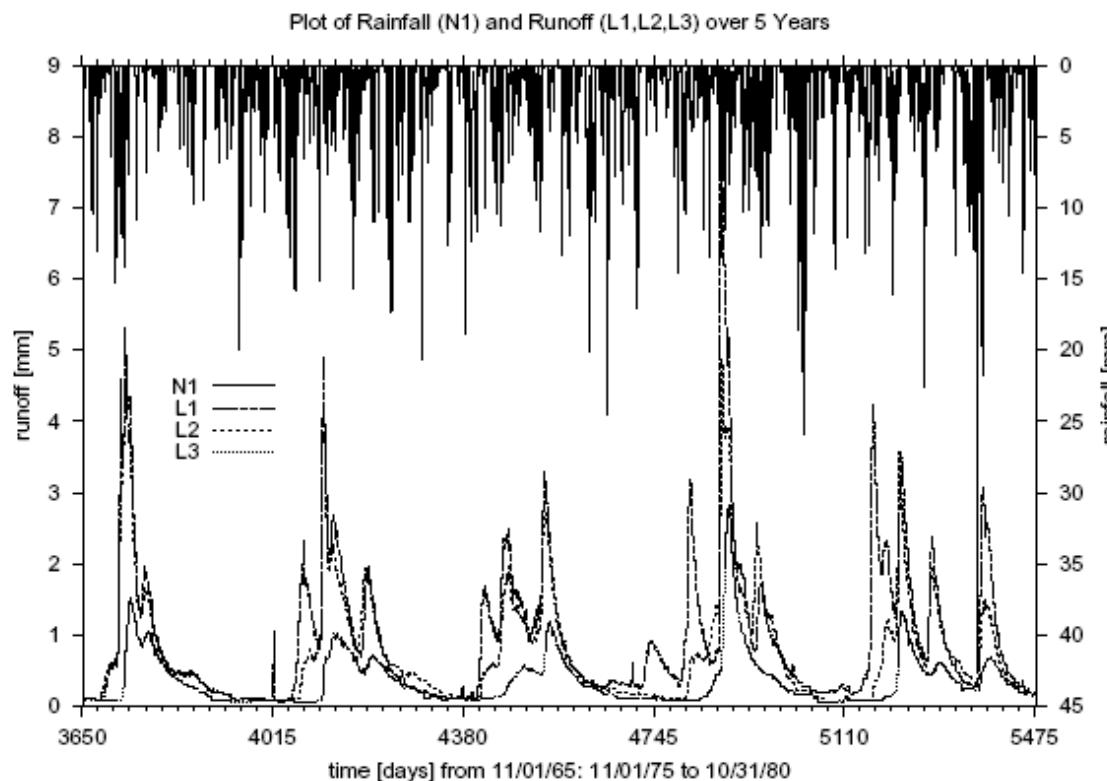
Zerlegung des Wertebereichs der Funktion in Intervalle (Bins), die üblicherweise gleiche Länge haben

Zuordnung der Anzahl von Stützstellen, deren Funktionswert in ein gegebenes Intervall fällt, an diese Intervalle

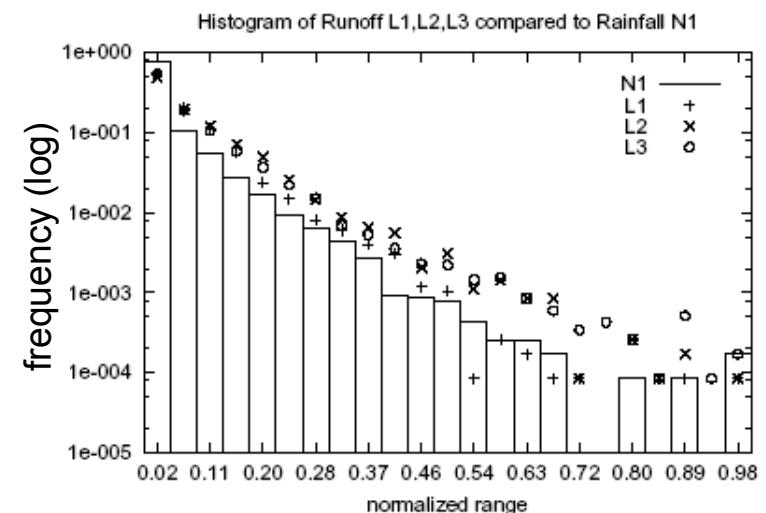
Regel zur Wahl der Anzahl der Intervalle:

$1/\sqrt{n}$, n die Anzahl der Stützstellen, jedoch nicht weniger als 5

Beispiel: Ökologische Zeitreihe mit vier Komponenten



Histogramme von N1, L1, L2, L3:



Zeitreihe mit vier Komponenten über fünf Jahre:

- Niederschlag (N1)
- Laubbaum-Lysimeter (L2)
- Gras-Lysimeter (L1)
- Nadelbaum-Lysimeter (L3)

Statistische Merkmale

- **Moment der Ordnung** $q, q > 0$, einer Zufallsvariablen X mit Dichtefunktion $p : I\!\!R \rightarrow I\!\!R$

$$m_q = E(X^q) := \int_{-\infty}^{\infty} x^q p(x) dx \text{ (kontinuierlich)}$$

$$m_q := \sum_{i=0}^{n-1} x_i^q p_i \text{ (diskret)}$$

für $q = 1$: **Mittelwert**

- **Zentralmoment der Ordnung** $q, q > 0$, von

$$\hat{m}_q = E((X - m_1)^q) = \int_{-\infty}^{\infty} (x - m_1)^q p(x) dx \text{ (kontinuierlich)}$$

$$\hat{m}_q := \sum_{i=0}^{n-1} (x_i - m_1)^q p_i \text{ (diskret)}$$

für $q = 2$: **Varianz**

Anwendung auf Funktionen $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$

Berechnung statistischer Merkmale der Verteilung der auftretenden Funktionswerte:

Mittelwert von f :

$$m_1 := \sum_{i=0}^{n-1} f(i) \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=0}^{n-1} f(i)$$

Varianz von f :

$$\hat{m}_2 := \frac{1}{n} \sum_{i=0}^{n-1} (f(i) - m_1)^2$$

Anwendung:

Definition von Stationarität von $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$

- *Mittelwert-Stationarität*: Der gefensterte Mittelwert von f ist konstant.
- *Varianz-Stationarität*: Die gefensterte Varianz von f ist konstant.
- *Strikte Stationarität*: Alle gefensterten Momente von f sind konstant.

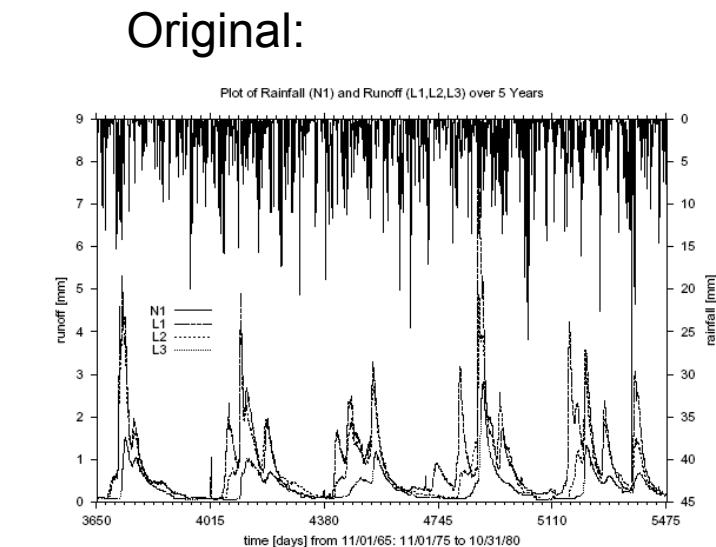
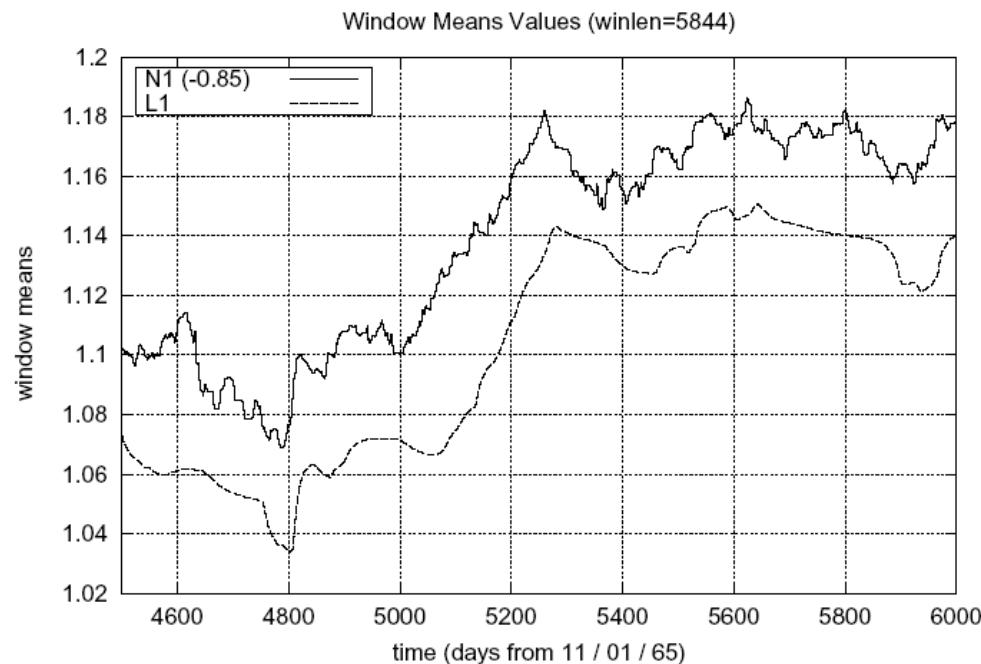
Fensterung eines Operators G :

$G(j, k)[f] :=$ das Ergebnis von G angewendet auf den Ausschnitt der Funktion, der durch Funktionswerte $f(i)$ mit $j - k \leq i \leq j + k$ gegeben ist.

Beispiele für Operatoren: Momente, Zentralmomente

Problem: Erkenntnis über Stationarität von der Fenstergröße abhängig.

Beispiel: Gefensterter Mittelwert von N1 (-0.85) und L1 des Öko-Beispiels



- **Auto-Kovarianz einer diskreten Funktion** $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$:

$$c(j) := \frac{1}{n} \sum_{i=0}^{n-1-j} (f(i) - \bar{f})(f(i + j) - \bar{f}), j = 0, \dots, n - 2$$

\bar{f} der Mittelwert von f , bestimmt den Grad der inneren linearen Korrelation der Funktion durch Vergleich mit einer Verschiebung von sich selbst.

- **Auto-Korrelation einer Funktion**

$$f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}: \\ r(j) := \frac{c(j)}{c(0)}$$

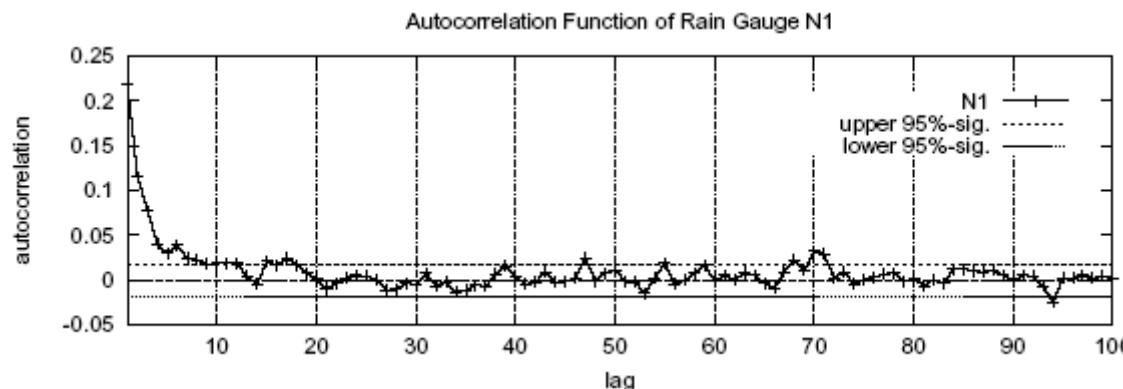
skaliert die Auto-Kovarianz auf das Intervall $[-1, 1]$:

Wert 0: keine Korrelation vorhanden

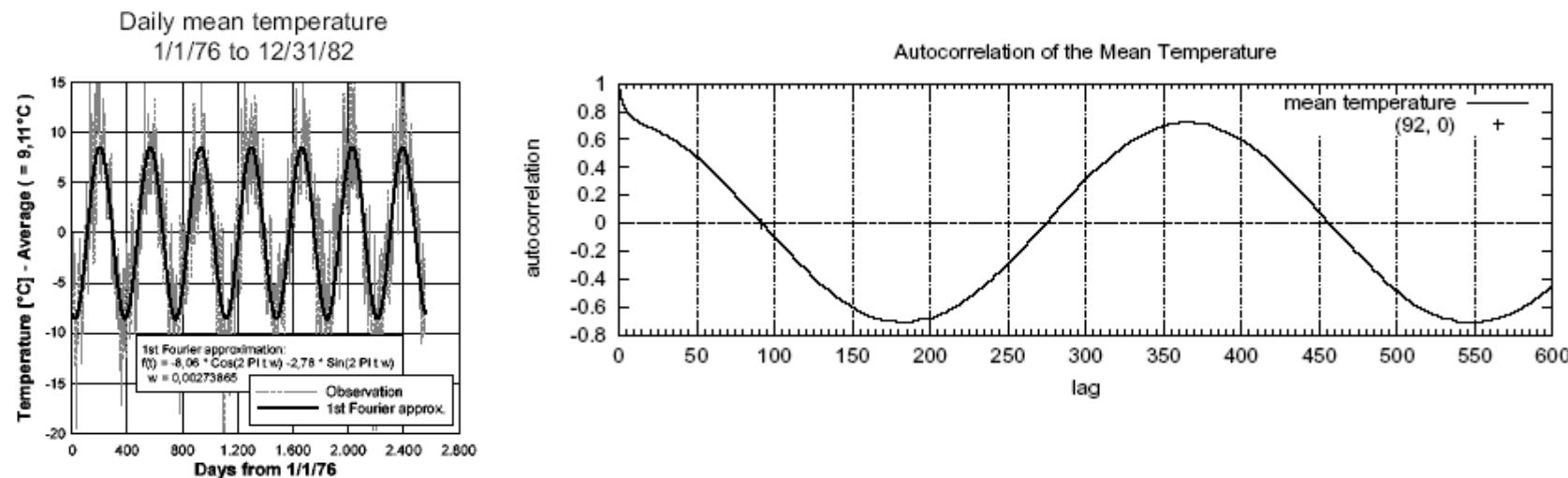
Wert -1: Antikorrelation

Wert 1: maximale Korrelation.

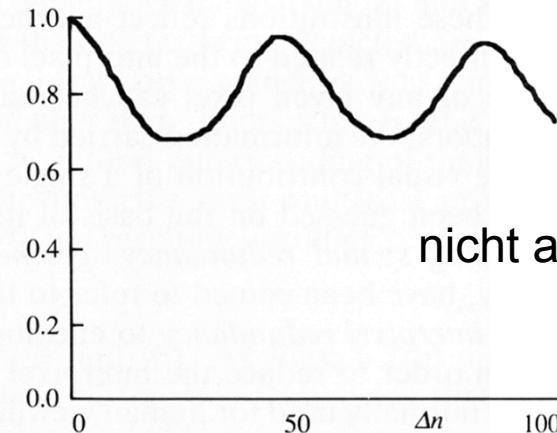
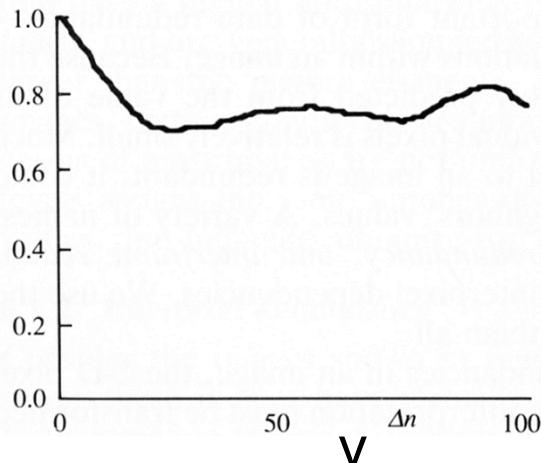
Beispiel: Autokorrelationsfunktion der Niederschlagsmenge N1 des Öko-Beispiels



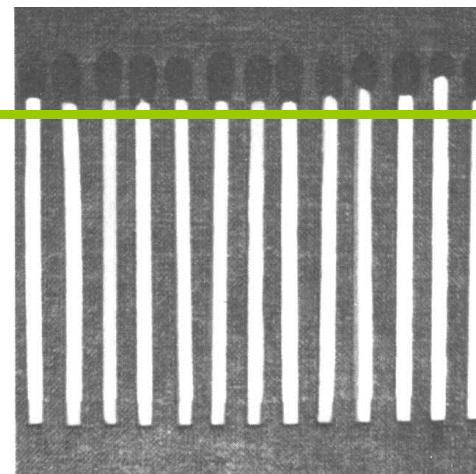
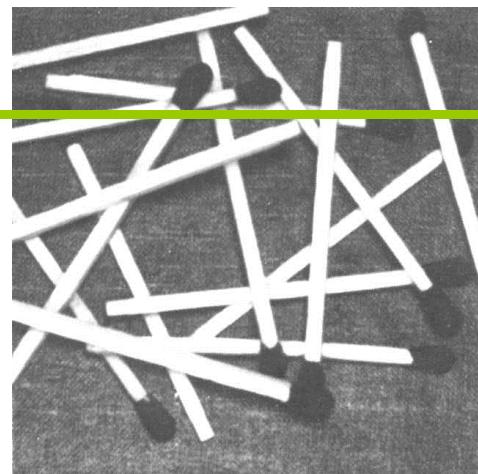
Beispiel: Autokorrelationsfunktion einer Außentemperaturmessung über 600 Tage



Beispiel: Autokorrelationsfunktion der Grauwertfunktionen längs der markierten Zeile in den beiden Bildern



nicht auf \bar{f} normiert



Interpretation der Autokorrelationsfunktion

- **Periodische Autokorrelationsfunktionen:**

lassen in der Regel auf periodische Funktionen schließen

- **Nichtperiodische Autokorrelationsfunktionen:**

Nach 0 abfallende Autokorrelationsfunktionen ergeben sich üblicherweise für verrauschte oder chaotische Funktionen.

Kriterium für Intervalle von Verschiebungen, in denen praktisch keine Verschiebung mit Korrelation vorhanden ist:

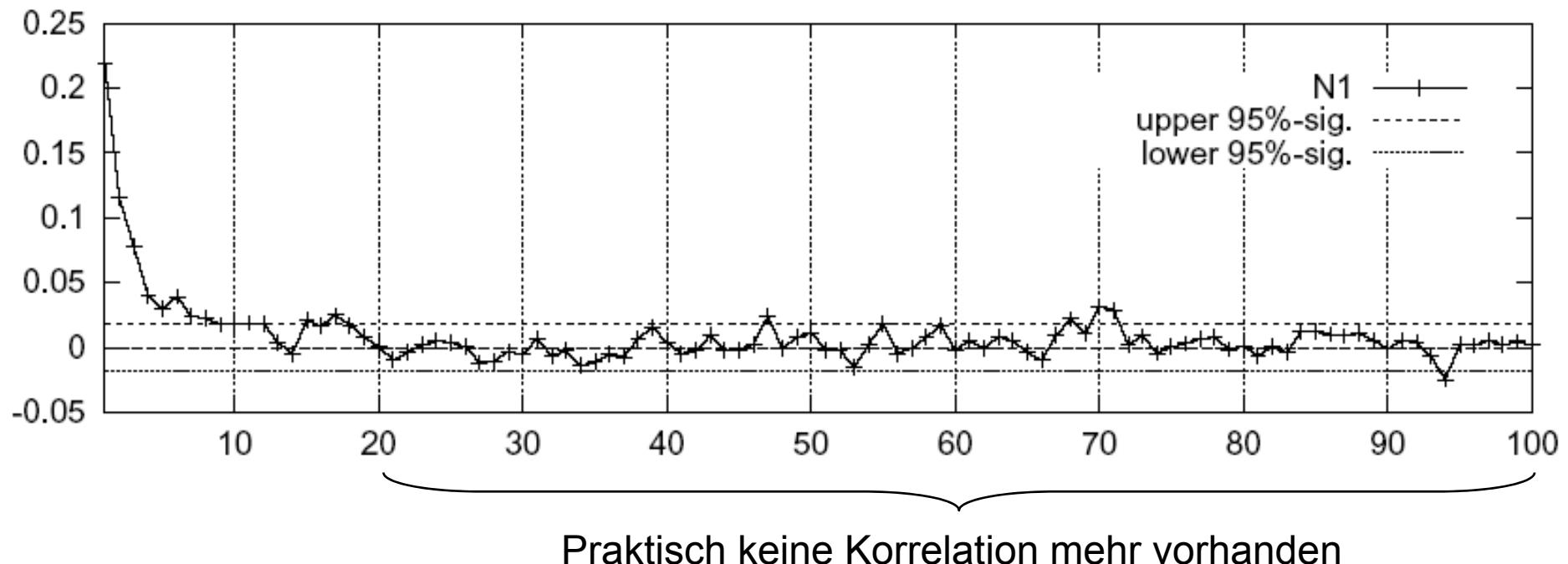
95% der Autokorrelationswerte liegen in

$[-1/n - 1.96/\sqrt{n}, -1/n + 1.96/\sqrt{n}]$, n die Länge des Intervalls.

Begründung: Die Autokorrelationsfunktion eines weißen Rauschens hat diese Eigenschaft.

Beispiel: Autokorrelationsfunktion der Niederschlagsmenge und das 95%-Intervall

$$l_{lo|up} = -1/11641 \mp 1.96/\sqrt{11641} \rightarrow l_{lo} \approx -0.0183, l_{up} \approx 0.0181$$



Eindimensionale diskrete Fourier-Transformierte (DFT)

einer Funktion $f : \{m \mid m = 0, 1, \dots, M - 1\} \rightarrow \mathbb{R}$:

$$F(k) := \sum_{m=0}^{M-1} f(m) \cdot \exp[-2\pi i(km/M)],$$

$$\begin{aligned} k &= 0, \dots, M - 1, \quad i := \sqrt{-1} \\ \exp[i\alpha] &= \cos \alpha + i \sin \alpha. \end{aligned}$$

Inverse eindimensionale DFT:

$$f(m) := \frac{1}{M} \sum_{k=0}^{M-1} F(k) \cdot \exp[2\pi i(km/M)],$$

$$m = 0, \dots, M - 1.$$

Interpretation der Fourier-Transformierten:

Falls die Funktion f reell ist, beschreibt die diskrete Fourier-Transformation die Darstellung von f als Summe von diskretisierten sinusförmigen Schwingungen unterschiedlicher Frequenz und Phase:

$F(k)$ bedeutet, dass die Schwingung der Frequenz k mit der Amplitude $|F(k)|/M$ und der Phasenverschiebung $\phi(k)$ auftritt, wobei

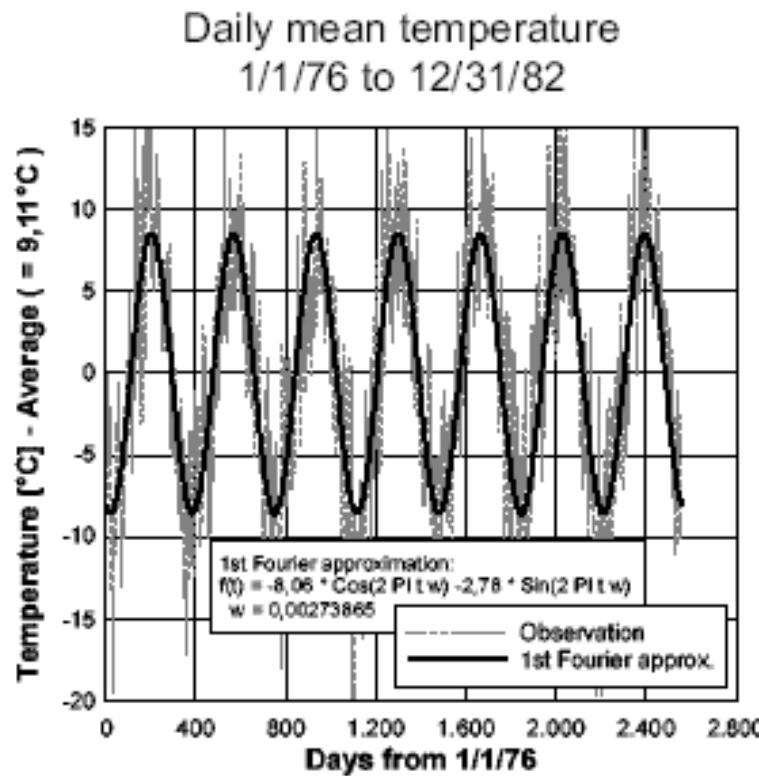
$$F(k) = |F(k)| \exp[i\phi(k)].$$

Periodogramm: Quadrat der Beträge der Fourier-Transformierten: $I(k) := |F(k)|^2$

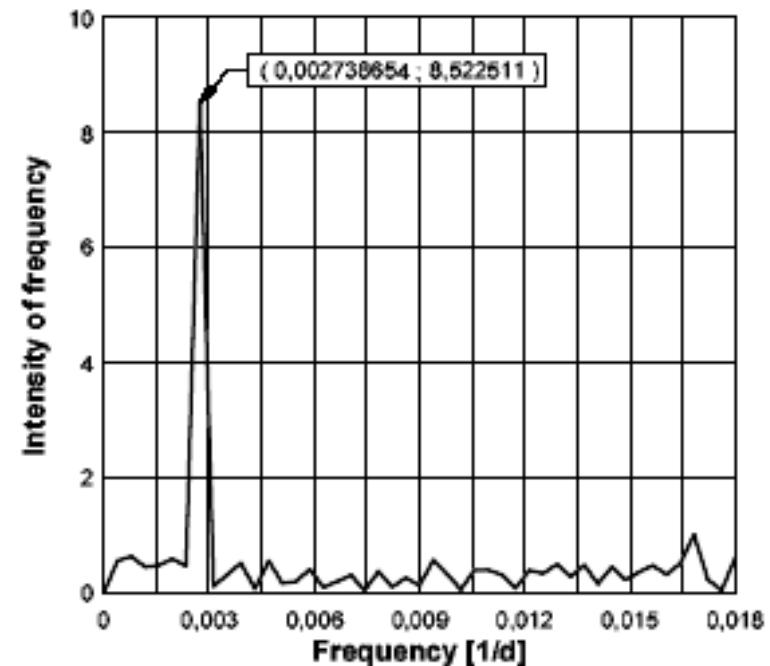
Ortsbereich: Definitionsbereich der gegebenen Funktion

Frequenzbereich: Definitionsbereich der Fourier-Transformierten

Beispiel:



Periodogram (Fast Fourier Transform)
Daily mean temperature
1/1/76 to 12/31/82



Darstellung der Originalfunktion und einer Approximation durch die Sinusfunktion (linkes Bild), deren Wert im Periodogramm (rechtes Bild) am größten ist.

Berechnung der diskreten Fourier-Transformation

Fast-Fourier-Transformationsalgorithmus (FFT-Algorithmus) ($\omega_M = \exp(-2\pi i/M)$):

$$F(k) = \begin{cases} F_g(k) + \omega_M^k F_u(k) & \text{für } 0 \leq k \leq \frac{M}{2} - 1, \\ F_g(k - \frac{M}{2}) - \omega_M^{k - \frac{M}{2}} F_u(k - \frac{M}{2}) & \text{für } M/2 \leq k \leq M - 1 \end{cases}$$

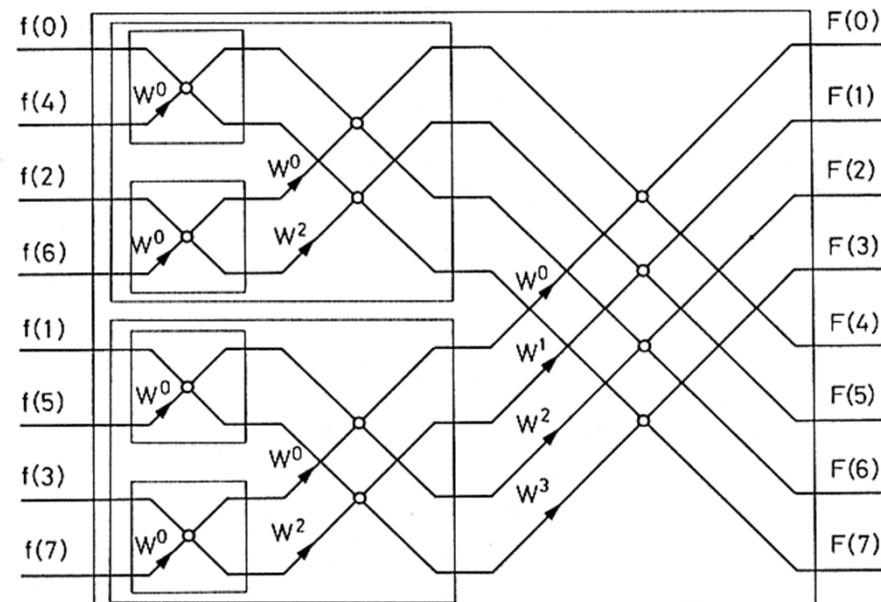
F_g : DFT über den „geraden“ Funktionswerten

F_u : DFT über den „ungeraden“ Funktionswerten

Für $M = 2$:

$$\begin{aligned} F(0) &= f(0) + f(1) \\ F(1) &= f(0) - f(1) \end{aligned}$$

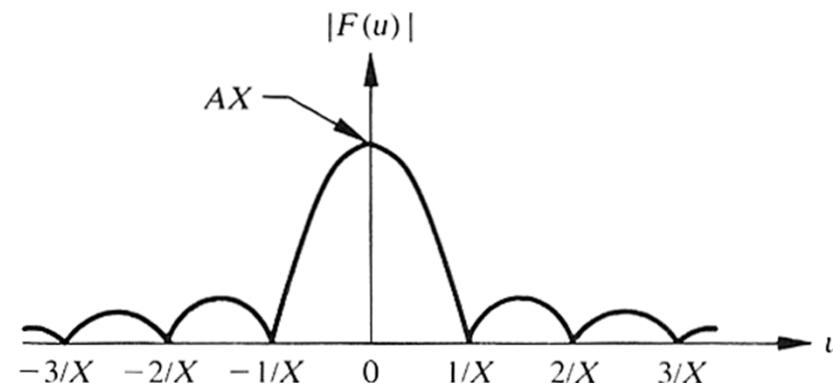
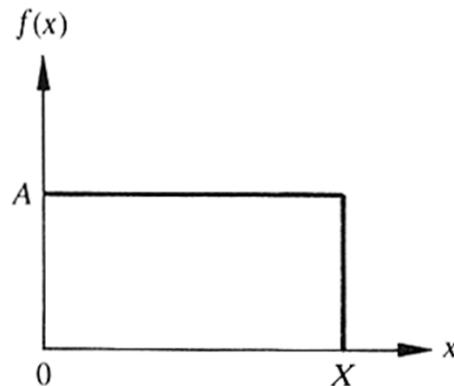
Zeitaufwand: $O(n \log n)$



Kontinuierliche Fourier-Transformierte (FT) einer Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$:

$$F(u) := \int_{-\infty}^{\infty} f(x) \exp[-iux] dx, \quad u \in \mathbb{R}, \quad i := \sqrt{-1}.$$

anschaulich: F liefert die Spektralzerlegung des „Signals“ f .



Inverse kontinuierliche FT: $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(u) \exp[iux] du, \quad x \in \mathbb{R}$

Diskrete Fourier-Transformierte als Approximation der kontinuierlichen Fourier-Transformierten

$$\begin{aligned}
 F(u) &= \int_{-\infty}^{\infty} f(x) \exp[-iux] dx \\
 F(u) &= \int_{-\infty}^{\infty} f(x) \exp[-iux] dx \\
 &= \int_0^{2\pi} f(x) \exp[-iux] dx \\
 &\approx \sum_{m=0}^{M-1} f(2\pi m/M) \exp[-iu(2\pi m/M)] / M \\
 &= \frac{1}{M} \sum_{m=0}^{M-1} f(2\pi m/M) \exp[-2\pi i u m / M]
 \end{aligned}$$

wobei $f : [0, 2\pi] \rightarrow \mathbb{R}$.

Bemerkungen:

- Bei gegebenem Wert von u hängt der Approximationsfehler von der Größe von M ab. Bei zu kleinem M werden die Sinus- und Cosinus-Terme so ungenau durch die Abtastwerte repräsentiert („Alias-Effekt“), dass der Fehler bei der Integralannäherung sehr groß wird.
- Das erklärt die Beobachtung, dass für $u > M - 1$ die Approximationswerte an u und $u - M$ übereinstimmen (Periodizität), wohingegen $F(u)$ und $F(u - M)$ sehr unterschiedliche Werte haben können.

Abtasttheorem:

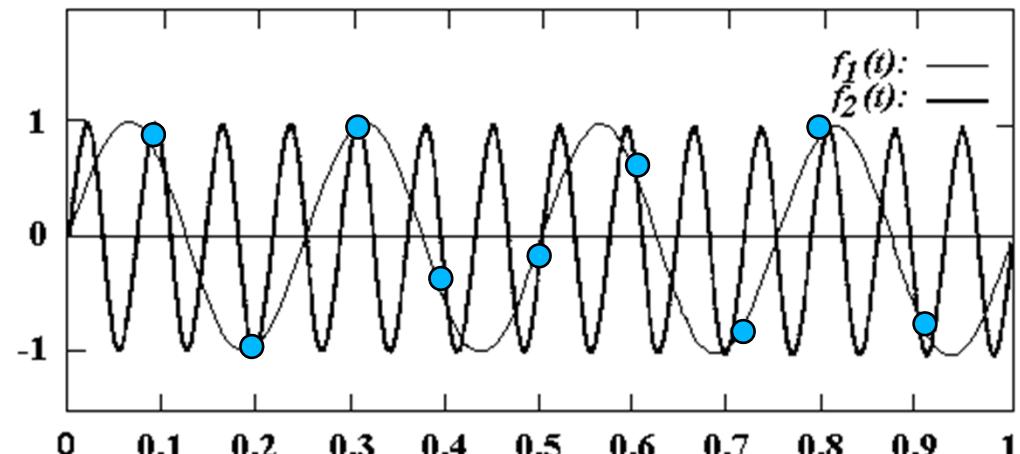
Die ursprüngliche kontinuierliche Funktion ist aus der diskreten Abtastfunktion rekonstruierbar ist, wenn die Abtastung mit **mindestens einer doppelt so hohen Frequenz wie der** in der gegebenen Funktion auftretenden **höchsten Frequenz** geschieht. Andernfalls können Störungen, sogenannte Aliaseffekte, auftreten.

Beispiel:

$$f_1(t) = \sin 8\pi t,$$

$$f_2(t) = \sin 28\pi t$$

stimmen an den Abtaststellen $i/10, i = 0, \dots, 9$ überein und können daher abgetastet nicht unterschieden werden.

**Bemerkung:**

Falls das ursprüngliche Signal beliebig hohe Frequenzen enthält, können Aliaseffekte dadurch vermieden werden, dass die Frequenzen durch einen Tiefpassfilter begrenzt werden. Dadurch kann das Bild jedoch unscharf werden.

Tiefpassfilter:

- **Im Ortsbereich:**

Ersetzung des Funktionswerte $f(i)$ durch ein gewichtetes Mitteln der Funktionswerte in einem Intervall $i - k, \dots, i + k$ um i .

Die Gewichte sollten positiv sein und die Summe 1 ergeben.

- **Im Frequenzbereich:**

Verminderung der Amplituden der hohen Frequenzen in der Fourier-Transformierten und anschließendes Rücktransformieren in den Ortsbereich.

Diskrete Wavelet-Transformation einer diskreten Funktion

$f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$:

$$W(j, s) = \sum_{k=0}^{n-1} f(k) \cdot \overline{\psi}\left(\frac{k-j}{s}\right), \quad j = 0, \dots, N - 1, 1 \leq s \leq S$$

wobei das sogenannte Mutter-Wavelet $\psi : \mathbb{R} \rightarrow \mathbb{R}$

1. quadratisch integrierbar mit endlichem Wert ist
2. eine kontinuierliche Fourier-Transformierte Ψ besitzt mit

$$0 < \int_{-\infty}^{\infty} |\Psi(u)|/|u| du < \infty$$

3. einen verschwindenden Mittelwert $\int_{-\infty}^{\infty} \psi(u) du = 0$ hat

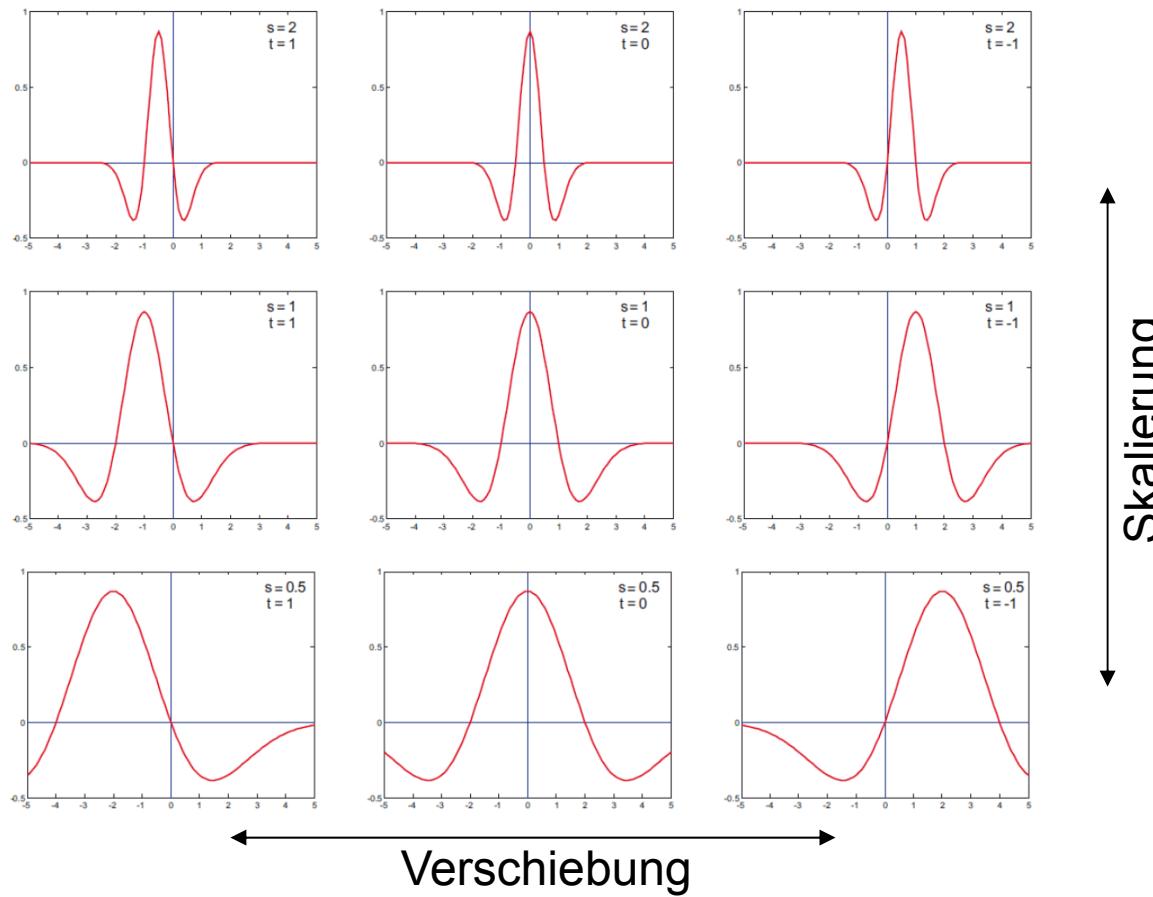
$S \in \mathbb{N}$ bezeichnet die Anzahl der analysierten Skalen,

$\overline{\psi}$ das konjugiert Komplexe bei einer komplexwertigen Funktion.

Bedingungen 1 und 3 bedeuten, dass $\psi(u)$ für betragsmäßig große u sehr klein ist und, dass negative und positive Funktionswerte angenommen werden.

Beispiel für eine kontinuierliche Wavelet-Funktion: Mexikanischer Hut

$$\psi(t) := \frac{2}{(9\pi)^{1/4}}(1 - t^2)e^{-t^2/2}$$



Bemerkungen

- **Randartefakt:**

Im Bereich des linken und rechten Randes des Definitionsbereichs der Funktion f, k nahe bei 0 oder nahe bei $n - 1$, ragt der signifikante Bereich des Definitionsbereichs von $\bar{\psi}$ über den Definitionsbereich von f hinaus.

Dadurch wird f nur unvollständig mit $\bar{\psi}$ gefiltert.

Dieser Bereich wird als Randartefakt bezeichnet und ist für die Datenanalyse kritisch.

- **Leistungsspektrum:** $P(j, s) := |W(j, s)|^2$

Interpretation der diskrete Wavelet-Transformation

- Die Wavelet-Transformation in der hier vorgestellten Definition ist eine Kreuzkorrelation der gegebenen Funktion f mit skalierten Wavelet-Funktionen.

Kreuzkorrelation:
$$f \circ \circ g(j) := \sum_{k=0}^{p-1} f(k) \cdot \overline{g(k-j)}$$

wobei

$$f : \{0, \dots, n-1\} \rightarrow \mathbb{R}, \quad g : \{0, \dots, n-1\} \rightarrow \mathbb{R}, \quad p := \max\{m, n\}.$$

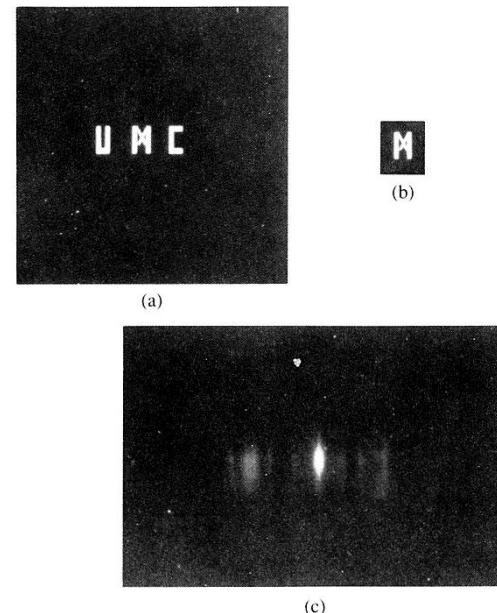
- Die Wavelet-Transformation kann damit so verstanden werden, dass sie die lokale Korrelation der gegebenen Funktion f mit der Wavelet-Funktion in unterschiedlichen Skalierungen ausdrückt.

Ein hoher Wert von $W(j, s)$ deutet eine große Übereinstimmung von f und ψ , skaliert mit s , an der Stelle j an.

Kreuzkorrelation:

$$\phi_{fh}(m, n) := f(m, n) \circ \circ h(m, n) := \sum_{\xi=0}^{M-1} \sum_{\eta=0}^{N-1} f(\xi, \eta) \bar{h}(\xi - m, \eta - n)$$

Anwendung der Korrelation:



Berechnung mit dem Faltungssatz:

$$\phi_{fh}(m, n) = f(m, n) \circ \circ h(m, n) \stackrel{M, N}{\circ \circ \bullet} F(k, l) \overline{H}(k, l) =: \Phi_{fh}(k, l)$$

Berechnung der Wavelet-Transformation

1. Die Interpretation als Faltung erlaubt die effiziente Berechnung der Wavelet-Transformation durch Anwendung des Faltungssatzes und des Fast-Fourier-Transformationsalgorithmus.

Faltungssatz: Die Fourier-Transformation der Faltung zweier Funktionen ist gleich dem Produkt der Fourier-Transformation der beiden Funktionen.

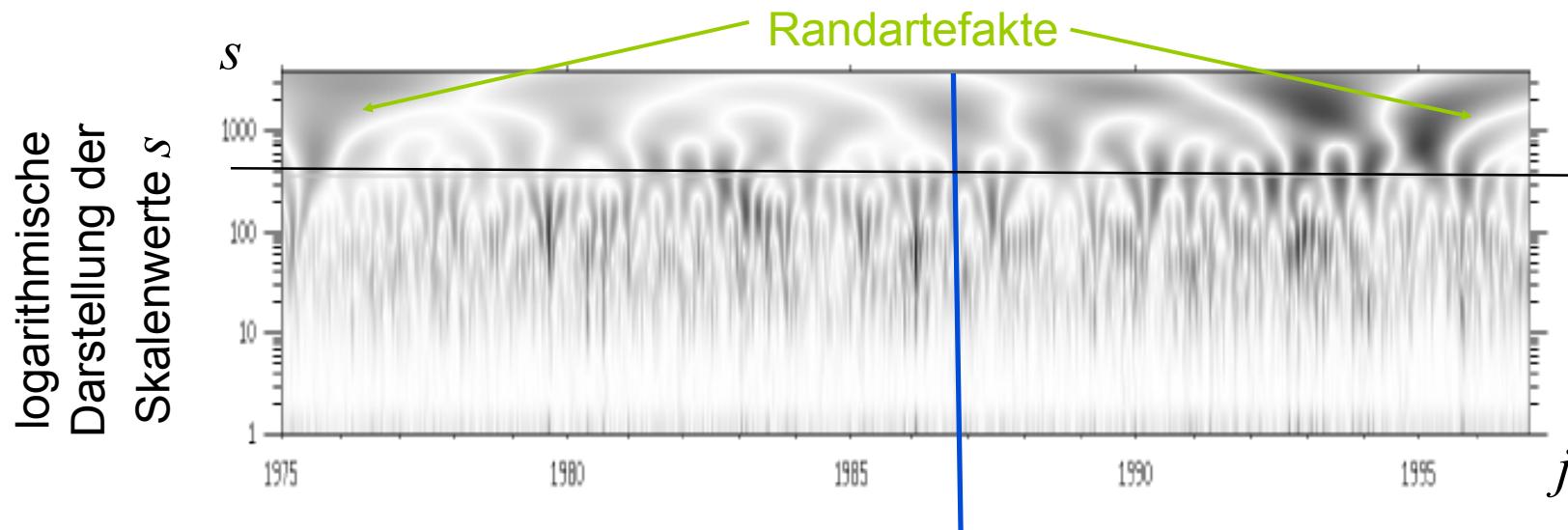
siehe Vorlesungen „Mensch-Maschine-Interaktion“ oder „Digitale Bildverarbeitung“

2. Mittels eines speziellen Wavelet-Transformationsalgorithmen ist die diskrete Wavelet-Transformation mit linearem Zeitaufwand zu berechnen.

Literatur

Gonzalez, Woods, Digital Image Processing, 2. Auflage, Kap. 7.3.2

Visualisierung der Wavelet-Transformation

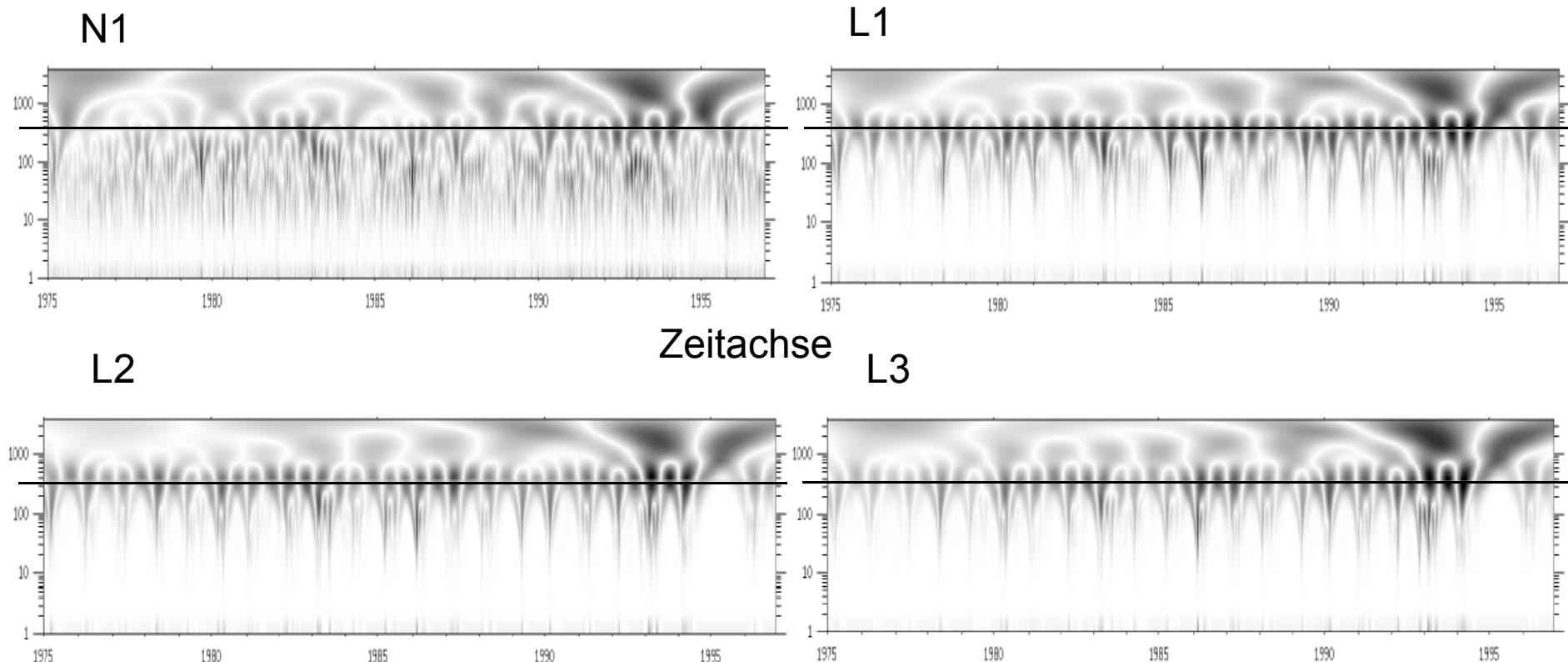


Darstellung der Stärken $|W(j,s)|^2$ durch einen Grauwert:
hell = geringer Wert, dunkel = hoher Wert. Dabei werden jeweils
sieben Spalten zu einem Wert pro Zeile zusammengefasst.

Die waagrechte Linie markiert die 365-Tage-Wavelet-Auflösung.

Beispiel:

Darstellung der Wavelet-Transformation der Öko-Daten, Zeitraum 8036 Tage

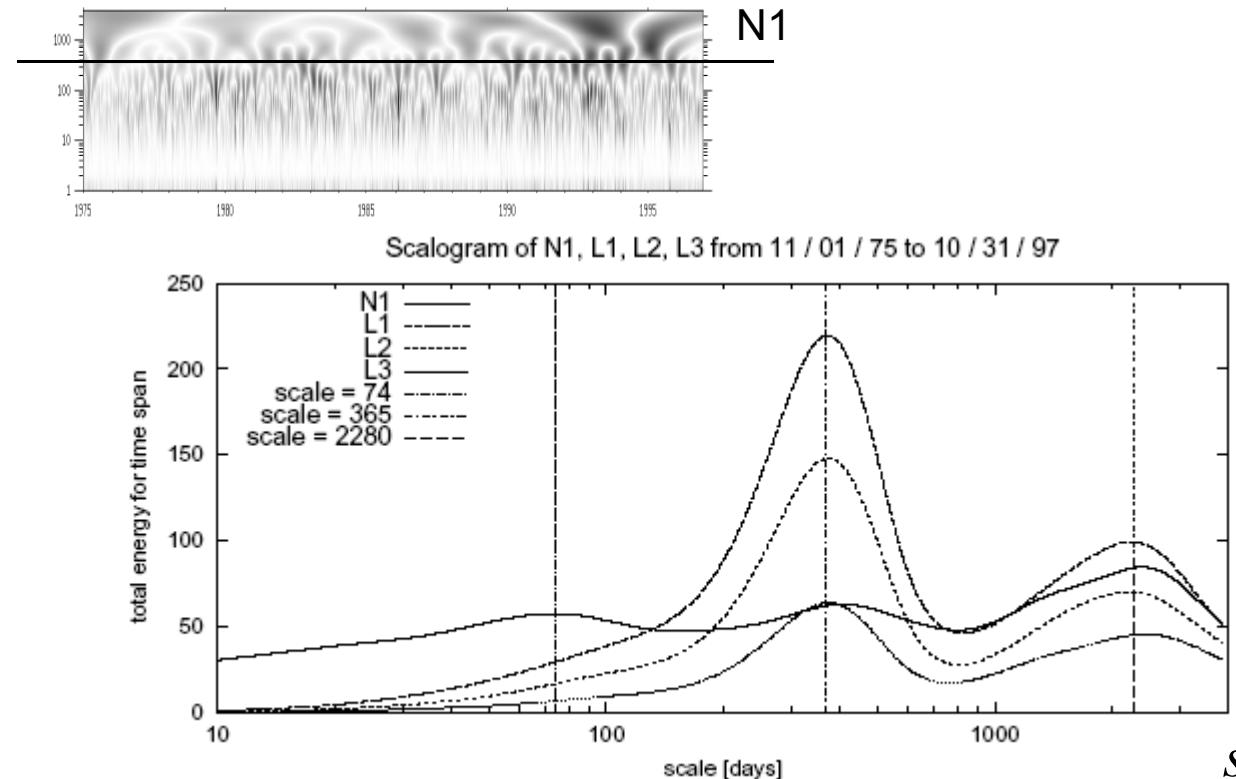


Auf der Linie der 365-Tage-Wavelet-Auflösung ist ein periodisches Verhalten mit 365-Tage-Periode wahrnehmbar.

Skalogramm: Summe (Integral) der $|W(j, s)|^2$ über j für festgehaltenes s

Beispiel:

Skalogramme
der Ökodaten



Interpretation: Die Maxima geben die Skalen wieder, deren Auflösungen am stärksten in der gegebenen Funktion präsent sind.

Im Beispiel sind beim Niederschlag neben der 365-Tage-Skala auch noch zwei weitere Skalen als relevant erkennbar (6 Jahre bzw. 74 Tage).

Phasenraumdarstellung und Phasenraumeinbettung

- Phasenraumdarstellung
- Verschiebungseinbettung
- Weitere Visualisierungsverfahren für Phasenraumdarstellungen
- Filterung einer diskreten Funktion im Phasenraum

Phasenraumdarstellung einer kontinuierlichen Funktion:

Gegeben sei eine Funktion $f : I \rightarrow \mathbb{R}^d, I \subseteq \mathbb{R}$.

Dann heißt die Kurve in Parameterdarstellung, die von f im d -dimensionalen Raum erzeugt wird, die **Phasenraumtrajektorie** von f .

Der Raum, dessen Koordinatenachsen den Komponentenfunktionen von f entsprechen, heißt **Phasenraum**.

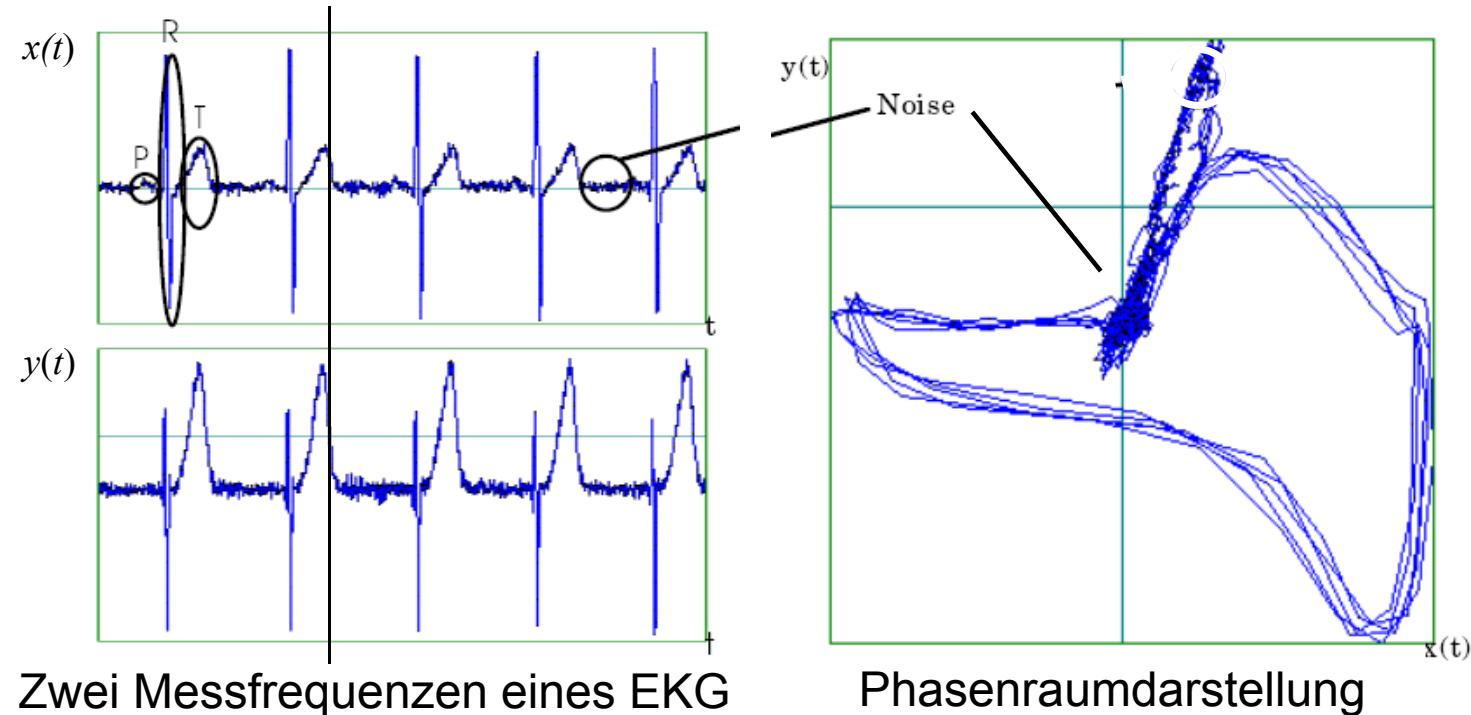
Beispiel: Der Phasenraum der Funktion $f(t) := \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, t \in [0, 2\pi)$

ist zweidimensional. Die Phasenraumtrajektorie ist der Einheitskreis.

Phasenraumdarstellung einer diskreten Funktion:

Gegeben sei eine diskrete Funktion $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}^d$.

Dann heißt der Polygonzug mit den Knoten $f(k), k = 0, 1, \dots, n - 1$, die **Phasenraumtrajektorie** von f .

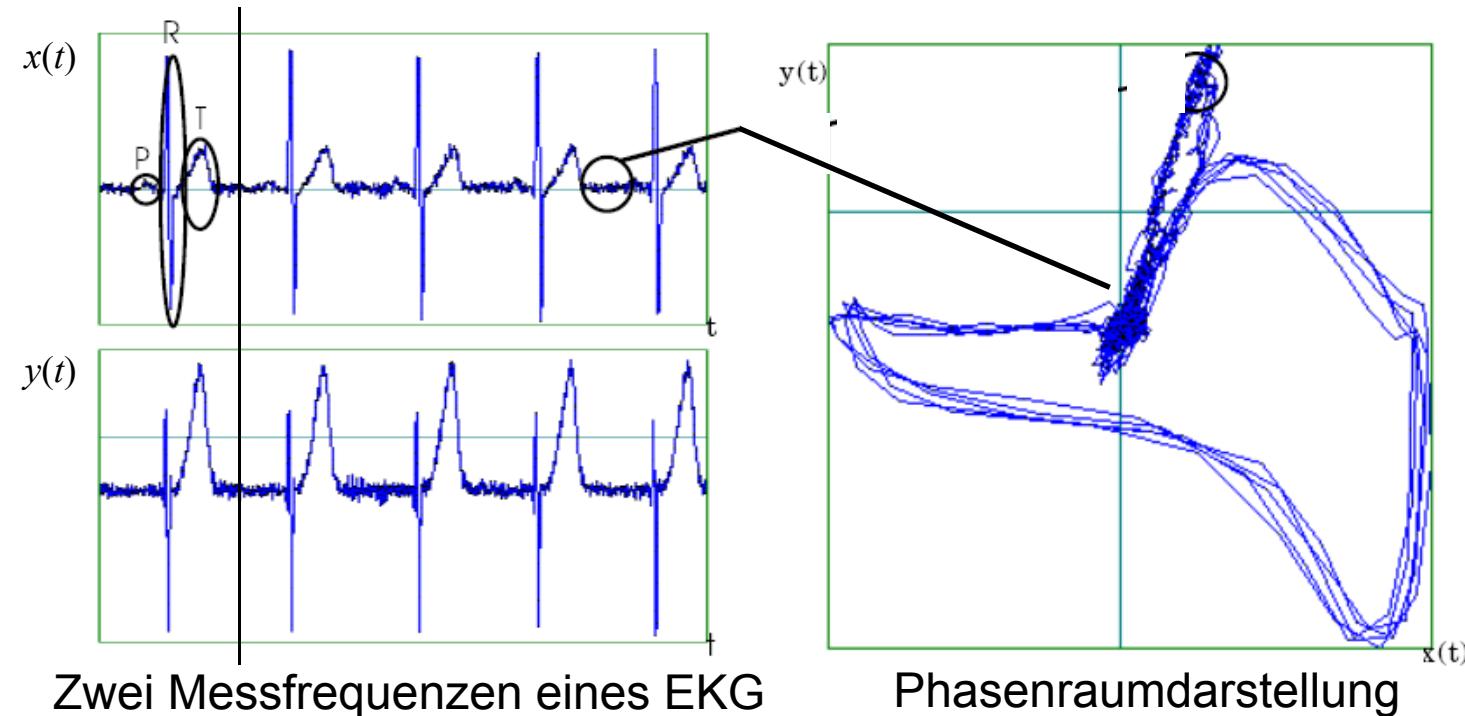
Beispiel:

Anwendung der Phasenraumdarstellung

- **Auffinden von Attraktoren**

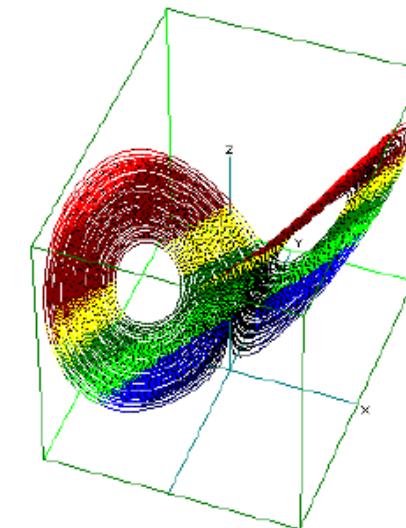
Attraktor einer diskreten Funktion: Gebiet im Phasenraum mit Anziehungscharakter für die Phasenraumtrajektorie.

Beispiel:



Typen von Attraktoren:

- Gewöhnliche Attraktoren:
 - anziehende Punkte
 - isolierte geschlossene Grenzzyklen
- Seltsame Attraktoren:
fraktale Punktmenge



Beispiel: Lorenz-Attraktor

Verschiebungseinbettung

Gegeben seien

- eine diskrete Funktion $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$.
- eine Dimension $d > 0$,
- eine Verschiebung $j > 0$.

Dann heißt die d -dimensionale Trajektorie

$$\mathbf{V}(d, j, i) := (f(i), f(i+j), \dots, f(i + (d-1) \cdot j)), \quad i = 1, \dots, n-1-(d-1) \cdot j$$

Verschiebungseinbettung von f in einen d -dimensionalen Phasenraum.

Anwendungen der Verschiebungseinbettung

a) Autokorrelationsanalyse:

Verschiebungseinbettung einer Funktion $f : \{0, 1, \dots, n - 1\} \rightarrow \mathbb{R}$.
in einen zweidimensionalen Phasenraum:

$$\mathbf{V}(j, i) := (f(i), f(i+j)), \quad i = 1, \dots, n - 1 - j$$

An der Verteilung der Punkte lässt sich der Grad der Korrelation bei der Verschiebung j erkennen: große Streuung entspricht geringer Korrelation

b) Analyse von dynamischen Systemen

I.5.6. Phasenraumdarstellung

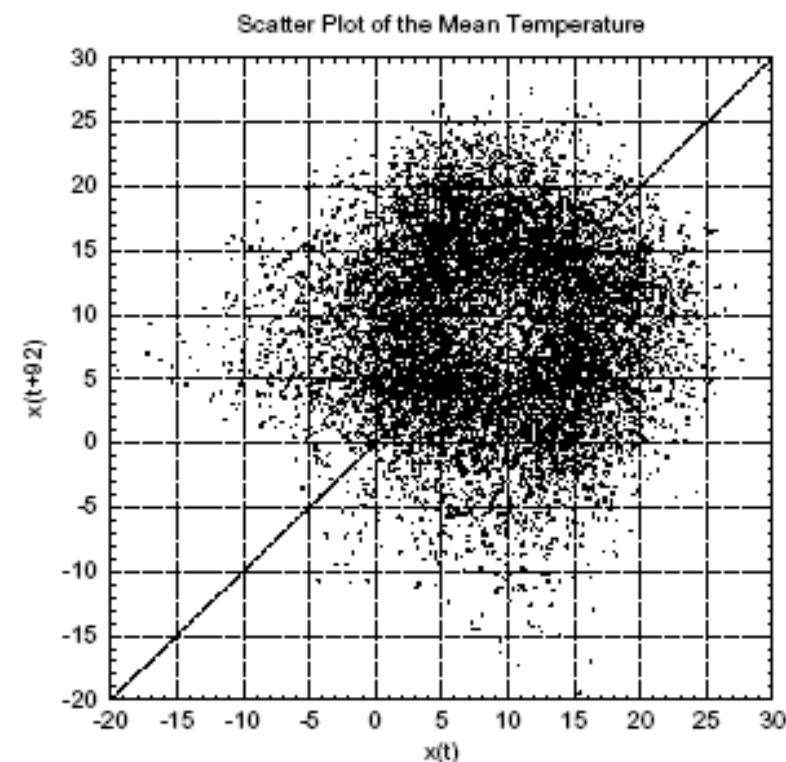
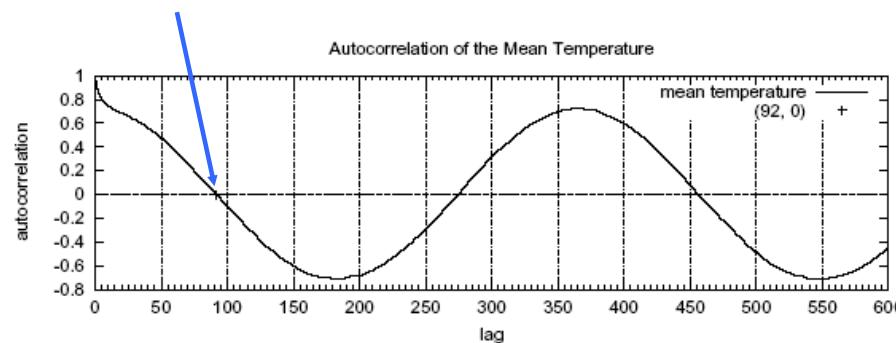
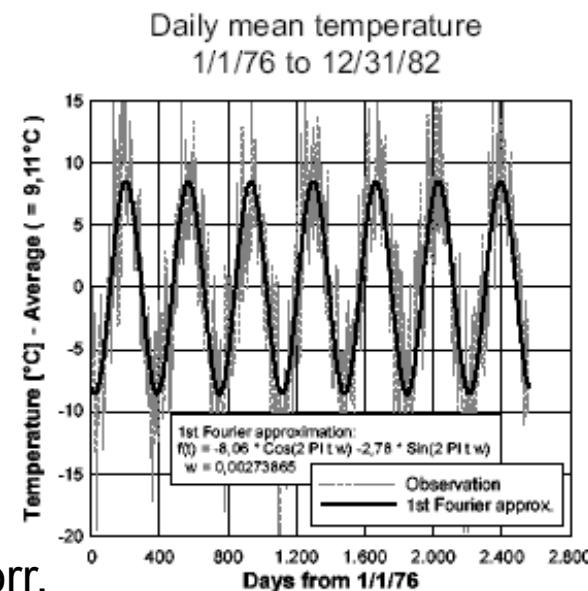
I.5.6.2. Verschiebungseinbettung

Beispiel:

Phasenraumeinbettung der Temperaturfunktion mit $d = 2$ und $j = 92$.

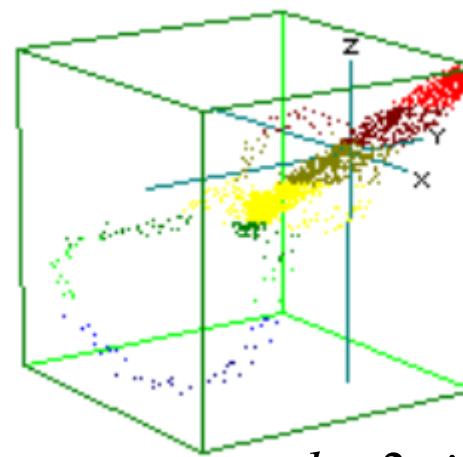
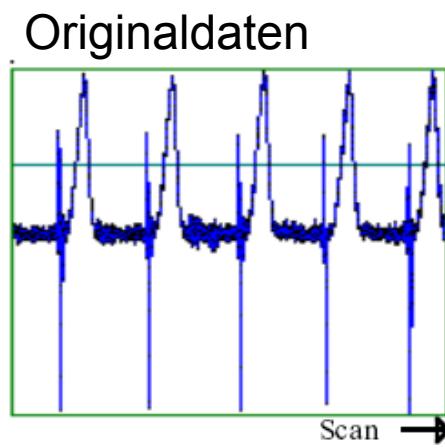
Temperatur-
funktion:

Wahl von j mit
schlechter Autokorr.

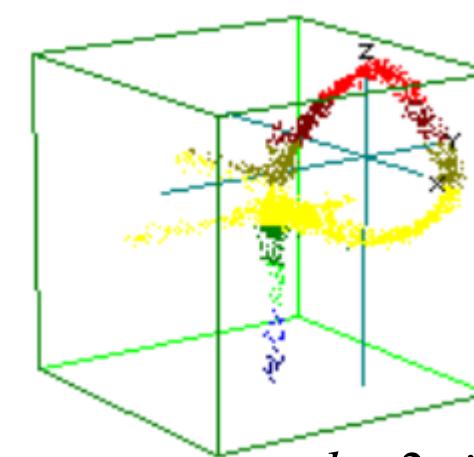


Phasenraumeinbettung

Beispiel: Verschiebungseinbettungen einer Komponente der EKG-Daten

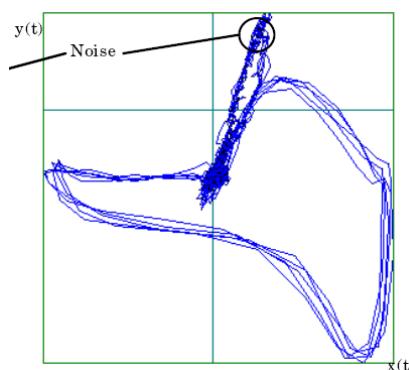


$$d = 3, j = 2$$



$$d = 3, j = 25$$

Die Farbe repräsentiert eine redundante Kodierung der z-Komponente durch deren Abbildung auf eine Farbtabelle.



Die rekonstruierte Trajektorie für $d = 3, j = 25$ ähnelt der Phasenraumtrajektorie der zwei Messfrequenzen des EKG.

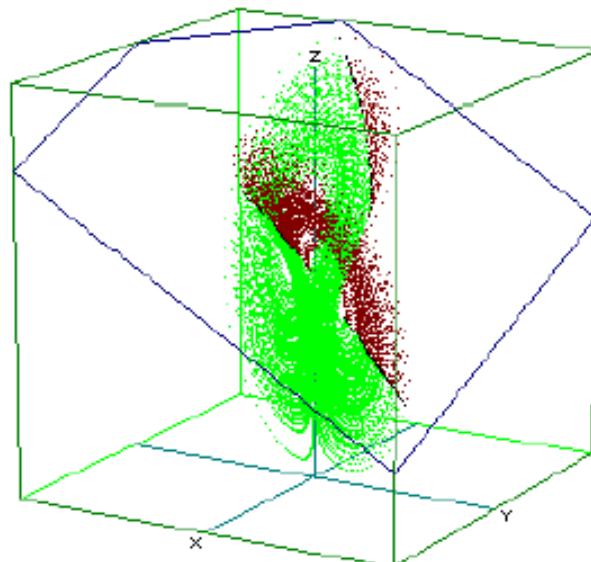
A.M. Kirchberger, I. Gröller, 7th Eurographics Workshop on Visualization in Scientific Computing, Springer-Verlag, 1996

Weitere Visualisierungsverfahren für Phasenraumdarstellungen

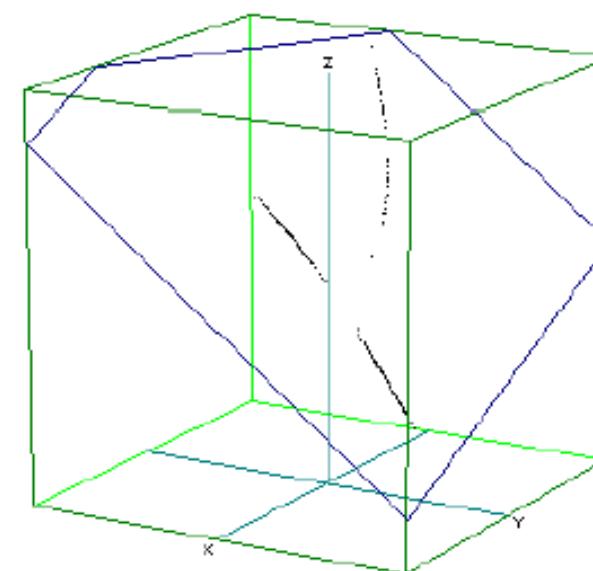
Poincaré-Abbildung

Folge der Schnittpunkte der Phasenraumtrajektorie mit einer Hyperebene im Phasenraum

Beispiel: Original-Phasenraumtrajektorie mit Schnitthyperebene



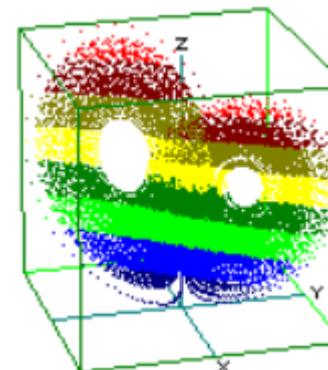
Poincaré-Abbildung



A.M. Kirchberger, I. Gröller, 7th Eurographics Workshop on Visualization in Scientific Computing, Springer-Verlag, 1996

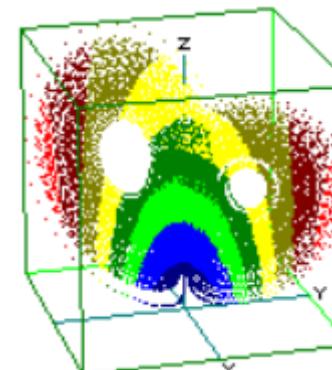
Visualisierung der Bewegungsdynamik im Phasenraum

Zuweisung der dritten Komponente der gegebenen Funktion an die z-Achse und auch an eine Farbtabelle. Dies wird durch gleichgefärzte Schichten entlang der z-Achse sichtbar.

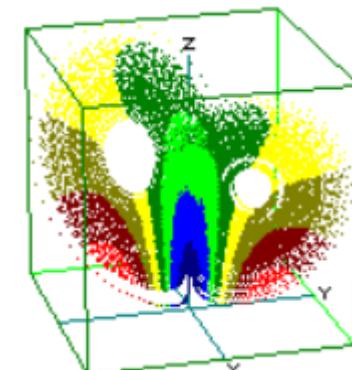


$$\Delta\tau = 0$$

Zuweisung der dritten Komponente an die z-Achse und an eine Farbtabelle, wobei die Zuweisung um $\Delta\tau$ zeitversetzt geschieht.
Bei stetiger Änderung von $\Delta\tau$ „fließt“ die Farbe durch den Phasenraum.



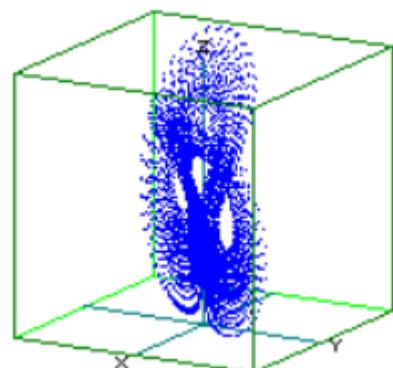
$$\Delta\tau = 10$$



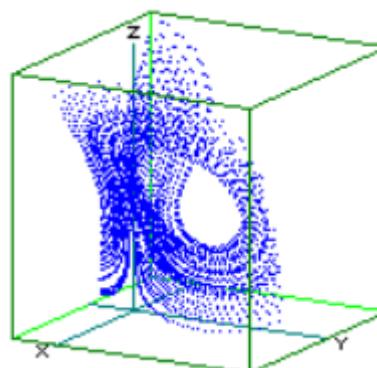
$$\Delta\tau = 20$$

Zoom-Operationen

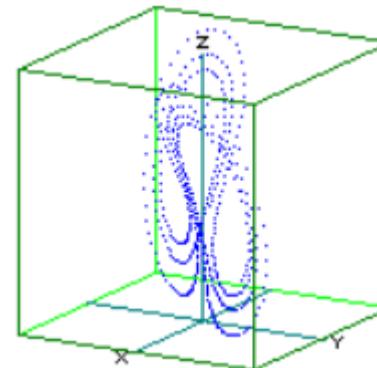
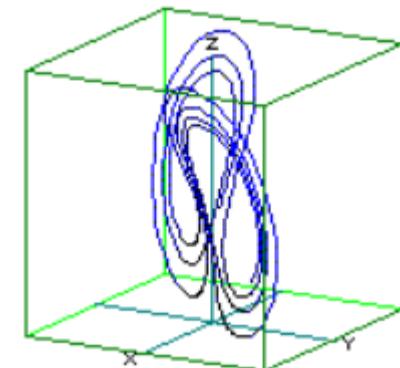
- **Räumliches Zoomen:** Einschränkung der Visualisierung auf einen quaderförmigen Ausschnitt des Phasenraums
- **Zeitliches Zoomen:** Einschränkung der Visualisierung auf ein Intervall des Definitionsbereichs der Zeitreihe



Originaldaten



räumliches Zoomen

zeitliches Zoomen
Polygonzugdarstellung
des zeitlichen Zooms

Rekurrenz-Plot

Darstellung der Abstände der Punkte des Phasenraums durch eine graphische Repräsentation der Distanzmatrix bei

- **Schwarzweißdarstellung:**
Schwarzer Punkt genau dann, wenn die Distanz kleiner als ein gegebener Schwellwert ist, siehe Beispiel
- **Grauwertdarstellung:**
Visualisierung der Größe der Distanz durch einen Punkt entsprechender Intensität
- **Farbdarstellung:**
Verwendung einer geeigneten Farbskala, z.B. aus dem ISO-Farbmodell

Anwendungssituation: hochdimensionale Phasenräume

Beispiel:

Rekurrenz-Plot der Phasenraumeinbettung der Niederschagsfunktion N1 aus dem Öko-Beispiel

$d = 7, j = 8$

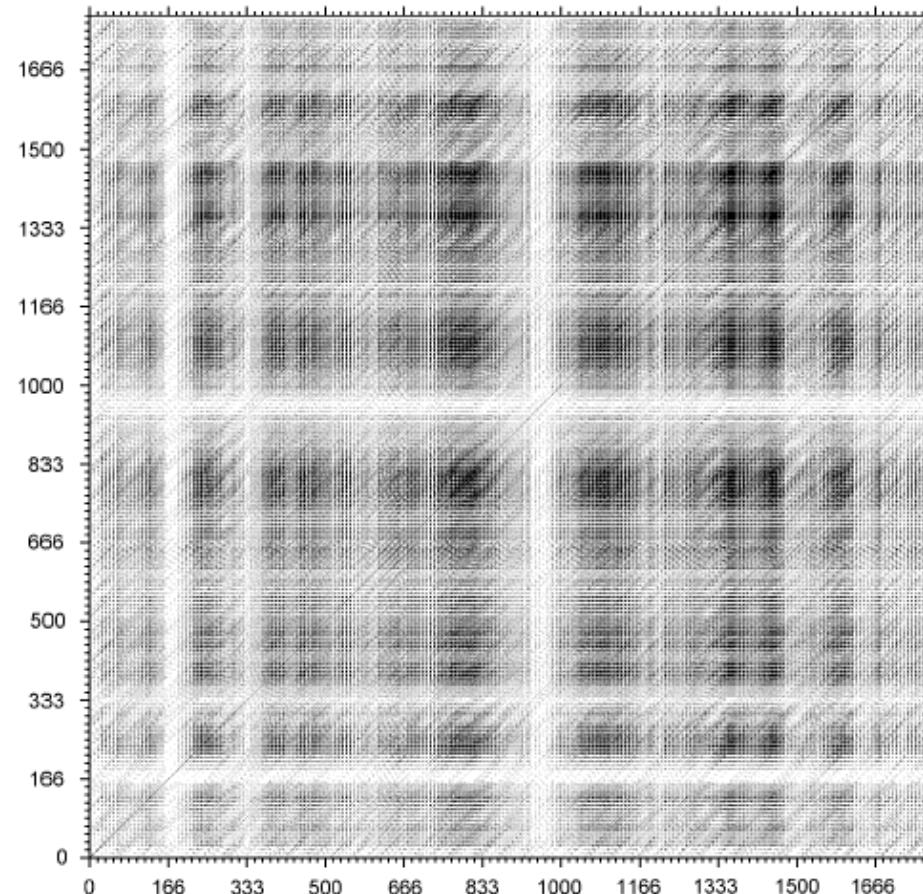
Euklidscher Abstand

Distanzschwellwert 8.84

Zeitraum zwischen 18.12.1995
und 17.12.1970 (1777 Punkte)

Bemerkung:

Geringe Korrelation ist an einem unruhigen Rekurrenz-Plot erkennbar.



Bemerkung

Die Datenreduktionsreduktions- und Visualisierungsverfahren für hochdimensionale Punktmengen können auf die Visualisierung hochdimensionaler Phasenräume übertragen werden.

Beispiel:

Elimination der Kohärenz in multidimensionalen Funktionen durch die Hauptachsentransformation

Filterung einer diskreten Funktion im Phasenraum

- **Tiefpassfilterung im Phasenraum:** Zerlegung des Phasenraums in kleine Hyperwürfel und Ersetzung der Trajektorienpunkte in jedem Hyperwürfel durch deren Mittelwert.
- **Punktdichtenfilterung im Phasenraum:**
z.B. Anwendung von alpha-shape-basierter Dichtefilterung im Phasenraum

Beispiel: Punktdichtenfilterung im Phasenraum

Ausdünnung der Punktfolge zu näherungsweise einem Kreis (Idealform angedeutet)

