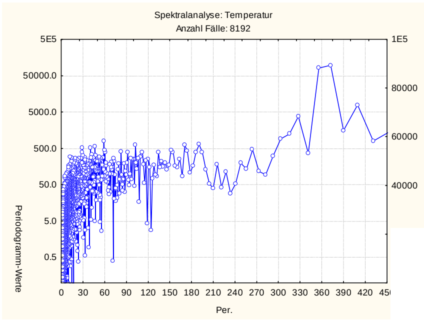
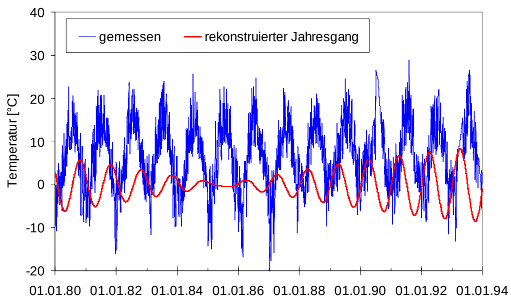
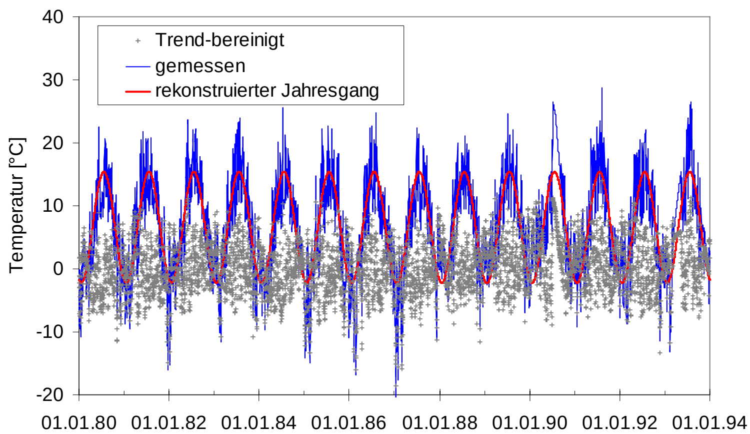
Desaisonalisierung: Temperatur





Alternative: Anpassen einer sin-Funktion

Funktionsgleichung: ***Tmp****=a+b*(sin((2*Pi****Datum****/365.25)-c))*



Wiederkehrdiagramme (*Recurrence Plots*)

Ziel: Visualisierung von Zeitreihen zur Identfizierung von lokalen Auffälligkeiten (Trends, Periodizitäten, etc.)

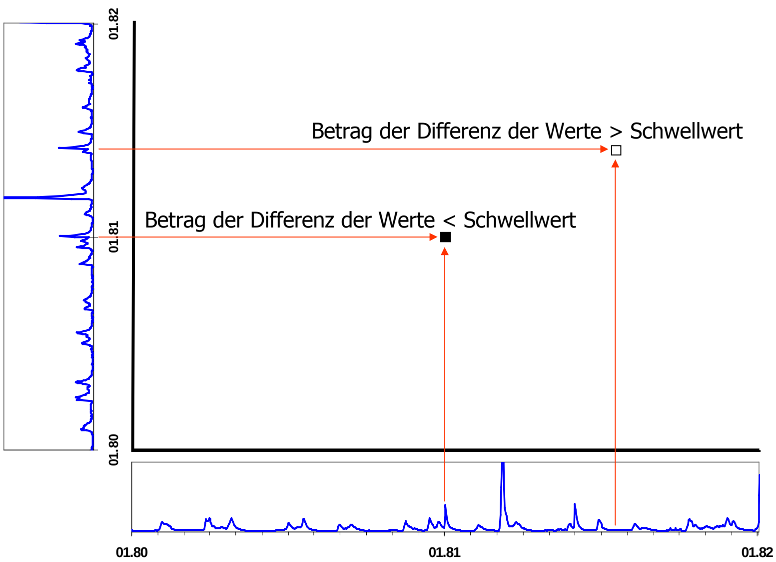
Motivation: geschickte Visualisierung gibt Hinweise auf lokale Auffälligkeiten im Datensatz

Prinzip:
- Quantifizierung der Ähnlichkeiten zwischen verschiedenen Fenstern des Datensatzes
- farbkodierte Visualisierung der Matrix der Ähnlichkeiten (Schwellwert)

Voraussetzungen: keine

Erweiterungen:
- Kreuzwiederkehrdiagramme zum Vergleich zweier Variabler
- RQA = ***R**ecurrence **Q**uantification **A**nalysis*

Prinzip



Recurrence

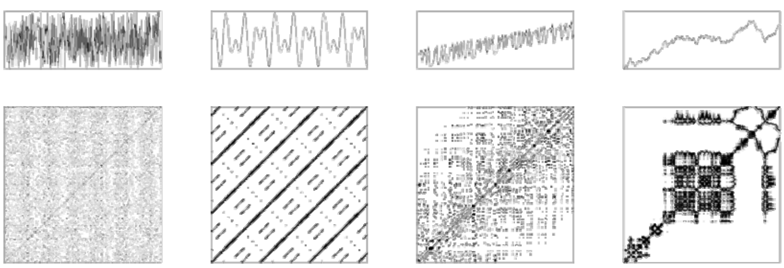
= Wiederkehr: $R_{i,j} = \Theta\left(\varepsilon - \left\| x_i - x_j \right\|\right) \quad i, j = 1, \ldots, n$

mit

- *Heaviside-Funktion* $\Theta(y) = \begin{cases} 0 & \text{für } y \leq 0 \\ 1 & \text{für } y > 0 \end{cases}$
- *Norm (Abstand)* $\|\ldots\|$ *(z.B. Euklidische Distanz)*
- *Schwellenwert* ε

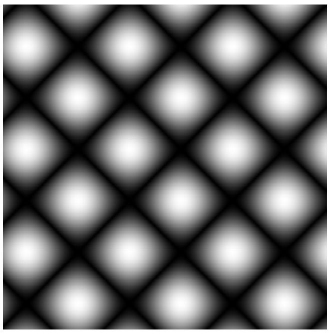
Beispiele für Wiederkehrdiagramme

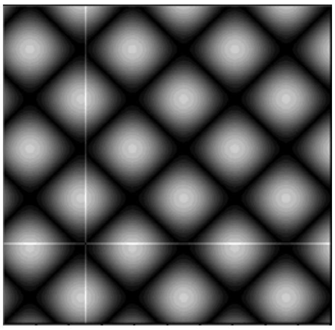
Weißes Rauschen	Harmonische Oszillation	Trend	Fluktuationen der Dynamik



Norbert Marwan (http://en.wikipedia.org/wiki/Recurrence_plot)

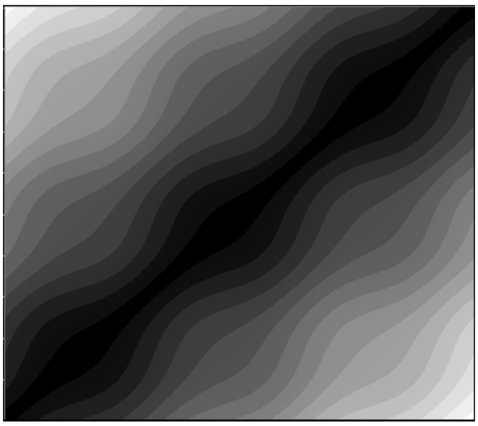
Wiederkehrdiagramme bekannter Funktionen





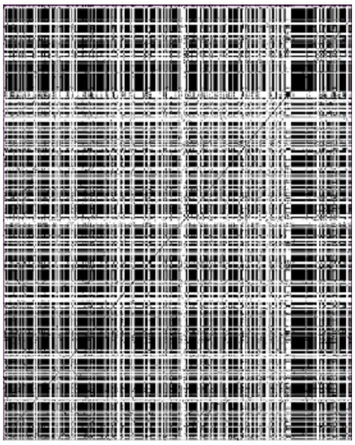
$y = \sin x$ mit Extremereignis/Ausreißer

RP eines Trends

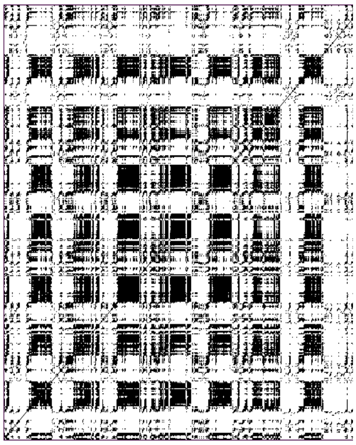


$$y = \sin x + 0.02x$$

Regen und Abfluss Lehstenbach (Fichtelgebirge)

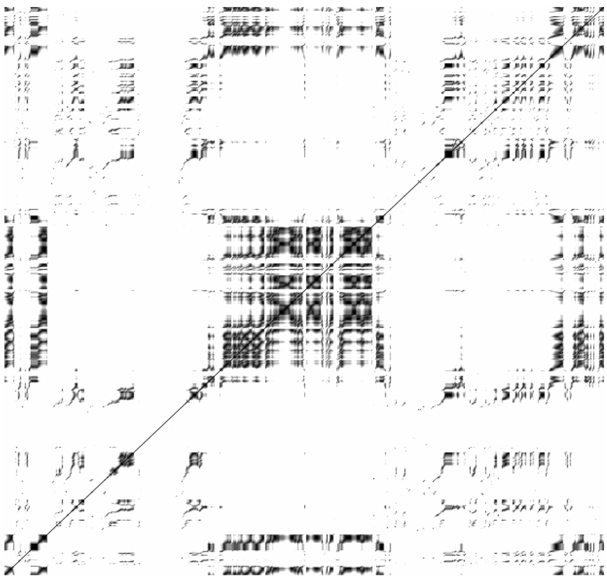


Regen

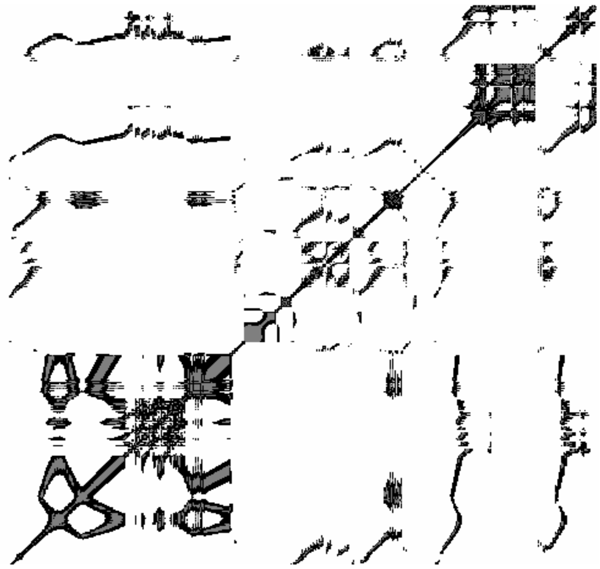


Abfluss

Lehstenbach Abfluss (Detail)



Steinkreuz Abfluss



Erweiterungen

- **Einbettungsdimension** (*Fensterlänge*) m > 1:

$$R_{i,j} = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{x}_j\right\|\right) \qquad i, j = 1, \ldots, n$$

z.B. für Einbettungsdimension m =2:

$$R_{i,j} = \Theta\left(\varepsilon - \left\|\begin{pmatrix} x_i \\ x_{i+1} \end{pmatrix} - \begin{pmatrix} x_j \\ x_{j+1} \end{pmatrix}\right\|\right) \qquad i, j = 1, \ldots, n$$

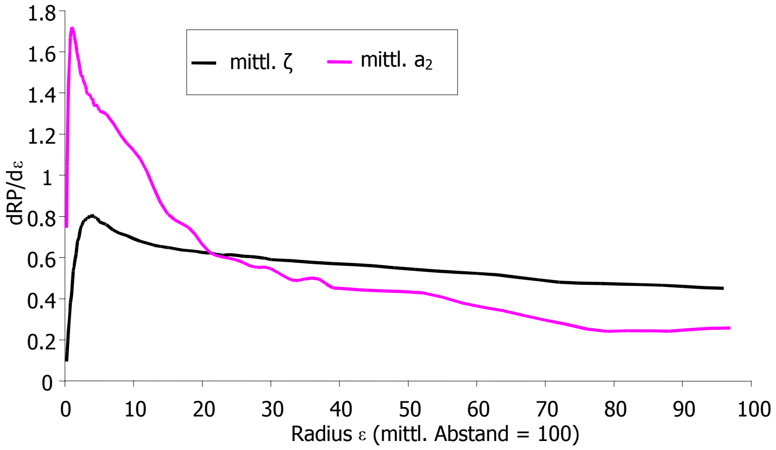
- **Delay** (Verzögerung) τ für $\vec{x} = (x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\cdot\tau})$

=> Wiederkehrmatrix ***R*** = $(n-(m-1)\tau)\times(n-(m-1)\tau)$ -Matrix

Ermittlung des optimalen Schwellenwertes ε

- Bestimmung der Zahl der Wiederkehrpunkte *RP* als Funktion des Schwellenwertes (Radius) ε
- Berechnung des Zuwachses $dRP/d\varepsilon$
- Maximum beim Überschreiten des „noise floors"
    - danach Plateau? Wähle Beginn des Plateaus
    - kein Plateau? dann halber Wert des Maximums
- Faustregel: RP ca. 30-50%

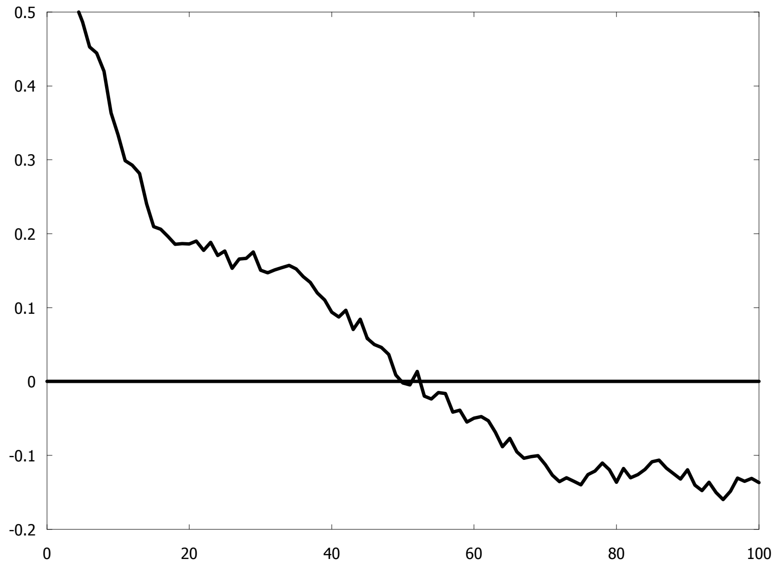
Ermittlung des optimalen Schwellenwert-Radius



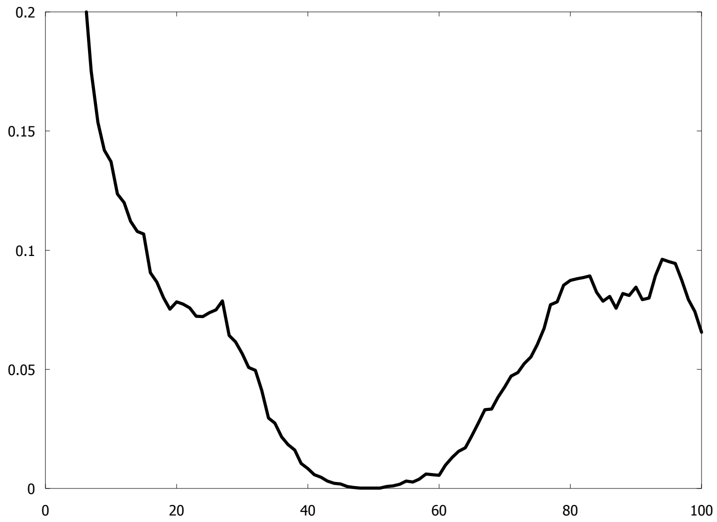
Ermittlung des Delays τ

- Der Attraktor sollte nicht zu dicht „abgetastet" werden
- Aufeinanderfolgende Einbettungsvektoren sollten nicht zu stark autokorreliert sein
- Ermittlung der ersten Nullstelle der Autokorrelation (linear) oder des ersten Minimums der wechselseitigen Information (nichtlinear)
- Wahl des Delays dort in der Nähe

Optimales Delay: Nullstelle der Autokorrelation



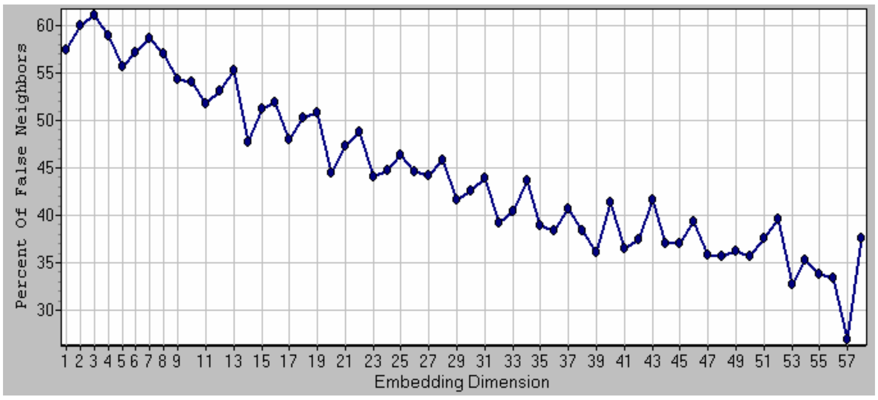
Optimales Delay: Minimum der wechselseitigen Information



Ermittlung der Einbettungsdimension *m*

- Bestimme zu jedem Vektor $\vec{x} = (x_1, x_2, ..., x_n)$
  seinen nächsten Nachbarn $\vec{y} = (y_1, y_2, ..., y_n)$
- Bestimme den Abstand der Werte
  zum nächsten Zeitpunkt: $R_{pred} = |x_{n+1} - y_{n+1}|$
- Bestimme den Abstand im Originaldatensatz:
  $R_{orig} = |x_{n+1} - x_n|$ („trivialer Prädiktor")
- Ist $R_{orig} \leq R_{pred}$, zählt $\vec{y}$ als „falscher" (zufälliger) Nachbar
- Wähle die Einbettungsdimension mit der geringsten Zahl von falschen Nachbarn

Einbettungsdimension nach modifizierten „false nearest neighbors“



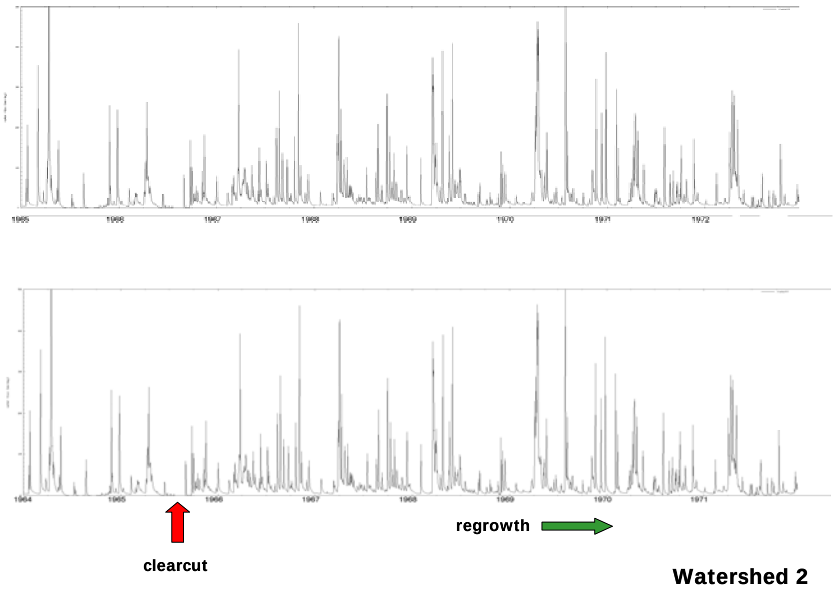
Abfluss Lehstenbach/Fichtelgebirge

Hubbard Brook Experimental Forest

- 43°56′ N, 71°45′ W (New Hampshire)
- established 1955 by the U.S.D.A. Forest Service
- size 31.6 km^2, altitude range 222 m - 1015 m a.s.l.
- **Watershed 6:** undisturbed control
- **Watershed 2:** clearcut 1965-66, regrowth prohibited until 1968

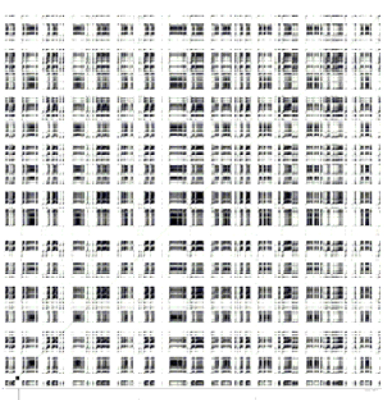
Abflüsse

Watershed 7

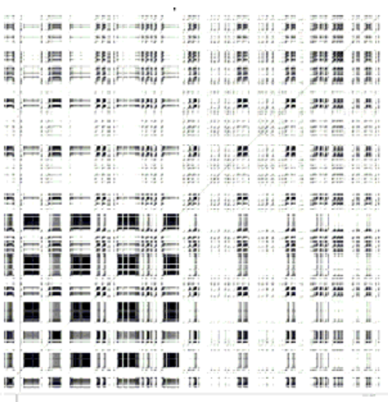


Watershed 2

Abflüsse

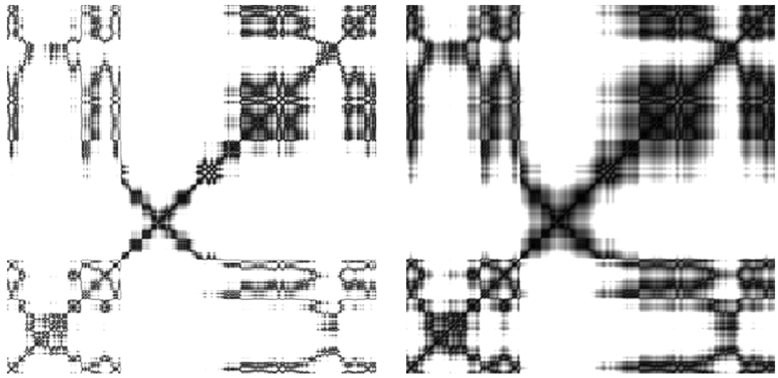


Hubbard Brook Nr. 6



Hubbard Brook Nr. 2

Unterschiedliche Zeitauflösung

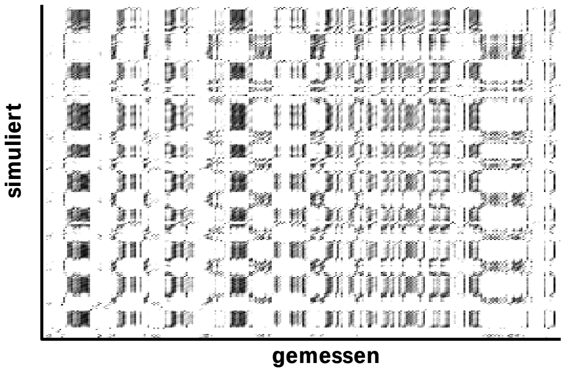


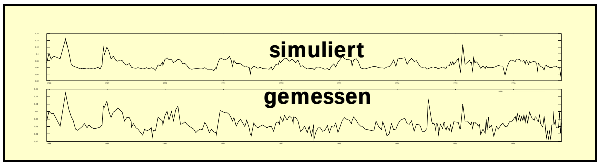
Original 10-fach vergröbert

Kreuzwiederkehrdiagramme

- Vergleich zweier unterschiedlicher Datenreihen
- Vorher Daten auf einen einheitlichen Wertebereich skalieren (z.B. [0,1])
- Interpretation bzw. Quantifizierung analog zu RP bzw. RQA

Nitrat: Vergleich Modell-Messung





Quantitative Analyse der Wiederkehrdiagramme (*RQA = Recurrence Quantification Analysis*)

Ziel: Anhand der Recurrence Points (***RP***) Maßzahlen für Muster bestimmen

Wiederkehrrate = Anteil der RP

$$RR = \frac{1}{n^2} \cdot \sum_{i,j=1}^{n} R_{i,j}$$

Determinismus = Anteil der RP, die auf Linien parallel zur 1:1-Diagonale entfallen *(P(l): Häufigkeit)*

$$DET = \sum_{l=l_{\min}}^{n} l \cdot P(l) \Big/ \sum_{i,j=1}^{n} R_{i,j}$$

Linienlängen-Entropie = Shannon-Entropie der Verteilung der Längen der Linien parallel zur 1:1-Diagonalen (*p(l): Wahrscheinlicheit*)

$$ENTR = -\sum_{l=l_{\min}}^{n} p(l) \cdot \ln p(l)$$

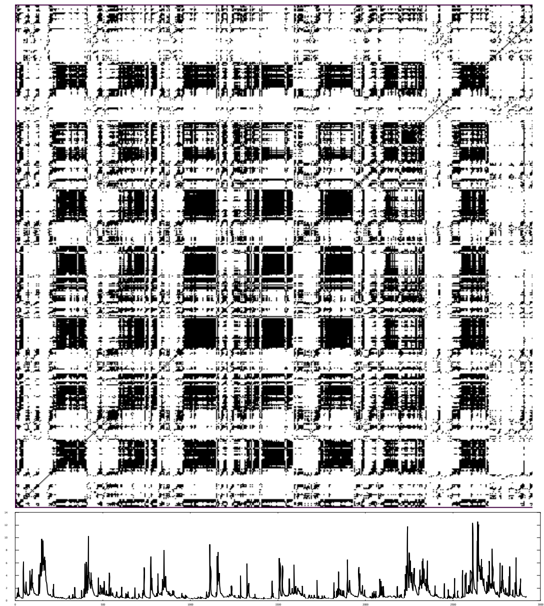
Maximale Linienlänge parallel zur 1:1-Diagonale

$$L_{\max} = \max(\{l_i; i = 1, \dots, n_l\})$$

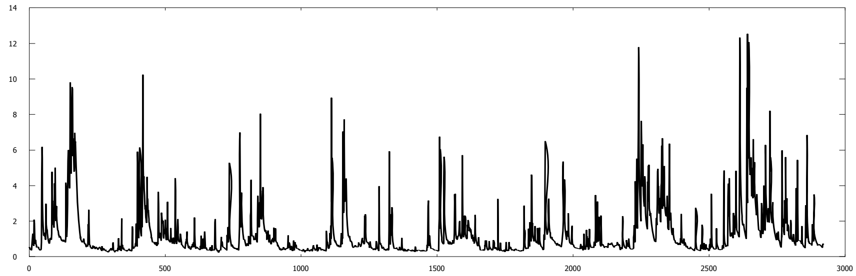
Trend = Ausdünnen der RP zu den Ecken hin

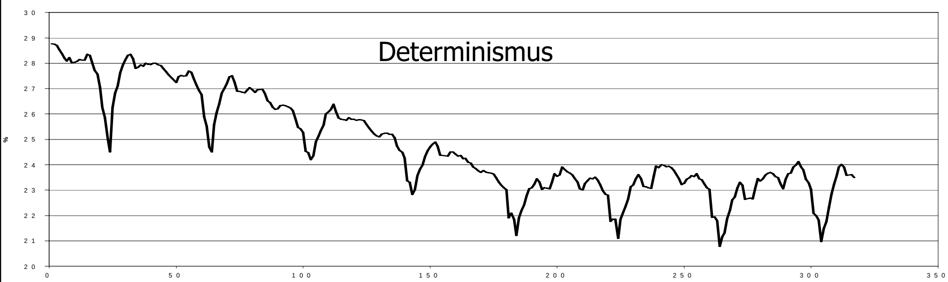
$$TREND = \frac{\sum_{i=1}^{\tilde{n}} (i - \tilde{n}/2) \cdot (RR_i - \overline{RR_i})}{\sum_{i=1}^{\tilde{n}} (i - \tilde{n}/2)^2}$$

RQA Lehstenbach-Abfluss 1987-1995

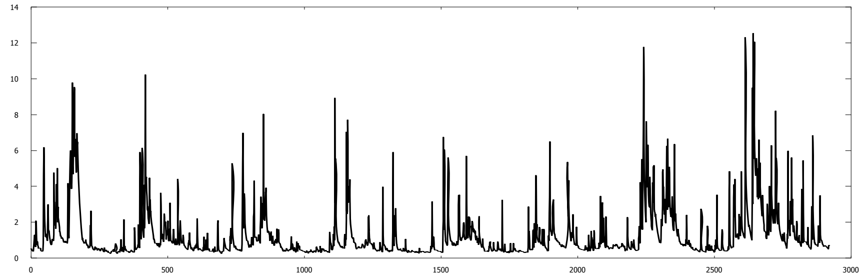


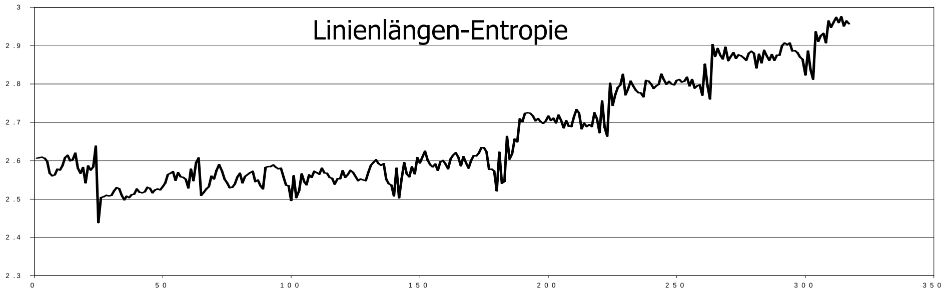
RQA am Beispiel Lehstenbach



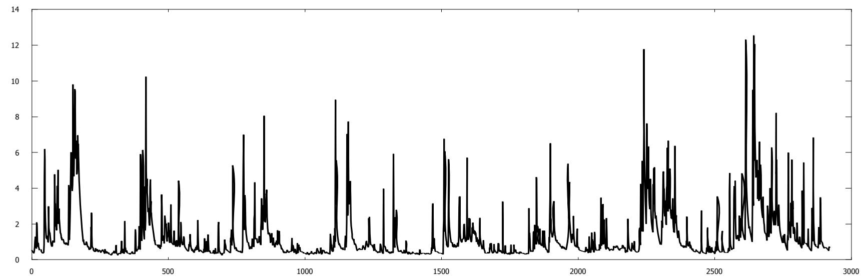


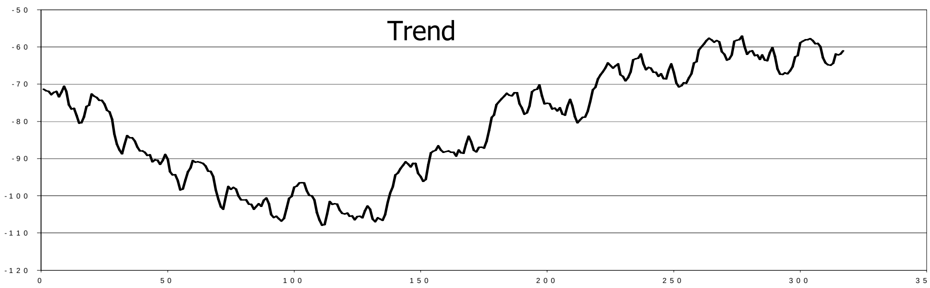
RQA am Beispiel Lehstenbach



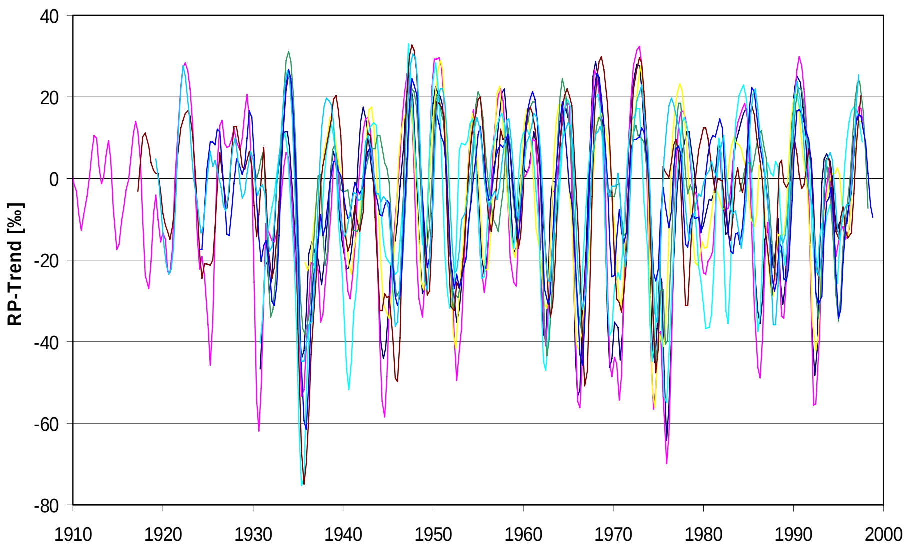


RQA am Beispiel Lehstenbach





RP-Trend für süddeutsche Abflüsse



Aufgabe

1. Erstellen Sie jeweils Wiederkehrdiagramme für die 30-Tages-Werte von Niederschlag, Temperatur und Abfluss (Einbettungsdimension m = 1, Verzögerung τ = 1).
2. Variieren Sie die Schwellenwerte ε für die farbige Kodierung der Differenzen.
3. Identifizieren Sie anhand der Wiederkehrdiagramme "ruhige" und "unruhige" Perioden und vergleichen Sie diese Darstellung mit den Grafiken der Zeitreihen.