

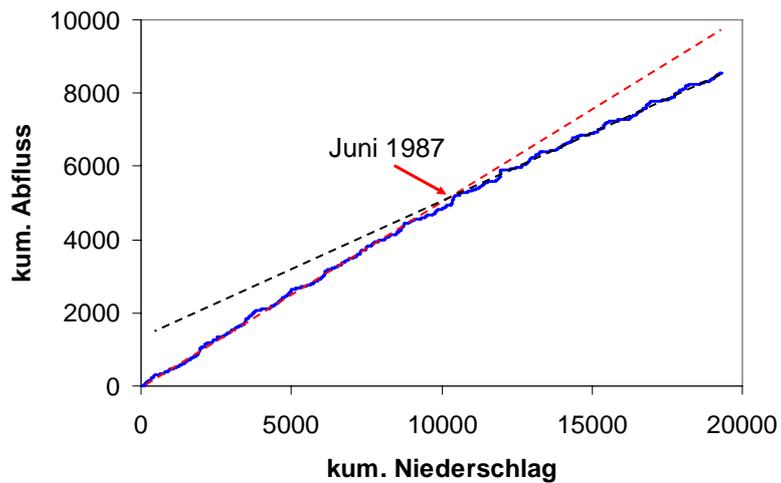
# Korrektur: Lineare Regression in Excel

Microsoft Excel - PDFA\_Datensatz.xls

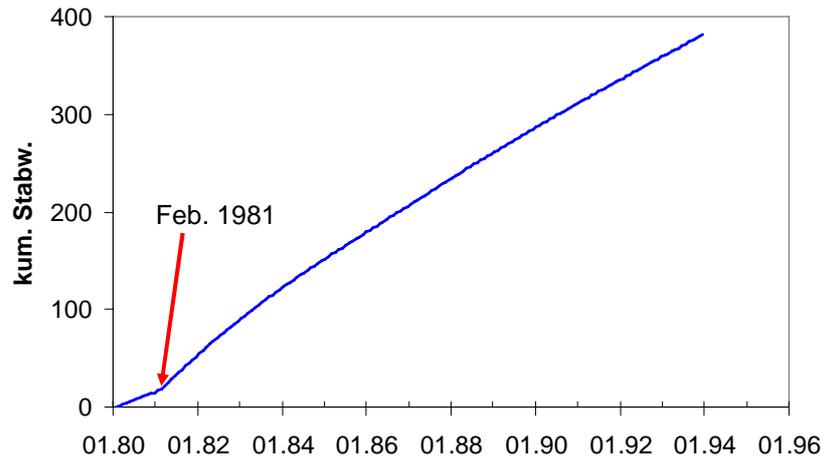
Formelzeile: E31 =RGP(D2:D31;A2:A31;WAHR;WAHR)

	A	B	C	D	E	F	G	H
1	Datum	Niederschlag [mm/d]	Temperatur [°C]	Abfluss				
21	20.01.80	0	-5.80	0.432	-0.0157634	461.439528	-0.0768085	
22	21.01.80	0	-1.20	0.41472	-0.0157634	461.439528	-0.0783251	
23	22.01.80	0	0.40	0.40608	-0.0157634	461.439528	-0.0712017	
24	23.01.80	0	-0.10	0.40608	-0.0157634	461.439528	-0.0554382	
25	24.01.80	0	0.00	0.39744	-0.0157634	461.439528	-0.0483148	
26	25.01.80	0	-0.90	0.40608	-0.0157634	461.439528	-0.0239114	
27	26.01.80	0	-6.00	0.40608	-0.0157634	461.439528	-0.0081479	
28	27.01.80	0.2	-7.30	0.39744	-0.0157634	461.439528	-0.0010245	
29	28.01.80	0.9	-7.40	0.3888	-0.0157634	461.439528	0.00609894	
30	29.01.80	9.9	-3.30	0.44064	-0.0157634	461.439528	0.07370237	
31	30.01.80	0	1.40	0.89856	-0.0157634	461.439528	0.54736581	0.1301001
32	31.01.80	0	3.20	0.9504	-0.0670729	1965.50475	-2.5706694	
33	01.02.80	0.2	-4.40	0.69984	-0.0670729	1965.50475	-2.7541565	
34	02.02.80	7.7	-0.60	0.864	-0.0670729	1965.50475	-2.5229236	

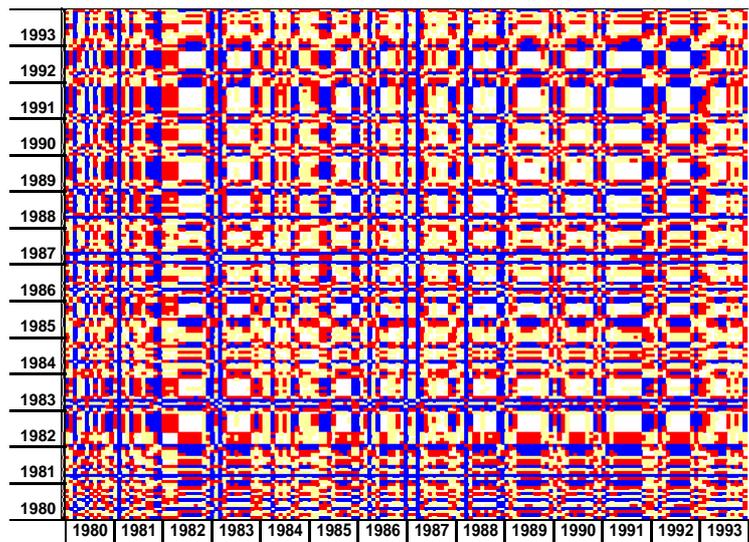
## Doppelsummenkurve



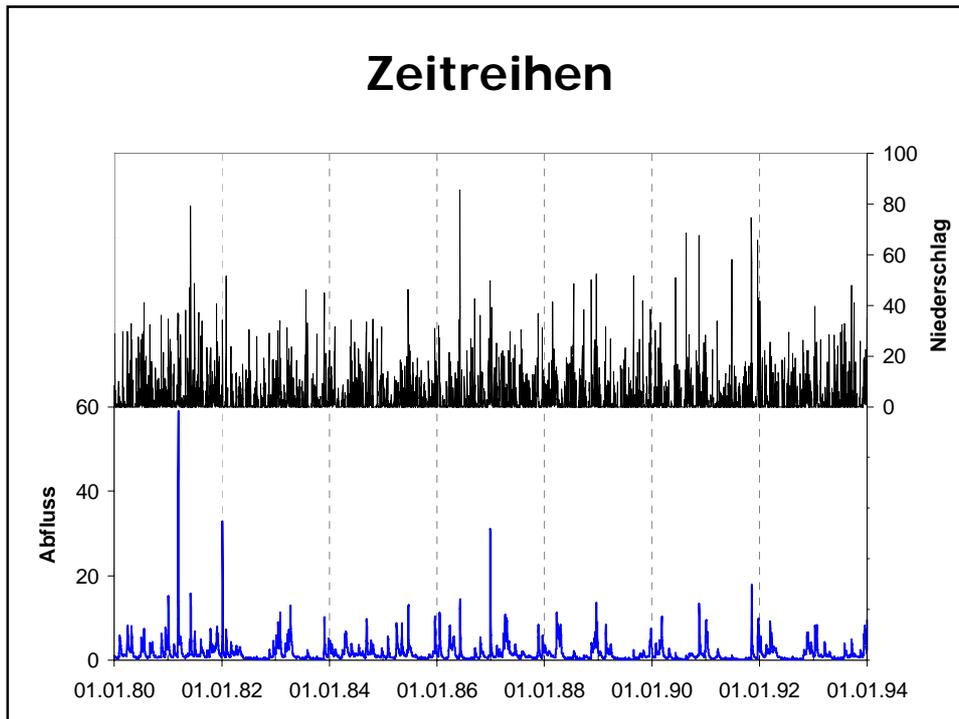
## PDFA Abfluss Lange Bramke



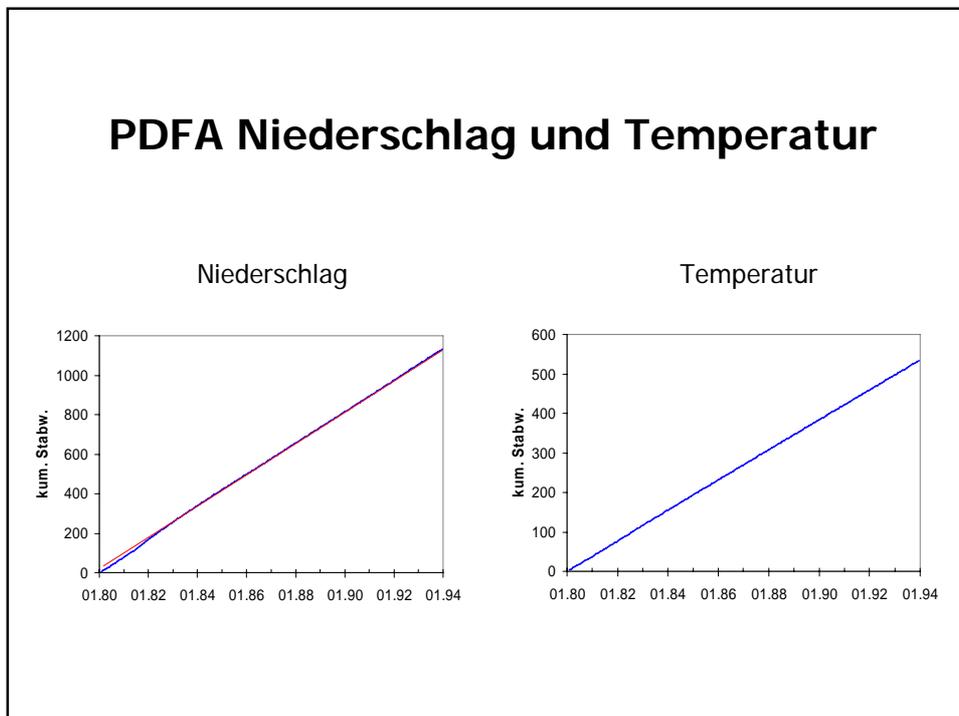
## Zum Vergleich: Wiederkehrdiagramm



## Zeitreihen



## PDFA Niederschlag und Temperatur



# Extremwerte

**Motivation:** Oft sind nicht Mittelwerte, sondern die Extrema von Interesse

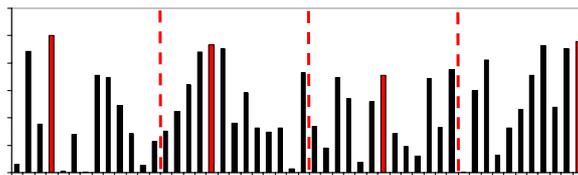
**Fragestellung:** Lässt sich aus der Verteilung der beobachteten (Jahres-)Maxima die Wahrscheinlichkeit extrem seltener Maxima (z.B.  $HQ_{100}$ ) abschätzen?

**Problem:** Extrema sind sehr schwer zu schätzen

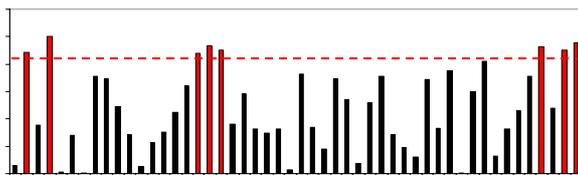
- Spannweite der Werte steigt mit zunehmendem Stichprobenumfang (keine Sättigung der Statistik)
- Kleine Änderungen der Parametrisierungen angepasster Häufigkeitsverteilungen wirken sich sehr stark auf die Wahrscheinlichkeit der Extremwerte aus
- Extrema sind selten

## Zwei verschiedene Ansätze

1. Block-Extrema (Fenster gleicher Länge; Bsp.: Jahresmaxima): *GEV*



2. Schwellwertüberschreitungen (*POT = Peak over Threshold*): *GPD*



## Annahmen

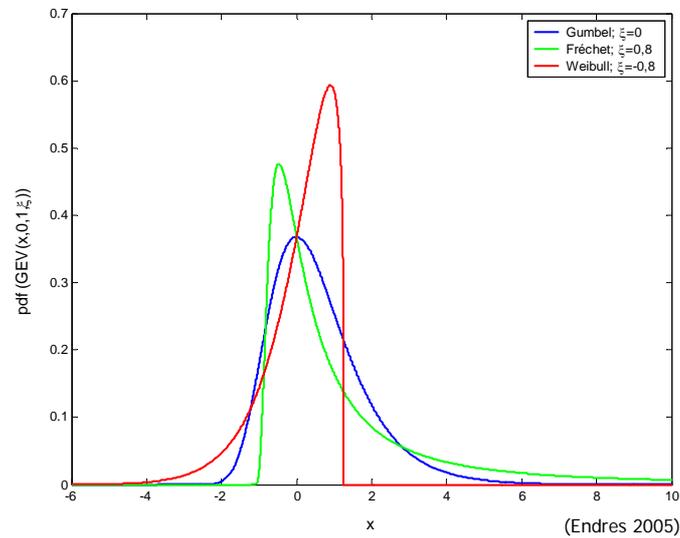
- Verteilung der  $x(t)$  ist nicht bekannt
- unabhängige und gleichverteilte **Maxima** (*i.i.d.* = *independent and identically distributed*)
- Extreme Ereignisse sind unkorreliert: **Asymptotic Independence of Maxima Theorem** (AIM-Theorem)
- Approximation für  $n \rightarrow \infty$

## Allgemeine Extremwertverteilung (*GEV = General Extreme Value Distribution*)

$$GEV_{\xi, \bar{x}, \sigma}(x) = \begin{cases} e^{-\left(1 + \xi \left(\frac{x - \bar{x}}{\sigma}\right)^{-1/\xi}\right)} & \text{für } \xi \neq 0 \\ e^{-e^{-\left(\frac{x - \bar{x}}{\sigma}\right)}} & \text{für } \xi = 0 \end{cases}$$

- $\bar{x}$ : Mittelwert (Lageparameter)
- $\sigma$ : Standardabweichung (Skalenparameter)
- $\xi$ : Formparameter
  - für  $\xi = 0$ : *Exponential Tail* (**Gumbel**-Verteilung)
  - für  $\xi > 0$ : *Heavy Tail* (**Fréchet**-Verteilung)
  - für  $\xi < 0$ : *Finite Tail* (neg. **Weibull**-Verteilung)

## GEV: 3 Typen



## Verallgemeinerte Pareto-Verteilung (GPD = Generalized Pareto Distribution)

- für  $x' =$  Schwellwert:

$$GPD_{\xi, \sigma}(x - x') = 1 - \left(1 + \xi \cdot \frac{x - x'}{\sigma}\right)^{-1/\xi}$$

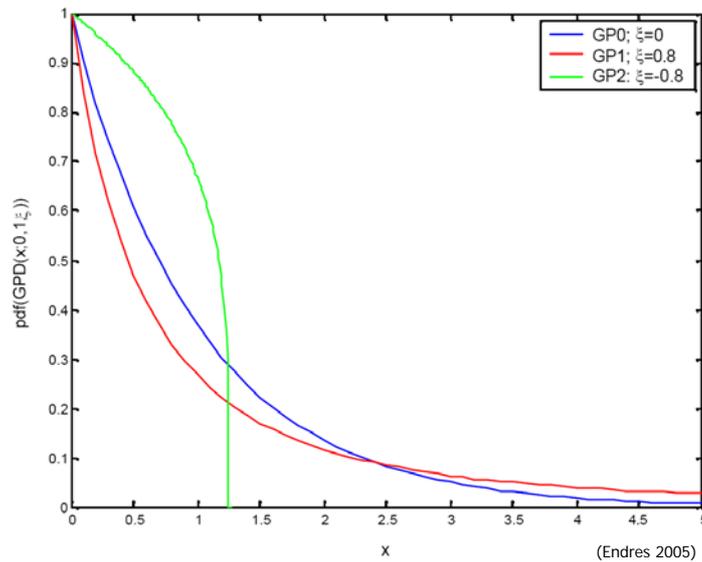
- 3 Spezialfälle:

für  $\xi = 0$ : *Exponential Tail* (**Exponential**-Verteilung, **GPO**)

für  $\xi > 0$ : *Heavy Tail* (**Pareto**-Verteilung, **GP1**)

für  $\xi < 0$ : *Finite Tail* (**Beta**-Verteilung, **GP2**)

## GPD: 3 Typen



## GEV vs. GPD

- **Zusammenhang:**

$$GPD(x) = 1 + \log GEV(x) \quad \text{für} \quad \log GEV(x) > -1$$

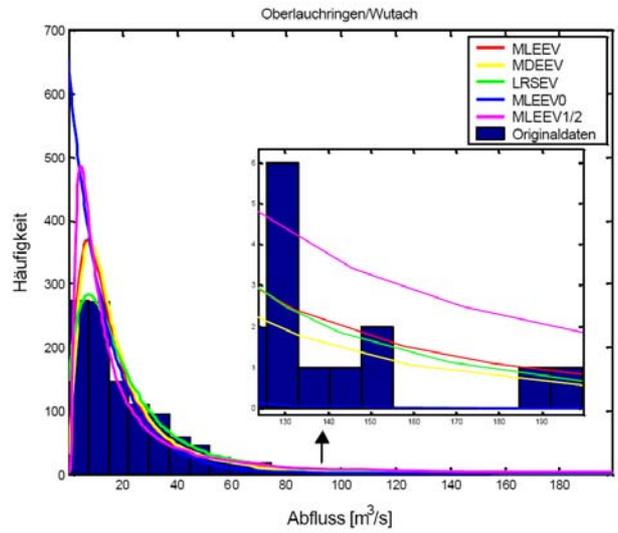
- **Ausnutzung der Daten:**

- bestimmt durch Fensterlänge (*GEV*) oder Schwellwert (*GPD*);
- jeweils so zu wählen, dass die Extrema voneinander unabhängig sind

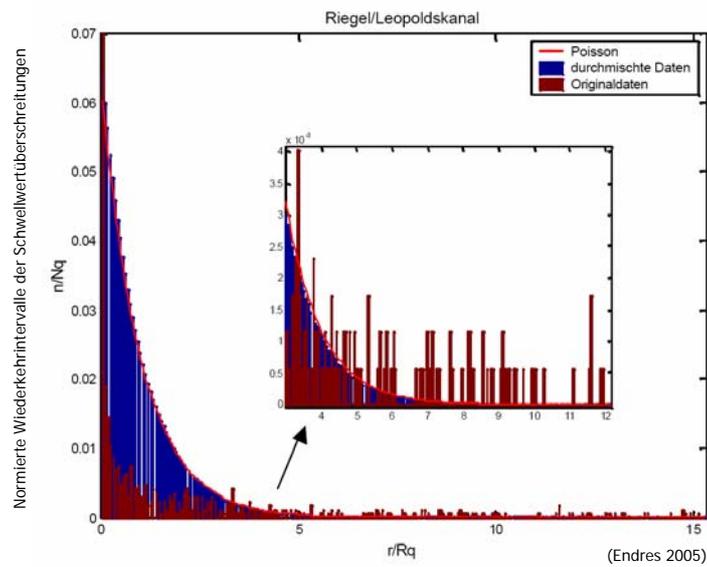
- **Präferenzen:**

- Kontinuierliche Messungen: *GEV*, z.B. *Abflussganglinien*
- Protokollierung der Extremereignisse: *GPD*, z.B. *Schadensmeldungen der Versicherungen*

## Anpassung der theoretischen Verteilung



## Problem: Langreichweitige Korrelationen



## Übersicht

1. Definition und Eigenschaften von Zeitreihen
2. Tests und Trenderkennung für Zeitreihen
3. Fourriertransformation, Powerspektrum
4. Wavelets
5. Zeitreihenmodellierung der ARMA-Klasse (kurzes Gedächtnis)
6. Modellierung von Zeitreihen mit langem Gedächtnis
7. Komplexität und Information von Zeitreihen
8. Wiederkehranalyse
9. Singuläre System-Analyse
10. Bruchpunktanalyse
11. Extremwertanalyse

## Zeitreihen

- Eine Zeitreihe ist eine Menge von Werten, die in einer festgelegten (und bekannten!) Reihenfolge vorliegen.
- Zeitreihen rechtfertigen eigene statistische Verfahren (und eine eigene Vorlesung), weil sie in der Regel autokorreliert sind.

*Gegenbeispiel:* Weißes Rauschen

## Korrelation etc.

	Maß für	Mathematisch
<b>Varianz</b>	Streuung um den Mittelwert	mittlere quadrierte Abweichung
<b>Kovarianz</b>	<i>Enge</i> des linearen Zusammenhangs zwischen zwei Variablen	mittlere quadrierte Abweichung zweier Variablen jeweils von ihrem Mittelwert
<b>Korrelation</b>	<i>Enge</i> des linearen Zusammenhangs zwischen zwei Variablen	auf [-1;1]-normierte Kovarianz (Teilung durch das Produkt der Standardabweichungen)
<b>Regression</b>	<i>Art</i> des Zusammenhangs zwischen mehreren Größen	Funktion
<b>Faltung</b>	<i>Enge</i> des Zusammenhangs zweier Zeitreihen bzw. Funktionen	gemittelttes Produkt zweier gegeneinander verschobener Zeitreihen bzw. Funktionen

## Übersicht

Verfahren	Ziel	global/lokal	Instationaritäten	Periodizitäten
(saisonaler) Kendall-Test	Analyse	global	des Mittelwerts	-
Bartlett-, Levene-Test	Analyse	global	der Varianz	-
Stationaritätstest nach Witt	Analyse	lokal	der Verteilung	-
Fourierspektrum, Periodogramm	Analyse	global	-	Frequenz-scharf (harmonische)
Power-Spektrum	Analyse	global	-	Frequenz-scharf (allgemein)
Wavelet-Analyse	Analyse	lokal	des Frequenzspektrums	Frequenz-scharf (harmonische)
ARIMA-Modelle	Modellierung	global	-	-
Hurst-Koeffizient	Analyse	global	(langreichweitige Korrelationen)	-
Komplexität und Information	Analyse	global	-	-
Wiederkehranalyse	Analyse	lokal	der Dynamik (Attraktoren)	annähernd
Singuläre System-Analyse	Analyse	lokal	der Dynamik (Attraktoren)	annähernd
Doppelsummenkurve	Analyse	lokal	des Verhältnisses zweier Variablen	-
Bruchpunktanalyse (i.e.S.)	Analyse	Lokal	des Mittelwertes	-
Progressive Detrended Fluctuation Analysis	Analyse	lokal	der Varianz, der Korrelationen	-
Extremwertanalyse	Analyse	global	-	-