Universität Augsburg
Mathematisch-Naturwissenschaftlich-Technische Fakultät

Klimarekorde und die Statistik von Extremen

Stefan Großkinsky
Public Climate School
24.11.2023

Universität Augsburg
Mathematisch-Naturwissenschaftlich-Technische Fakultät

- Temperaturdaten – Trends und Schwankungen
- Rekorde für iid ZV
- Rekorde für ZV mit Drift
- Asymptotische Verteilung von Extremwerten - Niederschläge
- Zusammenfassung und Diskussion

Klimarekorde und die Statistik von Extremen

Stefan Großkinsky
Public Climate School
24.11.2023

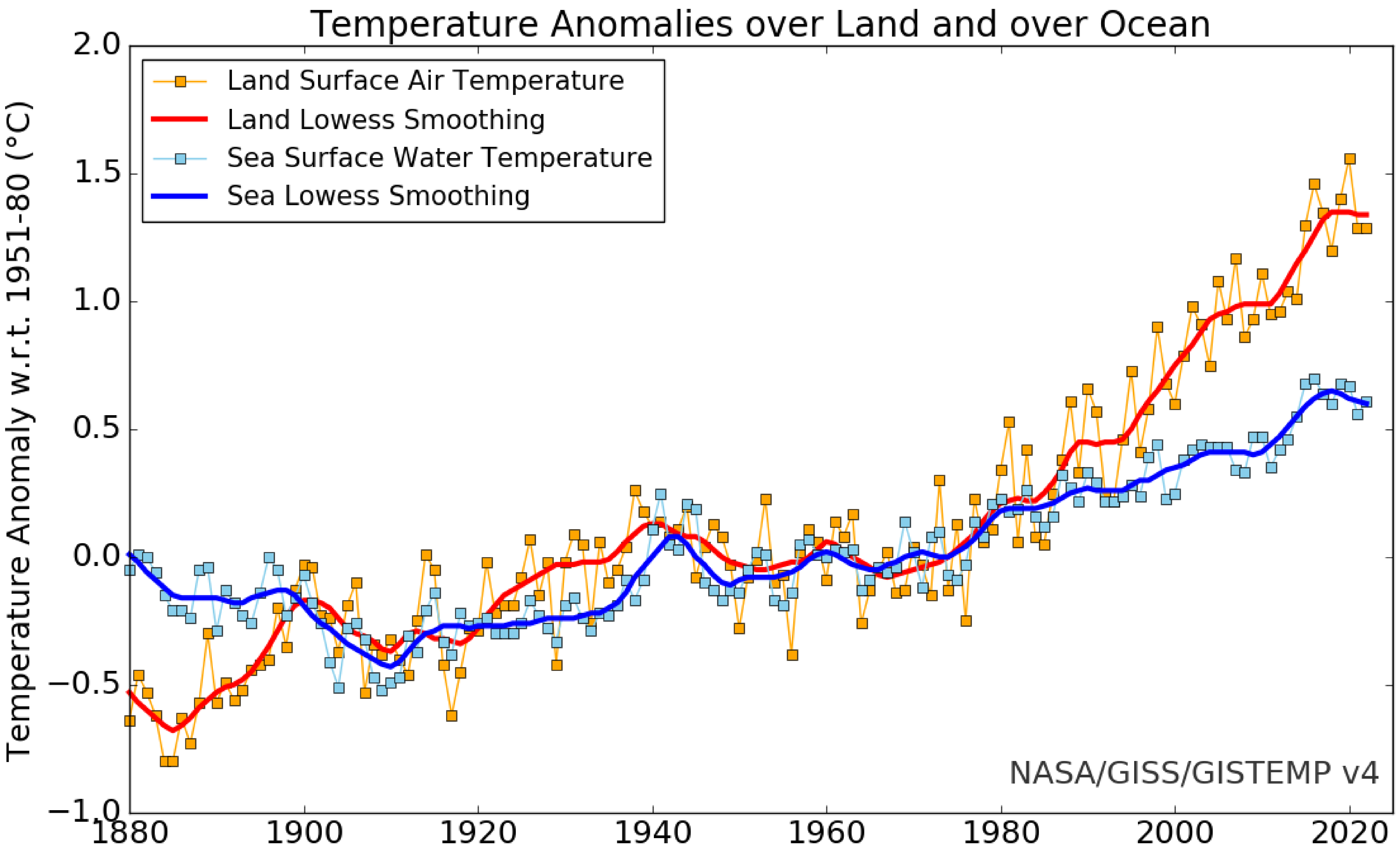
Quellen

- Wergen, Krug: Record-breaking temperatures reveal a warming climate EPL, 92 (2010) 30008
- Franke, Wergen, Krug: Records and sequences of records from random variables with a linear trend, JSTAT (2010) P10013
- G. Wergen, J. Krug, S. Rahmstorf: Klimarekorde, Spektrum der Wissenschaft (2014)

Daten:

- Project team ECAD, European climate assessment and dataset, Technical Report, Royal Netherlands Meteorological Institute, KNMI, 2008. https://www.ecad.eu/
  EI: 43 Stationen (1906-2005), EII: 187 Stationen (1976-2005)
- https://wiki.bildungsserver.de/klimawandel/index.php
  https://data.giss.nasa.gov/gistemp/graphs_v4/

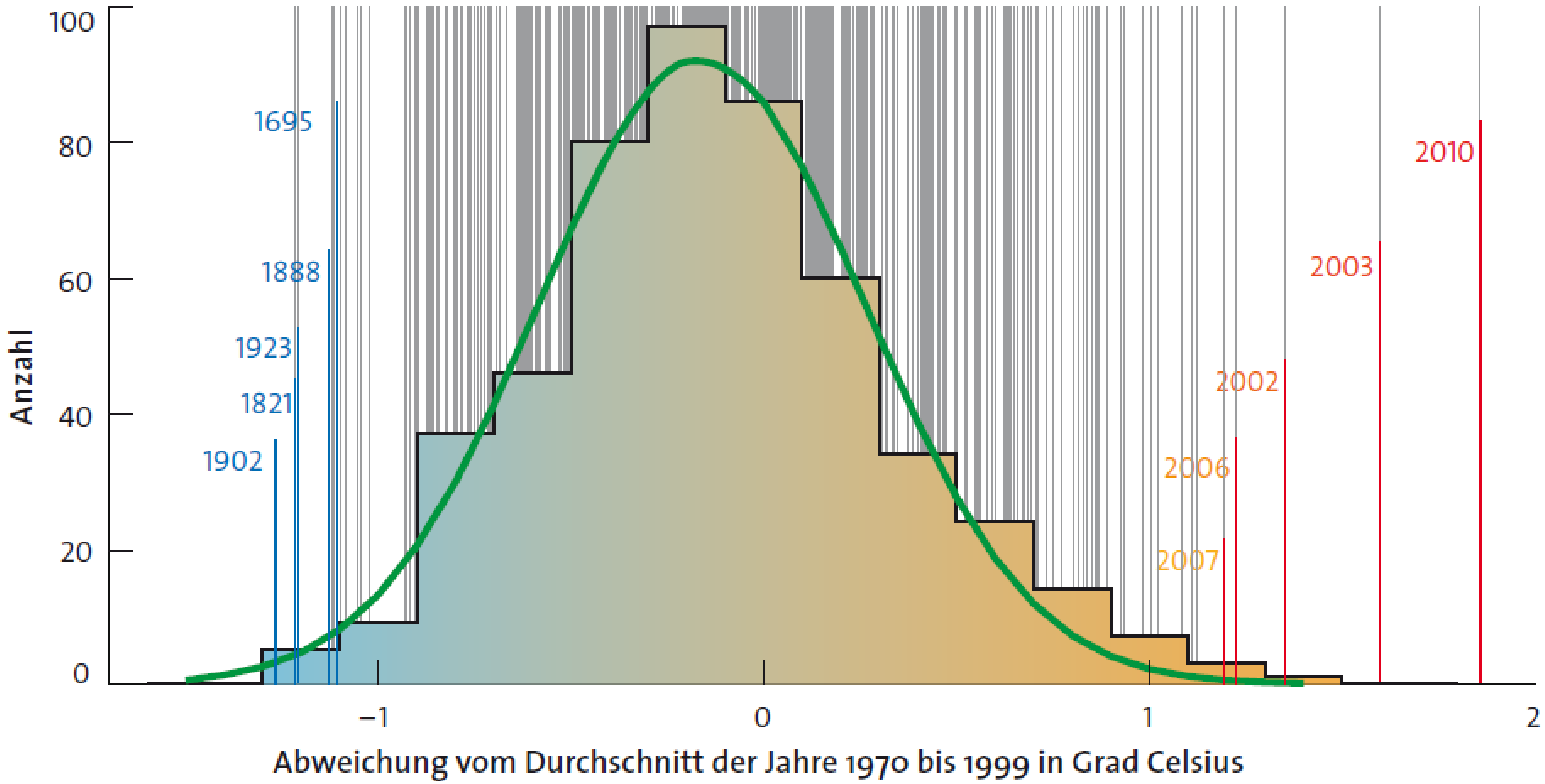
Temperaturmessungen – Trends und Schwankungen



LOWESS
locally weighted scatterplot smoothing

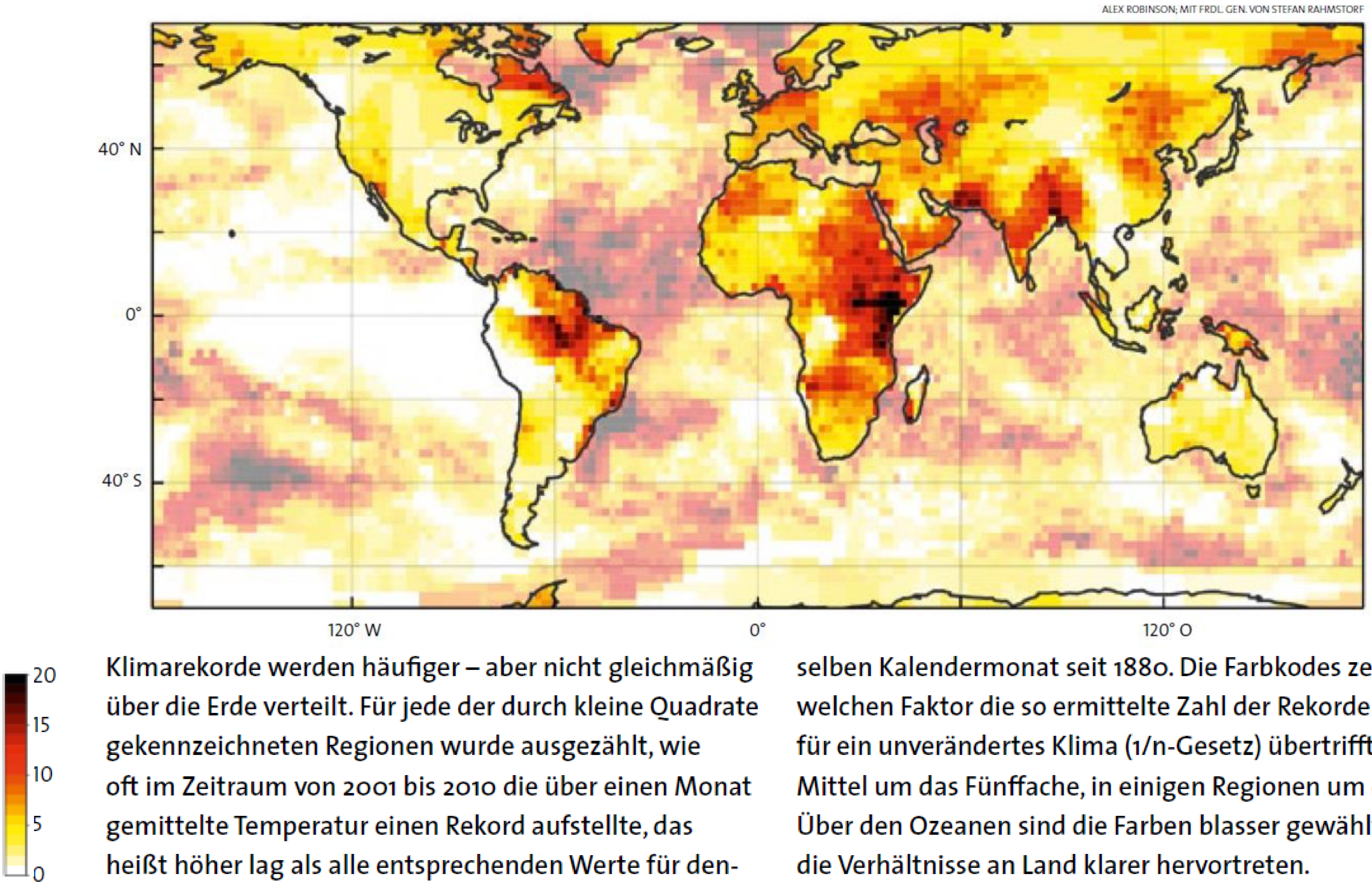
LOESS
locally estimated scatterplot smoothing

Temperaturmessungen – Trends und Schwankungen



Sommertemperaturen in Europa von 1500 bis 2010. Das Histogramm zeigt an, wie oft die Temperatur in jedes Intervall der Breite 0,2 Grad fiel; grün eingezeichnet ist die Gaußkurve, die am besten zu den Daten passt. Die Jahreszahlen der fünf kältesten und der fünf wärmsten Sommer sind angegeben (Balkenlänge hat keine Bedeutung).

Temperaturmessungen – Trends und Schwankungen



Klimarekorde werden häufiger – aber nicht gleichmäßig über die Erde verteilt. Für jede der durch kleine Quadrate gekennzeichneten Regionen wurde ausgezählt, wie oft im Zeitraum von 2001 bis 2010 die über einen Monat gemittelte Temperatur einen Rekord aufstellte, das heißt höher lag als alle entsprechenden Werte für denselben Kalendermonat seit 1880. Die Farbkodes zeigen, um welchen Faktor die so ermittelte Zahl der Rekorde die Erwartung für ein unverändertes Klima (1/n-Gesetz) übertrifft: im globalen Mittel um das Fünffache, in einigen Regionen um das 20-Fache. Über den Ozeanen sind die Farben blasser gewählt, damit die Verhältnisse an Land klarer hervortreten.

Grenzwertsätze für Summen

Warum sind die über Stationen gemittelten Temperaturdaten normalverteilt?

$$\boldsymbol{X_1, X_2, \ldots} \in \mathbb{R} \quad \textbf{uiv} \text{ ZV} \quad \text{mit } \boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X_i}] \in \mathbb{R} \text{ und } \boldsymbol{\sigma^2} = \mathbb{V}[\boldsymbol{X_i}] \in \boldsymbol{(0, \infty)}$$

$$\textbf{Gesetz der großen Zahlen} \quad \boldsymbol{S_m} = \sum_{i=1}^{m} \boldsymbol{X_i}\ , \quad \frac{\mathbf{1}}{\boldsymbol{m}} \boldsymbol{S_m} \to \boldsymbol{\mu} \quad \text{für } \boldsymbol{m} \to \infty$$

$$\text{denn} \quad \mathbb{E}\left[\frac{\mathbf{1}}{\boldsymbol{m}} \boldsymbol{S_m}\right] = \frac{\mathbf{1}}{\boldsymbol{m}} \boldsymbol{m\mu} = \boldsymbol{\mu} \quad \text{und} \quad \mathbb{V}\left[\frac{\mathbf{1}}{\boldsymbol{m}} \boldsymbol{S_m}\right] = \frac{\mathbf{1}}{\boldsymbol{m^2}} \boldsymbol{m\sigma^2} = \frac{\boldsymbol{\sigma^2}}{\boldsymbol{m}} \to \mathbf{0}$$

Grenzwertsätze für Summen

Warum sind die über Stationen gemittelten Temperaturdaten normalverteilt?

$$\boldsymbol{X_1, X_2, \ldots} \in \mathbb{R} \quad \textbf{uiv} \text{ ZV} \quad \text{mit } \boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X_i}] \in \mathbb{R} \text{ und } \boldsymbol{\sigma^2} = \mathbb{V}[\boldsymbol{X_i}] \in (\boldsymbol{0}, \boldsymbol{\infty})$$

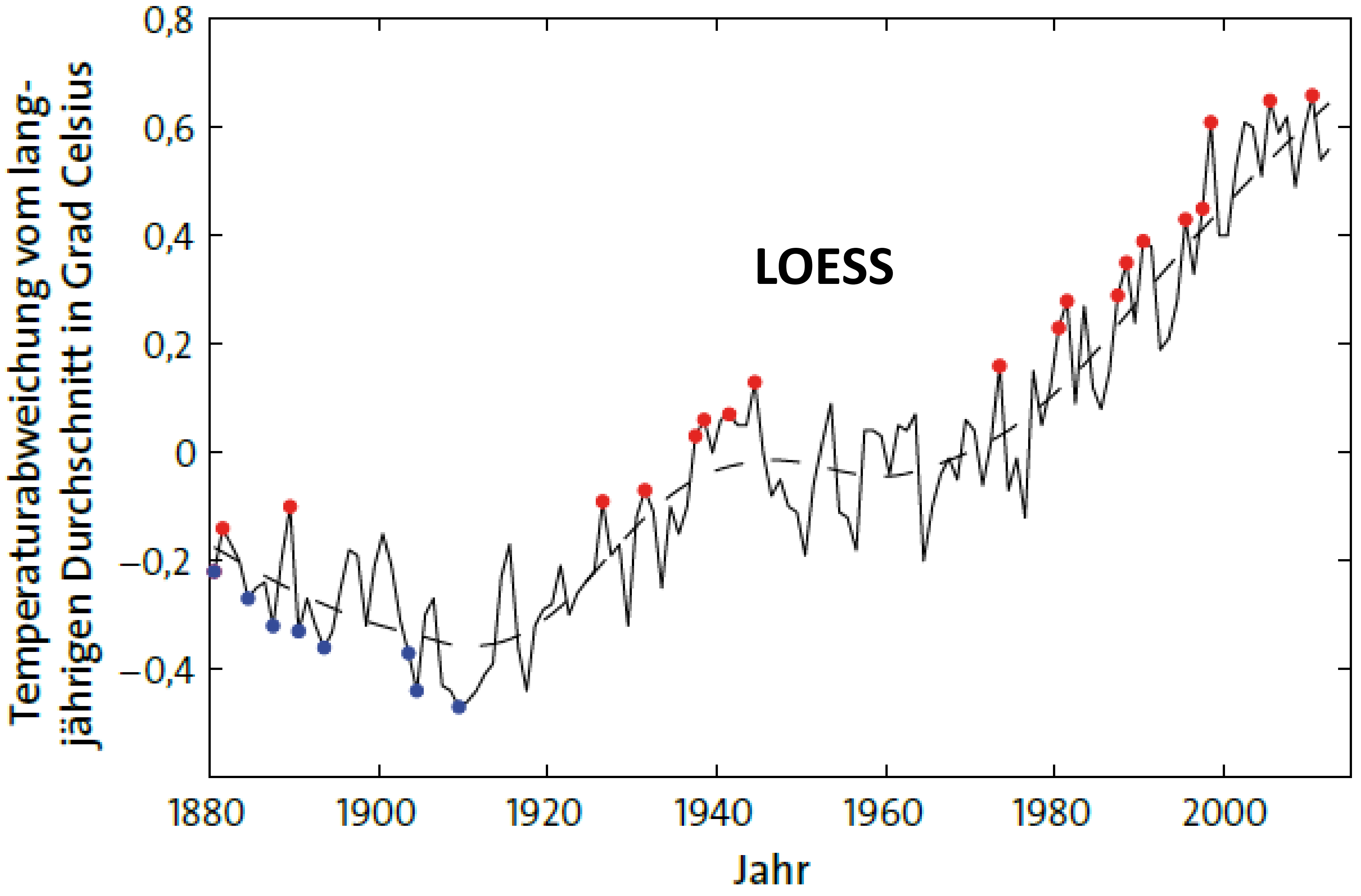
Gesetz der großen Zahlen $\boldsymbol{S_m} = \sum_{i=1}^{m} \boldsymbol{X_i}\ , \quad \frac{1}{m}\boldsymbol{S_m} \to \boldsymbol{\mu}$ für $\boldsymbol{m} \to \infty$

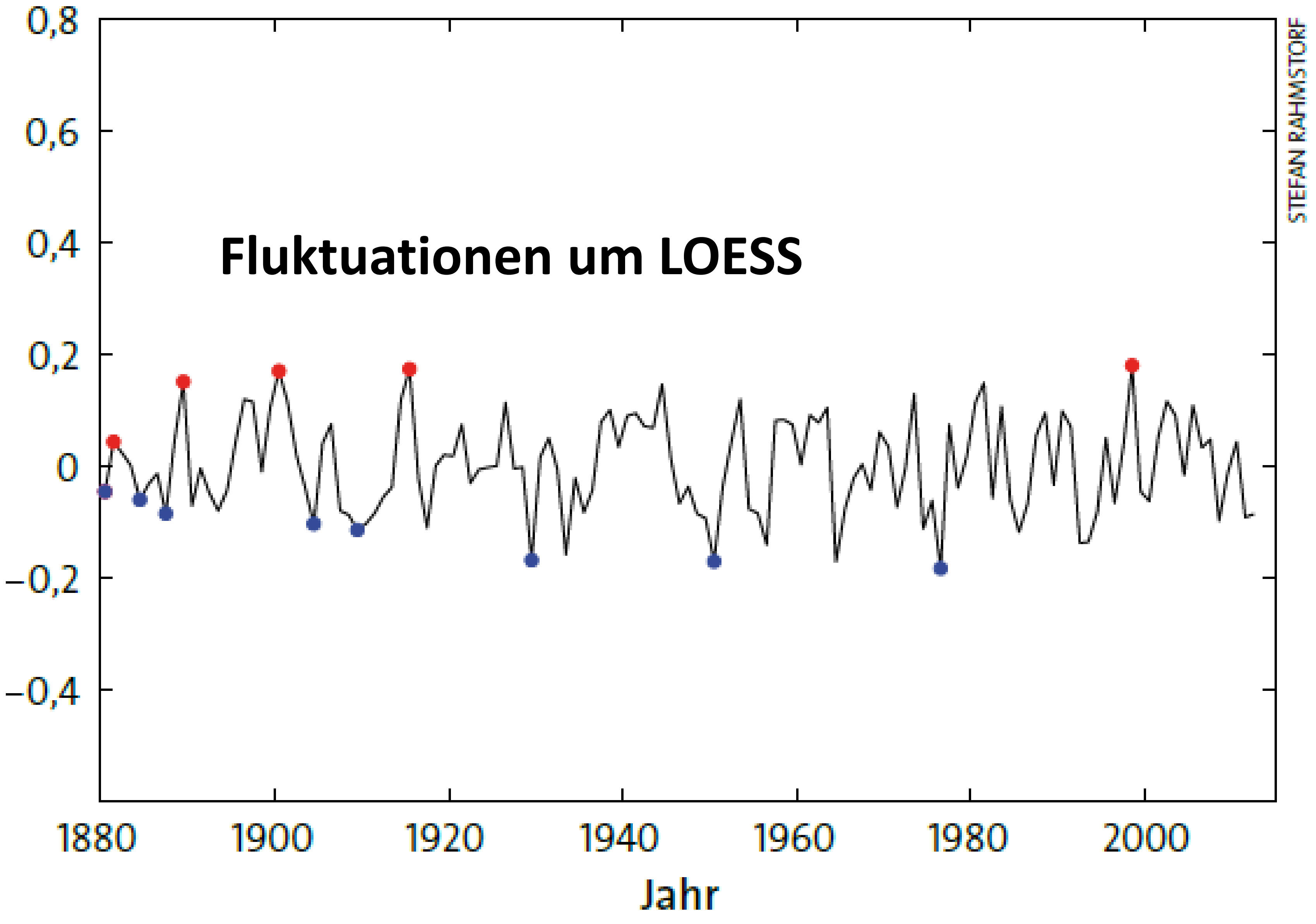
$$\text{denn} \quad \mathbb{E}\left[\frac{1}{m}S_m\right] = \frac{1}{m}m\mu = \mu \quad \text{und} \quad \mathbb{V}\left[\frac{1}{m}S_m\right] = \frac{1}{m^2}m\sigma^2 = \frac{\sigma^2}{m} \to 0$$

Zentraler Grenzwertsatz $\frac{S_m - m\mu}{\sqrt{m\sigma^2}} \to \xi \sim \mathcal{N}_{0,1}$ für $\boldsymbol{m} \to \infty$

$$\text{also} \quad \frac{1}{m}S_m \simeq \mu + \frac{\sigma}{\sqrt{m}}\xi + O(1/m) \quad \text{und Fluktuationen sind normalverteilt}$$

Temperaturmessungen – Trends und Schwankungen





Rekorde für unabhängige Zufallsvariablen (uiv)

$\boldsymbol{X_1, X_2, \ldots} \in \mathbb{R}$ **uiv** ZV, kont. Verteilung , d.h. $\mathbb{P}[\boldsymbol{X_i} = \boldsymbol{X_j}] = \boldsymbol{0},\ \boldsymbol{i} \neq \boldsymbol{j}$

$\boxed{\textbf{Ordnungsstatistik} \quad \boldsymbol{X^n_{(1)} < \ldots < X^n_{(n)}}}$ der ersten $\boldsymbol{n}$ Elemente

$$\mathbb{P}\left[\boldsymbol{X_1 = X^n_{(1)}, \ldots, X_n = X^n_{(n)}}\right] = \boldsymbol{1/n!}$$

Rekorde für unabhängige Zufallsvariablen (uiv)

$X_1, X_2, \ldots \in \mathbb{R}$ **uiv** ZV, kont. Verteilung , d.h. $\mathbb{P}[X_i = X_j] = 0,\ i \neq j$

$\boxed{\textbf{Ordnungsstatistik}\quad X^n_{(1)} < \ldots < X^n_{(n)}}$ der ersten n Elemente

$$\mathbb{P}\big[X_1 = X^n_{(1)}, \ldots, X_n = X^n_{(n)}\big] = 1/n!$$

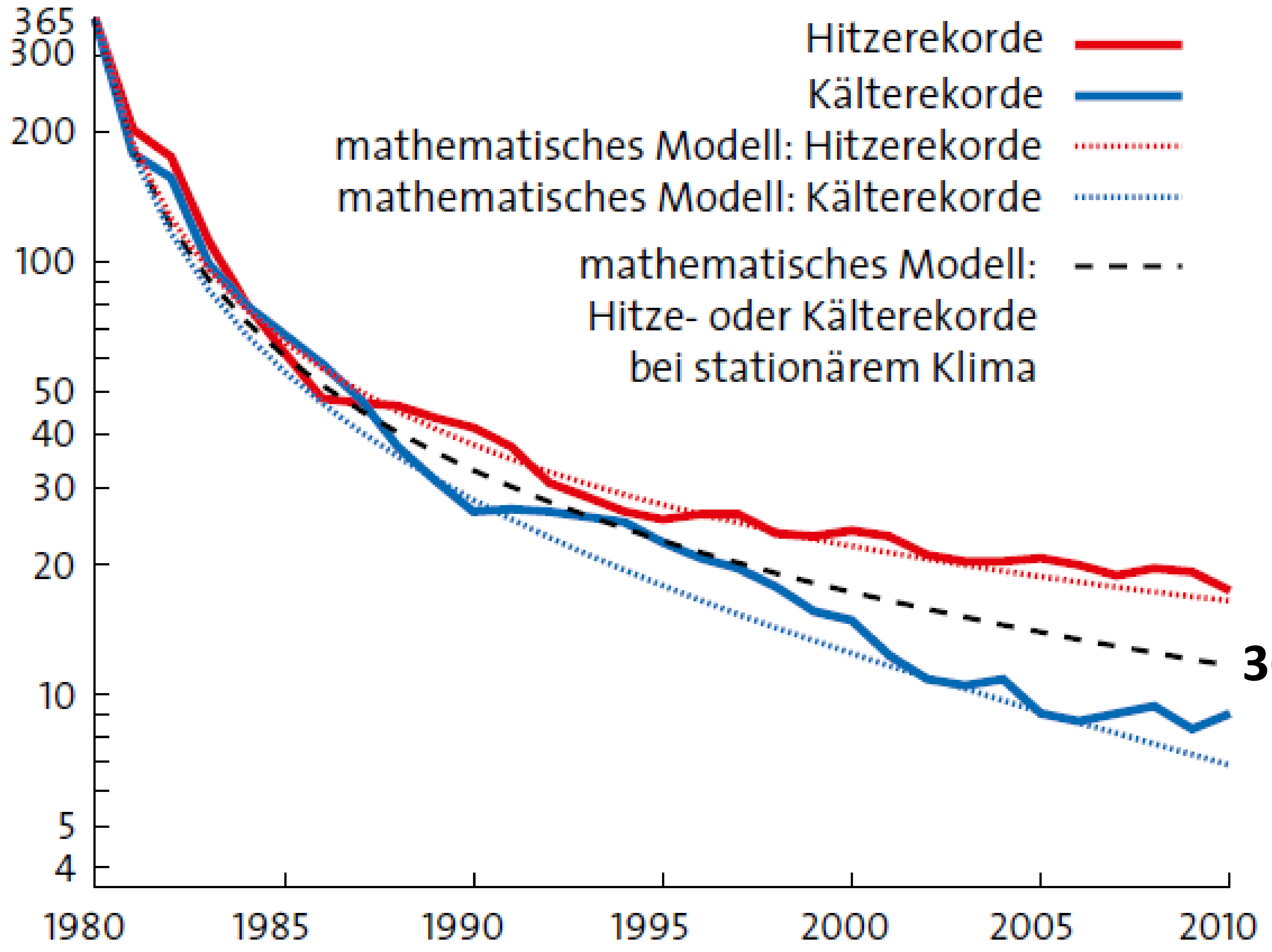
$$\boxed{\textbf{Positiv-Rekord}\quad \mathbb{P}\big[X_n = X^n_{(n)}\big] = \frac{(n-1)!}{n!} = \frac{1}{n}}$$

Rekord Ereignisse $A_n := \{X_n = X^n_{(n)}\}$ sind **unabhängig**

\# der Rekorde $R_n = \sum_{k=1}^{n} \mathbb{I}_{A_n}$ hat $\boxed{\mathbb{E}\big[R_n\big] = \sum_{k=1}^{n} \frac{1}{k} \simeq \ln n} + \gamma + O(\tfrac{1}{n})$

$$\mathbb{V}[R_n] = \mathbb{E}\big[(R_n - \mathbb{E}[R_n])^2\big] = \sum_{k=1}^{n} \frac{1}{k}\Big(1 - \frac{1}{k}\Big) \simeq \ln n - \frac{\pi^2}{6} + \gamma + O(\tfrac{1}{n})$$

Vergleich der Modellvorhersagen mit Klimadaten



Rekorde auf Tagesbasis

Für jede von über 200 europäischen Wetterstationen wurde ein Hitzerekord notiert, wenn die Maximaltemperatur eines Tages höher war als alle seit 1980 am selben Platz und am selben Kalendertag gemessenen Temperaturen; entsprechend zählte man Kälterekorde. Die dicken Linien zeigen für jedes Jahr die über alle Wetterstationen gemittelte Anzahl dieser Ereignisse.

365/n = erw. # der Rekordtage pro Jahr im Jahr 1979+n

Rekorde für unabhängige Zufallsvariablen mit Trend

$\boldsymbol{X_1, X_2, \ldots \in \mathbb{R}}$ **unabh.** mit Dichte $\boldsymbol{\rho_i(x)}$ und $\boldsymbol{F_i(x) = \int_{-\infty}^{x} \rho_i(y)dy}$

$$\boldsymbol{p_n := \mathbb{P}\big[X_n = X_{(n)}^n\big] = \int_{-\infty}^{\infty} \rho_n(x)\Big(\prod_{k=1}^{n-1} F_k(x)\Big)\,dx}$$

$\boxed{\textbf{Drift.}\quad \boldsymbol{X_k = X'_k + kv}}$ mit $\boldsymbol{X'_1, X'_2, \ldots}$ uiv , Entwicklung in $\boldsymbol{v} \in \mathbb{R}$

Rekorde für unabhängige Zufallsvariablen mit Trend

$$X_1, X_2, \ldots \in \mathbb{R} \quad \textbf{unabh.} \text{ mit Dichte } \rho_i(x) \text{ und } F_i(x) = \int_{-\infty}^{x} \rho_i(y)dy$$

$$p_n := \mathbb{P}\left[X_n = X_{(n)}^n\right] = \int_{-\infty}^{\infty} \rho_n(x)\left(\prod_{k=1}^{n-1} F_k(x)\right) dx$$

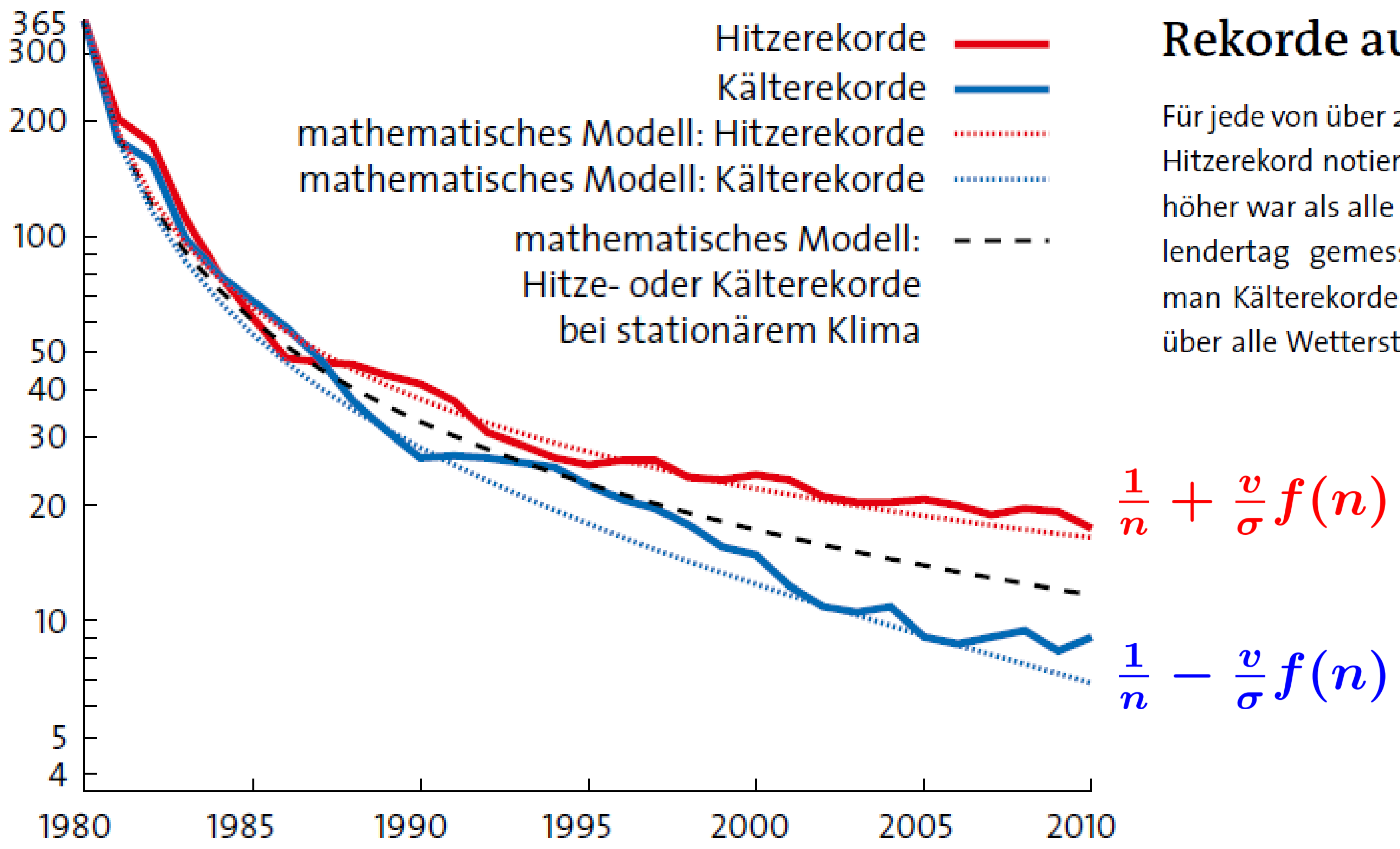
$$\boxed{\textbf{Drift.} \quad X_k = X_k' + kv} \quad \text{mit} \quad X_1', X_2', \ldots \text{ uiv }, \quad \text{Entwicklung in } v \in \mathbb{R}$$

$$p_n = \int_{-\infty}^{\infty} \rho(x)\left(\prod_{k=1}^{n-1} F(x+kv)\right) dx \simeq \boxed{\frac{1}{n} + v\frac{n(n-1)}{2}\int_{-\infty}^{\infty} \rho(x)^2 F(x)^{n-2} dx}$$

$$\boxed{X_k' \sim \mathcal{N}_{0,\sigma^2} \quad \text{normalverteilt}} \quad \Rightarrow \quad p_n \simeq \frac{1}{n} + \frac{v}{\sigma}\frac{2\sqrt{\pi}}{e^2}\sqrt{\ln\left(\frac{n^2}{8\pi}\right)}, \quad n \geq 7$$

$$\boxed{p_n \simeq \tfrac{1}{n} \pm \tfrac{v}{\sigma} f(n)}, \quad f(30) = 0.91, \quad f(100) = 1.17$$

Vergleich der Modellvorhersagen mit Klimadaten



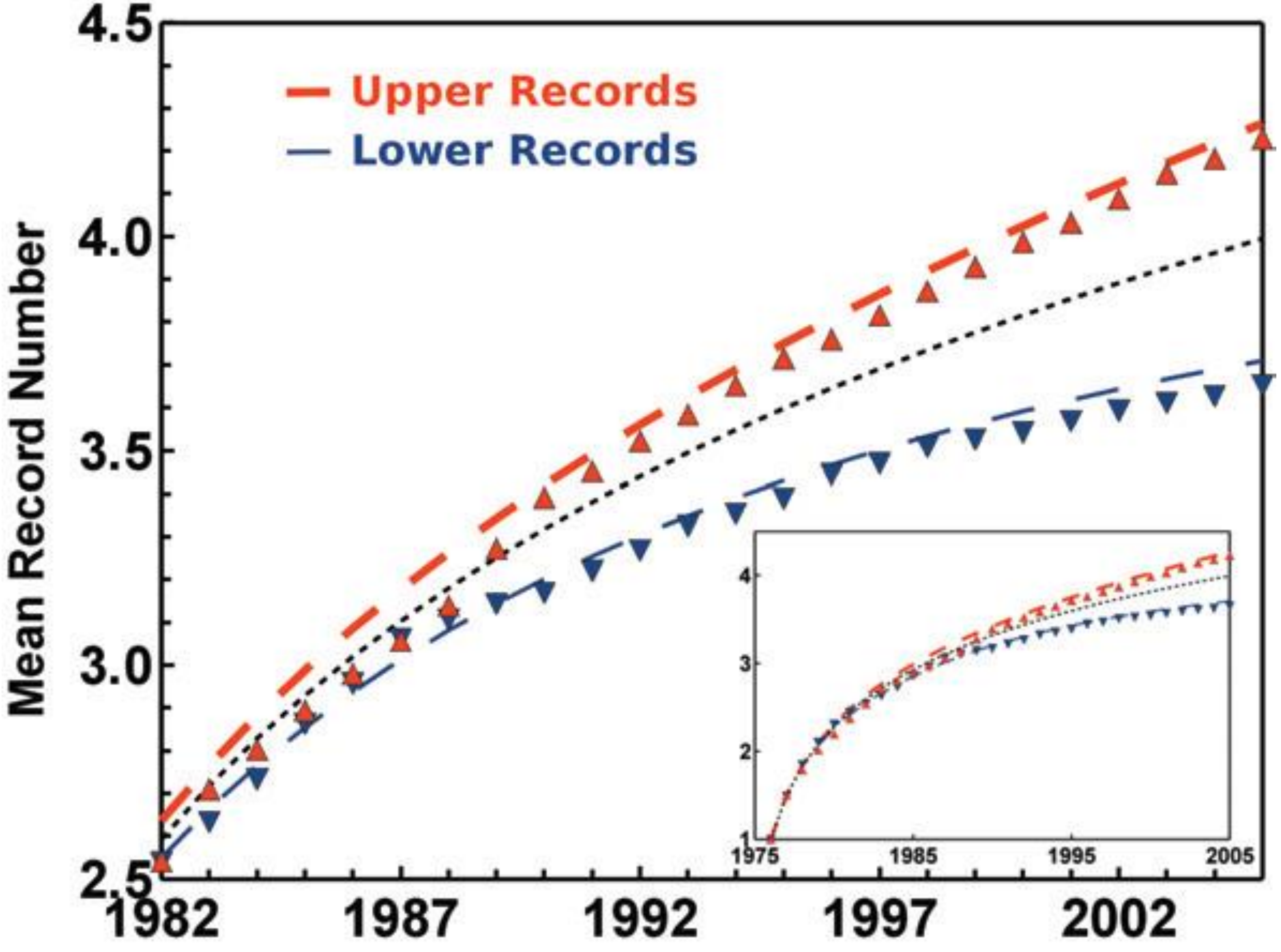
Rekorde auf Tagesbasis

Für jede von über 200 europäischen Wetterstationen wurde ein Hitzerekord notiert, wenn die Maximaltemperatur eines Tages höher war als alle seit 1980 am selben Platz und am selben Kalendertag gemessenen Temperaturen; entsprechend zählte man Kälterekorde. Die dicken Linien zeigen für jedes Jahr die über alle Wetterstationen gemittelte Anzahl dieser Ereignisse.

1980-2010:

v=0.045 Grad / Jahr
σ=3.3 Grad
v/σ=0.014 / Jahr

Vergleich der Modellvorhersagen mit Klimadaten



$$\mathbb{E}[R_n] = \sum_{k=1}^{n} 1/k \simeq \ln n + \gamma + O(1/n)$$

$$R_n = \text{Rekorde pro Tag im Jahr } 1975 + n$$

Grenzwertsätze für Maxima

$X_1, X_2, \ldots \in \mathbb{R}$ **uiv** mit $\overline{F}(x) = \mathbb{P}[X > x] \to 0,\ x \to \infty$ "exponenentiell"

Maximum $M_n = \max\{X_1, \ldots, X_n\}$, $\boxed{\mathbb{P}[M_n \leq x] = \left(1 - \overline{F}(x)\right)^n}$

wähle $x = x_n$ so dass $\mathbb{P}[M_n \leq x_n] \to e^{-c},\ c > 0 \quad \Leftrightarrow \quad \overline{F}(x_n) \simeq \dfrac{c}{n}$

Grenzwertsätze für Maxima

$X_1, X_2, \ldots \in \mathbb{R}$ **uiv** mit $\overline{F}(x) = \mathbb{P}[X > x] \to 0,\ x \to \infty$ "exponenentiell"

Maximum $M_n = \max\{X_1, \ldots, X_n\}$, $\boxed{\mathbb{P}[M_n \le x] = \left(1 - \overline{F}(x)\right)^n}$

wähle $x = x_n$ so dass $\mathbb{P}[M_n \le x_n] \to e^{-c},\ c > 0 \quad \Leftrightarrow \quad \overline{F}(x_n) \simeq \frac{c}{n}$

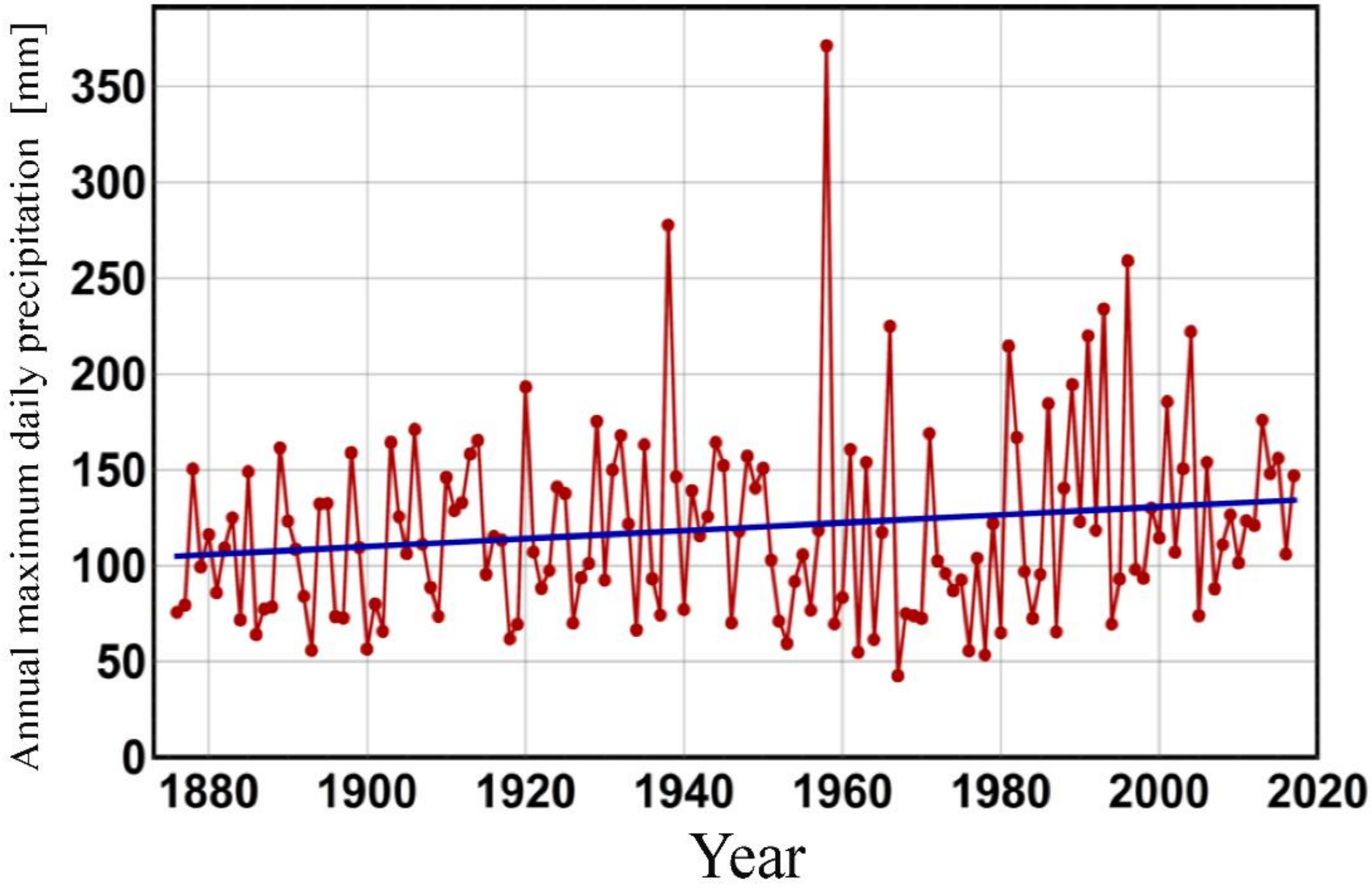
$$\boxed{\textbf{Extremwerttheorem} \quad \mathbb{P}\left[\frac{M_n - b_n}{a_n} \le x\right] \to \exp\left(-e^{-x}\right)}, \quad n \to \infty$$

standard **Gumbel Verteilung** mit Dichte $\rho(x) = \exp\left(-x - e^{-x}\right),\ x \in \mathbb{R}$

Extremwerttheorem (EVT) [Fisher, Tippet, Gnedenko]

- **Fréchet Verteilung** für Variablen mit schwerer Randverteilung (z.B. "power law")
- **Weibull Verteilung** für beschränkte Variablen (z.B. Uniform auf [a,b])

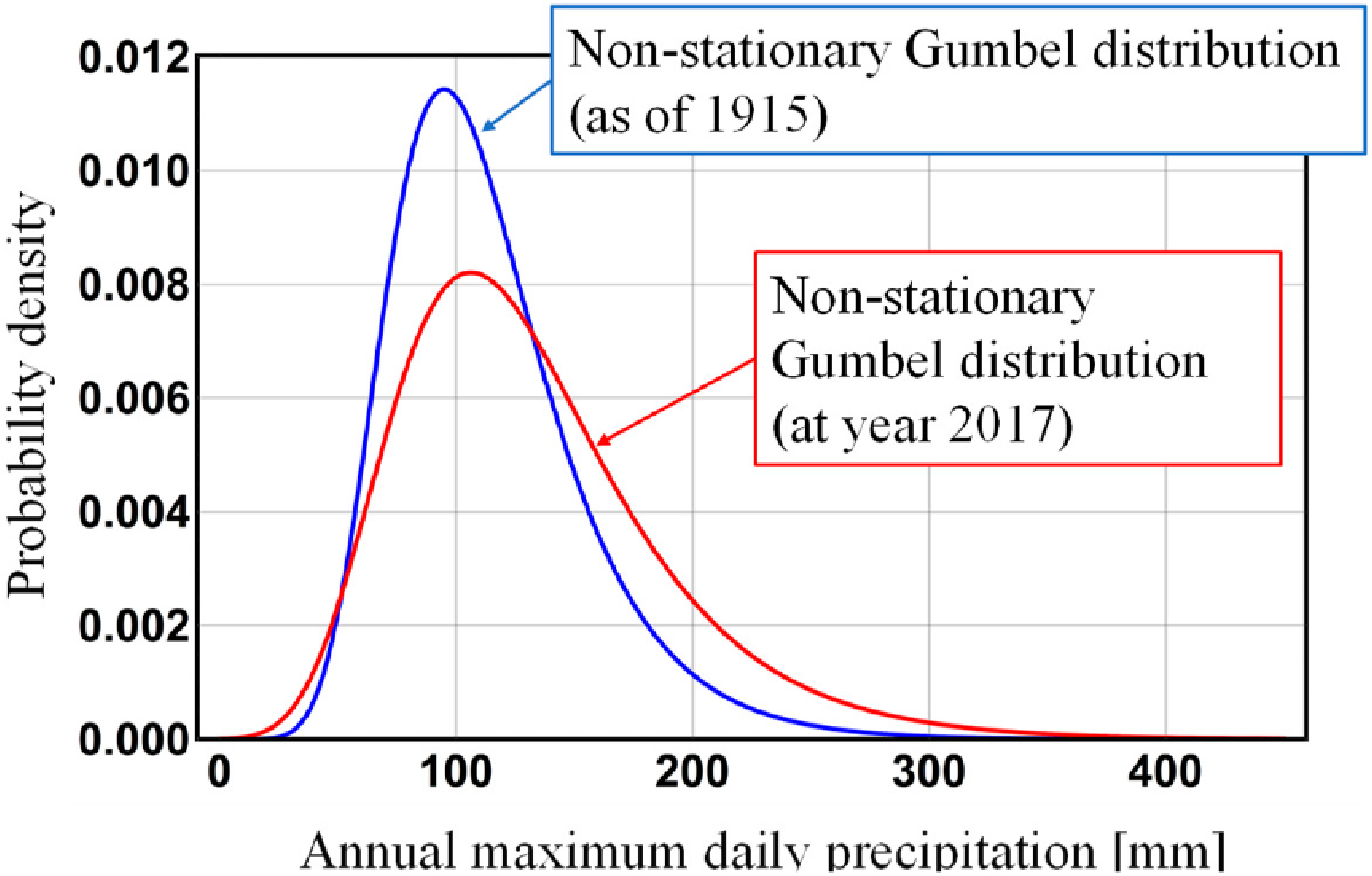
Anwendung auf Niederschläge



Time series showing observed data of the annual maximum daily precipitation over 142 years from **1876 to 2017** at the Meteorological Agency ground observation station **Tokyo** and the regression line fitted to this time series.

[Shimizu, Yamada, Yamada: Introduction of Confidence Interval Based on Probability Limit Method Test into Non-Stationary Hydrological Frequency Analysis, *Water* **2020**, *12*(10), 2727]

Anwendung auf Niederschläge



Gumble Verteilung

$$\boldsymbol{F(x) = \exp\left(-e^{-(x-\mu)/\beta}\right)}$$

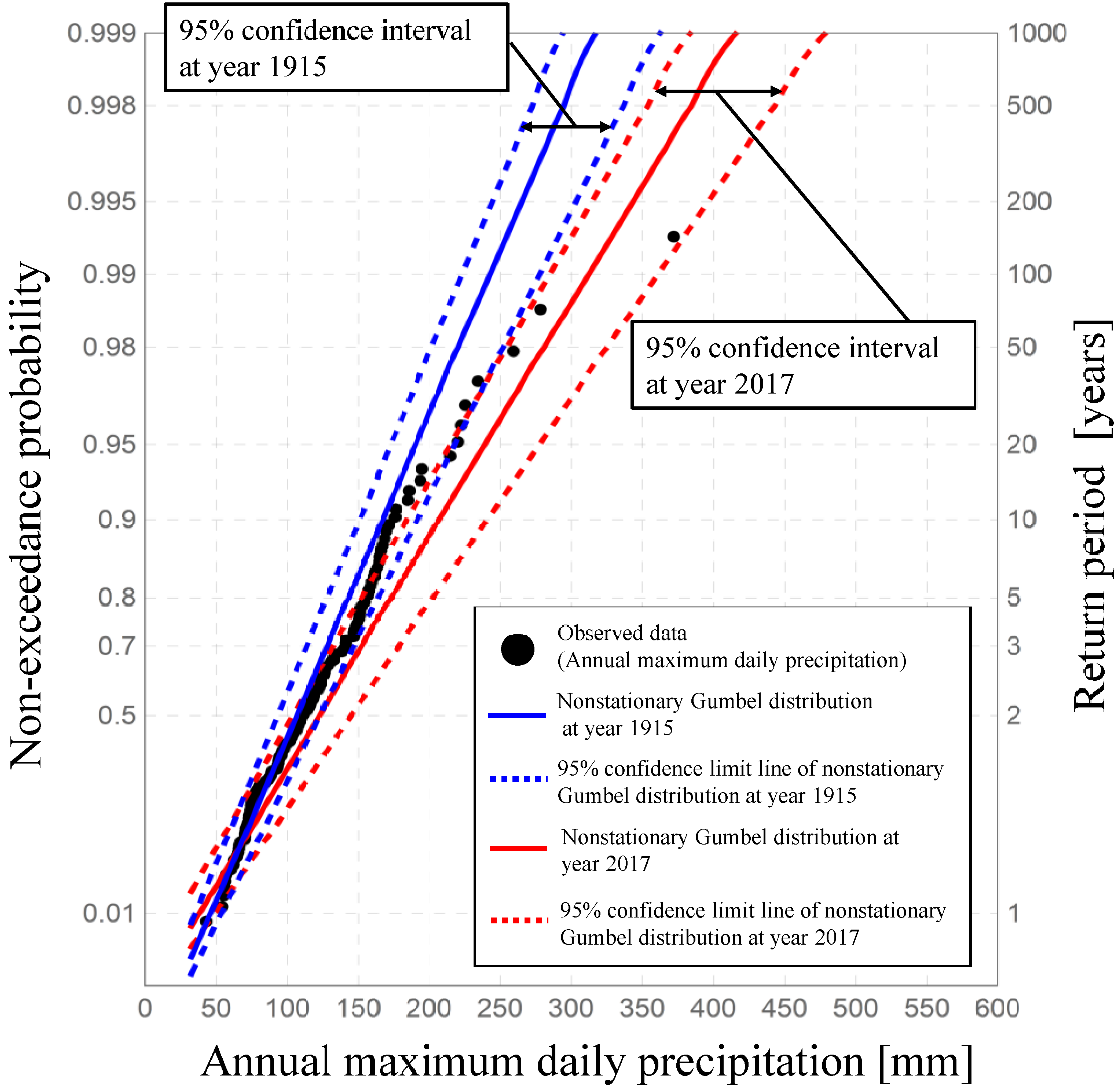
Lageparameter $\boldsymbol{\mu} \in \mathbb{R}$

Skalenparameter $\boldsymbol{\beta} > \mathbf{0}$

Anwendung auf Niederschläge

Jährlichkeit, Annuität, oder Frequenz

Observed data of the annual maximum daily precipitation over 142 years (from 1876 to 2017. at the ground observation station Tokyo, together with the non-stationary Gumbel distribution fitted to these 142 observed data.



Zusammenfassung

- Die Statistik von Extremwerten und Rekorden in Zeitreihen zeigen viele **universelle Eigenschaften**, die mit einfachen stochastischen Modellen untersucht werden können.

- Die Rate von Extremereignissen und die globale Temperatur nimmt in den letzten Jahrzehnten messbar zu. Mittels stochastischer Modelle kann die erhöhte Rate **quantitativ** auf die Erwärmung zurückgeführt werden.

- (Durch etablierte physikalische Modelle kann die Erderwärmung auf die Emission von Treibhausgasen zurückgeführt werden.)

- **Es ist jedoch im Allgemeinen nicht möglich, einzelne Extremereignisse als direkte Folge des Klimawandels zu identifizieren.**