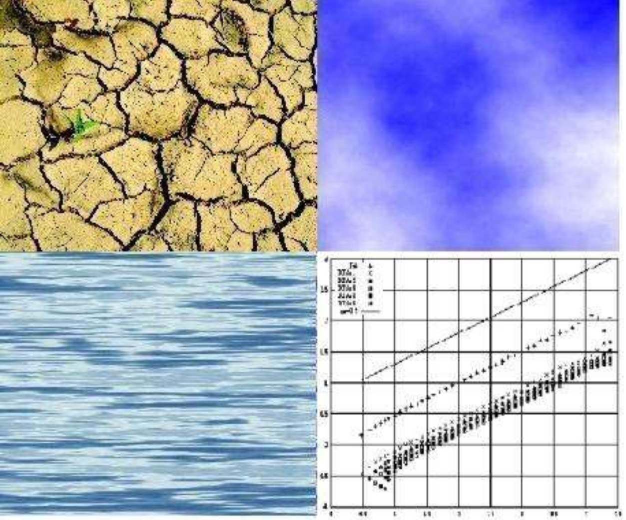


R&D Research Project: Scaling analysis of hydrometeorological time series data

Extreme Value Analysis considering Trends: Methodology and Application to Runoff Data of the River Danube Catchment

M. Kallache, H.W. Rust, H. Lange and J. Kropp

P I K

Bayerisches Landesamt für Umwelt

Bayreuther Zentrum für Ökologie und Umweltforschung

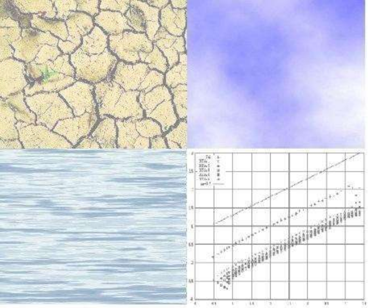
UNIVERSITÄT BAYREUTH

Bayceer

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

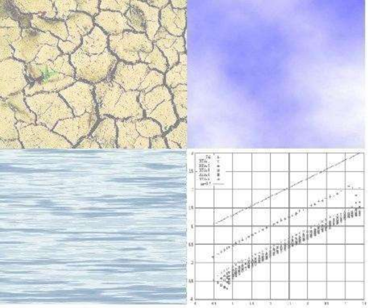
SKOG FORSK

Bar-Ilan University



Structure

1) Motivation
2) Generalized Extreme Value Distribution (GEV)
3) Non-Stationary GEV
4) Trend Estimation
5) Significance of Trend
6) Simulation Studies
7) Catchment Danube Example
    a) Shape parameter estimation
    b) Trend form
    c) Prediction
    d) Seasonal Effects
8) Discussion and Outlook

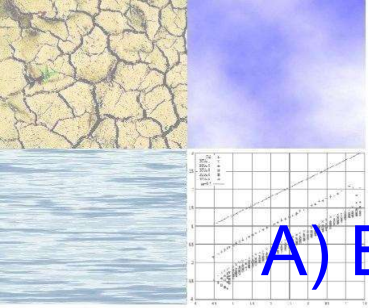


Motivation

How do trends in extreme values influence extreme value statistics?

Do streamflow peaks reveal a regionwide or even countrywide common trend?

Does global warming have an influence on the hydrological cycle?

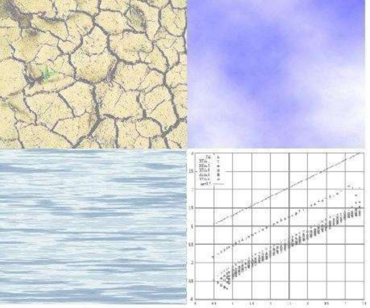


Extreme Value Statistics

A) Evaluate empirical data (order statistics, ...)
- No distribution assumption necessary
- Few data points available

B) Fit theoretical distributions to extrema
- Calculation for arbitrary positions possible
- Extrapolation possible
- Distribution assumption necessary, e.g.:
  - Log-Normal
  - Pearson
  - Generalized Pareto (GPD)
  - Generalized Extreme Value (GEV)

Is the assumed disribution
- good enough (goodness-of-fit)?
- better than other ones (Maximum Likelihood)?



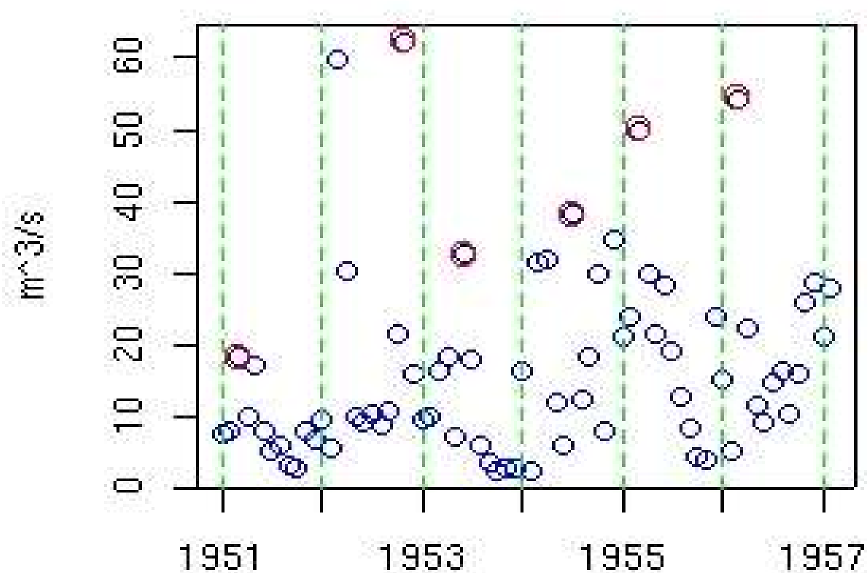
GEV Approximation of Maxima

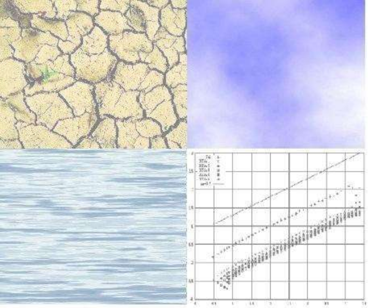
$$X_{1,} X_{2,} \ldots.. , X_n \qquad F(z) = Pr\{X_j \leqslant z\} \qquad \boxed{i.i.d}$$

$$M_n = max\{X_1, \ldots.., X_n\}$$

$$Pr\{\frac{M_n - b_n}{a_n} \leqslant z\} \rightarrow G(z) \quad for \quad n \rightarrow \infty$$

$$G(z; \mu, \sigma, \xi) = \exp\{-[1 + \xi(\frac{z-\mu}{\sigma})]_+^{-1/\xi}\}$$

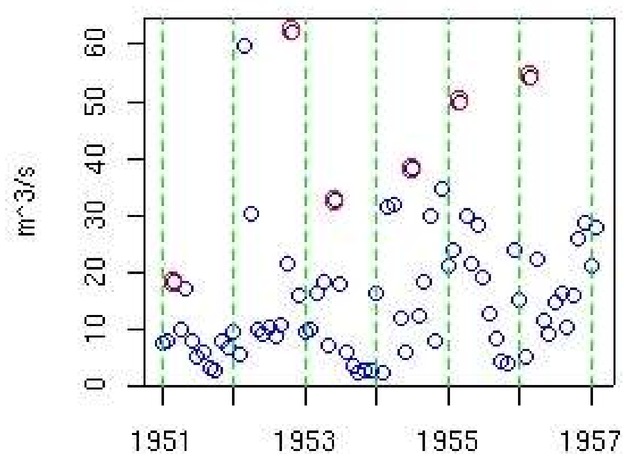


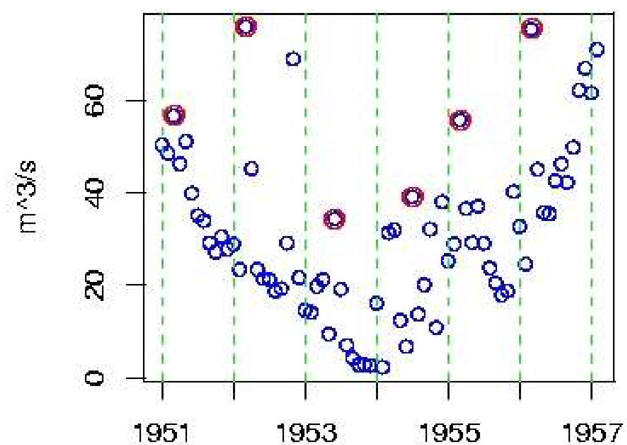


Non-Stationary GEV

What happens if data are not identically distributed?

$$Pr\{\ddot{M}_n \leqslant z\} \neq F(z)^n \qquad Pr\left\{\frac{\ddot{M}_n - b_n}{a_n} \leqslant z\right\} \not\to G(z) \quad for \quad n \to \infty$$

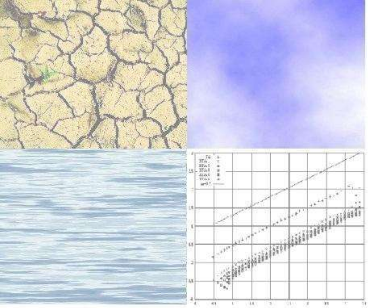




➡ Model time dependence of distribution explicitely, e.g.:

$$Z_t \sim G(\mu(t), \sigma(t), \xi(t))$$

$$\mu(t) = \beta_0 + \sum_{i=1}^{k} \beta_i t^i$$



Non-Stationary GEV II

Parameter estimation using Maximum Likelihood:

$$Z \sim f_{\mu,\sigma,\xi}(z) \;\Leftrightarrow\; \int_{-\infty}^{a} f_{\mu,\sigma,\xi}(z)\,dz \;=\; P(Z \le a)$$

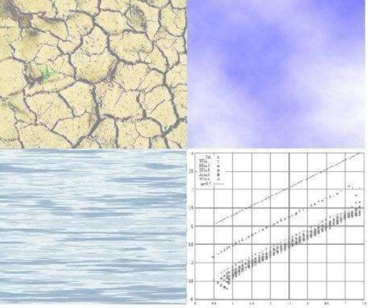
$$f_z(\mu,\sigma,\xi) \;=\; Likelihood$$

Example: μ dependent on time, which results in a different GEV for every single time point: G(μ (t),σ,ξ)

$$Log\ Likelihood$$

$$l(\mu,\sigma,\xi) = -m\log(\sigma) - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{z_t-\mu(t)}{\sigma}\right)\right]$$

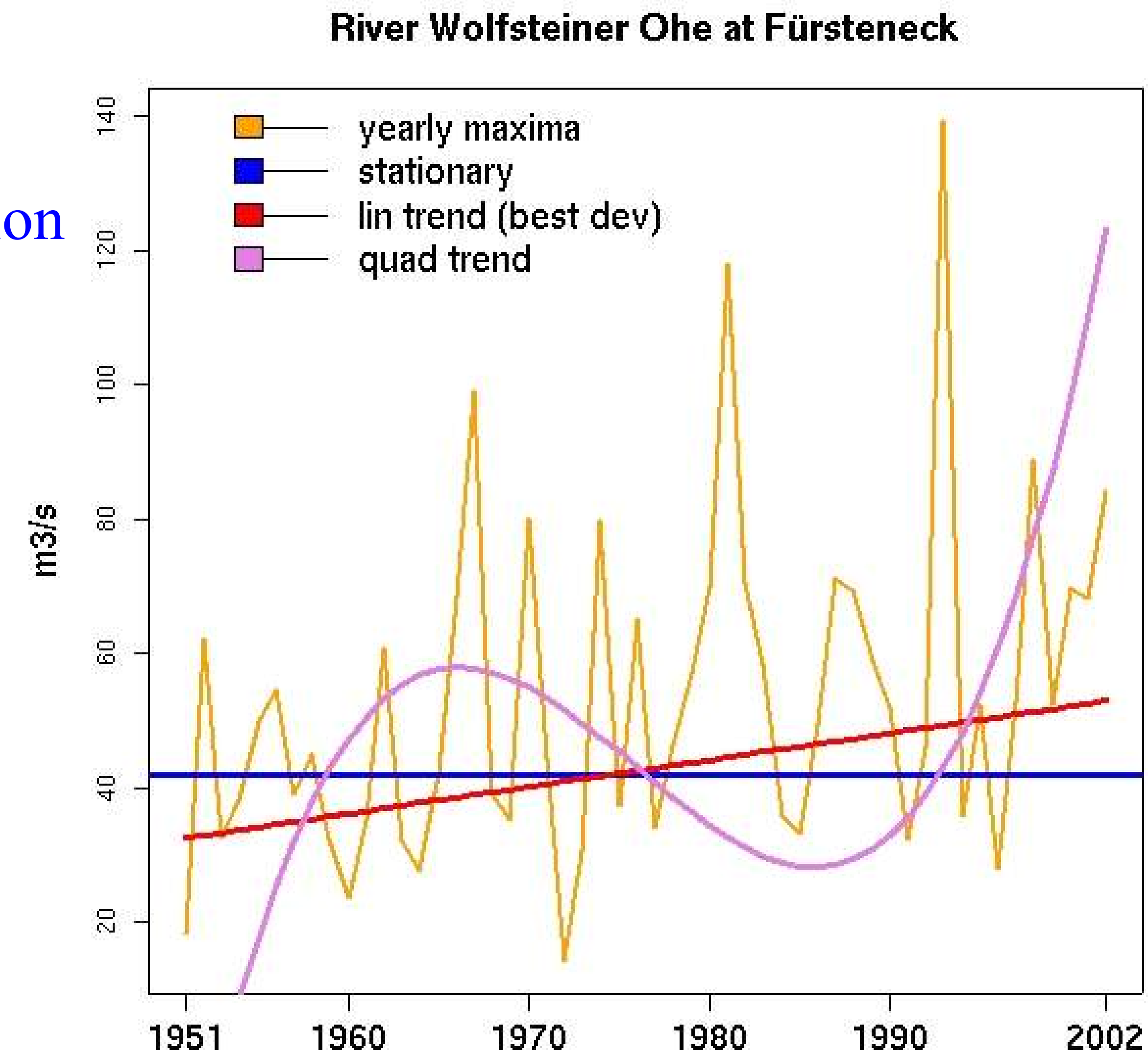
$$-\sum_{i=1}^{m}\left[1+\xi\left(\frac{z_t-\mu(t)}{\sigma}\right)\right]^{-1/\xi} \}$$

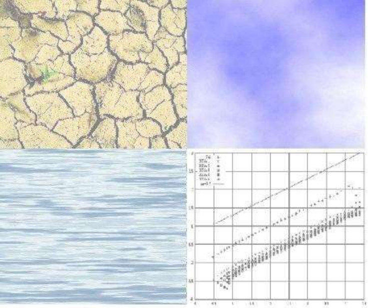


Trend Estimation

Polynomial Assumption

- Trend and GEV parameters fitted simultaneously.
- Distribution of data considered.
- Explicit assumption shape of the trend.





Test for Significance of Trend

Deviance Statistic

- Comparison of nested models (GEV with and without trend).
- Likelihood based.

$$D = 2\{l_c(M_c) - l_s(M_s)\}$$

$$M_c \quad complex\ model$$

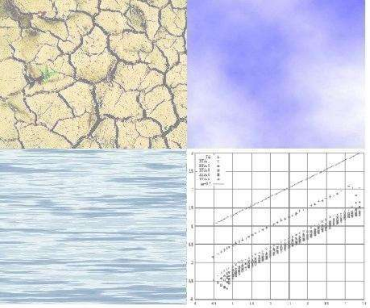
$$M_s \quad simple\ model$$

D is χ^2_m distributed (*m* = difference in degrees of freedom).

- Criterion: explanation of variation in data versus complexity of model.

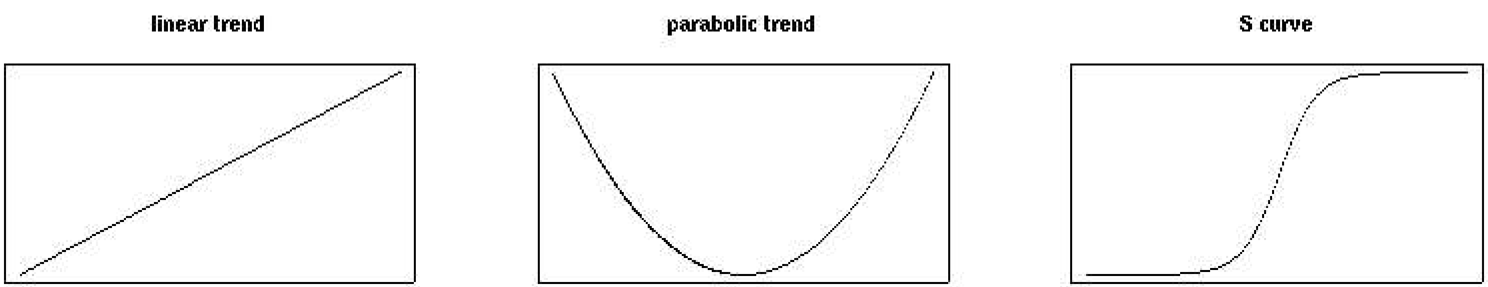
Goodness-of-Fit

- KS Test, Quantile Plots, Probability Plots, ...



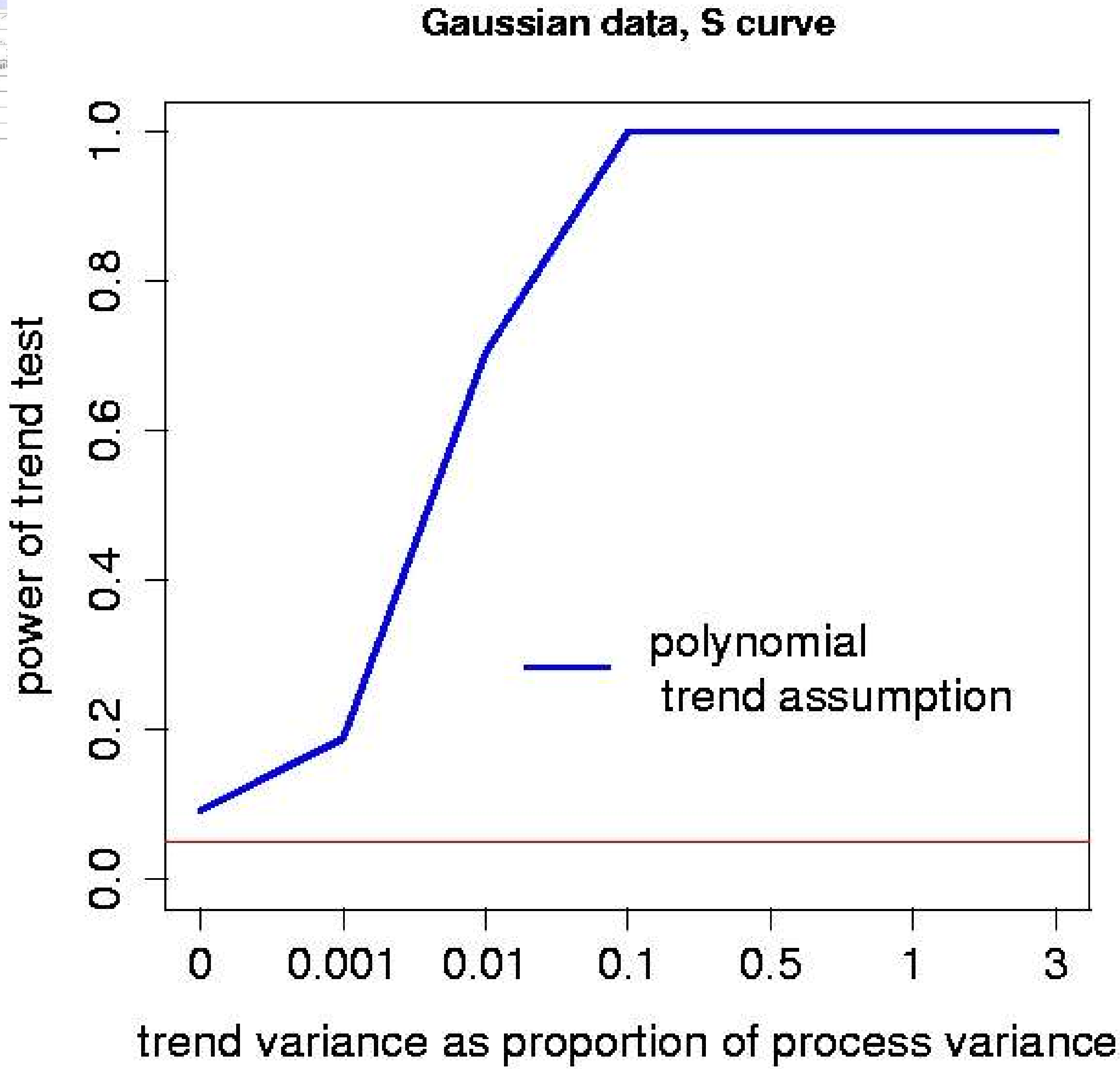
Simulation Settings

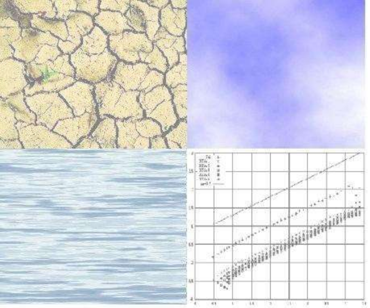
- Yearly maxima (528 years long).
- Distributions:
  - Gaussian, independent,
  - Pareto, independent,
  - Pareto, long-range correlated (δ = 0.2).
- Trends of various shapes added to the data.



- Variance of trend augmented: 0-3 times of process variance.
- Trend assessed with non-stationary GEV fit with polynomial trend assumption.

Power of Trend Test

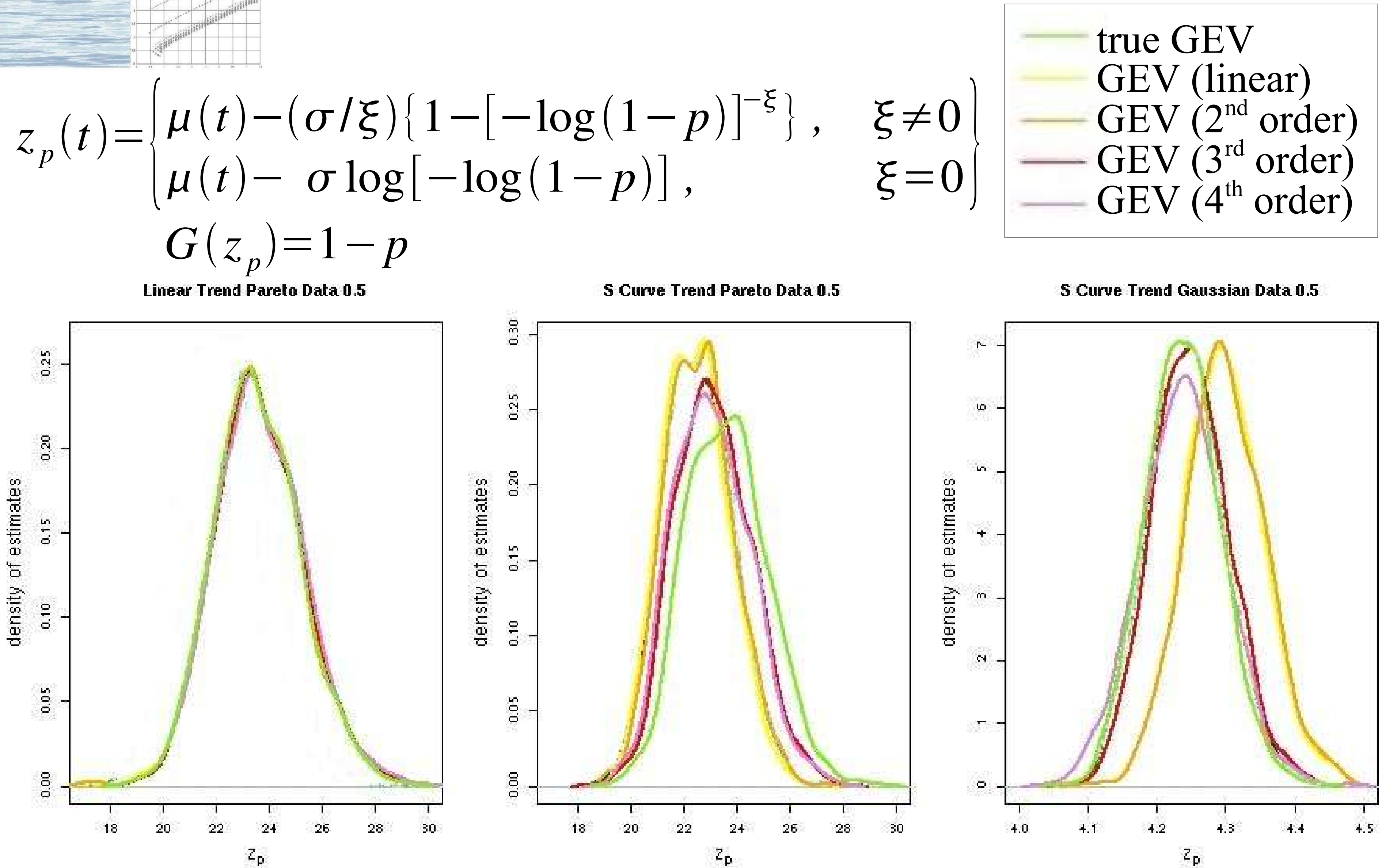




Quantile Estimation

$$z_p(t)=\begin{cases}\mu(t)-(\sigma/\xi)\{1-[-\log(1-p)]^{-\xi}\}, & \xi\neq 0\\ \mu(t)-\sigma\log[-\log(1-p)], & \xi=0\end{cases}$$

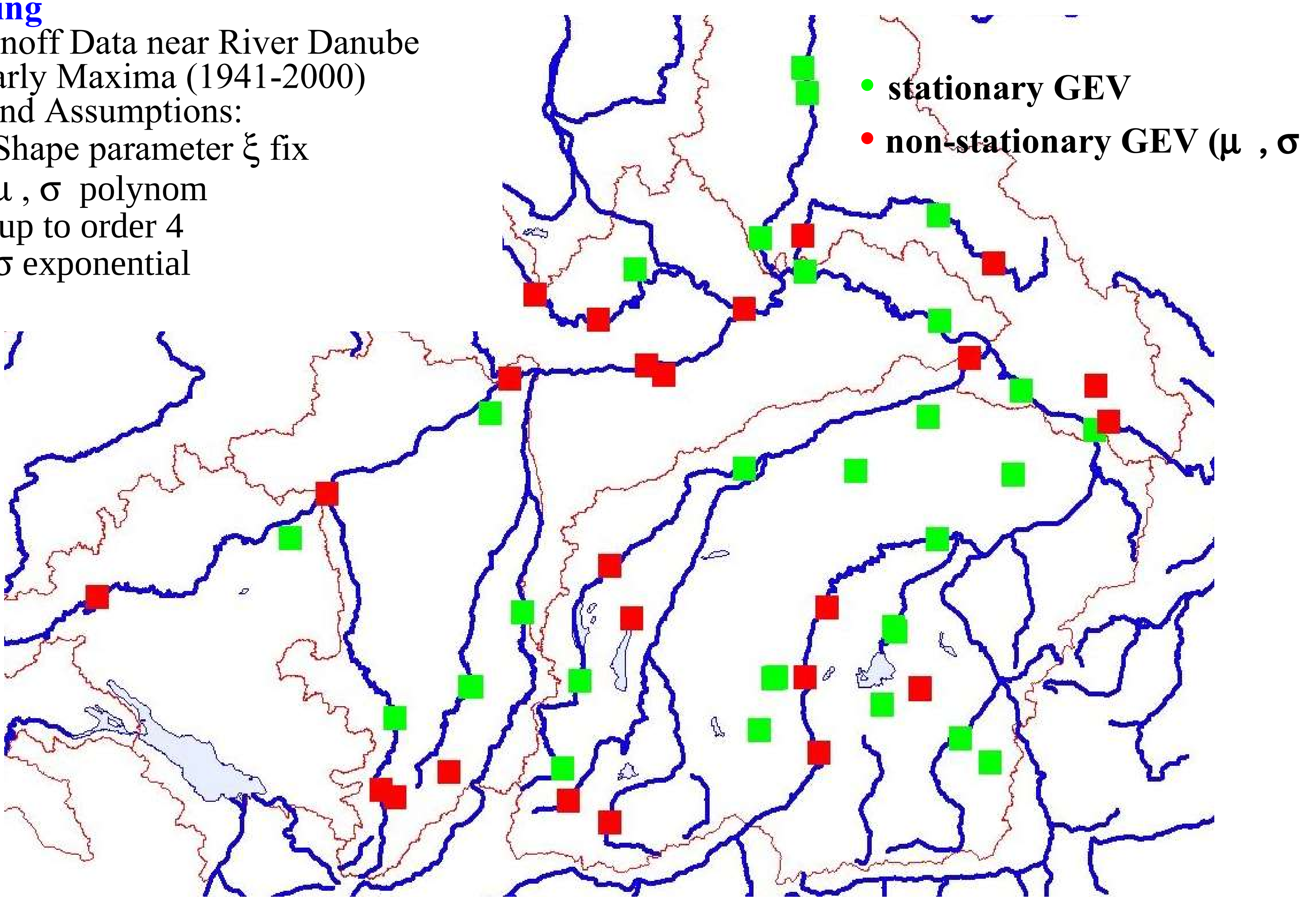
$$G(z_p)=1-p$$



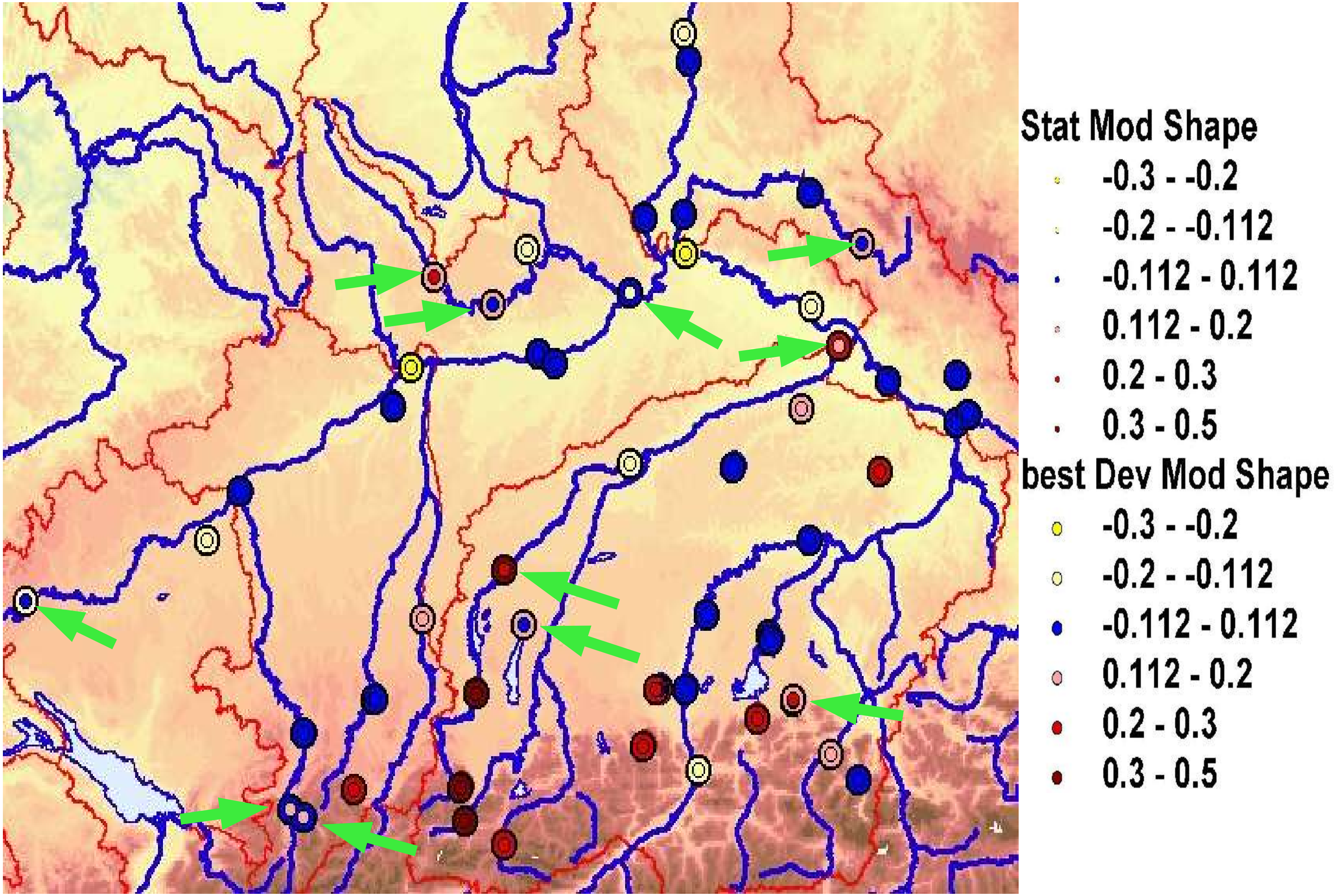
Yearly Maxima 1941-2000

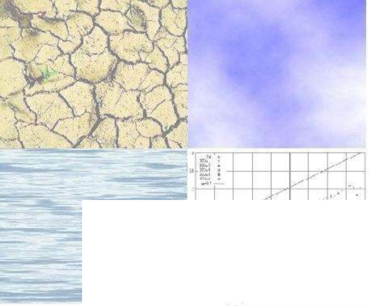
Setting

- Runoff Data near River Danube
- Yearly Maxima (1941-2000)
- Trend Assumptions:
  - Shape parameter ξ fix
  - μ , σ polynom up to order 4
  - σ exponential



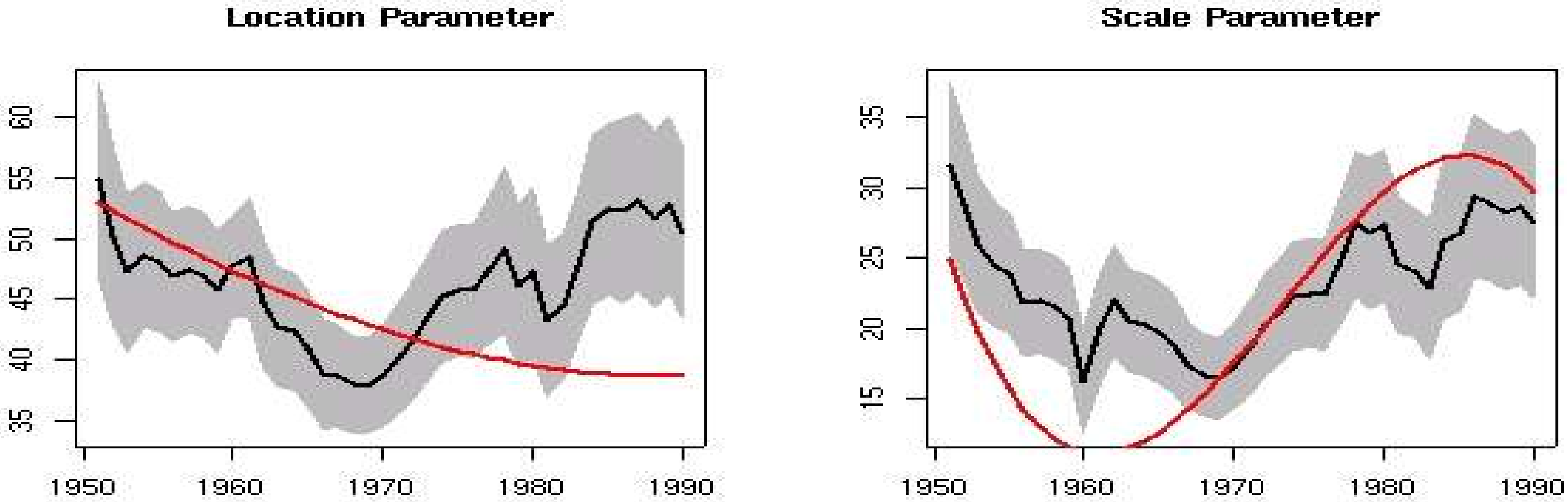
Shape Parameter Danube Catchment



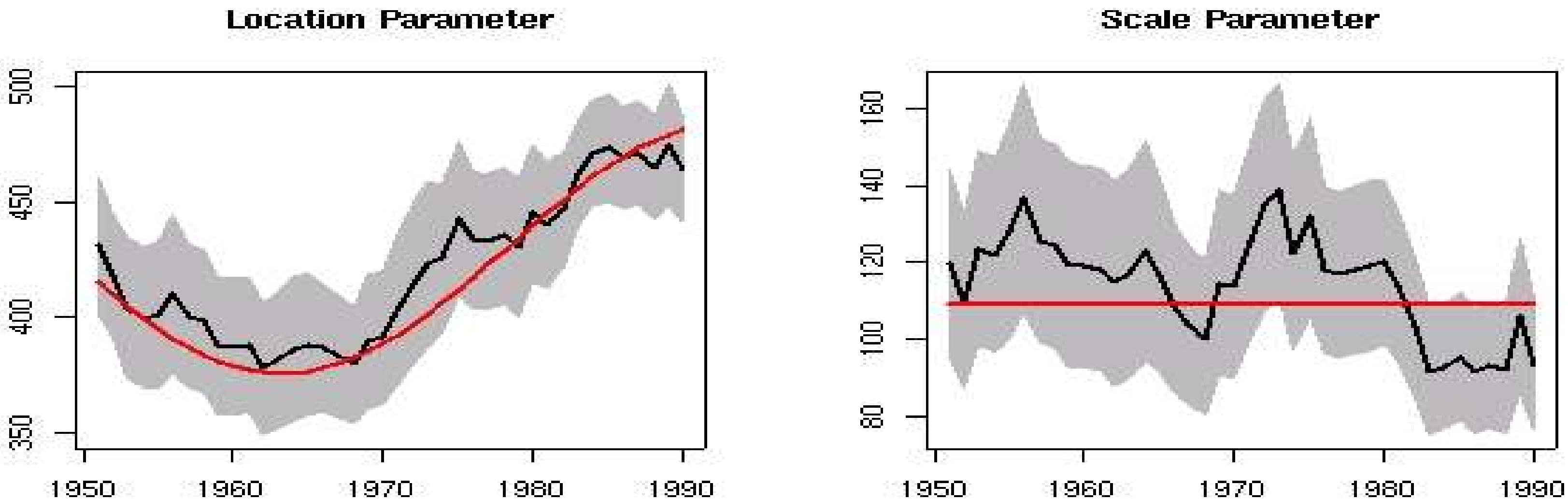


Plausability of Trend Form

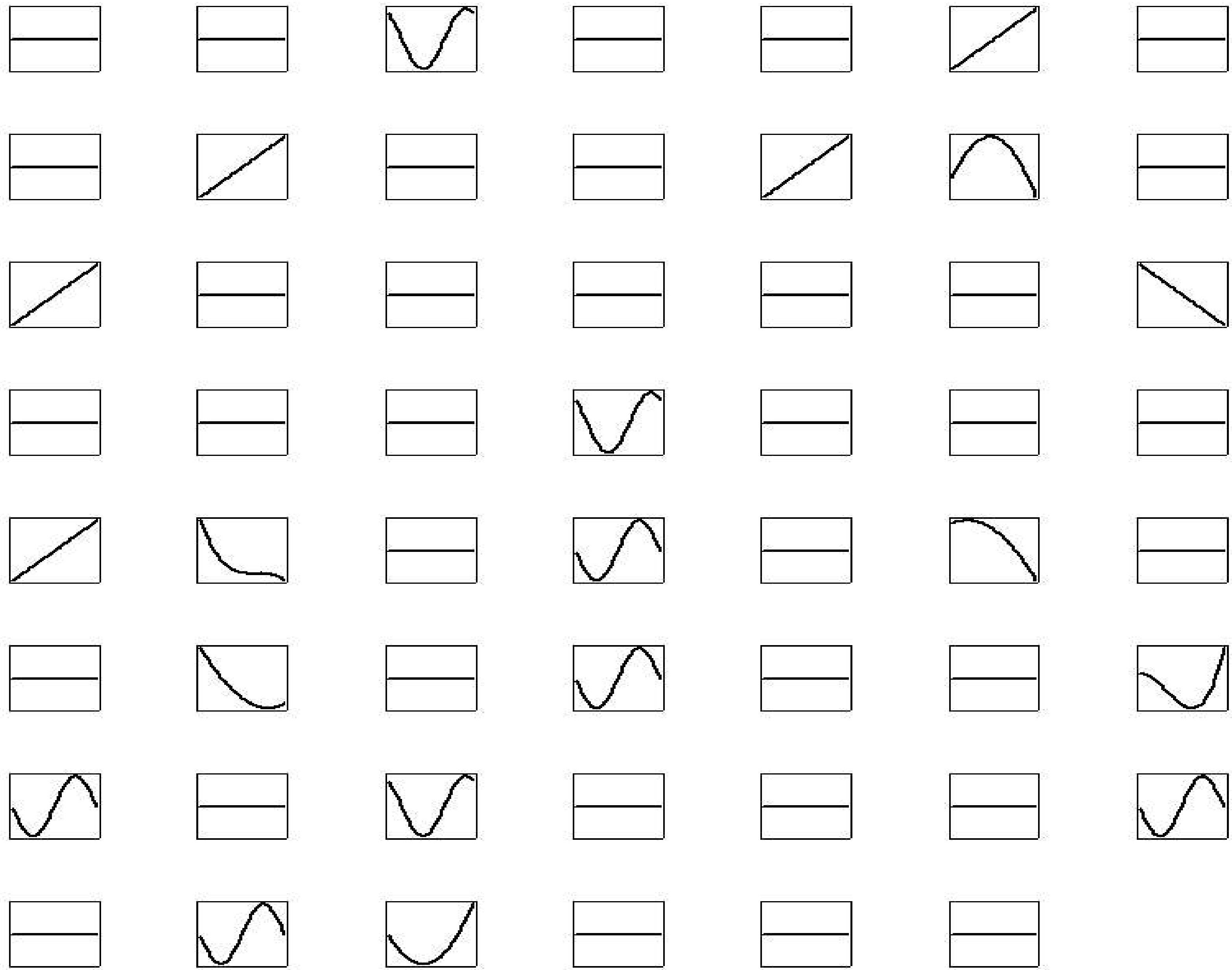
River Isar Mühlbächen /Plattling



River Altmühl / Eichstätt

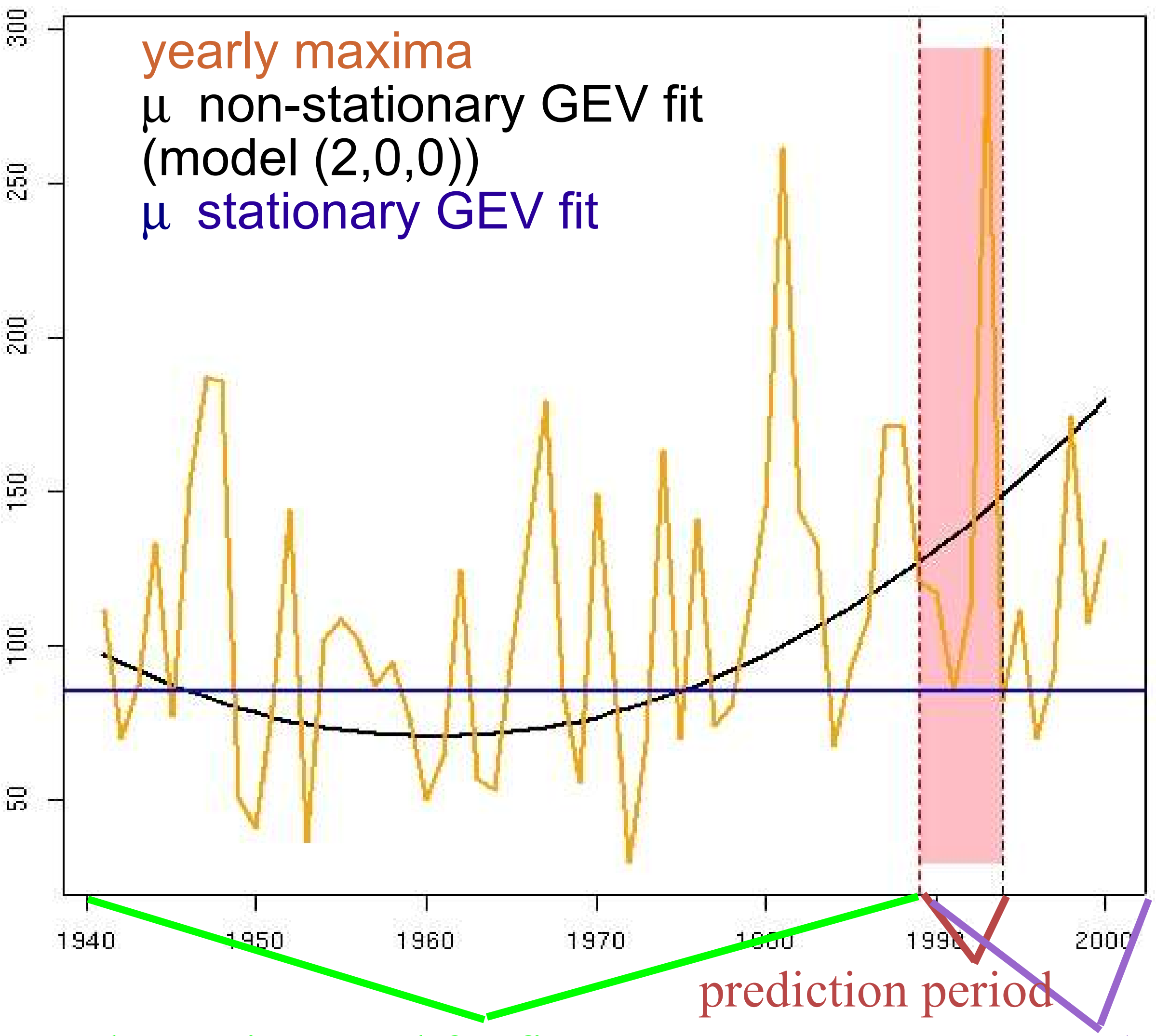


Trend in Location Parameter



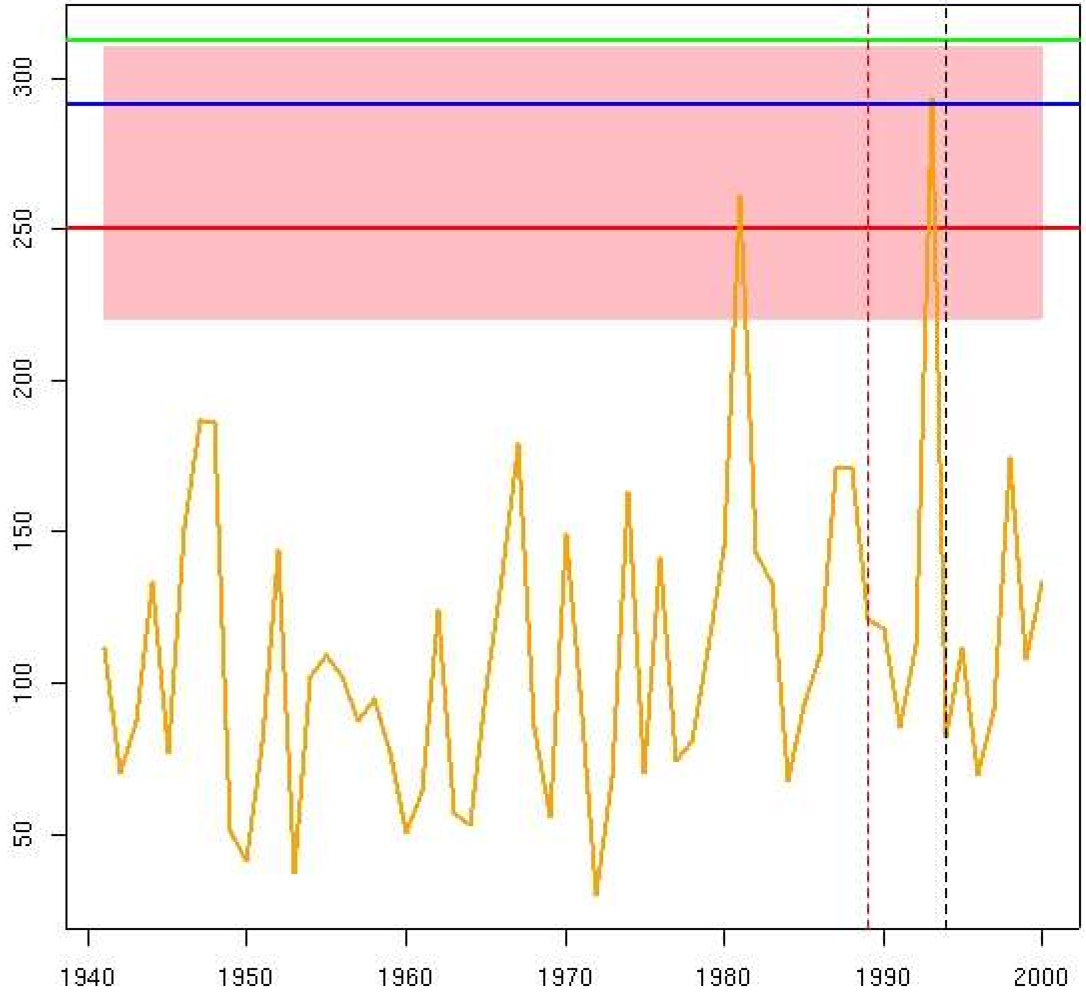
Prediction

Ilz/Kalteneck: Time Dependence of μ

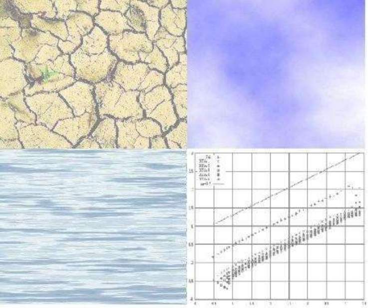


best model according to deviance statistics: (2,0,0)

Prediction II

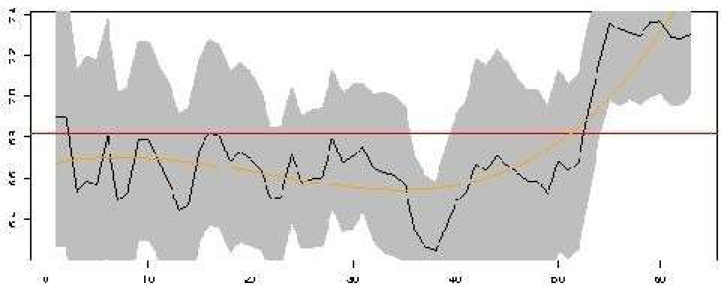


HQ 100 return levels calculated at 1989 for
− stationary GEV
− non-stationary GEV model (2,0,0)
− non-stationary GEV model (2,0,0) trend extrapolated, prediction period 5 years (until 1994)

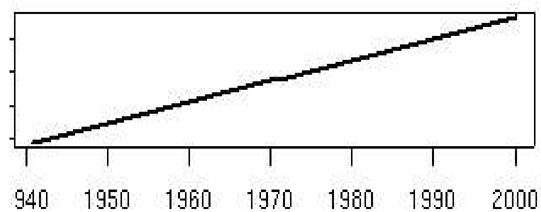


Seasonal Effects

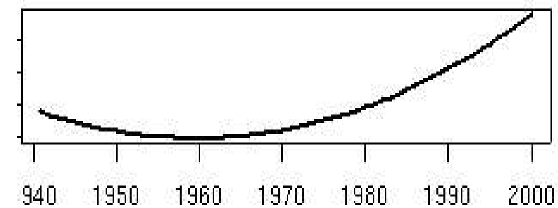
Yearly Maxima Location (windows)



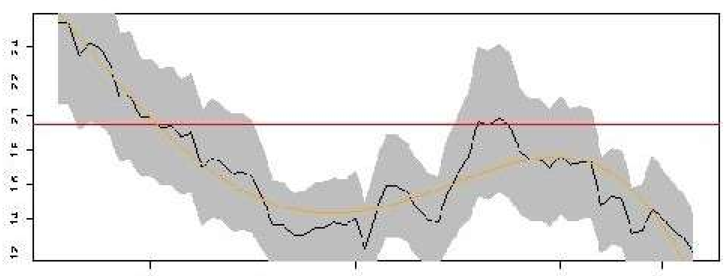
Yearly Maxima Location



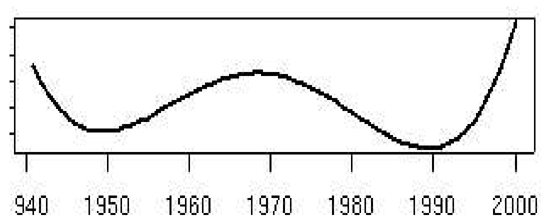
Max Nov-Feb Location



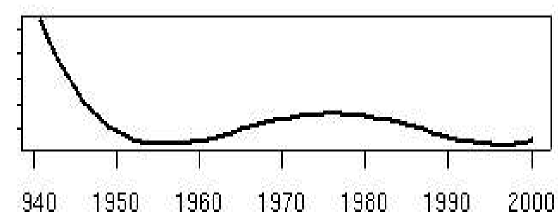
Yearly Maxima Scale (windows)



Yearly Maxima Scale



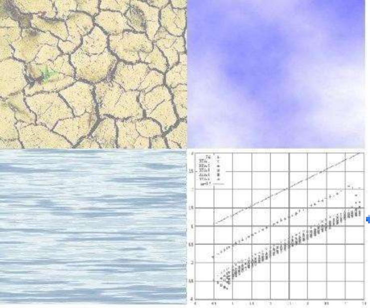
Max Nov-Feb Scale



River Leutstetten at Würm

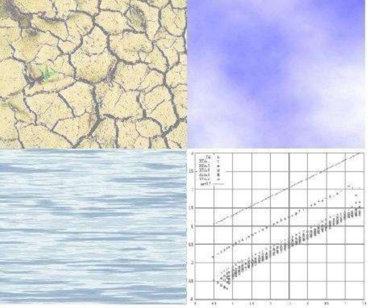
best model
yearly maxima:
(1,4,0)

Nov-Feb:
(2,4,0)
Mar-Jun:
(0,0,0) [stationary]
Jul-Oct:
(0,0,0) [stationary]



Discussion

- Trends in extreme values have to be considered when fitting theoretical distributions to empirical data:
  - they occure empirically,
  - otherwise (depending on the form of the trend) the estimated parameters may have a bias, even the important shape parameter.
- Maximum likelihood is applicable as formal criterion to detect trends (deviance statistics).
- Simulation studies reveal a good performance of non-stationary GEV fits with a polynomial trend assumption – even when analysing S-curved trends.
- Extreme values of river runoff do not show a regional trend tendency (as shown for the river Danube catchment).
- Nevertheless: regionwide trend analysis is useful to
  - detect patterns,
  - analyse underlying processes and
  - compensate scarcity of empirical data.



Outlook

- Consideration of correlations by using the General Pareto Distribution (GPD, peak over threshold)
- Usage of a semi-parametric approach incorporating GPD and splines (only local trend assumptions necessary).
- Incorporation of various covariates like the NAO index or weather regime data to analyse their influence on river runoff.

Thank you very much for your attention!