Hydrological Risk:
Recent advances in peak river flow modelling, prediction and real-time forecasting - Assessment of the impacts of land-use and climate changes
ESF LESC Exploratory Workshop, Bologna, Italy, October 24-25 2003

# On the appropriateness of the Gumbel distribution for modelling extreme rainfall

Demetris Koutsoyiannis
Department of Water Resources, School of Civil Engineering,
National Technical University, Athens, Greece

# Introduction

- **1914 (Hazen):** Empirical foundation of hydrological frequency curves known as "duration curves"
- **1922, 1923 (von Bortkiewicz, von Mises):** theoretical foundation of probabilities of extreme values
- **1958 (Gumbel):** convergence of empirical and theoretical approaches
- **Today**: the estimation of hydrological extremes continues to be highly uncertain

  *"… the increased mathematisation of hydrological frequency analysis over the past 50 years has not increased the validity of the estimates of frequencies of high extremes and thus has not improved our ability to assess the safety of structures whose design characteristics are based on them. The distribution models used now, though disguised in rigorous mathematical garb, are no more, and quite likely less, valid for estimating the probabilities of rare events than were the extensions 'by eye' of duration curves employed 50 years ago."* (Klemeš, 2000)

# The notion of distribution of maxima

- Parent variable: $Y$ (e.g. the daily rainfall depth)
- Parent distribution function: $F(y)$
- Variable representing maximum events

$$X := \max\{Y_1, Y_2, \ldots, Y_n\}$$

- Distribution function of maxima: $H_n(x)$
- Exact distribution of maxima for constant $n$:

$$H_n(x) = [F(x)]^n$$

- Exact distribution of maxima for randomly varying $n$, following a Poisson process

$$H'_\nu(x) = \exp\{-\nu[1 - F(x)]\}$$

# The notion of asymptotic or limiting distribution of maxima

- Asymptotic or limiting distribution for $n \rightarrow \infty$ or $\nu \rightarrow \infty$ (General extreme value distribution – GEV)

$$H(x) = \exp\{-[1 + (\kappa/\lambda)(x - \varepsilon)]^{-1/\kappa}\} \qquad (\kappa x \geq \kappa \varepsilon - \lambda)$$

- In hydrology, un upper bound of $x$ is not realistic, so $\kappa \geq 0$
- In case $\kappa > 0$, $H(x)$ represents the (three-parameter) extreme value distribution of maxima of type II (EV2)
- In the special case $\kappa = 0$, $H(x)$ represents the extreme value distribution of maxima of type I (EV1 or Gumbel)

$$H(x) = \exp\{-\exp[-(x - \varepsilon)/\lambda]\} \qquad (-\infty < x < +\infty)$$

- In the special case where the lower bound is zero ($\kappa \varepsilon = \lambda$), $H(x)$ is two-parameter EV2 (Fréchet distribution)

$$H(x) = \exp\{-(\varepsilon/x)^{1/\kappa}\} \qquad (x \geq 0)$$

# Why EV1 is so common in hydrology?

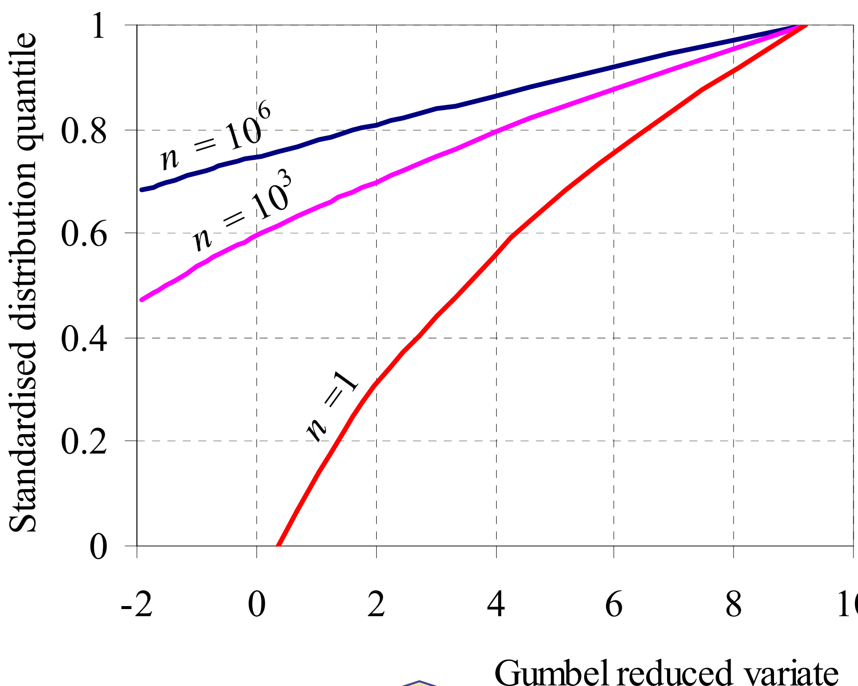
- **Theoretical reasons:** Most types of parent distributions functions used in hydrology, such as **exponential**, **gamma**, **Weibull**, **normal**, **lognormal**, and **EV1** itself belong to the domain of attraction of the Gumbel distribution
- **Simplicity:** The mathematical handling of the two-parameter EV1 is much simpler than that of the three-parameter EV2
- **Accuracy of estimated parameters:** Two parameters are more accurately estimated than three
- **Practical reasons:** EV1 offers a **linear probability plot** (Gumbel probability plot) of observed $x_H$ vs. observed $z_H := -\ln(-\ln H)$ (Gumbel reduced variate); in contrast, a linear probability plot is not possible for the EV2, unless it is two-parameter (Fréchet), which offers a linear plot of $\ln x_H$ vs. $z_H$
- Note: Empirical evidence shows that, in most cases (especially in rainfall maxima) plots of $\boldsymbol{x_H}$ **vs.** $\boldsymbol{z_H}$ give more straight-line arrangements than plots of $\boldsymbol{\ln x_H}$ **vs.** $\boldsymbol{z_H}$

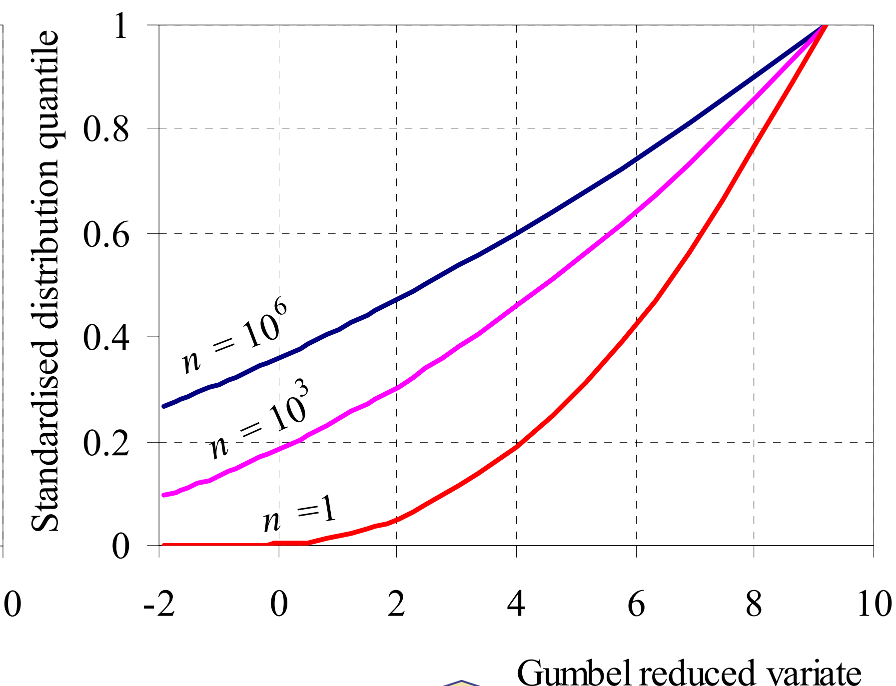
# The disadvantage of EV1

- It results in risk significantly higher than EV2 for engineering structures
  (i.e., for small probabilities of exceedence, or large return periods $T = 1 / (1 – H)$, it yields the smallest possible quantiles $x_H$ in comparison to those of the three-parameter EV2 for any value of the shape parameter $\kappa$)
- Normally, this would be a sufficient reason to avoid the use of EV1 in engineering studies
- However, EV1 has been the prevailing model for rainfall extremes

**Question**: Are the theoretical reasons favouring EV1 and the empirical evidence from hydrological records strong enough to counterbalance this disadvantage and justify the use of EV1?

# How fast is the convergence of the exact distribution of maxima to EV1?




Convergence of distribution of maxima for parent distribution standard normal

Convergence of distribution of maxima for parent distribution Weibull with shape parameter $k = 0.5$

Note: The distribution quantiles have been standardised by $x_{0.9999}$ corresponding to $z_H = 9.21$

# Do parent distributions of hydrological variables belong to the domain of attraction of EV1?

- Rainfall depth at fine time scales (hourly, daily) has been modelled by the gamma or Weibull distributions
- Both these distributions belong to the domain of attraction of EV1
- However, the parameters of distributions are not constant all the time but vary due to:
  - seasonal effects
  - overyear (large scale) fluctuations
- Parameter variations may change the domain of attraction to EV2

# Theoretical example of the shift of domain of attraction

**Assumptions:**

Probability density function of $Y_i$ conditional on $\alpha_i$:

$$f_i(y \mid \alpha_i) = \alpha_i^{\theta}\, y^{\theta - 1}\, \mathrm{e}^{-\alpha_i y} / \Gamma(\theta)$$

Probability density function of $\alpha_i$:

$$g(\alpha_i) = \beta^{\tau}\, \alpha_i^{\tau - 1}\, \mathrm{e}^{-\beta \alpha_i} / \Gamma(\tau)$$

**Results:**

Unconditional density function of $Y$:

$$f(y) = [1/\, \beta\, B(\theta + \tau)]\, (y/\beta)^{\theta - 1} / (1 + y/\beta)^{\tau + \theta}$$

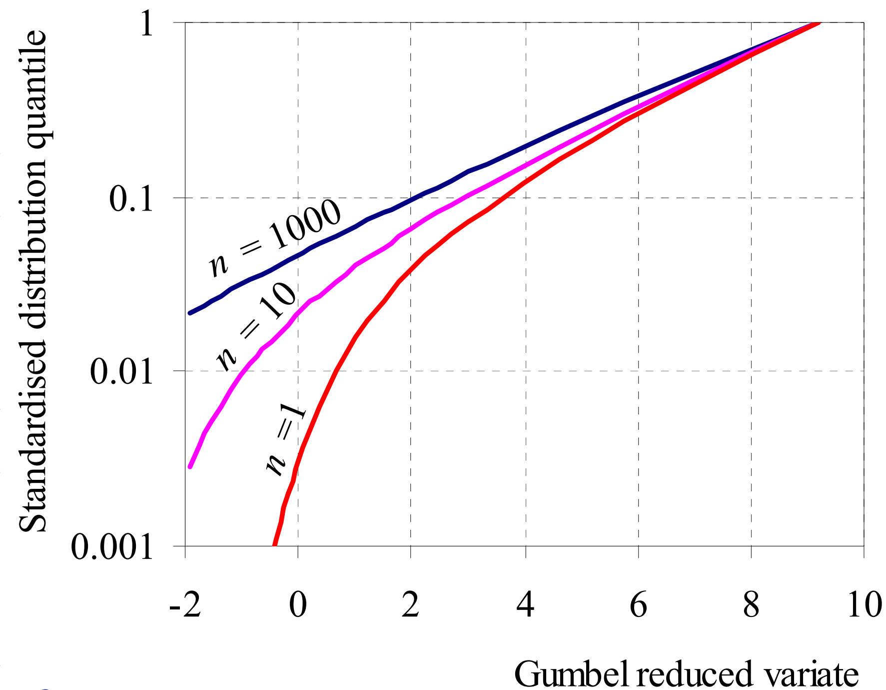
(Beta distribution of the second kind)

Exact distribution of maxima for constant $n$:

$$H_n(x) = [B_{x/(x + \beta)}\, (\theta, \tau) \,/\, B\, (\theta, \tau)]^n$$

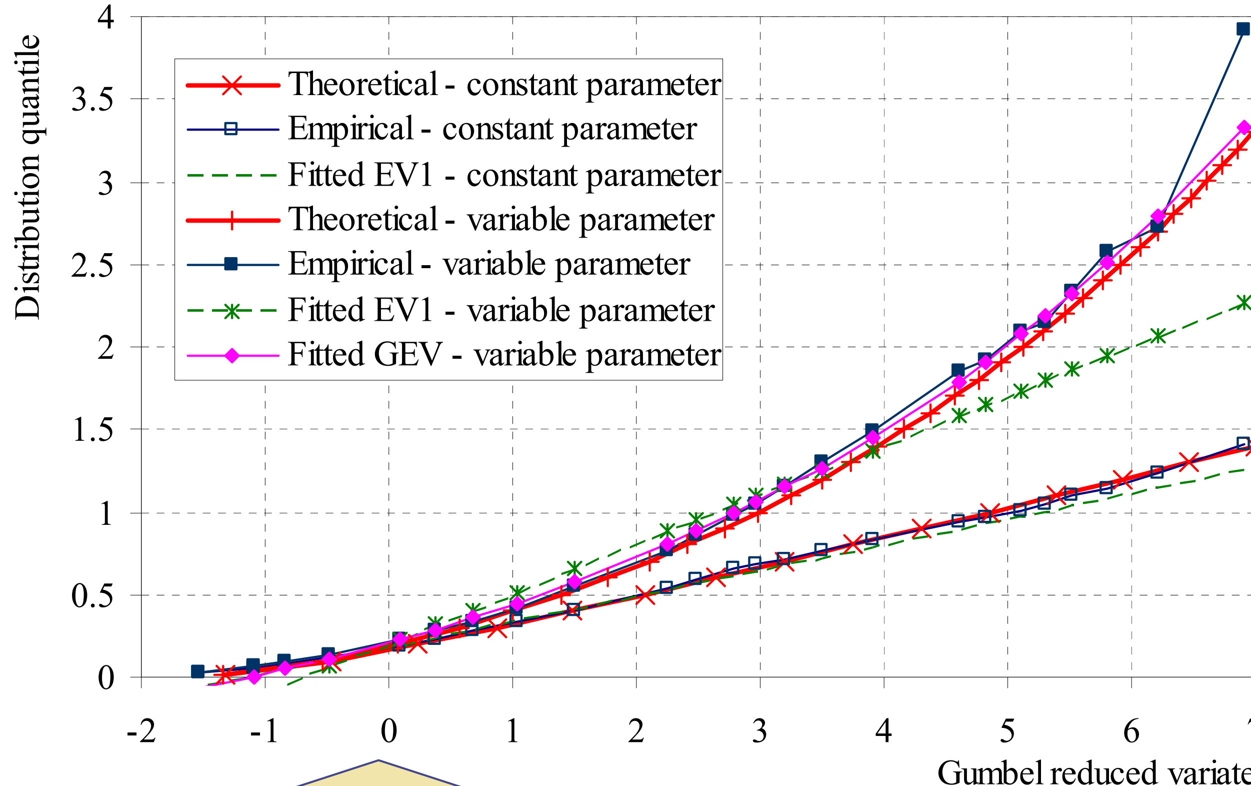
Asymptote:

EV2 (Fréchet distribution )




Convergence of distribution of maxima for gamma parent distribution with shape parameter $\theta = 0.5$ and scale parameter randomly varying following a gamma distribution with $\tau = 3$ and $\beta = 1$ (Fréchet probability plot)

# Numerical example of the shift of domain of attraction



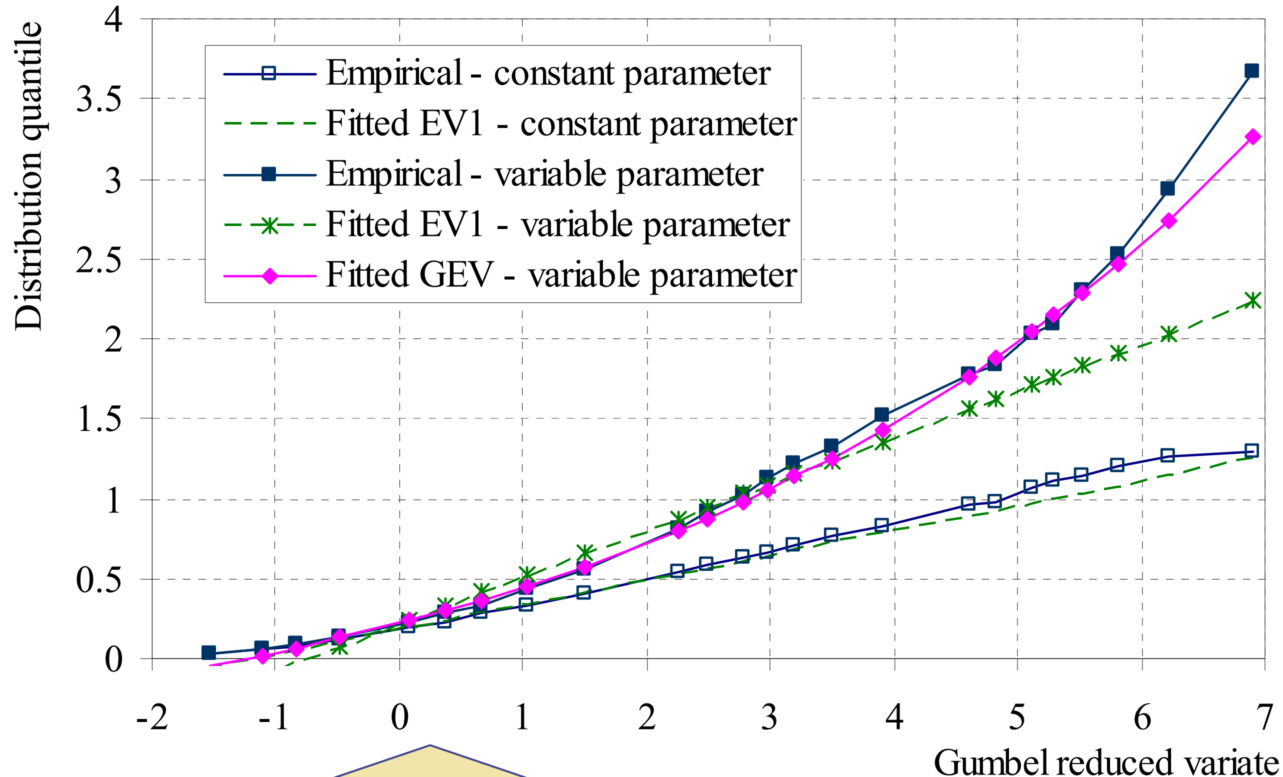

Exact distribution function of maxima $H_5(x)$ for gamma parent distribution with shape parameter $\theta = 0.5$ and scale parameter either:

- constant $\alpha = 5$
- randomly varying with gamma distribution with shape parameter $\tau = 5$ and scale parameter $\beta = 1$

Also plotted:

- empirical distribution functions from synthesised series of length 4000
- fitted to these series EV1 and EV2 distribution functions

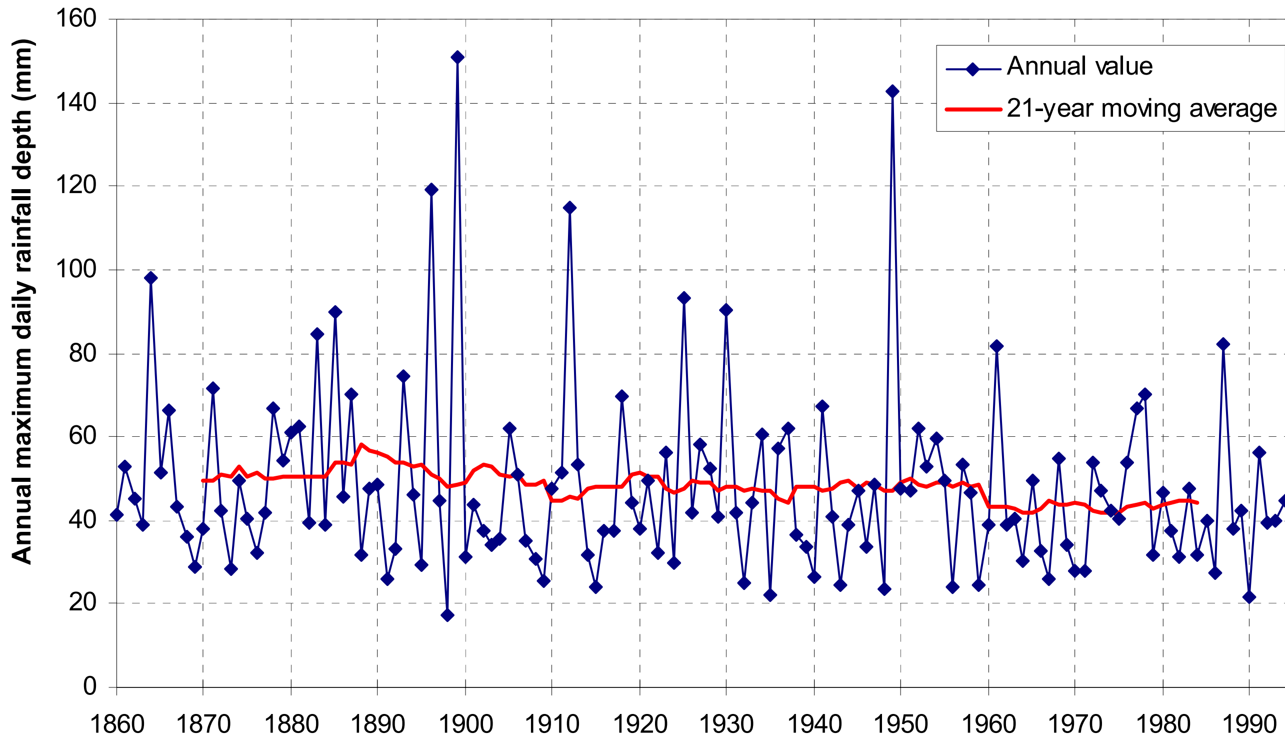
# A second numerical example of the shift of domain of attraction




Empirical distribution functions of maxima $H_5(x)$ and fitted EV1 and EV2 distribution functions, as they result from synthesised series of length 4000 assuming gamma parent distribution with shape parameter $\theta = 0.5$ and scale parameter either:

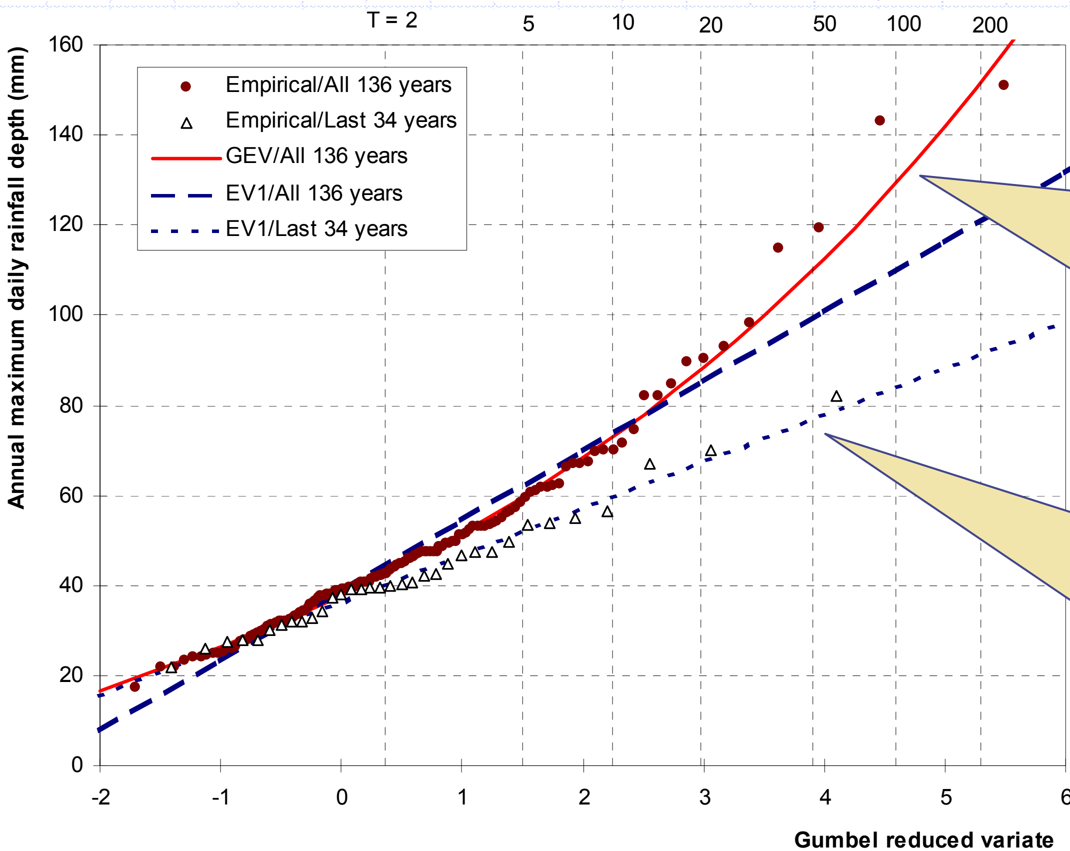
- constant $\alpha = 5$
- shifting at random between the values $\alpha_1 = 2$ and $\alpha_2 = 6$ with probabilities 0.25 and 0.75, respectively

# Empirical study: An example based on a long record




Time series of the annual maximum daily rainfall depth at Athens, Greece (station of the National Observatory of Athens, 1860-1995, 136 years; data analysed by Koutsoyiannis and Baloutsos, 2000)

# Demonstration of the effect of the record length




Distribution of the complete series (136 years)

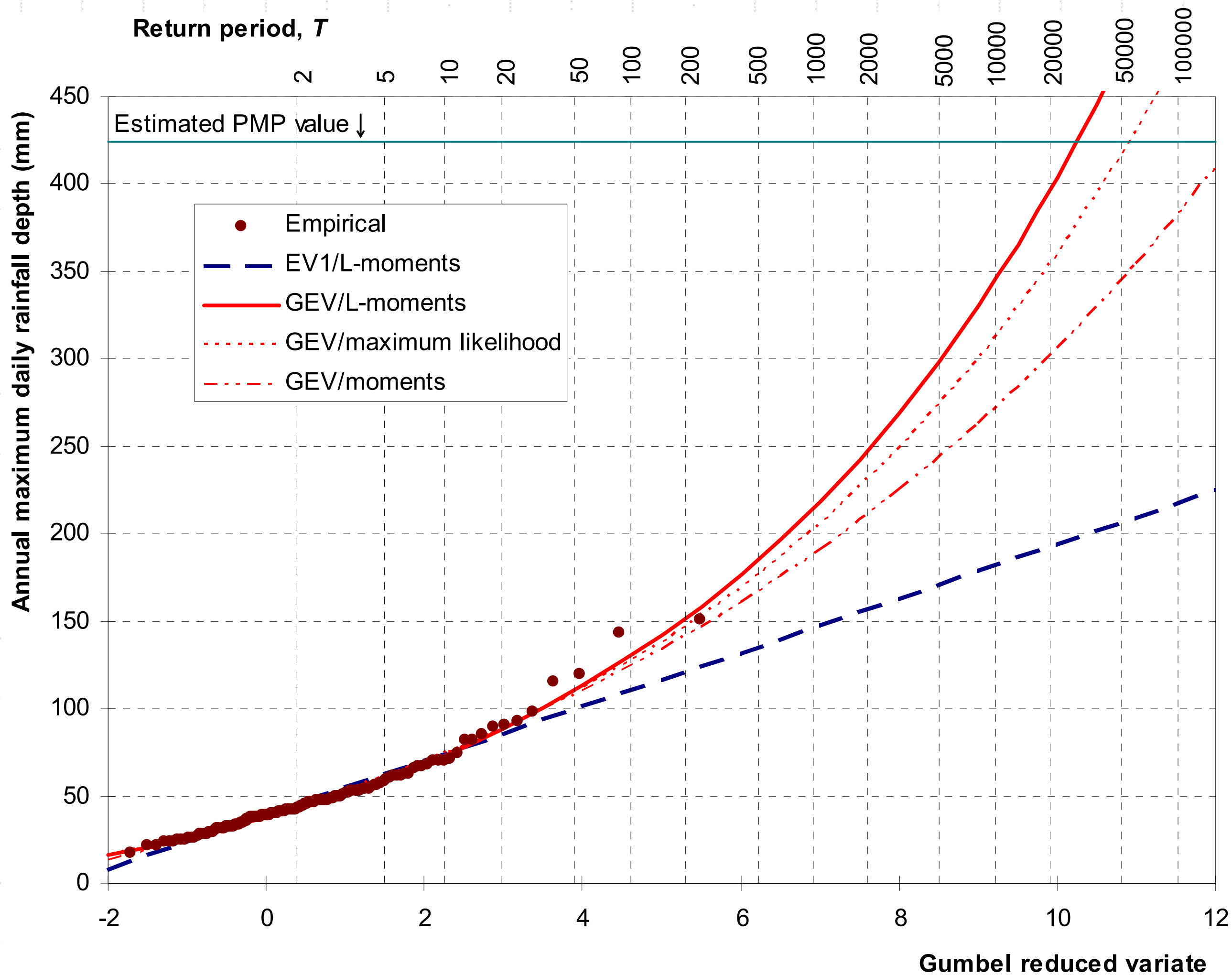
EV1 distribution rejected ($\alpha$ = 0.2%)

EV2 distribution ($\kappa$ = 0.185) not rejected

Distribution of a sub-series corresponding to the fourth quarter (last 34 years) of the record length

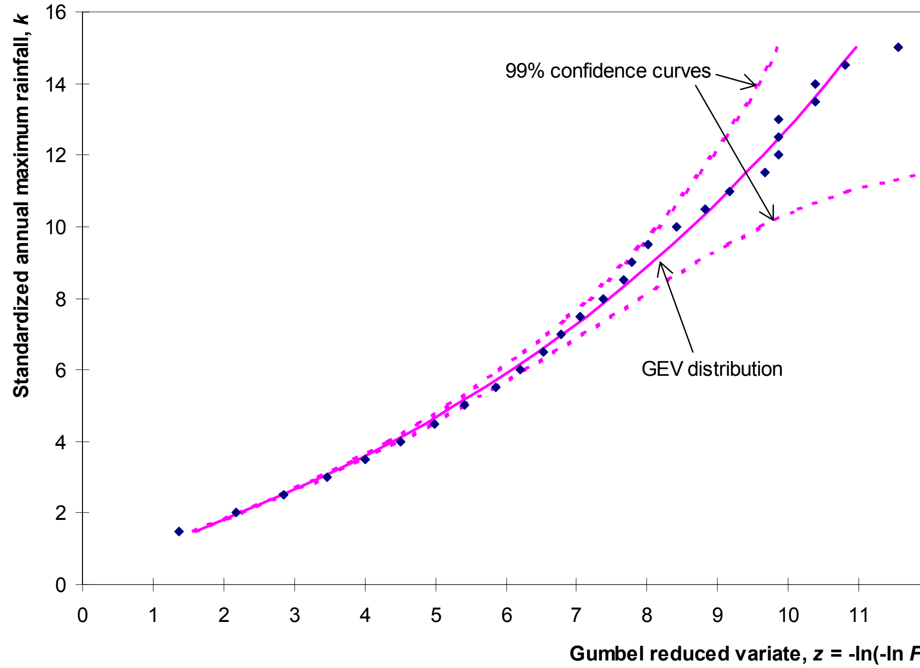
EV1 distribution not rejected

# Demonstration of the differences of EV1 and EV2 estimates for high return periods



# Recovery from the high-risk estimations of the Gumbel distribution

1. Assume EV2 rather than EV1
2. To increase the record length so as to estimate the shape parameter $\kappa$ more accurately, "substitute space for time", i.e., incorporate in the analysis information from other rainfall data sets
3. Utilise results of analyses of global data sets, e.g. Hershfield's (1961) data set comprising 95 000 station-years. According to a later study by Koutsoyiannis (1999) this results in:
   average $\kappa = 0.13$ or
   $\kappa = \max(0.183 - 0.00049\,\mu, 0)$




Gumbel probability plots of the empirical and GEV distribution functions of standardised rainfall depth $k$ for Hershfield's (1961) data set (from Koutsoyiannis, 1999)

# Conclusions

- The Gumbel distribution should be avoided when studying hydrological extremes because it may underestimate seriously the largest extreme rainfall or discharge amounts
- The theoretical and empirical reasons that made the Gumbel distribution prevail in hydrological studies may be not valid
- The three-parameter EV2 distribution (GEV bounded from the left) is a better alternative