Analyzing Extreme Rain

Jürgen Grieser [1]

July 12, 2012

[1] j.grieser@gmx.de, www.juergen-grieser.de

Contents

Preface

Dr. Bruno Rudolf was the head of the Global Precipitation Climatology Centre (GPCC) for more than 15 years when he recommended that I should give up trying to work as a meteorologist in 2005. He asked me to privately publish wrong information about my work in the GPCC and not to resist since "It will be worse for you if you resist". At that time I was 41 years old, held a phd in meteorology, co-authored the Nobel-price winning fourth assessment report of intergovernmental Panel on Climate Change (IPCC) and was employed by the GPCC for four years as a research scientist.

Bruno Rudolf was right. He could afford lying for years (and still does) and even got promoted to become a division director with the German Met Service (Deutscher Wetterdienst, DWD).

I had to give up my profession as a result of just saying and proofing that they are lying. The DWD claimed that they have checked the proofs and that I am lying, which I know I don't and can prove. However, no governmental authority wanted to investigate and watch the proof. The Federal Ministry for Research even claimed that I am wrong, although they avoided getting hold of the proof. Lawyers recommended not to sue the DWD since they are well protected by the state. So I had something like no chance at court. On the other hand I could not resist from saying the truth which I consider one of the duties of holding a phd in science. Finally the DWD sued me for being offensive and publishing wrong statements. At court it was again not a question whether I was telling the truth or not. That there was simple proof that I was right was not a matter of discussion. But since Dr. Bruno Rudolf had the right to feel offended by me telling the truth I was asked to stop doing so and to pay for my lawyer.

So I don't work anymore as a meteorologist. I turned from a scientist to a hobby-meteorologist and a spare-time researcher. This did, however, not alter my interest in meteorology in general and precipitation specifically.

Furthermore, having a decent job as a risk modeller with Risk Management Solutions in London keeps me in touch with some meteorological aspects. In this respect Bruno Rudolf did not meet one of his goals. I am not completely out of business and he and his folks always have to fear that I blame them of lying whenever they climb a stage to cheat the public or the scientific community. [1]

[1] For more than half a decade you find the statement 'The VASClimO project developed an observational database (Europe and Global) for statistical analyses with regard to climate variability on a decadal to centennial time scale' at http:gpcc.dwd.de. Two post docs worked for 5 years on the development of this database. At least this was told to the federal administration supporting this project with about EU100,000 a year. So you might just ask the GPCC about this database. What is in it? Who has ever seen/used it? Can you use it? This simple test will show you how arbitrary the public

So, having less time and less access to data than the GPCC experts I did some thinking, data analysis and modelling in order to better understand extreme rain. This text aims for a couple of reasons. Firstly, writing things down deepens understanding. Secondly, getting older means that I forget things easier so instead of keeping them in mind it is safer to keep them in a file. Thirdly, since I travel a lot, my web-page is the simplest one location to find my text. Fourthly, having this text at a public location ensures that I can - if asked - just point to it. Finally, if you read it and find something of which you think it is outer rubbish (or, less likely, you might find it interesting and/or helpful) you can tell me and I can learn from the discussion.

I have to admit it: I wrote this text for myself. Nevertheless, I hope that this text finds some readers and that they do not hesitate to inform me about their views, ideas and the errors they find at j.grieser@gmx.de.

Jürgen Grieser, Bonn, June 2012

statements of the Global Precipitation Climatology Centre are.

Chapter 1

Introduction

Rain is fascinating. It feeds the rivers with water, makes flowers blossom and let crops grow. Failure of rain let people suffer - even starving in parts of the world. Extreme rain, on the other hand can be very powerful and destructive. Of course, it is not the rain itself which destroys (unless it comes together with strong winds or falls as hail stones) but the accumulation of quite some rain water over a limited area which then leads to floods and flash floods. Therefore one of the most interesting questions in applied meteorology may be: How likely is a certain rain amount over a certain period of time and a certain area. Others than meteorologists have to think about the impact this might have on human-made and natural structures.

Engineers, architects, designers of flood defenses and political decision makers have to take extreme rain into account. They need estimates of the probabilities of extreme rain of different durations and magnitudes.

This text is not comprehensive but reflects some basics and my experience in statistical analysis and modelling of extreme rain.

The first chapter provides some numbers about water in the atmosphere and rain as a frame for the later chapters. In the second worldwide observed extreme rains of different durations are provided and a very interesting behaviour (scaling) is discussed.

The third chapter discusses various attempts to model extreme rain from physical considerations. Some very interesting conceptual results follow from this exercise. It gives hints on the shape of the intensity distribution of extreme rain. It determines which meteorological variables are important drivers of extreme rain. But the major result, however, may be that physical considerations do not solve the question on what the maximum possible rain at a location and for a given duration can be. Deterministic physical approaches do not yield a maximum possible precipitation.

The fourth chapter discusses a very useful and widely used tool: The Intensity-Duration-Frequency plots (IDF plots). A more general discussion of the scaling found in chapter 2 is given. Statistical features of rain intensities over different durations are linked in a peculiar way which provides strong tools for the estimation of extreme rain during different durations.

The classical tool investigating extreme rain is extreme value theory. It is discussed and applied in chapter 5. The strong impact of the seasonal cycle is shown. Maybe it

is the only really important information of this text, that ignoring the seasonal cycle overestimates return periods considerably.

Extreme value theory applied to precipitation has one disadvantage. According to the results arbitrary high rain amounts are possible everywhere. They might become very unlikely, though, but not impossible. This is not satisfying since one may assume that there is an upper limit, i.e. that in reality there is a maximum possible precipitation. Chapter 3 showed that we were not capable of determining it from physical considerations. Therefore a probable maximum precipitation PMP was defined as early as 1959. In chapter 7 the different methods of estimating PMP are discussed, their weaknesses are revealed and a definition in line with the value at risk (VaR) in the insurance business is provided. It might be the most interesting point of this section that given the rain observations of Dresden (Germany) from 1917 to 2001, the likelihood of an Elbe-flooding rain within 95 years was about 15%. This is not that low.

Chapter 8 is a wild conglomeration of different things that I came across. It is therefore named 'This and That', and shortly discusses how monthly rain observations can be transformed to be close to Gaussian distributed. It shortly discusses self-organized criticality (SOC) as a conceptual model of extreme rain, the standardized precipitation index (SPI) as well as some minor points.

Chapter 2

Water in the Atmosphere

The discussion of this section follows closely to Pruppacher and Klett [15]. Water enters the atmosphere due to evaporation and sublimation from the ground. It is transported through motion on various scales and leaves the atmosphere intermittently due to precipitation. The atmospheric water content is about $1.3 \cdot 10^{16} kg$ ($1300 km^3$) which is about $1 \cdot 10^{-5}$ of the surface water storage of the earth. The mixing ratio of water in the atmosphere decreases with height. This is not only due to the fact that the source and sink is at the ground but also that the temperature decreases with height and thus, according to the Clausius-Clapeyron Equation, less water can be stored in higher altitudes. Water is hardly transported through the tropopause. The stratosphere is relative dry. A temperature of $-60°C$ at $40 hPa$ means a saturation mixing ratio of $2.75 \cdot 10^{-3}$ according to the Magnus formula, while the observed mixing ratio at that altitude is about $2 \cdot 10^{-6}$, which means that the relative humidity is extremely low.

The annual average global rain is about $1m$ which, by the way, means that the earth surface looses about 80 W/m^2 due to evaporation. Given the earth surface ($4 \cdot \pi R_E^2$, with the earth radius $R_E = 6.37 \cdot 10^6 m$) the total amount of rain water is $5 \cdot 10^{17} kg/year$. The global human fresh water demand in 1995 was $3906 km^3 = 3.906 \cdot 10^{15} m^3$ and is expected to increase by 50% until 2025. Therefore the freshwater supply by precipitation is only about 100 times higher than the demand.

The atmospheric amount of water ($1.3 \cdot 10^{16} kg$) equals a surface water layer of $25 mm$. A residence time of water in the atmosphere can easily be defined as

$$T = \frac{\text{Storage}}{\text{Flux}} = \frac{25 mm}{1000 mm/\text{year}} \approx \frac{1}{40}\text{years} \approx 9\text{days}. \tag{2.1}$$

Assuming a liquid water content of clouds of about $0.5 g/m^3$ and an average cloud depth of $3 km$ we get an average cloud water content per square meter of $1.5 kg$. Assuming further that half of the globe is covered with clouds, the liquid water content of the atmosphere becomes about $0.75 kg/m^2$ or $3.8 \cdot 10^{14} kg$ ($380 km^3$) meaning that 1/30 of the atmospheric water is liquid. The residence time of cloud water is therefore 9 days divided by 30, i.e. 7 hours, meaning that the liquid water in the atmosphere is replaced about 1200 times a year. Only 10% of the clouds are raining and only 50% of the rain produced reaches the ground (the rest is re-evaporated between cloud base and ground). Thus the average duration of precipitation is $7 hrs/10 \cdot 0.5 = 20 min$. Assuming that all

the liquid water falls out within this 20 min the average rain rate becomes $1.5mm/hr$. This is in reasonable agreement with observations of single cells. A cloud lifetime is about 30 to 60 minutes and a liquid water residence time is 7 hours. This means that the number of condensation-evaporation cycles required for a precipitating system is of the order 10.

Chapter 3

Observed Extreme Precipitation

The rainiest location on earth is Mt. Waialeale (Hawaii) with $11.84m$ mean annual precipitation. Precipitation shows high spatial variability. And the rain gauge at station Kekala which is about $20km$ away from the summit of Mt. Waialeale gets only $0.553m$ rain a year - a factor 21 less.

Keeping in mind that average annual rain can show such high spatial variability we might expect that extreme rain is a very difficult variable. So the first thing to do is to look at the observations. Fig. 3.1 shows the highest ever recorded rain depths over rain duration. The data are taken from [21].

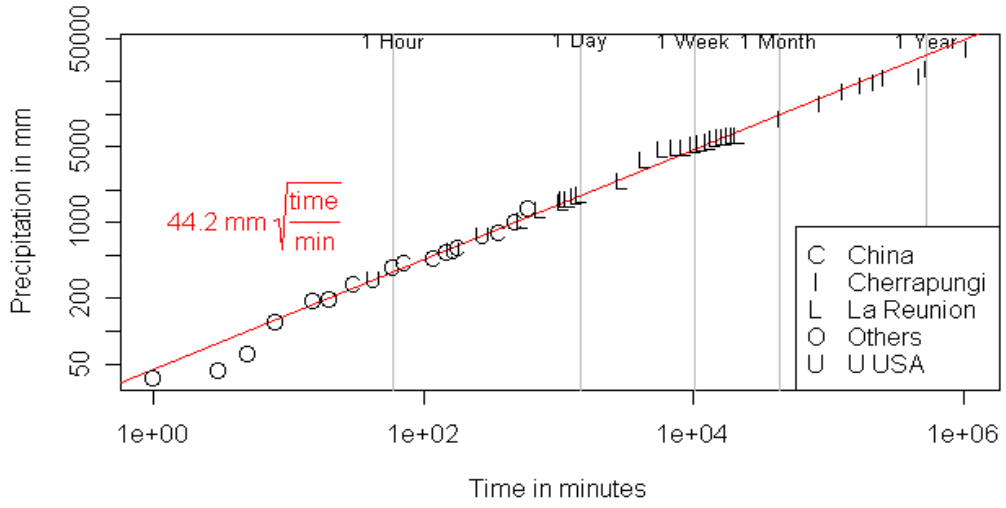

Figure 3.1: Highest ever observed (and recorded) rain depths for different durations worldwide [21].

Fig. 3.1 shows that the highest recorded rain depth R scale over duration D as

$$R = 44.217mm\,(D/1min)^{0.5064}. \tag{3.1}$$

The attributed maximum rain rates R/D scale therefore with $D^{-1/2}$.

The scaling of extreme rain observations worldwide should be surprising since rain extremes on different temporal scales (durations) and regions should be triggered and driven by different mechanisms. And indeed for durations of more than a month all extremes happen at the same location (Cherapunji, India) and are triggered by monsoon hitting a mountain range. For durations of 12 hours to 2 weeks all registered extremes occurred on La Reunion, a volcanic island in the Indian ocean. And they are all linked to tropical storms. So we had monsoon in one time scale, tropical storms in another time scale and mountains in both. Rain extremes below a time scale of 12 hours may be due to synoptic rain at fronts, mesoscale systems (especially when locked by mountain ranges) and - for the shortest time scales - local severe convection. Mountains seem to become less important on these time scales. The 15min record of nearly 200mm is observed at the coast of Jamaica. On this short time scales mountains seem to become negligible compared to the strong updrafts that can be produced by convection. Given the different forcing mechanisms on the different time scales it seems absolutely surprising that the scaling of eq. (3.1) describes the observations so well. One possible explanation of this scaling is provided by Galmarini et al. 2004 [6]. In order to get a better understanding of rain, the next section deals with the question: what produces rain?

Chapter 4

Modelling Rain

4.1 What produces Rain?

When an air mass becomes supersaturated, condensation starts and water vapor is transformed to cloud drops. These cloud drops are too small to overcome the airs viscosity and thus do not fall to the ground. However, when drops or ice crystals get into contact with each other they stick together and build larger drops or ice crystals. In a certain range of particle size this process is more efficient than growing due to condensation. Most efficient however is the droplet growth of already large droplets when they fall through air with small droplets and finally are becoming rain drops. So the whole process happens in stages which finally can lead to rain when a critical state is reached. Here we don't discuss cloud microphysics, which is a very complicated field of parametrization of processes. Self-organized criticality (SOC) is sometimes proposed to be the formal frame to model rain as an intermittent process (see e.g. [13]). SOC is shortly discussed in section 9.2. In the next section a conceptual model is discussed which can explain the form of observed rain intensity distributions despite its simplicity.

4.2 A Conceptual Model

Rain is produced by clouds which consist of moist rising air. We want to get an idea of the total rain falling out of a raining cloud. This can be expressed as the average rain rate times the life time of the cloud. Air rises due to a convergence in column integrated horizontal mass flux, $-\nabla \cdot (\rho \vec{v})$, where ρ is the air density and $\vec{v}$ is the horizontal wind vector, between the surface and the top of the moist layer z_m. The average precipitation rate R is proportional to the column integral over the convergence of the horizontal water mass flux $-\nabla \cdot (q\rho\vec{v})$ (see Fig. 4.1), with the specific humidity q and a precipitation efficiency κ and thus can be expressed as

$$R \approx -\kappa \overline{\int_0^{z_m} \nabla \cdot (q\rho\vec{v})\, dz} \tag{4.1}$$

where the overline denotes temporal average. In this approach all cloud-microphysical processes are combined and discribed by the precipitation efficiency κ. We assume

further that evaporation from the surface can be neglected during the life cycle of a cloud and that the specific humidity is constant within the moist layer. Using the inelastic continuity equation directly leads to

$$R \approx -\overline{\kappa q \int_0^{z_m} \frac{\partial(\rho w)}{\partial z} dz} = \overline{\kappa\, q\, \rho w_m} \tag{4.2}$$

with the updraft speed w and index m indicating the top of the moist layer.

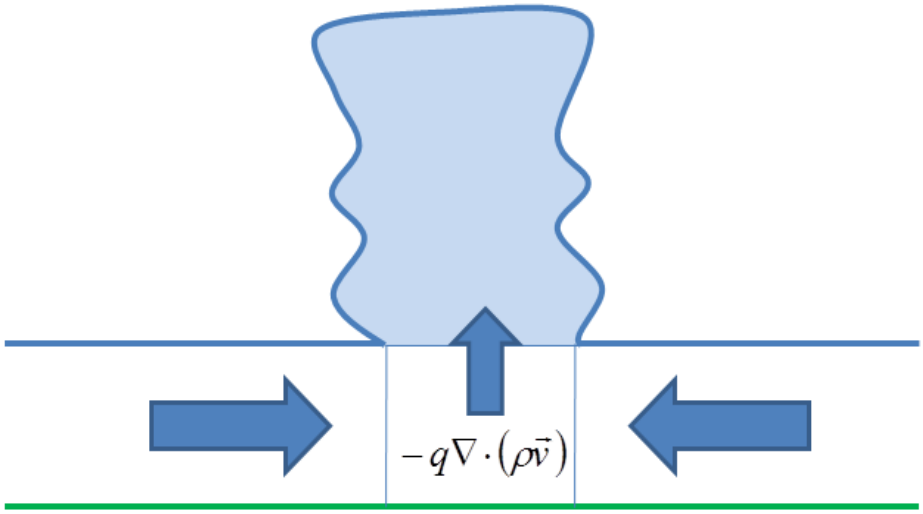

Figure 4.1: A conceptual model for rain intensity as a function of the horizontal convergence of moist air.

This conceptional model is published 1978 by Stevens and Lindzen [17]. Wilson and Tsuomi (2005) [18] assumed that the three variables (column-integrated horizontal mass flux convergence $-\nabla \cdot (\rho\vec{v})$, average specific humidity q, and precipitation efficiency κ) are independent and normally distributed and further simplified

$$R \approx -\overline{\kappa q \rho w_m} = \mathcal{N}(\kappa)\,\mathcal{N}(q)\,\mathcal{N}(\rho w_m) \tag{4.3}$$

where $\mathcal{N}$ means the normal (Gaussian) distribution. A product of three random variables is large if all three variables are large. If we assume all three variables to be equally large they have to have the magnitude $R^{1/3}$ each. Therefore, for the tail of the distribution of R, each of the normal variables to be multiplied has the form

$$\mathcal{N}(\kappa) \approx \mathcal{N}(q) \approx \mathcal{N}(\rho w_m) \approx \mathcal{N}(R^{1/3}) \propto \exp\left(-\left(R^{1/3}\right)^2\right). \tag{4.4}$$

The cumulative probability distribution of R, i.e. $P(R > r)$ follows from integration and approximation for large r to be of the form

$$P(R > r) \propto \exp\left(-\left(\frac{r}{R_0}\right)^c\right) \tag{4.5}$$

with the scale parameter R_0 and the exponent $c = 2/3$. The result of this discussion is that, under the assumptions made, the mean precipitation rate (and thus the precipitation sum of extreme rain events) should be distributed according to the stretched exponential distribution provided in eq. (4.5). Furthermore, we expect the exponent to be 2/3. Such events are individual clouds. Therefore in order to test this hypothesis, rain data of individual precipitating clouds are necessary. However, Wilson and Tsuomi [18] applied this method to daily data of many stations worldwide and found a best-fit exponent c varying between 0.2 and 1.

We now apply this model to test its general applicability. The first question is how high the value of a product of three Gaussians has to be to make a stretched exponential a good approximation. To test this we multiply 1 million realizations of three independent Gaussians and set the threshold to 0. The result of a best fit of a stretched exponential to the realizations is shown in Fig. 4.2. We see that even for the threshold zero, which is exceeded by half of the realizations, we get a very good fit of the stretched exponential. The experiment is repeated 20 times with 10 million realizations each. The mean exponent derived from this is 0.64 with a standard deviation of 0.16. We see that estimates of the exponent from data vary considerably even for long records.

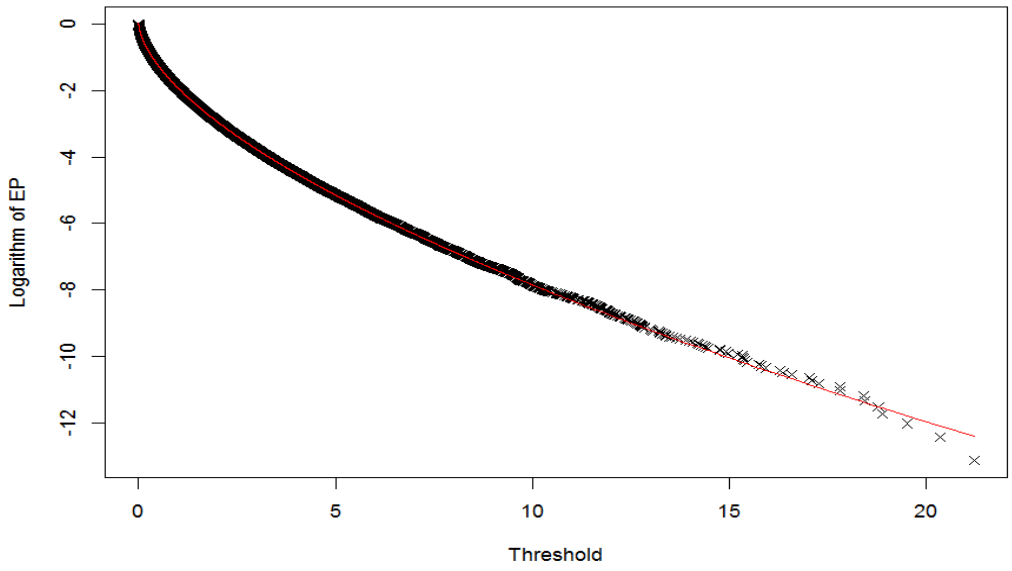

Figure 4.2: Fit of a stretched exponential distribution to positive realizations of a product of three independent Gaussian variables (1 million realizations each).

We now test whether the stretched exponential distribution is capable of describing the frequency distribution of daily rain observations of Dresden (see appendix A for details about the data). We calculate the coefficients R_0 and c by non-linear least square fits of the transformed empirical cumulative density function of all values $r > r_0$ to the term $\exp\left[\left(-\frac{r}{R_0}\right)^c\right]$ as

$$1 - \hat{F}(r > r_0) = \exp\left(-\frac{(r - r_0)}{R_0}\right)^c \tag{4.6}$$

which follows directly from eq. (4.5). $\hat{F}$ is estimated by the rank numbers n of the N sorted observations as $\hat{F} = \frac{n}{N+1}$. If the hypothesis of a stretched exponential is true for $r > r_0$ we expect seeing a straight line with intercept zero and slope 1 if $\log\left(1 - \hat{F}(r > r_0)\right)$ is drawn against $\left(-\frac{(r-r_0)}{R_0}\right)^c$. Results are shown in Fig. 4.3 for the observations until 2001. The straight-line fits look very good though slope and intercept significantly deviate from the expected ones in all seasons except fall.

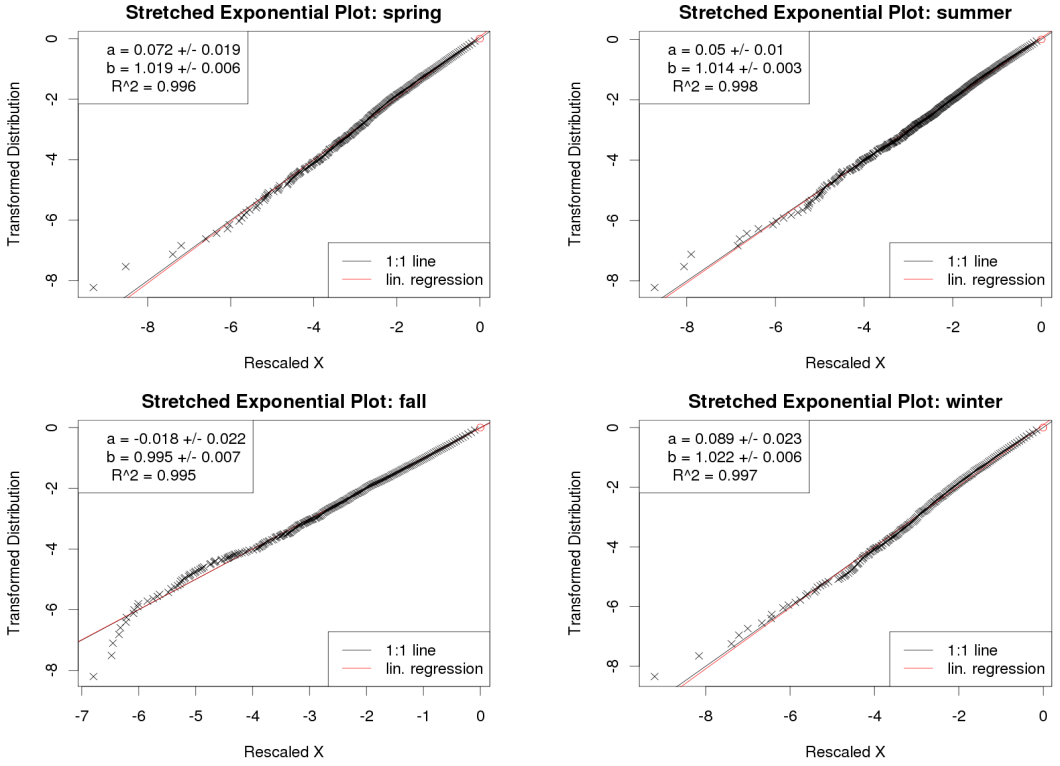

Figure 4.3: 'Stretched-Exponential' plot of the intensity distribution of observed daily rain in Dresden from 1917 to 2001 for spring (upper left), summer (upper right), fall (lower left) and winter (lower right).

The exceedance probabilities that follow from the best-fit stretched exponentials are shown in Fig. 4.4 for the observations until 2001. Including the exceptionally high rain observation in August 2002 does not alter the results qualitatively. We see that the fits are very good on a seasonal scale. The best fit for the annual data looks very similar to the one for summer and is therefore not shown here.

We see that the stretched exponential slightly underestimates the exceedance probability of observed extraordinary large rain events in all seasons except for fall. We can now use the best fit stretched exponentials to estimate return periods for the highest observed daily rain in Dresden, i.e. 158mm/day. If we do that for the best fit to the whole data set we get about 960 years. If we do that for the four seasons we get 480 summers, 110E3 springs, 2447E3 falls or 10E6 winters. This shows that such a high rain should occur much more likely in summer than in winter. It also shows that mixing the rain of different seasons leads to considerable underestimation of the risk (factor 2) compared to only analyzing summer data. We should keep in mind that daily rain sums of different

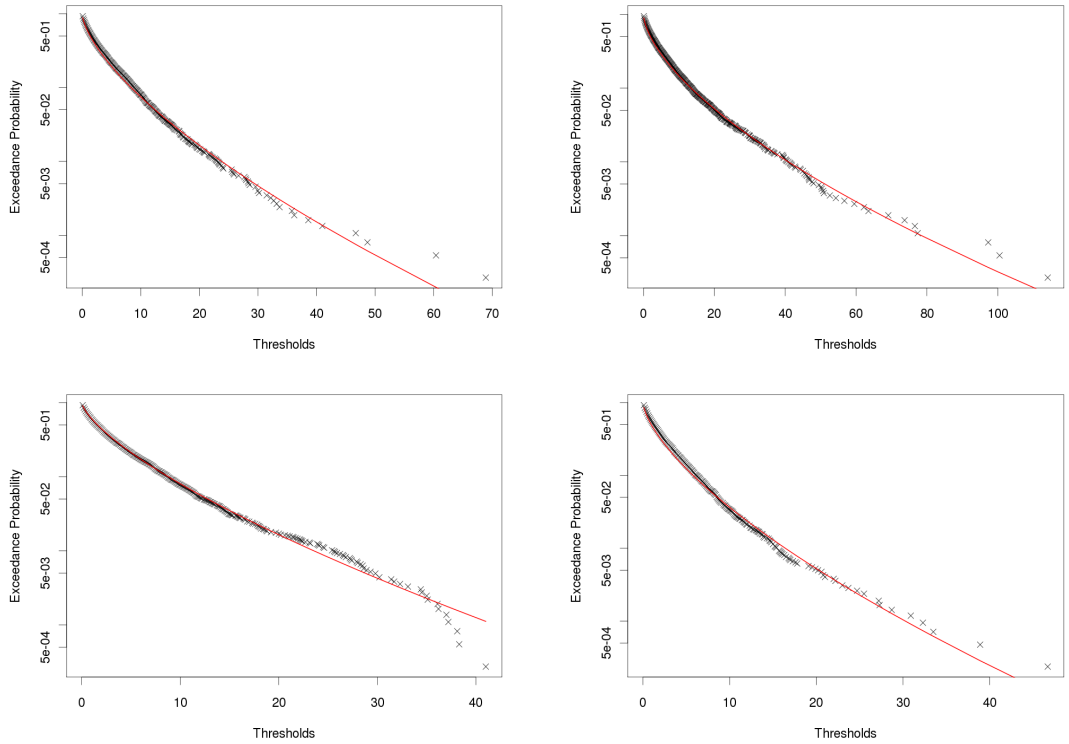

Figure 4.4: Exceedance probability plots of fit of a stretched exponential distribution to the intensity distribution of observed daily rain in Dresden from 1917 to 2001 for spring (upper left), summer (upper right), fall (lower left) and winter (lower right).

seasons are realizations of different random variables and should not be mixed when calculating return periods. Table 4.1 provides the parameter values of the best fits and their standard errors. The latter are small compared to the best estimates. This is especially impressive for the stretching exponent.

So far we have used the threshold zero for precipitation, i.e. we used all precipitation observations of days where at least 0.1mm rain was observed. We now ask how the exponent of the stretched exponential varies with the choice of the threshold. Results for thresholds between 0 and 20mm are presented in Fig. 4.5.

It might be surprising that the estimated stretching exponents are closer to 2/3 for lower thresholds than for larger ones. The reason might be that for larger thresholds the tail in the observed data becomes more important. We already saw in Fig. 4.4 that the curvature from the stretched exponential (which determines the deviation from the pure exponential which would be a straight line) is strongest for small precipitation values. For high thresholds and fall data we even get exponents larger unity representing a compressed exponential tail as it was visible in Fig. 4.4.

The fact that the stretching exponent significantly deviates from 1 means that a pure exponential distribution cannot describe severe rain events. This has implications for the extreme value analysis performed later. It also means that some common distributions with exponential tails like the Γ distribution cannot describe the whole intensity distribution of daily rain. However, it can easily be shown that the stretched exponen-

Table 4.1: Coefficients of the fit of a stretched exponential return period $\tau = \exp\left[-\left(\frac{r}{R_0}\right)^c\right]$ to daily rain observations in Dresden from 1917 to 2001.

Season	R_0	c
all	2.053	0.586
stdErr	0.02554	0.002636
spring	2.375	0.662
stdErr	0.035467	0.004432
summer	3.524	0.623
stdErr	0.031389	0.002439
fall	2.677	0.702
stdErr	0.046295	0.005841
winter	1.643	0.663
stdErr	0.029362	0.005131

tial and the Weibull distribution have a common tail. The difference is that the Weibull distribution can be provided analytically for the whole range of the variable. This is possible since the pre-factor to the stretched exponential in the density function is not a constant but itself a function of the independent variable. The Weibull distribution is one of the three limit distributions of peaks over threshold and thus will cross our way later again.

Since the Weibull distribution has a stretched-exponential tail it might be a good candidate for describing the whole range of the precipitation intensity distribution. Therefore, in appendix B a Weibull distribution is fit to the daily rain observations of Dresden and different fitting methods are compared.

We end this discussion here. In the following section we approach the question of modeling extreme rain by estimating how much water fits into a column of air.

4.2.1 How much water fits into air?

The total amount of water vapor S_x (with units $\frac{kg}{m^2} = mm$) that fits into a column of air is defined as the integral over the saturation water vapor density (with units kg/m^3) $s^* = q^*\,\rho$ as

$$S_x = \int_0^\infty q^*\,\rho\,dz = -\frac{1}{g}\int_{p_0}^0 q^*\,dp \tag{4.7}$$

with the saturation specific humidity q^* as the ratio of water vapor mass to air mass [1]. This ratio can be replaced by the product of the ratio of the *mol* mass of water vapor to air ($\approx .622$) and the ratio of saturation vapor pressure e^* and air pressure p as

$$q^* = .622\frac{e*}{p} \tag{4.8}$$

and we get

$$\frac{s^*}{s_0^*} = \frac{\rho\,q*}{\rho_0\,q_0^*} = \frac{\rho\,p_0\,e*}{\rho_0\,p\,e_0^*} \tag{4.9}$$

[1] The right hand side of eq. 4.7 follows from the hydrostatic equation

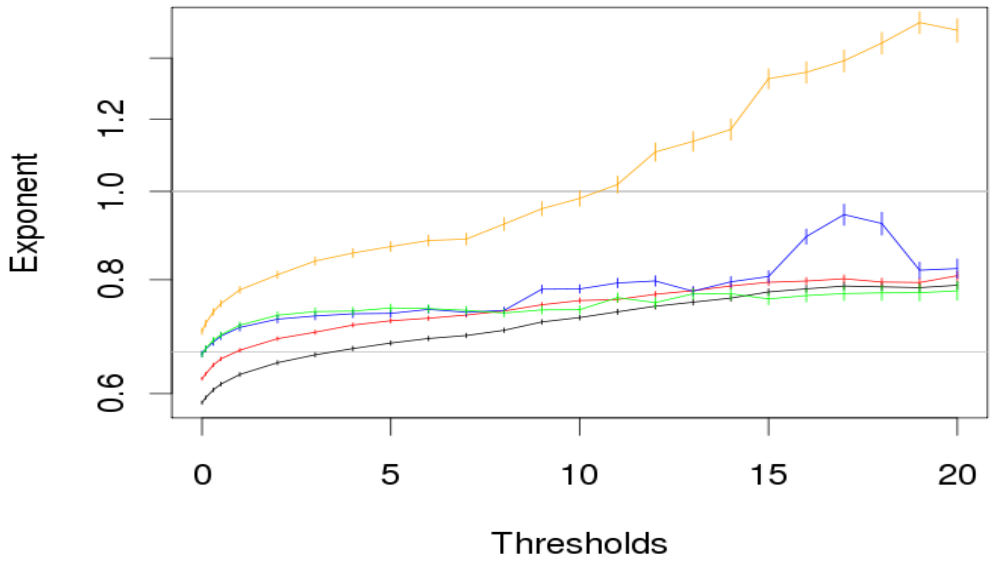

Figure 4.5: Dependence of the exponent of the stretched exponential as function of rain threshold between 0 and 10mm for daily rain observations in Dresden (1917-2001) if all data are used (black line), for spring (green line), summer (red line), fall (orange line), and winter (blue line). Standard errors of the Maximum-Likelihood fit are indicated as vertical lines.

where the index 0 stands for surface height. Assuming that both, p and ρ, decrease exponentially with height with the same rate, we get

$$s^* \approx s_0^* \frac{e^*}{e_0^*}. \tag{4.10}$$

Inserting the Magnus formula for e^* we get

$$s^* \approx s_0^* \exp\left(\frac{a(T_0+\gamma z)}{b+T_0+\gamma z} - \frac{aT_0}{b+T_0}\right). \tag{4.11}$$

where we have assumed a linear vertical temperature profile with a lapse rate γ. We further simplify $b+T_0+\gamma z \approx b+T_0$ and $b+T_0 \approx b$ to get a simple altitude dependency of s^* as

$$s^* = s_0^* \exp\left(\frac{a\ \gamma\ z}{b}\right) = s_0 \exp\left(\frac{-z}{H_m}\right) \tag{4.12}$$

with the e-folding height for atmospheric moisture

$$H_m = \frac{-b}{a\gamma}. \tag{4.13}$$

The saturation water vapor density at ground level $s_0^* = \rho_0\, q_0^* = \frac{\rho_0}{p_0}\, 0.622\, e_0^* = \frac{0.622}{R\,T_0} e_0^*$ depends only on the air temperature at the surface T_0 and the gas constant of air $R = 287 \frac{J}{kgK}$. Therefore we can write

$$s_0^*(T_0) = \frac{.622\ \ 6.112hPa}{R\ (T_0 + 273°C)} \exp\left(\frac{a\ T_0}{b+T_0}\right). \tag{4.14}$$

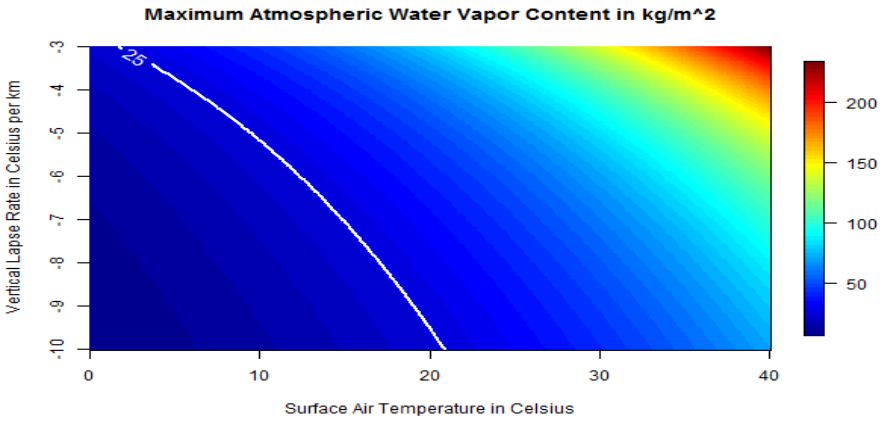

Figure 4.6: Maximum water vapor content of an atmospheric column in $\frac{kg}{m^2} = mm$ as a function of the vertical lapse rate and the near-surface temperature. The global average atmospheric water vapor content of $25mm$ is drawn as a white line.

Equation (4.14) reveals an inconsistency in our formulation so far. Within the Magnus formula temperatures have to be provided in $°C$ while outside they are absolute temperatures in K. This is why we have to add $273°C$ in order to use T_0 in $°C$ within eq. (4.14).

Equation (4.12) allows for the simple calculation of the saturation water vapor density in any arbitrary height within the atmosphere. So we can easily calculate the maximum water vapor content $S_{1,2}$ in kg/m^2 of an air column between two altitude levels z_1 and z_2 as

$$S_{z_1,z_2} = s_0^* \int_{z_1}^{z_2} \exp\left(\frac{-z}{H_m}\right) dz = s_0^* \, H_m \left\{\exp\left(\frac{-z_1}{H_m}\right) - \exp\left(\frac{-z_2}{H_m}\right)\right\} \tag{4.15}$$

or the maximum water vapor content of an entire air column as

$$S_X(T_0, \gamma) = s_0^*(T_0) \, H_m(\gamma). \tag{4.16}$$

The water vapor content above cloud-base level z_{cb} is then given by

$$S_{cb}(T_0, H_m, z_{cb}) = s_0^* \, H_m \exp\left(\frac{-z_{cb}}{H_m}\right). \tag{4.17}$$

The water vapor content below cloud-base level may be approximated by a constant mixing ratio represented by the one at the surface. With the surface height z_s we get

$$S_b = \int_{z_s}^{z_{cb}} s_0 dz = \int_{z_s}^{z_{cb}} q_0 \rho \, dz = \frac{-q_0}{g} \left(p_{cb} - p_0\right) \tag{4.18}$$

where we used the hydrostatic equation $dp = -g\rho \, dz$. Assuming the vertical temperature gradient below the cloud to be dry-adiabatic we finally get

$$S_b = \frac{q_0}{g} p_0 \left[1 - \exp\left(\frac{-2g\left(z_{cb} - z_s\right)}{RT_0\left(1 - \Gamma_d(z_{cb} - z_s)\right)}\right)\right]. \tag{4.19}$$

Since the lapse rate is assumed to be dry adiabatic below the cloud base but not so above, we also need to alter the T_0 used in eq. (4.17). Instead of using T_0 we need to use a T_0^* (see next section) with

$$T_0^* = T_0 - (z_{cb} - z_s)\,(\Gamma_d - \gamma). \tag{4.20}$$

Eq. (4.17) and (4.19) allow for the estimation of the maximum water vapor content of an entire atmosphere as a function of surface height z_s, surface air temperature T_0, cloud-base height z_{cb}, the average lapse rate within the cloud γ and the specific humidity at ground level q_0. In the limiting case of $z_{cb} = z_s = 0$ we get the results shown in Fig. 4.6.

4.3 How much rain can be produced above a cloud level?

As a first proxy one might assume that the precipitable water content (PWC) of an atmospheric column equals the maximum amount of rain water that can be produced within that column over time. As we saw in section 4.2 this is not exactly true. Rain is produced by upwind which entrains humid air into the air column in low levels and detrains dryer air out the air column in high levels. This means that the amount of rain that can be produced in an air column within a duration D can be larger than the water content of the column.

So how is water vapor transformed to rain water? Water vapor condenses when an air parcel is cooled below its dew point temperature. The small droplets of cloud water form clouds or fog. Droplets can grow by collision among each other to rain drops. So how can an air mass be cooled to trigger condensation? By updraft. Rising air expands due to lower air pressure in higher altitudes. Expansion leads to cooling and colder air can hold less water vapor. Thus lifting of saturated air produces cloud water. The local condensation rate $c = -ds/dt$ (with the units $kg/m^3/s$) can be expressed as

$$c = -\frac{ds}{dt} = -\frac{ds}{dz}\frac{dz}{dt} = -w\frac{ds}{dz} \tag{4.21}$$

which shows that the condensation rate is linked linearly to the updraft velocity w. Inserting eq.(4.12) yields

$$c = c(T_0, w, z, H_m) = \frac{w\ s_0}{H_m} \exp\left(\frac{-z}{H_m}\right). \tag{4.22}$$

Eq. (4.22) provides information about the amount of cloud water built per air volume and time interval as a function of altitude and updraft velocity. The condensation of cloud water in a given column of saturated air between 2 altitude levels z_1 and z_2 (expressed in $kg/m^2/s$) due to a constant updraft w can now be calculated as

$$C_{z_1,z_2} = \frac{w\ s_0}{H_m} \int_{z_1}^{z_2} \exp\left(\frac{-z}{H_m}\right) dz = w\ s_0 \left\{\exp\left(\frac{-z_1}{H_m}\right) - \exp\left(\frac{-z_2}{H_m}\right)\right\}. \tag{4.23}$$

For the condensation above and below a given level, C_{up} and C_{down} respectively, the following expressions are obvious

$$\begin{aligned} C_{up} &= w\ s_0 \exp\left(\tfrac{-z_{up}}{H_m}\right) \\ C_{down} &= w\ s_0 \left\{1 - \exp\left(\tfrac{-z_{down}}{H_m}\right)\right\}. \end{aligned} \tag{4.24}$$

Up to now, in this section it is assumed that the complete atmosphere is saturated and thus lifting yields to condensation in all altitudes. However, condensation occurs only above the cloud base $z = z_{cb}$. Condensation rates above this altitude can be calculated with eq. (4.24) if $s_0(T_0)$ is known. T_0, however, is the surface air temperature (at $z = z_s$) under the condition that condensation happens to occur in all altitudes of the atmosphere, i.e. that the saturation lapse rate γ_{sat} can be applied everywhere. Since usually $z_{cb} > z_s$, T_0 has to be replaced by an effective surface air temperature T^* which together with the dry-adiabatic lapse rate Γ_d leads to the same temperature at cloud base T_{cb} as it is in case of saturation lapse rate γ_{sat} and T_0.

Given

$$T_{cl} = T_0 + \gamma_{sat}\, z_{cl} = T^* + \Gamma_d\, z_{cl} \tag{4.25}$$

one gets

$$T^* = T_0 - (\Gamma_d - \gamma_{sat})\,(z_{cl} - z_s). \tag{4.26}$$

With $\Gamma_d = -1K/100m$ and $\gamma_{sat} \approx -.5K/100m$ the final equation

$$T^* = T_0 + .5K/100m\, z_{cl}. \tag{4.27}$$

follows.

The lower the cloud base, the higher are the rain rates produced. In order to estimate the maximum rain rate we now make a couple of further assumptions:

1. we assume a cloud base at $z_{cb} = 0$,
2. we assume that all condensed water falls out as rain, i.e. that there is no re-evaporation
3. we assume a constant updraft speed w within the whole cloud column, knowing that this is hardly possible with a cloud base at $z = 0$.

From these assumptions we get a maximum rain rate R in m/s of

$$R = w \cdot \frac{.622\ \ 611.2Pa}{R\,(T_0 + 273°C)} \exp\left(\frac{a\,T_0}{b + T_0}\right) \tag{4.28}$$

where we have used eq. (4.14) and eq. (4.24). In this formulation the maximum rain intensity is only a function of the surface air temperature and the column updraft speed.

Fig. 4.7 provides rain rates according to equation (4.28) and shows the maximum observed one-hour and one-day rain rate for comparison. It shows that the equation can be used to get realistic extreme rain amounts.

4.4 Orographic Rain

4.4.1 A Simple Model

On its windward side, orography can transfer horizontal wind into updraft. Therefore, rain can be produced or enhanced if humid air is advected towards a mountain range. In

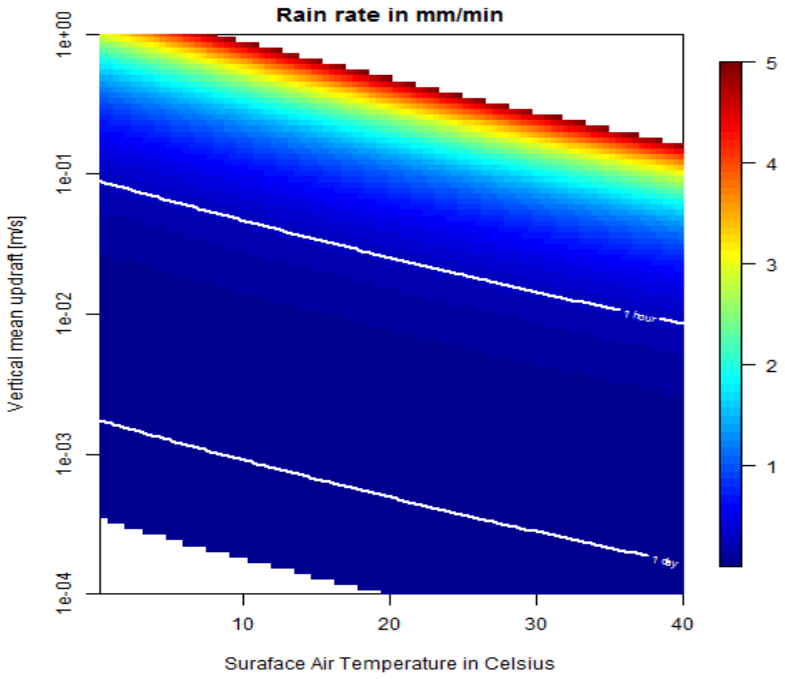

Figure 4.7: Maximum rain rate following a simple model of rain as a function of surface air temperature T_0 and updraft speed w. For comparison the highest ever observed one-hour rain rate and one-day rain rate are included as white lines.

order to model this we can estimate the updraft w as the dot product of the horizontal wind times the gradient of the orography

$$w = H(\vec{v_h} \cdot \nabla z_s) \tag{4.29}$$

where $H()$ denotes the Heaviside function ensuring that negative values of the argument are set to zero. Inserting eq. (4.29) into eq. (4.28) provides an estimate of the highest possible orographic rain rates as a function of horizontal wind speed and surface air temperature. If we know the slope $\frac{dz}{dx}$ in wind direction we can write eq. (4.29) as

$$w = |\vec{v_h}| \frac{dz}{dx} = |\vec{v_h}| \sin\alpha \tag{4.30}$$

where α is the angle of the slope of the orography in windward direction.

The maximum local rain rate follows as

$$R_{oro} = \frac{|\vec{v_h}|\, s_0(T_0) \sin\alpha}{H_m} \int_{max(z_{cb}, z_0)}^{z_t} \exp\left(\frac{-z}{H_m}\right) dz \tag{4.31}$$

where z_0 is the local surface height, z_{cb} and z_t are cloud base and cloud top height, respectively and T_0 is the surface air temperature.

Fig. 4.8 shows the rain rate at cloud base (in case rain is not advected) for a slope of $50km$ length, rising by $2.5km$. The cloud base is assumed to be $500m$. Note that for elevations higher than the cloud base rain rates reduce exponentially with height in the cloud. This result is bassed on a constant slope and a constant horizontal wind speed.

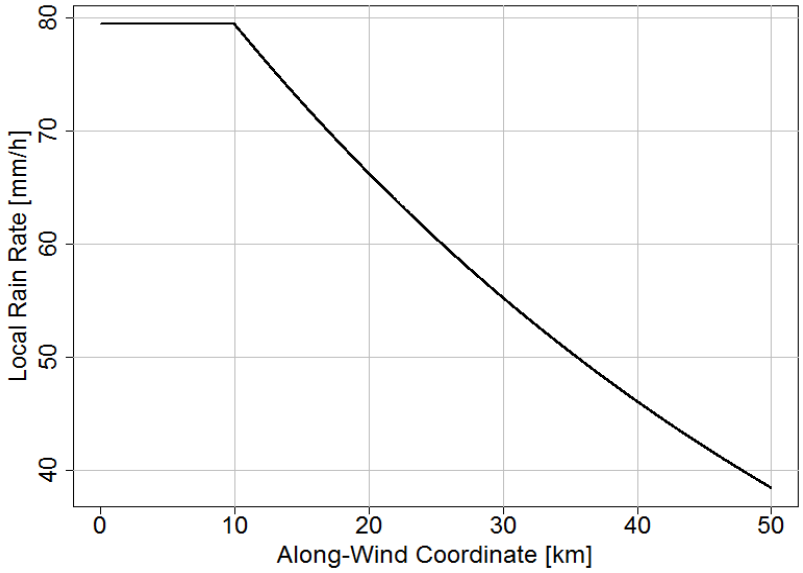

Figure 4.8: Maximum rain rate following a simple model of orographic rain as a function of surface air temperature T_0 for a horizontal wind speed of $20m/s$, a hill slope of $dz/dx = 0.05$, a cloud base height of $z_{cb} = 500m$ and a surface air temperature of $30°C$ at $z = 0$.

4.4.2 Rain and Height

Based on the observation of higher rain depths in mountainous areas than in flat terrains, it is a common misunderstanding that rain amounts increase with increasing height. Although this might be observed in some cases, it is not true in general as Figure 4.8 reveals. This is for several reasons. One reason is that the maximum available water vapor decreases with height as the air temperature decreases with height.

For a single event, rain intensity is highest at the cloud base z_{cb} since rain is produced above this level and accumulates to it while it partly evaporates below the cloud base. Using eq. (4.24) we can build the ratio of condensation above an arbitrary level z and total condensation above the cloud base as

$$\gamma = \frac{\exp\left(-\frac{z}{H_m}\right)}{\exp\left(-\frac{z_{cb}}{H_m}\right)} = \exp\left(-\frac{(z - z_{cb})}{H_m}\right). \tag{4.32}$$

This decreasing condensation rate makes the rain amount of a rain event smaller the higher the elevation. In fact extreme mountaineers climbing in the Himalaya prefer days on which rain is forecasted to those on which strong wind is forecasted for their attempts of the summit since substantial precipitation affects only lower altitudes.

4.5 Convective Rain

The maximum updraft within a convective cloud due to the release of latent heat by condensation of water vapor is a function of the convective available potential energy

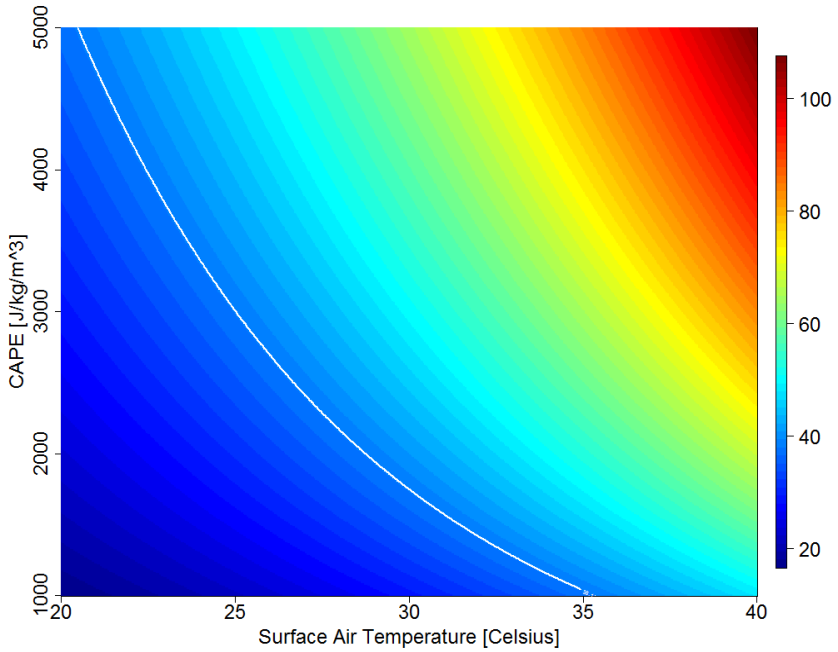

Figure 4.9: Convective rain rate at cloud base in mm/min of a cloud with cloud base $z_{cb} = 500m$, cloud top $z_t = 8000m$ and $w_{cb} = w_t/10$ as function of surface air temperature and CAPE. For comparison the observed maximum one-minute rain rate of 38.1mm is drawn as a white line.

$CAPE$. We call this updraft speed which is reached at the top of the cloud

$$w_t = \sqrt{2 \cdot CAPE}. \tag{4.33}$$

We furthermore assume that the updraft speed increases linearly from its value w_{cb} at cloud base z_{cb} to $w_{cb} + w_t$ at cloud top z_t. This allows to write the updraft speed w within a convective cloud as a liner function of z

$$w(z) = w_{cb} + \frac{z - z_{cb}}{z_t - z_{cb}} w_t = w_a + w_b \cdot z. \tag{4.34}$$

Inserting this updraft speed into equation (4.22) and integrating over cloud depth yields the column condensation rate above cloud base

$$\begin{aligned} C &= s_0(T_0) \cdot w_a \left[\exp\left(\tfrac{-z_{cb}}{H_m}\right) - \exp\left(\tfrac{-z_t}{H_m}\right) \right] \\ &+ s_0(T_0) \cdot w_b \left[(H_m + z_{cb}) \exp\left(\tfrac{-z_{cb}}{H_m}\right) - (H_m + z_t) \exp\left(\tfrac{-z_t}{H_m}\right) \right] \end{aligned} \tag{4.35}$$

which we interpret as the rain rate at cloud base. Fig. 4.9 shows results of eq. (4.35) for given cloud base and cloud top heights as function of $CAPE$ and surface air temperature and a simple relation between updraft at cloud base and at cloud top ($w_{cb} = 0.1 w_t$). Equation (4.35) seems to provide a realistic order of magnitude of maximum convective rain.

Chapter 5

Intensity-Duration-Frequency and Scaling

5.1 IDF Plots

Plots with Intensity-Duration-Frequency curves, so-called IDF plots, are a common tool to investigate the frequency (or return period) of exceedances of certain thresholds within certain durations. These plots usually show scaling. As an example, Figure 3.1 shows the maximum observed rain for different durations from one minute to 2 years worldwide. It can be interpreted as an IDF plot approximating an infinite return period since the observed frequency of exceedance is 0.

Let I_D be the rain depth for duration D and $I_{n\cdot D}$ the rain depth for duration $n \cdot D$. Both rain depths have cumulative distributions $F(I_D)$ and $F(I_{n\cdot D})$, respectively. Scaling means that

$$F_D(I_D) = F_{n\cdot D}(n^{-\beta}\, I_{n\cdot D}). \tag{5.1}$$

β is the scaling exponent and relation (5.1) is called simple scaling in the strict sense. If scaling does not hold for the distribution but for moments, i.e.

$$E\left[I^q_{n\cdot D}\right] = n^{\beta_q} E\left[I^q_D\right] \tag{5.2}$$

with a scaling exponent β_q for the q-th moment $E[I^q]$ and β depends linearly on q then it is called simple scaling in the wide sense. The β_q can be calculated by linear regression from

$$\log\left(E\left[I^q_{n\cdot D}\right]\right) = \beta_q \log(n) + \log\left(E\left[I^q_D\right]\right). \tag{5.3}$$

Scaling holds for parameters of fitted CDFs (e.g. Gumbel) and for quantiles, see Gupta and Waymire, 1990 [7]. Therefore we might assume that it also holds for extremes.

Let us take the time series of Dresden (see section A) and apply eq. (5.3) to some moments, quantiles and the maxima. We chose the first three moments, i.e. n-day precipitation mean, mean of squared precipitations and mean of cubed precipitations as well as the n-day precipitation standard deviation. Our duration unit D is a day and n is $1, 2, 3, \cdots, 100$, meaning we search for scales in the range of one day to 100 days. Fig. 5.1 shows that the moments scale pretty well in the range of 1 to 100 days

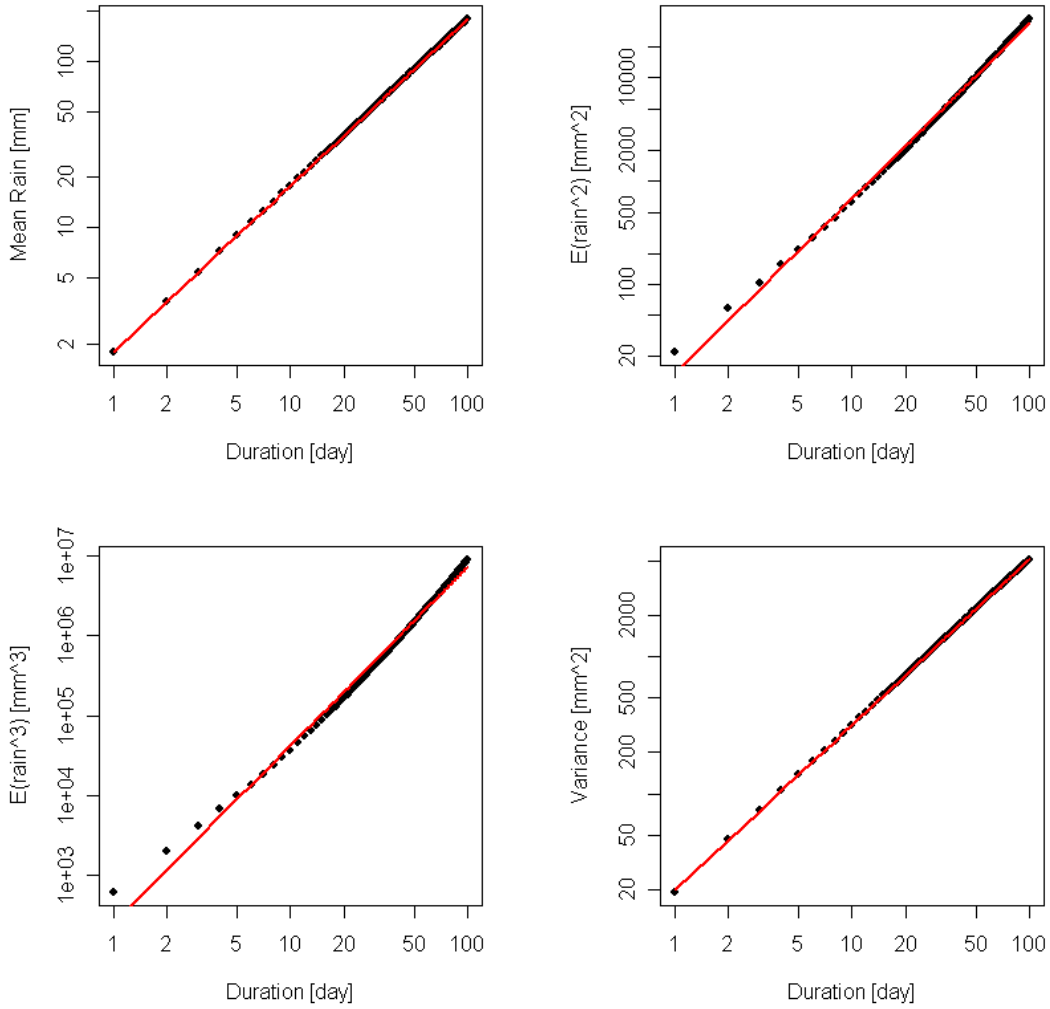

Figure 5.1: Scaling of moments with duration from 1 to 100 days for rain observations in Dresden from 1917 to 2001. First moment (expectation) is shown in upper left plot, second moment (expectation of squared observations) in upper right plot, third moment (expectation of cubed observations) in lower left plot, and variance in lower right plot.

for precipitation observations in Dresden from 1917 to 2001. Small deviations from the scaling law are visible in the second and third moments, though.

Fig. 5.2 shows that also the 90%, the 95% and the 99% quantiles scale very good. And even for the observed n-day maxima, scaling is found in the precipitation observations of Dresden from 1917 to 2001, even though with much more noise.

Table 5.1 shows the scaling coefficients found for the moments and quantiles of rain observations in Dresden. The analysis is done for two periods, 1917-2001 and 1917-2002, respectively, since in the year 2002 an extreme daily rain event was observed in Dresden.

Table 5.1: Scaling parameters for moments and quantiles of observed rain in Dresden for two periods and a scaling range from 1 to 100 days.

Parameter	1917-2001		1917-2002	
	$D = 1$	exponent	$D = 1$	exponent
Mean $[mm]$	1.784	1	1.789	1
$E(\text{Rain}^2)$ $[mm^2]$	13.888	1.693	14.251	1.688
$E(\text{Rain}^3)$ $[mm^3]$	252.711	2.226	284.091	2.2
Variance $[mm^2]$	19.711	1.209	20.431	1.203
q90 $[mm]$	5.479	0.85	5.456	0.852
q95 $[mm]$	8.319	0.789	8.266	0.792
q99 $[mm]$	19.117	0.653	19.462	0.65
Maximum $[mm]$	103.233	0.322	112.807	0.299

In order to check whether we see considerable deviations from scaling in the wider sense, we calculate the expectations of rain depths to the power of q for q in 1 to 6. Fig. 5.3 shows the estimated exponents β_q of eq. (5.2) as a function of the order of the moment q. We see that the first six moments of aggregated rain in Dresden do not show significant deviations from the linear relationship and thus that we find no evidence on which we could reject the hypothesis of scaling in the wider sense. The linear relations are

$$\begin{aligned} \beta_q &= 0.5191 + 0.5602\, q \text{ for the period 1917-2001} \\ \beta_q &= 0.5577 + 0.5352\, q \text{ for the period 1917-2002.} \end{aligned} \tag{5.4}$$

The correlations are 0.998597 for the period of 1917 to 2001 and 0.9979286 for the period of 1917 to 2002.

5.2 Envelope Curves

The structure of eq. (3.1) is well known to hydrologists who use it for regional envelope curves (REC) as

$$Q = k\, A^{\gamma} \tag{5.5}$$

with the maximum flux Q in m^3/s, the size of the region A and the regional constants k and γ. Castellarin et al. (2009) transferred this structure to point rainfall by defining

$$h_{D,max} = k\, MAP^{\gamma} \tag{5.6}$$

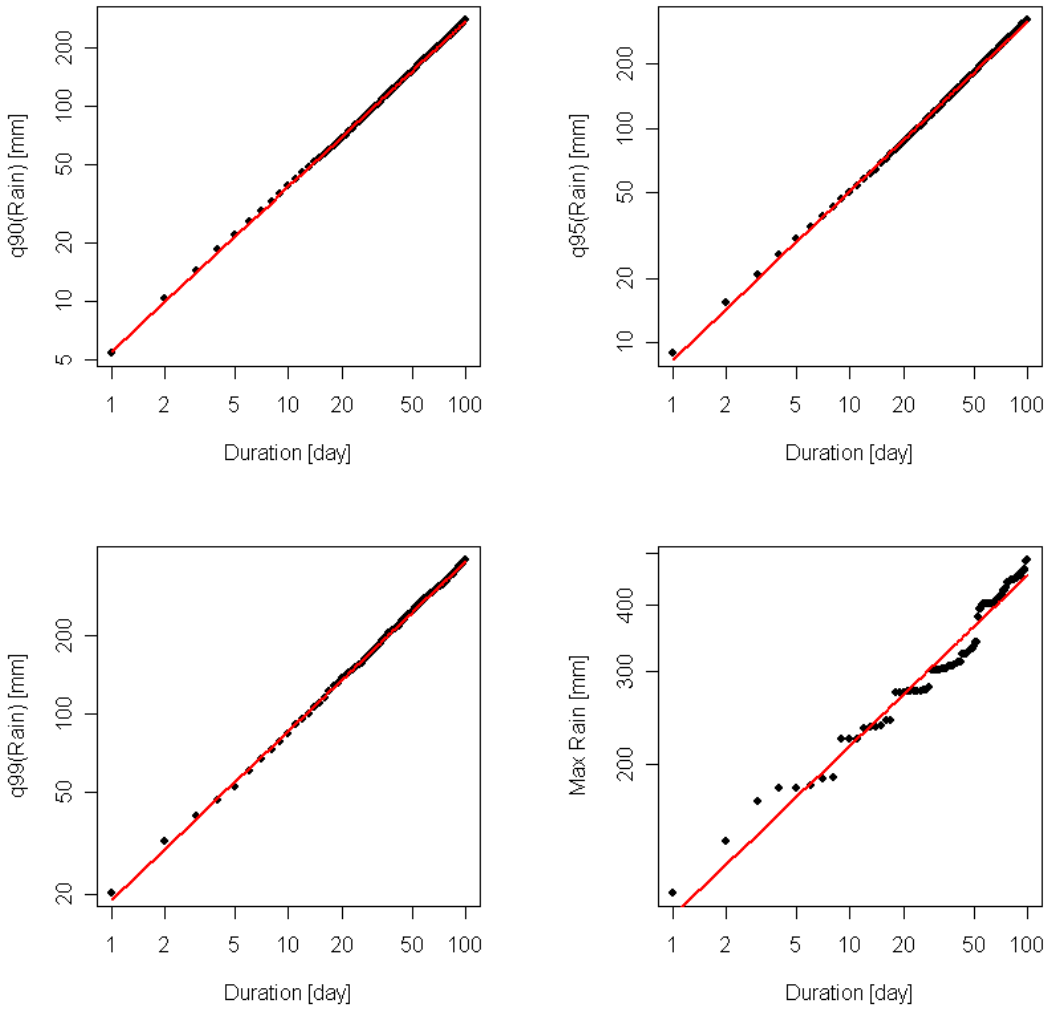

Figure 5.2: Scaling of quantiles and maxima with duration from 1 to 100 days for rain observations in Dresden from 1917 to 2001. 90% quantile is shown in upper left plot, 95% quantile in upper right plot, 99% quantile in lower left plot, and maximum in lower right plot.

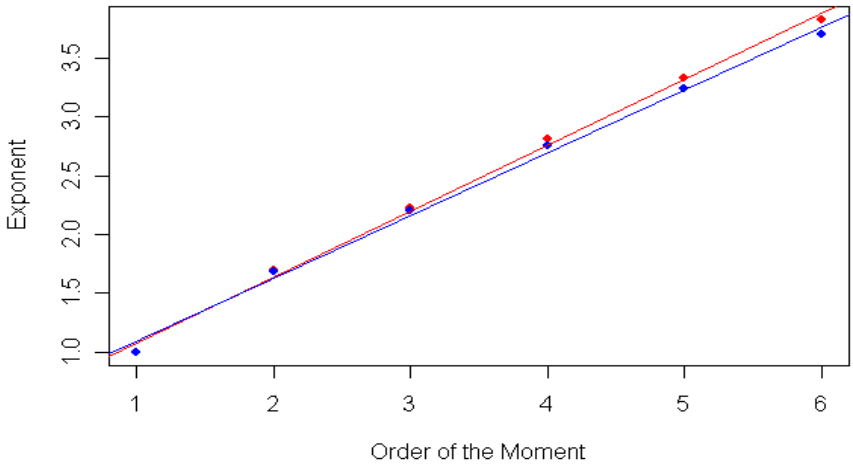

Figure 5.3: Scaling exponents of moments of observed precipitation for durations of one to 100 days as function of the order of the moment for rain in Dresden observed from 1917 to 2001 (red dots) and from 1917 to 2002 (blue dots). Lines are the best fit regression lines.

where $h_{D,max}$ is the proxy of the highest rain rate with duration D, MAP is the mean annual precipitation at the location of interest and k and γ are regional constants. Equation (5.6) can be rewritten as

$$\ln h_{D,max} = \ln k + \gamma \ln MAP \tag{5.7}$$

or for the ratio of the highest rain rate with duration D to the mean annual precipitation as

$$\ln\left(\frac{h_{D,max}}{MAP}\right) = a + b \ln MAP \tag{5.8}$$

with the constants a and $b = \gamma - 1$. Equation (5.8) is called Depth-Duration Envelope Curve ($DDEC$).

The structure of eq. (5.8) is similar to the one of eq. (5.3) which, as we saw, can also be applied to maxima. The difference is that while eq. (5.3) can be used to estimate rain maxima for durations D as function of the maximum rain of one specific duration, eq. (5.8) provides estimates of maximum rain for any duration if the mean annual precipitation of the location is known and the regional constants k and γ.

The advantage is in the regionalisation. Estimating k and γ from some stations allows estimates of maximum rain for locations for which only monthly or annual precipitation is observed or in case only long-term averages are available. The estimation of the regional coefficients allows even to estimate maximum rain of a location in case no observation at all is available. An interpolated annual average rain depth is the only necessary variable.

Let us test the quality of this approach for Germany as an example. Daily rainfall observations covering the 60 years from 1952 to 2011 are available from EC&D [5]. We use the 20 stations Bamberg, Berlin-Dahlem, Bremen, Dresden, Frankfurt am Main,

Hamburg Fuhlsbuettel, Hohenpeissenberg, Karlsruhe, Muenchen, Potsdam, Schwerin, Stuttgart, Zugspitze, Feldberg im Schwarzwald, Wendelstein, Oberstdorf, Wasserkuppe, Brocken, Fichtelberg and Kempten. These stations are distributed all over Germany and span an altitude range of 4m to 2964m. Hohenpeissenberg, Zugspitze, Feldberg im Schwarzwald, Wendelstein, Wasserkuppe, Brocken and Fichtelberg are mountain stations. All stations have 21914 observations from 1952 to 2011 except Bamberg (21911 observations), Muenchen and Zugspitze (21913), and Stuttgart (21887 observations).

We do the fit for durations from 1 to 30 days and for the maximum, second highest, 5th highest and 10th highest observation, assuming that observed maxima are very noisy and that the noisiness reduces for observations closer to the centre of the distribution. For the same reason we also look at the 90, 95, 99, and 99.9% quantile. Given that the 60-year period consists of 21914 days these quantiles reflect the 23rd, 220th, 1097th and 2192nd highest observations, respectively. Fig. 5.4 shows the explained variance (r^2) of the best linear fit to equation 5.7 for the different quantiles and extremes. It can clearly be seen that the correlation is poorer for shorter durations than for longer and for larger quantiles. It is poorest for the extreme observations. It seems that the envelope curve approach is justified for the 99% quantile and lower quantiles. The 99% quantile however, is exceeded on average in one of 100 days with rain. Since it rains on about 50% of the days in Germany this quantile is exceeded nearly twice a year.

The variance however shows good scaling with explained variances exceeding .98 for all durations longer than 3 days. This clearly indicates the heteroscedasticity of precipitation, i.e. higher mean precipitation is linked to higher variance in precipitation. This can be clearly seen in Fig. 5.5 where the coefficient k (values are multiplied by 10 to plot them together with the other two variables), the exponent γ and the explained variance of a best fit of eq. 5.7 for the standard deviation is drawn. We see that the factor k increases for durations longer than 5 days while the exponent γ increases only slowly with duration. It may even converge to a certain value. This is, however, not investigated here.

The mean daily precipitation μ_d is just MAP/n_R where n_R is the average number of days with rain per year. The variance of daily rain scales with MAP and can be deduced from it with good accuracy. This can be used to generate daily data from MAP if a two-parameter (e.g. Weibull or Γ) distribution is assumed.

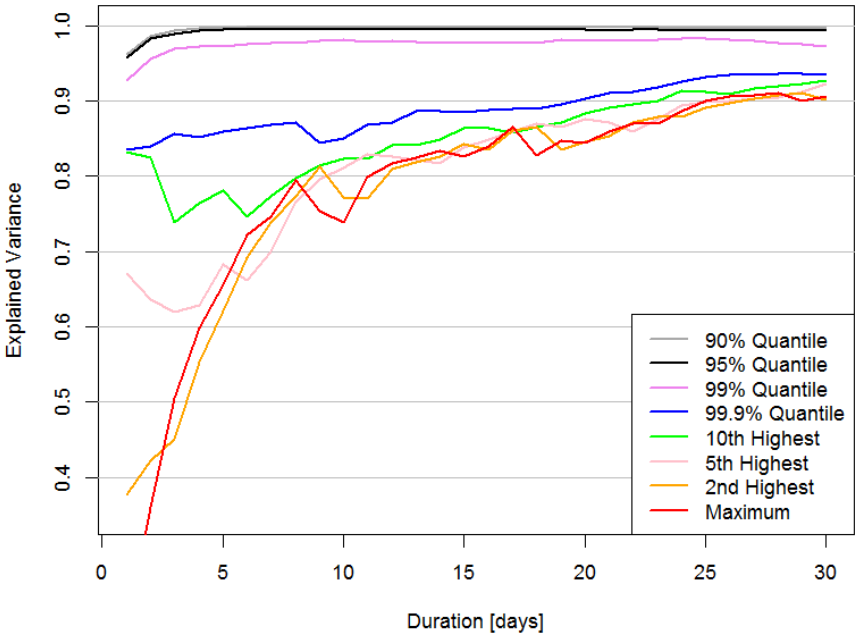

Figure 5.4: Explained Variance of best fit of eq. 5.7 to daily rain observations of 20 stations in Germany from 1952 to 2011.

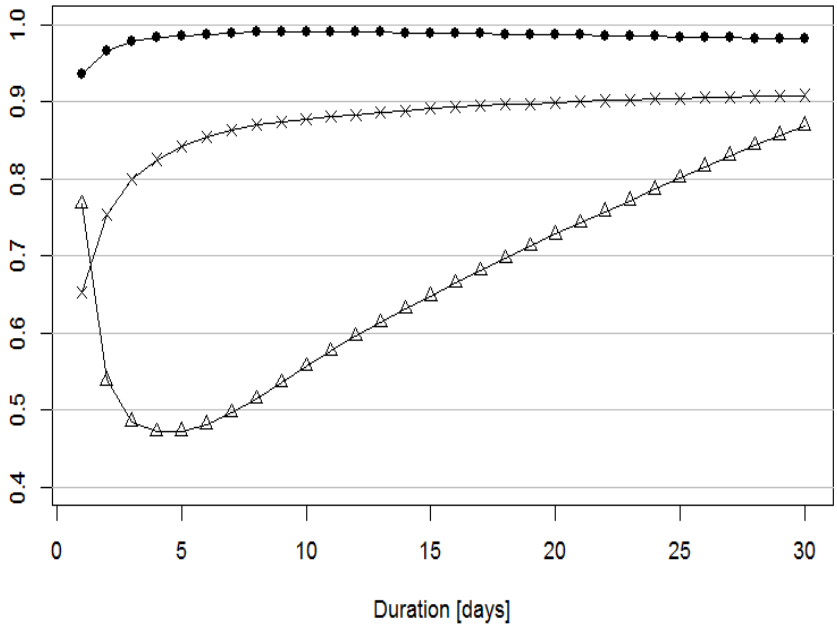

Figure 5.5: Explained Variance (black dots) and estimates of the factor $10 \cdot k$ (triangles) and the exponent γ (crosses) of best fit of the standard deviation of rain of different durations to the scaling law eq. 5.7 for daily rain observations of 20 stations in Germany from 1952 to 2011.

Chapter 6

Extreme Value Theory

6.1 Peaks over Threshold

It is a classical but naïve approach to estimate the probability of excess of a threshold by counting the fraction of exceedances from observations and using this frequency as a proxi for probability. This is generally no problem for low thresholds, which are exceeded often. Those estimates are, however, pretty useless as one focuses on the most extreme values. Applying this approach to high thresholds yields pretty bad results since the number of exceedances becomes small and the uncertainty of the results high. In the most extreme case, the probability of an exceedance of the highest observed value would be estimated to be zero. This would exclude unprecedented high rain events which contradicts experience.

Fig. 6.1 shows that the estimates become more uncertain the higher the return periods are one is interested in. However, one can also see that there seems to be a functional relation between the threshold and the probability with which it is exceeded.

Balkema and de Haan (1974) [2] and Pickands (1975) [14] showed that for a broad class of distributions the distribution of their peaks over threshold converges towards the generalized Pareto distribution GPD

$$GPD(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\theta}\right)^{-1/\gamma} & , \text{ for } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\theta}\right) & , \text{ for } \gamma = 0. \end{cases} \tag{6.1}$$

This distribution is max-stable, meaning that the exponent γ (the shape parameter) does not depend on the chosen threshold, as long as it is large enough so that the convergence to the GPD is ensured. This actually is the problem. How can we know whether a threshold is large enough? A simple straight-forward test is the fit of a GPD to various thresholds and to test convergence. However, for very large thresholds only few exceedances are observed and the error variance may increase considerably. Fig. 6.2 shows the best estimates of scale θ and shape $1/\gamma$ for the time series of observed daily precipitation in Dresden using different thresholds. We see that for thresholds larger 20mm the scale parameter stays around 10mm while the shape parameter varies between 0.1 and 0.3. This can be considered a large range. Fig. 6.2 also shows the estimated return period in years for the exceedance of 100mm given the estimated

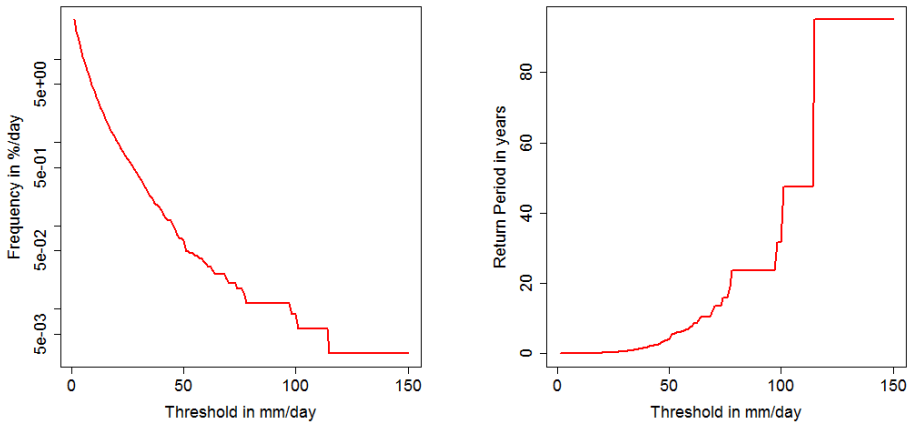

Figure 6.1: Frequency (left graph) and estimated return period (right graph) of daily precipitation peaks over threshold in Dresden from 1917 to 2002.

parameters. We clearly see a strong bias if thresholds are chosen too low but high variability for high thresholds. Furthermore Fig. 6.2 shows that the relative error in scale stays more or less constant for all thresholds larger 24 mm while the relative error in shape further increases. All fits are done with the R package 'evd'.

Given the high uncertainty of the POT method with respect to the observations in Dresden we now focus our attention on another method for the estimation of return periods.

6.2 Extreme Value distribution and Annual Maxima

6.2.1 Return Period, Return Level and the Generalized Extreme Value Distribution

One way to attribute return periods to rain amounts is the fit of the generalized extreme value distribution (GEV) to the annual series of maxima. This approach can be applied under the assumption that daily rain is identically and independently distributed. We fit GEV here to the time series of Dresden (see appendix A) as a test case.

The generalized extreme value distribution describes the probability for so-called block maxima to stay below a certain threshold z. If we are interested in maxima of daily rain we can use a year as a block and thus consider only the highest observation per year, i.e. neglect 364.25 out of 365.25 data. To put it in other words: we use only less than 0.3% of the data. Fig. A.2 shows the time series of annual maxima of daily rain (the annual series) for the observations in Dresden.

According to extreme value theory, the pdf of this time series of maxima can be described by the distribution $G(z)$ with

$$G(z) = \begin{cases} \exp\left\{-\left[1+\zeta\left(\frac{z-\alpha}{\beta}\right)\right]^{-1/\zeta}\right\} & \text{for } \zeta \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\alpha}{\beta}\right)\right]\right\} & \text{for } \zeta = 0 \end{cases} \tag{6.2}$$

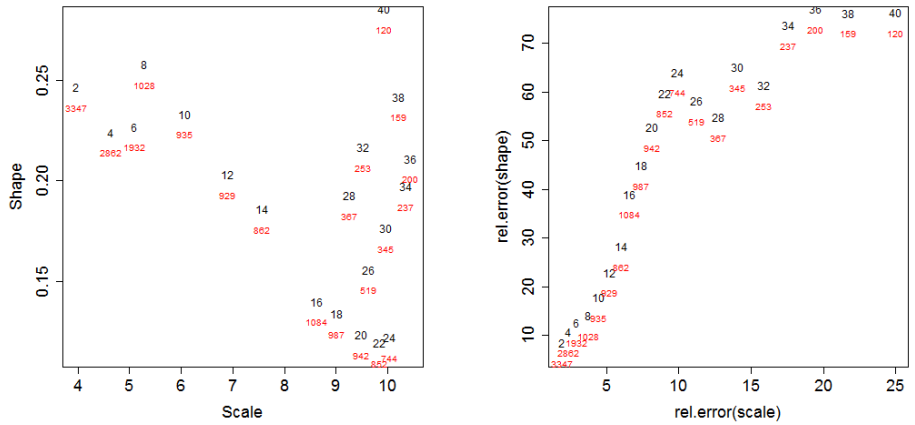

Figure 6.2: Fit of GPD to daily precipitation observations in Dresden. Best parameter estimates as function of threshold (black numbers) are shown in the left graph. The right graph shows the relative error of the estimates in %. Red numbers indicate the return period for the exceedance of 100mm rain in years determined by the best fit parameter values.

where the case $\zeta = 0$ is the Gumbel distribution as a special (2 parameter) case of the 3-parameter GEV.

The probability to exceed a threshold is $1 - G(z)$ per year. If the extremes per year are independent we can define the return period as $\tau = \frac{1}{1-G(z)}$. τ is the average waiting time to the next exceedance of the attributed return level z. We can invert eq. (6.2) and calculate the thresholds as function of G as

$$z_p = \begin{cases} \alpha - \frac{\beta}{\zeta}\left\{1 - [-\ln(G)]^{-\zeta}\right\} & , \text{ for } \zeta \neq 0 \\ \alpha - \beta \ln\left(-\ln(G)\right) & , \text{ for } \zeta = 0. \end{cases} \tag{6.3}$$

With $G = \frac{\tau - 1}{\tau}$ eq. (6.3) links return periods and return levels $z(\tau)$ uniquely.

We can use the 'evd' library in R to easily fit a GEV and a Gumbel distribution to the Dresden time series of yearly maxima of daily rain. And we can do this for parts of the time series in order to see how stable the results are. Fig. 6.3 shows the fit of a Gumbel and a GEV distribution to the data of Dresden in case the time series from 1917 to 2001 is used and in case the time series from 1917 to 2002 is used. The latter includes the record-breaking rain of the year 2002.

Fig. 6.3 shows that the best-fit GEV looks considerably different than the best-fit Gumbel distribution. We also see that the Gumbel distribution seems to be more sensitive to the inclusion of the most extreme value of year 2002. The major differences between the 4 distributions however is invisible in Fig. 6.3. They differ strongly in the tail which cannot be resolved in the representation we have chosen. Therefore we look also at the return period as function of the return level as shown in Fig. 6.4.

While all four curves are pretty similar for return periods up to the order of the length of the observed time series there are major differences for longer return periods. For long return periods (larger 100 years) the GEV seems to react more sensitive to the inclusion

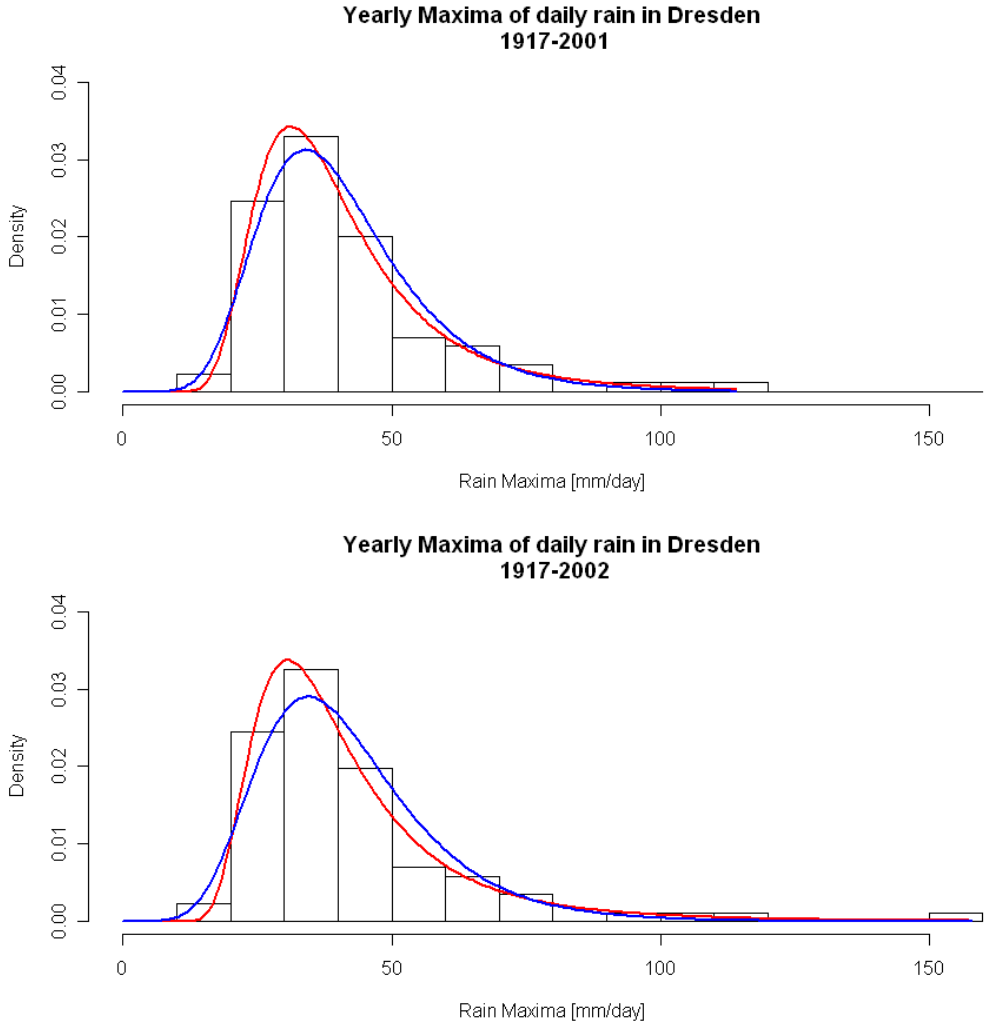

Figure 6.3: Histogram and fitted GEV (blue line) and Gumbel distribution (red line) to observed yearly maxima of daily rain in Dresden from 1917 to 2001 (upper panel) and from 1917 to 2002 (lower panel).

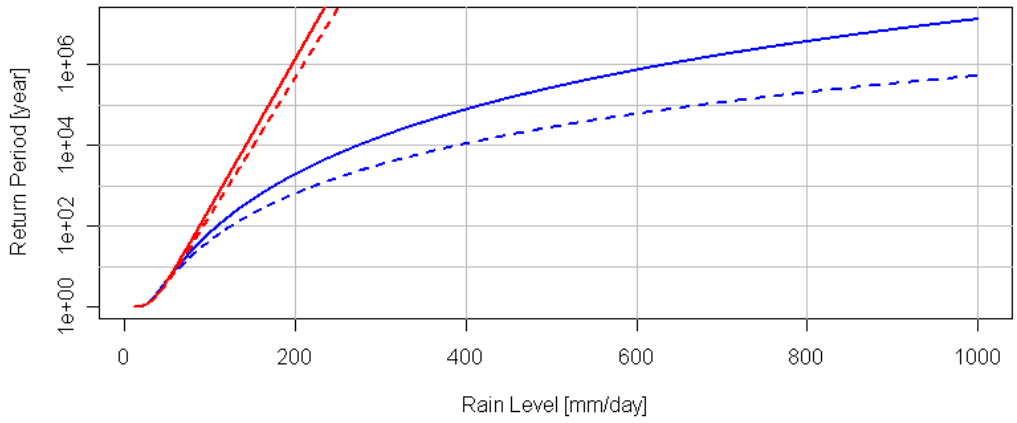

Figure 6.4: Return periods as function of return level for daily precipitation in Dresden and fitted GEV (blue line) and Gumbel distribution (red line) for observations from 1917 to 2001 (dashed line) and from 1917 to 2002 (solid line).

of the observation of the year 2002. For a return level of 200mm we see a difference in the estimated return period of factor 1000 between the GEV and the Gumbel estimates.

Usual ways to visually check the quality of the fit of a distribution to a sample of data points are the quantile-quantile plot (QQ plot) and the probability-probability plot (PP plot). We can easily calculate the quantiles and probabilities from the best-fit distribution. We can also attribute empirical quantiles and probabilities to the observations. For a time series of length L it is usually assumed that the l-th smallest observation reflects the $l/(L+1)$ quantile. And the empirical probability to get a value lower than or equal to the $l/(L+1)$ quantile is just $l/(L+1)$. The empirical quantiles are drawn vs. those of the best-fit distribution in a QQ plot. The PP plot shows the empirical probabilities vs. those of the best-fit distribution. We expect all data points to line up along the 1:1 line if the best-fit distribution is adequate. Systematic deviations from this 1:1 line indicate that the best-fit distribution is not an adequate representation of the observations.

Looking at the PP plots in Fig. 6.5 gives confidence to all 4 distributions since the points line up at the 1:1 lines. However, looking at the QQ plots reveals some systematic discrepancies which are larger for the Gumbel distribution then for the GEV. Especially in case of the Gumbel distribution one is attempted to see a linear dependency for quantiles lower $60mm$ and a linear range with a different slope for quantiles above $60mm$ indicating that the true distribution of extremes may be the sum of two distributions which might reflect different processes leading to the extremes. This would mean that extreme daily rain is driven by a different mechanism for about a dozen of the 95 years - or about one in 8 cases. However, this is not supported by the fit of the GEV where the 4th or 5th highest value is pretty close to the 1:1 line.

So what is the estimated return period for the observed rain maximum of $158mm/day$? And what is the return period for a day with 200mm rain according to the 4 different

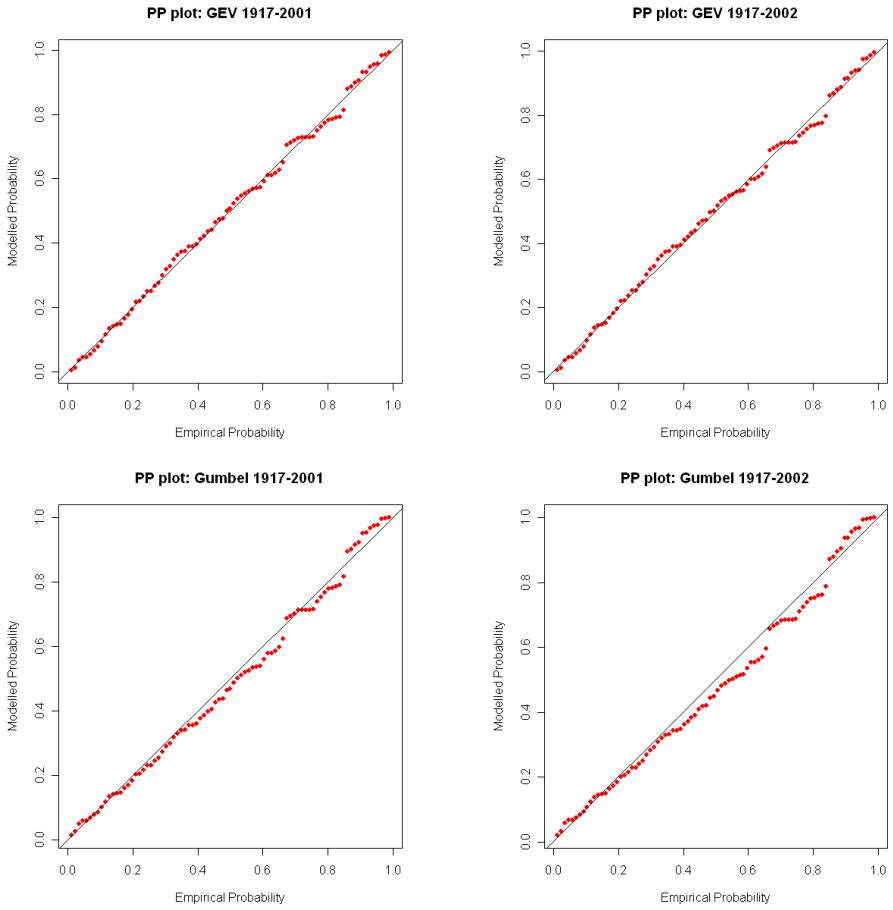

Figure 6.5: PP plots of daily precipitation in Dresden and fitted GEV (upper panel) and Gumbel distribution (lower panel) for observations from 1917 to 2001 (left side) and from 1917 to 2002 (right side).

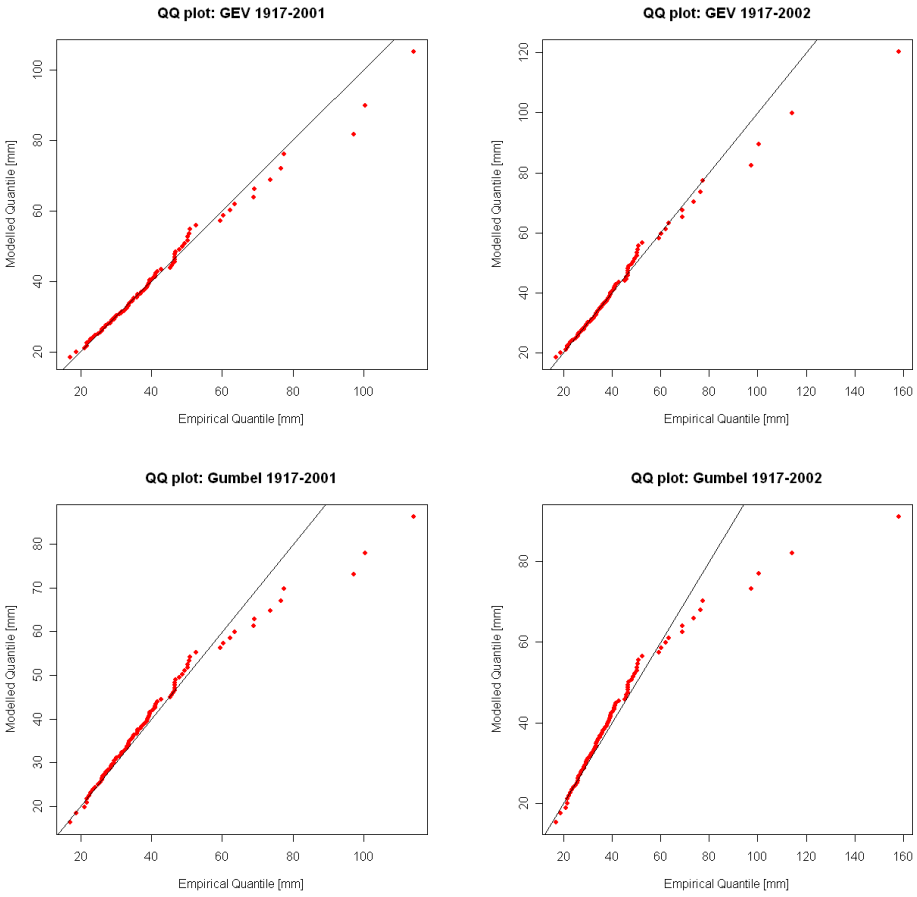

Figure 6.6: QQ plots of daily precipitation in Dresden and fitted GEV (upper panel) and Gumbel distribution (lower panel) for observations from 1917 to 2001 (left side) and from 1917 to 2002 (right side).

fits? What amount of rain can we expect every 200 years, or every 500 years. Table 6.1 provides the parameters of the fits and these return periods and return levels.

Table 6.1: Parameters of fit of Gumbel distribution and GEV to yearly maxima of daily rain for observation period from 1917 to 2001 as well as 1917 to 2002 and estimates of return periods (RP) for 158mm and 200mm daily rain. Return levels (RL) for return periods of 200 and 500 years are provided as well.

Distribution	Gumbel		GEV	
Period	1917-2001	1917-2002	1917-2001	1917-2002
α	34.39	33.92	32.8702	32.8601
β	12.68	11.77	10.8822	11.1899
ζ	0	0	0.1692	0.2313
RP158 [years]	37938	17072	593	251
RP200 [years]	1345899	468071	1934	639
RL200 [mm]	96	102	126	149
RL500 [mm]	107	113	153	188

What can we learn from Table 6.1? Either the highest observed daily rain in Dresden in the year 2002 was a very unlikely event with return period of at least 17000 years or the fit of the Gumbel distribution is not adequate to describe the most extreme events. The latter alternative is supported by the fact that the fit of the GEV leads to exponents of 0.17 or larger which seems to be too far away from zero to be neglected. The Gumbel distribution seems not to be the adequate distribution for the estimation of return periods of extreme daily rain.

However, how reliable is the fit of the GEV? Fitting the GEV to a time series of 85 years means estimating 3 parameters (instead of 2 in the Gumbel case) based on 85 observations, i.e. less than 30 observations per parameter. This is especially important for the shape parameter as a measure of skewness of the distribution. This parameter is very sensitive to single observations. Table 6.1 shows that taking the observation of the year 2002 into account changes the coefficient α by 0.03%, the coefficient β by 2.8% but the exponent γ by 36.7%. This shows the high sensitivity of the shape parameter on observed extremes which in turn lead to high sensitivities of estimated return periods. Table 6.1 shows that the return period of 200mm changed by factor 3 just by taking one more observation into account. We deal with this problem in section 6.2.3.

6.2.2 Seasonality

When applying extreme value theory we silently assume that rain is an iid variable. In fact, it is not. Rain shows seasonality and persistence. However, persistence is usually low (a couple of days). But seasonality is often pronounced. In order to deal with seasonality we might chop a year into four seasons and apply extreme value theory to each of these seasons. This results in 4 GEVs, G_1 to G_4, one per season. We now use these four GEV in order to create a distribution for the annual extreme values. The probability that a threshold is exceeded within the first season is $1-G_1$. The probability that it is not exceeded in the first but in the second season is $G_1 \cdot (1 - G_2)$. Applying this to all four seasons leads to the annual cumulative distribution function for extreme

rain as

$$G_y = 1 - [(1 - G_1) + G_1(1 - G_2) + G_1G_2(1 - G_3) + G_1G_2G_3(1 - G_4)] = G_1\, G_2\, G_3\, G_4. \tag{6.4}$$

Fig. 6.7 shows that the recognition of seasonality alters the results in terms of return periods slightly. It cannot resolve the issue with the four highest observed extreme daily rain sums.

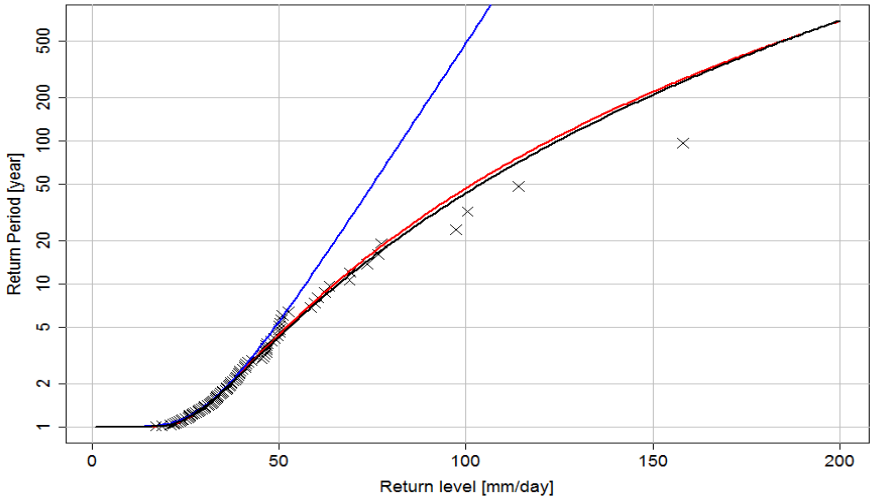

Figure 6.7: Return periods as function of return level for daily precipitation in Dresden from fitted GEV (red line), Gumbel distribution (blue line) and seasonal fit combined following eq. (6.4) for observations from 1917 to 2011. Empirical return periods are provided as crosses.

Instead of fitting GEVs to each season we can also fit them to each month and combine the results similarly as in eq. 6.4. Fig. 6.8 shows the exceedance probability of the combination of monthly and seasonal fits as well as the fit of a GEV to the whole dataset. It is obvious that the monthly approach is a better fit to the observations if the whole time series of daily rain observations is used. It is not clear whether this approach is superior in case only the data until 2001 are used.

As a final experiment regarding seasonality we compare the estimates of return periods of the observed highest daily rain of 158mm in Dresden with estimates from the monthly, seasonal, and direct approach of the best-fit GEV distribution (fit to the annual series) and the Weibull distribution (fit to the whole series). Table 6.2 reveals considerable differences. Keeping in mind that the Weibull distributions are fit to the whole dataset it might be assumed that they are a less good approximation of the tail compared to the GEV.

Finally we can conclude from Table 6.2 that estimates of the return period depend strongly on the representation of seasonality. In the discussed example, neglecting seasonality gives an estimated return period which is more than 5 times higher compared to the case where monthly resolution is used to represent the seasonal cycle. The question, whether a month is long enough to ensure the applicability of the GEV is not discussed here. Of course any other resolution between seasonal and monthly can be used.

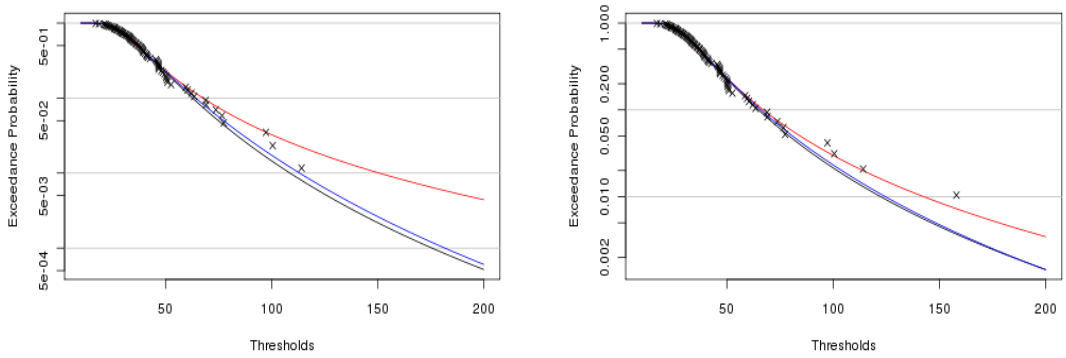

Figure 6.8: Exceedance probability curves of daily precipitation in Dresden for observations until 2001 (left graph) and until 2011 (right graph); Observations (crosses) and best fit of GEV to whole dataset (black lines), seasonal sets (blue lines) and monthly sets (red line).

Table 6.2: Estimates of return period (in years) of daily rain exceeding 158mm in Dresden from observations from 1917 to 2001 based on different distributions and considerations of annual cycle.

Distribution	Annual Fit	Seasonal Fit	Monthly Fit
GEV	593	493	115
Weibull	1613	1542	1203

6.2.3 Buying Time for Space

In section 6.2.1 we saw that in case of the time series of daily rain in Dresden taking the observation of the year 2002 into account (instead of stopping the observation in 2001) alters the coefficients of the fitted GEV very differently. The coefficient α changed by 0.03%, the coefficient β by 2.8% but the exponent γ by 36.7% (see Table 6.1). We see that the estimate of the shape parameter (γ) of the GEV is much more uncertain than the location and the scale parameter, α and β, respectively. The reason is simply that the location parameter is closely linked to the first moment, the scale parameter to the second moment and the shape parameter to the third moment. The higher the order of a moment the more observations are necessary to estimate it with low error variance. The reason is that the moments are sums over the power of observations and powers stretch the distribution of observations.

In order to get less noisy estimates of the shape parameter one can try and combine the observations of neighboring stations and by that 'buy' observation time for 'space'. To do so all observations should be standardized in a way. One way to do that is by substracting the mean and dividing by the standard deviation so that all observed time series have the same mean (zero) and the same standard deviation (unity). As an example we do this for six German stations (all in the east) and fit a GEV to each of the resulting time series. The results are shown in Table 6.3. The time series cover a different number of years ranging from 61 years in the case of Erfurt to 136 years in case of Berlin. The estimates of the location parameter are $-.470 \pm .028$. The scale parameter is 0.601 ± 0.086 and the shape parameter is 0.181 ± 0.056. We can roughly

say that the coefficient of variation is twice as high in the scale parameter than in the location parameter and increases by another factor two for the shape parameter. The latter has a coefficient of variation of about 31%. Note that this variation is not only due to spatial variability but also due to the different length of the time series.

Table 6.3: well.

Station	location	scale	shape	years
Berlin	-0.4676319	0.5187817	0.2607286	136
Dresden	-0.4499417	0.5156023	0.2321149	95
Erfurt	-0.5014258	0.7108575	0.1232623	61
Goerlitz	-0.5077772	0.6936420	0.1468584	64
Halle	-0.4484008	0.6098957	0.1361418	112
Rostock	-0.4431606	0.5572314	0.1840686	61
mean	-0.4697230	0.6010018	0.1805291	88.17
std.dev	0.0283163	0.0856722	0.0556579	31.5
cv	6.03%	14.25%	30.83%	35.7%
All	-0.4627846	0.6010495	0.1665042	529

We now buy time for space by combining all 6 standardized time series to one series of length 529 years and fit a GEV to it. The resulting shape parameter is 0.1665 or about 1/6. We might assume that this is a representative estimate for the whole area and fit only the scale and shape parameter to each observed time series hoping that this is a more reliable local estimate. Fig. 6.9 shows the estimated standardized exceedance probability curves for the six individual time series as well as the combined series and the fitted GEV.

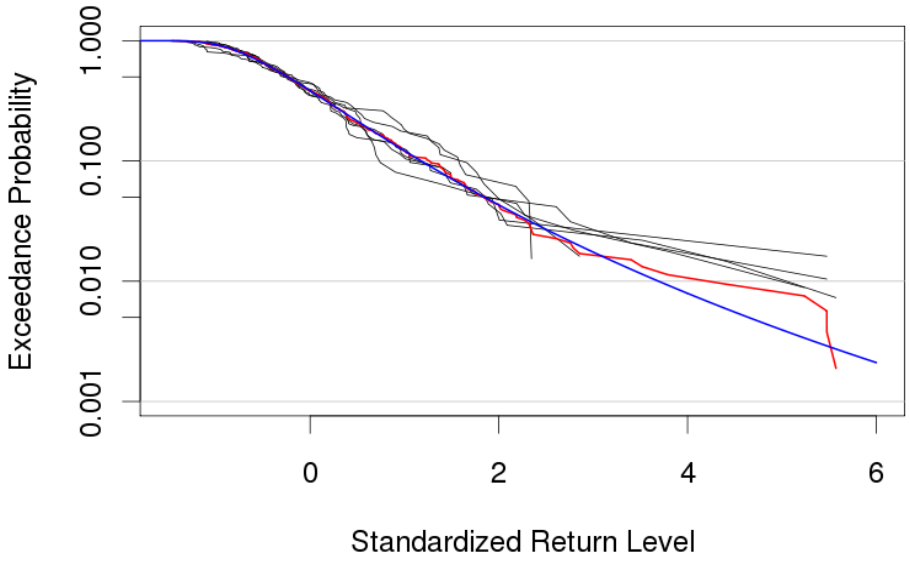

Figure 6.9: Standardized empirical exceedance probability curves for six stations in the eastern part of Germany (black lines), for the combined time series (covering 529 years, red line) and best fit of a GEV distribution.

In a last experiment we look how sensitive the results are with respect to outliers, i.e. the torrential rain observed in Dresden in summer 2002. We therefore do the analysis for the period from 1952 to 2001 and 1952 to 2011 separately. The results of both experiments are shown in Fig. 6.10.

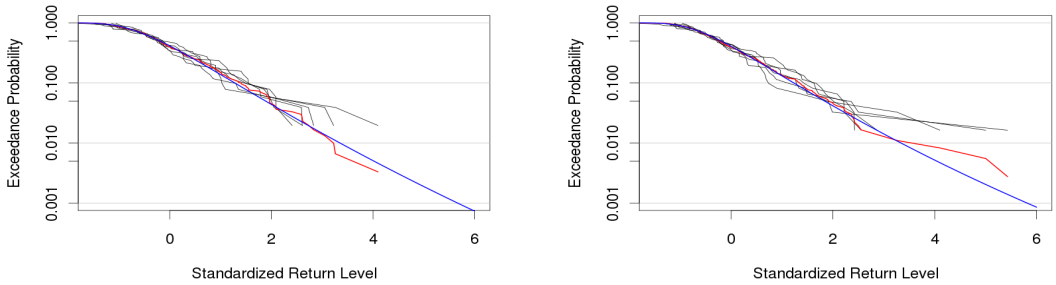

Figure 6.10: Standardized empirical exceedance probability curves for six stations in the eastern part of Germany (black lines), for the combined time series (covering 529 years, red line) and best fit of a GEV distribution. The left graph shows results for the period from 1952 to 2001, while the right one shows results for the period 1952 to 2011.

We see that the fitted GEV shows nearly identical results for both periods and therefore is less sensitive to outliers. The differences in the individual exceedance-probability curves within each plot are not affected by different time series length, though. The results look quite robust.

6.2.4 EVT and Scaling in the IDF Plot

In chapter 3 we saw that the global observed rain extremes over different durations scale with the square root of the duration. For rain intensities I this means

$$I_D = \frac{a}{D^{0.5}}. \tag{6.5}$$

Scaling in IDF plots is generally described by a simplification of

$$I_D = \frac{a}{(D^\beta + \gamma)^\eta} \tag{6.6}$$

where either $\beta = 1$ or $\eta = 1$ is to be used to avoid overparameterization. We know that we get different IDF curves for different return periods τ. Furthermore we see that the denominator is only a function of the duration D. Therefore we have to conclude that the numerator a must be a function of τ. This means we can write

$$I_D = \frac{a(\tau)}{b(D)}. \tag{6.7}$$

Different approximations for $a(\tau)$ can be found in the literature where the oldest ($a = \lambda\tau^k$) may be from Bernard (1932). Here we follow an approach by Koutsoyiannis et al. 1998 [11] to show that the form of $a(\tau)$ follows from the underlying distribution or the extreme value distribution for the tail. We start by asking for the probability that I_D does not exceed a given threshold i and write the identity

$$P(I_D \leq i) = P(I_D \cdot b(D) \leq i \cdot b(D)) \tag{6.8}$$

which can be written in terms of the cumulative density function as

$$F_I(I_D) = F(I_D \cdot b(D)). \tag{6.9}$$

Inserting eq. 6.7 yields

$$F(I_D \cdot b(D)) = F(a(\tau)) = 1 - \frac{1}{\tau}. \tag{6.10}$$

This can be used to define $a(\tau)$ to finally write the scaling law

$$I_D = \frac{F^{-1}\left(1 - \frac{1}{\tau}\right)}{b(D)}. \tag{6.11}$$

Comparing equations (6.7) and (6.11) shows that the function $a(\tau)$ can be calculated directly from the return period if the pdf is known or assumed. If $b(D) = D^\beta$ is a good approximation (as we showed it is for the global extreme observations and the time series of Dresden) we might use eq. 6.11 to calculate IDF curves for any return period. However, as we saw in Table 5.1, β is far from constant. It was about 0.85 for the 90% quantile, .79 for the 95% quantile, .65 for the 99% quantile and about 0.3 for the maximum. This shows that this nice approach is not applicable for the time series of daily rain in Dresden without further knowledge of $\beta(\tau)$.

6.2.5 General Limitation of Extreme Value Theory

A general problem with extreme value distributions for non-limited variables is that every extreme can happen. It might be very unlikely but according to the distribution not impossible. It is just a matter of time to observe an arbitrary high daily rain amount. This is very unsatisfying. In fact we should expect a physical upper limit. The amount of water on the earth is limited. The amount of water that fits into the atmosphere is limited. The maximum evaporation rate as a source of water in the air is limited. Therefore we should expect that there is a maximum of rain that can fall within a certain period at a given location. Can we estimate this maximum from observations?

Of course not more rain than is available on earth can fall during one day at one location. So this is an estimate of the maximum. However, this is a bad estimate. It is certainly far to high. So what we search for is the lowest estimate of the maximum, given some physical considerations and the local conditions. We want to estimate a probable maximum precipitation. And this is the subject of the following chapter.

Chapter 7

A Glimpse of Meteorology

So far we used conceptual physical as well as statistical methods to analyze and describe the observed daily precipitation time series of Dresden. Different views led to different estimates of return periods. However, we did not distinguish different meteorological situations, although we saw that maybe the largest observations may not fit to the rest of the distribution. Here we take another standpoint arguing that there are usual extremes and unusual extremes. We assume that they belong to different statistical populations. If this is true they should not be mixed but dealt with separately to get a full and consistent picture. While the usual extremes may be in line with the distribution of rain or extreme rain, the unusual ones might result from a specific weather pattern. We do not advocate this approach but have a look at the 850hPa wind pattern for four of the 8 highest rain observations. We chose 4 of the 8 events since only these 4 are in the time range covered by NCEP CFSR data. Aside 8^{th} August 2002 these are the 18^{th} July 1987, the 2^{nd} August 1998, and the 6^{th} July 1999. The wind patterns of 12.00 noon are shown in Fig. 7.1.

In all four cases warm and humid air from the Adriatic Sea is conveyed northward where some convergence becomes visible. In case of the extraordinary high rain of 8^{th} August 2002 we see a looping so that the humid air arrives in Saxony from the north hitting the mountain ranges there, which dramatically enhanced the rain. We needed to look at much more information in order to get a full picture of what was going on. Instead of going down to details, it might be a conclusion that identifying the meteorological situation that yields extraordinary extreme rain may add value to the overall analysis of extreme rain. It might allow to do a statistical analysis for all but the most extreme rain. This might be more robust than using all data in case that the latter can be explained by a specific meteorological situation. If finally the probability of such a meteorological situation can be estimated (say from reanalysis data or GCM runs) a full picture can be drawn.

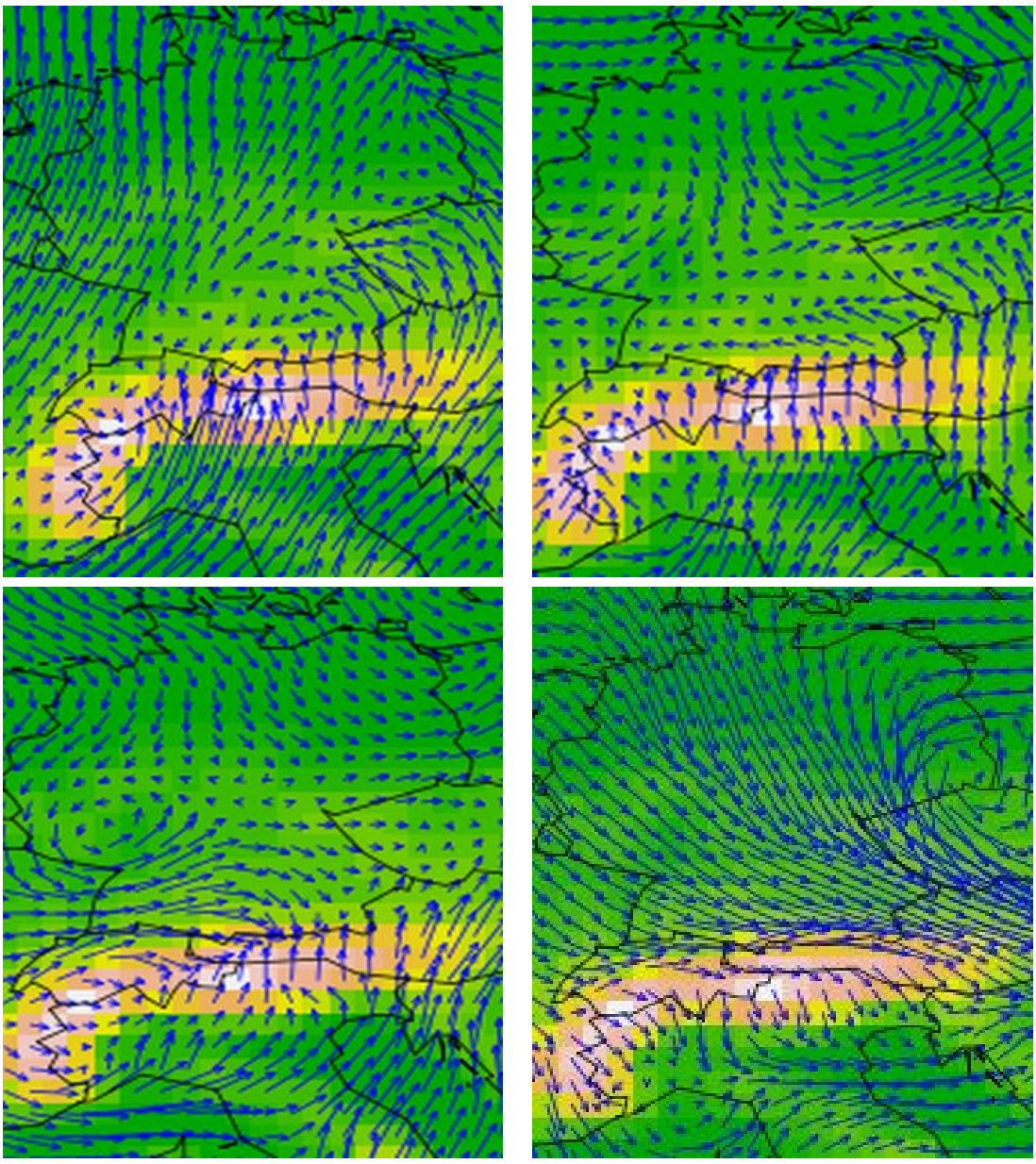

Figure 7.1: 850hPa reanalysis wind field from CFSR for 12 noon of the 18^{th} July 1987 (upper left image), the 2^{nd} August 1998 (upper right image), the 6^{th} July 1999 (lower left image), and the8^{th} August 2002 (lower right image).

Chapter 8

Probable Maximum Precipitation PMP

8.1 Definition of PMP

Probable Maximum Precipitation (PMP) is defined by the AMS (1959) [1] as "the theoretically greatest depth of precipitation for a given duration that is physically possible over a particular drainage area at a certain time of year". In 1986 the WMO [19] defined it very similar. Analog to the PMP a PMF (Probable Maximum Flood) can be defined as well. The definition of PMP, however, gives no aid for the estimation of it. As early as 1942 Showalter and Solot [16] published three limiting conditions for maximum precipitation. These three limitations are

1. The amount of water vapor that fits into an air column,
2. the rate with which water vapor can be transformed into rain, and
3. the rate with which new water vapor can be advected.

Note, that these conditions are very close to the assumptions used in section 4.2.

All efforts to estimate PMP theoretically by formulating and combining the three limits led to unsatisfying high and very sensitive results. Therefore, PMP is usually estimated by statistical means from observations although the definition of PMP asks for a theoretical approach for what is physically possible.

According to Benson (1973) [3] the concept of PMP is based on the earlier definition of a maximum possible precipitation, which proofed to be not achievable. Benson suggested that the whole idea of providing an estimate of a probable upper limit of precipitation was not dropped because it '... removes responsibility for making important decisions as to degree of risk or protection'. It is therefore desirable to attribute probabilities of exceedance to estimates of PMP. We discuss this in section 8.2.

Scaling laws (see chapter 5) can be used to estimate PMP for different durations D if it is known for one duration. The dependence of PMP from duration D is often modelled according to the observed most extreme rain worldwide as the scaling law

$$PMP = aD^b. \tag{8.1}$$

Different authors provide different values for the coefficients a (the one-hour rain depth) and the exponent b. Chow et al. (1988) estimated $a = 422mm$ and $b = 0.475$ for D in hours from global data. Dingnom (1994) estimated $a = 425mm$ and $b = 0.47$. Winsley (1982) gave $a = 417mm$ and $b = 0.48$ for observed maximum rainfall. Viessman and Lewis (1996) gave $a = 15.3inch$ and $b = 0.486$. Note that while the exponents are always lower than the one found for the observed global maxima, the hourly PMP is higher than the observed one of $408mm$.

8.2 Statistical Approach to PMP

In the sixties of the last century a maximum credible earthquake (MCE) [20] was defined as an earthquake which has a probability of 10% to occur at least once within the next 50 years. In the same way also probable maximum loss (PML) [20] is defined in the insurance industry as a loss which is exceeded with a certain probability within a certain time frame. The finance industry introduced the value at risk (VaR) in 1987 as the amount of money that can be lost with a given probability within a certain period. We can use the same approach to estimate a probable maximum precipitation $R_n(p_n)$ as the rain threshold which is exceeded with a probability p_n within the next n years.

If the probability p to exceed a threshold within a year is constant in time (Poisson process) we can give the waiting-time distribution

$$g(k) = (1-p)^{k-1}p \tag{8.2}$$

that we have to wait k years until the threshold is exceeded first. $(1-p)^{k-1}$ is the probability that the threshold is not exceeded within the first $k-1$ years while p is the probability that it is exceeded in the k-th year. The expectation of this distribution is the return period τ. The risk p_n of exceeding the threshold within the next n years is just the sum of the probabilities of exceeding it first in any of the years and thus

$$p_n = \sum_{k=1}^{n} g(k) = 1 - (1-p)^n. \tag{8.3}$$

p is the probability to exceed the threshold within a year and is therefore the reciprocal of the return period τ. Therefore we can write

$$p_n = 1 - \left(1 - \frac{1}{\tau}\right)^n \text{ or } \tau = \frac{1}{1-(p_n-1)^{1/n}}. \tag{8.4}$$

One might not like to use the power n especially if the time horizon regarded is long, i.e. $n = 200$ years. We can rewrite the first of eqs. (8.4) as

$$p_n = 1 - \exp\left(n \ln\left(1 - \frac{1}{\tau}\right)\right) \tag{8.5}$$

and approximate $\ln x = x - 1$ to get

$$p_n \approx 1 - \exp\left(\frac{-n}{\tau}\right).^{1} \tag{8.6}$$

[1] Note that a further approximation $\exp(x) = 1 + x$ would yield $p_n = \frac{n}{\tau} = p \cdot n$ which is only reasonable if $p \cdot n \ll 1$.

Using the same approximation we get

$$\tau \approx \frac{-n}{\ln(1-p_n)}. \tag{8.7}$$

We can use eq. (8.6) to calculate how likely it was to observe a day with 158mm rain within a time series of Dresden rainfall covering $n = 95$ years from 1917 to 2011. According to Table 6.1 we get $\tau = 593$ years from observations within 1917 to 2001 and $\tau = 251$ years if the year 2002 is considered as well. The probability to observe a day with $158mm$ rain in Dresden in a time series of 95 years is therefore 14.8% in the first case and 31.5% in the latter. These probabilities are not low at all.

We can generally use eq. (8.7) to estimate the return period an event must have to occur with a certain probability within a certain period. If we want a 10% probability to occur within 100 years we get a return period of 949 years. An event that we want to happen with a probability of 1% within 50 years needs a return period of 4975 years.

We can use eq. (8.6) together with eq. (6.3) and $G = \frac{\tau-1}{\tau}$ to calculate the probability of a maximum daily rain within a certain period. This can be treated as a maximum probable precipitation in the sense of a maximum precipitation which is not exceeded with a certain probability within a given period of time.

Figures 8.1 and 8.2 show the probability of exceeding a given daily rain sum within a given period of time in Dresden based on observations from 1917 to 2001 and 1917 to 2002.

8.3 Estimation of PMP based on humidity

One way of estimating PMP is to define it based on the highest observed rain $I_{max,observed}$ times the ratio of highest possible humidity to actual humidity during the observed highest rain. The humidity should be measured as specific humidity q in kg/m^3. We can write

$$PMP = I_{max,obs} \cdot \frac{q_x}{q(I_{max,obs})} \tag{8.8}$$

with the maximum observed rain rate $I_{max,obs}$ and the humidities q, where $q(I_{max,obs})$ is the humidity when $I_{max,obs}$ occurred and q_x is an estimate of the maximum possible specific humidity. q_x can either be approximated by the highest observed value or by the highest possible value given the highest observed temperature $q_x = q^*(T = T_{max})$. For Dresden we have observations of daily relative humidity and mean temperature. From that we can calculate daily specific humidity. The maximum within the 95 years from 1917 to 2011 is $14.7g/m^3$. The warmest day had a mean temperature of $30.4°C$ which leads to a maximum possible daily mean specific humidity of $43.4g/m^3$.

Applying equation (8.8) we get PMP values of $198mm$ according to the highest observed daily specific humidity and $585mm$ according to the highest possible specific humidity if humidity is independent of temperature.

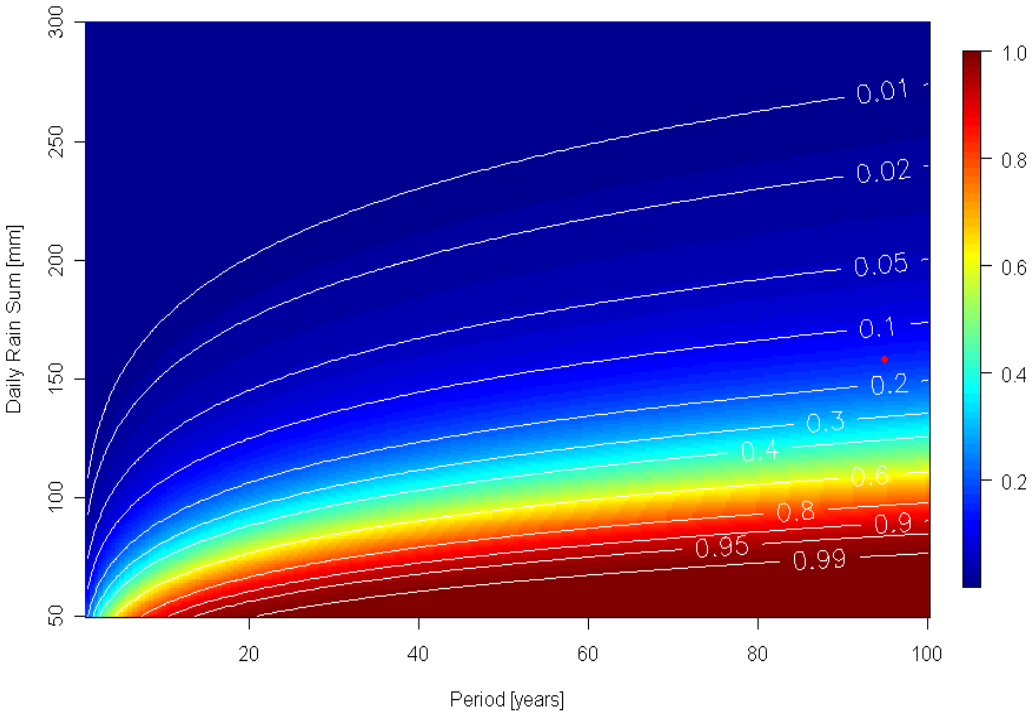

Figure 8.1: Probability to exceed a given threshold within a given period of time for daily precipitation in Dresden based on observations from 1917 to 2001 and a GEV distribution. The red dot marks the observed daily rain of 12th August 2002.

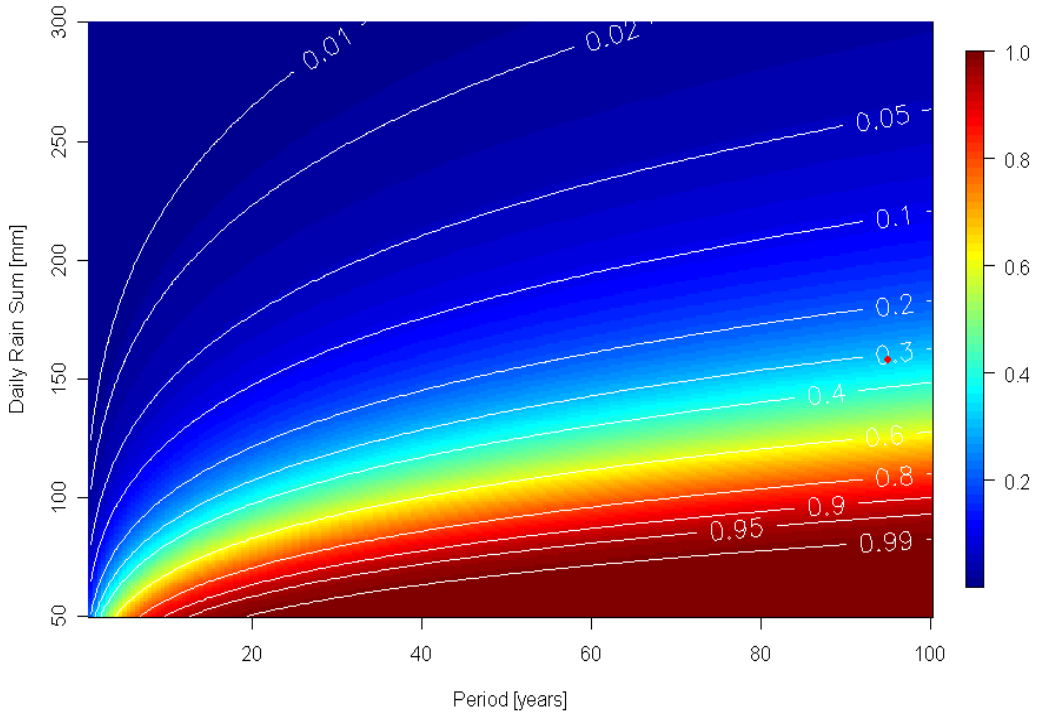

Figure 8.2: Same as Fig. 8.1 but based on observations from 1917 until 2002. The red dot marks the observed daily rain of 12th August 2002.

8.4 Hershfield method for maximum one-day rain

Hershfield [8] analyzed 2645 time series of observed precipitation in order to estimate PMP from observations. He standardized all maxima of annual maximum daily rain by

$$k_m = \frac{x_m - \overline{x}}{\sigma} \tag{8.9}$$

where $\overline{x}$ and σ are the mean and the standard deviation of the observed annual maxima time series and x_m is the highest observation of the time series. k_m is therefore the standardized rain record for each station. The highest value of k_m can be seen as a lower limit of PMP from observation.

Hershfield called k_m the frequency factor. From the 2645 time series he got 2645 frequency factors which he used to investigate the influence of different time series lengths and climate regions. He found a range of k_m values between 1 and 15 and took the maximum value of 15 to be used to estimate PMP for all stations. Therefore we can also use this value in order to estimate a PMP for Dresden and get $310.54mm$ for the time series from 1917 to 2001 and $370.24mm$ for the time series from 1917 to 2002. Based on the fit of a GEV, these thresholds can be attributed to return periods of 19425 and 7914 years, respectively.

However, in a further analysis in 1965 Hershfield [10] showed that $k_m = 15$ seems to be too high for areas with heavy rain and too low for arid areas. He also showed that k_m depends on duration. For shorter durations than 24 hours k_m should take smaller values. His nomogram with values in the range of 5 to 20 became a WMO standard in 1986 [19], although stations with realizations of extreme rain that can be attributed to k_m values of 25 and even 30 became available.

In the literature also an adaptive version of Hershfield's method is propagated where k_m is not predefined but estimated from the time series itself as

$$k_m = \frac{x_m - \overline{x_{N-1}}}{\sigma_{N-1}} \tag{8.10}$$

with the largest observed value x_m and the average and standard deviation of the annual maximum series excluding the largest observation, $\overline{x_{N-1}}$ and σ_{N-1}, respectively. We can easily apply Hershfield's adaptive method to the observations of Dresden. When regarding only the years 1917 to 2001 we get a k_m of about 4.5 and a PMP of $123.1mm$ which is exceeded considerably in 2002 (see Table 8.1 for all data). If we include the year 2002 we get another k_m of about 6.5 and a PMP of $184.6mm$ (see Fig. 8.3). It seems that the adaptive estimation of k_m from observations at a single station is rather sensitive to observed extremes.

According to our discussion in section 8.2 we can attribute probabilities to these PMPs based on GEV and Gumbel distribution. These probabilities are provided in Table 8.2 showing that the probability to exceed these estimates of maximum precipitation within a time series of 95 years is far from zero.

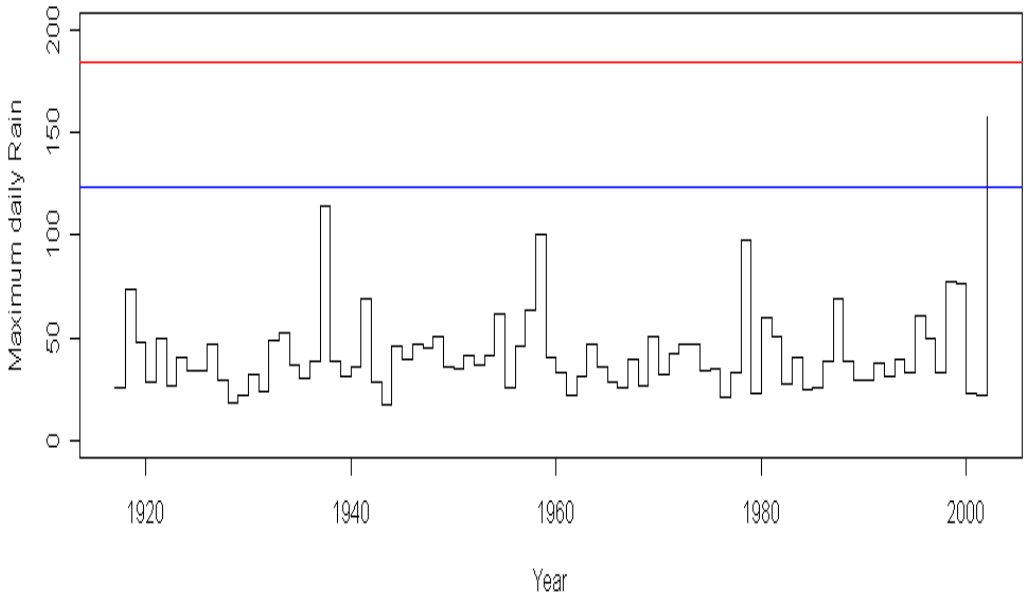

Figure 8.3: Time series of observed yearly maxima of daily rain from 1917 to 2002 and estimates of probable maximum precipitation by the Hershfield method based on this data set (red line) and when omitting the observations of 2002 (blue line).

Table 8.1: Parameters of the Hershfield method for estimating probable maximum precipitation based on daily rain observations in Dresden for two periods.

Variable	1917-2001	1917-2002
$PMP[mm]$	123.08	184.63
$\overline{x}[mm]$	41.29	42.64
$\sigma[mm]$	17.95	21.84
k_m	4.557	6.502
$x_m[mm]$	114.1	158
$\overline{x_{N-1}}[mm]$	40.42	41.29
$\sigma_{N-1}[mm]$	16.17	17.95

Table 8.2: Probability of exceeding the PMP estimate by Hershfield based on Gumbel and GEV distribution.

Distribution	Gumbel		GEV	
Period	1917-2001	1917-2002	1917-2001	1917-2002
$p_n(RL = 123.1mm)$	4.75%	8.36%	41.3%	63.1%
$p_n(RL = 184.6mm)$	0.03%	0.07%	7.1%	18.5%

8.5 Link between Hershfield Method and GEV

The GEV introduced in eq. (6.2) has the mean

$$\mu = \begin{cases} \alpha + \beta\frac{\Gamma(1-\zeta)-1}{\zeta} & , \text{ for } \zeta \neq 0 \\ \alpha + \beta\gamma & , \text{ for } \zeta = 0. \end{cases} \tag{8.11}$$

with the Euler-Mascheroni Constant $\gamma = .57721566490153286\overline{06}$ and standard deviation

$$\sigma = \begin{cases} \frac{\beta}{\zeta}\sqrt{\Gamma(1-2\zeta) - \Gamma(1-\zeta)^2} & , \text{ for } \zeta \neq 0 \\ \frac{\beta\pi}{\sqrt{6}} & , \text{ for } \zeta = 0. \end{cases} \tag{8.12}$$

If we now assume that the rain maxima are GEV distributed, we can use the means and standard deviations as derived from fitted GEV within the Hershfield formula $PMP_H = \mu + k_m\sigma$. The results are provided in Table 8.3.

In case of a Gumbel distribution we can directly estimate a return period attributed to the Hershfield PMP. This is possible since we can provide quantiles of the Gumbel distribution in the form $q = \mu + k_m\sigma$. We do not have to know the parameters of the best-fit Gumbel distribution to do that. Using eq. (8.11) and (8.12) for $\zeta = 0$ we get a return level of

$$\begin{aligned} z &= \alpha - \beta\ln(-\ln G) \\ &= \mu - \gamma\beta - \beta\ln(-\ln G) \\ &= \mu - \sigma\frac{\sqrt{6}}{\pi}(\gamma + \ln(-\ln G)) \\ &= \mu - \sigma\frac{\sqrt{6}}{\pi}\left(\gamma + \ln\left(-\ln\frac{\tau-1}{\tau}\right)\right) \end{aligned} \tag{8.13}$$

which we can compare to $PMP_H = \mu + k_m \cdot \sigma$ to get

$$k_m = -\frac{\sqrt{6}}{\pi}\left(\gamma + \ln\left(-\ln\frac{\tau-1}{\tau}\right)\right) \text{ or } \tau = \left(1 - \exp\left(-\exp\left(\frac{-k_m\pi}{\sqrt{6}} - \gamma\right)\right)\right)^{-1}. \tag{8.14}$$

This approach allows to attribute the results of the Hershfield method to a return period of the best-fit Gumbel distribution. The results seem to be very sensitive to the data used (see Table 8.3). In case of the time series of daily precipitation in Dresden the attributed return period is about 615 years for the observation period from 1917 to 2001 and changes to more than 7400 years if the year 2002 is included.

If we assume that the extremes are GEV distributed, we get the following link between the distribution G and a threshold z expressed as function of μ and σ

$$z = \mu + \sigma \left(\frac{-g_1 + (-\ln G)^{-\zeta}}{\sqrt{g_2 - g_1^2}} \right) \tag{8.15}$$

with $g_k = \Gamma(1 - k\zeta)$. Eq. (8.15) shows that it is not possible to express a quantile of the GEV in the form $z = \mu + k_m\sigma$ without knowing ζ. Therefore it is not possible to attribute a return period of a best-fit GEV to a k_m estimated by the Hershfield method without estimates of the shape parameter ζ. The reason is the higher degree of freedom of the GEV with the variable shape expressed by the shape parameter ζ.

Table 8.3: Means μ and standard deviations σ of best fit GEV and Gumbel distribution of observed time series of yearly maxima of daily rain in Dresden for two different periods.

Distribution	Gumbel		GEV	
Period	1917-2001	1917-2002	1917-2001	1917-2002
$\mu[mm]$	40.7107	41.7128	41.3185	42.599
$\sigma[mm]$	15.094	16.268	18.565	22.130
k_m	4.557	6.502	4.557	6.502
$\mu + k_m\sigma$	109.5	147.5	125.9	186.5
$\tau(k_m)[years]$	615.6	7453.4		

8.6 Is there a maximum standardized rain?

The basic assumption of the Hershfield method is that there is a maximum possible precipitation at each location which can be estimated from observations. This contradicts the assumption that the daily maximum precipitation of an annual series is GEV distributed. Accepting the assumption that maximum rain is GEV distributed leads to the fact that the standardized maxima $k = \frac{x-\mu}{\sigma}$ are GEV distributed as well. This is easily seen since the transformation

$$\frac{k - \alpha}{\beta} = \frac{x/\sigma - \mu/\sigma - \alpha}{\beta} = \frac{x - \mu - \alpha\sigma}{\beta\sigma} = \frac{x - a}{b} \tag{8.16}$$

should only change the scale and location parameter. This contradicts the assumption of Hershfield that there is a maximum k_m.

8.7 PMP from Scaling

Fig. 5.2 shows that the observed maximum rain depth over durations from 1 to 100 days follow a scaling law. A closer look to the figure reveals that there is structure in

the deviations from the scaling line. For some durations D the observed maxima make marked steps compared to durations $D-1$. This can be attributed to the effect that long periods of relatively strong rain are rare and thus undersampled. If we assume that this has no effect on the scaling exponent we can use the largest positive deviation from the scaling line in Fig. 5.2 and draw another line parallel to the original scaling line. This means that we apply the same scaling but anker it to the observed most rainy duration D. The observed rain depth for this duration D is the one that came closest to the PMP in the observation record. Therefore we can use the new scaling line as a lower estimate of $PMP(D)$ for all durations. Fig. 8.4 shows the results for Dresden for the two periods from 1917 to 2001 and 1917 to 2002.

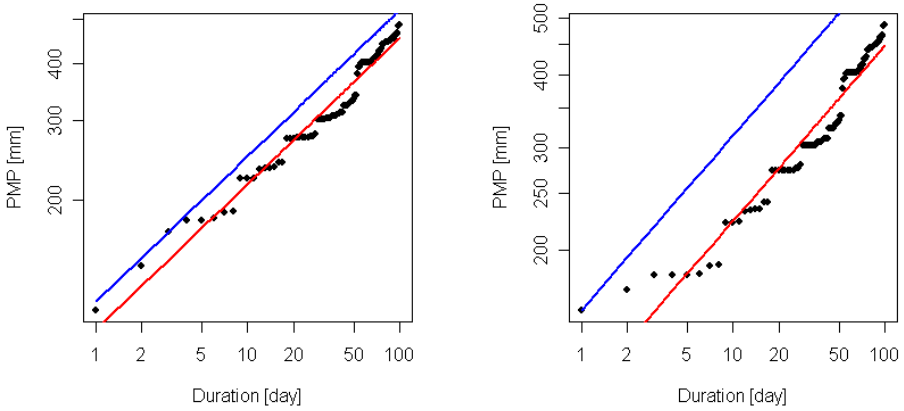

Figure 8.4: Estimates of lower limit of PMP (blue line) vs. duration D based on scaling (red line) of maximum observed rain for durations from 1 to 100 days for observations from 1917 to 2001 (left side) and from 1917 to 2002 (right side).

The resulting estimates of $PMP(D)$ are

$$PMP = \begin{cases} 118.97mm \cdot D^{0.322} & \text{for 1917 to 2001} \\ 158mm \cdot D^{0.299} & \text{for 1917 to 2002.} \end{cases} \tag{8.17}$$

We see clearly that the extreme one-day rain of 2002 changed both, the exponent and of course the estimate of the one-day lower limit of PMP. We also see that taking only observations until 2001 into account the estimated one-day lower PMP limit from scaling is higher than the observation of $114mm$ but lower than the Hershfield estimate of $123mm$.

Instead of ankering the scaling to the observed most rainy duration it can also be ankered to any estimate of the 1-day PMP in order to estimate PMP for any duration of interest.

Chapter 9

This and That

9.1 Monthly Rain and the Weibull Distribution

Many statistical questions can easily be answered if the pdf of a variable is known. If the pdf can be transformed into a Gaußian things become even more handy. Monthly rain observations, however, are not Gaußian distributed. Therefore, often transformations are applied to rain observations x to achieve a variable which is closer to a Gaußian. The most popular transformation is the power transformation $y = x^c$. The exponent c is usually pre-given, e.g. 0.6. Here we show how monthly precipitation observations can automatically be transformed into a distribution that is very close to a Gaußian.

Observed as well as gridded precipitation time series can be closely described by a best-fit Weibull distribution

$$w = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \exp\left(-\left(\frac{x}{b}\right)^{c}\right) \tag{9.1}$$

with the shape parameter c and the scale parameter b (a location parameter a can be introduced if x is replaced by $x - a$). The Weibull distribution with $c = 1$ is the exponential distribution. The transformation

$$y = \left(\frac{x}{b}\right)^{c} \tag{9.2}$$

transforms an arbitrary Weibull distribution into an exponential distribution. This can easily be shown since for a variable transformation from x to y the distribution function changes from W to E according to

$$\frac{dE}{dy} = \frac{dW}{dx}\,\frac{dx}{dy}. \tag{9.3}$$

From eq. (9.2) we get

$$x = b\,y^{1/c} \longrightarrow \frac{dx}{dy} = \frac{b}{c}y^{1/c-1}. \tag{9.4}$$

Inserting this into eq. (9.3) with $\frac{dW}{dx} = w$ yields

$$\frac{dE}{dy} = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \exp\left(-y\right)\frac{b}{c}\left(\left(\frac{x}{b}\right)^{c}\right)^{1/c-1} \tag{9.5}$$

and finally

$$\frac{dE}{dy} = \exp(-y). \tag{9.6}$$

The transformation from the Weibull distribution to the exponential distribution can be applied in both directions. This allows for the transformation of any Weibull distribution into any other Weibull distribution. The necessary coordinate transformation is

$$x_2(x_1) = x_2(y(x_1)) = b_2 y^{1/c_2} = b_2 \left(\frac{x}{b}\right)^{c_1/c_2}. \tag{9.7}$$

A Weibull distribution with $c = 3.60232$ has skewness zero and kurtosis -0.2831404. It is practically indistinguishable from a Gaußian. Therefore eq. (9.7) can be used to transform time series of monthly rain to nearly normally distributed time series. Eq. (9.7) is open with respect to b_2. This means that we can choose b_2 in order to meet another target. This could e.g. be that the mean or the variance of a sample of time series is set constant. For Weibull distributions these moments are

$$\begin{aligned} m &= b\Gamma_1 \\ s^2 &= b^2(\Gamma_2 - \Gamma_1^2) \end{aligned} \tag{9.8}$$

with $\Gamma_i = \Gamma(1 + i/c)$. We see that both moments depend on both parameters. If we choose one b_2 for all transformations all time series have the same mean and variance. However, one might be interested in keeping the local variance, i.e. $s_2^2 = s_1^2$ or

$$b_2 = b_1 \sqrt{\frac{(\Gamma_{1,2} - \Gamma_{1,1}^2)}{(\Gamma_{2,2} - \Gamma_{2,1^2)}}} \tag{9.9}$$

with $\Gamma_{j,i} = \Gamma(1 + i/c_j)$. The transformation for this case results to

$$x_2 = b_1^{1-c_1/c_2} \sqrt{\frac{(\Gamma_{1,2} - \Gamma_{1,1}^2)}{(\Gamma_{2,2} - \Gamma_{2,1^2)}} x_1^{c1/c2}}. \tag{9.10}$$

It transforms a Weibull distributed variable with (b_1, c_1) into a nearly normally distributed variable with the same variance. This transformation is very useful before EOF analysis is done. Sometimes power transformations are done in order to normalize rain data. A traditionally used exponent is 0.6. From the view of this section it transforms a Weibull distribution with $c_2 = 2.161392$ into a distribution with skewness zero.

9.2 Rain and SOC

SOC (self-organized criticality) is a theory about the behavior of systems far from equilibrium. Such systems display scale-free energy releases over a wide range. In 1987 Bak, Tang and Wiesenfeld introduced a model (widely known as BTW model) for avalanches on sand piles [4]. The model is based on a lattice. Each gridbox has a state characterized by an integer between zero and four representing the local gradient of the sand pile surface. During a model run the state of an arbitrary gridbox is increased by one representing additional sand corns. If this leads to a state variable larger than a critical value (4) an avalanche is initiated by reducing the state of this gridbox by

4 but adding 1 to each of the four neighboring gridboxes. In some cases this can lead to surpasses of the critical value in one or more of the neighboring boxes and by that finally can lead to larger avalanches if the system is in a critical state, i.e. receptive to react dramatically to small changes due to the fact that many cells have high values of their state variable.

Some authors analyzed precipitation data on the background of SOC. The idea is that water is constantly evaporating from the earth's surface and moved by atmospheric motion. If the water vapor content in a certain area exceeds a local threshold the area becomes susceptible to precipitation which in the metaphor of the BTW model replaces an avalanche. Ole Peters and Kim Christensen [13] analyzed rain time series observed at the Baltic sea with a temporal resolution of 1min. They counted the number N of events per year with a size M provided in mm rain. They found

$$N(M) \propto M^{\tau}, \text{ with } \tau \approx 1.4 \tag{9.11}$$

for a range of $M_{min} = 5 \cdot 10^{-3} mm$ to $M_{max} = 35 mm$.

They suggested to model the probability of rain $p(R)$ as a function of humidity h minus critical humidity h_c as

$$p(R) = \alpha(h - h_c)^{\beta}. \tag{9.12}$$

9.3 Rain Indices

9.3.1 Standardized Precipitation Index SPI

The standardized precipitation index SPI is introduced by McKee at all in (1993) [12]. It is meant as a means for the comparison of wet or dry phases even across different climate zones and for any duration. In order to achieve this goal the local probability density function of rain is estimated and transformed into a normal (Gaussian) distribution. This allows to provide the deviation of rain from the mean in terms of standard deviation regardless of the original local distribution of rain. Note that the mean of the normal distribution is also its median, which means that the SPI is negative when observed rain is below the median of the original distribution and positive otherwise.

The transformation from the original distribution to a normal distribution can either be performed by fitting a continuous distribution function (e.g. Weibull) or by attribution of Gaussian percentiles to the observed frequency distribution. The SPI can be calculated for any aggregation period of rain and is usually calculated on monthly to biannual scale which allows for the evaluation of regional dry and wet periods on different temporal scales.

Values of the SPI between ± 0.74 are usually called near normal while values ≥ 0.75 are called moderately wet, ≥ 1.25 are called very wet, ≥ 2 extremely wet, and ≥ 3 exceptionally wet. The same thresholds are applied for negative values to define dry states.

9.4 Estimating short-duration maxima from long-duration observations

Hershfield [9] provided simple factors in order to estimate short-duration rain maxima from hourly maxima. According to him the 30min maximum rain is 0.79 of the hourly maximum. He even went to shorter durations providing the factors 0.37, 0.57, and 0.72 for the conversion from 30min maximum rain to 5, 10, and 15 minutes with relative errors of 8, 7, and 5%, respectively.

9.5 Daily vs. 24-hr Data

The maximum rain of time series of daily data should not equal the maximum rain to fall within any 24 hour period. Instead for arbitrary 24 hour intervals higher maximum rain sums should be expected. In fact the analysis by Hershfield [9] gave a 13% higher maximum in observations of hourly data aggregated to arbitrary 24-hour intervals compared to daily data. He found the same factor when comparing 60min aggregates with clock-hourly data.

Appendix A

Observed Daily Rain in Dresden and the Year 2002

Daily rain observations for Dresden are freely available at ECA&D [5] for the period covering the years 1917 to 2011. They are shown in Fig. A.1. In summer 2002 some consecutive days brought high rain amounts in south-east Germany. The record observation for Dresden of 114.1mm in one day was exceeded by about 38.5%. 158mm of rain fell within one day on 12 August 2002.

The time series of daily rain in Dresden offers the opportunity to estimate the risk of a rain event like the 158mm/day given the time series from 1917 to 2001. It furthermore allows analyzing how the estimates of risk change if we consider the observation of the day on which the 158mm were observed. Therefore the time series of daily rain in Dresden is used as a test case for the analyses.

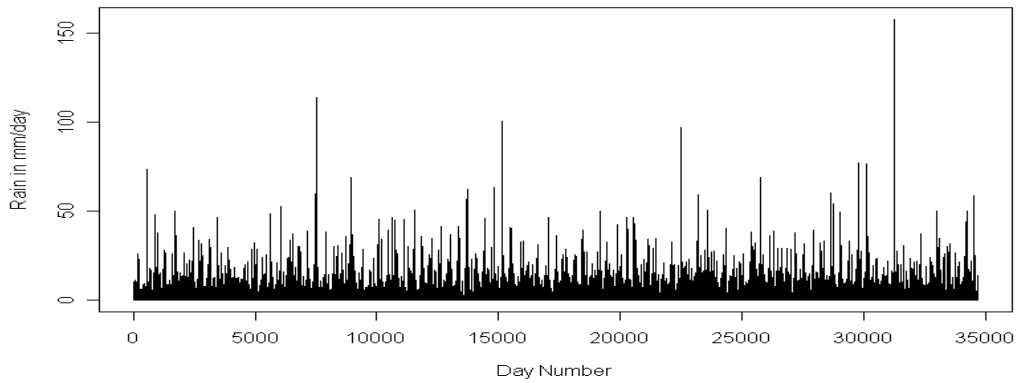

Figure A.1: Daily rain observations from 1917 to 2011.

A classical way to analyze extreme rain is to construct a time series of extremes. Two ways are possible: firstly, one can build the time series of annual maxima (sometimes called annual series), secondly one can build the time series of rain days above a given threshold, the partial series. In the latter case the threshold is usually chosen in a way

that partial series of a time series of n years consists of n extremes, i.e. the n largest observations. Both series of extremes are drawn in Fig. A.2 for comparison.

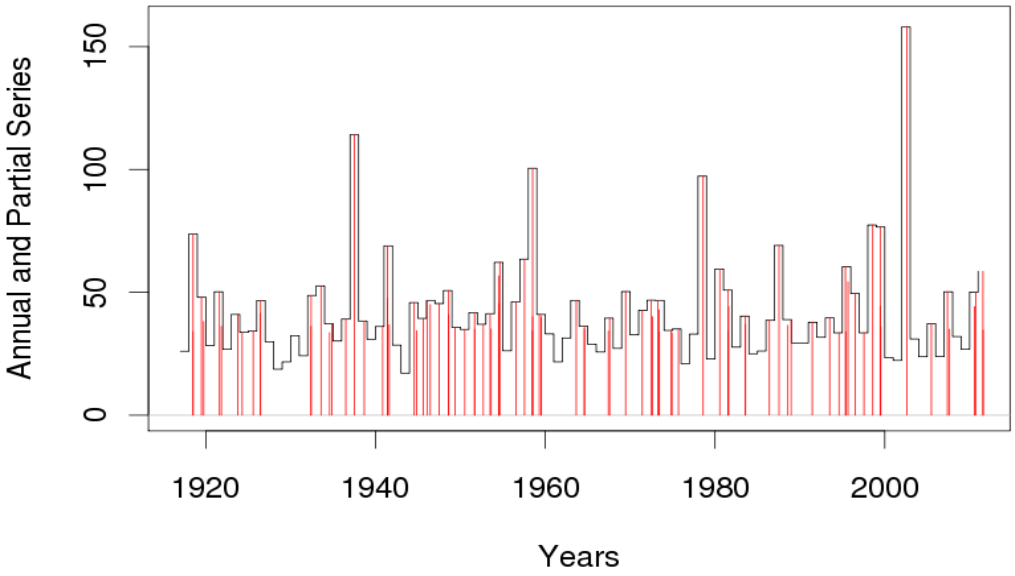

Figure A.2: Annual and partial series of daily rain observations in Dresden from 1917 to 2011.

Both time series consist of 85 observations. However, the partial series is more irregularly distributed while the annual series attributes one observation to each year. The largest value of the annual series is also the largest value of the partial series. However, this need not to be true for any of the other extremes. The annual series can contain lower values than the partial series since there may be years in which none of the n largest observations occur. On the other hand, large values within a year can be hidden by even larger observations within the same year and thus may not be considered in the annual series. Fig. A.3 shows that there is a systematic difference between the annual and the partial series.

The more clustered rain extremes occur the more clustered is the partial series. Consecutive days with extreme rain are usually within one year. On the other hand, several consecutive years may occur without any extreme rain (as defined by the partial series). In the case of the Dresden series we see a three-year gap in the sixties and a five-year gap in the thirties of the last century. Significant clustering can be identified by overdispersion. The number of extreme-rain days of the partial series should be Poisson distributed if they occur independent of each other. Therefor we close this section with a look at the distribution of the partial series per year. Fig. A.4 shows both, the observed and the Poisson distribution. There are no obvious differences and the overdispersion. The rate of the Poisson process is $\lambda = 1$ and therefore the expected mean and variance should be one as well. The observed mean is of course one, since we chose as many extreme events as years of observations by definition of the partial series. The observed variance, however, is 1.0238. This gives an overdispersion of 1.011835 which I consider to be low. The observations give no evidence of clustering of days with extreme rain in Dresden.

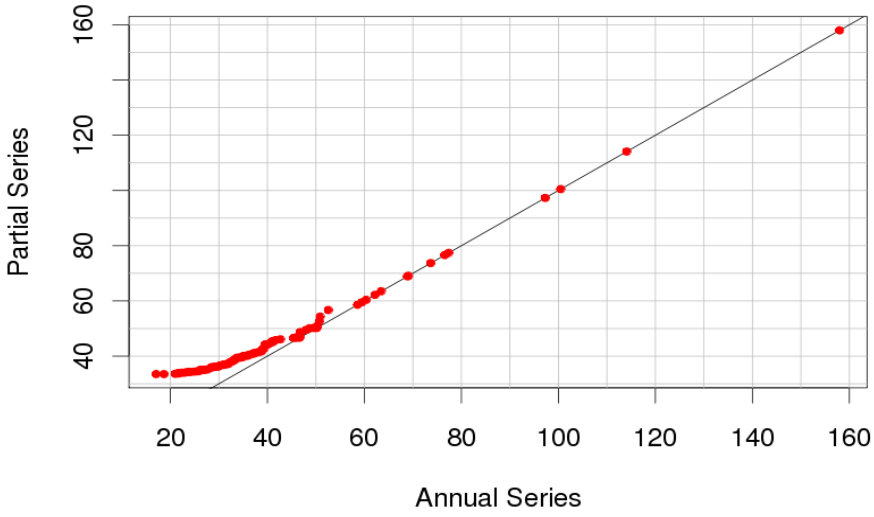

Figure A.3: Annual vs. partial extremes of daily rain observations in Dresden from 1917 to 2011.

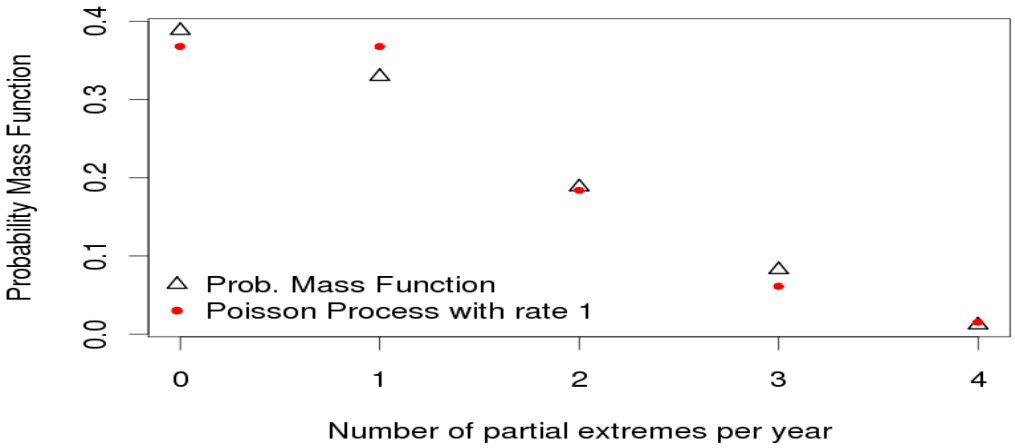

Figure A.4: Distribution of number of extremes in the partial series of Dresden from 1917 to 2011 per year (triangles) and expected distribution (Poisson distribution, red dots).

Appendix B

Estimating Intensity Distributions

There are different ways to estimate the probability function (also called cumulative density function, cdf) from an observed sample. Unfortunately, the results of the different methods can be pretty different, especially in case of very skewed variables like rain. The most widely used method is the maximum-likelihood estimate (MLE) which often needs to be solved numerically. This however is no limitation since usually the necessary subroutines are available. For many distributions the moment estimator (ME) can be used which ensures that some of the moments of the fitted distribution equal the observed moments. As an example, the normal distribution is fully characterized by only 2 moments, the mean μ and the variance σ^2 which can easily be estimated from the observations. The same can be done for the Γ-distribution. A third way of fitting a theoretical distribution to a sample is by transforming the cdf and the observations to get a linear function which can be used to estimate the parameters from linear regression. This approach can be applied e.g. for the Weibull distribution F_W and the Gumbel distribution F_G. For these two cases we get

$$\begin{aligned} \ln\left[-\ln\left(1-F_W\right)\right] &= a\ln x - a\ln b \text{ and} \\ \ln\left(-\ln F_G\right) &= \tfrac{\alpha}{\beta} - \tfrac{x}{\beta}. \end{aligned} \tag{B.1}$$

We abbreviate this approach as transformed linear model (TLM). One of it's advantages is its simpleness. Another is that we can plot the sample and the best fit distribution and by that see systematic deviations from the straight line which helps checking whether the choice of the distribution is justified. A third advantage is that in case of very skewed distributions the transformation brings all values to the same range of magnitude. This justifies the use of least-squares for fitting the straight line. Generally, one needs not to transform the cdf to do a linear regression but can use a nonlinear regression to fit the empirical cdf to a theoretical one. This latter approach is used in section 4.2 to fit a stretched exponential.

Independent of the appraoch for fitting a theoretical to the empirical cdf (the ecdf) we need an estimate for the ecdf from the observations. The cdf is an estimate of the probability that an observed value is not larger than a threshold x. A naïve approach would therefore be to sort the sample of size N and attribute the ecdf $\hat{F(x)} = n(x)/N$

to the n-th smallest observation. However, the disadvantage of this method is that we would attribute the ecdf value 1 to the largest observation claiming that a larger value than the largest observation is not possible. Gumbel proposed the use of $\hat{F} = \frac{n}{N+1}$, which (among other approximations) is widely used.

In case of a sample of limited accuracy only a finite number of values are possible. Rain is often provided in tenth of millimeters. This means that precipitation between 0.1 and x mm can only take one of $10 \cdot x$ values. It is therefore inevitable to observe some values more than once in case of a time series of more observations is to be analyzed. In this case at least one threshold x would get different occurrence probabilities if $\hat{F} = \frac{n(x)}{N+1}$ is applied to each observation. To deal with such bindings, only the largest n for each x is to be used and we use the estimate

$$\hat{F}(x) = \frac{\max(n(x))}{N+1}. \tag{B.2}$$

15306 days with rain larger 0mm are observed in Dresden during the 85 years from 1917 to 2001. 1114 of these days had an observation of 0.1mm, 986 days occurred with 0.2mm rain. We therefore get the first 2 values of the ecdf as 7.28% and 13.72% as the empirical probability that 0.1mm and 0.2mm, respectively, is not exceeded on a rainy day in Dresden.

Finally instead of fitting a cdf to the ecdf one can also fit to a cumulative histogram of e.g. unique class width. This reduces the number of points used for the fit. But why should this make sense? The reason is that this method may reduce the bias that can occur if many small observations but only few large observations exist. In this case a non-classified approach would give much more weight to the many small observations than to the few large ones which in fact may be the ones one like to model best. This approach has its main advantage when the sample is highly skewed. Alternatively to using this simple approach based on classified data one may use the full detailed range of data and attribute weights to the optimization which might be inversely proportional to the data density or alternatively proportional to the observed rain.

From the wide range of opportunities we only compare three here. We fit a Weibull distribution to the observed daily rainfall in Dresden and use a) the maximum-likelihood method, b) the TLM (which is also used to get start values for the numerical estimation with MLE), and c) the TLM fit to the cumulative histogram of the observations using 12 classes of 10mm width each.

Fig. B.1 shows clearly that the simple TLM fit is superior to the MLE even though the TLM fit is used as the starting setting for the optimization. What is the reason for this? As we have seen before, about 14% of the observed values are either 0.1mm or 0.2mm. The MLE results may represent these values better than the TLM. But for large return periods results get worse by the optimization. The best fit is achieved with the simplest method by fitting the cumulative histogram of 12 classes by TLM. Fig. B.2 shows that this does not hold for all seasons.

The use of the transformed linear model allows to plot the best fit as a straight line together with the transformed sample data. Therefore it allows to see whether the model is biased. In this case the data points do not follow the straight line of the model. The Weibull plots for the annual and the four seasonal series of observed rain in Dresden from 1917 to 2011 are shown in figures B.3 and B.4. Although the general agreement

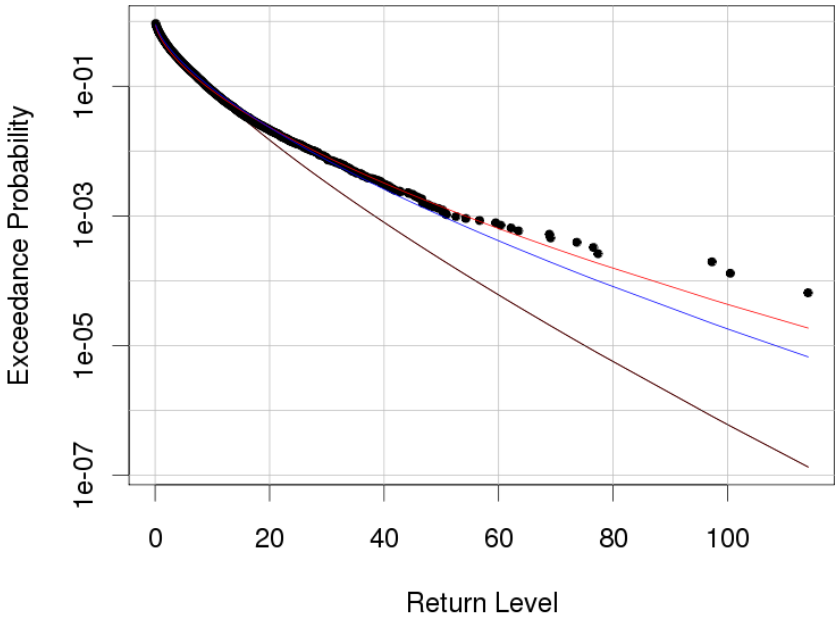

Figure B.1: Exceedance probability of daily rain in Dresden estimated from observations from 1917 to 2011 (black dots) and from different fits of a Weibull distribution: Maximum-Likelihood Estimate (black line), transformation and linear regression of the ecdf according to eq. (B.1) (blue line) and of the cumulative histogram (red line).

looks promising, they reveal systematic deviations in both tails of the distributions. This is problematic since the tails are the most interesting part. It seems that a stretched exponential distribution reproduces the data better (see Fig. 4.3).

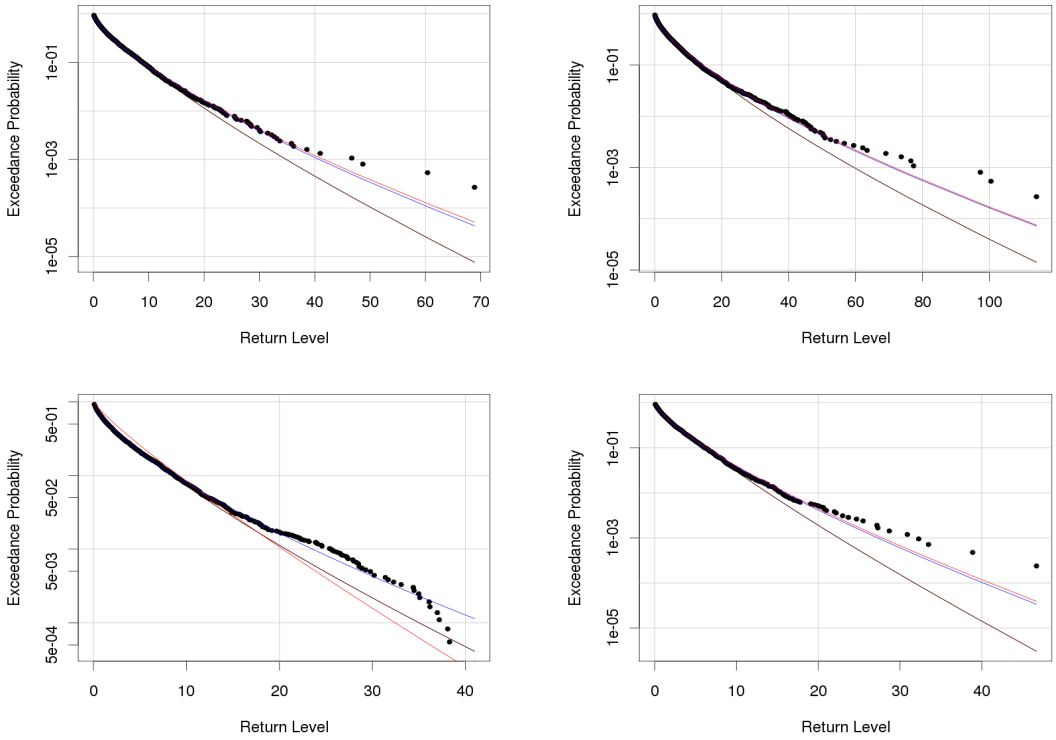

Figure B.2: Seasonal exceedance probability of daily rain in Dresden estimated from observations from 1917 to 2011 (black dots) and from different fits of a Weibull distribution: Maximum-Likelihood Estimate (black line), transformation and linear regression of the ecdf according to eq. (B.1) (blue line) and of the cumulative histogram (red line). Upper left plot: spring, upper right plot: summer, lower left plot: fall, lower right plot: winter.

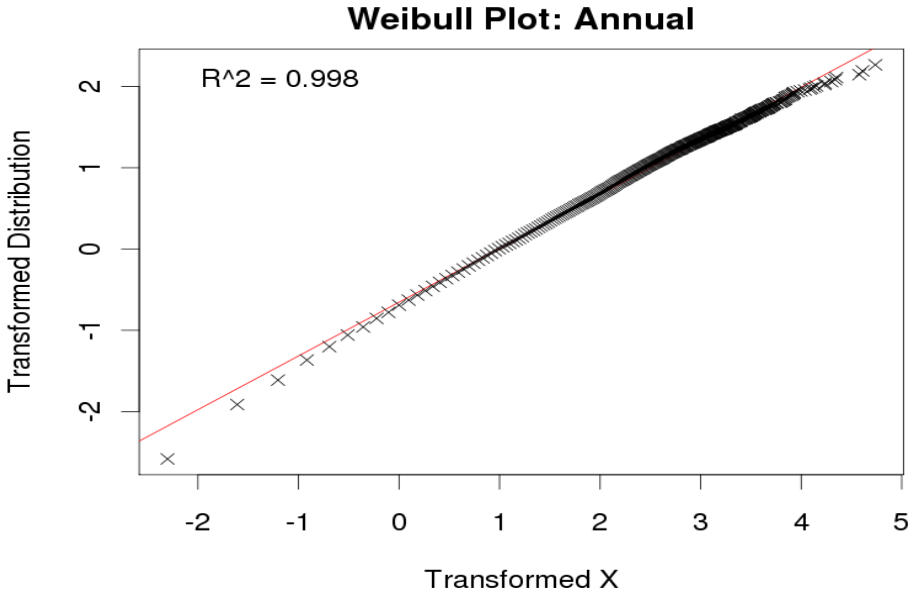

Figure B.3: Weibull plot of observed daily rain in Dresden from 1917 to 2001.

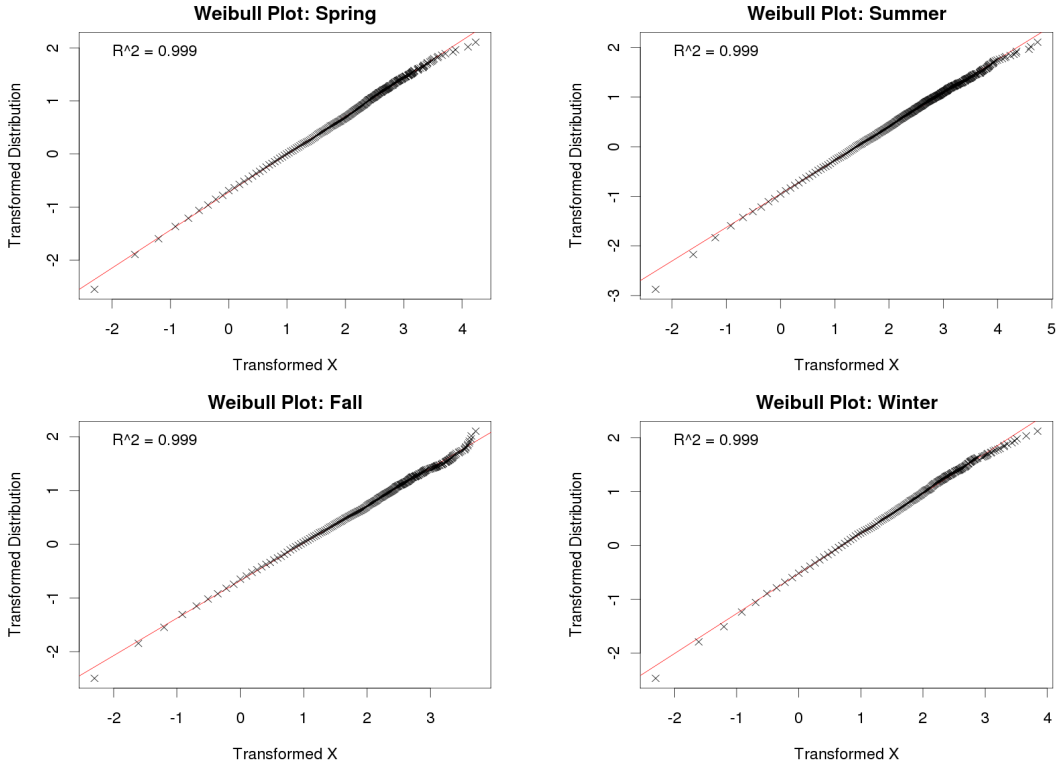

Figure B.4: Weibull plots of observed daily rain in Dresden from 1917 to 2001 for different seasons.

Bibliography

[1] American Meteorological Society (AMS), 1959: Glossary of Meteorology,R. E. Huschke, ed., AMS, Boston, 638 pp.

[2] Balkema, A. and de Haan, L., 1974: Residual Life at Great Ages, Annals of Probability, 2, 792-804.

[3] Benson M. A., 1973: Thoughts on the design of design floods. In: Floods and Droughts (Proc. Second Int. Symp. in Hydrology), 27–33. Water Resources Publications, Fort Collins, Colorado, USA.

[4] Bak, P., Tang, C., and K. Wiesenfeld, 1987: Phys. Rev. Lett. 59, 381.

[5] http://eca.knmi.nl/dailydata/index.php

[6] Galmarini S., Steyn D.G., Ainslie B., 2004: The scaling law realting world point-precipitation records to duration. Int. J. Climatol. 24, 533-546.

[7] Gupta V.K. and E.C. Waymire, 1990: A simple rainfall model for extreme rainfall. Water resource research, 35(1): 335-339.

[8] Hershfield, D.M., 1961: Estimating the probable maximum precipitation. J. Hydraul. Div., Amer. Soc. Civ. Eng., 87, 99-106.

[9] Hershfield, D. M., 1961: Rainfall frequency atlas of the united states for durations from 30 minutes to 24 hours and return periods from 1 to 100 years. Weather Bureau Technical Paper No. 40, U.S. Weather Bureau, Washington D.C.

[10] Hershfield, D.M., 1965: Method for estimating probable maximum precipitation. J. American Waterworks Association, 57, 965-972.

[11] Koutsoyiannis D., D. Kozonis, and A. Manetas, 1998: A mathematical framework for studying rainfall intensity-duration-frequency relationships. Department of Water Resources, national Technical University, Heroon Polytechniou 5, GR-157 80 Zografou, Athens, Greece.

[12] McKee, T.B., N.J. Doesken, and J. Kleist, 1993. The relationship of drought frequency and duration ot time scales. Eighth Conference on Applied Climatology, American Meteorological Society, Jan 17-23, 1993, Anaheim CA, pp. 179-186.

[13] Peters, O., Christensen, K., 2002: Rain: relaxations in the sky. Physical Review E 66: 036120-1-036120-9.

[14] Pickands, J., 1975:Statistical inference using extreme order statistics. Annals of Statistics 3, 119-131.

[15] Pruppacher H.R. and J. D. Klett, 1997: Microphysics of Clouds and Precipitation. Kluwer Academic Publishers, Dordrecht, The Netherlands. 954 pp.

[16] Showalter, A. K. and S.B. Solot, 1942: Computation of maximum possible precipitation, Trans. Am. Geophys. Union 1942, 258–274.

[17] Stephens, D.E., and R.S. Lindzen, 1978: Tropical wave-CISK with a moisture budget and cumulus friction. J. Atm. Sci., 35, 940-961.

[18] Wilson, P.S., and R. Toumi, 2005: A fundamental probability distribution for heavy rainfall. GRL, 32, doi:10.1029/2005GL022465.

[19] World Meteorological Organization, 1986: Manual for Estimation of Probable Maximum Precipitation,2nd ed., Operational Hydrology Report No. 1, WMO No. 332, Geneva, 270 pp.

[20] Woo, G., 2011: Calculating Catastrophe. Imperial College Press, London. 355 pp.

[21] http://www.nws.noaa.gov/oh/hdsc/record_precip/record_precip_world.html