Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

MeteoSwiss

Technical Report MeteoSwiss No. 255, 3rd Edition

MeteoSwiss extreme value analyses: User manual and documentation

Sophie Fukutome and Anne Schindler, Andrea Capobianco (Chapter 5)

ISSN: 2296-0058

Technical Report MeteoSwiss No. 255, 3rd Edition

First Edition: 2015

MeteoSwiss extreme value analyses: User manual and documentation

Sophie Fukutome and Anne Schindler, Andrea Capobianco (Chapter 5)

Recommended citation:
Fukutome S, Schindler A, Capobianco A : 2018, MeteoSwiss extreme value analyses: User manual and documentation, *Technical Report MeteoSwiss*, 255, 3rd Edition, 80 pp.

Editor:
Federal Office of Meteorology and Climatology, MeteoSwiss, © 2018

MeteoSwiss
Operation Center 1
8058 Zürich-Flughafen
T +41 58 460 91 11
www.meteoswiss.ch

Abstract

Extreme value analyses of meteorological quantities, such as precipitation, or temperature, are of great importance in design engineering and risk management. Since 2016, MeteoSwiss has made extreme value analyses of precipitation publicly available for more than 300 stations of its observational network. The present report provides users with a “manual” for understanding and interpreting these analyses; it documents the data used and the statistical methods employed. It is meant for engineers and other experts familiar with basic statistical concepts. The report is structured in several parts. The first chapter describes and explains the contents of the analysis sheets, the second documents the underlying data, and the third is dedicated to the theoretical background. Further chapters are short analyses addressing issues related to the contents of the web platform.

Zusammenfassung

Extremwertanalysen von meteorologischen Grössen, wie Niederschlag oder Temperatur, sind von grossem Interesse für Dimensionierung und Risikomanagement. Seit 2016 stellt MeteoSchweiz Extremwertanalysen von Niederschlag an mehr als 300 Stationen ihres meteorologischen Messnetzes der Öffentlichkeit zur Verfügung. Dieser Bericht liefert eine “Bedienungsanleitung”, um diese Analysen zu verstehen und zu intepretieren und dokumentiert die zugrundeliegenden Daten und Methoden. Er richtet sich an Ingenieure und andere Experten, die statistisches Grundlagenwissen besitzen. Der Bericht ist wie folgt gegliedert: Das erste Kapitel beschreibt und erklärt die Inhalte der Analyse-Blätter. Das zweite Kapitel dokumentiert die zugrundeliegenden Daten und das dritte Kapitel befasst sich mit den nötigen theoretischen Grundlagen. Weitere Kapitel sind kurze Analysen im Zusammenhang mit dem Inhalt der Webplattform.

Contents

1 Introduction

Switzerland has witnessed a number of weather events causing devastating floods in the last decades, as for instance in August 2005, when a large part of the Swiss northern Alpine rim was affected. Such events raise questions regarding adequate planning of infrastructure based on reliable statistical estimates. The latest official documents on the matter were compiled by the Federal Institute for Forest, Snow and Landscape Research (WSL) in the 70s and 80s, and are outdated, both in terms of precipitation data and of statistical methods.

As a federal office operating a broad network of meteorological observation stations, MeteoSwiss has a public responsibility to provide information on the severity of rare events. In a first step, a web page (`http://www.meteoswiss.ch/home/klima/vergangenheit/extrem-niederschlaege.html`) was set up on the website of MeteoSwiss at the end of 2013, presenting extreme value analyses of 1-day, 2-day, and 3-day precipitation at 27 stations of the MeteoSwiss observational network. On the instigation and with the support of the Federal Office for the Environment (FOEN), a more comprehensive platform (`http://climate-extremes.ch`) has been launched in 2016 with analyses at more than 300 (nearly 60) stations for daily (subdaily) precipitation, in which state-of-the-art statistical methods were applied.

The extreme value analyses made available on the web platform include a 2-page sheet in pdf-format. They are intended for engineers and other experts requiring quantitative information for design and risk management. The present report explains how to read and interpret these analyses, and documents the underlying data and methods. The methods employed here are long established, and climatological aspects are not addressed. Thus, the report must be seen as a ”user manual” for a person familiar with basic statistical concepts, wishing to understand the contents of the analyses, and how they came into being. The exact mathematical expressions have been added for completeness in separate boxes, but can be ignored, as they are not essential for understanding the text.

The report is divided as follows. The first part provides a detailed description (chapter 2) of the contents of the analysis sheet. The second is dedicated to the data used (chapter 3), their measurement, eventual manipulation, and quality issues. The third part (chapter 4) describes the theoretical background: the relevant aspects of extreme value statistics, the assumptions that need to be made, and when these may be considered appropriate in this particular context. Finally, the fourth part (chapter 5) compares analyses by *Zeller et al.* (1976-1992), widely used until now, with the results on the MeteoSwiss web platform.

”Graphical” table of contents (pages 9 - 10):
Example of a 2-page extreme value analysis sheet (as appears on the web page).

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Zürich / Fluntern: 556m, 47.38N, 8.57E

Extreme Value Analysis

1-day precipitation, 5:40-5:40 UTC

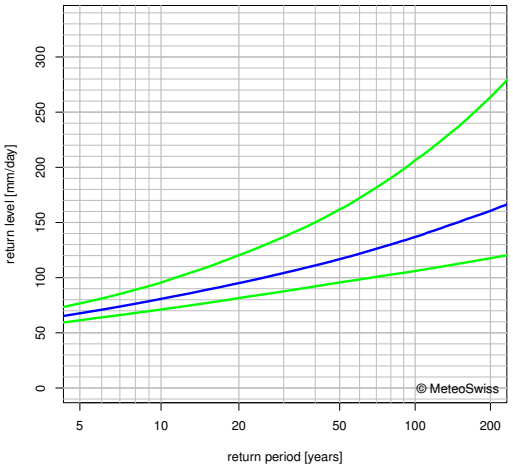
1966 - 2015 (number of missing years: 0)

Bayes (seasonal maxima, GEV). **Reliability of results: good**.

Short summary
Interpretation ..... p. 15

Plot of annual return levels and their uncertainty (ordinate) for a given return period (abscissa).

The estimated return levels are colored blue. The return level 95% confidence intervals are given in green.



Return level plot
Interpretation ..... p. 11
Definition ..... p. 49

Table of the largest annual extrema in the period analysed.

If two large events occur in the same year, only the largest will appear in this table.

date	precipitation [mm/day]	estimated return period [years]
1968-09-21	103.1	29
2007-08-08	97.8	23
1978-08-07	97.2	22
1973-06-23	97.1	22
1999-05-12	94.2	19

Table of annual return levels for a selection of return periods.

The 95% confidence intervals are shown in brackets.

return period [years]	return value [mm/day]	confidence interval [mm/day]
2.33	53.6	(49.5 - 58.6)
5.00	67.7	(61.4 - 76.7)
10.00	80.8	(71.1 - 95.6)
20.00	95.1	(81.5 - 120.3)
30.00	104.2	(87.6 - 136.1)
50.00	116.8	(95.7 - 161.9)
100.00	136.9	(1060 - 206.4)

Operation Center 1, P.O. Box 257, CH-8058 Zurich Airport
Contact: customerservice@meteoswiss.ch
rre150mx; xs_6.5.3; September 13, 2018

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Method information
Documentation ... p. 31

Bayesian Analysis

- Assumption: seasonal maxima follow a Generalized Extreme Value (GEV) distribution.
- Annual return periods, return levels and their confidence intervals are derived numerically from the seasonal posterior predictive distributions.

Parameter Prior Distributions

- Location and scale: uninformative normal distribution.
- Shape: Beta(2,2) on the interval [-0.5,0.5].

Estimation information
Interpretation ..... p. 39

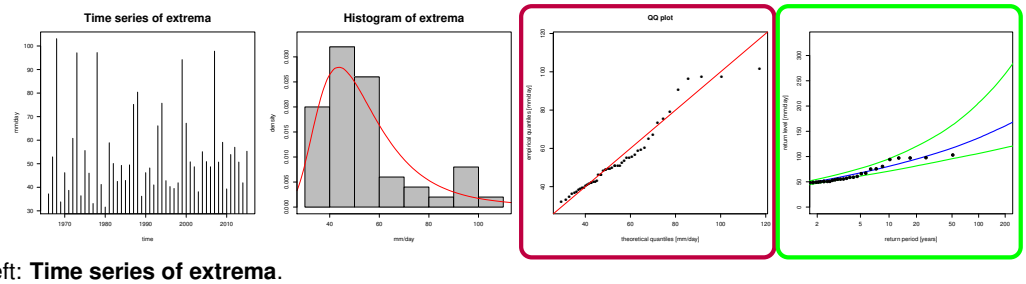
Data and data quality

- The raw data is quality-checked, but not homogenized.
- Missing data: none.

Data information
Interpretation ..... p. 16
Documentation ... p. 21

Reliability
Interpretation ..... p. 13;
Documentation ... p. 44

Additional information



Left: **Time series of extrema**.

Plotting points ..... p. 12

Center left: **Histogram of extrema**. Red line: derived density distribution.

Center right: **QQ plot**. Plot of empirical vs. theoretical quantiles of the observations. The theoretical quantiles are derived numerically from the seasonal posterior distributions. Should the dots align on the diagonal (red line), the fit would be perfect.

Right: **Return level plot** of the annual return levels and their uncertainties derived from the seasonal posterior distributions (blue and green lines, respectively). The black points provide information on the largest observation of each year in the analyzed period. The position of the point on the y-axis corresponds to the measured precipitation sum. The position on the x-axis derives from the number of years within the analyzed period. The return period of the observation can be inferred by extending a horizontal line from the black point to the blue line. The position of the intersection point on the x-axis corresponds to the return period.

Operation Center 1, P.O. Box 257, CH-8058 Zurich Airport
Contact: customerservice@meteoswiss.ch
rre150mx; xs_6.5.3; September 13, 2018

2 Interpretation

The station-based extreme value analysis is presented on two pages. The first page (shown on page 9 for Zurich) states briefly the data employed and the fit quality, and displays the return level plot, a table of return levels for a selection of return periods, and a table of return values for the five yearly maxima in the data series. The second page (shown on page 10) contains more detailed information on the data, the distribution and the estimation.

The extreme value analysis presented on the web platform of MeteoSwiss is based on the premise that the generalized extreme value (GEV) distribution (see section 4) describes the behavior of yearly (or seasonal) maxima of daily precipitation. The estimation procedure is more sophisticated, however, than a conventional, direct estimation of the GEV from yearly maxima. In a first step, Bayesian inference is applied to seasonal maxima, then yearly information is reconstructed from the seasonal analyses. Thus, the interpretation of the analyses on pages 9 and 10 presented below departs from the interpretation of a straightforward analysis of yearly maxima.

2.1 Return level plots

Return level plots (figure 1) display the probability that a given value is exceeded. The probability is expressed in terms of years. Thus a value that has a probability of 1% of being exceeded in any given year is - in the (very) long term average - expected to be exceeded once in a 100 years. Then, the value, or "return level", is said to have a "return period" of 100 years. In fact, return levels are extreme quantiles: the 100-year return value has a probability of 0.99 of *not* being exceeded, and therefore corresponds to the customary 0.99-quantile (see figure 2).

To focus the attention on the behavior of the rarest events, return periods are transformed in such a way that the relation to the return levels will appear as a straight line for a Gumbel distribution (see section 4) of the yearly maxima.

Figure 3 illustrates the meaning of the return levels from an empirical point of view with 100 yearly maxima. Sorting the yearly maxima by their magnitude yields an empirical quantile function. Transforming them produces an empirical return level plot. The transformation squeezes together the central part of the distribution and stretches the extremities.

In real life applications, the true probability distribution describing the behavior of extremes is unknown, and must be estimated from the observations. Due to the finite size of the sample, this estimation is inherently uncertain, but the uncertainty can be quantified, and translates to uncertainty of the estimated return levels (green lines, figure 1).

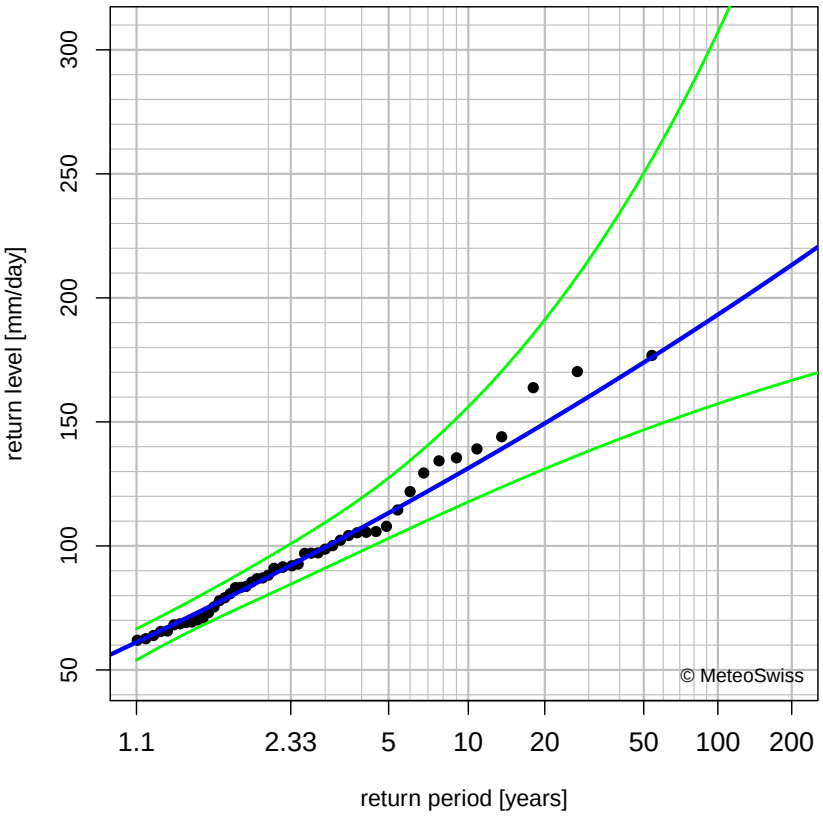


Figure 1: Return level plot of 1-day precipitation at station Lugano for the period 1961-2013. The blue line represents the return level estimates based on the estimated GEV-distribution; the green lines represent their 95% confidence interval; and the y-coordinate of the black points represent data used for estimation of the GEV-distribution; the plotting position (x-coordinate) of the points is solely determined by the length of the underlying record (and is not robust).

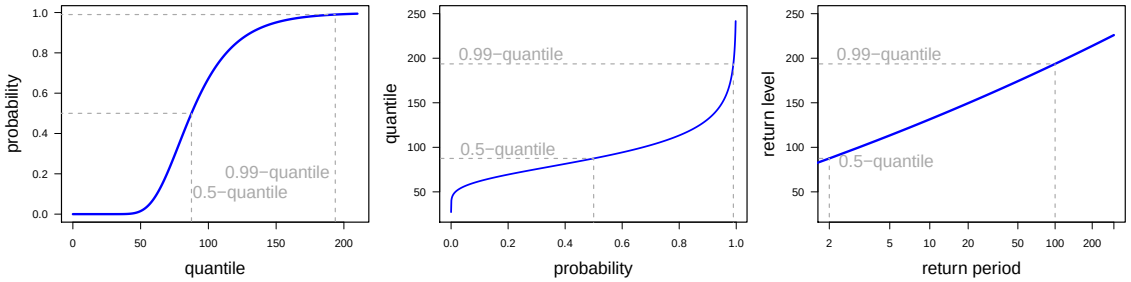


Figure 2: Correspondence between cumulative probability distribution function (CDF, left), quantile function (center) and return level plot (right). The distribution shown here was estimated for Lugano 1-day precipitation extremes for the period 1961-2013 (cf. figure 1).

Remark. It is important to keep in mind that the plotted uncertainties only describe the uncertainty in return **level** estimates. No statement can be made regarding an uncertainty in return **period** based on the uncertainties represented in figure 1.

Plotting points The return level plot sometimes displays so-called "plotting points" (black points, figure 1). The position of these plotting points on the y-axis corresponds to the values of the observed sample maxima. Since the true return period of these events is unknown, however, the position on the x-axis is determined empirically from the sample size using the Weibull estimator (see definition 6, section 4.3.1). Thus, for a sample size of 50 years, for instance, the largest observation will be assigned a return period of approximately 50 years, the second largest a return period of approximately 25 years, and so on. In other words, it does not represent the local behavior of the quantity of interest (e.g. daily

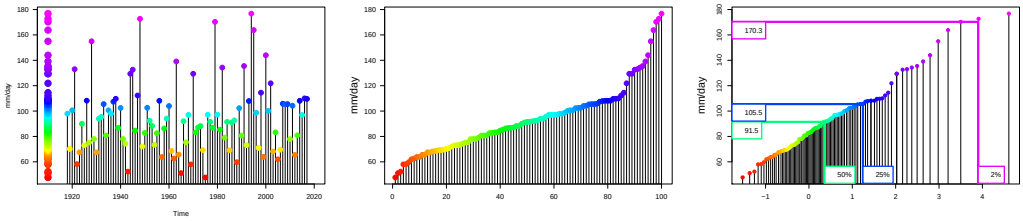


Figure 3: Yearly maxima of daily precipitation at station Lugano, 1918-2017 (100 years). Left: yearly maxima ordered chronologically (left) with rainbow colors from red to magenta as they increase. Center: sorted in ascending order; can be understood as an empirical version of the quantile function in fig. 2, center. Right: sorted in ascending order, and transformed as in a return level plot. There are 2 (25;50) measurements of daily precipitation above 170.3 mm/day (105.5 mm/day; 91.5 mm/day). Since we have 100 values in total, this means 2 (25;50) in 100 or 1 in 50 (4; 2) values. Thus 170.3, 105.5, and 91.5 mm/day can be taken as 50, 4, and 2 year "empirical" return levels, respectively. Of course, this is only one sample: we would need a large number of samples to verify that these values truly correspond to the return levels. Note also that, on this empirical view, the return level is not unique, as, for instance, any value between the third and the second largest would have had the same number of values above it.

precipitation) at the particular station under consideration, since the return period would be the same regardless of the geographical location of the station. Some conclusions can, however, be drawn from the comparison between plotting points and best estimate (blue line, figure 1): any substantial or systematic disagreement between empirical and GEV estimates, after allowance for sampling error, suggests an inadequacy of the GEV model.

For a more thorough analysis of the fit, it is preferable to examine the agreement between empirical and modeled estimates in a quantile-quantile plot, which is easier to interpret correctly (see sec. 4.3.1).

On the product presented on the web platform (`http://climate-extremes.ch`), exemplified on pages 9 and 10, the return level plot is *not* a transformed cumulative distribution function as in fig. 2. The return levels depicted in blue are derived from the return levels of seasonal analyses (see sections 4.5 and 4.2.3), the description of which was not included. Nonetheless, the plot is read in exactly the same way as described in this section.

2.2 Reliability

Each extreme value analysis is tested for its reliability, and the verdict is stated on the analysis sheet. This verdict does not pertain to the inherent uncertainty of the estimation due to the limited sample size. Rather, it describes the possibility that the assumptions motivating the choice of the statistical model may not be fulfilled. If the reliability is poor, for instance, the model does not represent well the observed extreme values. Three degrees of reliability have been defined: poor, questionable, and good. Their meaning and what to do about them is shown in the following table.

Remark. "The only justification for extrapolating an extreme value model is the asymptotic basis on which it is derived." (*Coles*, 2001, page 3)

"However, if a model is found to perform badly in terms of its representation for extreme values that have already been observed, there is little hope of it working well in extrapolation." (*Coles*, 2001, page 3)

Verdict	Interpretation for the user	Action
poor	Observed extreme values are **badly** represented by the model.	Do not use this statistical model! Use largest observed events or closest reliable station instead.
questionable	Observed values are **not well** represented by the model. Underlying assumptions might be violated.	Careful assessment necessary. Use visual guides (sec. 4.3.1) to decide: if points line up - proceed as usual; if not - contact klimainformation@meteoswiss.ch.
good	Observed extreme values are **well** represented.	The statistical model can be used.

The QQ plot presented on page 10 may not suffice to judge the quality of the fit, as the estimates are reconstructed from seasonal estimates (see sections 4.5 and 4.2.3), hence the recommendation to seek advice from klimainformation@meteoswiss.ch in cases of questionable estimation quality.

2.3 Example stations

In the following, we illustrate the interpretation of a typical extreme value analysis product for two stations, Genève-Cointrin (abbreviation GVE) and Zürich / Fluntern (abbreviation SMA), that are subject to different climatic conditions in terms of heavy precipitation, and therefore lend themselves well to highlight concepts and problems in interpreting extreme value analysis. The period of analysis for these examples is 1961-2013.

Geneva is located at the south-western tip of Switzerland, in a flat plain, between the Jura mountains and the foothills of the Alpine arch, while Zurich lies in the north-eastern part of the Swiss Plateau. Heavy precipitation is generally due to summer thunderstorms on the background of a weather configuration characterized by moist, south-westerly flow. Geneva is comparatively dry because of its protected position in the lee of the Jura mountains. The seasonal cycle of daily precipitation is very modest in Geneva, with the largest monthly maxima in September; in Zurich on the other hand, it is very pronounced and peaks in August.

The differences in precipitation regime between Genève-Cointrin and Zürich/Fluntern are manifest in the statistical characterization of their extremal behavior. Average heavy precipitation is of similar magnitude at both stations: we can expect events of 50mm per day to be exceeded on average every other year, with similar ranges of deviation from this average. Yet, they differ in the behavior of the very rare events, and thus illustrate two regimes of extremal behavior.

2.3.1 Genève-Cointrin

Here, we refrain from displaying the extreme value analysis product for Genève-Cointrin, as its form is the same as for Zürich/Fluntern on pages 9 and 10. All plots, tables, and explanatory text relevant to the discussion will be shown as separate figures.

Short summary To define the product at hand, the first few lines provide succinct information regarding the data and statistical method (figure 4):

- The station name, here **Genève-Cointrin**, its altitude (**411m**) and geographical coordinates (**46.25N, 6.13E**).

Genève / Cointrin: 411m, 46.25N, 6.13E
Extreme Value Analysis
1-day precipitation, 5:40-5:40 UTC
1961 - 2013 (number of missing years: 0)
Bayes (seasonal maxima, GEV). **Reliability of results: good**.

Figure 4: Short summary of extreme value analysis.

- The MeteoSwiss product name, here **Extreme Value Analysis**.
- The variable analyzed, here **1-day-precipitation**, recorded daily at **5:40 UTC**, observed from **1961 to 2013 with no missing years** (see section 3).
- The extreme value approach employed: **Bayes**ian inference based on block maxima data, where **seasonal maxima** are assumed to follow a **generalized extreme value** distribution (see section 4).
- A summary of the reliability of the results based on the quality of the fit, here **good** (see section 4.3.2 and page 49).

Return level plot The return level plot summarizes the extremal behavior at a glance (figure 5). A range of 1-day-precipitation amounts (y-axis) is plotted against their exceedance probabilities expressed in terms of average number of years (x-axis).

The return level estimate (blue line) is curved upwards. This means that regardless which amount of precipitation we are looking at, the probability that a larger one occurs is non-zero; the probability of larger and larger amounts becomes smaller and smaller, and the rate at which this probability decays at GVE is relatively fast. The fastest possible decay that never reaches a probability of zero would be exponential, and would look like a straight line on the return level plot. Indeed, the return level plot has purposely been transformed in such a way that an exponential decay looks like a straight line.

The green lines on the return level plot (figure 5) show the 95% confidence intervals. Note that these confidence intervals only describe the uncertainty of the return levels, expressed in mm.

The strong positive curvature of the upper confidence bound (green line, figure 5) indicates that this extremal behavior is subject to uncertainty: a much slower decay of the exceedance probability is also quite possible. This would imply that very severe events would have a higher probability of being exceeded than is described by the blue line. The curvature of the lower confidence bound is difficult to see. It could be a straight line, or be very slightly negatively curved, which would correspond to a bounded distribution. A bounded distribution is therefore possible but very unlikely. A negative curvature would suggest that there exists a daily precipitation amount above which any amount has a zero probability of being exceeded.

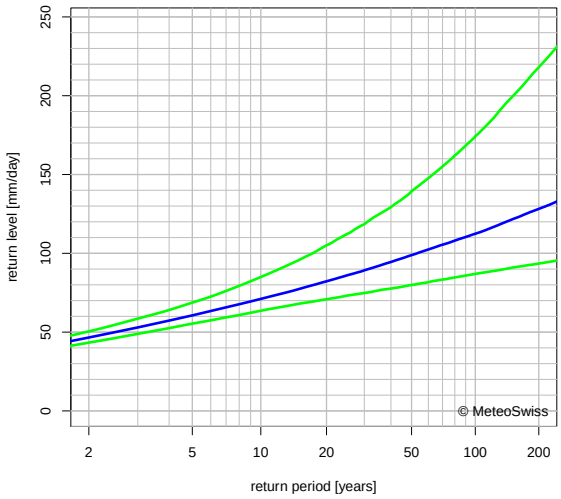


Figure 5: Return level plot of 1-day precipitation in Geneva for the period 1961-2013. The blue line represents the return level estimates; the green lines represent their 95% confidence intervals.

The blue line appearing on the return level plot is *not* a GEV. It was calculated based on the return levels of the seasonal maxima (see section 4.5). For this reason, there are no GEV parameters with which to describe extreme behavior succinctly.

Tables The table of the largest annual extrema and the table of return levels contain the same information as the return level plot. The first (table 1, left) "reads" the return periods corresponding to the 5 largest maxima, the second (table 1, right) "reads" the precipitation amounts corresponding to given return periods.

date	precipitation [mm/day]	estimated return period [years]
2002-11-14	92.6	36
1993-09-09	85.2	24
1978-08-07	76.1	14
1975-09-14	75.7	14
1999-09-25	72.9	11

return period [years]	return value [mm/day]	confidence interval [mm/day]
2	46.6	(43.3 - 50.5)
5	60.6	(55.4 - 68.7)
10	71.1	(63.6 - 84.1)
20	82.2	(70.9 - 105.1)
30	89.2	(74.8 - 118.5)
50	98.8	(79.9 - 139.3)
100	1124	(86.9 - 174.2)

Table 1: Extreme value analysis at Genève-Cointrin (GVE) over the period 1961-2013: largest 1-day precipitation events with estimated return period (left) and return level estimates for a selection of return periods (right).

Data information The time series of annual maxima (figure 6, left) informs visually about missing data or possible trends or cycles. Here, there could be a decadal oscillation or a trend towards higher values at the end of the record, although it may be entirely by chance that the two largest events occur in the second half of the period of analysis. Further inquiries are therefore necessary to rule out random effects. The histogram (figure 6, right) shows the frequency with which the annual maxima have been recorded, and can be understood as the empirical probability density function (PDF). The red line is calculated from the seasonal analyses: it is *not* a GEV.

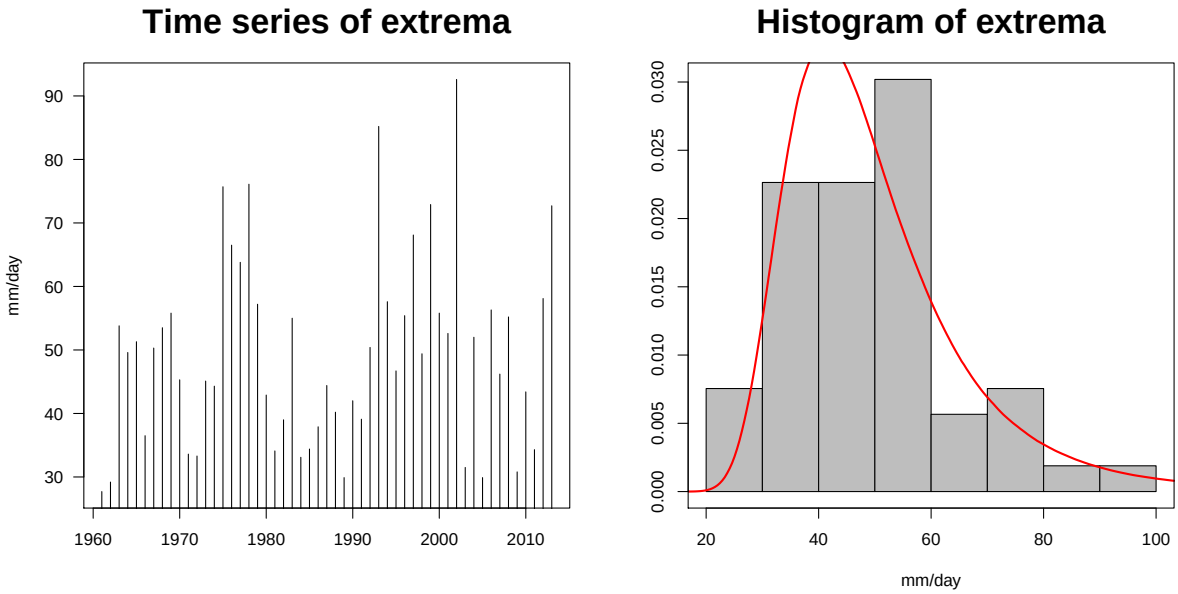


Figure 6: Time series (left) and histogram (empirical density) (right) of annual maxima with estimated density function (red line) taken from the extreme value analysis of 1-day precipitation at station Genève-Cointrin for the period 1961-2013.

Reliability Genève-Cointrin is an excellent example for the problems experienced when applying extreme value statistics: although the fit is declared as "good", the QQ-plot of the yearly maxima (fig. 7) displays an odd pattern for values that are not really extreme, i.e. up to about 60 mm/day (and return periods below 10 years). In addition, the very few high values seem systematically a little off, and suggest that we overestimate the high return levels, especially if we extrapolate beyond the range of observed values.

As it turns out, on the QQ-plots for each (separate) seasonal analysis (not shown) the dots closely follow the red line, indicating quite reliable fits. The pattern we observe on the yearly QQ-plot may therefore be the result of an unfortunate grouping of seasonal maxima in certain years. Thus, despite the somewhat questionable aspect of the yearly QQ-plot, the analysis is reliable and can be used for extrapolation at unobserved levels.

Remark. What have we learned:
At Genève-Cointrin, the probability of very extreme precipitation decays polynomially with the severity of events. Statistically speaking, the data is consistent both with a much slower decay and with an exponential decay (faster) of the tail of the distribution. On the other hand, a bounded distribution is very unlikely. Thus, any precipitation amount, however high, has a non-negligible probability of being exceeded (there exists no upper bound).

The largest 1-day precipitation event (92.6mm) recorded between 1961 and 2013 was November 14$^{\text{th}}$, 2002, which corresponds roughly to a 90-year event.

The reliability of the fit is good. Eventually, further investigation of the seasonal analyses underlying the present results could clear any suspicions regarding long-term trends. All in all, extrapolation to unobserved levels up to a reasonable return period compared to the length of the data set (perhaps 200 years) is justifiable.

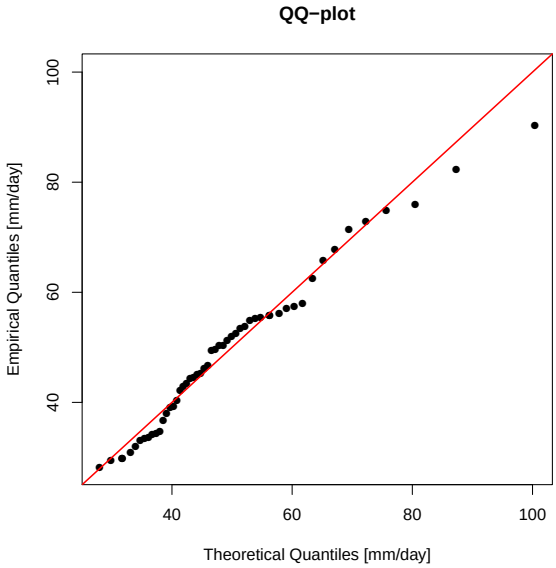


Figure 7: Quantile-quantile plot of 1-day precipitation at station Genève-Cointrin for the period 1961-2013.

2.3.2 Zürich / Fluntern

The extreme value analysis for 1-day precipitation observed at Zürich / Fluntern during the period 1961-2013 is shown on pages 9 and 10. Here, we will highlight differences to the statistical and extremal behavior of the extreme value analysis for Genève-Cointrin.

Reliability As at Genève-Cointrin, the statistical model in Zürich / Fluntern is judged as “good”. This means that any mismatch between model and observations can be attributed to chance. Yet, the yearly QQ plot (fig. 9) would suggest a different regime for the lower values. Up to nearly 70 mm/day (i.e. return periods below 5 to 10 years), the empirical quantiles increase at a systematically slower rate than the theoretical ones, indicating a steeper empirical cumulative distribution function in that area. The four largest values on the yearly QQ plot hint at a systematic difference, but a glance at the histogram casts doubt on this conclusion, as these 3 values could simply be nearly equal by accident. Also, we should keep in mind that measurement errors can never be completely ruled out.

The seasonal QQ plots (not shown) show no irregularities for lower values, but display the same odd pattern for the highest values. On the other hand, none of the seasonal time series (not shown) indicate anything worthy of notice. Thus, a closer look at the seasonal analyses clears doubts regarding the quality of the estimates at low and medium levels, but the evidence remains inconclusive for the range of the highest values. A larger sample would certainly be helpful.

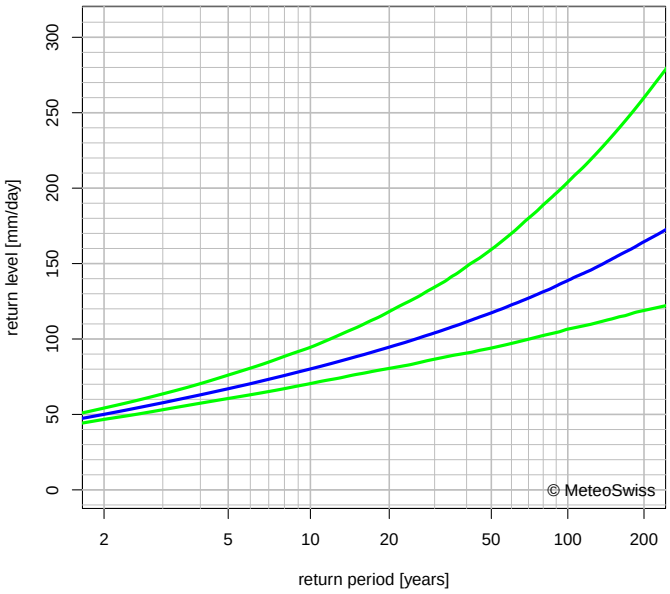


Figure 8: Return level plot of 1-day precipitation in Zürich for the period 1961-2013. The blue line represents the return level estimates, and the green lines the 95% confidence intervals.

Remark. At Zürich / Fluntern, the probability of very extreme precipitation decays polynomially with the severity of events, and much more slowly than at station Genève-Cointrin. This means that any precipitation amount has a larger probability of being exceeded than at Genève-Cointrin. The largest event in Geneva, which had an estimated return period of approximately 90 years, would have a return period below 20 years at station Zürich / Fluntern. Here too, the uncertainties are large, but it is very unlikely that the distribution should be bounded or even have an exponentially decaying tail.

The largest 1-day precipitation event between 1961 and 2013 was recorded on September 21st, 1968 and amounted to 103.1mm. Three events of similar magnitude (97.8mm, 97.2mm, 97.1mm) were recorded 08/2007, 08/1978 and 06/1973.

The reliability of the fit is good, but the visual guides for reliability revealed some cause for concern that could not be completely dispelled by a close examination of the seasonal analyses. Thus, caution should be exercised when extrapolating return values to unobserved levels. In particular, consideration of the confidence intervals is recommended.

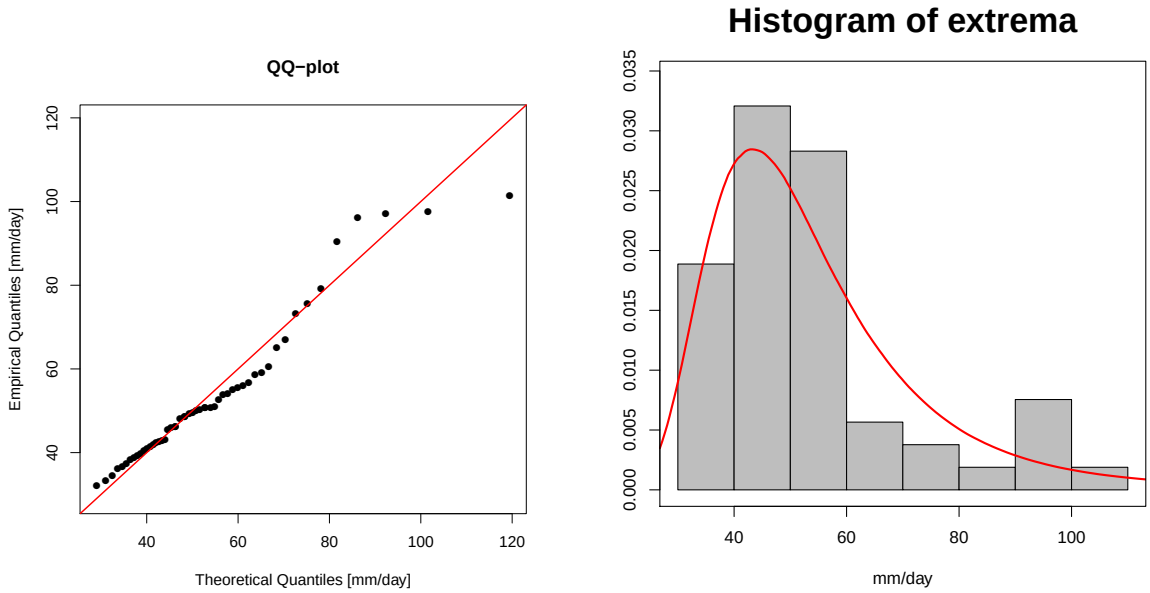


Figure 9: Quantile-quantile plot (left) and histogram (right) of 1-day precipitation at station Zürich / Fluntern for the period 1961-2013. The red line on the histogram represents the probability density function (PDF) calculated from the seasonal analyses. It is *not* a GEV.

3 Data

3.1 Stations and Analysis Period

The data used for the extreme value analyses on the web platform http://climate-extremes.ch come from in situ measurements at stations of the MeteoSwiss observational network. In chapter 5, the extreme value analyses presented on http://climate-extremes.ch are compared with the analyses by *Zeller et al.* (1976-1992). The stations selected for this comparison are part of the Swiss National Basic Climatological Network (NBCN), which groups the most important climatological stations within the observational network of MeteoSwiss (*Begert et al.*, 2007). Most of these stations have very long records: data at some of them date back to the 19th century. The NBCN stations are spread out over Switzerland, and should be to some extent representative of the regional climatic variations in monthly precipitation. The subset selected for the comparison in chapter 5 is shown in fig. 10.

The MeteoSwiss observational network encompasses different types of stations. NIME stations observe precipitation only, and measurements are carried out manually, at 7:30 local time every day. Climate Stations and SwissMetNet stations measure a wide range of meteorological parameters. Today, there are no more manual Climate Stations in operation: they were gradually changed to SwissMetNet stations, the last one in 2013. At SwissMetNet stations, precipitation is measured automatically in 10-minute intervals, starting on the hour. At Climate Stations, manual precipitation measurements took place daily at 6:00 and 18:00 UTC. Note that for manual measurements, there is a time window of approximately 70 minutes for the observer to carry out the reading. The NBCN stations used for the comparison in chapter 5 were all Climate Stations until conversion to SwissMetNet stations.

Only SwissMetNet stations can provide precipitation data at durations shorter than one day. We will call these durations “sub-daily” and, by extension, refer to such stations as sub-daily stations. All stations that yield daily precipitation data will be referred to as “supra-daily” stations, as precipitation data for durations of one to several days can be derived. The vast majority of stations of the MeteoSwiss observational network can be classified as supra-daily. The earliest sub-daily measurements took place in 1981, while some supra-daily stations were already in use in 1864.

In addition to customary quality standards, extreme value analysis requires data to cover a large number of years for its estimates to offer any credibility, which limits the number of adequate stations. Since the sub-daily and supra-daily measurements cover different periods, the selection criteria differ accordingly. The station selection criteria are listed in section 3.3. They resulted in a set of over 300 supra-daily stations (fig. 11) and nearly 60 sub-daily stations (fig. 12).

Clearly, the coverage for supra-daily data is quite comprehensive. Nevertheless, it is uneven both in horizontal (see for example the empty area in the Ticino) and vertical extent. Only 10% of the

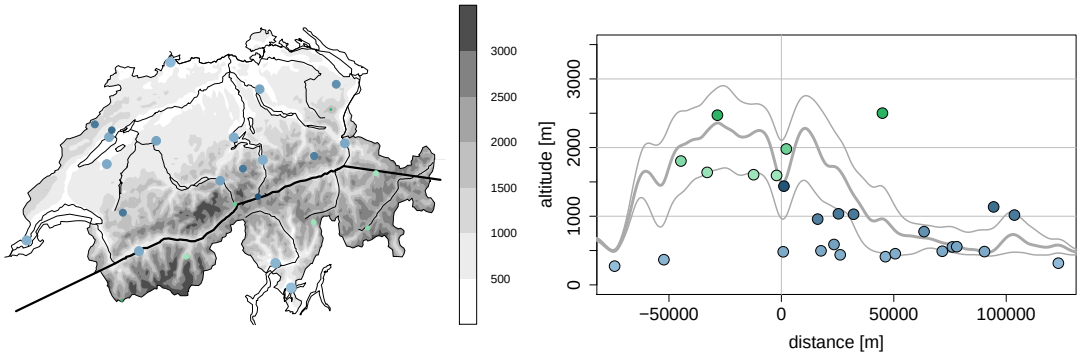


Figure 10: Map (left) and vertical cross-section (right) of NBCN station locations. The list of station names, geographical location, and altitude can be found on page 69. The vertical cross-section across the Alpine ridge is calculated as the distance to the **inner-Alpine valleys (black line on the map)**, defined as the Rhone and the beginning of the Rhine valleys, connected to each other and prolonged on either side by straight lines. To the east of the Rhine valley, the straight line is made to pass through the station Scuol. The **northern (southern) Alps** are defined here as to the north (south) of the inner-Alpine valleys. The cross-section of the Alpine topography is computed with the USGS GTOPO30 (http://eros.usgs.gov) digital elevation model, and is the minimum air-line distance of each grid-point center to the inner-Alpine valleys. The thick (thin) grey line(s) represent(s) the smoothed median (10% and 90% quantiles) of the altitude of all Swiss grid-points in a 100m bin of distances. Note that all bins do not contain the same number of grid-points. In fact, at the furthest distance to the south, there are so few grid-points per bin that the altitude quantiles all give the same value.

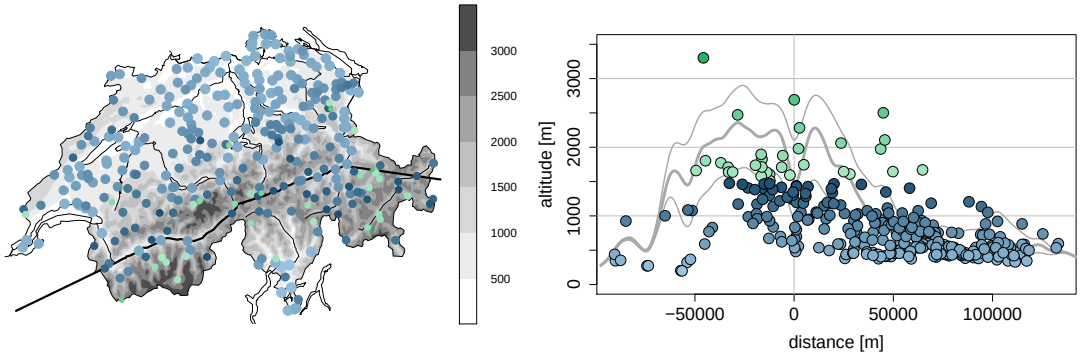


Figure 11: supra-daily stations used on the web platform (July 2017). Blue (green): stations below (above) 1500m. The size (color) of the dots decreases (intensifies) with altitude.

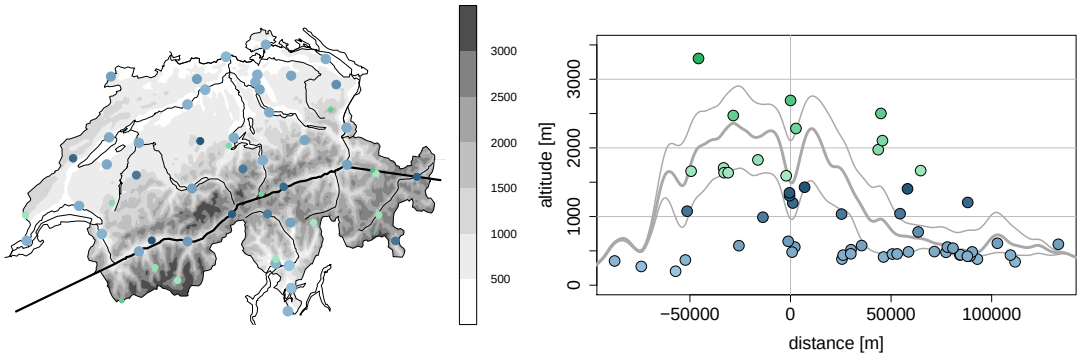


Figure 12: sub-daily stations used on the web platform (July 2017). Blue (green): stations below (above) 1500m. The size (color) of the dots decreases (intensifies) with altitude.

stations are above 1500 m (compared to 28% of the surface of Switzerland), and 60% of these are in the southern Alps (see fig. 10 for the precise definition of inner-Alpine valleys and southern/northern Alps). The great majority of stations (nearly 80%) are on the northern side of the inner-Alpine valleys (compared to 70% of the surface of Switzerland), and at least 60% are located in the Plateau, Jura, and northern Alpine rim, at more than 30 km to the north of the inner-Alpine valleys (compared to 52% of the surface Switzerland).

In the inner Alps, within a 30km ribbon on either side of the inner-Alpine valleys (36% of Swiss land), only one station (Weissfluhjoch) is above the 90^{th} percentile of the altitudes of all grid-points within a distance bin, almost all (92%) are below the median, and a considerable amount (nearly 60%) are below the 10^{th} percentile. While it would be very useful to have better coverage in elevated areas, we should keep in mind that measuring precipitation in mountainous terrain remains a challenge, both because of the necessity for regular human intervention, and because of difficulties in maintaining adequate measuring conditions (for instance, strong winds may prevent rain drops from entering the receptacle).

The extreme value analyses on the web platform `http://climate-extremes.ch` are based on two time periods, each tailored to different user's needs: the Standard Period, covering 50 years, and the Longest Period, extending over the longest possible period, but with at least 40 available years of data within the most recent period of 50 (completed) years.

Standard Period: For users who need to qualify a precipitation event in terms of its "extremeness" or compare the results at different stations (e.g. annual reports on the return period of events in the vicinity of a barrage, reports on precipitation events with impacts at regional or national level, expert opinions, claims), the web platform provides return values and return periods estimated for the standard period. A common time period is necessary for comparing return periods of events regardless whether they took place during the standard period. It is also necessary for comparing return values at different stations. The standard period has a length of 50 years, and is shifted forward in time by 5 years every 5 years.

Longest Period: For users who need a return value at a specific location or from a specific data series, the web platform provides return values estimated for the longest possible period with available data. While these estimates are based on more information, they may suffer from a long-term trend, as non-stationarities have been ignored.

Estimating return levels with respect to different time periods inevitably leads to different results. In addition, both the return values and the goodness-of-fit vary from standard period to standard period in case (a), and from year to year in case (b) due to the annual update. At the Lugano station (LUG), for example, the 100-year return value for 1-day precipitation events is 16 mm/day lower if the standard period (1966-2015) is used instead of the longest possible period (1864-2015), and at the Schaffhausen station (SHA) it is 22.5 mm/day higher. These differences are not statistically significant, however. To ensure consistency, MeteoSwiss will use the results estimated for the standard period for the qualification of precipitation events.

Different analysis periods are only available for supra-daily precipitation events. For sub-daily precipitation, the time period used is prolonged by one year every year until it reaches a coverage of 40 years and can be considered as a standard period. Until then, "Standard Period" and "Longest Period" will

display the same results for sub-daily precipitation events.

3.2 Data quality

The measuring instruments of the MeteoSwiss observational network have changed over time (table 2). The first ombrometers measured precipitation by means of an open cylindrical receptacle with a horizontal cross-section of 400cm^2. In the early 20th century, these were replaced with Hellmann rain gauges with a horizontal cross-section of 200cm^2. Both of these systems require manual reading. At the beginning of the 1980s, about two thirds of the MeteoSwiss stations were equipped with automatic (heated) tipping bucket rain gauges. Since 2012, a further automatic ombrometer has been introduced that uses a weighing mechanism that is also used to partly automate NIME stations (see section 3.1).

measuring instrument	time period	reading time
ombrometer 400 cm^2	19th / early 20th century	6.00 UTC or 7.30 local time (m)
Hellmann rain gauge 200 cm^2	early 20th century till today	6.00 UTC or 7.30 local time (m)
tipping bucket rain gauges	since 1980s	every 10min., starting on the hour (a)
weighing precipitation gauge	starting 2012	every 10min., starting on the hour (a)

Table 2: Precipitation measuring instruments of the MeteoSwiss observational network with manual (m) and automatic (a) reading.

All of these ombrometers have their weaknesses, and precipitation measurements suffer from both systematic and non-systematic errors. The Hellmann rain gauges require manual reading at regular times, and sometimes the measurements are carried out at incorrect reading times. If the reading time has been skipped, an accumulated value is provided for the following day, and the amount of precipitation belonging to the proper time interval is unknown. The tipping bucket systems systematically underestimate (overestimate) precipitation at high (low) intensities, while the weighing rain gauges occasionally yield ghost precipitation due to pressure fluctuations caused by gradients in temperature. The latter errors are, at the time of this writing, being adressed by correcting the data during the quality assessment procedure. The consequences of these measurement errors for the analysis of extremes is not always straightforward, and we should keep in mind that they inevitably contribute to the uncertainty of the results.

3.3 Heavy precipitation data

The supra-daily data analyzed on the web platform `http://climate-extremes.ch` consists of 1 day precipitation and of 2, 3, 4, and 5 day sliding sums of daily precipitation. The sub-daily data analyzed consists of 10 minute precipitation and of 20 and 30 minute sliding sums of 10-minute precipitation; as well as of 1 hour precipitation and 2, 3, 4, 6, 8, 12, and 16 hour sliding sums of hourly precipitation.

When building sliding sums of 2 to 5 day precipitation, all missing days but 1 are ignored. For instance, one missing daily observation is treated as zero precipitation and the two 2 day sums it should contribute to correspond to precipitation on the previous and on the following day. Thus, it will not lead

to two missing entries for 2 day precipitation. At sub-daily stations, daily precipitation is the sum of all available 10 minute measurements, provided there are no more than 24 missing values, of which no more than 12 are consecutive. Sub-daily sliding sums handle missing values differently before and after 2004, when daily verification of 10 minute data quality was introduced. From 1981 to January 2003, up to 1 missing 10 minute value is ignored for all sliding sums. Since February 2004, up to 1, 2, 2, 3, 4, 5, and 8 missing 10 minute values are ignored in 2, 3, 4, 6, 8, 12, and 16 hour sums.

For extreme value analysis with the block maxima approach (see section 4 for details), only the maximum value per year is considered, and thus the sample size corresponds to the number of years of observation available. At MeteoSwiss, all yearly maxima are computed from daily precipitation values, be it the precipitation sum for supra-daily durations, or the daily maximum of a sub-daily precipitation sum. The computation of a yearly maximum ignores up to 22 missing values, provided no more than 11 of them are consecutive.

In turn, the daily maximum of the sub-daily precipitation durations accepts no more than 1 missing entry (if, for instance, two 1-hour or two 6-hour sums - consecutive or not - in a day are missing, the daily maximum will be missing). Up to January 2004, this pertained to 10 minute values as well, because the daily maximum 10 minute precipitation was computed from the hourly maximum 10 minute precipitation. Since February 2004, the daily maximum 10 minute value is computed directly from the quality-checked 10 minute data, and ignores up to 9 missing 10 minute sums per day, provided no more than 5 are consecutive.

Since long records are essential for the credibility of extreme value analyses, a minimum number of available yearly maxima must be imposed. For supra-daily data, 40 years of data must be available in the most recent 50 years of the analysed period (Standard or Longest Period). For sub-daily data, a minimum of 25 years is required.

The selection criteria for the stations are summarized below for version 2016 of the web platform.

Remark. Data requirements on version 2016 of `http://climate-extremes.ch`.

- Standard Period / Longest Period: at least 40 (25) years of observations in the period $1966-2015$ $(1981-2015)$ for supra-daily (sub-daily) data.
- quality checked
- still operated today
- non-homogenized data
- well-documented station history

Station abbreviation	1-day record (date)	1-day record (amount) in [mm]	2-day record (date)	2-day record (amount) in [mm]	3-day record (date)	3-day record (amount) in [mm]
ALT	31.07.1874	147.2	31.07.1874	194.7	04.05.2002	204.3
ANT	04.07.1916	185.3	03.05.2002	224.4	16.11.2002	269.5
BAS	25.05.1872	94.8	26.05.1872	139.4	26.05.1872	150.8
BER	14.08.2010	90.3	03.10.1888	139.4	03.10.1888	141.4
CDF	25.08.2002	100.8	19.01.1910	170.4	20.01.1910	206.7
CHD	28.12.1947	96.9	19.01.1910	167.3	20.01.1910	207.6
CHM	25.09.1987	102.9	26.09.1987	175.8	27.09.1987	175.8
DAV	18.11.1874	95.0	19.11.1874	143.0	20.11.1874	198.0
ELM	29.08.1890	147.3	04.07.1891	187.4	15.02.1990	205.6
ENG	31.07.1874	153.8	31.07.1874	226.5	01.08.1874	226.5
GRC	02.11.1968	170.3	15.10.2000	198.4	15.10.2000	226.8
GRH	21.12.1991	151.4	22.12.1991	236.2	22.12.1991	292.8
GSB	12.11.1996	159.6	12.11.1996	273.2	13.11.1996	321.5
GVE	14.11.2002	92.6	09.09.1993	111.3	09.09.1993	129.2
LUG	21.08.1911	262.8	21.08.1911	342.5	22.08.1911	368.4
LUZ	06.06.2002	111.8	06.06.2002	123.3	27.07.1976	133.2
MER	22.08.2005	110.9	22.08.2005	205.0	22.08.2005	218.0
NEU	08.10.1949	116.1	26.09.1987	125.7	20.01.1910	126.1
OTL	26.09.1991	317.9	22.09.1981	377.8	23.09.1981	416.4
PAY	26.09.1987	78.4	26.09.1987	106.8	27.09.1987	106.8
RAG	29.08.1890	175.0	29.08.1890	217.0	31.08.1890	241.0
SAE	22.08.2005	186.7	03.11.1921	290.0	04.11.1921	321.4
SBE	07.08.1978	175.5	15.11.2002	258.5	16.11.2002	367.7
SIA	03.11.2000	108.0	15.11.2002	159.3	16.11.2002	223.0
SIO	14.02.1990	79.3	14.02.1990	138.5	15.02.1990	170.2
SMA	11.06.1876	171.5	12.06.1876	244.5	12.06.1876	272.5
STG	01.09.1881	250.0	02.09.1881	309.1	03.09.1881	338.9

Table 3: NBCN stations with more than 40 available years of daily precipitation data (up to 2010) with the records (in mm) of 1-day, 2-day, and 3-day precipitation.

4 Theoretical background: Extreme Value Statistics

Univariate extreme value statistics offers three different approaches to describe extreme events. The approaches differ in the way they characterize extreme events: via the largest/smallest values; via the excess of a pre-defined threshold value; or, more generally, via the count of events falling into a pre-defined set of values. They are called block maxima (BM), peaks over threshold (POT) or point process (PP) approach, respectively. In the following we will describe the BM approach closely following *Coles* (2001).

Here, we first present some general properties of the generalized extreme value distribution (GEV), the "extremal" distribution that - so extreme value theory tells us - represents the behavior of the maxima. Then, we explain different methods to infer this distribution from the data at hand, and assess the goodness of the estimated fit. Finally, we close the section with the notion of return levels, a means to illustrate the behavior of extreme values.

4.1 The GEV distribution

Using the maxima of a block of observations to characterize the behavior of rare events finds its justification in extreme value theory. *Fisher and Tippett* (1928) and *Gnedenko* (1943) prove that the distribution of the largest values of any[1] random variable converges towards one of three types of limit distributions, and that these three types are the only possible limit distributions. This law is known as the extremal types theorem or Fisher-Tippett-Gnedenko theorem. Strictly speaking, the maxima must first be appropriately re-scaled for the limit distribution to be non-degenerate (see *Coles*, 2001, p.46 for details).

[1] see remark 4.1 and *Leadbetter et al.* (1983)

Remark. The extremal types theorem - as does the central limit theorem for sample means - justifies the practice of approximating the distribution of sample maxima via extreme data only; thus, it is not necessary to estimate the population's distribution F in order to describe the behavior of the population's rare events.

Notation. Given a sequence of independent random variables $X_1, \ldots, X_n$ with a common distribution function F, we denote with $M_n = \max\{X_1, \ldots, X_n\}$ the maximum of the sequence. For example, M_{365} would be the annual maximum of daily observations for a block of one year.

In this chapter, we use the following conventions:

- random variables are symbolized by capital letters, for instance Z, or Θ.
- observations, i.e. realizations of these random variables are symbolized by small letters, such as z, or θ. For instance, $z = (z_1, \ldots, z_m)$ could be the m yearly maxima of daily precipitation.
- The probability that z occurs is expressed, for continuous probability distributions, as $Pr\{Z \leqslant z\}$, the "probability that Z does not exceed z". The function describing this quantity for all z is the Cumulative Distribution Function (CDF), and is symbolized by a capital letter, such as $F(z)$, or $G(z)$ for the GEV distribution.
- The probability density function (PDF) is the first derivative of the CDF and can be understood as a weighting function, large for values that occur frequantly, small for values that are rare. It is designated by small letters, such as $f(z)$, or $g(z)$ for the PDF of the GEV distribution. When "some" probability density function (not necessarily GEV) is meant, we will use $p(z)$.

Details. *Let M_n be a sequence of maxima, a_n and b_n appropriate re-scaling sequences of constants and G the limit distribution of $(M_n - a_n)/b_n$, then the probability that M_n does not exceed a given value z is given by*

$$\Pr\{M_n \leqslant z\} \approx G^*(z)$$

with G^* another member of the GEV family.

In practice, we do not need to estimate the re-scaling sequences, a_n and b_n, but only the distribution G^*.

Remark. Note that the observations from which we take the maximum (e.g. daily precipitation) should not display seasonal behaviour within a block. If they are not independent (e.g. precipitation on a given day depends on precipitation on previous days), the series should be stationary with very weak dependence between large observations (e.g. of daily precipitation) that are very far apart (see *Coles*, 2001, section 5.2 for details).

In addition, block sizes must be sufficiently large for the block maximum to converge towards the limit distribution. In particular, if $X_1, \ldots, X_n$ is not independent (but stationary), an even larger block may be necessary for convergence to be achieved.

Remark. For certain distributions F, a non-degenerate limit distribution function does not exist; this is the case, for instance, for the Poisson and geometric distribution, which both describe discrete counting processes. For such distributions, the maxima cannot be described by a GEV.

The three types of limit distributions are the Weibull, Gumbel, and Fréchet families of distributions. They can be summarized in one parameterization of the limit model - in exchange for an additional parameter (*von Mises*, 1954; *Jenkinson*, 1955). This generalized limit distribution, known as the GEV distribution (definition 1), will be used for the subsequent analyses.

Definition 1 (GEV distribution)**.** The family of distributions defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$ via

$$G(z) = \exp\left(-\left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}\right)$$

with $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \xi < \infty$, is called the generalized extreme value family of distributions (GEV distribution). The parameter μ is a location parameter, the parameter σ is a scale parameter and ξ is a shape parameter. The subset of the GEV family with $\xi = 0$ is interpreted as the limit as $\xi \to 0$.

Example (GEV distribution)**.** Examples of

Weibull ($\mu = 45$, $\sigma = 11$, $\xi = -0.1$),
Gumbel ($\mu = 45$, $\sigma = 11$, $\xi = 0$), and
Fréchet ($\mu = 45$, $\sigma = 11$, $\xi = 0.25$)

probability density functions. These distributions and their samples will be used throughout the document to illustrate extremal behavior, estimation behavior, uncertainties, and exceedance probabilities.
They loosely resemble the 1-day-accumulated precipitation annual maxima distributions at Fribourg (not shown), Genève, and Zürich-Fluntern, respectively.

The cumulative distribution function (CDF) and probability density function (PDF) of a Weibull (red), a Gumbel (black) and a Fréchet (blue) distribution are shown in figure 13. All three have a location parameter $\mu = 45$, and a scale parameter $\sigma = 11$. They differ by their shape parameter ξ, which is negative (in this example $\xi = -0.1$) for a Weibull, zero for a Gumbel, and positive (in this example $\xi = 0.25$) for a Fréchet distribution (see generalized parameterization of definition 1).

Figure 14 shows the PDFs of the three example distributions seen in figure 13, here in separate plots, with the histograms of the respective samples appearing in figure 15. Figure 15 shows random samples (of equal length) drawn from these three distributions: these can be imagined as time-series of annual maxima of daily precipitation.

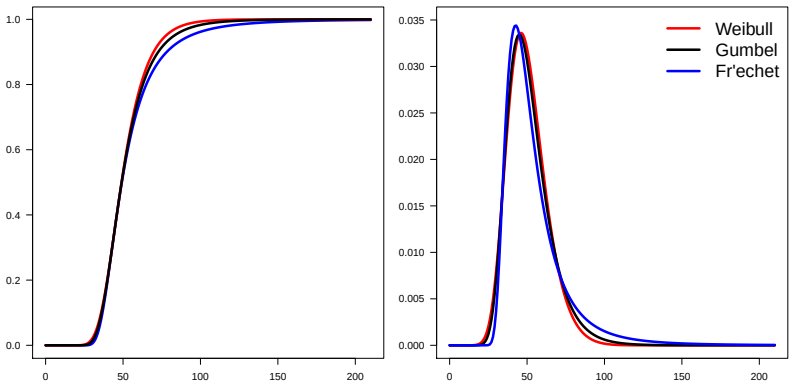


Figure 13: Three extreme value distribution functions (left: cumulative distribution function (CDF); right: probability density function (PDF)) belonging to the Weibull (red; $\xi = -0.1$), Gumbel (black; $\xi = 0$), and Fréchet (blue; $\xi = 0.25$) distribution families, respectively. All three have the same location parameter ($\mu = 45$) and scale parameter ($\sigma = 11$).

Remark. The three different families of distributions are distinguishable by their tail behavior, and in particular by the rate of decay in the tail:

- The Weibull family, i.e., the subset of GEV distributions with $\xi < 0$, has a finite upper end-point (figure 13 red line - also called type III or bounded). Thus, there is 0 probability for values above that end-point to occur.
- The Gumbel, family, i.e., the subset of GEV distributions with $\xi = 0$, has an infinite upper end-point. The tail of the probability density distribution decays exponentially (figure 13, black line - also called type I or light-tailed). Thus, regardless the value, there is a non-vanishing probability for a higher value to occur.
- The Fréchet family, i.e., the subset of GEV distributions with $\xi > 0$, has an infinite upper end-point. The tail of the density distribution decays polynomially (figure 13, blue line - also called type II or heavy-tailed). Here too, whatever the magnitude of the event, a larger one has a non-negligible probability to occur.

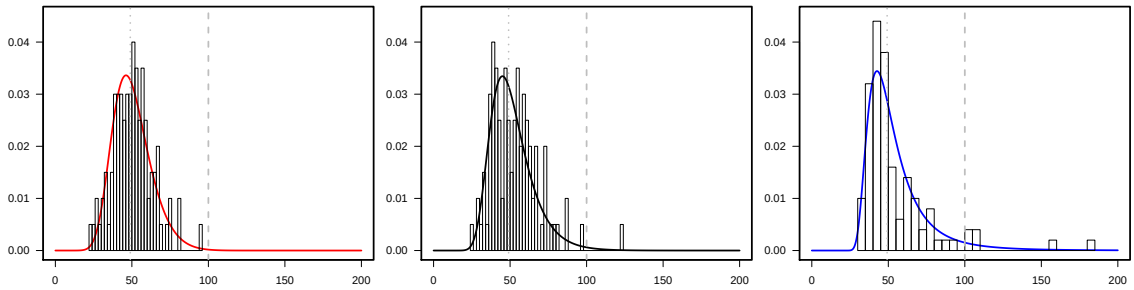


Figure 14: Probability density functions of figure 13 (solid lines) of the Weibull (left, red, $\xi = -0.1$), Gumbel (center, black, $\xi = 0$), and Fréchet (right, blue, $\xi = 0.25$) distribution, with a location parameter of 45 and a scale parameter of 11. The underlying histograms (with 50 breaks) show the empirical densities of samples drawn from these three distributions. The samples can be viewed in the form of the fake time series in figure 15. The vertical dotted line represents the sample median, while the dashed vertical line at 100 is meant to ease comparison between panels, and with figure 15.

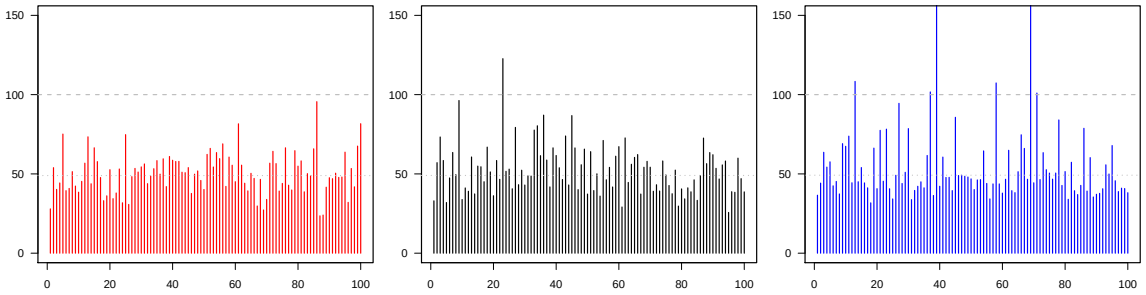


Figure 15: Three samples of length 100 drawn from the Weibull (left, red, $\xi = -0.1$), Gumbel (center, black, $\xi = 0$), and Fréchet (right, blue, $\xi = 0.25$) distribution, with a location parameter of 45 and a scale parameter of 11. The horizontal lines indicate, as in figure 14, the sample median (dotted line) and the value 100 (dashed line).

4.2 Inference for the GEV distribution

Inference is - one could say - an educated guess of the distribution, based on the available sample.

According to Karl Pearson (1857-1936), observations are only a random manifestation of a deeper, but abstract reality, the probability distribution of our quantity of interest (*Salsburg* (2001)). This revolutionary view took shape towards the end of the 19th century, when, despite an increase in measurement accuracy of physical instruments, observations appeared to maintain a degree of randomness. However precisely we measure, we cannot predict the exact outcome of an individual measurement. In other words, the true quantity we are looking for is invisible to our eyes. In fact, it is not even a fixed value, but a probability distribution characterized by fixed parameters; the nature of the randomness we witness is encapsulated in these parameters. They cannot be measured (para-meter = almost-measurement in Greek), but must be inferred from the observations and the way they scatter.

Along with this revolutionary view came the frequentist definition of probability proposed by John Venn in 1872: if we can run identical trials an illimited number of times, the proportion of times an event occurs ***is*** the probability that this event occurs (*Salsburg* (2001)). Together, these ideas form the foundation of frequentist inference. This is of consequence for the interpretation of confidence intervals, an important aspect of inference. The 95% confidence interval estimated from our particular observed sample either will or will not contain the true value, and we do not know which of the two. In fact, we do not even know the probability of finding the true value within the confidence bounds. The 95% probability statement pertains to the chance, in a very large number of hypothetical, similar experiments (parallel universes with similar climates, perhaps), that one of the calculated intervals actually contains the truth; or, to formulate it in a frequentist way, it pertains to the proportion of these hypothetical intervals that will contain the truth.

Bayesian inference views things from another angle. The parameters of the distribution are random, and are conditional on the observed measurements (*Salsburg* (2001)). This means the parameters have a probability distribution, which represents their uncertainty. We ***propose*** a probability distribution for each parameter, ***measure*** the quantity we are interested in, then ***update*** the probability distributions of the parameters accordingly. Uncertainty is explicitly taken into account from beginning to end, and

confidence intervals (or credible intervals in Bayesian jargon) can be formulated in such a way that they contain the ”true” value with a given probability. It is interesting to note that having the results in the form of a probability distribution is very convenient for decision making, which is generally not based on the value of the parameter itself, but on its consequences for some other quantity (often a currency).

The extreme value analyses presented today on the web platform (`http://climate-extremes.ch`), are carried out within the Bayesian framework. Previous versions of the analysis relied, however, on frequentist inference. In the following, we discuss several approaches of parameter estimation, both within the frequentist and the Bayesian inferential frameworks.

Several frequentist methods exist to estimate the underlying GEV distribution from the data *at hand*. Here, we present two of them: the Maximum Likelihood (ML) method and the method of L-moments. Bayesian inference, which can be seen as as an extension of Maximum Likelihood estimation, is explained in section 4.2.3.

In each section, we discuss first parameter, then uncertainty estimation. In the frequentist framework, the only source of uncertainty that we can quantify is the one due to the finite size of the sample of observations, compared to the hypothetical, infinite population. The uncertainty then essentially describes a range of values that would also be consistent with the data at hand. For some frequentist estimation methods, as is the case for Maximum Likelihood estimation, the quantification of uncertainties is a consequence of the theoretical foundation. For others, such as L-Moments, an alternative can be found in bootstrapping methods.

Remark. As was customary at the time, the approach of *Zeller et al.* (1976-1992) was to first opt for one of the extremal distribution families, and then estimate the scale and location parameter using regression methods.

In practice, this meant selecting a particular sheet of paper with the corresponding pre-drawn graphon it, either with a log or a log-log scale. Thus, the estimation hinges on the preliminary (educated) guess of the appropriate family.

“But”, as pointed out by (*Coles*, 2001, p. 47) “there are two weaknesses: first, a technique is required to choose which of the three families is most appropriate for the data at hand; second, once such a decision is made, subsequent inferences presume this choice to be correct, and do not allow for the uncertainty such a selection involves, even though this uncertainty may be substantial.”

With the adoption of the generalised extreme value distribution parametrization, “the uncertainty in the inferred value of ξ measures the lack of certainty as to which of the original three types is most appropriate for a given dataset.“ (*Coles*, 2001, page. 48)

Remark. There exist several sources of uncertainty, not all of which can be quantified:

- the data quality, measurement errors, or the representativeness of the data
- the model choice (GEV distribution)
- the violation of model assumptions, such as that the maxima are independent and identically distributed (i.i.d.), or that the selected blocks are sufficiently long to ensure convergence toward the GEV
- the estimation of model parameters
- the interpretation of probabilistic results (stochastic uncertainty / sampling uncertainty)

Notation. To avoid multiple indices we denote with Z_i the maximum M_n of block i, and omit the information on the block length n. You can imagine $\{Z_1, \ldots, Z_m\}$ as the set of m annual maxima of daily precipitation.

In the Bayesian framework, the parameters of the GEV distribution are random variables. Let Θ denote the (vector) random variable for the three GEV parameters, and $\theta = (\mu, \sigma, \xi)$ one particular realization of Θ.

4.2.1 Maximum likelihood estimation

Maximum likelihood (ML) estimation looks for the (distribution) parameters that are most likely to have led to the sample at hand, assuming that the largest values of the process of interest are indeed described by the chosen distribution family, here the GEV distribution.

The likelihood that an arbitrary set of GEV parameters led to the observations can be written as the probability that all of these observations $\{Z_1, \ldots, Z_m\}$ should occur jointly. Letting this probability depend on the set of GEV parameters and checking whether the likelihood increases or decreases as the parameters vary allows us to look for the set of parameters most likely to have produced our observations $\{Z_1, \ldots, Z_m\}$.

The likelihood of drawing the recorded values from one particular distribution of the GEV family can be represented as an n-dimensional surface, where n is the number of parameters of the distribution. Maximizing this likelihood function or its logarithm (the log-likelihood) yields an estimate of the parameters.

Parameter estimation The log-likelihood function, ℓ, derives from the distribution function. In our analyses, the numerical optimization for obtaining the minimum of $-\ell$ is done with Nelder-Mead at initial values estimated by L-moments with a fixed $\xi = 0.1$ (see section 4.2.2 or *Hosking* (1990) for details).

Definition 2. Log-likelihood function of the GEV, at $z_1, \ldots, z_m$,

$\ell(\mu, \sigma, \xi | z_1, \ldots, z_m) =$

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

Notation. To distinguish between theoretical and estimated parameters, one uses the notation of a small hat above the parameter name, e.g., ξ denotes the theoretical shape parameter, while $\hat{\xi}$ denotes the estimated shape parameter.

Let us estimate the parameters using ML estimation for the example distributions in example 4.1. For each of the samples in figure 15, ML estimation was used to estimate the distribution of the population as if the sample were a set of maxima of blocks of observations. Figure 16 shows the histograms of our samples with the true distribution (solid lines) as in figure 14, as well as the estimated distribution (dashed lines) of the population. The histogram can be understood as the empirical "PDF", i.e., without the assumption of a model. The dashed lines show the distribution estimated by maximum-likelihood from the samples. The vertical lines represent the median of the theoretical (solid line) and of the estimated (dashed line) distribution. Keep in mind that this is a sandbox example in which we know the distribution of the population the sample was drawn from (solid lines). In real life, of course, the population is unknown. We call the estimated distributions (dashed lines) fitted models, because they are adapted to the observed data.

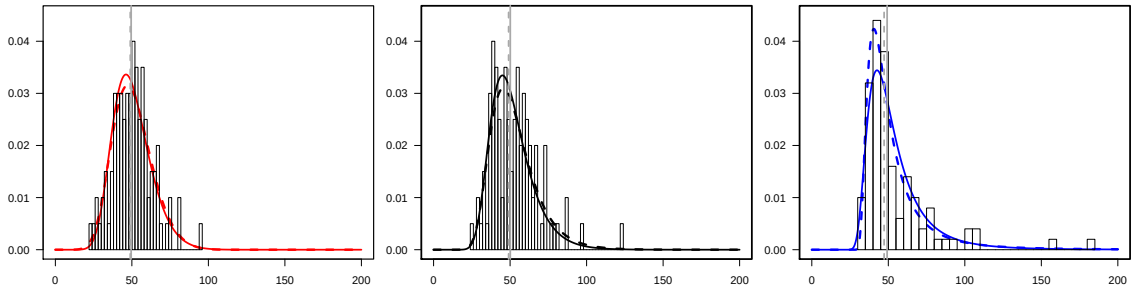


Figure 16: Histograms of the Weibull ($m = 100$, left, red), Gumbel ($m = 100$ center, black), and Fréchet sample data ($m = 100$, right, blue) seen in figure 15. Overlaid are the PDFs fitted by Maximum Likelihood to the sampled data (dashed lines), and the theoretical PDFs (solid lines, see figure 14) of the true distributions from which the data was sampled.

Both the PDFs and the medians of the distributions differ (figure 16), the more so for the heavy-tailed distribution (right panel). Thus, inferring the underlying distribution from a small sample is difficult, and the inferred information is inherently uncertain, because it depends on the particular sample we used. In the setting of maximum likelihood estimation, the uncertainty of the estimates can be reduced by increasing the number of independent maxima.

For the interpretation of uncertainties it is very important to keep in mind that it is an uncertainty of the estimates - similar to the uncertainty in the assessment whether a coin is fair based on a sample of coin tosses - and not an uncertainty about the "prediction" of an event - as the uncertainty whether the outcome of tossing a coin is heads or tails.

The example (figure 16) illustrates the inherent uncertainty of inference with few observed extremes.

Here, we have estimated the distribution from $m = 100$ years; by construction, we can be certain that all assumptions - data quality, representativeness (but for sample size), Z_i independent and identically distributed (iid), model choice - are fulfilled, and only the estimation uncertainty still remains (and stochastic uncertainty, which does not play a role in this section). Of course, this is not the case in real life.

Even though all our assumptions are correct, the fitted model deviates from the true distribution. This remaining uncertainty is the estimation uncertainty.

Uncertainty estimation ML estimation provides a theoretically founded approximation for the estimation of uncertainties. Standard errors of maximum likelihood estimates can be calculated easily. Nevertheless, using the profile likelihood function is recommended, as it allows for asymmetric confidence intervals (CIs).

Definition 3. Profile log-likelihood confidence intervals for $\hat{\xi}$, $\hat{\sigma}$ or $\hat{\mu}$.
For uncertainty estimates of $\hat{\xi}$, one fixes $\xi = \xi_0$ for a range of ξ_0, e.g. from -1.5 to 2, and for each ξ_0 maximizes the log-likelihood function with respect to μ and σ, given the observations. Thus one obtains a concave function of ξ with a maximum at $\hat{\xi}$. Approximate confidence intervals can be derived based on the fact that two times the difference in the log-likelihood function at ξ_0 and $\hat{\xi}$ follows a χ^2_1 distribution (see *Coles*, 2001, section 2.6.6 p.35 for details). The same can be done for $\hat{\mu}$ and $\hat{\sigma}$.

In figure 17 we see the profile log-likelihood functions (black solid lines) for all parameters. The vertical solid lines represent the true parameters and the three grey dashed vertical lines indicate the ML estimate, with the 95% confidence intervals ($1.96 \times \text{se}_{mle}$) on either side, where se_{mle} denotes the standard error estimated by Maximum Likelihood. The profile log-likelihood CIs are represented by the blue dashed vertical lines. They mark the intersection of the profile log-likelihood function with the critical value of the χ^2_1 distribution (lower blue horizontal line).

As expected, the true values for bounded (Weibull) or light-tailed (Gumbel) type samples are well within the CIs (regardless of the approximation method) and the two methods for approximating the CIs hardly make a difference. For the sample of Fréchet type, on the other hand, even though the CIs are larger than in the other two cases, the true value is just barely within the CI based on the standard errors. Thus, taking confidence intervals into consideration in practical applications is essential if 'capturing' the true value is of any importance, and extreme value analyses should mention these inherent uncertainties explicitly.

The width of the CI depends on sample size. The example presented here had a sample size of 100. In real life situations, however, we often have only 40 years of observations at our disposal. For comparison, we have plotted in figure 18 the profile log-likelihood functions for sample size 40 drawn from the same population distribution as in figure 13. Evidently, the confidence intervals for each parameter are considerably larger than with 100 years of data. Thus, shorter data records yield larger CIs.

In addition, both the inferred distribution and the approximated confidence intervals depend on the sample at hand. Figure 19 illustrates the fact that subsamples of equal size (m = 40) result in different

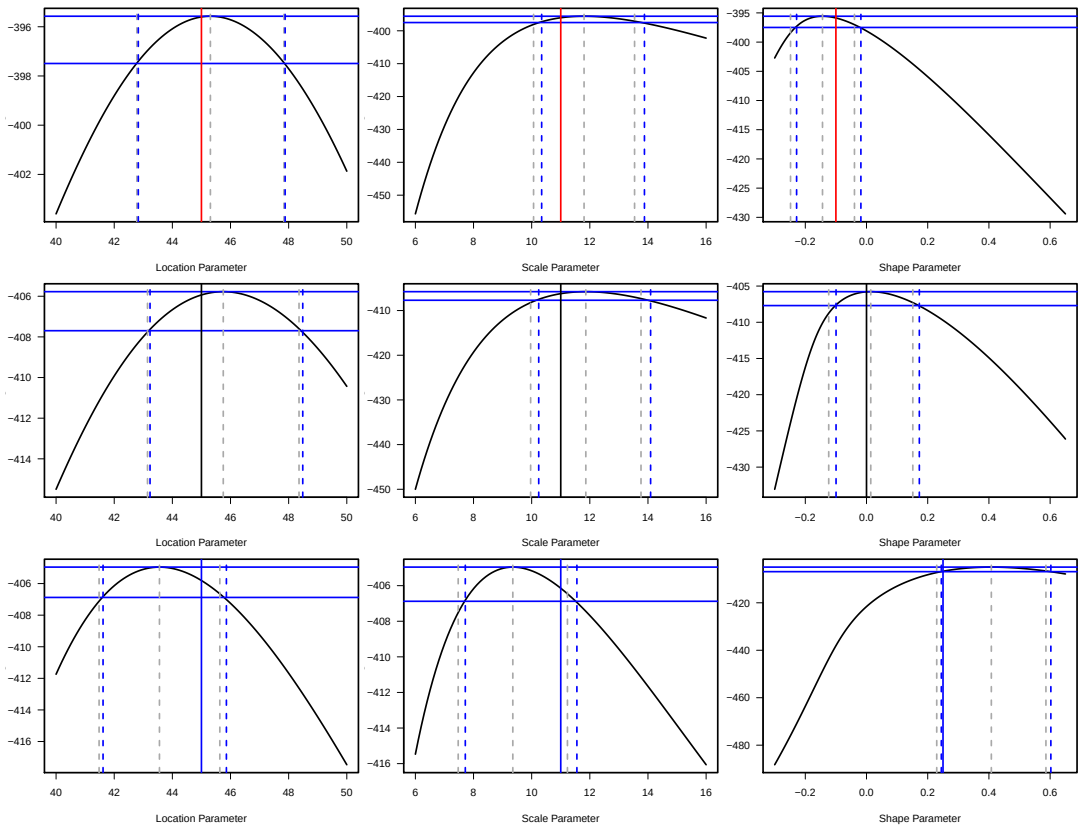


Figure 17: Profile log-likelihood functions (solid black curves) for each parameter - location (left), scale (center), shape (right)- estimated from the Weibull (top, $m = 100$), from the Gumbel (center, $m = 100$) or from the Fréchet (bottom, $m = 100$) sample data. The blue horizontal solid lines mark the maximum value of the profile log-likelihood function (corresponding to the ML estimate) and the critical value (95%) of the χ^2_1 distribution. The profile likelihood 95% confidence intervals are indicated by the 2 blue dashed vertical lines, the maximum likelihood estimate and 95% confidence intervals are indicated by the 3 grey dashed vertical lines. Finally, the true value is indicated by the solid vertical lines (Weibull in red, Gumbel in black, and Fréchet in blue).

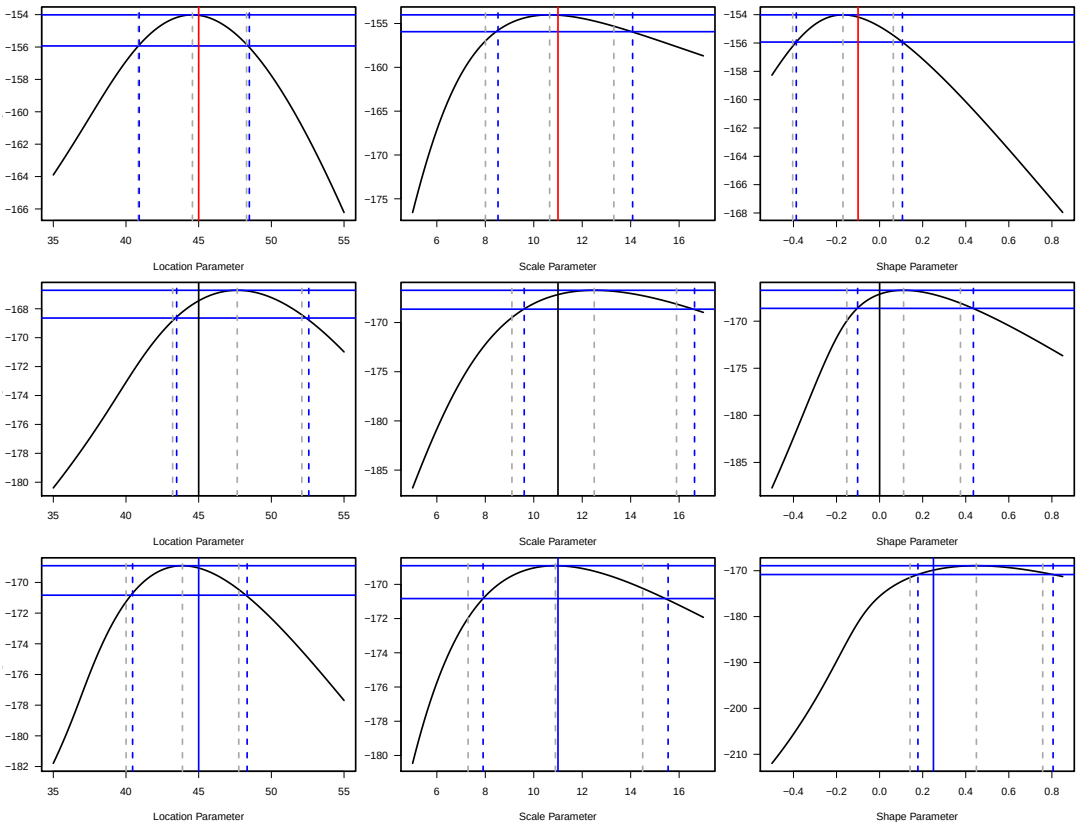


Figure 18: Profile log-likelihood functions (solid black curves) for each parameter - location (left), scale (center), shape (right) - estimated from the Weibull sample data ($m = 40$, top), from the Gumbel sample data ($m = 40$ center), and from the Fréchet sample data ($m = 40$, bottom). The profile likelihood 95% confidence intervals are indicated by the blue dashed vertical lines, the maximum likelihood best estimate and 95% confidence intervals are indicated by the grey dashed vertical lines and the true value is indicated by the solid vertical lines (Weibull in red, Gumbel in black, and Fréchet in blue).

inferred distributions, although they were drawn from the same population. The difference is particularly striking when drawing from the Fréchet family.

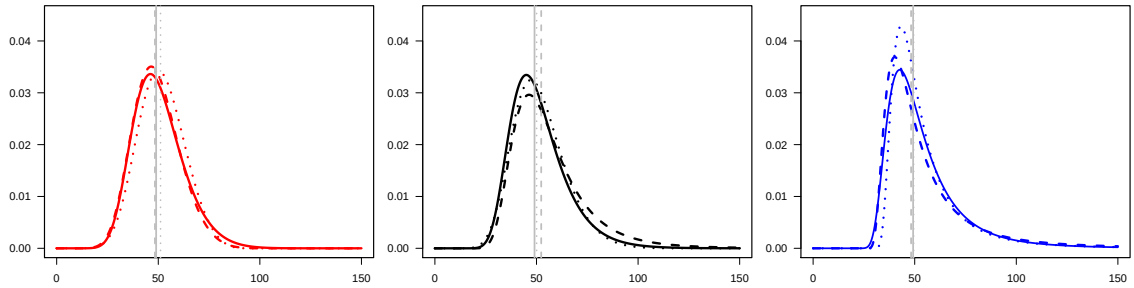


Figure 19: Six fitted PDFs (dashed and dotted lines) estimated from two subsets of Weibull sample data ($m = 40$, left, red), two subsets of the Gumbel sample data ($m = 40$, center, black), and two subsets of the Fréchet sample data ($m = 40$, right, blue). The theoretical PDFs of figure 14 are represented by solid lines and are considered as "truth".

Conclusions regarding the temporal evolution of the behavior of extreme events therefore cannot be based on fitting distributions within different time periods (for instance 1901 - 1950, 1951 - 2000) and comparing the estimates without taking into account the uncertainties. For such a task, rather sophisticated tests designed specifically for extremes are necessary.

4.2.2 L-moments estimation

The L-moments (*Hosking*, 1990) belong to the class of moment estimators, which are quite popular among hydrologists for the estimation of the GEV parameters. One minor drawback of this method is the lack of theoretically derived uncertainty estimates, but this can be circumvented by parametric resampling (see below).

Parameter estimation

Definition 4. L-moment estimation of the GEV parameters (*Hosking*, 1990)

$\hat{\xi} \approx -7.8590\left(\frac{2}{(3+t_3)} - \frac{\ln 2}{\ln 3}\right) - 2.9554\left(\frac{2}{(3+t_3)} - \frac{\ln 2}{\ln 3}\right)^2$ with $t_3 = l_3/l_1$, [a]

$\hat{\sigma} = l_2 \frac{-\hat{\xi}}{(1-2^{\hat{\xi}})\Gamma(1-\hat{\xi})}$ with $l_2 = 1/2 \frac{1}{\binom{n}{2}} \sum\sum_{i>j}(x_{(i)} - x_{(j)})$, and

$\hat{\mu} = l_1 - \frac{\hat{\sigma}}{\hat{\xi}}\left(\Gamma(1-\hat{\xi}) - 1\right)$ with $l_1 = \frac{1}{n}\sum_i x_i$.

[a] $l_3 = 1/3 \frac{1}{\binom{n}{3}} \sum\sum\sum_{i>j>k}(x_{(i)} - 2x_{(j)} + x_{(k)})$

Here, in figure 20, we see the L-moments estimates in addition to the maximum likelihood estimates. The difference between the estimates is very small.

Uncertainty estimation For confidence interval estimations, we choose a **parametric resampling** approach with a sample size of 3000.

The idea is to generate surrogate data from the estimated distribution, representing possible, additional samples. Samples of the same size as the original sample of observations are drawn from the estimated distribution, which is entirely characterized by the estimated parameters, hence the name

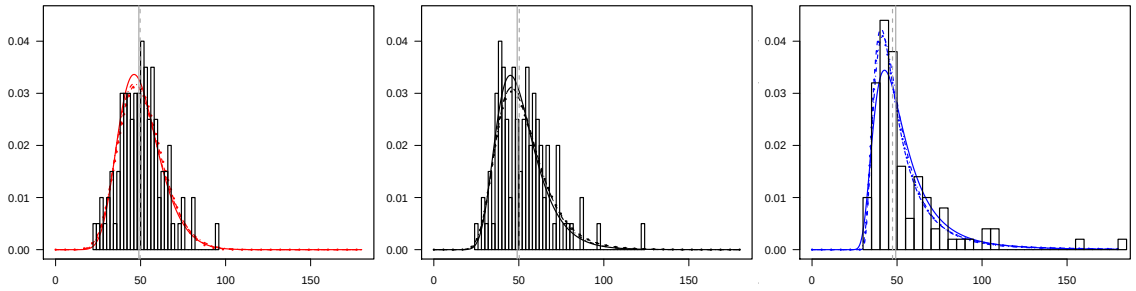


Figure 20: Same as fig. 16, with, in addition, the L-moments estimated PDFs (dotted lines).

"parametric resampling". (Some bootstrapping methods resample the observed data, rather than sampling randomly from the estimated GEV).

The new, surrogate, samples then serve as starting points for new estimates of the underlying distribution. From the resulting, here 3000, simulated distribution estimates, we construct 2-sided 95% confidence intervals, by taking the 2.5% and the 97.5% quantiles of the estimated quantity.

4.2.3 Bayesian Inference

In Bayesian inference, rather than estimating the parameter θ of the distribution that we believe generates the observations, we estimate a *distribution* for the parameter θ *itself*, conditional on the observed values z. The probability density function of θ given the observations z is called the ***posterior*** distribution, because it incorporates information available only ***after*** the values z have been observed. It is the ingredient we need to calculate any statistics that describe the behavior of θ (means, moments, quantiles, etc.), as well as any probabilistic quantities necessary for decision making.

We can obtain the posterior distribution easily by using the rules of conditional probability. If, for any θ, we know the following probability distributions, we can simply multiply them:

- the probability that our observations were produced by this particular θ,
- the probability that this particular θ should occur, i.e. that it was this particular set of parameters that governed the processes generating the observations.

Strictly speaking, this product should be normalized by the probability density of the observed quantity. As we shall see below, however, the product, as it is, suffices in practice to draw a large sample of θs distributed approximately as the posterior distribution.

The first probability distribution is easy to formulate: it is the likelihood function mentioned in section 4.2.1. The second is a distribution that reflects previous experience regarding the range of parameter values, i.e. knowledge available before having inspected the data. Accordingly, it is named ***prior*** distribution, as it represents information gathered ***before*** observing. The estimate of the posterior distribution of θ given z will then derive information both from the observations and from the prior distribution. The prior distribution expresses the general experience of the analyst, and is, to some extent, subjective. In many cases, the analyst's knowledge is vague, and the prior distribution is purposely shaped so as to express this vagueness. In such a case, all the information comes from the observa-

tions.

Notation. In this section, we need to introduce a few more notations:

- $\theta = (\mu, \sigma, \xi)$: a parameter triplet of the GEV.
- z: an observation. For instance $\{z_1, \ldots, z_m\}$, the m yearly maxima of daily precipitation.
- $p(\theta)$: probability density function of θ regardless of any observations, or before any information on the quantity Z was available.
- $p(z)$: probability density of z, regardless of θ.
- The vertical line in $p(a \mid b)$ means "conditional" on. Read: the probability density of a, given that b has occurred.
- $p(z \mid \theta)$: probability density of z, given that θ is the parameter triplet of the GEV. If z is fixed and θ varies, this is the likelihood function (and is no longer a probability density).
- $p(\theta \mid z)$: probability density of θ, given that z has been observed.
- θ-space: 3-dimensional space of possible values of the GEV parameters.

Details. *Bayes' rule*
Let θ denote the parameter of the distribution family assumed to have led to the observations z. Let $p(\theta, z)$ denote the joint distribution of θ and z.

Bayes' rule states that

$$p(\theta \mid z)p(z) = p(\theta, z) = p(z \mid \theta)p(\theta)$$

$$\Rightarrow p(\theta \mid z) = \frac{p(z \mid \theta)p(\theta)}{p(z)}$$

Note that $p(z)$ does not depend on θ, and is therefore a constant. In practice, as we shall see below, $p(\theta \mid z)$ will be sampled using the product $p(z \mid \theta)p(\theta)$, so that it will not be necessary to calculate $p(z)$ explicitly.

Bayesian inference proceeds in three steps. First, a model must be selected to represent the process producing the observed data. Second, a (joint) prior distribution must be formulated for the parameter(s) of the model distribution. Third, the (joint) posterior distribution of the parameter(s) is calculated or sampled.

Step 1: select a model We assume that the yearly (or seasonal) maxima follow a GEV distribution (definition 1 in section 4.1). Thus, our task is to estimate the joint distribution of the three GEV parameters μ, σ, and ξ, given the observed yearly (or seasonal) maxima.

Step 2: select and "customize" the prior distribution Since the GEV has three parameters, the prior distribution must be a joint probability distribution for the triplet $\theta = (\mu, \sigma, \xi)$. For simplicity, we assume that, in the prior distribution, μ, σ, and ξ are independent of each other. This allows us to define a prior distribution for each parameter separately, and multiply them to obtain the prior distribution of θ.

For describing the prior probability distribution of the shape parameter, we choose the beta distribution, a family of continuous probability distributions defined on the interval $[0, 1]$, parameterized by two shape parameters α and β. For our purposes, we choose the beta distribution $B(\alpha = 2, \beta = 2)$ and define it on the interval $[-0.5, 0.5]$. It is represented by the blue line in Fig. 21. It is symmetric, centered on zero, and the probability that ξ is outside of $[-0.5, 0.5]$ is 0. In other words, we assume the distributions to be mostly close to a Gumbel distribution, allowing for deviations up (down) to 0.5 (-0.5). The symmetry of the prior distribution expresses our belief that the Weibull and the Fréchet form are equally likely (see fig. 13 for examples). The decreasing form towards the limits 0.5 and -0.5 translates our past experience with extreme value analyses of precipitation that very small and large values of the shape parameter are less compatible with observed data, and therefore less likely. As it turned out, this constraint was instrumental in reducing unrealistic estimates of the shape parameter, as were often encountered in the frequentist framework.

For the location and scale parameters, we assume a non-informative Gaussian, i.e. a normal distribution with a standard deviation so large that the distribution is practically flat. Here, we take the standard deviation of the Maximum Likelihood (see section 4.2.1) estimate $\hat{\mu}$ inflated by a factor 1000. This choice expresses our ignorance about the location and scale parameter. In other words, although technically we define a prior distribution, it carries essentially no information. The estimates will then be completely determined by the observations. If we had selected an uninformative prior distribution for all three parameters, the triplet that maximizes the posterior density would be the same as that obtained by Maximum Likelihood estimation.

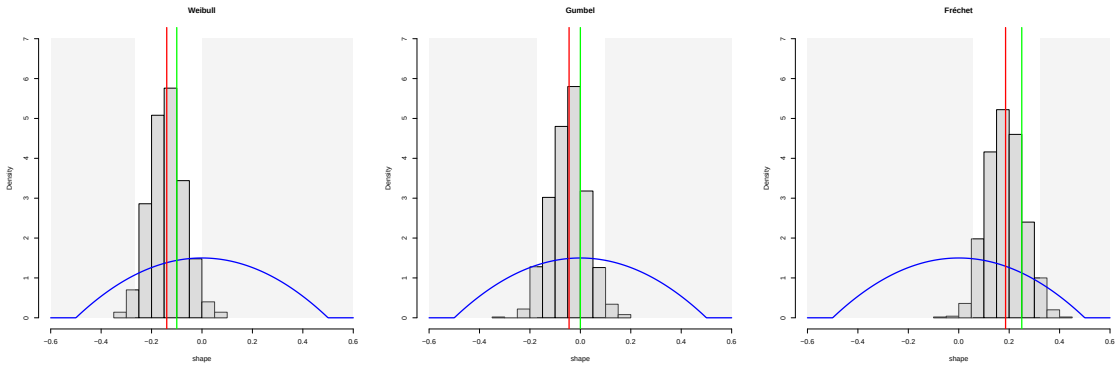


Figure 21: Histogram of the estimated shape parameters. Blue line: prior distribution for the shape parameter. Red line: median of the estimated shape parameters. Green line: stipulated true shape parameter. Pseudo-data samples of sample size 100 are drawn from the three GEV distributions of example 1 in section 4.1 ($\mu = 45$, $\sigma = 11$, and $\xi = -0.1, 0, 0.25$, respectively for the Weibull, Gumbel, and Fréchet distributions). Left: Weibull; center: Gumbel; right: Fréchet. The shaded areas are outside of the 95% confidence intervals. The pseudo-data samples do not appear on the graph. Note that they differ from the samples in fig. 13.

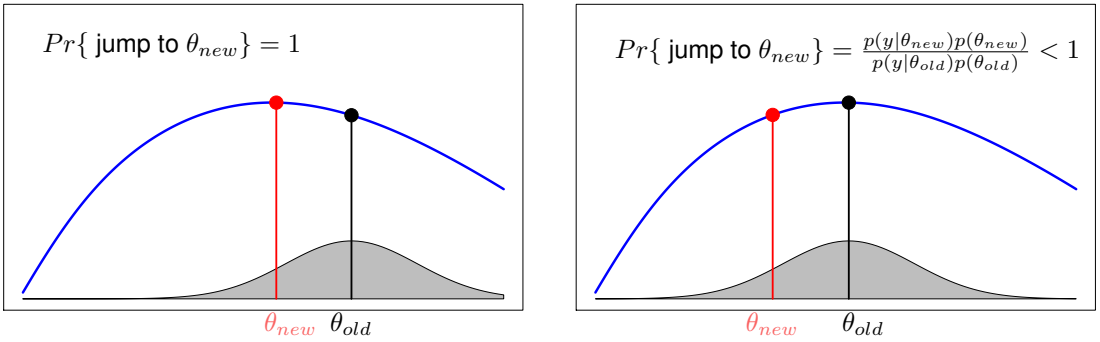


Figure 22: Illustration of the Metropolis algorithm. The blue line represents the function $f(\theta) = p(z \mid \theta)p(\theta)$. It is essentially the posterior distribution $p(\theta|z)$ we are looking for, except that it is not strictly speaking a probability distribution because it has not been normalized by $p(z)$. The value θ_{new} is randomly drawn from a normal distribution centered on θ_{old} (grey shading; corresponds to a random Gaussian walk). The probability of jumping from θ_{old} to θ_{new} (and adding it to the sample) depends on whether $f(\theta_{new}) \geqslant f(\theta_{old})$. On the left, θ_{new} is accepted, as would any value for which $f(\theta_{new}) \geqslant f(\theta_{old})$. On the right, since $f(\theta_{new}) < f(\theta_{old})$, we accept θ_{new} with a probability $p = f(\theta_{new})/f(\theta_{old}) < 1$: we "roll the dice" by drawing from a uniform distribution on $[0, 1]$. Suppose we have $p = 0.6$; if we draw a number between 0 and 0.6, we jump to θ_{new}; otherwise, we stay at θ_{old} and draw a new proposal θ from the normal distribution.

Step 3: determine the posterior distribution We are looking for $p(\theta \mid z)$, the probability density function of θ, given the occurrence of observations z. In other words, we are looking for the joint distribution of the GEV parameter triplet θ, now that we know the values z have been observed.

According to Bayes' rule, the posterior probability density is proportional to the prior probability density $p(\theta)$, constrained by the likelihood $p(z \mid \theta)$ that the observed values z could have been produced by θ. In our case, the likelihood is that of the GEV (see definition 4.2.1 for the expression of the GEV log-likelihood).

In real life problems, $p(\theta \mid z)$ can rarely be determined in analytical form. An alternative is to obtain an empirical distribution based on a large sample of θs drawn directly from $p(z \mid \theta)p(\theta)$. Unfortunately, sampling directly from $p(z \mid \theta)p(\theta)$ is not possible if the model has several parameters, as is the case for the GEV. In such a case, one resorts to a procedure called Markov Chain Monte Carlo (MCMC) (see for instance *Gelman et al.*, 2013, ch.11). The purpose is to sample lots of θs in areas of θ-space where $p(z \mid \theta)p(\theta)$ is large (high probability density), and few of them where $p(z \mid \theta)p(\theta)$ is small (low probability density). The term Markov Chain refers to the fact that each sampled θ will depend, for technical reasons, on the previously sampled one.

Starting at a given value of θ, say θ_{old} (for instance the Maximum Likelihood estimate $\hat{\theta}$), we randomly propose a new value of θ, say θ_{new} in the vicinity. This can be done for example by sampling from a normal distribution centered on θ_{old} (Gaussian random walk). If the product $p(z \mid \theta)p(\theta)$ is larger for θ_{new} than for θ_{old}, we jump to θ_{new} and add it to our sample. If it is smaller for θ_{new} than for θ_{old}, we only sometimes jump to θ_{new} and add it to our sample: we "roll the dice" and only jump at a rate that reflects how much smaller the probability density for θ_{new} is than for θ_{old} (i.e. we jump often if it is not so much smaller, and rarely if it is much smaller). If we happen not to jump, we stay where we are without re-adding θ_{old} to our sample. The procedure is illustrated in fig. 22.

Details. *The result of the bayesian analysis is a large set of MCMC-sampled triplets* $\{\theta_1, ..., \theta_S\} = \{(\mu_1, \sigma_1, \xi_1), ..., (\mu_S, \sigma_S, \xi_S)\}$ *that approximate the joint probability distribution of* μ*,* σ*, and* ξ *given our observations* z*. In our case,* $S = 1000$*. Each of these location, scale, and shape triplets* $\theta_s, s = 1, ..., S$ *represents a different GEV distribution* $G_s = G(z; \theta_s = (\mu_s, \sigma_s, \xi_s))$*.*

As we shall see below, this MCMC-sampled set of triplets and their corresponding GEV distributions allow us to derive confidence intervals for the GEV parameters, return levels and return periods, and even the goodness of fit results.

Figure 21 shows the histogram of ξ values thus "MCMC-sampled" for artificial data generated with the Weibull, Gumbel, and Fréchet distributions of example 1. The stipulated true values of ξ for each of these population distributions are represented by the vertical green lines. The histogram approximates the posterior distribution of the shape parameter. Its median (vertical red line) can be used as the estimate for ξ. We can see that, after having included the information provided by the observed data, the posterior distribution is considerably narrower than our prior distribution. Depending on the set of artificial data, the Weibull and Fréchet form are no longer equally likely, except approximately so in the case of Gumbel artificial data. In fact, for Weibull (Fréchet) artificial data, the probability that the distribution is of form Fréchet (Weibull) is 2.7% (0.3%). (The possibility to calculate such probabilities so easily is a useful consequence of Bayesian inference).

In the analyses on the web platform (`http://climate-extremes.ch`), the MCMC simulation chain consists of a Metropolis algorithm using a Gaussian random walk for proposing new parameter values. The standard deviation of the random walk is taken to be the asymptotic standard error of the Maximum Likelihood estimate.

Uncertainty estimation In the Bayesian framework, the uncertainties are explicitly quantified in the posterior distribution of θ given our observed yearly (or seasonal) maxima z. Traditionally, they are called *credible intervals*, no doubt to distinguish their interpretation from that of confidence intervals in the frequentist framework. For convenience, however, we will keep the albeit incorrect appellation *confidence intervals*.

Details. *By definition, equal-tailed* 95% *confidence intervals contain* 95% *of the posterior probability distribution, and put an equal amount of probability in each tail:* 0.025*.*

The bounds of the 95% *confidence intervals of* μ*,* σ*, and* ξ *are therefore the* 2.5% *and* 97.5% *quantiles of the MCMC-sampled values* $\{\mu_1, ..., \mu_S\}$*,* $\{\sigma_1, ..., \sigma_S\}$*, and* $\{\xi_1, ..., \xi_S\}$ *respectively.*

In fig. 21, the area outside of the 95% confidence intervals is depicted on a grey background. The confidence intervals are not symmetric with respect to the median of ξ (red vertical line), as the posterior distributions of ξ are somewhat skewed to the right.

4.3 Assessing the fit of the model

The validity of the model assumptions usually cannot be checked in a systematic manner. However, it is possible to compare the empirical distribution $\tilde{G}$, that makes no assumptions regarding an underlying model, with the estimated distribution $\hat{G}$.

Notation. $\tilde{G}$: empirical distribution, i.e., for each value of the sample, the proportion of the total number of values in the sample that is below it.

$\hat{G}$: estimated distribution.

This can be done with diagnostic plots or with goodness-of-fit tests. Diagnostic plots are visual guides to the validity of model assumptions, and will be introduced in section 4.3.1 below. Goodness-of-fit tests quantify the difference between empirical and estimated distribution and are presented in section 4.3.2.

4.3.1 Visual guides

Probability-probability (empirical vs. estimated probability) and quantile-quantile (empirical vs. estimated quantile) plots are instruments for visual inspection of fit reliability. The empirical and estimated probabilities (quantiles) are plotted against each other. Should the estimated distribution describe the observed data well (up to sampling uncertainty), the points should line up along the line of slope 1 and intercept 0.

Although they offer essentially the same information, probability-probability (PP) plots and quantile-quantile (QQ) plots highlight different areas of the distribution.

PP-plots The probability-probability plot plots the estimated probability against the empirical probability.

Definition 5 (probability-probability-plot.)**.** A probability-probability plot consists of the points

$$\left\{\left(\hat{G}(z_{(i)}), \tilde{G}(z_{(i)})\right) : i = 1, \ldots, m\right\},$$

where $z_{(i)}$, $i = 1, \ldots, m$, are the ordered observed values, $z_{(1)} \leq \ldots \leq z_{(i)} \leq \ldots \leq z_{(m)}$, $\tilde{G}$ is the empirical distribution function and $\hat{G}$ the estimated distribution function.

In the context of extreme values, the PP plot has a weakness: By definition, both the estimated and the empirical probabilities are bound to approach 1 as the quantile increases. Therefore, the PP plot offers little (or no) information in the region of most interest.

The QQ plot circumvents this weakness by changing the scale.

QQ-plots The quantile-quantile plot associates to each value of the ordered sample (e.g., a maximum over a block) the corresponding quantile of the estimated distribution.

By definition, the cumulative distribution function at a given value z_p gives the probability p that any of the values up to (and including) z_p may be realized. The value z_p is then called the p-quantile: $\Pr\{Y \leq z_p\} = p$. The median of a distribution, for instance, is its $\frac{1}{2}$-quantile or 50%-quantile; if $z_{0.5}$ is the median, there is a 50% chance for a realization of the random variable to be below $z_{0.5}$. If $z_{0.75}$ is the 75%-quantile, there is a probability of 75% that any value smaller or equal $z_{0.75}$ will be be realized. Thus, if a sample value is such that 75% of sample values are below it, it will be associated on the QQ-plot with the 75%-quantile of the estimated distribution.

If the model is appropriate for the data at hand, the empirical quantiles of the data should approximately agree with the estimated quantiles, the more so for a large sample.

For the upper quantiles, empirical quantiles that are systematically smaller than the fitted ones suggest that the tail of the true distribution may decay more slowly than that of the fitted distribution.

Definition 6 (quantile-quantile-plot.)**.** A quantile-quantile plot consists of the points

$$\left\{\left(\hat{G}^{-1}(\frac{i}{m+1}), z_{(i)}\right) : i = 1, \ldots, m\right\},$$

where $z_{(i)}$, $i = 1, \ldots, m$, are the ordered observed values, $z_{(1)} \leq \ldots \leq z_{(i)} \leq \ldots \leq z_{(m)}$, $\hat{G}$ is the estimated distribution function, and $\frac{i}{m+1}$ is an empirical estimate of the return period of observation of order (i).
(This particular empirical estimator of the return period is called Weibull estimator and is one of many suggested in the literature).

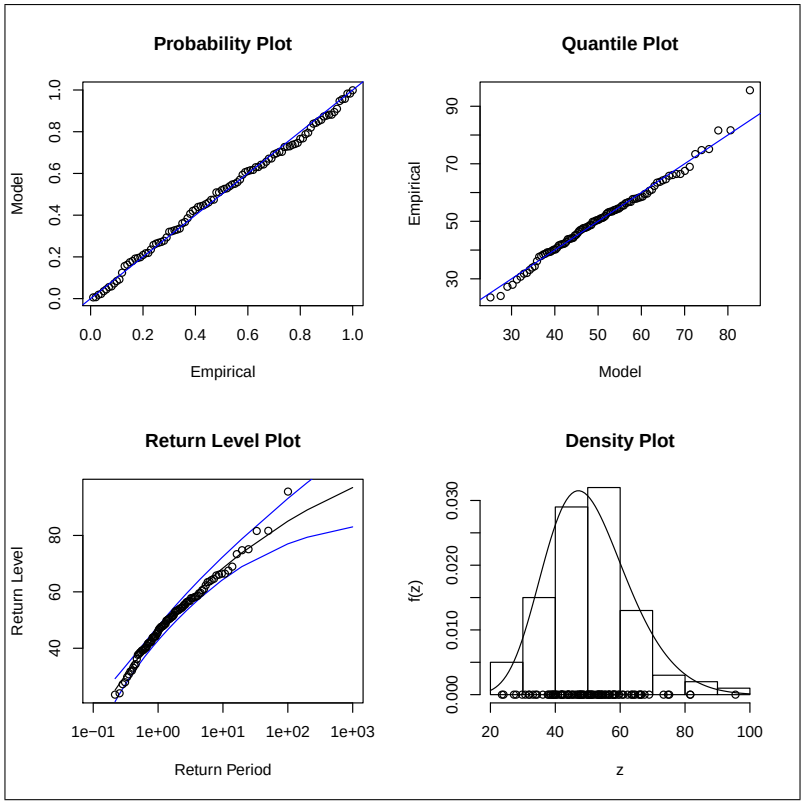


Figure 23: Suite of diagnostic plots for the Weibull sample ($m = 100$, left panel on fig. 15) as provided by the R-package ismev. Note that the CIs are based on the ML estimation standard errors and therefore the upper limit of the CI is underestimated.

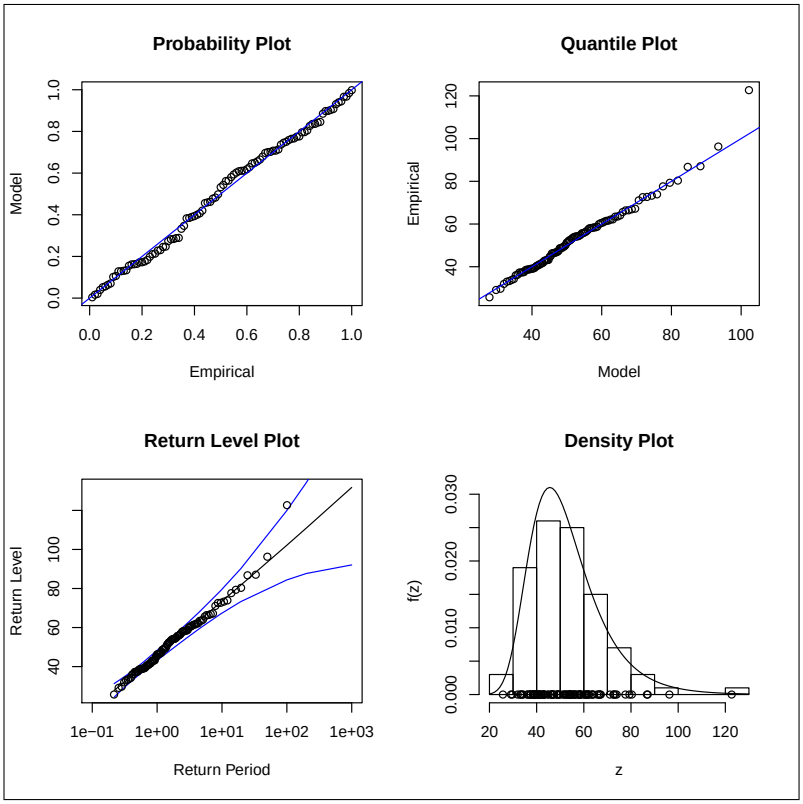


Figure 24: Suite of diagnostic plots for the Gumbel sample ($m = 100$, center panel on fig. 15) as provided by the R-package ismev. Note that the CIs are based on the ML estimation standard errors and therefore the upper limit of the CI is underestimated.

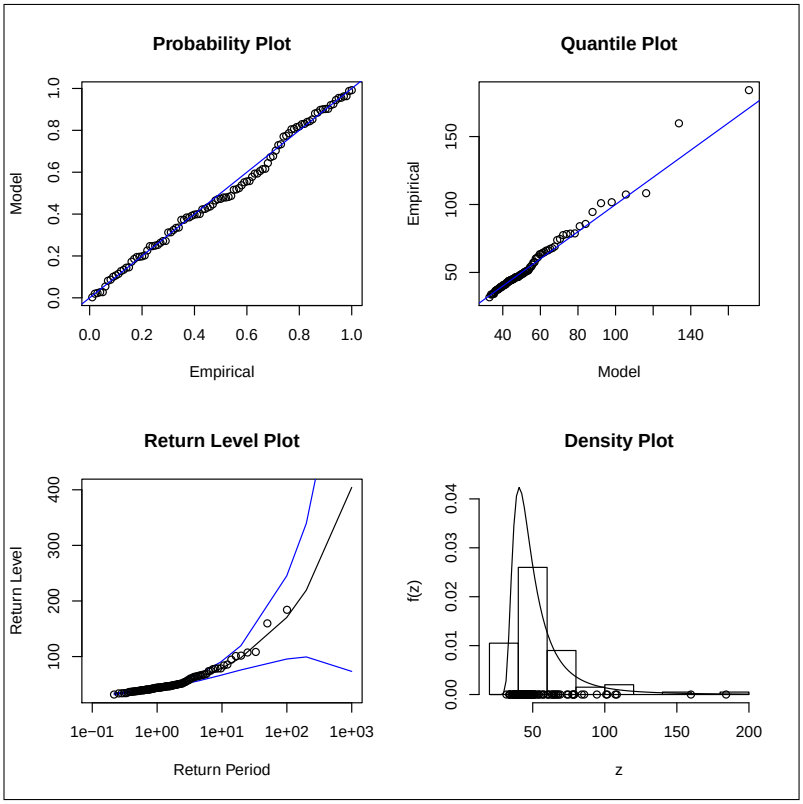


Figure 25: Suite of diagnostic plots for the Fréchet sample ($m = 100$, right panel on fig. 15) as provided by the R-package ismev. Note that the CIs are based on the ML estimation standard errors and therefore the upper limit of the CI is underestimated.

Bayesian framework In the Bayesian framework, rather than one GEV distribution indexed by a specific parameter triplet θ, we estimate a (posterior) distribution for θ. In order to obtain quantiles for the QQ plots, we make use of the large posterior sample of GEV triplets resulting from the MCMC procedure (see section 4.2.3), which approximates the posterior joint distribution of GEV parameters. For each of these $S = 1000$ MCMC-sampled GEVs, we calculate the m quantiles corresponding to the m ordered observed values $z_{(1)} \leq \ldots \leq z_{(i)} \leq \ldots \leq z_{(m)}$, as in definition 6. Each of the m quantiles appearing on the QQ plot is the median of the $S = 1000$ quantiles thus calculated.

Seasonal analysis If the analysis was applied to seasonal —rather than yearly —maxima, the quantiles must be derived numerically from the seasonal quantiles. Details can be found in section 4.5, where the procedure is described for return levels.

4.3.2 Goodness-of-fit criteria

We employ goodness-of-fit methods based on empirical distribution statistics, such as the well-known Kolmogorov-Smirnoff (KS) and Cramer-von Mises (CM) statistics but also statistics that put more weight on the tail, i.e., on the rarer events of the distribution.

In a first step, we consider the root-mean-square-error (RMSE) between empirical and estimated distribution (see table 5). In addition, we assess the appropriateness of the fit of the model via goodness-of-fit tests. Goodness-of-fit tests test if the overall distance between the empirical distribution and the estimated distribution is significantly different from zero.

As with the visual guides, one can perform the tests on different scales, and / or put weight on areas of the distribution one is interested in. As we would like to draw conclusions regarding the behavior of extremes, we need the right tail of the distribution to be reliable.

To this end, we apply the suite of tests listed in table 5 (see *Luceno* (2006) for further details). The null hypothesis (H_0) is that the empirical and estimated distribution are the same. Rejecting the null hypothesis indicates that the estimated distribution deviates from the empirical distribution more than to be expected by chance.

Frequentist framework For each test, the critical values are obtained by parameteric resampling: a set of 3000 samples of the same size as the observed sample is drawn from the estimated distribution; a GEV is fitted to each of these simulated samples, and the value of the test statistic is estimated; the 95% quantile of these test statistic values is the critical value of the test.

Bayesian framework To set up the goodness of fit criteria in the Bayesian framework, we followed chapter 6, “Model Checking”, of *Gelman et al.* (2013). The question we want to answer is: “Could our data have been produced by the range of parameters of the posterior distribution we have estimated?”.

In order to test that, we make use of the large posterior sample of GEV triplets obtained by MCMC that approximate the joint posterior distribution of GEV parameters: for each of these GEVs, we can calculate one of the goodness-of-fit tests described above.

Suppose now, that for each of these MCMC-sampled GEVs we drew a sample of artificial observations (of the same size as the original sample), and calculated the same test, but with the artificial data. The value of these tests will reflect the typical test results if the data are actually generated by these GEVs.

Will we be able to distinguish the actual from the artificial test results in any way? If we can't, we can assume that our model represents the observed data to an acceptable degree.

Details. *Let $z = \{z_1, ..., z_m\}$ be the observed sample. We have $S = 1000$ MCMC-sampled GEV distributions that approximate the posterior distribution. Each of these S distributions $G_s = G(z; \theta_s = (\mu_s, \sigma_s, \xi_s))$, $s = 1, ..., S$ is indexed by a location, scale, and shape triplet that we call θ_s. Let T represent any of the tests listed in table 5 (KS, CR, AD, ...); it depends on the data and the GEV distribution the data is meant to originate from.*

- *Step 1: For each s, we calculate the value of the test $T(z; \theta_s)$.*
- *Step 2: We draw a "replicate" sample $z_s^* = (z_{s\,1}^*, .., z_{s\,m}^*)$ of length m from each MCMC-sampled GEV G_s.*
- *Step 3: For each of these replicate samples, we calculate the value of $T(z_s^*; \theta_s)$.*
- *Step 4: We compare the distributions of $T(z)$ and $T(z^*)$. Since T is essentially a distance, the model should be rejected if $T(z) > T(z^*)$ systematically, that is if the distance from the estimated distributions to the empirical distribution of observed values is systematically greater than the distance between the estimated distributions and the empirical distribution of an artificial sample drawn from these distributions. The fraction of times $T(z; \theta_s) < T(z_s^*; \theta_s)$ can be used as a p-value: a very small fraction indicates that it is unlikely to be by chance that the observed sample is poorly represented by the model.*

The p-value can be determined for each test, for each season separately, as well as for the yearly analysis derived from the seasonal ones.

Final verdict For the overall decision whether the fitted distribution (frequentist approach) or the median of the posterior distribution (Bayesian approach) is suited for extrapolation outside the observed range, we use the set of rules given in table 4. These affect essentially the cases of questionable reliability: they are to be understood as a "red flag" pointing to some irregularity in the fit that must be looked into carefully before proceeding any further. In particular, they are meant to attract our attention to much weaker anomalies - that would normally not indicate a rejection of the null hypothesis - and to unusual values of the shape parameter.

Verdict	Rule
poor	any H_0 rejected at $\alpha = 5\%$
questionable	at least one H_0 rejected at $\alpha = 20\%$ or at least 3 H_0 rejected at $\alpha = 30\%$ or $\xi \notin [-0.15, 0.19]$ (supra-daily) $\xi \notin [-0.1, 0.3]$ (sub-daily)
good	otherwise

Table 4: Set of rules defining the verdict, where H_0 denotes the null hypothesis of any of the tests given in table 5 (i.e., H_0: empirical and estimated distribution are the same).

RMSE	$RMSE = \sqrt{1/n \sum \lvert F(x) - S_n(x)\rvert^2}$	$RMSE = \sqrt{1/n \sum_{i=1}^{n} \lvert z_i - S_n(x_{(i)})\rvert^2}$
KS	$D_n = \sup_x \lvert F(x) - S_n(x)\rvert$	$D_n = \frac{1}{2n} + \max_{1 \leq x \leq n} \lvert z_i - \frac{i-1/2}{n}\rvert$
CM	$W_n^2 = n \int_{-\infty}^{\infty} \lvert F(x) - S_n(x)\rvert^2 dF(x)$	$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(z_i - \frac{i-1/2}{n}\right)^2$
AD	$A_n^2 = n \int_{-\infty}^{\infty} \frac{\lvert F(x)-S_n(x)\rvert^2}{F(x)\{1-F(x)\}} dF(x)$	$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1)\{\ln z_i + \ln(1 - z_{n+1-i})\}$
ADR	$R_n^2 = n \int_{-\infty}^{\infty} \frac{\lvert F(x)-S_n(x)\rvert^2}{\{1-F(x)\}} dF(x)$	$R_n^2 = \frac{n}{2} - 2\sum_{i=1}^{n} z_i - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \ln(1 - z_{n+1-i})$
AD2R	$r_n^2 = n \int_{-\infty}^{\infty} \frac{\lvert F(x)-S_n(x)\rvert^2}{\{1-F(x)\}^2} dF(x)$	$r_n^2 = 2\sum_{i=1}^{n} \ln(1 - z_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{2i-1}{1-z_{n+1-i}}$

Table 5: Empirical distribution statistics (*Luceno*, 2006) used for goodness-of-fit assessment (left) and their computational form (right). $F(x)$ is the theoretical Cumulative Distribution Function (CDF), and $S_n(x)$ is the empirical CDF, which is a step function with jumps at the order statistics. We use the notation $z_i = F(x_{(i)})$. **RMSE**: Root mean square error; **KS**: Kolmogorov-Smirnoff; **CM**: Cramer-von Mises; **AD**: Anderson-Darling; **ADR**: Right-tail Anderson-Darling; **AD2R**: Second order right-tail Anderson-Darling.

4.4 Return levels

Return levels are the name given to extreme quantiles. A return level with return period T corresponds to a quantile with probability $1 - 1/T$. The equation defining return levels expressed in terms of the GEV parameters is given in definition 7.

Definition 7 (return level)**.** The return level z_p with return period $T = 1/p$ is given by

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\ln(1-p)\}^{-\xi}], & \text{for } \xi \neq 0 \\ \mu - \sigma \ln\{-\ln(1-p)\}, & \text{for } \xi = 0. \end{cases}$$

where $\Pr\{Z \leqslant z_p\} \approx G(z_p) = 1 - p$, with G a member of the GEV family.

As seen in figures 2 and 13, displaying the probabilities on the original scale of the data does not give sufficient information on the tails of the distribution. Figure 26 illustrates why the return level plot (see definition 8) is an intuitive way to present extreme value statistics. It focuses on the tail behavior of the

distribution, by log-log transforming the x-axis. Without any additional tools, one can see if the distribution is bounded or heavy tailed: if the function has a positive curvature (blue line), the distribution is heavy tailed; if the function has a negative curvature (red line) the function is bounded. In case the function is linear (black line), the distribution belongs to the Gumbel family.

Definition 8 (return level plot)**.**

$$\{(ln(y_p), z_p) : 0 < p < 1\}$$

with $y_p = -ln(1-p)$.

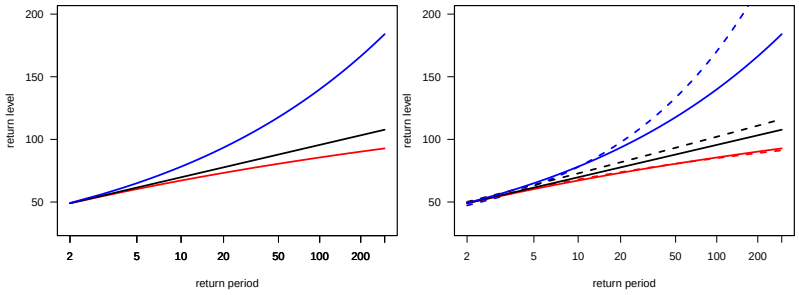


Figure 26: Theoretical return levels (solid lines) of Weibull (red), Gumbel (black), and Fréchet (blue) distributions (figure 14). Right panel: same as left with the estimated return levels (dashed lines - right panel) of the respective fitted distributions (figure 16).

4.4.1 Return level estimation

Frequentist framework The estimation of the return levels is straightforward if the parameters of the distribution are already estimated: one simply replaces the parameters of definition 7 with their respective estimates and computes the return level of interest.

In general, $G(z_p) = 1 - p$ has to be solved for z_p. In some cases, for instance when non-stationarities are modeled explicitly, this must be done numerically as no analytic expression exists (see section 4.5 below).

Bayesian framework In the Bayesian framework, the return level for a given return period is not just one value, but a probability distribution —the posterior distribution —of return levels. Once again, we make use of the large posterior sample of GEV triplets obtained via MCMC, which approximates the posterior joint distribution of GEV parameters. For each of the $S = 1000$ MCMC-sampled GEVs, we calculate a return level (as in definition 7) for the return period of interest $1/p$ (for instance the 50-year return level), thus obtaining S return levels $\{z_{1_p}, ..., z_{S_p}\}$. The return level z_p is then taken to be the median of the MCMC-sample of return levels $\{z_{1_p}, ..., z_{S_p}\}$. Keep in mind that the resulting curve appearing on the return level plot is *no longer* a GEV.

4.4.2 Uncertainty estimation

Frequentist framework As for the parameter uncertainties, the uncertainties of the return level estimates can usually be estimated using the profile log-likelihood function.

Definition 9. Profile log-likelihood confidence intervals for z_p.
For uncertainty estimates of $\hat{z_p}$, one needs to reparametrize the likelihood function. For instance one can replace μ with $z_p + \frac{\sigma}{\xi}\left[1 - \{-\ln(1-p)\}^{-\xi}\right]$ in the likelihood function. Then one maximises the likelihood function with respect to σ and ξ given the observations. Approximate confidence intervals can be derived via the χ^2 distribution.

We see in figure 27 that the CIs are not symmetric around the return level estimates: the smaller the probability, i.e., the longer the return period, the less symmetric the CIs. It is important to keep in mind that the uncertainties can only be read in one direction: the confidence interval describes the uncertainty of the return level, and not of the return period.

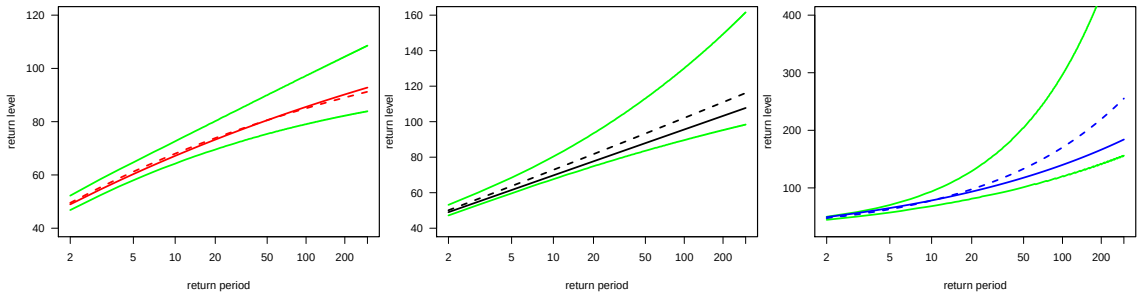


Figure 27: Return level plots of theoretical distribution (solid line, figure 14), and fitted distribution (figure 16) from samples (figure 15) of size m = 100 (dashed lines) with 95% CI via profile likelihood function (green lines).

The uncertainties of the return levels can always be estimated by parametric resampling (see section 4.2.2 for a description of the method). In particular, this is a good option when no analytic expression of the return level exists (as is the case when non-stationarities are modeled explicitly).

Bayesian framework In the Bayesian framework, the confidence intervals are straightforward to calculate. To obtain the return level, we calculated For each MCMC-sampled GEV $G_s, s = 1, ..., S$ the return level corresponding to the return period of interest (for instance the 50-year return level), and took the median of the S return levels. The 95% confidence bounds are then the 2.5% quantile and the 97.5% quantile of the same S return levels $\{z_{1_p}, ..., z_{S_p}\}$.

Figure 28 shows the return level plots for artificial data sampled from the three distributions in example 4.1. As expected from the shape parameters estimated in fig. 21, the estimated return levels are lower than the true return levels in all three cases. The green lines are GEV distributions (the theoretical GEVs from which we drew the artificial samples), while the red lines representing the estimated return levels, are *not*.

4.5 Seasonal non-stationarities

An often tacit assumption for estimating the GEV is that the series of values within a block (of which we will take the maximum) are independent, or to the very least, stationary with very weak long-range dependence at extreme levels (see *Coles*, 2001, ch.5).

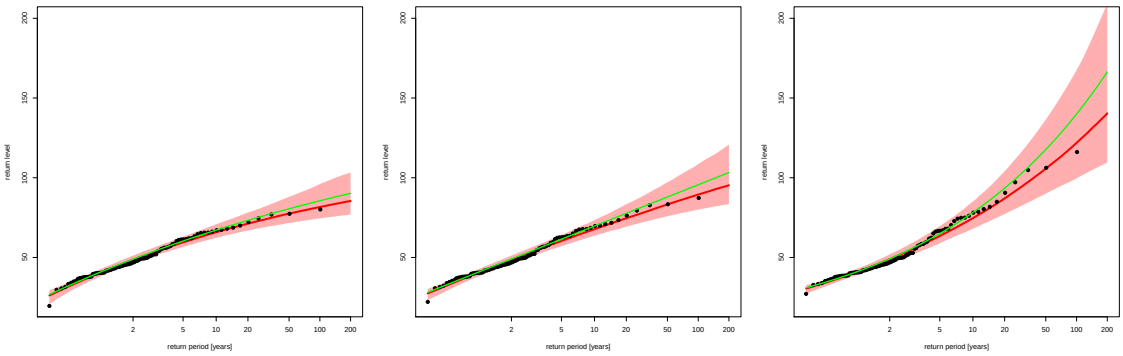


Figure 28: Return level plots for the estimates based on artificial data samples drawn from the example Weibull ($\mu = 45$, $\sigma = 11$, and $\xi = -0.1$; left), Gumbel ($\mu = 45$, $\sigma = 11$, and $\xi = 0$; center), and Fréchet ($\mu = 45$, $\sigma = 11$, and $\xi = 0.25$; right). Red line: median of the estimated return levels. Pink shading: 95% confidence intervals. Green line: true return level.

Definition 10 (Stationarity)**.** A series $X_1, X_2, ..., X_n$ (for instance daily precipitation) is said to be stationary if its stochastic properties are homogeneous through time.

If the $X_1, X_2, ..., X_n$ series is stationary, then X_1 must have the same distribution as X_{51}, or X_{101}, and the joint distribution of X_1, X_2 must be the same as that of X_{51}, X_{52} and X_{101}, X_{102}. But X_1, X_2, and X_{102} can be dependent.

Non-stationarity, however, is very common in physical processes: we need only think of the daily or seasonal cycles of sunshine duration, or of long-term trends in temperature. We have reasons to believe that precipitation processes in Switzerland cannot always be represented by a stationary process, as processes that generate rain depend on the season. In winter, precipitation is usually brought about by large-scale synoptic cyclones sweeping over the continent. In summer, it is often thunderstorms, which are strongly influenced by local conditions, that cause heavy precipitation. Depending on the geographic location, not all seasons are equally likely to host a yearly maximum, and the magnitudes of monthly maxima of daily precipitation reveal a distinct seasonal cycle (*Umbricht et al.* (2013)).

From a statistical point of view, non-stationarities must be dealt with case by case, because the form of the dependence on time is specific to the physical processes under consideration: it requires expert knowledge and must be inferred beforehand, by other means. A common approach, once such a form has been hypothesized, is to make the parameters of the GEV depend on time accordingly (see *Coles*, 2001, ch.6).

Here, we address only seasonal non-stationarities in precipitation. We adopt a pragmatic approach: we divide the year in 4 rigid seasons (December-February, March-May, June-August, and September-November), and perform the extreme value analysis on each season separately. We then deduce the yearly return period from the seasonal return periods, and derive numerically the yearly return level from the seasonal return levels.

Details. *Let Z^{year} denote the random variable associated with yearly maxima, and Z^{DJF}, Z^{MAM}, Z^{JJA}, Z^{SON} those associated with seasonal maxima in winter, spring, summer, and autumn respectively. Let $Pr\{Z^{block} \leq z\}$ denote the probability that the largest value in the block is no larger than z, where "block" means either the entire year, or one of the four seasons.*

If the largest value in the year does not exceed an amount z, then neither does any of the four seasonal maxima. We can therefore write

$$Pr\{Z^{year} \leq z\} = Pr\{Z^{DJF} \leq z\} \times Pr\{Z^{MAM} \leq z\} \times Pr\{Z^{JJA} \leq z\} \times Pr\{Z^{SON} \leq z\},$$

assuming that the seasonal maxima are independent of each other. In other words,

$$Pr\{Z^{year} \leq z\} = G_{DJF}(z) \times G_{MAM}(z) \times G_{JJA}(z) \times G_{SON}(z)$$

If p denotes the probability of exceedance of a value z_p (for instance a specific amount of daily precipitation), the yearly return period $T = 1/p$ is very easy to calculate:

$$p = Pr\{Z^{year} > z_p\} = 1 - Pr\{Z^{year} \leq z_p\}$$

$$\Rightarrow p = 1 - G_{DJF}(z_p) \times G_{MAM}(z_p) \times G_{JJA}(z_p) \times G_{SON}(z_p)$$
$$\Rightarrow T = \frac{1}{1 - G_{DJF}(z_p) \times G_{MAM}(z_p) \times G_{JJA}(z_p) \times G_{SON}(z_p)}$$

*Deriving the return levels, on the other hand, requires more work. Since the product of GEV distributions with different parameters is **not** itself a GEV, this equation cannot be solved analytically for z_p. To obtain the yearly return level, we must resort to a numerical method, the Newton-Raphson algorithm. It is also not possible to provide a location, scale, or shape parameter to describe extreme behavior.*

Frequentist framework For each season separately, the yearly return periods and return values are obtained as described in the technical remark above (4.5). Confidence intervals for the return levels are obtained by parametric resampling of each season separately, deriving the artificial yearly return values, then taking the 2.5% and the 97.5% quantiles thereof as confidence bounds.

Bayesian framework For each season separately, the posterior distribution of the GEV parameters must be estimated following the Bayesian inference procedure described in section 4.2.3. Recall that the posterior distribution of the GEV parameters is approximated by a large set of MCMC-sampled parameter triplets. The GEVs from different seasons can be put together to form an artificial year, and the yearly return period or return level derived as in the technical remark above (4.5). Since we obtain a distribution of yearly return levels, the 95% confidence intervals are delimited by the 2.5% and the 97.5% quantiles.

Details. *Bayesian inference applied separately to each season results in $S = 1000$ GEV distributions in each season, for instance $\{G_1^{JJA}, ..., G_S^{JJA}\}$ for summer.*

For convenience, we assume that $\{G_s^{DJF}, G_s^{MAM}, G_s^{JJA}, G_s^{SON}\}$ belong to the same artificial year; as the seasonal maxima are assumed to be independent from each other, it doesn't matter which members of the series we glue together to form a year.

The probability of exceedance of z in a given year p^{year} can be obtained as follows:

For each $S = 1, ..., 1000$, we calculate the yearly probability of exceedance ${p_s}^{year} = Pr\{Z_s^{year} > z\}$ using $\{G_s^{DJF}, G_s^{MAM}, G_s^{JJA}, G_s^{SON}\}$ as described in the technical remark above (4.5). Then p^{year} is the median of the probabilities of exceedance $\{{p_1}^{year}, ..., {p_S}^{year}\}$, and the yearly return period is $T^{year} = 1/p^{year}$.

To obtain the yearly return level z corresponding to a yearly return period $T^{year} = 1/p^{year}$, we derive $\{z_1, ..., z_S\}$ using the Newton-Raphson algorithm, then take the median thereof. The 95% confidence bounds are the 2.5% and the 97.5% quantiles of $\{z_1, ..., z_S\}$.

4.6 Extreme Value Analysis example

Step by step extreme value analysis of 1-day precipitation recorded in Zürich / Fluntern (station abbreviation: SMA) between 1961 and 2013.

Figure 29 shows the daily precipitation recorded during 43 years at Zürich / Fluntern.

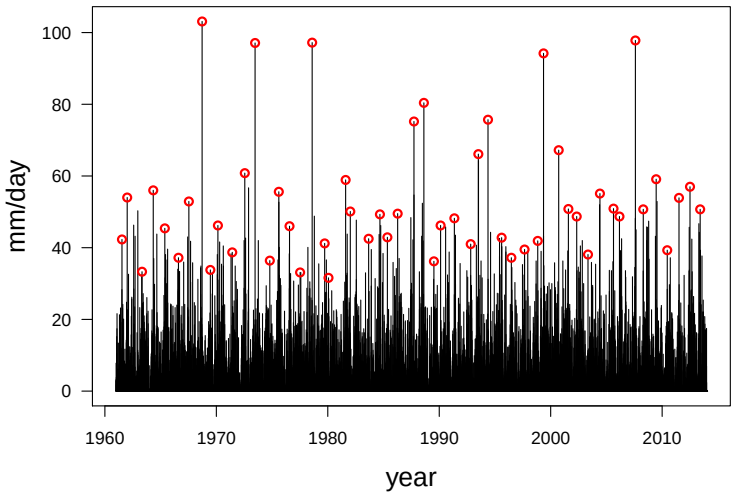


Figure 29: Daily precipitation time-series recorded at station Zürich / Fluntern (SMA) between 01.01.1961 and 31.12.2013. The largest values per year are circled in red.

For the block maxima approach we need to extract the annual maxima (figure 30).

As we can see, most annual maxima occur during the summer season (table 6).

To fit the GEV we plug all annual maxima into the negative log-likelihood function and use a numerical optimization routine (in most cases Nelder-Mead) to find the combination of location, scale, and shape parameters that maximize the likelihood surface. The negative log-likelihood maximum of **215.4663** is reached at location: **44.25**, scale: **10.39**, and shape: **0.26**. These are our maximum likelihood

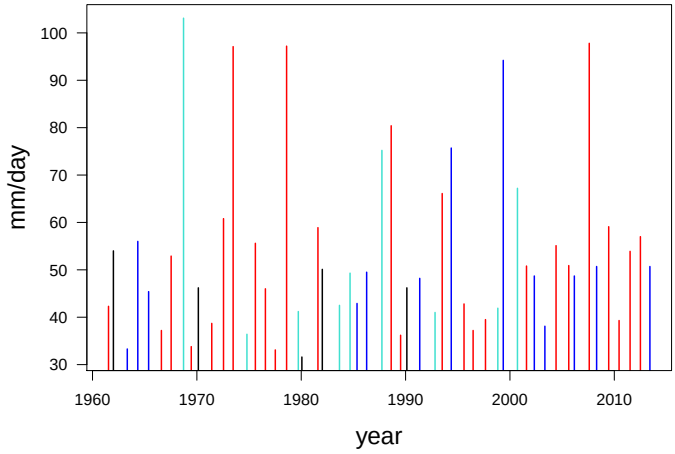


Figure 30: Time-series of annual maxima recorded at station Zürich / Fluntern (SMA) between 01.01.1961 and 31.12.2013. The climatological season in which the maximum occurs is indicated by color: December, January, February (DJF) - black, March, April, May (MAM) - blue, June, July, August (JJA) - red, and (September, October, November) SON - turquoise.

rank	amount	date	rank	amount	date
1	103.1 mm	1968-09-21	6	80.4 mm	1988-08-15
2	97.8 mm	2007-08-08	7	75.7 mm	1987-09-25
3	97.2 mm	1978-08-07	8	67.2 mm	2000-09-20
4	97.1 mm	1973-06-23	9	66.1 mm	1993-07-05
5	94.2 mm	1999-05-12	10	60.8 mm	1972-07-21

Table 6: Top 10 annual maxima with the occurrence date, ordered by decreasing amount. The climatological season in which the maximum occurs is indicated by color: DJF - black, MAM - blue, JJA - red, and SON - turquoise.

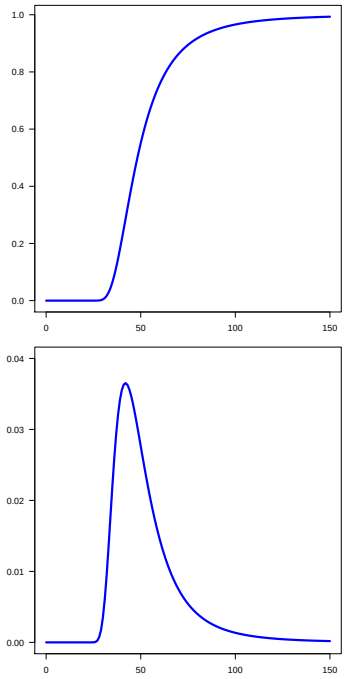
estimates for the GEV parameters. The corresponding standard errors are 1.65, 1.36, and 0.13, respectively.

- We have already discussed the goodness-of-fit of the statistical model in chapter 2, where we showed the QQ-plot (figure 9).
- The cumulative probability function and the corresponding probability density function are shown in the left panel of figure 31a.
- As we are interested in the extremes, we calculate return levels (figure 31b) with their confidence intervals.

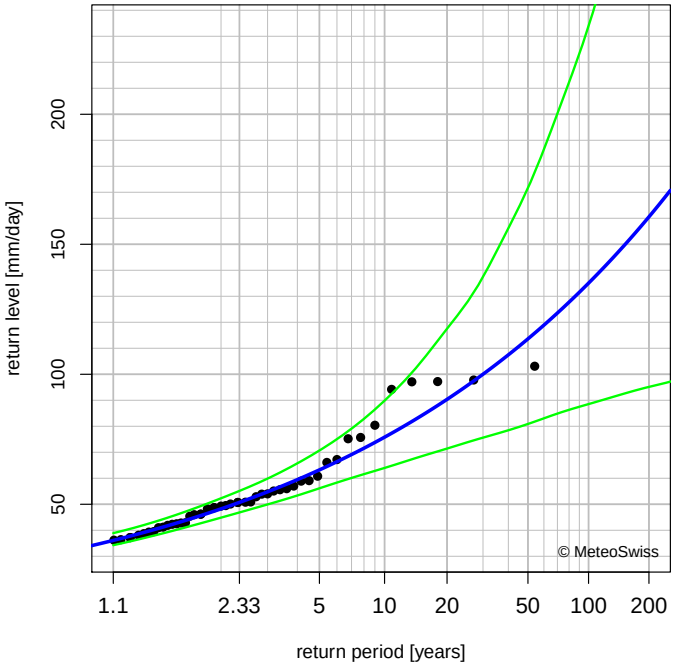
Coming back to the top 10 events, we can now calculate the probability of exceeding these 10 events. We again express the probability in terms of return periods.

date	amount	return period	date	amount	return period
1968-09-21	103.1mm	34 years	1988-08-15	80.4 mm	12 years
2007-08-08	97.8 mm	27 years	1987-09-25	75.7 mm	10 years
1978-08-07	97.2 mm	27 years	2000-09-20	67.2 mm	6 years
1973-06-23	97.1 mm	27 years	1993-07-05	66.1 mm	6 years
1999-05-12	94.2 mm	24 years	1972-07-21	60.8 mm	4 years

Table 7: Top 10 annual maxima at station Zürich / Fluntern (SMA) between 1961 and 2013 and their estimated return period.



(a) Cumulative distribution function (top) and probability density function (bottom).



(b) Return level plot: best estimate (blue) with 95% confidence intervals (green) based on the profile likelihood.

Figure 31: Extreme value analysis of 1-day precipitation at station Zürich / Fluntern (SMA) for the period 1961-2013.

5 Comparison with Zeller, Geiger, Röthlisberger Volumes

5.1 Introduction

Until recently, several sources of varying quality were available for looking up return levels of precipitation. One of these was the collection of analyses by *Zeller et al.* (1976-1992) in 9 volumes published by the Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL. This source will be referred to hereafter as ZGR volumes, or simply ZGR. It is no longer in print, and while most engineers are aware of the existence of these analyses, some do not know where they can be found. Since 2016, the MeteoSwiss web platform documented in the present report (`http://climate-extremes.ch`) has put state-of-the-art extreme value analyses of precipitation at the disposal of the public. It will subsequently be referred to as the CE web platform, or simply CE.

The question may arise whether infrastructure designed with the help of return values provided in the ZGR volumes is still safe. In the following, we conduct a systematic comparison of return levels in the ZGR volumes and on the CE web platform at a subset of stations. Our purpose is to increase awareness of differences between the ZGR volumes and the CE web platform, and encourage users to contact MeteoSwiss for further details.

Note that at the time of writing, the CE analyses were performed using Maximum Likelihood Estimation. In 2018, the estimation method was changed to the Bayesian approach (see section 4.2.3) applied to seasonal maxima (see section 4.5), thereby increasing the number of stations and precipitation durations for which the return level estimates were reliable. Nevertheless, the present comparison was not repeated with the new results, and is based on analyses that are no longer available on CE.

5.2 Background

5.2.1 The Zeller, Geiger, Röthlisberger volumes

Zeller et al. (1976-1992) is a collection of analyses based on precipitation data gathered partly between the mid 19th century and 1987, but for most stations between 1901 and 1970. The analyses were conducted in the late 1970s up to the late 1980s.

The method used for estimating return levels is described in detail in *Geiger et al.* (1991). It estimates Gumbel distributions for the largest value of the year by first assigning empirical return periods to the

observed return levels, then plotting these return levels against the appropriately transformed empirical return periods on a graph, and finally using linear regression to obtain return periods of unobserved return levels. Should the plotted points diverge systematically from a straight line, a statistical test depending on the differences in slope from point to point is applied to determine whether a Fréchet distribution might be more appropriate (for details regarding the Fréchet distribution, see section 4.1). In the following comparisons, the method used in ZGR will be referred to as LR_{ZGR} (linear regression).

Linear regression is very easy to use, and at the time could be performed without much computational power. Unfortunately, however, it assumes that the empirical return periods are true, and that the error for all data points is normally distributed. Thus, it disregards the large uncertainty associated with the estimation of the tail.

Although more recent analyses exist, those presented in ZGR are still widely used in hydrology, hazard prevention, and engineering. They do not make use of precipitation data in the last 30 years, however, and the statistical methods are outdated.

5.2.2 climate-extremes.ch

The CE web platform provides state-of-the-art statistical analyses of heavy and extreme precipitation for a large nation-wide network of measuring stations. The measurement period currently used for the analyses presented on the website is 1966 to 2015 for supra-daily precipitation, also referred to as the *standard period*. Sub-daily precipitation analyses are based on data gathered from 1981 to the most recent complete year. The detailed methodological background can be found in chapter 4.

Essentially, the largest values in a year are assumed to follow a Generalized Extreme Value distribution (GEV). This approach is generally referred to as Block Maxima, as each observation used for the estimation of the GEV is itself the maximum value of a quantity (say 1-day precipitation), over a block, i.e. a fixed number of observations (say all days in a year). The parameters of the GEV are estimated by maximum likelihood. Uncertainties are estimated by parametric resampling. In the following comparisons, the method used on the CE web platform will be referred to as ML (maximum likelihood). Although the estimation method is inadequate, ZGR can also be said to use the Block Maxima approach in the sense that it bases its estimates on the largest value of the year.

Theoretically, the GEV distribution is the limiting distribution of the maxima of an infinite number of observations. In addition, the observations within a block, for instance daily precipitation, are assumed to result from the same physical process, so that they can be regarded as realizations of independent and identically distributed random variables.

In practice, the maximum is taken over a season, or a year, which limits the number of observations within a block to 366 or perhaps 90. Under certain circumstances (nature of the observed quantity, type of distribution of the observed quantity,...), this may not be sufficient for the largest value to converge towards the distribution of the maxima. In addition, the probability of rain on a given day, for instance, is known to depend on the occurrence of precipitation on the previous day. Also, precipitation processes are subject to seasonal variations. Thus, it is by no means unlikely that the assumptions underlying the present estimates should be violated. In such cases, the observations might not be represented well by the estimated GEV distribution.

To address this issue, a goodness-of-fit (GoF) flag was introduced to attract the attention of users to the reliability of the estimation, or lack thereof. A detailed description of this flag can be found in section 4.3.2. Where the reliability is ”poor”, the results are withheld. Where it is ”questionable”, the user is encouraged to take a closer look at the analysis of the quality of the estimate that is available on the web page under ”further information”.

5.3 Comparison and Interpretation

Differences between CE and ZGR estimated return levels can stem either from the estimation method used, or from the time period covered by the data. For supra-daily precipitation, data is available over the same period as analyzed in ZGR, and thus it is possible to disentangle the causes underlying the differences. To this end, the method used on CE was applied to the time period covered in ZGR, and the results subsequently compared both to ZGR, and to CE. For sub-daily precipitation, on the other hand, this was not possible, as in ZGR the sub-daily return levels are extrapolated from the 1901-1970 supra-daily (1 to 7-day precipitation) return levels. The extrapolation method itself has its weaknesses, and is irrelevant here, since today the time series of sub-daily data are sufficiently long to estimate the return levels directly. Thus, no effort was made to reproduce an elaborate but questionable extrapolation, and the return levels of sub-daily precipitation estimated with CE and ZGR are simply compared directly.

Since the difference between return level estimates depends on the return period in question, it is convenient to have a quantitative ”summary difference” for all six return periods extracted from ZGR at once. We will use the root mean square difference $RMSD$, which is defined as follows. Let $z_t^{(est_a)}$ and $z_t^{(est_b)}$ denote two different estimates est_a and est_b of the return levels of the t^{th} return period $T \in \{2.33, 10, 20, 30, 50, 100\}$. The root mean square difference between $z_t^{(est_a)}$ and $z_t^{(est_b)}$ is defined as

$$RMSD(est_a, est_b) = \sqrt{\frac{1}{6}\sum_{t=1}^{6}(z_t^{(est_a)} - z_t^{(est_b)})^2}.$$

In summary, the following comparisons have been made:

- **Raw difference: Maximum Likelihood applied to the standard period (ML_{SP}) on CE vs. Linear Regression (LR_{ZGR}) in ZGR.**
- **Difference due to different methods: Maximum Likelihood applied to the period covered by ZGR (ML_{PerZGR}) vs. Linear Regression in ZGR (LR_{ZGR}).** This comparison was carried out on supra-daily data only.
- **Difference due to different periods: Maximum Likelihood applied to the standard period (ML_{SP}) on CE vs. Maximum Likelihood applied to the period covered by ZGR (ML_{PerZGR}).** This comparison was carried out on supra-daily data only.

We first present a detailed analysis at two stations, showing the different estimates on the same return level plot. Then, the results of the comparison are displayed in the form of maps of the 100-year return level. The comparison is carried out for 10 minute, 1 hour, 1 day, 2 day, and 5 day precipitation.

The stations chosen for the comparisons form only a small subset of the stations of the MeteoSwiss observational network. The selected subset is part of the Swiss National Basic Climatological Network

(NBCN), which represents the climatic regions of Switzerland. More information about the Swiss NBCN can be found in *Begert* (2008). For supra-daily (sub-daily) analyses, 25 (33) stations of the Swiss NBCN were used.

5.3.1 Supra-daily Analyses

In order to illustrate the differences in return level estimates on CE and ZGR and look into the role played by method or period, we consider 2 stations, selected for their contrasting results: Altdorf (ALT), and la Chaux-de-Fonds (CDF). Altdorf is situated in a Föhn valley of the northern Alps and is exposed to Föhn activity from fall to spring. La Chaux-de-Fonds is in a high valley of the Jura mountains, where the topographic configuration favors the formation of cold air lakes. Figure 32 shows the 1 day and Figure 33 the 1 and 5 day precipitation return levels read on the graphs of ZGR (LR_{ZGR}) at return periods of 2.33, 10, 20, 30, 50 and 100 years on the one hand, and the return levels estimated by Maximum Likelihood for the standard period (ML_{SP}) and the period covered by the ZGR data (ML_{PerZGR}) on the other hand. The 95% confidence intervals for the Maximum Likelihood method are also displayed, and allow us to gauge the significance of the differences. Recall that the confidence intervals do not give us boundaries between which there is a 95% chance of finding the true value. They only tell us how likely it is for a value to be the estimate, given the data we could observe (the likeliest one being the best estimate).

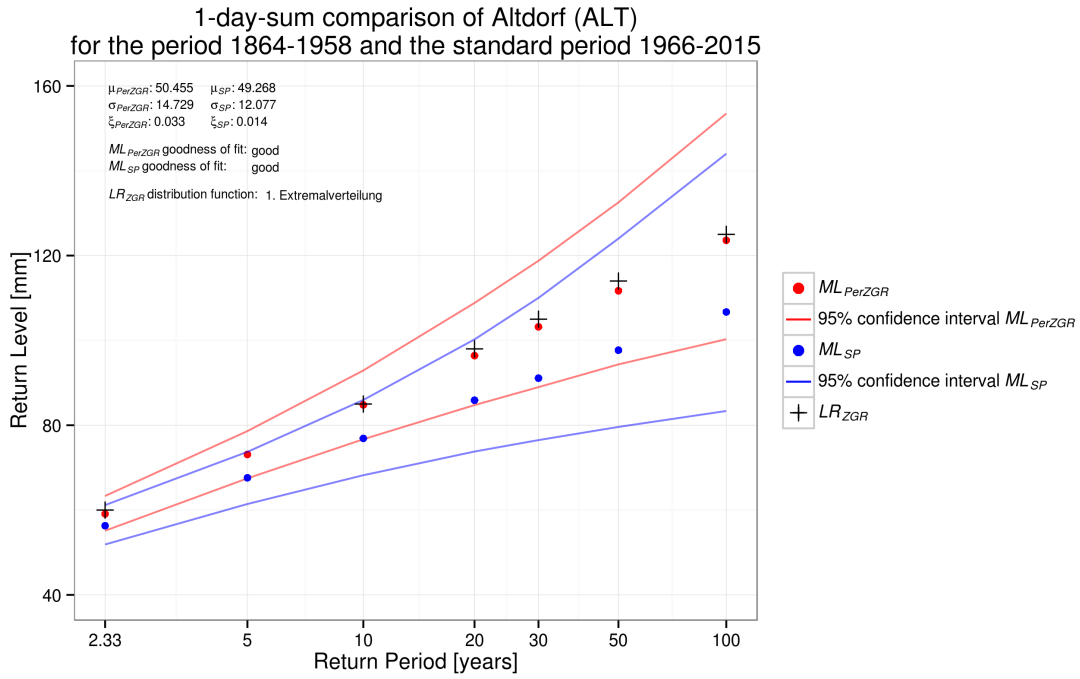


Figure 32: Station Altdorf (ALT), 1-day precipitation: return levels estimated by Maximum Likelihood applied to the standard period 1966-2015 on CE (blue dots, ML_{SP}) and to the period covered by ZGR (red dots, ML_{PerZGR}), and their respective 95% confidence intervals (blue and red lines). Corresponding ZGR return levels (LR_{ZGR}) estimated by linear regression of plotting points are indicated by little crosses.

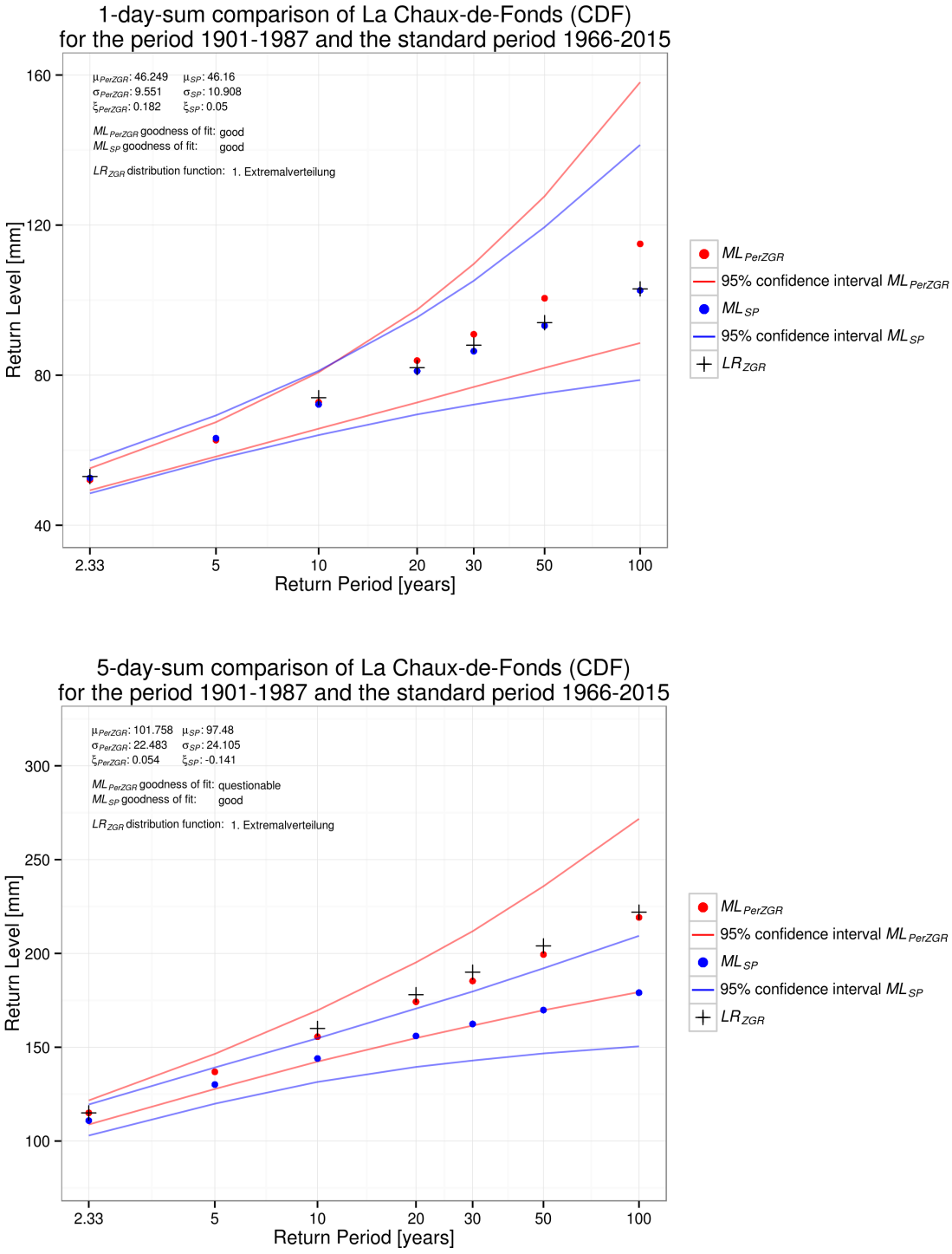


Figure 33: Station La Chaux-de-Fonds (CDF), 1-day (above) and 5-day (below) precipitation: return levels estimated by Maximum Likelihood applied to the standard period 1966-2015 on CE (blue dots, ML_{SP}) and to the period covered by ZGR (red dots, ML_{PerZGR}), and their respective 95% confidence intervals (blue and red lines). Corresponding ZGR return levels (LR_{ZGR}) estimated by linear regression of plotting points are indicated by little crosses. Note the different scale for the return levels.

For 1 day precipitation at Altdorf and la Chaux-de-Fonds, the raw differences in estimates (LR_{ZGR} vs.

ML_{SP}) can be seen as non-significant, as they are within the confidence intervals of the CE estimates. In practice, this means that if, for instance, the ZGR 100-year return value of about 125 mm/day at Altdorf or 100 mm/day at la Chaux-de-Fonds had been used for design, it would still be valid today, because it is not incompatible with the ML_{SP} estimates obtained by Maximum Likelihood from the data of the standard period. Although the ZGR 100-year return value is less likely than the ML_{SP} best estimate, it is by no means impossible, given the climate observed during the standard period. For 5-day precipitation at la Chaux-de-Fonds, on the other hand, ZGR estimates (LR_{ZGR}) differ significantly from CE estimates (ML_{SP}). This suggests that the ZGR estimates could be unnecessarily high.

For both 1 day precipitation at Altdorf, and 5 day precipitation at la Chaux-de-Fonds, the discrepancy between ZGR and CE is essentially due to the difference in observation period, rather than to the difference in method. This is the case for a large number of NBCN stations examined: the root mean square difference $RMSD(ML_{SP}, ML_{PerZGR})$ between Maximum Likelihood estimates over the different time periods is larger than the root mean square difference $RMSD(LR_{ZGR}, ML_{PerZGR})$ between return levels estimated over the same period with different methods at 88%, 75% and 75% of stations for 1, 2, and 5 day precipitation respectively. In the three examples presented here, the return levels appear to have decreased between the ZGR period and today. Given the small sample, however, this could very well be due to chance, rather than to a change in the processes leading to extreme events.

A striking feature of 1 day precipitation at la Chaux-de-Fonds is that ZGR return level estimates (LR_{ZGR}) are much lower at high return periods than the Maximum Likelihood estimates over the same period (ML_{PerZGR}). Here, the difference due to the method used is as large as the difference due to the period of observation, and in this case, the differences compensate each other. Over the ZGR period, the GEV distribution ML_{PerZGR} is evidently heavy-tailed, and the empirically estimated return periods used by ZGR prior to applying linear regression are overestimated. In this particular case, ZGR estimates turn out to be nearly the same as CE estimates over the standard period ML_{SP}.

The spatial distribution of the raw differences between ZGR and CE 100-year return levels can be viewed for 1-day precipitation on figure 34. The differences are not very large - at least 2/3 of the stations display a difference below 10%, regardless of the sign - and there is no clear spatial pattern. For two of the stations, Basel/Binningen and Col du Grand St-Bernard, the raw difference is significant. At both of these stations, the reliability of the CE estimate is questionable, although this may be irrelevant, because it concerns only the part of the distribution corresponding to relatively low return periods (not shown). As it turns out, for all return periods $T \in \{2.33, 10, 20, 30, 50, 100\}$ put together, ZGR return levels differ significantly from CE return levels only for 11%, 20% and 21% of the observations, for 1 day, 2 day, and 5 day precipitation respectively. Only at Col du Grand St-Bernard is the raw difference significant for all three supra-daily durations.

Return levels for 100-year return periods are higher on CE than in ZGR at more than half of the stations, and the same holds true for 2 and 5 day precipitation (not shown). For station Säntis (the dark orange circle in the north-east with a 100-year return level 20-30% lower than on CE), a remark in ZGR states that the data is unreliable and not representative. Generally, stations at high altitude such as Säntis and Col du Grand St-Bernard suffer from measurement difficulties related to wind and snow. The stations Altdorf for 1-day precipitation seen in fig. 32 and La Chaux-de-Fonds for 5-day precipitation in fig. 33 display relatively large differences: the 100-year return levels are 10-20% lower on CE than in ZGR.

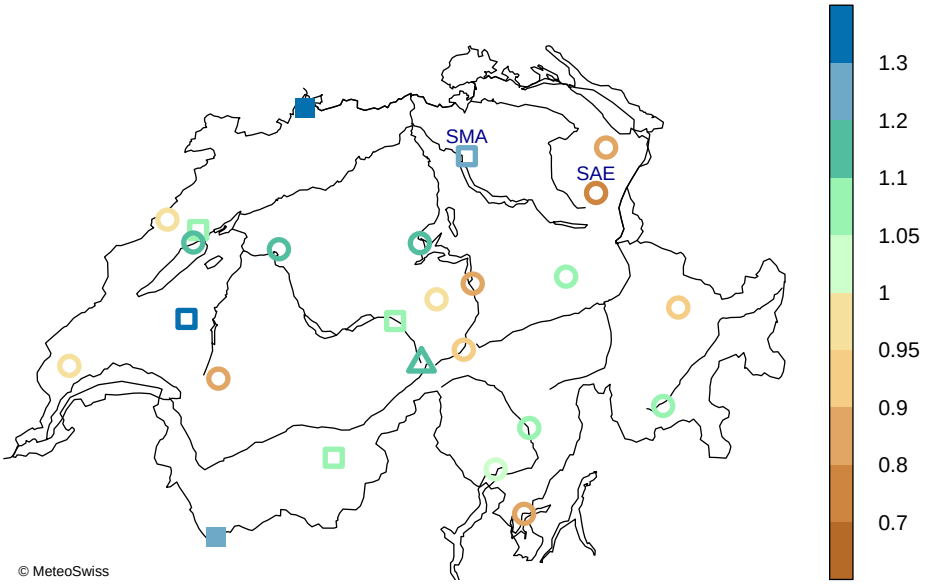


Figure 34: Ratio ML_{SP}/LR_{ZGR} of the 100-year return levels of 1-day precipitation estimated by Maximum Likelihood over the standard period 1966-2015, and by linear regression of plotting points in ZGR (mostly 1901-1970). Example: in Zürich / Fluntern, with station code SMA (Säntis, station code SAE), ZGR would underestimate (overestimate) the 100-year daily precipitation by 20-30%. Circles/squares/triangles: good/questionable/poor reliability of the ML_{SP} estimates. Filled (empty) shapes: the LR_{ZGR} return level is outside (inside) of the ML_{SP} confidence intervals and the difference is therefore significant (non-significant).

The discrepancy between ZGR and CE can also be expressed in terms of return periods. For both 1 day precipitation at Altdorf, and 5 day precipitation at la Chaux-de-Fonds for instance, switching from ZGR to CE corresponds to a substantial increase in return period. Figure 35 shows the “new” return period the ZGR 100-year return level has when estimated with Maximum Likelihood applied to the standard period ML_{SP} on CE. It appears that over a large part of the Plateau, heavy precipitation events are considerably more frequent than suggested by ZGR.

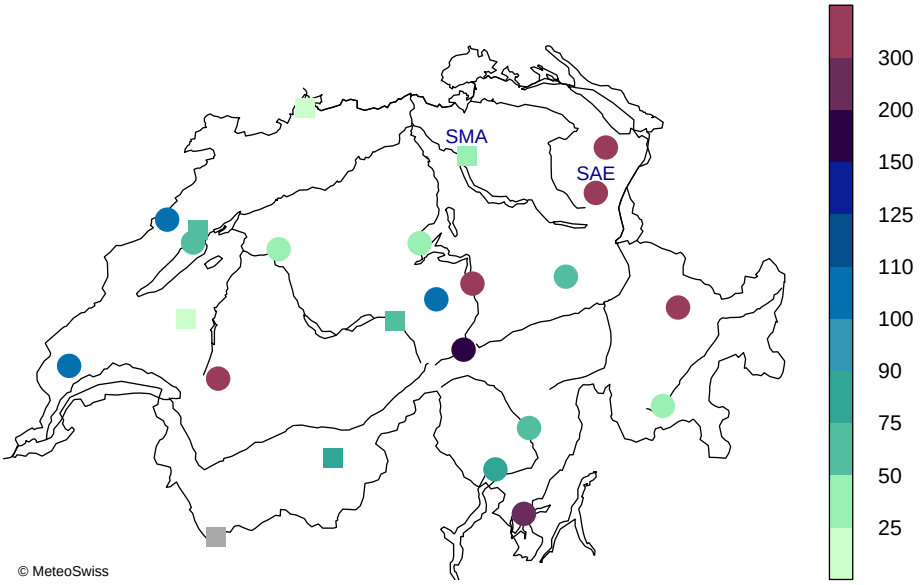


Figure 35: Return period according to CE of the ZGR 100-year return levels for 1-day precipitation. Grey indicates a return period $<$ 10 years. Example: in Zürich / Fluntern, station code SMA (Säntis, station code SAE) a precipitation amount that was exceeded on average every 100 years according to ZGR would be exceeded on average every 25-50 years (every 300 years or less) according to CE. Filled circles/squares: good/questionable reliability of the ML_{SP} estimates (stations with estimates of poor reliability are omitted). The significance of the difference in return periods cannot be assessed because confidence intervals are only estimated for return levels.

5.3.2 Sub-daily Analyses

The comparisons of sub-daily analyses look quite different from the ones for supra-daily precipitation. As can be seen in figure 36 for 1 hour precipitation, the return level on CE is lower than the return level in ZGR at more than half the stations. For 10-minute precipitation, this proportion is nearly 80%. In addition, the differences tend to be large everywhere, exceeding 20% (regardless of sign) at nearly 70% (80%) of stations for 1 hour (10 minute) precipitation. Viewed in terms of return periods, figure 37 shows that a 100 year event in ZGR often corresponds to an event with a return period above 300 years for ML_{SP}. These raw differences between 100-year return levels on ZGR and CE are significant at about 20% of the stations for 1 hour precipitation, but at more than half the stations for 10 minute precipitation.

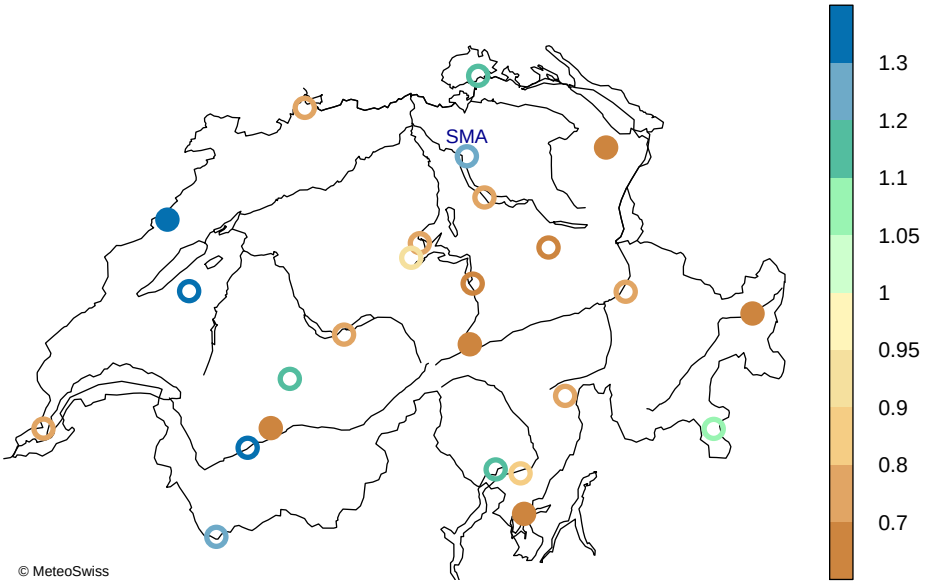


Figure 36: Ratio ML_{SP}/LR_{ZGR} of the 100-year return levels of 1-hour precipitation estimated on CE by Maximum Likelihood over the period 1981-2015, and by extrapolation of daily intensities in ZGR (mostly 1901-1970). Example: in Zürich / Fluntern (station code SMA), ZGR would underestimate the 100-year hourly precipitation by 20-30%. Only stations with a reliable estimation of ML_{SP} are represented. Filled (empty) circles: the LR_{ZGR} return level is outside (inside) of the ML_{SP} confidence intervals and the difference is therefore significant (non-significant).

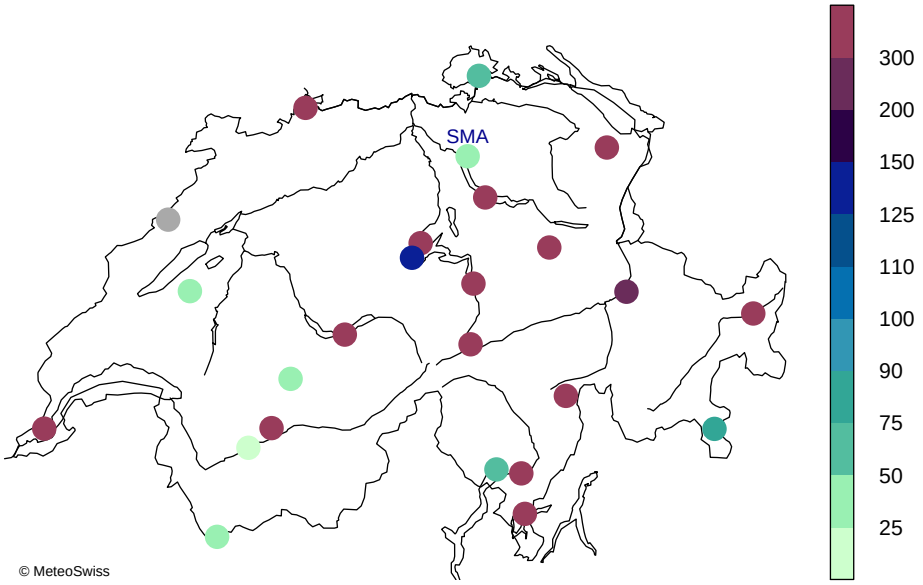


Figure 37: Return period according to CE of the ZGR 100-year return levels for 1-hour precipitation. Grey indicates a return period $<$ 10 years. Example: in Zürich / Fluntern (station code SMA), a 100-year return level of 1-hour precipitation according to ZGR has a return period of 25-50 years according to CE. Only stations with a reliable estimation of ML_{SP} are represented. The significance of the difference in return periods cannot be assessed because confidence intervals are only estimated for return levels.

5.4 Conclusion

The comparison of return levels presented on CE with those in ZGR at 25 (33) stations of the MeteoSwiss observational network for supra-daily (sub-daily) precipitation has shown that differences can be large, especially for sub-daily precipitation. For supra-daily precipitation, only few of the raw differences are significant, while for sub-daily precipitation a large proportion - up to half, depending on the precipitation duration considered - is.

For supra-daily precipitation, further investigation in the causes underlying the raw difference in return levels revealed that the difference due to the time period analysed is generally much larger than the difference due to the method. Surprisingly, the naïve linear regression approach usually yields very similar results to the maximum likelihood best estimate over the same time period. This should not, however, justify the use of linear regression of plotting points for the estimation of return levels. Firstly, the estimate depends on the successful guess of the type of distribution. Secondly, maximum likelihood takes into account the lack of information in the tail of the distribution, and provides a more reliable framework for estimating uncertainties. Even if ZGR provided information on uncertainties, these would be strongly underestimated, especially for high return periods.

Although we can only speculate, it is probable that the method used to derive the sub-daily return levels in ZGR plays an important role in the resulting difference with CE results. These are extrapolated from supra-daily return levels of durations from 1 day to a week. The extrapolation is based on the assump-

tion that intensities at different precipitation durations are related by a scaling factor depending on the ratio between these durations. Above all, the information comes entirely from supra-daily precipitation, which is governed by different processes than short-term precipitation.

For supra-daily precipitation, no obvious spatial pattern could be discerned. For sub-daily precipitation, on the other hand, a large number of stations in the Prealps and south-eastern Switzerland display lower return levels on CE than than in ZGR. This would suggest that, in these regions, design based on ZGR information was and still is safe, but was perhaps more costly than necessary. In the Plateau, on the other hand, the few stations examined hint at a possible underestimation of return values in ZGR.

We strongly caution the reader not to assume that stations located between the 25 (or 33, for sub-daily data) stations analysed here behave in one way or the other. MeteoSwiss has implemented a tool for rapid analysis of the differences in return levels between ZGR and CE at all ZGR stations that are also on CE. Interested users are kindly requested to order ZGR-vs-CE comparisons for stations of their choice at klimainformation@meteoswiss.ch.

Abbreviations

CI	confidence interval
EVS	extreme value statistics
EVT	extreme value theory
GEV	generalized extreme value (distribution)
GPD	generalized Pareto distribution
BM	Block Maxima
ℓ	log-likelihood function
mle	maximum likelihood estimation
ML	maximum likelihood
LR	linear regression
PP	probability-probability
QQ	quantile-quantile
GoF	goodness-of-fit
RMSD	root mean square difference
CE	web platform `http://climate-extremes.ch`
ZGR	collection of analyses by *Zeller et al.* (1976-1992) in 9 volumes published by WSL
SP	Standard Period

station abbreviation	station name	longitude	latitude	altitude [m]
ALT	Altdorf	8.62	46.89	438.00
ANT	Andermatt	8.58	46.63	1438.00
BAS	Basel / Binningen	7.58	47.54	316.17
BER	Bern / Zollikofen	7.46	46.99	552.84
CDF	La Chaux-de-Fonds	6.79	47.08	1018.00
CHD	Château-d'Oex	7.14	46.48	1029.00
CHM	Chaumont	6.98	47.05	1136.00
DAV	Davos	9.84	46.81	1594.16
ELM	Elm	9.18	46.92	958.00
ENG	Engelberg	8.41	46.82	1035.66
GRC	Grächen	7.84	46.20	1605.00
GRH	Grimsel Hospiz	8.33	46.57	1980.00
GSB	Col du Grand St-Bernard	7.17	45.87	2472.00
GVE	Genève-Cointrin	6.13	46.25	420.00
LUG	Lugano	8.96	46.00	273.00
LUZ	Luzern	8.30	47.04	454.00
MER	Meiringen	8.17	46.73	588.60
NEU	Neuchâtel	6.95	47.00	485.00
OTL	Locarno / Monti	8.79	46.17	366.75
PAY	Payerne	6.94	46.81	490.00
RAG	Bad Ragaz	9.50	47.02	496.00
SAE	Säntis	9.34	47.25	2502.00
SBE	S. Bernardino	9.18	46.46	1638.70
SIA	Segl-Maria	9.76	46.43	1804.00
SIO	Sion	7.33	46.22	482.00
SMA	Zürich / Fluntern	8.57	47.38	555.95
STG	St. Gallen	9.40	47.43	775.66

List of Figures

List of Tables

References

Begert, M. (2008), Die Repräsentativität der Stationen im Swiss National Climatological Network (Swiss NBCN), *Technical Report MeteoSwiss, 217*, 40 pp.

Begert, M., G. Seiz, N. Foppa, T. Schlegel, C. Appenzeller, and G. Müller (2007), Die Überführung der klimatologischen Referenzstationen der Schweiz in das Swiss National Climatological Network (Swiss NBCN), *Technical Report MeteoSwiss, 215*, 43 pp.

Coles, S. (2001), *An introduction to statistical modelling of extreme values*, Springer Series in Statistics.

Fisher, R., and L. Tippett (1928), On the estimation of the frequency distributions of the largest or smalles member of a sample, *Proceedings of the Cambridge Philosophical Society*, *24*, 180 – 190.

Geiger, H., J. Zeller, and G. Röthlisberger (1991), *Starkniederschläge des schweizerischen Alpen- und Alpenrandgebietes. Einführung, Methoden, Spezialstudien*, vol. 7, p. 334, Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft, WSL, Birmensdorf.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013), *Bayesian Data Analysis*, 675 pp., Chapman and Hall, Boca Raton.

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, *44*, 423 – 453.

Hosking, J. (1990), L-moments: analysis and estimation of distributions using linear combinations of order statistics, *J. R. Statist. Soc. B*, *52*(1), 105–124.

Jenkinson, A. (1955), The frequency distribution of the annual maximum (or minimum) values of meteorological events, *Quarterly Journal of the Royal Meteorological Society*, *81*, 158–172.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983), *Extremes and related properties of random sequences and processes*, Springer Verlag, New York.

Luceno, A. (2006), Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators, *Computational Statistics and Data Analysis*, *51*, 904–917.

Salsburg, D. (2001), *The lady tasting tea*, 312 pp., Owl Books, New York.

Umbricht, A., S. Fukutome, M. A. Liniger, C. Frei, and C. Appenzeller (2013), Seasonal variation of daily extreme precipitation in Switzerland, *Scientific Report MeteoSwiss, 97*, 122 pp.

von Mises, R. (1954), *La distribution de la plus grande de n valeurs.*, vol. II, pp. 271–294, American Mathematical Society, Providence, RI.

Zeller, J., H. Geiger, and G. Röthlisberger (1976-1992), *Starkniederschläge des schweizerischen Alpen- und Alpenrandgebietes*, 9 vols, Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft, WSL, Birmensdorf.

Acknowledgments

We thank Thomas Schlegel and Stephan Bader for sharing their knowledge of the Swiss Climate in numerous discussions, Michael Begert whose expertise on data quality was essential for describing MeteoSwiss stations and measurements, and Deborah van Geijtenbeek for patiently providing the details on the data handling. We are indebted to all those who have contributed in some way to creating and maintaining a dense network of measuring stations, and to those who spend countless hours providing us with high quality data in the form that we need. We thank Moritz Bandhauer and Chritoph Frei for their careful reading and critical comments that helped improve the readability and consistency of the report.

MeteoSchweiz
Operation Center 1
CH-8044 Zürich-Flughafen
T +41 58 460 99 99
www.meteoschweiz.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno Monti
T +41 58 460 97 77
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genève 2
T +41 58 460 98 88
www.meteosuisse.ch

MétéoSuisse
Chemin de l'Aérologie
CH-1530 Payerne
T +41 58 460 94 44
www.meteosuisse.ch