https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions

Regionalisation of Rainfall Depth-Duration-Frequency curves in Germany

Bora Shehu[1], Winfried Willems[2], Henrike Stockel[2], Luisa Thiele[1], Uwe Haberlandt[1]

[1] Institute of Hydrology and Water Resources Management, Leibniz University Hannover Germany

[2] IAWG, Engineering Hydrology, Applied Water Resources and Geoinformatics, Ottobrunn Germany

Correspondence to: Bora Shehu (shehu@iww.uni-hannover.de)

Abstract.

Rainfall depth-duration-frequency (DDF) curves are required for the design of several water systems and protection works. These curves are typically generated from the station data by fitting a theoretical distribution to the annual extremes (AMS). The aim of this study is to investigate the use of different data types and methods for estimating reliable DDF curves covering whole Germany. The following three questions are investigated for the evaluation and regionalisation of the DDF curves in Germany: i) which is the best local estimation method, ii) which regionalisation method shows best performance, and iii) which data sets should be used and how they should be integrated. For this purpose, two competitive DDF-procedures for local estimation (Koutsoyiannis et al. 1998, Fischer and Schumann, 2018) and two for regional estimation (kriging theory vs index-based) are implemented and compared. Available station data from the German Weather Service (DWD) for Germany are employed, which includes; 5000 daily stations with more than 40 years available, 1261 high resolution (1min) recordings with observations period between 10 and 20 years, and finally 133 high resolution (1min) recordings with 60-70 years of observations. The performance of the selected approaches is evaluated by cross-validation, where the local DDFs from the long sub-hourly time series are considered the true reference. The results reveal that the best approach for the estimation of the DDF curves in Germany is by first deriving the local extreme value statistics based on Koutsoyiannis et al. 1998 framework, and later use the kriging regionalisation of long sub-hourly time series with the short sub-hourly time series acting as an external drift. The integration of the daily stations proved to be useful only for DDF values of very low return period (T<10 years), but not doesn't introduce any improvement for higher return periods (T≥10 years).

Keywords:

Depth-Duration-Frequency, Regionalisation, Disaggregation, Kriging, Index-based

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions

1. Introduction

Rainfall volumes at varying duration and frequencies are required for the design of water management systems and facilities, like dams or dikes, spillways, flood retention basins, urban drainage systems, etc. These design precipitation volumes are also known as IDF (Intensity-Duration-Frequency) or DDF (Depth-Duration-Frequency) curves. The main application of the DDF curves is the derivation of design discharge from design rainfall, when no sufficient discharge observations are available assuming that both rainfall and discharge events have the same recurrence interval (herein referred to as return period Ta). For sampling, the annual maximum series (AMS) or peak-over-threshold (POT) can be used, however for return periods greater than 10 years, there are hardly any differences in the results obtained from each method. Often the AMS are preferred over the POT because the methodology is more direct and easier, whereas the POT method needs a prior assumption on the threshold selection. The typical procedure includes fitting a theoretical probability distribution (PDF) to the observed rainfall extremes at a certain duration level, and based on the obtained PDF, compute the quantiles corresponding to different return periods. Most common distribution functions are Generalised Extreme Value (GEV), Gumbel, Log-Pearson-III and Lognormal distributions for AMS, with GEV and Gumbel being the most popular, and Generalised Pareto for POT. L-moments are primarily used for parameter estimation in recent national applications (Johnson and Sharma, 2017). Since the estimation of extreme design rainfall is done locally at each measurement station (rain-gauge), a regionalisation method, often the index-flood method (herein referred the index method) is employed to estimate design rainfall depth at ungauged location (Hosking and Wallis, 1997). In Germany, the Coordinated Heavy Rainfall Regionalisation Evaluation KOSTRA-DWD (Malitz and Ertel, 2015) from the German Weather Service (DWD) has been providing these design precipitation volumes for different application purposes since 1980. A revision of the KOSTRA-DWD is required, in order to consider the recent state-of-the-art and additional data. Therefore, it is the aim of this study to investigate the use of different methods for the estimation and regionalisation of the DDF curves and the best integration of different data types, in order to give the basis for the development of the new regional design rainfall catalogue for Germany. In this procedure, several research questions arise which are discussed below:

i) Local estimation

A prominent probability distribution that is frequently used in the statistical analysis of AMS of heavy rainfall is the Gumbel distribution. The Gumbel distribution is a special case of the three-parameter GEV distribution where the shape parameter is zero (γ=0) and the distribution follows an exponential tail behaviour. If the shape parameter is greater than zero, the distribution exhibits a so-called heavy-tail behaviour (also known as GEV type II), whilst if the shape parameter is less than zero no-tail behaviour is present (also known as GEV type III)(Coles, 2001). The GEV type III is not employed in rainfall extreme value statistics, as it is bounded from above. The Gumbel and the GEV type II (herein referred to as simply GEV) are almost similar for low percentiles, nevertheless diverge greatly for high return periods. Therefore, for the design rainfall at high return periods, the expression of the shape parameter is of decisive importance. Regarding this issue, extensive investigations were carried out to determine the role of the shape parameter in heavy precipitation data, both in a theoretical manner and on the basis of empirical findings. For instance, Koutsoyiannis (2004a) investigated the heavy-tail behaviour of extreme daily rainfall values at 169 worldwide locations with very long observations (100-150 years) and concluded that when only short observations are present (less than 50 years) the heavy-tail characteristics can be overlooked and the Gumbel distribution is chosen falsely as a good fit. This may be also the reason why for a long time in the literature mainly the Gumbel distribution was preferred. Koutsoyiannis (2004b) proposed a GEV distribution with a shape parameter fixed within the range γ=0.1-0.15 for all examined geographical zones (mainly in Europe and North America). Specifically, he proposes the value of 0.15 if very high return periods are of interest, and the value of

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

0.1 if the focus is also on low return periods. Later, Papalexiou and Koutsoyiannis (2013) analysed more than 15,000 stations worldwide with observation length from 40 to 160 years, and again the results favoured the implementation of heavy-tail GEV distribution instead of the Gumbel. A recent study by Papalexiou (2018) on hourly rainfall measurements in the USA, suggested that also for sub-daily durations, the rainfall extremes exhibit a heavy-tail (sub-exponential), much heavier than the exponential or the Gamma tails. Mountain areas tend to exhibit heavier tails; however, terrain is not the dominant factor influencing the tail behaviour. Overall, the analysis suggests that the shape parameter cannot be evaluated adequately when the station's recordings are short and from a Gumbel distribution, therefore the GEV should be used instead.

To determine the design rainfall, distribution functions are usually first fitted separately for each of the selected duration levels. This way, quantile crossing may arise between different duration levels (Cannon, 2018). Quantile crossing here refers to when the extreme rainfall volumes for a fixed return period (say Ta=100 years) are not increasing with longer duration levels. Theoretically, the rainfall volume is dependent on the duration and thus another step in the extreme value analysis is needed, to ensure that the extremes are consistently increasing with the duration levels. An empirical relationship was first developed by Bernard (1932), where the intensities at different duration levels are generalised by a power law depending on three location constants. (Koutsoyiannis et al., 1998) proposed a similar mathematical framework, where the AMS intensities are generalised based on two parameters ($\theta > 0$ and $0 < \eta < 1$) and a probability distribution function (PDF) is fitted based on these generalised intensities to estimate the quantiles for specific return periods. The generalised concept suggested by Koutsoyiannis has widely been implemented in the literature (Asikoglu and Benzeden, 2014; Muller et al., 2008; Ulrich et al., 2021; Van de Vyver, 2015). Ulrich et al. (2021) implemented such a framework in Germany for both monthly and annual IDFs curve, with a constant shape parameter of 0.11 for the annual estimation. Another alternative application is based on the wide sense scaling theory, where the PDF parameters or moments of each duration are dependable on a power law (Gupta and Waymire, 1990). Van de Vyver (2015) implemented a multi-scaling approach, where both the location and the scale parameters of duration specific GEV were related on a power law with the duration, while the shape parameter was kept constant. Similar approaches were also proposed and studied by Haktanir et al. (2010), Holešovský et al. (2016), Soulis et al. (2016), and are typically referred to as smoothing of extreme statistics over the duration levels.

Other solutions build also on the power law relationship between extremes and durations are for instance Bayesian distribution models (Boukhelifa et al., 2018; Lima et al., 2016; Roksvåg et al., 2021; Van de Vyver, 2018), marginal probabilities (Veneziano et al., 2007), and artificial intelligence (Cannon, 2018). An alternative approach for achieving a DDF based on data from example of such implementation in Germany was proposed by Fischer and Schumann (2018), where location and scale parameters are obtained by a regression model (based on a nonlinear least squares method), and the shape parameter is estimated indirectly by quantifying first the normalised scale/shape ratio with a robust linear regression. Here we consider the two approaches of Koutsoyiannis et al. (1998) and Fischer and Schumann (2018), as they have successfully been tested in Germany. Here, the question remains whether a homogenisation of intensities or a smoothing of GEV parameters across different duration levels is more appropriate for the estimation of the DDF curves in Germany.

ii) Regionalisation methods

Regionalisation of the design DDF curves provides estimation for unobserved locations, but also contributes to a more robust estimation, e.g. by using larger samples (Requena et al., 2019). Methodologically, a distinction can be made between two approaches: a) a direct regionalisation of quantiles, moments or parameters of distribution functions and b) a regional estimation of distribution functions for homogeneous regions. A direct regionalisation of quantiles may lead as well to quantile crossing across durations, and therefore mostly regionalisation of parameters is performed. Furthermore Borga et al. (2005) suggests the regionalisation of the parameters instead of the quantiles. For the direct regionalisation of parameters, regressions (Madsen et al., 2009; Smithers and Schulze, 2001), splines (Johnson and Sharma, 2017) or geostatistical methods (Ceresetti et al., 2012; Kebaili Bargaoui and Chebbi, 2009; Uboldi et al., 2014; Watkins et al., 2005) are applied. On the other hand, the estimation of regional distributions functions based on the index method proposed by Hosking and Wallis (1997), is one of the most used methods in the literature for the regionalisation of design precipitation (Burn, 2014; Durrans and Kirby, 2004; Forestieri et al., 2018; De Salas and Fernández, 2007). Many countries have actually employed the index-based regionalisation for estimation of regional IDF/DDF curves, for instance Canada (Burn, 2014), Denmark (Madsen et al., 2009) and USA (Perica et al., 2019). However, prior to the application of the index method, it is important to define adequately homogeneous regions where the rainfall statistics are similar, which can be a challenging task (De Salas and Fernández, 2007). Hosking and Wallis, (1997) recommend that site characteristics should be used for the identification of homogeneous regions instead of site statistics. Therefore, the second objective of this paper is to investigate whether a direct kriging interpolation of the GEV parameters or the application of the index-method on homogeneous regions is more suitable for the estimation of regional DDF curves in Germany.

iii) Combination of available datasets with different temporal resolution and observation length

As stated in Koutsoyiannis (2004a,b) short time series can choose Gumbel parameters falsely and hide the true heavy-tail behaviour of rainfall extremes. Thus, care should be taken when combining different statistics from different observation lengths. Madsen et al., (2017) investigated the IDF curves with long stations (more than 40 years) and short stations (less than 30 years) based on Generalised Pareto distribution with fix shape parameter, and concluded that the statistics are changing from one case to the other, with short series giving large estimates of the extreme intensities. This of course can be attributed to the non-stationarity of the IDF curves. Holešovský et al. (2016) separated the historical data into groups when estimating IDF curves for Czech Republic (long series with 35-40 and short series with 11-15 years of observations), and concluded that the uncertainty at estimating parameters for the short time series is quite high, especially for high return periods. In the index-based regionalisation, regional L-moments are averaged based on the observation length, which may lead to more stable results (Burn, 2014; Requena et al., 2019), however the interpolated index may still suffer from high uncertainties from pooling together short and long time series. This may also be the case when interpolating local GEV parameters with the kriging theory. Therefore, it is important to investigate which is the best combination of time series with different observation length: even though the short time series may be not adequate for high return period quantiles, they are much denser than the longer time series. Hence their information may be helpful in trading space for time.

In Addition to the high resolution (1-5min) network, the daily one is much denser and as well with very long observation lengths. Nevertheless, the temporal resolution is too coarse for the estimation of sub-hourly to sub-daily extremes. In such cases, GEV parameters for the sub-daily duration can be scaled from the GEV parameters of the daily extremes following the scale invariance principle of precipitation extremes. Bara et al. (2009) employed the scale invariance principle to derive DDF curves for sub-daily duration levels (5min – 3h) from daily observations in Slovakia, while Borga et al. (2005) applied two different scaling factors one for duration levels less than 1 hour and one for longer than 1 hour in northern-eastern Italy. A later study from Paixao et al. (2011) performed in Ontario Canada concluded that the scaling factors should not be used for reliable downscaling of daily extremes to durations less or equal to one hour. This is because the extremes at such short durations are governed by other rainfall mechanisms then the daily extremes, and hence a low dependency exists between the two extreme groups. Alternative to the scale invariance principle, disaggregation schemes can be applied to the daily data in order to obtain high resolution data. Various model approaches for disaggregation are

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions

described in the literature, and they mostly consist of a so-called cascade model (Müller and Haberlandt, 2018). Weather radar data can be used to estimate the probabilities in the individual levels and to derive the extensive parameter-sets suggested by Lisniak et al. (2013) for the disaggregation scheme. Therefore, the third objective of the paper is to investigate the value and the best combination of data from the long, short and disaggregated daily series for the regionalisation of the DDF curves in Germany.

The paper is structured as follows: first the available data sets for extreme value analysis are introduced in Section 2, then the methods selected for investigation of the local and regional estimation are presented respectively in Section 3.1 and 3.2, with performance assessment and validation explained in Section 3.3. The results are given for each objective as: best local estimation of extremes in Section 4.1, best regionalisation technique 4.2.1, best data integration 4.2.2. Finally, the obtained maps for Germany are discussed in section 4.3 and concludes in Section 5.

2. Study Area and Data

2.1 Available Data

The study area covers Germany and is illustrated in **Figure 1**. Three rainfall measuring networks are available from the German Weather Service (DWD); the daily network (DS) – typically Hellman devices recording the rainfall daily, the long network (LS) – mostly tipping bucket analogue sensors (before 2004) measuring rainfall at 1 min time steps with 0.1mm resolution and 2% uncertainty, and the most recent short network (SS) – digital sensors (after 2004) measuring rainfall also at 1min timesteps with 0.01mm resolution. The spatial distribution of these networks is shown in **Figure 1**, the observation length is given respectively in **Figure 2** and the number of stations available for each network is given in **Table 1**. The long network is the most appropriate data set for extraction and evaluation of extreme rainfall statistics, since on average it includes 65 years of observations (as shown in **Figure 2**– dark blue) and measures the rainfall at very fine temporal scales. Nevertheless, this network is sparse and only 133 stations in the whole Germany are available. On the other side the short network measures the rainfall as well at very fine temporal scales and is much denser than the long network (1261 stations excluding the LS locations), however on average it includes only 18 years of observations which is not enough for extreme value statistics. Lastly the daily network is much denser (with 4068 stations excluding LS and SS locations) and covers 40 years up to 120 years, but the temporal resolution of rainfall is too coarse to be useful for sub hourly extreme values analysis.

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

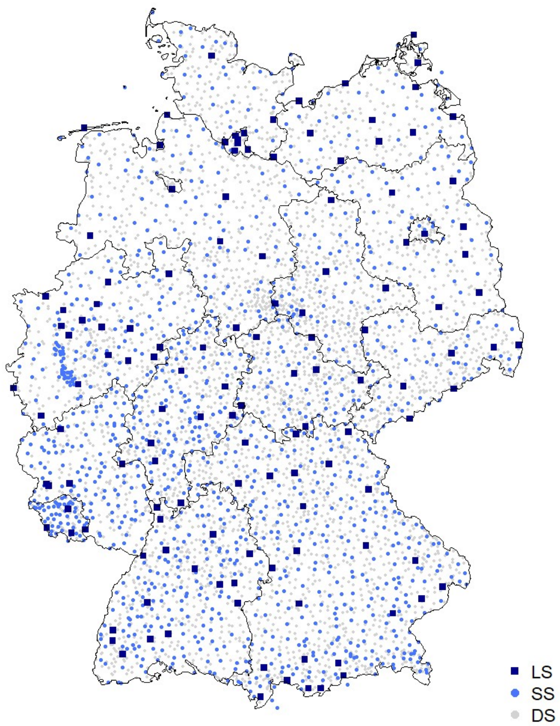

Figure 1 Available rainfall networks in Germany for different temporal resolution. The black lines illustrate the borders of German Federal States.

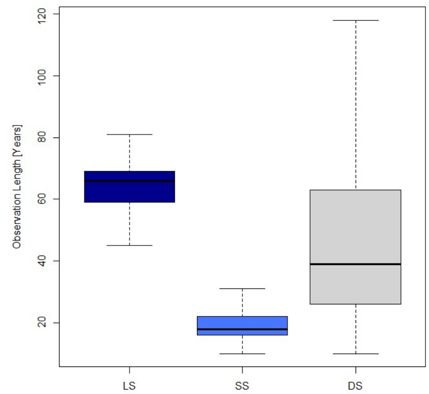

Figure 2 *Observation length of all stations grouped according to the three available networks in Germany.*

Table 1 *Number of stations for each of the available networks in Germany.*

Resolution	5min		1 day
Obs. Length	> 41y	> 10 y	>10 y
No. Gauges	133	+1261	+4068

2.2 Temporal Disaggregation of the Daily Network

The daily network is much denser than both long and short networks and includes even longer observation periods than the long network. If it is possible to disaggregate these data reliably, this will considerably increase the number of support points for the regionalisation of DDF curves. For the considerations presented here, the so-called cascade model first introduced by Olsson (1998) is employed. A more extensive parameterisation is implemented in the method according to Lisniak et al. (2013) which corresponds to a transfer of the Olsson method to a 3-fold distribution. To generate sub-hourly data, disaggregation parameters are derived from the RADOLAN radar time series of each grid cell (Bartels et al., 2004), and the daily observed volumes are disaggregated for the given durations as shown in **Table 2**. It is important to note that, due to the parameterisation using RADOLAN data, no parameter regionalisation is required, so that the parameter-rich disaggregation procedure in the Lisniak variant can be used. Moreover 30 realisations of disaggregated data were generated for each duration, in order to capture the uncertainty due to the disaggregation.

Table 2 *The disaggregation scheme applied to the daily network (DS) to obtain rainfall volumes at the given durations.*

Duration	12h	8h	6h	4h	3h	2h	1h	30min	15min
Disaggregation	24h /2	24h /3	$24h/2^2$	24h /3/2	$24h/2^3$	$24h / 3/ 2^2$	$24h/3/2^3$	$24h/ 3/2^4$	$24h/ 3/2^5$

To understand what errors can be introduced to the DDF curves when employing this disaggregation scheme, a direct comparison was conducted between the long series (LS) and the disaggregated series (DS) for the return periods 1, 10, 20, 50 and 100 years. For each station, duration level and return period, the relative error is calculated as the difference

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions

between the disaggregated and the original rainfall quantile. The resulting deviations for all stations are shown in **Figure 3**. The results indicate that at the longer duration levels (>6 hours), the DDF curves are captured quite well, and the main disadvantage of the disaggregation model (as expected) is for the very short duration. Below the duration of 4 hours, there is a clear tendency to underestimate the extremes. Thus, it is expected for the DS disaggregation scheme to be more useful for the longer duration extremes than the short ones. This is particularly true for extremes at very short duration (5min) as the disaggregation scheme estimates volumes only down to 15 min durations.

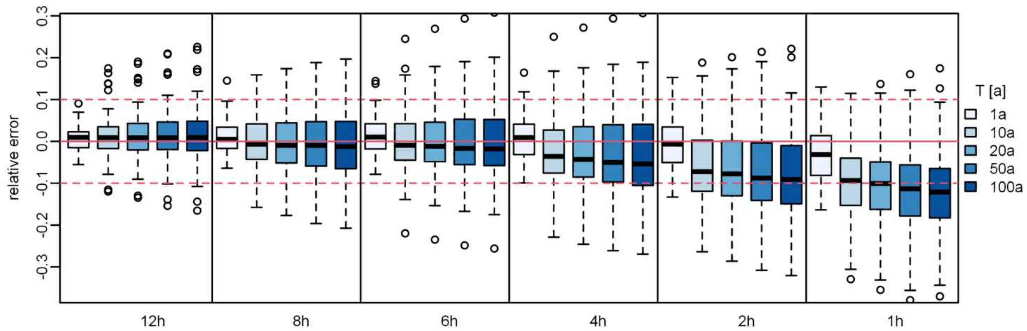

Figure 3 *The relative error of the disaggregated daily station data (30 realisations) based on radar parametrisation for different return periods and duration levels.*

2.3 Annual Maximum Series for Each Dataset

Using the five-minute time series, annual maximum series (AMS) are derived based on the calendar year for the duration levels 5min, 10min, 15min, 30min, 1h, 2h, 6h, 12h, 1d, 2d, 3d and 7d. A moving window with the length of each duration level is used to derive the annual maxima, considering a dry duration of 4 hours to ensure that the maxima selected are independent from one another. Additionally, following the guidelines given by DWA (2012) a scaling of the durations 5, 10 and 15 min AMS with the factors given in **Table 3** is performed. This is done to avoid the systematic underestimation of rainfall extremes at short duration caused by the deviation between i) the start of the actually largest rainfall sum of duration D, and ii) the fixed starting time of the 5 min time series (employed here).

Table 3 *Correction factors for the short duration AMS according to the DWA-531(DWA, 2012).*

Duration level	**5min**	**10min**	**15min**
Correction factor for AMS	1,14	1,07	1,04

2.4 Homogenisation of Long and Short Network

First plausibility and homogeneity checks were performed on the long and short data sets, herein referred to as respectively long series (LS) and short series (SS). An initial analysis of possible trends based on the quantile regression (Koenker, 2005) was carried out for the monthly 5min maximum intensities of the long series (LS). This method was chosen, as in comparison to the classical regression it is considerably more robust and it allows to obtain regression results for different non-exceedance probabilities. In **Figure 4**, the quantiles for the non-exceedance probabilities $\tau = 0.5$ (i.e. median), 0.8, 0.9 and 0.95 are considered. Quantile regressions for the four selected τ with time as the explanatory variable are implemented separately for each of the 133 measurement points. Each dashed line corresponds to a measuring station and each colour to a non-exceedance probability. Trend-like changes in the monthly five-minute maxima are visible with slopes that increase with τ. To understand why this trend is present in almost all long series, we investigated whether these instationarities are more trend-like or jump-like, with the latter assuming that the timing of jumps is associated with

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions

sensor changes in the measuring network. In the long network, a total of 19 different sensor types are distinguished simply by two states: analogue or digital.

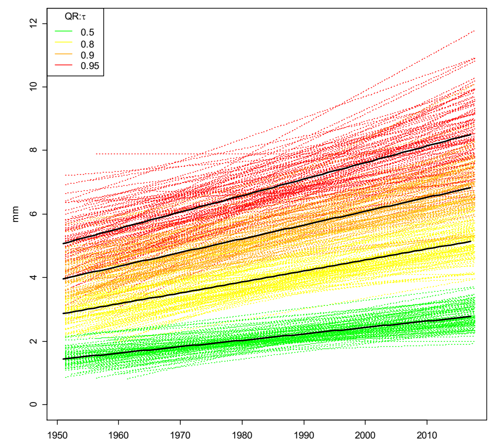

***Figure 4** Quantile regression on monthly maximum 5 min rainfall intensities for the long series (LS) for different non-exceedance probabilities (shown in coloured dashed lines). The fitted quantile regression is shown with solid black line.*

A test for trend, jump or stationarity based on in-stationary extreme value analysis (Coles, 2001) was performed for all 133 LS. We tested for linear trend in location parameter vs. jump at date of sensor change in the location parameter vs stationarity. The decision was based on Akaike Information Criterion. The results for different duration levels (x-axis) are shown in **Figure 5** –left. It is obvious that the majority of instationarity at short duration levels is better explained as a jump (with mostly positive sign) in the data. A possible reason could lie in the limited ability of analogue gauges to register abrupt intensity changes. Since the instationarities are usually jumps and not trends, a simple homogenisation of the data to a uniform sensor type is possible by raising to the mean value of the digital sensor type (DVWK, 1999). This jump correction is applied separately for each station and duration level. The results of applying the instationarity test to the homogenised series are shown in **Figure 5**– right. It can be seen that this approach can eliminate the instationarities at short duration levels significantly. About 30% of the stations show instationarities (either trend of jump), while the remaining part is considered stationary. Since only a small part of the stations show instationarities, here a stationary extreme value analysis is performed.

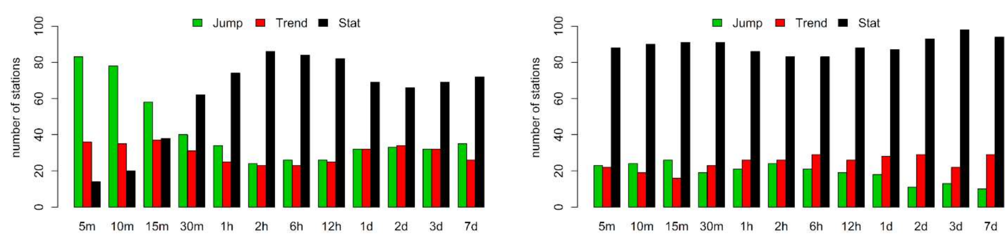

***Figure 5** Trend vs Jump Analysis (%) for left) - before jump elimination, right) after jump elimination.*

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

3. Methods

3.1 Local Estimation of Extreme Value Statistics

3.1.1 Reference Approach

Here, the Generalised Extreme Value (GEV) probability distribution is used for the statistical analysis of extreme rainfall and the derivation of the local DDF curves, that is described as following:

$$F(x;\mu,\sigma,\gamma) = exp\left\{-\left[1+\gamma\frac{(x+\mu)}{\sigma}\right]^{-\frac{1}{\gamma}}\right\},\ 1+\frac{\gamma(x-\mu)}{\sigma} > 0, \gamma \neq 0, \quad (1)$$

where μ is the location, σ the scale and γ the shape parameter. If the shape parameter is greater than zero, heavy-tail behaviour is present (GEV type II); if the shape parameter is less than zero, then it is the reverse Weibull distribution with no-tail behaviour (Coles, 2001). The GEV parameters are fitted to the AMS of each duration level and station separately, based on the L-moments method. For this purpose, the R-package "lmomco" was used (Asquith, 2021). A prior investigation on our study revealed that the L-moment approach led to more stable results than the method of Maximum Likelihood. The shape parameter was either estimated or fixed at 0.1 for estimation of return periods up to 100 years, approximately following the recommendation from Koutsoyiannis (2004a, b) and on a prior analysis conducted on LR series. Based on the parameters obtained the quantiles of return periods T1a, T10a, T20a, T50a and T100a were derived. Since the AMS-approach tends to underestimate quantiles at low return periods (Ta < 10 years), a correction of the AMS return periods according to the DWA 531-Regulations with factors given in **Table 4** was performed.

***Table 4** Correction of the Return Periods when fitting the GEV to the AMS adapted from (DWA, 2012).*

Return Periods for POT	Ta=1 year	Ta=5 years	Ta=10 years
Return Periods for AMS	Ta=1.6 years	Ta=5.5 years	Ta=10.5 years

As discussed previously in the introduction, because the parameters are fitted separately on each duration, quantile crossing may occur. **Figure 6** shows for different return periods T1a, T10a, T20a, T50a and T100a the number of stations affected by these crossings for the empirically calculated quantiles (left) and the quantiles fitted with the General Extreme Value (GEV) distribution (right). The empirical quantiles are calculated according to Hyndman and Fan (1996). It is clear that the number of stations with this problem increases significantly for larger return periods. In the empirical quantiles, especially the short series show quantile crossing at the long duration levels (D ≥ 24h). Here, the extremes of the duration levels D72h and D168h are lower than the extremes of the duration level D24h. With the GEV-fitted quantiles, significantly more stations show quantile crossings than with the empirically calculated quantiles. These problems occur for all return periods, however are more frequent for the return periods T50a and T100a. In order to avoid such problems two different methods are applied and compared here: the approach presented by Koutsoyiannis et al. (1998) and the approach presented by Fischer and Schumann (2018). These two methods are described below.

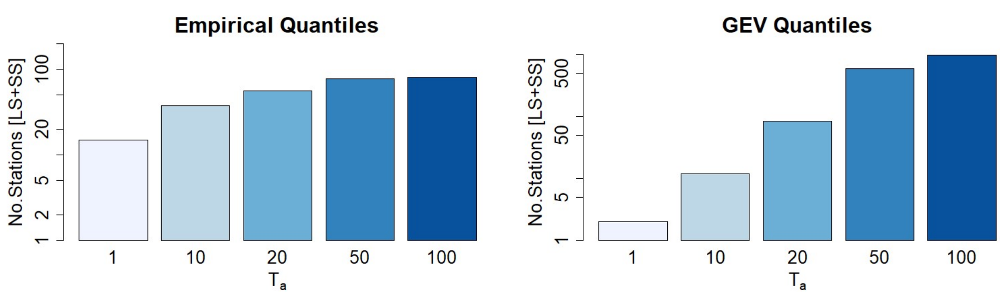

***Figure 6** Number of stations for different return periods showing quantile crossings in the empirically calculated quantiles (left) and the GEV-fitted quantiles (right) with increasing duration.*

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

3.1.2 Koutsoyiannis Approach

Koutsoyiannis et al. (1998) considers the intensity as a function of the duration level through two parameters (θ, η) and the generalised intensity can be calculated from duration specific intensity as described below:

$$i = i_d \cdot b_d \quad with\; b_d = (d+\theta)^{\eta}, \tag{2}$$

where i is the generalised intensity in mm/h, i_d is the intensity in mm/h observed at each duration level, d is the duration level in hours and Θ, η are the Koutsoyiannis parameters optimised for each station. Through this relationship a generalisation of the AMS intensities over all the chosen duration levels is possible. The parameters Θ (larger than 0) and η (within the range 0 to 1) are estimated for each station by minimising the Kruskal-Wallis statistic as indicated in Koutsoyiannis et al. (1998). The advantage of this optimisation method lies in its non-parametric character and robustness, as the Kruskal-Wallis statistics is not affected by the presence of extreme values in the sample. Once the parameters Θ and η are determined, the generalised intensities from all duration levels are pooled together (as the main assumption is now that they follow the same distribution) and a GEV distribution is fitted to this sample by the methods of L-moments. Lastly, to obtain DDF curves, the quantiles at specific return periods are estimated from the fitted GEV distribution, and are divided by the term b_d in Equation (2) (dependable on θ, η parameters and the duration level). This joint estimation of parameters over all durations should not only avoid the quantile crossings, but also make the estimation of DDF more robust.

3.1.3 Fischer/Schumann Approach

In contrast to Koutsoyiannis that treats the intensities of AMS as a function of the duration, Fischer and Schumann (2018) propose a new approach based on the GEV distribution, where the generalised GEV parameters are monotonically dependent on the GEV parameters determined for each duration level. Thus, as a first step the GEV parameters (as in Equation (1) are estimated from the L-moment methods for each duration level at each station, and then through a nonlinear regression (with two parameters α and β) each GEV parameter is related to the different duration levels as indicated by the following equations:

$$\mu_d = \frac{\alpha_\mu}{d^{\beta_\mu}},\; \sigma_d = \frac{\alpha_\sigma}{d^{\beta_\sigma}} \;\; and \;\; \frac{\sigma}{\gamma} = \alpha + \beta \cdot d, \tag{3}$$

where d is the duration level in 5min, μ_d, σ_d, γ are the GEV parameters of each duration, while α and β are the regression coefficients with $\alpha_\mu, \alpha_\sigma > 0$, $\beta_\mu, \beta_\sigma > -1$, $\beta \geq 0$. The parameters are obtained by nonlinear least-square-minimising. In addition to the shape parameter dependency shown in Equation (3), three alternative approaches are considered: a constant shape parameter over all durations, a shape parameter fixed at 0.1 and a quadratic relationship as in Equation (4).

$$\xi = a + P_1 \cdot log\,(d) + P_2 \cdot log\,(d)^{\,2}, \tag{4}$$

where P_1 and P_2 are estimated spanning across all stations and a is a station specific optimised parameter.

3.2 Regionalisation of Extreme Value Statistics

The local parameters estimated for each station (GEV parameters and generalisation parameters) make the base data set for the regionalisation of the extreme rainfall statistics. Each of these parameters is regionalised independently based on the regionalisation methods explained below, and later on, DDF maps for each duration and return period of interest are generated. The overall procedure for regionalisation is given in **Figure 7-a**, and the regionalisation methods are given in **Figure 7-b**.

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

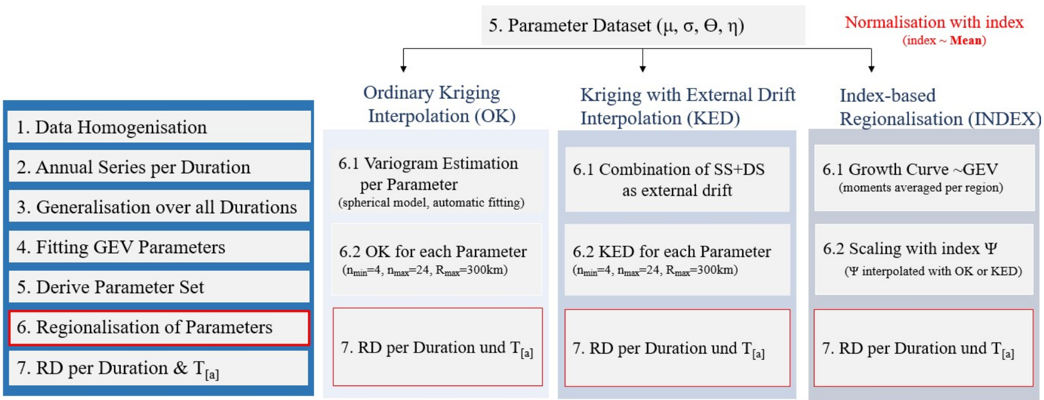

***Figure** 7 a) Overall methodology from the given data sets to DDF maps for Germany, b) a short description of the regionalisation methods applied here only for the KO.FIX local estimation of DDF; where RD is short for rainfall depth, and n_{min}, n_{max} and R_{max} are respectively the kriging parameters for minimum, maximum number of neighbours and maximum radius for neighbour search.*

3.2.1 Ordinary Kriging Interpolation

The regionalisation of extreme value statistics for Germany will first be carried out with Ordinary Kriging (OK) interpolation. Here, the extreme rainfall parameters are interpolated independently. The flow chart for this interpolation technique is shown in **Figure 7**-b. Ordinary Kriging is widely used for interpolation due to its simplicity in comparison to other kriging methods. The expected value of the random process being investigation *(E)* is treated as constant in space (as per Equation (5)), whereas the increase in variance of the target variable at any two location (u and $u+h$) depends only on the distance h. This increase in the variance is represented by the semi-variogram function $\gamma(h)$ (here called variogram). Therefore, in the first step, the empirical variogram is estimated by discrete point observations according to Equation (6).

$$E[Z(u+h)] = E[Z(u)] = m \tag{5}$$

$$\gamma\,(h) = \frac{1}{2N(h)} \sum\nolimits_{u_i-u_j=h} (Z(u_i) - Z(u_j))^2, \tag{6}$$

where N is the number of any two observed data pairs (u_i and u_j) at distance h. Since the empirical variograms are not continuous functions, theoretical variograms must be fitted to the observed values. To describe the spatial variance of the data, several theoretical variogram models can be used and fitted to the empirical variogram using the least squares method. For the interpolation of rainfall extremes a spherical variogram (as per Equation (7)) is chosen as more appropriate (Kebaili Bargaoui and Chebbi, 2009).

$$\gamma\,(h) = c_0 + c \cdot \left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) \; for\; h\; \leq a \; and \; \gamma\,(h) = \; c \; for\; h = a\;\;, \tag{7}$$

where c_0 is the nugget, c the sill and a the range of the variogram. The variogram describes the spatial variability of the target variable and the average dissimilarity between a known and unknown location. Once the theoretical variogram is known, it can be used as a basis for interpolating the statistical properties on a 5km² grid. Here, as indicated in Equation (8), the variable at an unknown location (Z') is estimated by the weighted average of the nearby known locations (Zu_i).

$$Z'(u_o) = \sum_{i=1}^{n} \lambda_i \cdot Z(u_i), \tag{8}$$

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

where the weights (λ_i) are derived from the theoretical variogram, and n is the number of selected neighbours. The R-package "gstat" is used to fit the variograms and interpolate the variables (Pebesma, 2004). An advantage of Ordinary Kriging interpolation is that the weights are determined in such a way that the difference between the estimate and the observed values is zero on average. However, this can lead to the interpolated variable being smoothed in space.

3.2.2. Kriging with External Drift Interpolation

In the Kriging with External Drift (KED), the expected value E of the target variable Z at any location u is linear dependent on secondary variables Y, and thus the Equation (5) takes the form of the Equation (9). Here the secondary variables (or the external drifts) reflect the spatial trend of the target variable. Theoretically, the variogram for KED interpolation is computed from the residuals between the target and the secondary variables. Here, for simplicity the OK variograms are used instead, since as shown in Delrieu et al. (2014) they can produce very similar results to the KED one.

$$E[Z(u) \mid Y_1(u), Y_2(u), \ldots, Y_m(u)] = b_0 + \sum_{k=1}^{m} b_k Y_k(u) \tag{9}$$

where Y represent the k secondary variables from 1 to m that is used as an external drift, and b_0 in the interception of the linear dependency and b_k the coefficient for each k drift. For this study different site characteristics (i.e. elevation) were investigated as external drift, however as indicated by the cross-correlation between the target variables (in this case the 4 parameters describing the local statistics) and the site characteristics, the linear dependency between them is not high (see in appendix **Figure A1**). Therefore, here only interpolated local parameters from the short or daily network are used as external drift information.

3.2.3 Index-based Regionalisation

The regionalisation of extreme rainfall statistics in Germany is as well carried out using the index method according to Hosking and Wallis (1997). The index method was originally developed for the regionalisation of flood quantiles, however found a wide application also for the regionalisation of extreme rainfall statistics. By pooling information in statistically homogeneous regions, a more robust estimate of extreme rainfall statistics can be made and on the basis of the regions the information can be transferred to other unobserved points. A homogeneous region exists if the distribution functions have the same shape at all points in the region. The homogeneity indicator H_1 presented by Hosking and Wallis (1997) is typically used to determine homogeneous regions. If the H_1 is lower than 1, the region is said to be homogeneous, if it is between 1 and 2 the region may be heterogeneous, and else, if it is higher than 2, the region is definitely not homogeneous. Here different site characteristics like the latitude, longitude, elevation, long term annual average of sunshine duration and mean annual precipitation were used as input to define homogeneous regions. Based on a k-clustering approach (Ward, 1963) 9 homogeneous regions were identified and are shown in **Figure 8**. The obtained homogeneous regions were tested for homogeneity for each data type combination and, as visible from **Figure A2** in appendix**,** all values are below 1, meaning that the regions selected are homogeneous and can be used for the index-based regionalisation. Note that the generalised statistics over all the durations from Section 3.1 are used as input for the homogeneity test. The R-package "nsRFA" is used to obtain the homogeneous regions (Viglione et al., 2020).

Once the homogeneous regions are determined, the different local statistics are normalised by a scaling factor, the index. In contrast to the previous regionalisation techniques discussed so far, the index-based regionalisation has an extra step – the normalisation of the general intensities with the index, which in this case is the mean generalised intensity. Next, the local L-moments are estimated on the basis of the normalised annual series and regional L-moments are derived for each region weighting the local L-moments according to their time series length. Finally, a GEV growth curve is fitted for each region and duration level via the regional L-moments. The R-package "lmomRFA" was employed for the application of the index method (Hosking and Wallis, 1997). In the final step, by back-scaling the normalised extreme rainfall for all

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

observed and unobserved points in the homogeneous region, estimates can be made about the extreme rainfall as a function of the duration and the return period. The geostatistical interpolation of the index makes it possible to transfer the extreme value statistical evaluations to unobserved points within the region.

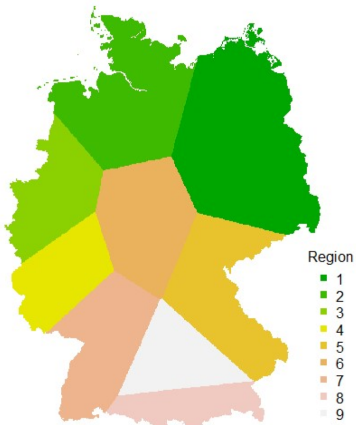

Figure 8 *Nine homogeneous regions implemented here for the index-based regionalisation. The regions shown here are a generalisation of the k-cluster results to account for spatial consistency.*

3.3 Performance Assessment and Comparison

3.3.1 Local Performance Assessment

For the local estimation of the GEV parameters that describe the extreme rainfall over all the selected duration levels, two different approaches were consulted: from Koutsoyiannis et al. (1998) (herein referred as KO) and from Fischer and Schumann (2018) (herein referred as FS). Before carrying on with the regionalisation it is important to investigate which of the methods is more adequate for the estimation of the GEV parameters over all the duration levels. Moreover, the two methods do not only distinguish in their approach of generalisation across duration, but they also include different variations on the calculation of the shape GEV parameter (γ). A review of the methods and shape parameters is given in **Table 5**, together with the respective optimised parameter set for each case. The obtained parameters for different data sets are shown in the appendix: **Figure A3** for KO and in **Figure A4** for FS.

Table 5 *A review of the methods and the different calculation of the shape parameter investigated for the local statistics.*

Method	Shape Parameter	Abbreviation	Optimised Parameter
KO	is constant per each station, as fitted by L-moments	KO.CON	μ, σ, γ, θ, η
	is fixed at all stations as $\gamma = 0.1$	KO.FIX	μ, σ, θ, η
FS	is calculated as proposed by Fischer and Schumann	FS.RLM	α_μ, β_μ, α_σ, β_σ, α, β
	is constant over all durations	FS.CON	α_μ, β_μ, α_σ, β_σ, γ
	a quadratic dependence on duration specific shape	FS.QUA	α_μ, β_μ, α_σ, β_σ, a
	is fixed at all stations as $\gamma = 0.1$	FS.FIX	α_μ, β_μ, α_σ, β_σ

The performance of the methods and the respective case of shape parameters as illustrated in **Table 5** is evaluated only at the location of the long series (LS) by comparing the normalised quantiles over all durations for return periods T1a, T10a, T20a, T50a and T100a with the GEV quantiles calculated separately at each duration level. Here the percentage RMSE (as per Equation (10)) was employed to assess the errors of the selected cases at each duration level and station with respect to the GEV duration specific quantiles:

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Open Access
Discussions

Percentage RMSE: $$RMSE_{d,st}[\%] = 100 \cdot \frac{\sqrt{\frac{1}{5}\sum_{i=1}^{5}\left(RD_{gen,st} - RD_{d,st}\right)^2}}{\overline{RD_{d,st}}}, \quad (10)$$

where i represents each selected return period T_a varying from 1 to 100 years, st varies from 1 to 133 available long series, $RD_{gen,st}$ corresponds to the derived rainfall depth from the generalisation method of duration d, $RD_{d,st}$ the derived rainfall depth from the GEV quantiles, and the $\overline{RD_{d,st}}$ is the mean rainfall depth from the GEV quantiles averaged over the return periods. Alternatively, the error for each return period and station can as well be calculated by Equation (10) by swapping the d with T_a, and where $\overline{RD_{T_a,st}}$ is the mean rainfall depth from the GEV quantiles at return period T_a averaged over the duration levels d (from 5min up to 7d).

Since the GEV quantiles fitted per each duration level cannot be considered the ground truth, a non-parametric bootstrap is performed when estimating the parameters of each method, in order to investigate the sampling uncertainty of derived DDF values. For this purpose, 100 randomisations of the observations were conducted and the uncertainty range of the derived rainfall depths is computed as following:

Normalised 95% Confidence Interval: $$nCI95\,[-] = 100\,\frac{CI95_{st,\,d,\,Ta}}{Mean_{st,\,d,\,Ta}} \quad (11)$$

where $CI95$ is the 95% confidence interval and Mean is the average of rainfall depth obtained from 100 realisations and expressed for each LS location st, duration level d and return period T_a. The smaller the uncertainty range, the more robust are the estimated parameters for the high return periods. Based on the two performance criteria, percentage RMSE and nCI95, all the methods in **Table 5** are compared in order to evaluate the best one for the estimation of rainfall DDF curves. The best method is selected as the one with the lowest RMSE and nCI95. The results of this comparison are given in Section 4.1.

3.3.2 Spatial Performance Assessment

In order to check which of the regionalisation approaches provides the best results, a leave-one out cross-validation was carried out at the locations of the long series (LS 133 stations). For each approach, the rainfall depth (RD) is determined from the return periods T1a, T10a, T20a, T50a and T100a and the selected duration levels. After regionalisation, the regionalised rainfall depths are compared with the local generalised GEV quantiles (here assumed to be the truth). The short series are omitted from the cross-validation, as no reliable extreme value statistics can be carried out for large return periods due to the very short observation length. The quality of the regionalisation approaches is evaluated using the following criteria:

Percentage Bias: $$PBIAS_{st,Ta}[\%] = 100 \cdot \frac{\frac{1}{D}\sum_{d=1}^{D}\left(RD_{regional,d} - RD_{local,d}\right)}{\sum_{d=1}^{D}\left(RD_{local,d}\right)}, \quad (12)$$

Percentage RMSE: $$RMSE_{st,Ta}[\%] = 100 \cdot \frac{\sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(RD_{regional,d} - RD_{local,d}\right)^2}}{\overline{RD_{local}}}, \quad (13)$$

Nash-Sutcliffe Criteria: $$NSC_{st,Ta}[-] = 1 - \frac{\sum_{d=1}^{D}\left(RD_{regional,d} - RD_{local,d}\right)^2}{\sum_{d=1}^{D}\left(RD_{local,d} - \overline{RD_{local}}\right)^2}, \quad (14)$$

where the d is the selected duration level, T_a the return period, st the respective LS, $RD_{regional}$ corresponds to the regionalised rainfall depth, RD_{local} the locally derived rainfall depth from the normalised GEV function and the RD_{local} is the mean local rainfall depth averaged over the 133 locations. The cross-validation only at the location of the LS makes it possible to investigate the value that the short (SS) and the disaggregated daily network (DS) are adding to each regionalisation method. For this purpose, the regionalisation methods are run first only with the LS as input, and the performance of such an application is considered the benchmark for improvement. Later on, the SS and DS are added

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

stepwise as input to the regionalisation, in order to assess the improvement, they introduce towards the benchmark. Additionally, one can calculate the expected performance when only the short or/and the disaggregated daily networks are available, and not the automatic one. An overview of these experiments and their aim is given at **Table 6**.

Table 6 *Overview of the experiments performed with different data sets for each regionalisation method.*

Input	Aim
Only LS	Benchmark for improvement
Only SS	The expected error from only short network
Only DS	The expected error from only disaggregated daily network
LS and SS	The added value from the short network
LS and DS	The added value from the daily network
SS and DS	The expected error from short and daily network
LS, SS and DS	The added value from the short and daily network

A directed comparison of the performance criteria between the different experiments and the benchmark is calculated here as per Equation (15).

$$Perf_{impr,Ta}\,[\%] = 100 \cdot \frac{(-Perf_{new,Ta} + Perf_{ref,Ta})}{Perf_{ref,Ta}}, \qquad (15)$$

where $Perf_{ref,Ta}$ is the performance criteria calculated for each return period Ta as per Equation (12)-(14) from the scenario with only LS as input, and $Perf_{new,Ta}$ is the performance of any other combination of input data as per Equation (12)-(14). A positive value for this criterion indicates an improvement in performance in comparison to the only LS scenario, while a negative value indicates a deterioration. Note that, the signs of the nominator are exchanged in the case of the improvement of the NSE. It is as well important to emphasise that the scenario *ref* corresponds to the best regionalisation method with only LS as input, namely ordinary kriging of LS based on results of Section 4.2.

Finally, based on different combinations of the available network as external drift in the kriging interpolation may help to shed light on which combination of the data is more useful for the regionalisation of the rainfall DDF values. Here the data to be used as external drift are first interpolated with ordinary kriging (also in cross-validation mode). A description of these different combinations for the KED interpolation is given is **Table 7.** The performance of the different combinations is evaluated only at the location of the LS, and the best integration is selected based on the highest improvement in comparison to regionalisation with only LS as input.

Table 7 *Overview of different integration of data types in the interpolation with KED. Pooling the data together with same importance is represented by (+) sign, whereas integration through an external drift (linear dependence) is represented by the (|) sign.*

Combination	Abbreviation
Interpolate LS with OK[SS] as external drift	KED[LS|SS]
Interpolate LS with OK[DS] as external drift	KED[LS|DS]
Interpolate LS with both OK[SS] and OK[DS] as external drift	KED[LS|SS+DS]
Interpolate LS and SS with OK[DS] as external drift	KED[LS+SS|DS]
Interpolate SS with OK[DS] as external drift	KED[SS|DS]

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

4. Results

4.1 Local Estimation of Extreme Statistics

Figure 9 illustrates the local percentage RMSE of each method in comparison to the duration specific quantiles (as per Equation (10)). The upper row of **Figure 9** shows the percentage RMSE calculated for each location and duration level over all the return periods and the lower row of **Figure 9** shows the percentage RMSE calculated for each location and return period over all the duration levels. The results from **Figure 9** – upper row indicate that the KO approaches (both fix and station constant shape parameter) have an almost constant RMSE over all durations under the value 10%. On the other hand, the FS approaches tend to have similar or little smaller RMSE for the longer duration (median RMSE under 8%), but are not able to represent well enough the very short durations. For the FS approaches, the RMSE median for duration levels up to 60 min, is higher than 10%, with the 5min RMSE being the highest (between 25-45%). The results from **Figure 9** – lower row illustrate that all the approaches manifest higher errors with higher return period. Both of the KO approaches (fix and station constant shape) show very similar behaviour. The KO.FIX performs slightly worse (1-4% higher RMSE) than the KO.CON, but this is expected as the duration GEV fitted per each duration independently favours the KO.CON (as the shape parameter is let free for the GEV parameter fitting). The FS approaches perform very similarly to one another, however here contrary to the KO.FIX approach, the performance of the FS.FIX seems better than the other approaches. Overall, the KO approaches have the priority at shorter durations and they can capture the volumes at specific durations better than the FS approaches. On the other side, the FS approaches can capture better extremes at longer durations. A unanimous selection is not yet possible from the obtained results so far, because the local GEV duration specific parameters may not represent the ground truth.

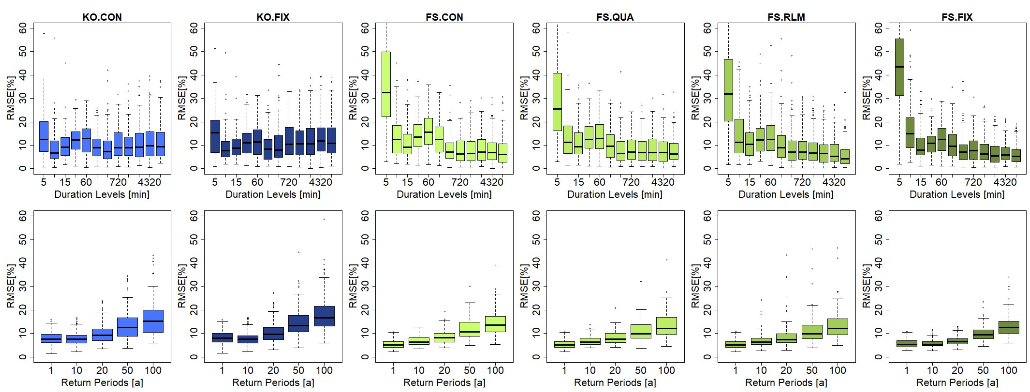

Figure 9 *RMSE (%) performance of the given generalisation methods over all the long stations (LS) in comparison to the duration specific GEV quantiles grouped: upper row - for different duration levels (calculated per station over return periods), and lower row - for different return periods (calculated per station over duration levels).*

To analyse which approach estimates more stable and representative parameters, a non-parametric bootstrap was performed (with 100 random realisations), and served as a basis for assessing the 95% confidence interval of the obtained DDF values. **Figure 10**-left shows the normalised 95% confidence intervals (nCI95) for the rainfall depth (as per Equation (11)) estimated for each of the selected approaches. A high value of the nCI95 indicates that the bootstrap yields very variable rainfall depths, and hence a higher uncertainty is associated with the method. Contrarily a low value of the nCI95 indicates that the rainfall depths have low variation across the random realisations, and thus the obtained DDF curves are considered more stable or robust. The results shown in **Figure 10** indicate that the KO.FIX exhibits the lowest variation (median nCI95~0.23), followed up by FS.FIX (~0.25), and by KO.CON, FS.CON, FS.QUA with slightly higher

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

variations (respectively ~0.3). Interesting is to see that the FS.RLM has a median nCI95 ~ 0.3, but can reach extreme values up to 2. **Figure 10**-right) shows the scatterplot of nCI95 obtained from the KO.FIX (x-axis) and FS.FIX (y-axis) for different duration levels and return periods (shown with different colours) at the LS locations. Except for very low return periods (T1a), FS.FIX exhibits on average higher values of nCI95 than KO.FIX. Based on these results, the KO.FIX was chosen as the best method and was used for the regionalisation of the DDF curves. The advantages of the KO.FIX are that: 1. It represents all duration levels similarly and fairly, 2. The parameter estimation is more robust than any of the other methods, 3. It uses a known and well-established method for the estimation of the DDF curves.

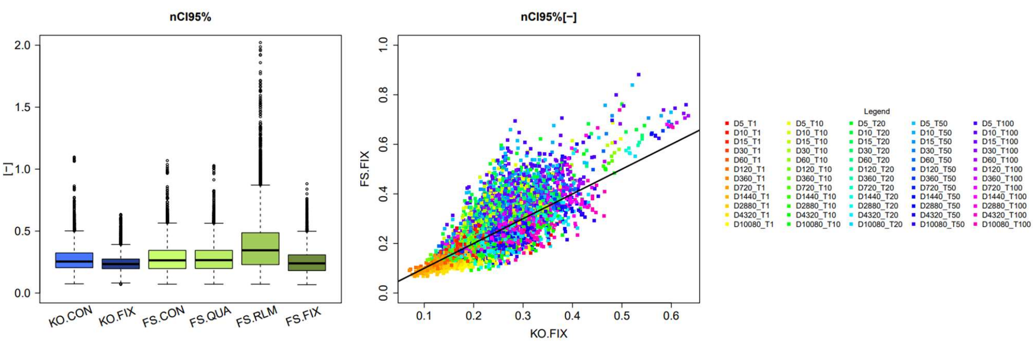

***Figure 10** left) comparison of confidence interval robustness for the methods and shape parameters selected for the generalisation of the DDF values over all the durations; right) a direct comparison of the confidence interval robustness for KO.FIX (x-axis) with FS.FIX (y-axis) for each duration and return period (shown in different colours).*

4.2 Regionalisation of Extreme Statistics

As discussed in the Section 4.1, the AMS at different duration levels were normalised according to Koutsoyiannis approach and the GEV parameters were fitted to the grouped generalised intensities. The shape parameter was kept fixed at 0.1. Ordinary Kriging (OK) and index-based (INDEX) regionalisation were run first only with the LR data as input – to decide about which of the two approaches will serve as a benchmark. A direct comparison based on Equation (15) is then performed for each of the selected performance criteria (where *new* is OK and *ref* is INDEX), to compute the improvement or deterioration of OK with only LS data compared to the INDEX. The median values for each return period, performance criteria and method, are given in **Table 8**. Here it becomes clear that the kriging approach exhibits lower RMSE for all return periods, worse BIAS for high return periods, and slightly better NSE than the index method. Based on these results, the kriging with LS as input (KRIGE[LS]) is used as a benchmark for calculating the improvement in performance by adding additional data types. Apart from the performance, the other advantage of kriging is that, it is more of a "pure" method, as it interpolates independently the 4 parameters, while the index approach is a "mixture" between the regional growth curve estimation, averaging θ and η parameters, and kriging to interpolate the index. For this reason, one may prefer the kriging regionalisation, as the errors are mainly from the kriging system, while the index method includes errors from the kriging system and from regional and averaged parameters.

***Table 8** Median performance improvement/deterioration (%) of ordinary kriging (OK) versus index-based (INDEX) interpolated calculated for different data as per Equation (15) (where new is OK and ref in INDEX), when only LS dataset is used as input. The performance is obtained by cross-validation over 133 LS stations. The colour green (+) indicates better performance by OK, red (-) indicates better performance by INDEX.*

	RMSE (%)					PBIAS (%)					NSE (%)				
	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a
LS	5.270	1.230	-0.268	0.015	1.510	2.500	-1.200	-1.440	-3.440	-2.469	0.250	0.010	0.002	0.002	0.006

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

4.2.1 Best Regionalisation for Different Data Combination

Kriging and index-based regionalisation was then performed for each data type experiment given in **Table 6**, and the cross-validation results for the 133 LS locations were compared to the benchmark (KRIGE[LS]) selected before as the best regionalisation with only LS as input. To enable an easy comparison between the two regionalisation methods, the difference between the improvements achieved between the kriging and the index-based regionalisation in comparison to the benchmark was calculated for each of the 133 LS locations. The median differences (in percent) for each data type experiment over the 133 locations for each performance criteria and return period are given in **Table 9**. A positive difference (dark green shade) means that the improvements reached by the kriging interpolation are higher than the index-based regionalisation. A negative difference (red shade) means vice versa. The data are combined by two operators: either (+) referring to pooling of the datasets together and the parameters and the index are interpolated with ordinary kriging, and (|) referring to a linear relationship between the datasets and the parameters and the index are interpolated through external drift kriging.

Table 9 *Median difference between kriging and index-based improvements calculated for different data as per Equation (15). The median is computed from 133 stations. The positive difference shown in green shades indicate that kriging introduces bigger improvements towards the benchmark than the index-based regionalisation. The negative differences shown in red shades indicate that the index-based regionalisation has the bigger improvements.*

	RMSE (%)					PBIAS (%)					NSE (%)				
	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a
SS	15.1	8.2	9.6	-0.1	0.4	6.5	10.4	4.8	1.5	-2.3	-0.1	0.6	0.0	0.0	-0.1
DS	19.4	4.8	6.1	10.1	12.2	-2.6	2.9	8.0	11.5	11.8	0.4	0.3	0.8	0.8	0.9
LS+SS	8.3	3.6	6.4	-2.3	-0.8	8.0	3.5	0.2	-6.7	-11.4	0.3	0.2	0.2	0.2	-0.1
LS|SS	5.5	11.6	12.3	9.8	10.8	13.0	8.6	3.6	6.1	6.0	0.2	0.3	0.5	0.5	0.5
LS+DS	101.2	90.4	75.3	77.3	76.9	157.5	162.9	154.7	134.1	130.5	10.1	10.0	10.1	10.1	10.0
LS|SS	20.7	16.6	16.1	15.5	12.8	27.6	12.6	10.5	3.9	1.4	0.7	0.4	0.4	0.4	0.3
SS+DS	111.0	97.5	82.5	79.0	82.6	176.0	194.6	188.7	157.2	150.8	10.3	9.8	9.8	9.8	9.4
SS|DS	10.6	6.8	8.8	4.0	5.1	9.9	-3.4	-2.8	-2.3	-5.9	0.2	0.4	0.3	0.3	0.2
LS+SS+DS	59.8	44.1	45.5	43.3	41.4	110.4	132.6	141.8	109.7	107.3	5.1	4.6	4.4	4.4	4.1
LS+SS|DS	13.1	12.2	13.2	10.6	11.9	10.4	2.0	-0.8	1.0	-2.8	0.2	0.5	0.5	0.5	0.5
LS|SS+DS	20.1	13.3	11.5	6.1	3.3	18.2	8.1	8.1	-0.2	-1.9	0.5	0.3	0.2	0.2	0.1

The results from the **Table 9** indicate that for the majority of the cases the kriging interpolation brings higher improvements to the benchmark than the index-based regionalisation. Exception are the regionalisation with only SS, LS+SS, SS|DS, LS+SS|DS and LS|SS+DS where the index-based regionalisation exhibits on median 2-12% higher PBIAS improvement for higher return periods than the kriging interpolation. However, for these cases, the RMSE and the NSE improvements are much higher for the kriging regionalisation. Therefore, it can be concluded that overall the kriging interpolation yields better results than the index-based regionalisation (lower RMSE and higher NSE), but may suffer depending on the combination of data types from slightly higher PBIAS. Also, it has to be mentioned, that when grouping the daily disaggregated time series directly (operator +) with the other data types (either LS and SS), the kriging performs up to 100% better than the index-based regionalisation. This suggests that the parameters from the disaggregation do not follow the same regions or growth curve as the high-resolution data (LS and SS), thus a kriging interpolation seems to be more reasonable when including these data as well.

The results of **Table 9** give a direct comparison between kriging and index-based regionalisation, nevertheless as they are relative to each case, do not give any information if ordinary kriging or external drift kriging is yielding better regionalisation results. For this purpose, the difference of improvements between KED and OK were calculated and shown as median over the 133 LS locations in **Table 10**. A positive difference (green shade) means that the improvements reached by KED are higher than the OK interpolation. A negative difference (red shade) means otherwise. The results

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

show that overall the KED exhibits higher RMSE and NSE improvements than the OK, but the KED tends to have lower PBIAS improvements than the OK. When only the high-resolution data sets are present (LS and SS), the KED behaves better than OK mainly for high return periods (50-100a), when LS and DS are present, KED clearly outperforms the OK. For all the remaining cases the OK outperforms the KED only for the PBIAS of high return periods.

Table 10 *Median difference between external drift kriging (KED) and ordinary kriging (OK) improvements calculated for different data as per Equation (15). The median is computed from 133 stations. The positive difference shown in green shades indicate that KED introduces bigger improvements towards the benchmark than the OK. The negative differences shown in red shades indicate that the OK regionalisation has the bigger improvements.*

	RMSE (%)					PBIAS (%)					NSE (%)				
	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a	T1a	T10a	T20a	T50a	T100a
LS and SS	-6.4	2.0	-1.9	7.8	8.8	-1.3	-4.9	-5.2	1.2	6.2	-0.5	-0.2	0.1	0.1	0.5
LS and DS	56.4	41.0	39.4	32.9	30.2	57.6	30.5	20.7	14.5	13.2	2.5	1.7	1.6	1.6	1.5
SS and DS	46.4	30.5	27.2	26.3	27.8	37.1	1.0	-8.1	-11.3	-14.9	1.9	1.4	1.3	1.3	1.4
LS+SS|DS	42.2	20.2	19.7	17.4	20.2	39.3	-0.5	-16.0	-18.6	-19.9	1.8	1.2	1.0	1.0	1.2
LS|SS+DS	40.0	20.6	16.3	16.4	16.4	37.0	-2.5	-21.5	-16.8	-17.7	1.6	1.0	0.9	0.9	1.0

4.2.2 Best Data Integration for Regionalisation

So far, the external drift kriging interpolation has shown superiority for regionalising DDF curves in comparison to the index-based and ordinary kriging regionalisation. Nevertheless, the question still remains, what is the best combination of the data sets for regionalising the DDF curves in Germany. Here it is interesting to see if all the three available data sets are useful for regionalisation, or if single or dual networks are enough. For this purpose, the performance improvement exhibited by different combinations of the data types in KED (as per **Table 7**) in comparison to the benchmark are visualised in **Figure 11**. Note that since there are 30 realisation of DS data, a boxplot is illustrating the performance spread over these 30 realisations. This affects regionalisation methods where DS data is present, otherwise a horizontal line indicates the performance of the regionalisation. For very low return periods (T1a), the integration of all data types of the form KED[LS+SS|DS] brings the best performance, with RMSE and BIAS up to 20% smaller and NSE 0.7% higher. For return period T10a, the KED[LS|SS], KD[LS|DS] and KED[LS+SS|DS] perform very similar: some random realisation from the disaggregated daily network (DS) introduce high improvement but as well low values, even though the median over the 30 realisation is at the same level as the KED[LS|SS] one. For high return periods (T100a), KED[LS|SS] introduces the highest improvement in all three performance criteria. Actually KED[LS|DS] is the second-best option, however the median over the 30 realisations is either lower or equal to the performance of the KED[LS|SS]. There are few realisations that introduce the highest improvements for RMSE and BIAS, nevertheless the computation time for the disaggregation scheme and the fitting of the Koutsoyiannis approach is also a disadvantage of using the DS data type. So finally, the kriging interpolation of the long network (LS) with the short network (SS) as an external drift, is chosen as an optimal method for the regionalisation of the GEV and Koutsoyiannis parameters. **Table 11** indicates the median performance criteria (RMSE, PBIAS, NSE) for different return periods reached by this method (KED[LS|SS]).

Table 11 *Median cross-validation performance over 133 stations for the final selected regionalisation method.*

	T1a	T10a	T20a	T50a	T100a
	KED[LS|SS]				
RMSE (%)	8.11	8.06	8.24	8.46	8.86
PBIAS (%)	1.00	1.10	0.80	1.00	0.80
NSE (-)	0.982	0.981	0.979	0.979	0.980

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

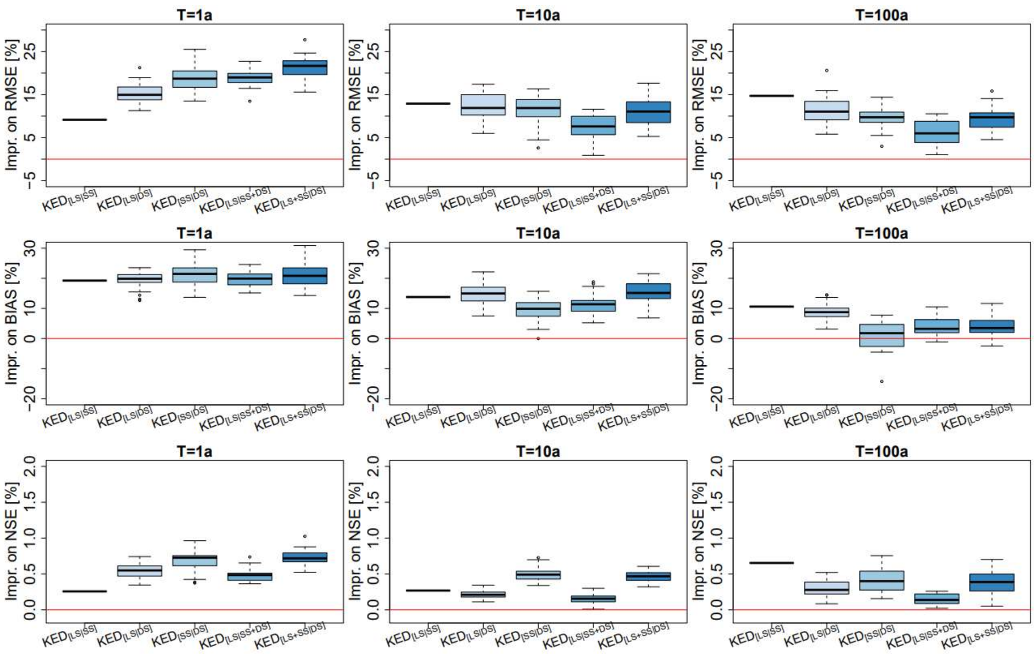

***Figure 11** Median performance improvements towards the benchmark from regionalising on different data combinations, as per **Table 7**, in kriging with external drift.*

4.3 Final Product

The obtained maps, on a 5 by 5 km raster, for the four regionalised parameters (location parameter – μ, scale parameter σ, Koutsoyiannis θ and η parameters) with the KED[LS|SS] approach, are illustrated in **Figure 12**. The spatial distribution of the location parameters follows partly the elevation information, with higher values in the south east, where the German Alps are located. The scale parameter values are independent of the elevation, with a high localised value near to Münster city. Recently, there has been a very extreme event in Münster which has affected the statistics of the station located in the vicinity. Currently it is not clear how to handle these singular extraordinary events in extreme value analysis in an optimal way. Both Koutsoyiannis parameters show similar spatial patterns with lower values in the Alp and other mountainous regions, as well as on the northern-west coast. These parameters exhibit higher variability in space than the GEV location or scale parameters. With these 4 interpolated maps, together with the shape parameter fixed at 0.1, DDF curves can be obtained for any location in Germany. Few examples of design rainfall maps for duration levels 5min, 1 hour and 1 day, and return period Ta=1,10,100 years, are given in **Figure 13**. For short durations (i.e. D=5 min) the spatial distribution of rainfall extremes is independent from the elevation and becomes more erratic with higher return periods. This is in accordance with the fact that the convective extreme events can happen anywhere and are very low correlated with the orography. With increasing duration level, the relationship between orography and extreme rainfall becomes stronger. As for instance in D=1h, the influence of the alpine regions is visible, which becomes even stronger for the duration of D=1d.

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Discussions
Open Access

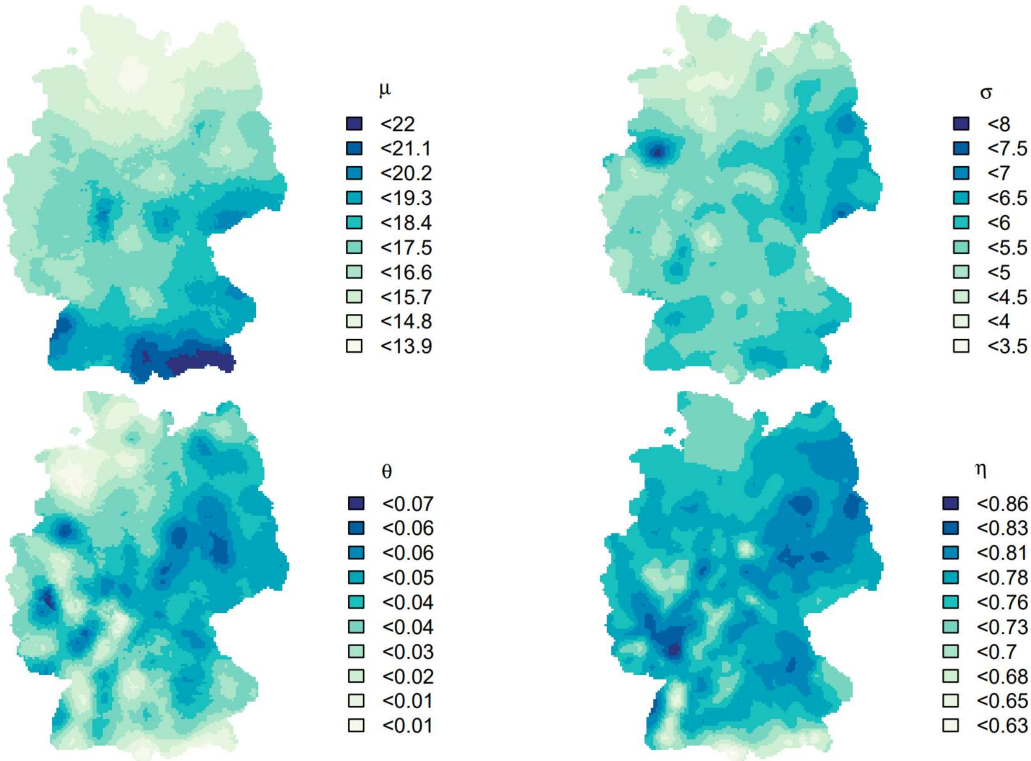

Figure 12 *Obtained interpolated maps from the KED[LS|SS] for each of the parameter: location parameter -* μ*, scale parameter -* $\boldsymbol{\sigma}$*, Koutsoyiannis* θ *and* η *parameters. The shape parameter* γ *is kept constant at 0.1.*

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

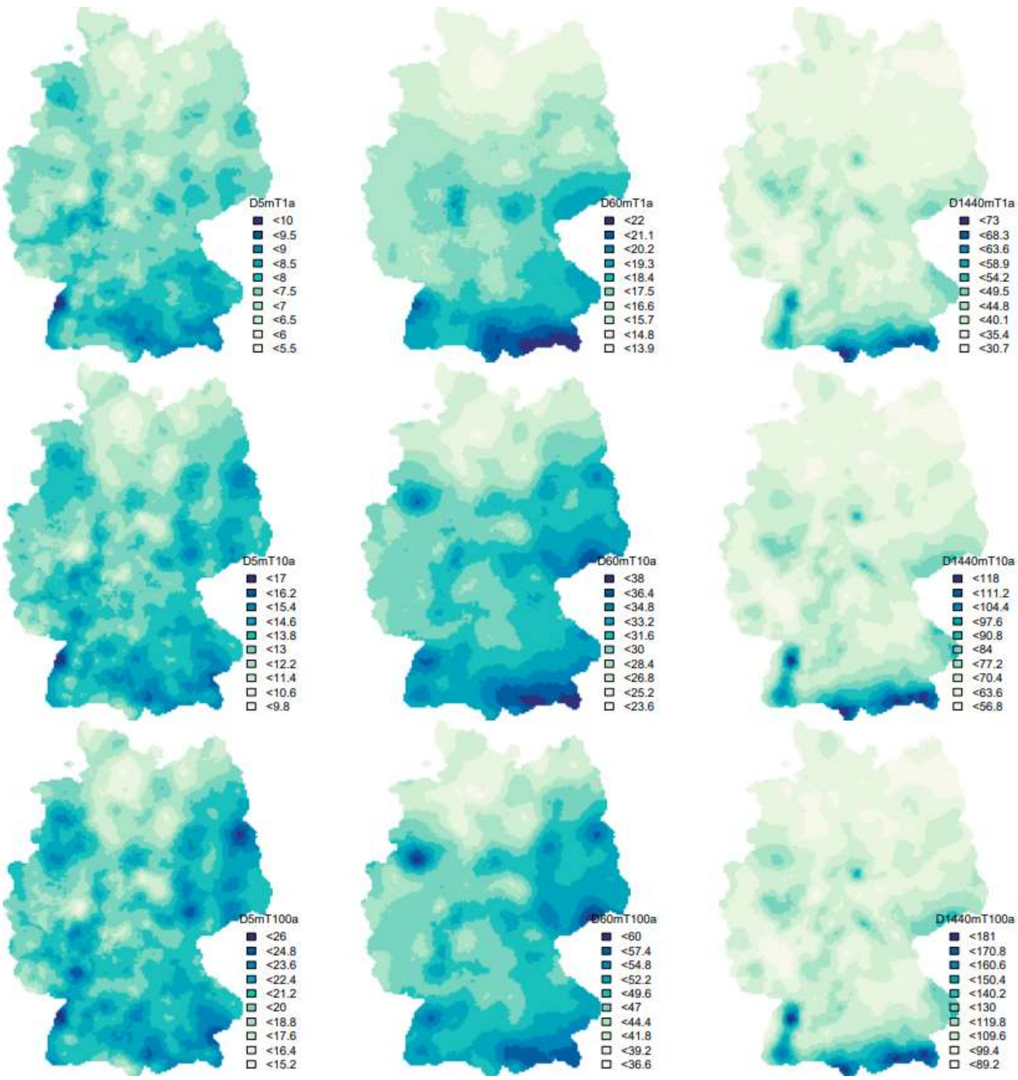

Figure 13 *Obtained design rainfall [mm] maps for whole Germany from the KED[LS|SS] regionalisation approach derived for different durations: first row – return period Ta=1-year, second row – return period Ta=10 years and third row – return period Ta=100 years.*

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

5. Conclusions

In this study the use of three ground measuring networks in Germany was investigated for the estimation of design rainfall maps. These networks included the long high-resolution network, with long observations at 5 min time steps from 60-70 years, the short high-resolution network with short observation also at 5 min time steps from 10 to 20 years, and the daily network with observations varying from 20 to 100 years. The purpose of the work was to review different methods for the estimation and regionalisation of the DDF curves and to investigate the value and the best integration of different data types for estimating DDF curves in ungauged locations. The results will provide the basis for a new update of the design storm maps for Germany, the KOSTRA-DWD2023. First, the long and short high-resolution networks were homogenised by performing a jump correction, with the jumps coinciding with sensor type changes. Second the daily network was disaggregated to sub hourly durations based on a cascade model parameterised according to Olsson, (1998) and Lisniak et al. (2013) from the RADOLAN data in Germany. Third, Annual Maximum Series (AMS) were derived for each station available in the three networks for duration levels ranging from 5 min to 7 days. This represents the main database for the present investigation. Two methods were investigated for local estimation of rainfall extreme statistics, adopted from Koutsoyiannis et al. (1998), and Fischer and Schumann (2018), and three different regionalisation approaches (ordinary kriging, external drift kriging and index-based regionalisation) were investigated for the spatial estimation of DDF curves in Germany. The conclusions derived, by considering the long high-resolution network as the truth, are summarised as:

- Both methods for local estimation of the rainfall extreme statistics behave quite similarly in capturing the local duration specific rainfall depths.

- Nevertheless, the estimation of parameters through the Koutsoyiannis approach is more robust in terms of data sampling uncertainties. Particularly the Koutsoyiannis approach combined with a Generalised Extreme Value (GEV) distribution with a fixed shape parameter value at 0.1 exhibited the highest robustness with tolerable decline in precision. Therefore, 4 parameters were used to describe the local statistics of extreme rainfall: the location and scale GEV parameters and the two Koutsoyiannis parameters θ and η. These 4 parameters represent the basis for the testing of different scenarios and regionalisation approaches.

- When only the long high-resolution network is present, both ordinary kriging and index-based regionalisation perform similarly, with ordinary kriging showing slightly better median performance. This result remains true as well for other data combination settings, with kriging methods exhibiting lower RMSE and NSE, but slightly higher PBIAS than the index-based regionalisation. The only case where the index-based regionalisation has slight superiority against kriging, is when only short high-resolution series are present.

- When more than two networks are available, kriging with external drift seems more adequate for the parameter interpolation than ordinary kriging, at least regarding the RMSE and NSE performance.

- A combination of long and short networks improves the performance of regionalisation considerably (up to 15% for Ta=100 years), but only when the data sets are combined with external drift kriging. Here the digital network is first interpolated with ordinary kriging, which later on, serves as an external drift for the kriging interpolation of the long network. This combination gave overall the best results at least for return periods higher than 10 years.

- A combination of the long high-resolution and daily networks improves the performance of regionalisation up to 10% being the second-best method for regionalisation. Here as well the best regionalisation was the external drift kriging, with the ordinary kriging interpolation of daily network serving as an external drift.
- A combination of the three networks improves the regionalisation considerably (up to 20%) only for low return periods (shorter or equal than 10 years).
- Overall, the best method for the regionalisation of the DDF curves in Germany, was the kriging interpolation of the long sub hourly stations, with the short sub hourly stations as an external drift. On average, this approach exhibited 8-9% RMSE (increasing with the return period) and up to 1% BIAS (decreasing with the return period) when compared to the locally estimated DDF curves.

The cross-validation implemented here can only describe the accuracy of the regionalisation methods when compared to the local estimation, but it does not say much about the precision of the predictions. Thus, it is important to perform an uncertainty analysis, which should include not only the local estimation of sample statistics (briefly discussed here) but as well the spatial uncertainty of the kriging interpolation. An investigation is currently going on for the integration of spatial uncertainty in the DDF design storms of Germany. Further improvements of the methodology, might include the validation of the methods on distinguished region. It has to be noted that the majority of the reference stations in Germany are located in the lowlands, thus the mountainous areas may be under-represented. It would be interesting to investigate if daily data or other site characteristics (like the elevation) are improving the performance of the chosen method in these regions. However, should one decide to perform region specific regionalisation, special care should be paid to the continuity of DDF values at the borders of the regions. Lastly, these conclusions are valid mainly for Germany, where dense networks are present. The advantage of each data set or approach may still change depending on the station density or study area location.

6. Data Availability

The daily and the short sub-daily network are made publicly available by the German Weather Service (DWD) and can be accessed at https://opendata.dwd.de/climate_environment/CDC/. The long sub-daily network has been digitalised and provided by the DWD. All R-codes can be provided by the corresponding authors upon request.

7. Authors Contribution

Supervision and funding for this research were acquired by UH and WW, the study conception, design and methodology were performed by all authors, while the software, data collection, derivation and interpretation of results were handled mainly by BS and WW (with support of the other authors). BS prepared the original draft, which is revised by all authors.

8. Competing Interest

The authors declare that they have no conflict of interest.

9. Funding

This research was funded by the German Ministry of Agriculture and Environment Mecklenburg-Vorpommern and the Federal State Funding Programme "Water, Soil and Waste”.

10. Acknowledgements

The results presented in this study are part of the research project “Investigating Different Methods for Revising and Updating the Heavy Rainfall Statistics in Germany (MUNSTAR)”, funded by the German Ministry of Agriculture and

Environment Mecklenburg-Vorpommern and the Federal State Funding Programme "Water, Soil and Waste" who are gratefully acknowledged. We are also thankful for the provision and right to use the data from the German National Weather Service (Deutscher Wetterdienst DWD), more specific Thomas Deutschländer and Thomas Junghänel.

11. References

Asikoglu, O. L. and Benzeden, E.: Simple generalization approach for intensity-duration-frequency relationships, Hydrol. Process., 28(3), doi:10.1002/hyp.9634, 2014.

Asquith, W. H.: Lmomco: L-moments, censoed L-moments, trimmed L-moments, L-comoments, and many distributions., 2021.

Bara, M., Kohnová, S., Gaál, L., Szolgay, J. and Hlavčová, K.: Estimation of IDF curves of extreme rainfall by simple Scaling in Slovakia, Contrib. to Geophys. Geod., 39(3), 2009.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., Gerlach, N. and MeteoSolutions GmbH: Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Offenbach am Main., 2004.

Bernard, M. M.: Formulas for rainfall intensities of long durations., Trans. Am. Soc. Civ. Eng., 96, 592–624, 1932.

Borga, M., Vezzani, C. and Fontana, G. D.: Regional Rainfall Depth-Duration-Frequency Equations for an Alpine Region, Nat. Hazards, 36, 221–235, 2005.

Boukhelifa, M., Meddi, M. and Gaume, E.: Integrated Bayesian Estimation of Intensity-Duration-Frequency Curves: Consolidation and Extensive Testing of a Method, Water Resour. Res., 54(10), doi:10.1029/2018WR023366, 2018.

Burn, D. H.: A framework for regional estimation of intensity-duration-frequency (IDF) curves, Hydrol. Process., 28(14), doi:10.1002/hyp.10231, 2014.

Cannon, A. J.: Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes, Stoch. Environ. Res. Risk Assess., 32(11), doi:10.1007/s00477-018-1573-6, 2018.

Ceresetti, D., Ursu, E., Carreau, J., Anquetin, S., Creutin, J. D., Gardes, L., Girard, S. and Molinié, G.: Evaluation of classical spatial-analysis schemes of extreme rainfall, Nat. Hazards Earth Syst. Sci., 12(11), 3229–3240, doi:10.5194/nhess-12-3229-2012, 2012.

Coles, S.: An Introduction to Statistical Modeling of Extreme., 2001.

Durrans, S. R. and Kirby, J. T.: Regionalization of extreme precipitation estimates for the Alabama rainfall atlas, J. Hydrol., 295(1–4), doi:10.1016/j.jhydrol.2004.02.021, 2004.

DVWK: Statistische Analyse von Hochwasserabflüssen, Merkblatt 251, Bonn, 62 S, 1999.

DWA: Arbeitsblatt DWA-A 531: Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer, DWA Arbeitsgruppe HW 1.1e, Hennef, Deutschland., 2012.

Fischer, S. and Schumann, A. H.: Berücksichtigung von Starkregen in der Niederschlagsstatistik, Hydrol. und Wasserbewirtschaftung, 62(4), 221–240, doi:10.5675/HyWa, 2018.

Forestieri, A., Lo Conti, F., Blenkinsop, S., Cannarozzo, M., Fowler, H. J. and Noto, L. V.: Regional frequency analysis of extreme rainfall in Sicily (Italy), Int. J. Climatol., 38(January), e698–e716, doi:10.1002/joc.5400, 2018.

Gupta, V. K. and Waymire, E.: Multiscaling properties of spatial rainfall and river flow distributions, J. Geophys. Res., 95(D3), 1999–2009, doi:10.1029/JD095iD03p01999, 1990.

Haktanir, T., Cobaner, M. and Kisi, O.: Frequency analyses of annual extreme rainfall series from 5 min to 24 h, Hydrol. Process., 24(24), doi:10.1002/hyp.7759, 2010.

Holešovský, J., Fusek, M., Blachut, V. and Michálek, J.: Comparison of precipitation extremes estimation using parametric and nonparametric methods, Hydrol. Sci. J., 61(13), doi:10.1080/02626667.2015.1111517, 2016.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Analysis, Cambridge University Press., 1997.

Hyndman, R. J. and Fan, Y.: Sample Quantiles in Statistical Packages, Am. Stat., 50(4), 361–365, doi:10.1080/00031305.1996.10473566, 1996.

Johnson, F. and Sharma, A.: Design Rainfall, in Handbook of Applied Hydrology, edited by V. P. Singh, pp. 125–3 to 125–13, McGraw-Hill, New York., 2017.

Kebaili Bargaoui, Z. and Chebbi, A.: Comparison of two kriging interpolation methods applied to spatiotemporal rainfall, J. Hydrol., 365(1–2), doi:10.1016/j.jhydrol.2008.11.025, 2009.

Koenker, R.: Quantile Regression, , doi:10.1017/CBO9780511754098, 2005.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation, Hydrol. Sci. J., 49(4), 575–590, doi:10.1623/hysj.49.4.575.54430, 2004a.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, Hydrol. Sci. J., 49(4), 591–610, doi:10.1623/hysj.49.4.591.54424, 2004b.

Koutsoyiannis, D., Kozonis, D. and Manetas, A.: A mathematical framework for studying rainfall intensity-duration-frequency relationships, J. Hydrol., 206(1–2), 118–135, doi:10.1016/S0022-1694(98)00097-3, 1998.

Lima, C. H. R., Kwon, H. H. and Kim, J. Y.: A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate, J. Hydrol., 540, doi:10.1016/j.jhydrol.2016.06.062, 2016.

Lisniak, D., Franke, J. and Bernhofer, C.: Circulation pattern based parameterization of a multiplicative random cascade for disaggregation of observed and projected daily rainfall time series, Hydrol. Earth Syst. Sci., 17(7), 2487–2500, doi:10.5194/hess-17-2487-2013, 2013.

Madsen, H., Arnbjerg-Nielsen, K. and Mikkelsen, P. S.: Update of regional intensity-duration-frequency curves in Denmark: Tendency towards increased storm intensities, Atmos. Res., 92(3), doi:10.1016/j.atmosres.2009.01.013, 2009.

Madsen, H., Gregersen, I. B., Rosbjerg, D. and Arnbjerg-Nielsen, K.: Regional frequency analysis of short duration rainfall extremes using gridded daily rainfall data as co-variate, Water Sci. Technol., 75(8), doi:10.2166/wst.2017.089, 2017.

Malitz, G. and Ertel, H.: KOSTRA-DWD-2010 - Starkniederschlagshöhen für Deutschland (Bezugszeitraum 1951 bis 2010). Abschlussbericht, Dtsch. Wetterd., 1–40, 2015.

Muller, A., Bacro, J. N. and Lang, M.: Bayesian comparison of different rainfall depth-duration-frequency relationships, Stoch. Environ. Res. Risk Assess., 22(1), doi:10.1007/s00477-006-0095-9, 2008.

Müller, H. and Haberlandt, U.: Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology, J. Hydrol., 556, 847–864, doi:10.1016/J.JHYDROL.2016.01.031, 2018.

Olsson, J.: Evaluation of a scaling cascade model for temporal rain- fall disaggregation, Hydrol. Earth Syst. Sci., 2(1), 19–30, doi:10.5194/hess-2-19-1998, 1998.

Paixao, ; E, Auld, H., Mirza, M. M. Q., Klaassen, J. and Shephard, M. W.: Regionalization of heavy rainfall to improve climatic design values for infrastructure: case study in Southern Ontario, Canada, Hydrol. Sci. Journal-Journal des Sci. Hydrol., 56(7), 1067–1089, doi:10.1080/02626667.2011.608069, 2011.

Papalexiou, S. M.: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency, Adv. Water Resour., 115, doi:10.1016/j.advwatres.2018.02.013, 2018.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions : A global survey on extreme daily rainfall, Water Resour. Res., 49(1), 187–201, doi:10.1029/2012WR012557, 2013.

Pebesma, E. J.: Multivariable geostatistics in S: The gstat package, Comput. Geosci., 30(7), 683–691, doi:10.1016/j.cageo.2004.03.012, 2004.

Perica, S., Pavlovic, S., St. Laurent, M., Trypaluk, C., Unruh, D., Martin, D. and Wilhite, O.: NOAA Atlas 14 Volume 10: Precipitation-Frequency Atlas of the United States, NOAA, Natl. Weather Serv. Silver Spring, MD, 1, 2019.

Requena, A. I., Burn, D. H. and Coulibaly, P.: Pooled frequency analysis for intensity–duration–frequency curve estimation, Hydrol. Process., 33(15), doi:10.1002/hyp.13456, 2019.

Roksvåg, T., Lutz, J., Grinde, L., Dyrrdal, A. V. and Thorarinsdottir, T. L.: Consistent intensity-duration-frequency curves by post-processing of estimated Bayesian posterior quantiles, J. Hydrol., 603, doi:10.1016/j.jhydrol.2021.127000, 2021.

De Salas, L. and Fernández, J. A.: "In-site" regionalization to estimate an intensity-duration-frequency law: a solution to scarce spatial data in Spain, Hydrol. Process. Hydrol. Process, 21, 3507–3513, doi:10.1002/hyp.6551, 2007.

Smithers, J. C. and Schulze, R. E.: A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments, J. Hydrol., 241(1–2), doi:10.1016/S0022-1694(00)00374-7, 2001.

Soulis, E. D., Sarhadi, A., Tinel, M. and Suthar, M.: Extreme precipitation time trends in Ontario, 1960–2010, Hydrol. Process., 30(22), doi:10.1002/hyp.10969, 2016.

Uboldi, F., Sulis, A. N., Lussana, C., Cislaghi, M. and Russo, M.: A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution, Hydrol. Earth Syst. Sci., 18(3), 981–995, doi:10.5194/hess-18-981-2014, 2014.

Ulrich, J., Fauer, F. S. and Rust, H. W.: Modeling seasonal variations of extreme rainfall on different timescales in Germany, Hydrol. Earth Syst. Sci., 25(12), doi:10.5194/hess-25-6133-2021, 2021.

Veneziano, D., Lepore, C., Langousis, A. and Furcolo, P.: Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall, Water Resour. Res., 43(10), doi:10.1029/2007WR006040, 2007.

Viglione, A., Hosking, J. R. M., Laio, F., Miller, A., Gaume, E., Payrastre, O., Salinas, J. L., N'guyen, C. C. and Halbert, K.: Non-Supervised Regional Flood Frequency Analysis., 2020.

Van de Vyver, H.: Bayesian estimation of rainfall intensity-duration-frequency relationships, J. Hydrol., 529, doi:10.1016/j.jhydrol.2015.08.036, 2015.

Van de Vyver, H.: A multiscaling-based intensity–duration–frequency model for extreme precipitation, Hydrol. Process., 32(11), doi:10.1002/hyp.11516, 2018.

Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, J. Am. Stat. Assoc., 58(301), 236–244, doi:10.1080/01621459.1963.10500845, 1963.

Watkins, D. W., Link, G. A. and Johnson, D.: Mapping regional precipitation intensity duration frequency estimates, J. Am. Water Resour. Assoc., 41(1), doi:10.1111/j.1752-1688.2005.tb03725.x, 2005.

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

12. Appendix

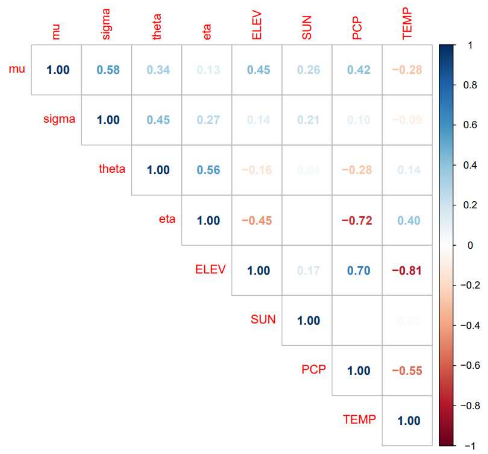

Figure A1 *Cross-correlation between the selected local parameters (Koutsoyiannis and GEV parameters) for regionalisation and useful site characteristics that might act as an external drift information.* Mu *is the location parameter,* sigma *the scale parameter,* theta *and* eta *the Koutsoyiannis parameters,* ELEV *is short for elevation information,* SUN *is short for long term average of annual sunshine duration,* PCP *is short for long term average of annual rainfall amount, and* TEMP *is short for the long-term average of annual mean temperature.*

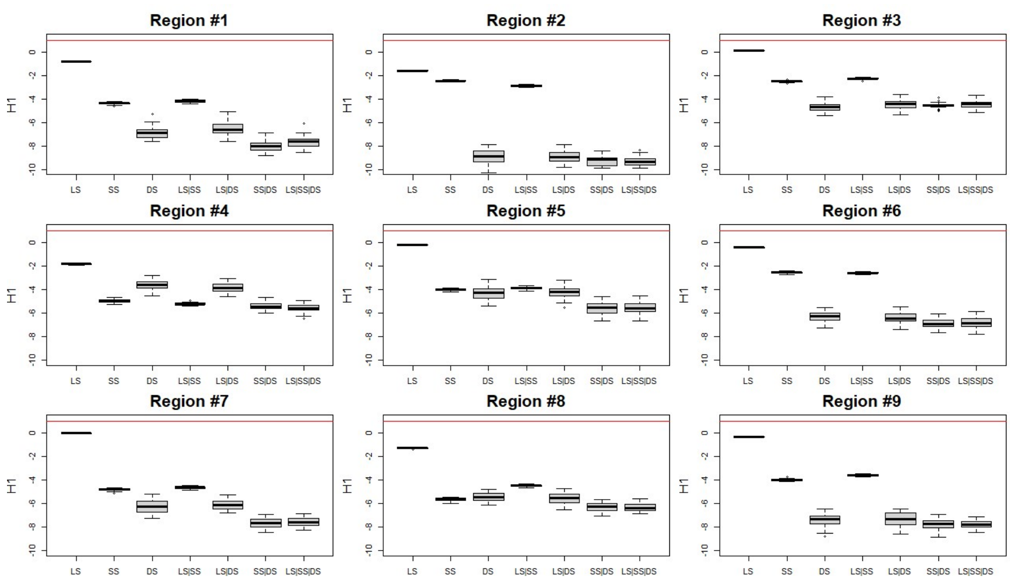

Figure A2 *The homogeneity index (H1) computed for each of the 9th selected regions for each of the dataset combinations.*

https://doi.org/10.5194/hess-2022-118
Preprint. Discussion started: 11 April 2022

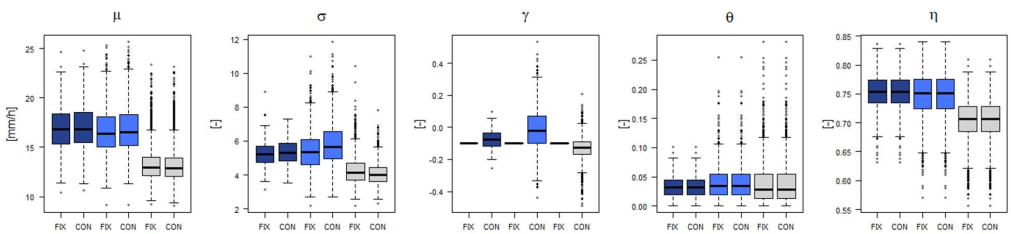

Figure A3 *Koutsoyiannis parameters obtained for each data set (LS in dark blue, SS in light blue and DS in grey) when fixing the shape parameter to 0.1 (FIX) or letting it free (FREE).*

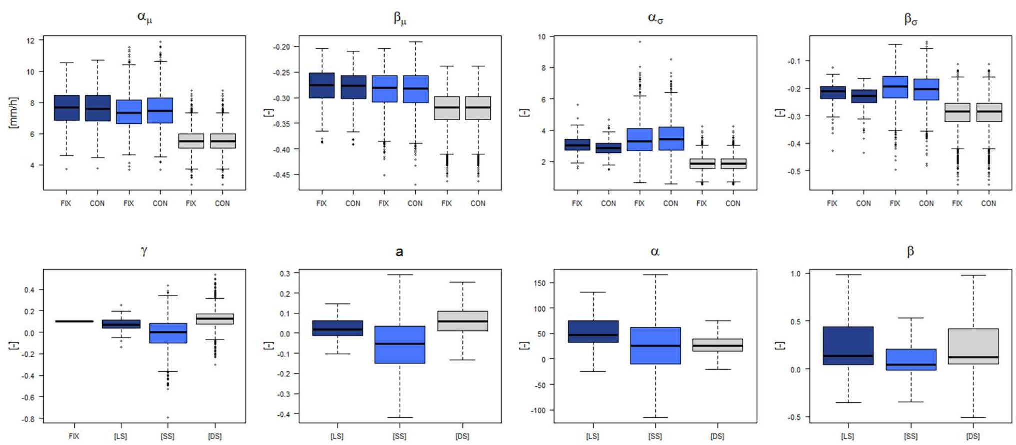

Figure A4 *Fischer/Schumann parameters obtained for each data set (LS in dark blue, SS in light blue and DS in grey) when fixing the shape parameter to 0.1 (FIX) or letting it free (FREE).*