ORIGINAL PAPER

Non-stationary large-scale statistics of precipitation extremes in central Europe

Felix S. Fauer[1] · Henning W. Rust[1]

Abstract

Extreme precipitation shows non-stationarity, meaning that its distribution can change with time or other large-scale variables. For a classical frequency-intensity analysis this effect is often neglected. Here, we propose a model including the influence of North Atlantic Oscillation, time, surface temperature and a blocking index. The model features flexibility to use annual maxima as well as seasonal maxima to be fitted in a generalized extreme value setting. To further increase the efficiency of data usage, maxima from different accumulation durations are aggregated so that information for extremes on different time scales can be provided. Our model is trained to individual station data with temporal resolutions ranging from one minute to one day across Germany. Models are chosen with a stepwise BIC model selection and verified with a cross-validated quantile skill index. The verification shows that the new model performs better than a reference model without large-scale information. Also, the new model enables insights into the effect of large-scale variables on extreme precipitation. Results suggest that the probability of extreme precipitation increases with time since 1950 in all seasons. High probabilities of extremes are positively correlated with blocking situations in summer and with temperature in winter. However, they are negatively correlated with blocking situations in winter and temperature in summer.

Keywords Extreme precipitation · Generalized extreme value distribution · Maximum likelihood · Non-stationary climate · Large scale · Intensity-duration-frequency

1 Introduction

Hydrologic extremes are changing. This is supported by the sixth IPCC assessment report (AR6) (Seneviratne et al 2021) which finds that the majority of measurement stations in Europe shows a significant increase in extreme precipitation over durations of 1 day and 5 days between 1950 and 2018. Trends might be variable in sign and value across regions and seasons (Croitoru et al 2013; Fischer et al 2015; Chiew et al 2009; Arnbjerg-Nielsen 2012). For example, a decreasing 5-day-maximum-precipitation (RX5day) by the year 2100 is reported (Iturbide et al 2021; Gutiérrez et al 2021) in a 2 °C warming scenario in summer and increasing RX5day in the

other seasons. These facts show the heterogeneity of developments in extreme precipitation. Furthermore, they emphasize that precipitation extremes are changing in a non-stationary fashion and the underlying distribution is subject to change with time and other large-scale variables (Schlef et al 2023; Rootzén and Katz 2013).

Extreme value statistics describes the relation between intensity and occurrence probability of extremes. One strategy is to describe block-maxima with the generalized extreme value distribution (GEV). Here, the block size is (1) one year for annual models or (2) one month for seasonal models. Even in a stationary setting, extremes are difficult to model since they are rare by definition. Increasing the complexity of the model by describing the dependence of extremes on other variables (covariates) is a challenge which can be faced by more efficient use of data. More information can be processed by using spatial models with GEV parameters depending on the location. Such a model has been created by Ulrich et al (2020) who were able to decrease the uncertainty, but not to generally

✉ Felix S. Fauer
felix.fauer@met.fu-berlin.de

Henning W. Rust
henning.rust@fu-berlin.de

[1] Institute of Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany

increase model performance score-wise. Here, nearby stations are modeled with a smooth transition and information gain, but over large distances it is difficult to capture the underlying patterns. Several more studies acknowledge the spatial dependence of extreme precipitation (Davison et al 2012; Schliep et al 2009; Blanchet et al 2016).

Another way of increasing data use efficiency is the inclusion of different duration accumulation steps. This is not only beneficial for the efficiency, but also because effects from covariates occur over different time scales. Therefore, precipitation data from different measurement resolutions (from minutes to days) can be accumulated to various durations (duration steps). With this data, duration-dependent GEV (d-GEV) distributions (Nguyen et al 1998) have been used so that more information of each year is processed, as maxima of different duration steps are fed into the model. The results of such analyses are often shown in Intensity-Duration-Frequency (IDF) curves (Chow 1953). The relation between duration and intensity can be described by different parametrizations, including multiscaling (Gupta and Waymire 1990), duration-offset (Koutsoyiannis et al 1998) and intensity-offset (Fauer et al 2021). Some of these duration-dependent approaches have been combined with large-scale influence on precipitation by Ouarda et al (2019). In their study, the d-GEV parameters depended on large-scale covariates, e.g., time and several teleconnection patterns. This resulted in statistics for three locations in the USA. However, there is no study known to us which covers Central Europe with such a model. Our approach uses a similar method as Ouarda et al (2019) and the main new aspects are: (1) We cover Germany with 199 precipitatin gauges. (2) We use large-scale information (NAO index, a blocking index, spatially and temporally averaged temperature and humidity) which might fit better to the atmospheric circumstances in Central Europe. (3) Our model features advanced flexibility regarding different durations and probes more potentially influencing covariates. (4) We use an advanced verification method to assess whether the use of large-scale information improves the model, aside from new insights into large-scale effects.

Cheng and AghaKouchak (2014) modeled extreme precipitation depending on large-scale covariates in a Bayesian setting which has the advantage that uncertainty of parameters can be estimated in a much more elaborated way. A disadvantage of Bayesian models is the need to choose a prior manually. The results might be sensitive to this choice of hyper-parameter. Another advantage of our study is the use of a consistent model that includes duration-dependence in one modeling step.

Our analysis aims for the identification of meaningful large-scale variables. Therefore, we investigate the influence of blockings, North Atlantic Oscillation (NAO), temperature, humidity and time.

A blocking situation is characterized by an interruption of the westerly flow due to persistent anticyclones (Otero et al 2022). The presence of a blocking situation can influence the appearance of heavy precipitation. The change of odds for heavy precipitation in presence of blocking depends heavily on season and region (Lenggenhager and Martius 2019). We will compare our findings with the literature with respect to our definition of blocking and choice of region in Sect. 4.

The NAO is the most important teleconnection pattern in Europe (Barnston and Livezey 1987). The change of extreme precipitation with respect to NAO has been investigated by Casanueva et al (2014) and the association between both variables is opposite in winter (positive) and summer (negative) in Germany. There, precipitation trend over time in Germany is mostly non-significant both in summer and winter.

Temperature and extreme precipitation show a correlation which has received considerable attention in the literature (Aleshina et al 2021; Westra et al 2014). The Clausius-Clapeyron scaling describes the dependence between potential water content and air temperature. It provides an explanation for increasing rain amounts in warmer air. However, the connection between extreme precipitation and temperature is more complex. After correcting for the Clausius-Clapeyron scaling, the sign of the correlation coefficient changes depending on the temperature regime and is negative (positive) for warmer (colder) temperatures in Australia (Hardwick Jones et al 2010). The same applies to Europe, where temperatures above 15 °C lead to less extreme precipitation (Drobinski et al 2016). In North-America, the correlation between both quantities is consistently positive (Mishra et al 2012), which is also known as Clausius-Clapeyron (C-C) scaling.

The temporal trend, described as change in time, has to be treated carefully as time in most cases is not physically influencing precipitation extreme, but it is a proxy for other effects that influence meteorological extremes. These effects are highly non-linear and therefore difficult to describe. Time is thus an interesting covariate as it represents multiple effects. The goal is, however, to integrate more physically relevant covariates and reduce the influence of time as covariate.

2 Data and methods

2.1 Data

We use precipitation data from three different sources. (1) The German meteorological service (DWD) provides data from stations across Germany and we use 86 stations that cover both daily and minutely resolution (see Fig. 1b). This

data is publicly available (DWD 2022). (2) Additionally, data from three DWD stations with long time ranges (longest with 57 years) and 5-minute resolutions were provided to us which are not publicly available (see Acknowledgements). (3) Furthermore, the Wupperverband provided data from 57 stations with daily data, 6 stations with hourly data and 18 stations with minutely data (see Fig. 1c). Stations vary in length of time series and availability of high-temporal-resolution measurements (Fig. 1a-c). Different stations that have a distance of less than 250 m were grouped together, since precipitation amount should not change considerably. Possible duplicates, i.e., more than one value for a specific station and duration and year might occur because different stations were merged or because both minutely and daily measuring devices will provide an accumulated rainfall value for durations $d \geq 24\,h$. In this case, values from the lower measuring frequency are omitted.

The NAO index is obtained from the National Oceanic and Atmospheric Administration (NOAA) and the Climate Prediction Center (CPC), where it is openly available (NOAA 2022). We use the dataset with monthly values which is based on a Rotated Principal Component Analysis, starting in 1950.

The mean surface air temperature (tas) and relative humidity over Germany are obtained from the ERA5 dataset with a daily resolution of 0.25°. The data are spatially averaged between 4°W and 15°W longitude and between 45°N and 55°N latitude. This way, one value per time step indicates the mean temperature on a large scale. Data is available from 1950 to 2021 (Bell et al 2020).

The blocking information is inferred from a binary blocking-index (BBI), using gridded daily ERA5 data (by 2.5°). It is based on the two-dimensional blocking index from Scherrer et al (2006); Schuster et al (2019) with minor modifications. The BBI of the grid fields is averaged over Scandinavia, because atmospheric blocking situations over this region are found to have an influence on convection in Central Europe (Mohr et al 2019). The blocking value that is used here ranges between 0 and 1 and indicates the spatial fraction of grid fields that were identified as blocked.

All daily values of the large-scale variables, i.e., NAO, temperature, humidity and blocking, are averaged over non-overlapping blocks of one month or one year, depending on the model (season or annual). Since all datasets of large-scale variables start in 1950, precipitation data of earlier years are omitted because our model cannot handle missing values in any of the predictor terms.

The data for temperature, humidity and blocking index has been accessed using the ClimXtreme Central Evaluation System framework (Kadow et al 2021).
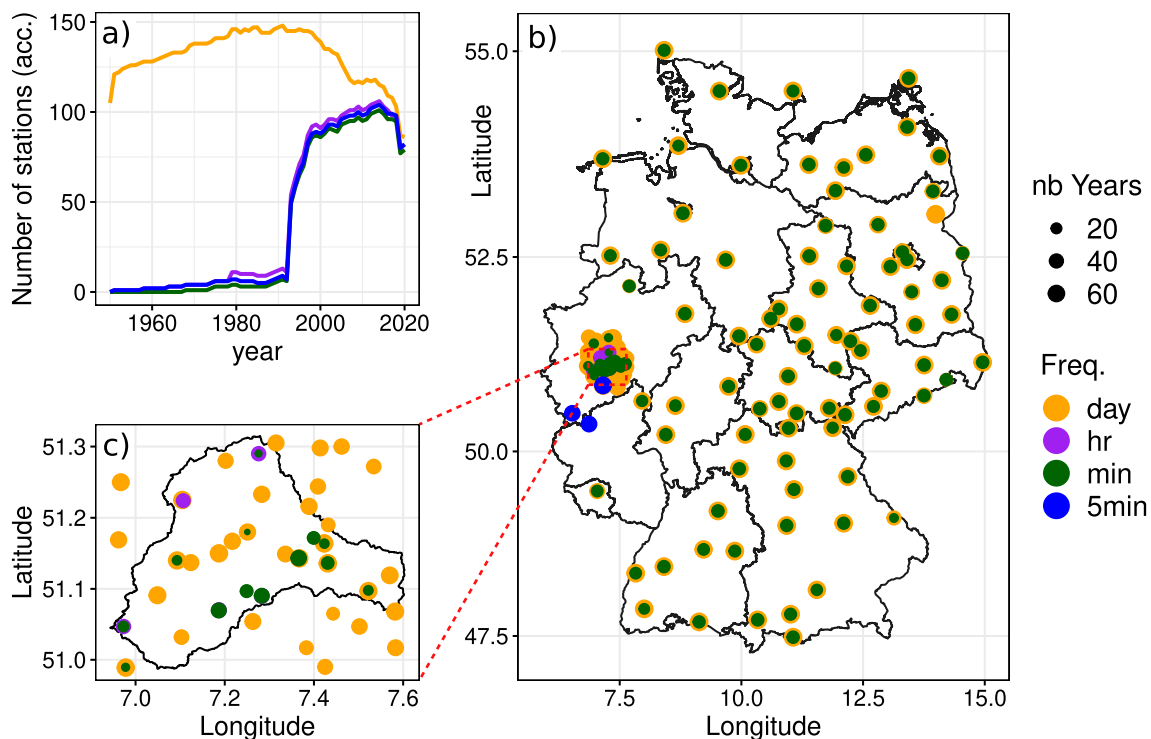

Fig. 1 **a**: Number of stations (accumulated) that provide data for each measurement frequency (color). **b** and **c**: Stations (dots) with measurement frequency (color) and length of time series (radius). **c**: Zoom into Wupper Catchment area

2.2 Flexible model for stationary GEV distribution

We model block-maxima of extreme precipitation with the GEV distribution. This distribution links probabilities or return periods to intensities or return levels. In this study, an extended version of the d-GEV distribution is used as proposed by Fauer et al (2021) whose study will be shortly summarized in this section. The used model shows a higher flexibility for very short ($d < 8$ h) and very long ($d > 24$ h) durations. This flexibility is introduced by a combination of existing features, namely curvature for short durations and multiscaling for medium durations, and an extension with an additional parameter τ which allows for return levels to deviate from the log-linear relation with duration (Fauer et al 2021; Ulrich et al 2021a). This flexible model is described by

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

$$\sigma(d) = \sigma_0(d + \theta)^{-(\eta + \eta_2)} + \tau, \tag{2}$$

$$\mu(d) = \tilde{\mu}(\sigma_0(d + \theta)^{-\eta} + \tau), \tag{3}$$

with the location parameter function $\mu(d)$, the scale parameter function $\sigma(d)$, the rescaled location parameter $\tilde{\mu}$, the scale offset $\sigma_0 > 0$, the shape parameter $\xi \neq 0$, the duration offset $\theta > 0$, the two duration exponents η and η_2, the intensity offset $\tau > 0$ and duration $d > 0$. The intensity z is restricted to $1 + \xi(z - \mu(d))/\sigma_0 > 0$. If $\xi = 0$, then $G(z) = \exp\{-[\exp((z - \mu(d))/\sigma(d))]\}$ applies.

The role of the different parameters has been explained in detail by Fauer et al (2021). The following paragraph provides a brief summary. Location μ, scale σ and shape ξ are characteristic distribution parameters, similar to many other distributions that describe the first three moments of the distribution. Adding duration-dependence to location and scale (Eqs. 2,3) requires additional parameters which have distinct effects on IDF or intensity-duration-variable (IDV) curves (Fig. 5, Sec. 3.3). Duration offset θ describes the curvature for short durations or how strong the curves deviate from a linear log-log relationship between duration and intensity. Therefore, this parameter is only necessary for stations with sub-hourly data. The intensity offset τ describes analogously the flattening of the relationship for long durations. This parameter is mainly important for annual models and only in combination with the duration offset (Fauer et al 2021). The Duration exponent η describes the slope of the relationship and the second duration exponent η_2 describes how the slope changes for different frequencies (multiscaling).

We estimate distribution parameters from the data with maximum likelihood estimation (MLE), meaning that the distribution parameters are chosen in a way such that the joined probability of all data points is maximized (Coles 2001).

The uncertainty of estimated intensities in the stationary model is obtained by parametric bootstrapping of the corresponding available years at each station. Years are sampled with replacement. When a year is chosen, data from all durations in this year are used. With this sample, the model is trained and return levels are estimated. This process is repeated 1000 times. Then, the 0.025- and the 0.975-quantile of the bootstrapped return levels determine the 95%-confidence interval. The uncertainty of estimated intensities in the large-scale model is obtained in the same way (see Sec. 3.3, last paragraph).

2.3 Motivation of non-stationary models

In this section, a sliding window approach motivates the need for a non-stationary model to describe the IDF relation. The methodology that is explained in this section will not be used for the final model of this study and will be presented in Sec. 3 (Results). Here, data points are grouped according to the value of a large-scale variable and d-GEV parameters are estimated for each group. This way, the change of d-GEV parameters can be shown with respect to a large-scale variable.

The dependence of d-GEV parameters $\tilde{\mu}$, σ_0, ξ, θ and η on the large-scale variable temperature is shown in Fig. 2 for the example station (Nürburg-Barweiler) in winter. Subsets of the data were created by choosing overlapping ranges of $4°$ around all possible centered temperature values in the data (first subset: $-6\,°C$ to $-2\,°C$, second subset: $-5.5\,°C$ to $-1.5\,°C$,...). Then, parameters are estimated for each of these subsets. The model parameters depending on the chosen subset with the centered temperature value on the abscissa are plotted as dots with vertical uncertainty bars. The four lines in different color represent a least-squares polynomial fit of degree 1 to 4. Solid lines indicate a significant coefficient of the covariate with the highest polynomial order according to a two-sided t test on a 0.05 level of significance. Dashed lines indicate polynomials with non-significant coefficients associated with the highest order. For example, the rescaled location parameter $\tilde{\mu}$ shows a significant dependence on temperature with polynomials of order 1 or 2 (blue and green solid lines). The histogram (Fig. 2f) shows that temperature values from all stations are not uniformly distributed and explains the higher uncertainty for very low temperatures which can also be seen in the sample sizes, given as small numbers below the bars in Fig. 2a).
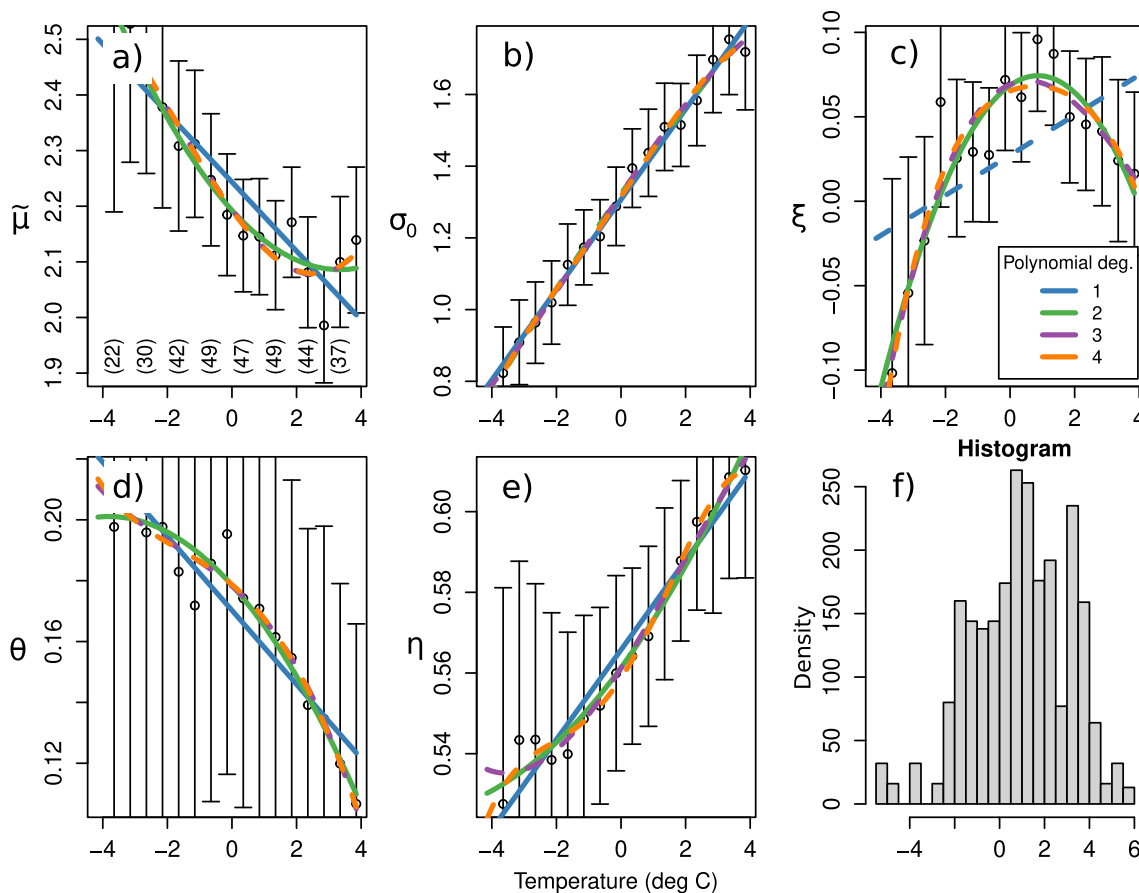
Fig. 2 An example of dependence of flexible d-GEV parameters $\tilde{\mu}$ (**a**), σ_0 (**b**), ξ (**c**), θ (**d**) and η (**e**) on large-scale variable temperature. Colored curves represent polynomial models to describe the parameter variability. Significance (t-test) is indicated with solid lines.

Sample size for each second bar is given in a). Data is from the example station Nürburg-Bahrweiler, for winter. **f** Histogram of temperature data for this station

2.4 Implications for non-stationary d-GEV model

The results of the previous section helps setting the boundaries for the model selection of the final model, i.e., restraining the d-GEV parameters depending on season, and which large-scale variables will potentially be used for estimating the d-GEV parameters. These implications are part of a pre-selection process and will be explained in this section. Afterwards, the systematic model selection will be explained in Sec. 2.5. A pre-selection is necessary to limit the computational costs.

The complex model with seven parameters (Eq. 1) is not used in all cases. For stations without sub-hourly data, we do not use the flexible IDF-model, because the more complex model is not expected to improve results (Fauer et al 2021). Here, only $\tilde{\mu}$, σ_0, ξ and η are allowed to vary; θ, η_2 and τ are held fixed at zero. For winter (DJF), the parameter τ is held fixed at zero, even when sub-hourly data is available. This parameter is particularly important, when the annual maxima potentially stem from different seasons (Fauer et al 2021; Ulrich et al 2021a) which is not

the case here. Moreover, in winter the parameter τ would have been chosen only twice, out of a possible maximum of 104 stations. This further justifies the exclusion of this parameter from the systematic model selection process. However, for summer (JJA) this parameter seems to improve the model since dependences of τ on large-scale variables were often significant (chosen 39 times out of 104). Hence, we allow τ to vary in summer. Alternatively, all possible combinations could have been probed and decided for with model selection. But, these limitations can be motivated by the previous arguments and dramatically reduce the computational costs of the following analysis. The maximum number of dependencies on large-scale variables for each parameter is set to two, e.g., the shape parameter can not depend on more than two different large-scale variables.

In this study, annual and monthly block maxima are used. Several other studies show that monthly maxima can be used to model GEV distributions (Ulrich et al 2021a; Rust 2009; Fischer et al 2019; Maraun et al 2009). Although, there is a debate whether a block size of one

month is sufficiently large to fulfill the requirements of a GEV distribution, since the length of droughts increase (Ionita et al 2022) and monthly precipitation sums might be zero in some cases. However, since this study aims for an analysis of different seasons which wouldn't be captured by annual maxima, we chose the monthly block size despite its drawbacks. Using the annual maxima, i.e., one value per year, easily enables the estimation of average return periods T since it is connected to the annual non-exceedance probability p from the GEV distribution function by $T = 1/(1 - p)$. Consequently, the exceedance probability is $p_e = 1 - p$. When using monthly maxima and three maxima for a season of 3 months, i.e., 3 values per year, the probability p_s from the distribution function has to be converted with $p = 1 - (1 - p_s)^{1/3}$ to get annual non-exceedance probabilities p, again.

In the final model, each d-GEV parameter will be modeled explicitly as a function of the large-scale covariates. The function will be a polynomial up to the fourth order and selected via a model selection process (Sec. 2.5). For future reference, we call the new model which contains large-scale information the *large-scale model*.

2.5 Systematic model selection

Not all d-GEV parameters show a significant dependence on large-scale variables and using too many parameters increases their uncertainty (Di Baldassarre et al 2006). Also, overfitting might be a potential problem. Therefore, we conducted a stepwise Bayesian information criterion (BIC) model selection for each station individually as follows: The initial reference model is a d-GEV model without any large-scale dependence. Then, all possible parameter-variable dependencies (combinations) of d-GEV parameters (7), large-scale variables (4) and order of polynomial (4) are added individually (7*4*4=112 possible models) in parallel. Whichever model scores the lowest two-fold cross-validated BIC is selected as the new reference model. Then, again all remaining possible model combinations are added to the new reference model in turns. This procedure is repeated until none of the new models has a lower BIC than the reference model.

This methodology is used for the final model. Please note that it differs from the methodology, presented in Sec. 2.3 which is not used for the final model but is meant to motivate the need of large-scale modeling.

2.6 Quantile skill index

We compare the new model with large-scale information to a reference model without large-scale information for verification. Therefore we use the Quantile Skill Index

$(-1 \leq QSI \leq 1)$ which is based on the quantile score $(QS > 0)$ (Bentzien and Friederichs 2014).

The QS compares the modeled quantile q with all data points z_n (see Eq. 5) and penalizes data points that are higher than the modeled quantile with a weight that scales with the non-exceedance probability p of the quantile (Eq. 4). This way, the model is penalized strongly, when data points exceed model quantiles with a high non-exceedance probability:

$$\rho_p(u) = \begin{cases} pu & , u > 0 \\ (p-1)u & , u \leq 0 \end{cases} \quad (4)$$

$$QS(p) = \sum_i^n \rho_p(z_i - q). \quad (5)$$

The quantile score is calculated for model (QS_M) and reference (QS_R). For a given probability p and duration d, the QSI shows whether a model yields more adequate p-quantiles (values close to 1) than the reference or worse (values close to -1)(cf. Fauer et al 2021, Section 2.5):

$$QSI = \begin{cases} 1 - QS_M/QS_R & , QS_M \leq QS_R \\ QS_R/QS_M - 1 & , QS_M > QS_R. \end{cases} \quad (6)$$

The QSI is cross-validated (CV) by using every possible three subsequent years as testing set and the remaining years as training set (test set in the first CV step: year 1 to 3, second CV step: year 2 to 4,...). The quantile score from all CV steps is averaged and the two QS from model and reference are used for the calculation of the QSI.

Summarizing the processes of model selection (Sec. 2.5) and verification (this section), please note that model selection is conducted with a two-fold cross-validated BIC and verification is done with cross-validated QSI.

3 Results

3.1 Overview of selected models

The number of models in which each parameter-variable combination has been chosen is shown in Fig. 3. The black horizontal lines show the mean proportion of models for this variable. However, the influence of d-GEV parameters on the model is very different and thus, the black lines just illustrate roughly the importance of a large-scale variable. All in all, large-scale dependencies were chosen most often by the rescaled location $\tilde{\mu}$ and scale offset σ_0 parameters.

The d-GEV parameters for stations where at least 30 years of sub-hourly data are available are shown in Table 1. This set of parameters is the result of the stepwise BIC-model selection process. Stationary parameters are combined in the vector $\phi = \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta, \eta_2, \tau\}$ for summer
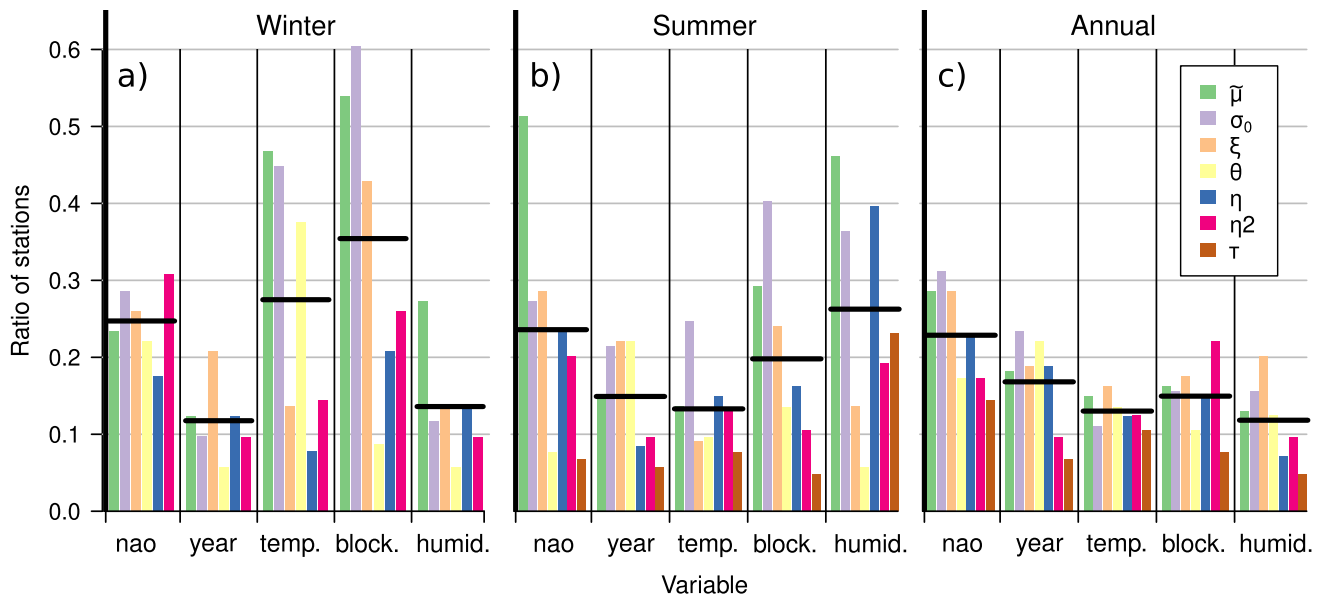
Fig. 3 For winter (**a**), summer (**b**) and annual maxima (**c**) and large-scale variable (abscissa) and d-GEV parameter (color), height of the bars show how often the stepwise regression with BIC chose a large-scale variable. Horizontal black lines show the average proportion of models for this large-scale variable and indicate the importance of each variable in the respective season

and annual models or $\phi_w = \{\tilde{\mu}, \sigma_0, \xi, \theta, \eta, \eta_2\}$ for winter, respectively. The other parameters show their functional dependency on large-scale variables in brackets, e.g., the shape parameter ξ depending on time t with a polynomial of third order notated as $\xi(t^3)$.

3.2 Verification

The large-scale flexible d-GEV models were verified against flexible d-GEV models without large-scale dependence using the QSI median over all stations. Figure 4 shows the QSI for all durations from 1 min to 5 days and non-exceedance probabilities p (return periods) up to 0.995 (200 years) and all seasons (a-c). Non-exceedance probabilities higher than $p_e = 0.98$ (50-year return period) have to be handled with care, because the quantile score cannot reasonably evaluate return periods much longer than the time range of the data. In this regime, a model is incentivized to yield larger values, since all data points are lower than the modeled quantile and the QS penalizes larger data points stronger. Therefore, black dots indicate whether the average number of years is equal or higher than the return period corresponding to the non-exceedance probability (vertical axis), but still might be unreliable for long return periods, e.g. 50 years.

For describing annual maxima (Fig. 4c), the large-scale model has a higher QSI in most durations d and non-exceedance probabilities p while in winter DJF (a) and summer JJA (b) there is no clear tendency. Despite there being no improvement of non-stationary modeling in some duration/probability regimes (blue), the new models gain

insight into dependencies (see Sect. 3.3). The color-scale exceeds the range of values in the plot because it is chosen consistently with previous studies evaluating the QSI of d-GEV models (Fauer et al 2021; Ulrich et al 2020).

3.3 Large-scale dependence of extreme precipitation

We present a visualization of modeling large-scale precipitation extremes which is an adaptation of known IDF curves (Fig. 5). The axes for intensity and duration stay the same, but different curves and colors show the range of a large-scale variable while the exceedance probability (average return period) is fixed to $p_e = 0.05$ (20 years) and the other large-scale variables are fixed to an average value. We call this visualization Intensity-Duration-Variable (IDV) curve. A stationary reference model without large-scale dependence is added (dashed line).

In a model where the duration offset θ depends on the year, intensities will vary for short durations (Fig. 5c). Dependence of rescaled location $\tilde{\mu}$, scale offset σ_0 or shape ξ (Fig. 5a-d) will let the intensities vary over the whole range of durations equally (on a log-scale) and produce a shift along the intensity-scale. Large-scale influence on the duration exponent might lead to opposing trends for both ends of the duration range (not shown). Dependence of the intensity offset τ will mostly effect the long-duration regime (Fig. 5c).

Another way of visualising the dependence of the exceedance probability (return-period) on large-scale variables is given in Fig. 6. It shows the dependence of

Table 1 d-GEV model parameters for selected stations

Station	Season	Dependencies
Buchenhofen	DJF	$\phi_w; \tilde{\mu}(T^1, h^4); \sigma_0(b^3); \xi(n^4, T^4); \theta(y^3); \eta_2(T^4, n^1)$
	JJA	$\phi; \tilde{\mu}(y^1, n^1); \xi(h^4, y^1); \theta(y^3, n^3); \eta(h^1, y^2); \eta_2(n^3); \tau(h^1, b^4)$
	Annual	$\phi; \tilde{\mu}(y^3, T^4); \xi(n^3, h^2); \theta(n^1); \eta(T^3); \eta_2(n^1, T^1)$
Leverkusen	DJF	$\phi_w; \tilde{\mu}(T^2, h^1); \sigma_0(y^1, n^2); \xi(h^3, y^3); \theta(y^4, T^2); \eta(y^3); \eta_2(n^4)$
	JJA	$\phi; \tilde{\mu}(h^2); \sigma_0(y^4); \xi(n^2, y^3); \theta(y^1); \eta(y^4); \eta_2(n^3)$
	Annual	$\phi; \sigma_0(y^1); \xi(T^1); \theta(h^3); \eta(T^4); \eta_2(y^4)$
Neumühle	DJF	$\phi_w; \tilde{\mu}(b^1, h^1); \sigma_0(T^1, y^4); \xi(h^2)$
	JJA	$\phi; \tilde{\mu}(T^1, h^4); \sigma_0(y^4, h^4); \xi(y^2); \theta(T^1); \eta_2(T^4); \tau(y^4)$
	Annual	$\phi; \tilde{\mu}(T^3, b^3); \xi(n^1, y^2); \theta(y^3, n^3); \tau(y^4, h^1)$
Solingen-Hohenscheid	DJF	$\phi_w; \tilde{\mu}(y^3, b^1); \sigma_0(T^4, h^1); \xi(y^3); \eta(y^2); \eta_2(y^1)$
	JJA	$\phi; \tilde{\mu}(n^1, y^3); \sigma_0(h^3); \theta(y^2); \tau(h^2)$
	Annual	$\phi; \sigma_0(y^4, b^2); \xi(b^1); \theta(y^3); \eta(T^4, y^4); \eta_2(b^2)$
Seehausen	DJF	$\phi_w; \tilde{\mu}(T^3); \sigma_0(b^1); \xi(y^1); \eta_2(n^2, h^4)$
	JJA	$\phi; \tilde{\mu}(h^3); \sigma_0(y^3, T^4); \xi(n^1); \eta(n^2, T^4)$
	Annual	$\phi; \tilde{\mu}(y^4, h^1); \sigma_0(h^4); \eta_2(n^1, b^1); \tau(n^3, b^1)$
Stötten	JJA	$\phi; \tilde{\mu}(T^2, b^2); \sigma_0(n^2, h^2); \xi(y^1); \eta(h^4); \eta_2(h^3)$
	Annual	$\phi; \tilde{\mu}(b^3, y^1); \sigma_0(b^1, T^4); \xi(h^2); \eta(b^1); \eta_2(n^2)$
Stuttgart-Echterdingen	DJF	$\phi_w; \tilde{\mu}(b^1); \tilde{\mu}(T^1); \sigma_0(T^4, n^2); \theta(T^3); \eta(y^2, T^3); \eta_2(n^2, b^2)$
	JJA	$\phi; \tilde{\mu}(h^3, T^1); \sigma_0(b^1, h^3); \theta(T^3, y^3); \eta(h^4); \eta_2(h^3); \tau(T^4)$
	annual	$\phi; \sigma_0(b^1); \theta(y^3); \eta_2(b^1, n^2); \tau(y^2)$

The variable ϕ_w contains all parameters without large-scale dependence for winter and ϕ for summer and annual models. The large-scale variables are nao N, year y, temperature T, blocking b and humidity h with their corresponding polynomials

extreme precipitation on the large-scale variable (abscissa) for many stations in one plot with an average over stations added (solid lines) to improve robustness.

For Fig. 6, an artificial reference event was defined which has an annual exceedance probability (average return period) of $p_e = 0.05$ (20 years). For this event, we use associated values for the large-scale parameter of the NAO-index $N = 0$, year $y = 1990$, temperature $T = 10°C$, blocking-index $b = 0$ and humidity $h = 75\%$. Note, that all curves intersect at these values. For each station (thin

lines), one large-scale variable has been varied (in each column of Fig. 6) while the others and the return level of the reference event are fixed to the large-scale reference values. Extrapolations outside of the data range of this parameter at this station are indicated as dotted lines. The thick solid lines show the median over all summer (red), winter (blue) and annual (black) models. In the following, only those median lines will be interpreted.

In most cases, there is no difference between the durations (rows in Fig. 6, 1 min to 3 days, d in hours). Only two

Fig. 4 Verification of large-scale model. **a–c** The QSI is shown for every probability and duration with the stationary model as reference model. Positive values of QSI (red color) indicate an improvement of the large-scale model over the reference model. **d** Histogram over all quantile skill indices. Most QSI values are between −0.05 and 0.05 (colored in white in a-c) and are considered non-relevant. Black dots indicate that the corresponding return period is longer than the average length of time series in the data
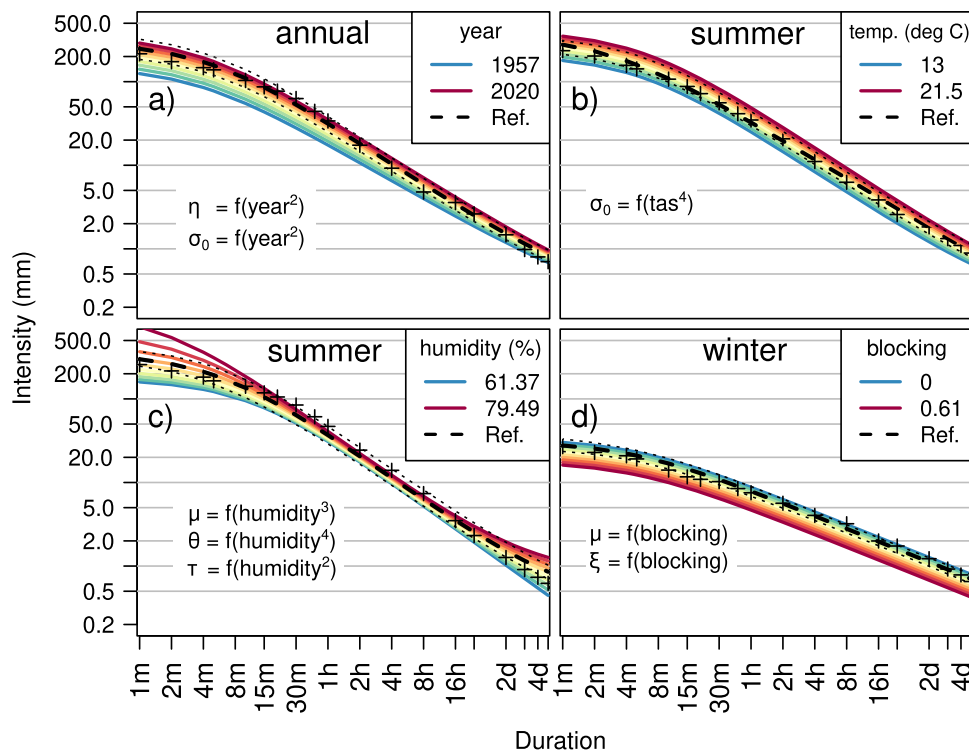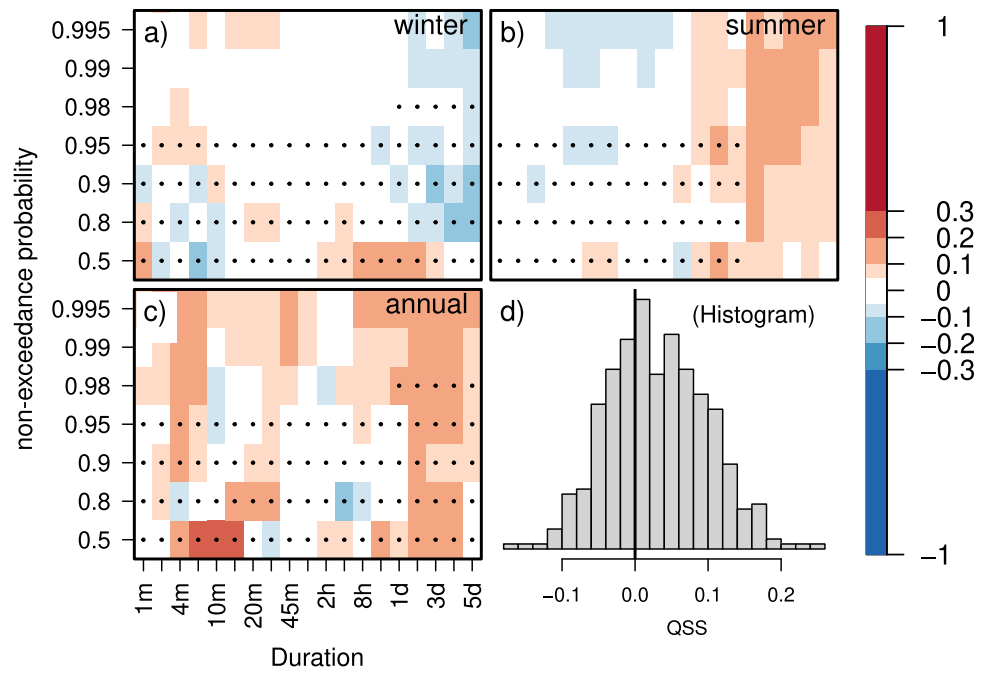
Fig. 5 Intensity-Duration-Variable (IDV) curve for selected stations and selected seasons. Intensity over duration is shown for different values of large-scale variables and for a fixed average return period of 5%. Stations are chosen for the purpose of visualizing the effect on d-GEV parameters. They do not necessarily represent a general trend over all stations. Dashed thick black curves show the reference without large-scale dependence. Dashed thin black curves show the confidence interval of the reference, obtained by bootstrapping. Black crosses show the empirical quantiles. **a** Annual model of station Lindscheid, visualizing the effect of year on duration exponent η and and scale offset σ_0. **b** Summer model of station Uckermünde, visualizing the effect of temperature on scale offset σ_0. **c** Summer model of station Angermünde, visualizing the effect of humidity on rescaled location $\tilde{\mu}$, duration offset θ and intensity offset τ. **d** Winter model of station Doberlug-Kirchhain, visualizing the effect of blocking on rescaled location $\tilde{\mu}$, and shape ξ (only daily data)
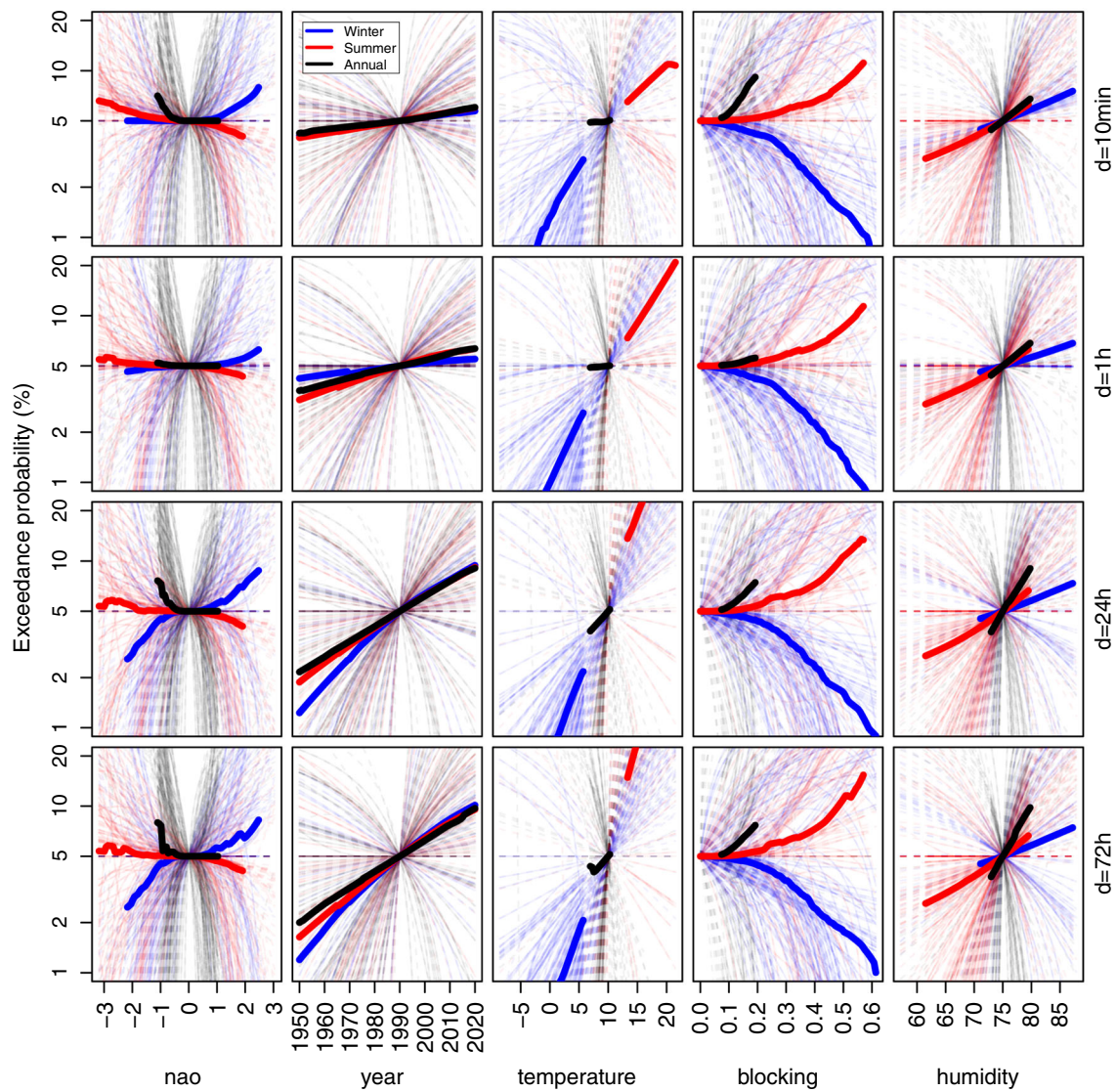
Fig. 6 Change of exceedance probability for extreme events with respect to a reference event with exceeding probability $p_e = 0.05$ in a situation which is defined as large-scale reference (see text). All curves meet at this value. Varying one large-scale variable (column) while fixing the other values allows to analyse the new exceedance probability p_e of the reference event in the new large-scale situation.

This figure shows how the probability of the reference event changes for different seasons (color) and durations (rows). Thin lines in the background represent individual stations/models while thick lines represent the median probability over all stations. Interpolations (extrapolations) are indicated as solid (dotted) thin lines for each station/model

clear duration-sensitive effects have been found: (1) The steepness of change with year increases for larger durations and (2) the steepness of change increases for the temperature for larger durations. The following results are similar for all durations. There is a positive effect of NAO on probability of an extreme event in winter and a slightly negative effect in summer. The trend over time (year) is almost always clearly positive, but smaller for short durations. Rising temperature has a positive effect on the probability of extreme rainfall in winter and summer. Blocking situations support extreme rainfall in summer and counteract extremes in winter. Higher humidity has a positive effect on the occurrence of extremes in all seasons.

The uncertainty of the 50-year return level over a duration of 24 h in the large-scale model is lower than in the stationary reference model in 28% (not shown) of all

stations and seasons (only DJF: 18%, only JJA: 21%, only annual: 44%). Over a duration of 1 h, this value drops to 19% (only DJF: 12%, only JJA: 15%, only annual: 30%).

4 Discussion and summary

The aim of this study was to investigate the dependence of precipitation extremes of different duration on large-scale variables. There was no particular focus on the physical dynamics, leading to precipitation extremes. That is why the independent variables (NAO, temperature, blocking) were used in a large-scale setting on purpose with no finer than monthly resolution. In future studies, we plan to investigate variables on different time scales like daily temperature or daily blocking index instead of monthly values and include seasonality in the model. Furthermore, we plan to create projections of extreme intensities in the future depending on large-scale covariates, however there are some challenges to address (Faulkner et al 2023).

According to the QSI most models with large-scale information outperform the reference without large-scale information (red regions in Fig. 4), meaning that quantiles estimated from the model with large-scale information in most cases are better than those from the simpler model. Additionally, the complex model is able to describe the influence of large-scale variables on extreme precipitation and provides new information and therefore has an advantage over the simple model. Furthermore, the fact that large-scale variables decreases the BIC during the model selection process shows that the model profits from this information. Still, the heterogeneous character of the out-of-sample-performance from the cross-validated QSI verification (Fig. 4) is noteworthy.

Large-scale influence only marginally depends on the duration (Fig. 6). But, using durations not only provides information about time scales in the final model, but also improves efficiency of data usage (Ulrich et al 2020).

A disadvantage of the new large-scale model is its increased uncertainty, due to the higher number of parameters that have to be estimated. Comparing this model extension to the step from a GEV model to a d-GEV model, there is a difference in efficiency gain. When letting GEV parameters depend on duration, the uncertainty decreases (Ulrich et al 2020). However, when using d-GEV parameters as functions of large-scale covariates, the uncertainty increases (see Sec. 3.3, last paragraph). In both cases, more information is used, but in the second case, the ratio of information usage and number of additional parameters is worse, so there seems to be no efficiency gain in the large-scale approach.

When comparing our results with Casanueva et al (2014), we find that both studies conclude to the same opposite association with NAO in winter and summer over Germany. Lenggenhager and Martius (2019, Fig. 12) find an increase of precipitation with blocking defined over a European sector (0°–30°W) in summer. In winter, the chance of precipitation is decreasing. Both these findings are in accordance with our results.

The aim of this study is to find meaningful large-scale variables that have an influence on extreme precipitation. Therefore, a parametrical duration-dependent GEV model includes the effect of large-scale variables and non-stationarity. A stepwise BIC model selection is conducted and the results are verified with a cross-validated QSI. The results are IDF-curves, depending on large-scale variables. Furthermore, the influence of large-scale effects on extreme precipitation can be investigated. We find that time (year) has a positive effect on exceedance probability of extremes for durations longer than 1 h while the effect of the NAO index, surface temperature averaged over Germany or the blocking index depend on the season. Especially the blocking index, the NAO index and the temperature are covariates that can change the exceedance probability of an extreme event by a factor of 2 or more. This shows that the non-stationary behavior of extreme precipitation should be acknowledged more. Our new large-scale model performs better than a stationary reference model in most duration-probability regimes and additionally is able to estimate probabilities of extreme precipitation in a changing climate.

Code availability The IDF model can be used with our R-package, available via CRAN (Ulrich et al 2021b). An example code script for the usage of the IDF package is available online (Fauer and Rust 2023).

Declarations

References

Aleshina MA, Semenov VA, Chernokulsky AV (2021) A link between surface air temperature and extreme precipitation over russia from station and reanalysis data. Environ Res Lett 16(10):105004. https://doi.org/10.1088/1748-9326/ac1cba

Arnbjerg-Nielsen K (2012) Quantification of climate change effects on extreme precipitation used for high resolution hydrologic design. Urban Water J 9(2):57–65. https://doi.org/10.1080/1573062X.2011.630091

Barnston AG, Livezey RE (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Mon Weather Rev 115(6):1083–1126. https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2

Bell B, Hersbach H, Berrisford P, et al (2020) Era5 monthly averaged data on single levels from 1950 to 1978 (preliminary version). copernicus climate change service (c3s) climate data store (cds). "Available online: https://cds.climate.copernicus-climate.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means-preliminary-back-extension?tab=overview, last access 09 August 2022"

Bentzien S, Friederichs P (2014) Decomposition and graphical portrayal of the quantile score. Q J R Meteorol Soc 140(683):1924–1934. https://doi.org/10.1002/qj.2284

Blanchet J, Ceresetti D, Molinié G et al (2016) A regional gev scale-invariant framework for intensity-duration-frequency analysis. J Hydrol 540:82–95. https://doi.org/10.1016/j.jhydrol.2016.06.007

Casanueva A, Rodríguez-Puebla C, Frías MD et al (2014) Variability of extreme precipitation over europe and its relationships with teleconnection patterns. Hydrol Earth Syst Sci 18(2):709–725. https://doi.org/10.5194/hess-18-709-2014

Cheng L, AghaKouchak A (2014) Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. Sci Rep 4(1):1–6. https://doi.org/10.1038/srep07093

Chiew FHS, Teng J, Vaze J et al (2009) Estimating climate change impact on runoff across southeast australia: Method, results, and implications of the modeling method. Water Resour Res. https://doi.org/10.1029/2008WR007338

Chow VT (1953) Frequency analysis of hydrologic data with special application to rainfall intensities. University of Illinois at Urbana Champaign, College of Engineering, Tech. rep

Coles S (2001) An introduction to statistical modeling of extreme values. Springer, London, https://primo.fu-berlin.de/FUB:FUB_ALMA_DS21803708050002883

Croitoru AE, Chiotoroiu BC, Ivanova Todorova V et al (2013) Changes in precipitation extremes on the black sea western coast. Global Planet Change 102:10–19. https://doi.org/10.1016/j.gloplacha.2013.01.004

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Statist Sci 27(2):161–186. https://doi.org/10.1214/11-STS376

Di Baldassarre G, Brath A, Montanari A (2006) Reliability of different depth-duration-frequency equations for estimating short-duration design storms. Water Resour Res. https://doi.org/10.1029/2006WR004911

Drobinski P, Alonzo B, Bastin S et al (2016) Scaling of precipitation extremes with temperature in the french mediterranean region: what explains the hook shape? J Geophys Res Atmos 121(7):3100–3119. https://doi.org/10.1002/2015JD023497

DWD (2022) Deutscher wetterdienst. Available online: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/, last access 09 June 2021

Fauer FS, Rust HW (2023). Maxima of station-based rainfall data over different accumulation durations and large scale covariates. https://doi.org/10.5281/zenodo.7822975

Fauer FS, Ulrich J, Jurado OE et al (2021) Flexible and consistent quantile estimation for intensity-duration-frequency curves. Hydrol Earth Syst Sci 25(12):6479–6494. https://doi.org/10.5194/hess-25-6479-2021

Faulkner DS, Longfield S, Warren S et al (2023) Modelling non-stationary flood frequency in england and wales using physical covariates. Hydrol Earth Syst Sci Discuss 2023:1–22. https://doi.org/10.5194/hess-2022-401

Fischer AM, Keller DE, Liniger MA et al (2015) Projected changes in precipitation intensity and frequency in switzerland: a multi-model perspective. Int J Climatol 35(11):3204–3219. https://doi.org/10.1002/joc.4162

Fischer M, Rust H, Ulbrich U (2019) A spatial and seasonal climatology of extreme precipitation return-levels: a case study. Spatial Stat 34(100):275. https://doi.org/10.1016/j.spasta.2017.11.007

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

Gutiérrez J, Jones R, Narisma G, et al (2021) Atlas. in climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. In: Climate Change 2021. Available from: http://interactive-atlas.ipcc.ch/

Hardwick Jones R, Westra S, Sharma A (2010) Observed relationships between extreme sub-daily precipitation, surface temperature, and relative humidity. Geophys Res Lett. https://doi.org/10.1029/2010GL045081

Ionita M, Nagavciuc V, Scholz P et al (2022) Long-term drought intensification over Europe driven by the weakening trend of the

Atlantic meridional overturning circulation. J Hydrol Reg Stud 42(101):176. https://doi.org/10.1016/j.ejrh.2022.101176

Iturbide M, Fernández J, Gutiérrez J, et al (2021) Repository supporting the implementation of fair principles in the ipcc-wg1 atlas. Available from:https://doi.org/10.5281/zenodo.3691645

Kadow C, Illing S, Lucio-Eceiza E et al (2021) Introduction to freva—a free evaluation system framework for Earth system modeling. J Open Res Softw 9(1):13. https://doi.org/10.5334/jors.253

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

Lenggenhager S, Martius O (2019) Atmospheric blocks modulate the odds of heavy precipitation events in Europe. Clim Dyn 53(7–8):4155–4171. https://doi.org/10.1007/s00382-019-04779-0

Maraun D, Rust HW, Osborn TJ (2009) The annual cycle of heavy precipitation across the united kingdom: a model based on extreme value statistics. Int J Climatol 29(12):1731–1744. https://doi.org/10.1002/joc.1811

Mishra V, Wallace JM, Lettenmaier DP (2012) Relationship between hourly extreme precipitation and local air temperature in the united states. Geophys Res Lett. https://doi.org/10.1029/2012GL052790

Mohr S, Wandel J, Lenggenhager S et al (2019) Relationship between atmospheric blocking and warm-season thunderstorms over western and central europe. Q J R Meteorol Soc 145(724):3040–3056. https://doi.org/10.1002/qj.3603

Nguyen V, Nguyen T, Wang H (1998) Regional estimation of short duration rainfall extremes. Water Sci Technol 37(11):15–19. https://doi.org/10.1016/S0273-1223(98)00311-4

NOAA (2022) Nao dataset. Available online: https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml, last access 09 August 2022

Otero N, Jurado OE, Butler T et al (2022) The impact of atmospheric blocking on the compounding effect of ozone pollution and temperature: a copula-based approach. Atmos Chem Phys 22(3):1905–1919. https://doi.org/10.5194/acp-22-1905-2022

Ouarda TBMJ, Yousef LA, Charron C (2019) Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. Int J Climatol 39(4):2306–2323. https://doi.org/10.1002/joc.5953

Rootzén H, Katz RW (2013) Design life level: quantifying risk in a changing climate. Water Resour Res 49(9):5964–5972. https://doi.org/10.1002/wrcr.20425

Rust HW (2009) The effect of long-range dependence on modelling extremes with the generalised extreme value distribution. Eur Phys J Spec Top 174(1):91–97. https://doi.org/10.1140/epjst/e2009-01092-8

Scherrer SC, Croci-Maspoli M, Schwierz C et al (2006) Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the euro-atlantic region. Int J Climatol 26(2):233–249. https://doi.org/10.1002/joc.1250

Schlef KE, Kunkel KE, Brown C et al (2023) Incorporating non-stationarity from climate change into rainfall frequency and intensity-duration-frequency (idf) curves. J Hydrol 616(128):757. https://doi.org/10.1016/j.jhydrol.2022.128757

Schliep EM, Cooley D, Sain SR et al (2009) A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. Extremes 13(2):219–239. https://doi.org/10.1007/s10687-009-0098-2

Schuster M, Grieger J, Richling A et al (2019) Improvement in the decadal prediction skill of the north atlantic extratropical winter circulation through increased model resolution. Earth Syst Dyn 10(4):901–917. https://doi.org/10.5194/esd-10-901-2019

Seneviratne S, Zhang X, Adnan M, et al (2021) Weather and Climate Extreme Events in a Changing Climate, Cambridge University Press, Cambridge, United Kingdom and New York, pp 1513–1766. https://doi.org/10.1017/9781009157896.013

Ulrich J, Jurado OE, Peter M et al (2020) Estimating IDF curves consistently over durations with spatial covariates. Water 12(11):3119. https://doi.org/10.3390/w12113119

Ulrich J, Fauer FS, Rust HW (2021) Modeling seasonal variations of extreme rainfall on different time scales in Germany. Hydrol Earth Syst Sci Discuss 2021:1–28. https://doi.org/10.5194/hess-2021-336

Ulrich J, Ritschel C, Mack L, et al (2021b) IDF: Estimation and Plotting of IDF Curves. https://CRAN.R-project.org/package=IDF, r package version 2.1.0

Westra S, Fowler HJ, Evans JP et al (2014) Future changes to the intensity and frequency of short-duration extreme rainfall. Rev Geophys 52(3):522–555. https://doi.org/10.1002/2014RG000464