The ability of a stochastic regional weather generator to reproduce heavy precipitation events across scales

Xiaoxiang Guan[1], Viet Dung Nguyen[1], Paul Voit[2], Bruno Merz[1,2], Maik Heistermann[2], Sergiy Vorogushyn[1]

[1]GFZ Helmholtz Centre for Geosciences, Section Hydrology, Potsdam, 14473, Germany
[2]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, 14469, Germany

Correspondence to: Xiaoxiang Guan (guan@gfz.de)

Abstract. We assess the ability of a regional weather generator to represent the extremity of heavy precipitation events (HPEs) across spatial and temporal scales. To this end, we implement the multi-site non-stationary Regional Weather Generator (nsRWG) for the area of Germany and generate 100 sets of synthetic daily precipitation data spanning 72 years. The weather extremity index (WEI) and its recent cross-scale modification (xWEI) are applied to quantify the cross-scale extremity of synthetic and observed HPEs and to compare their distributions. The results show that the nsRWG excels in replicating the extremity patterns for almost all 7 durations (ranging from 1 to 7 days) considered. The frequency of small-scale 1-day rainfalls is however slightly overestimated. nsRWG aptly reproduces the potential influential areas of HPEs, whether of short or long duration. It is capable of generating precipitation events mirroring the extremity patterns observed during past disaster-causing HPEs in Germany, while simultaneously accommodating their variations. This study demonstrates the potential of the nsRWG for simulating HPE-related hazard and assessing flood risks.

1 Introduction

Heavy precipitation events (HPEs) are rare weather phenomena which accumulate exceptional amounts of rainfall within hours to days, over areas ranging from a few to tens of thousands of square kilometers. As the main cause of damaging floods and landslides, HPEs are the costliest natural disasters in Europe (Gvoždíková et al., 2019; NatCatSERVICE, 2023; Banerjee et al., 2024). Climate change is expected to increase the frequency, intensity and spatial extent of HPEs (Christensen and Christensen, 2003; Lenderink and Fowler, 2017; Matte et al., 2022; Yang et al., 2023) and their associated impacts (Merz et al., 2021).

Germany has experienced several notable HPEs in recent years (Hu and Franzke, 2020). For example, in August 2002 heavy precipitation led to record-high water levels in the Elbe River and its tributaries (Kreibich, et al., 2017; Thieken et al., 2022). In July 2021, a widespread HPE hit the western and southern parts of Germany as well as neighboring countries and caused one of the most devastating flood events in German history with more than 180 fatalities and €33 billion damages (Apel et al., 2022, Szönyi et al., 2022, Mohr et al., 2023).

To understand physical mechanisms and to assess potential consequences of HPEs, past event analyses can be carried (e.g., Caldas-Alvarez et al., 2022, Mohr et al., 2023). Alternatively, large ensembles of climate model simulations (e.g., Ehmele et al., 2022) or synthetic weather fields can be generated to extract a set of plausible HPEs. The latter option represents a more computationally efficient approach by deploying stochastic Weather Generators (WGs). WGs are stochastic models that are capable to generate synthetic spatio-temporal fields of weather variables such as precipitation, temperature, humidity etc., retaining statistical properties of observed or climate model data on which WG is conditioned, such as autocorrelation, spatial covariance and multi-variable dependence. A large number of WG models have been introduced so far, based on various statistical methods among others reshuffling and perturbing analogue weather fields or applying multi-variate auto-regressive models (for a review, see Maraun et al., 2010, Haberlandt et al., 2011, Serinaldi and Kilsby, 2014, Benoit and Mariethoz, 2017, Nguyen et al., 2021). WGs can be instrumental in generating synthetic HPEs, thereby supporting flood risk management and climate adaptation (Breinl et al., 2013; Chen and Brissette, 2014; Harris et al., 2014, Sairam et al., 2021). WGs are widely used for estimation of hydrological design values (Winter et al., 2019), downscaling climate model output (Fatichi et al., 2011, Kiem et al., 2021), climate impact assessments on water resources (Harris et al., 2014, Najibi et al., 2024), and flood risk assessment (Sairam et al., 2021) providing long-term datasets for scenarios where observational data may be limited and downscaled future climate projections are needed. WGs are particularly effective when integrated with other models to better understand and prepare for HPEs and their consequences (Mehrotra and Sharma, 2010; Zhou et al., 2020). For instance, long-term synthetic weather fields (like precipitation and air temperature) generated by WGs can be used as meteorological forcing for hydrological models in order to quantify HPE-related floods and damages (Apel et al., 2016; Qin and Lu, 2014; Sairam et al., 2021). This approach allows to develop exceptional flood events needed for flood design or disaster management planning, as the generation of very long time series of flood flows and inundation increases the probability of obtaining situations where the unfortunate superposition of the different flood-generating processes leads to severe impacts (Falter et al., 2015). Such situations are rarely encountered in measured time series that are usually very limited in length.

Evaluating the performance of a WG is crucial to ensure that it accurately represents historical weather data and produces synthetic weather sequences that are adequate for the application context, e.g., flood estimation, drought assessment, climate change impact assessment (Tseng et al., 2020). The evaluation process helps to identify biases or limitations in the model output and fosters model improvements (e.g., Breinl et al., 2013, Serinaldi and Kilsby, 2014, Nguyen et al., 2021). Widely used metrics to evaluate synthetic precipitation data include mean, standard deviation and skewness, lag-1 autocorrelation, frequency of wet (or dry) days, and precipitation intermittency (Steinschneider and Brown, 2013; Tseng et al., 2020; Zhou et al., 2020; Nguyen et al., 2021). Additionally, the performance of WGs in simulating the extremity of precipitation events is of special interest. The extremity of precipitation events is usually described by intensity, duration and spatial extent statistics (Beniston and Stephenson, 2004; Jeferson de Medeiros et al., 2022; Müller and Kaspar, 2014; Zhang et al., 2011). The spatial consistency of multi-site WGs is sometimes evaluated based on the areal mean of the synthetic field within a region, e.g., catchment average rainfall (Ullrich et al., 2021). However, this method may underrepresent the variability

within the affected area, particularly when calculated within a fixed region, as the areal mean underestimates the extremity when only part of the region is heavily affected (Müller and Kaspar, 2014; Voit and Heistermann, 2022). To address these issues, the correlation between multiple sites can be used to measure the spatial dependence structure of precipitation amount across a region (Breinl et al., 2013; Gao et al., 2021; Tseng et al., 2020). Furthermore, the continuity ratio, expressed as the ratio between the average precipitation at one location when another location is dry or wet, can capture the spatial coherence of precipitation between adjacent locations (Wilks, 1998; Breinl et al., 2013). However, both the correlation coefficient and the spatial continuity ratio are dominated by the bulk of the precipitation events rather than by the extreme values.

HPEs can vary in duration, from short, intense downpours to prolonged periods of heavy rainfall. Quantile-based thresholds of block maxima of precipitation totals, and empirical probability distribution plots are typically used to characterize the WG performance with regards to extremes (Breinl et al., 2013; Zhou et al., 2020; Nguyen et al., 2021; Ullrich et al., 2021). Such methods are usually applied separately for different durations, neglecting the temporal coherence of HPEs which in reality can be extreme at different temporal and spatial scales simultaneously, triggering different types of flooding overlaying each other (Ramos et al., 2017). For instance, during the 2002 Elbe flood, small-scale extreme precipitation caused flash floods in several small Elbe tributaries. During the same event, extreme rainfall over large spatial scales and longer durations triggered fluvial flooding with widespread inundation and finally long-lasting groundwater flooding in the city of Dresden (Kreibich et al., 2005; Thieken et al., 2022). The ability of WGs to represent these cross-scale characteristics of precipitation events is thus essential for flood risk modeling. Up to now, no single measure has been used to evaluate the ability of WGs to capture the cross-scale extremity of rainfall.

In this study, we demonstrate a new evaluation approach for a WG based on the weather extremity index (WEI, Müller and Kaspar, 2014) and the cross-scale weather extremity index (xWEI) introduced by Voit and Heistermann (2022). These indices quantify the extremity of an event considering different duration levels and spatial extents. xWEI additionally integrates the extremity over different duration levels and spatial extents. Our aim is to evaluate how well the cross-scale extremity of precipitation events is captured by a WG, even if it is not specifically tailored or trained to do so.

2 Study area and data

The study area is Germany (Figure 1) given its exposure to hazards induced by heavy precipitation events (HPEs), as highlighted in the introduction. The non-stationary Regional Weather Generator (nsRWG) (Nguyen et al., 2024) is implemented for entire Germany and parts of the upstream neighboring countries, covering the five major river catchments Elbe, Rhine, Danube, Ems and Weser, to produce long-term spatially consistent synthetic precipitation and temperature data. In this study, we focus on the ability of the WG to represent the extremity of precipitation across scales. Weather generators that comprise several catchments and cover such a large spatial area (more than 650,000 km^2) are rare, and previous studies have commonly implemented WGs at smaller scales (Tseng et al., 2020; Ullrich et al., 2021; Gao et al., 2021).

Two datasets are used to parameterize nsRWG: (1) gridded precipitation data from E-OBS (version 25.0e; Cornes et al., 2018) and (2) mean sea level pressure and daily air temperature at 2 m height from the ERA5 reanalysis (Hersbach et al., 2020). Mean sea level pressure is used to classify circulation pattern types for which the local distributions of precipitation are conditioned. Regionally averaged daily temperature acts as a covariate of the local distribution in order to consider changes in precipitation for the same circulation pattern in a non-stationary warming world (Nguyen et al., 2024). Both E-OBS and ERA5 datasets are available at daily resolution, spanning from 1 January 1950 to 31 December 2021.

E-OBS is an ensemble gridded weather dataset and is available on a regular 0.25° grid covering Europe. It is based on station data collated by the European Climate Assessment and Dataset initiative (Cornes et al., 2018). To cope with the high computational demands, we have resampled the grid points to a reduced spatial resolution of 0.5°. The locations of the extracted E-OBS grid points are given in Figure 1.

The ERA5 dataset, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), is a state-of-the-art reanalysis dataset widely used in climate research (Hersbach et al., 2020). It provides a comprehensive and detailed representation of atmospheric conditions at the global scale. To classify the large-scale atmospheric situation, mean sea level pressure is extracted over an extent (25°N-70°N x 15°W-30°E, see Figure 1.a) that encompasses a substantial portion of Europe and adjacent regions. The regional average 2-m daily air temperature is computed for the domain 45.125°N – 55.125°N and 5.125°E – 19.125°E, which covers the nsRWG setup area.

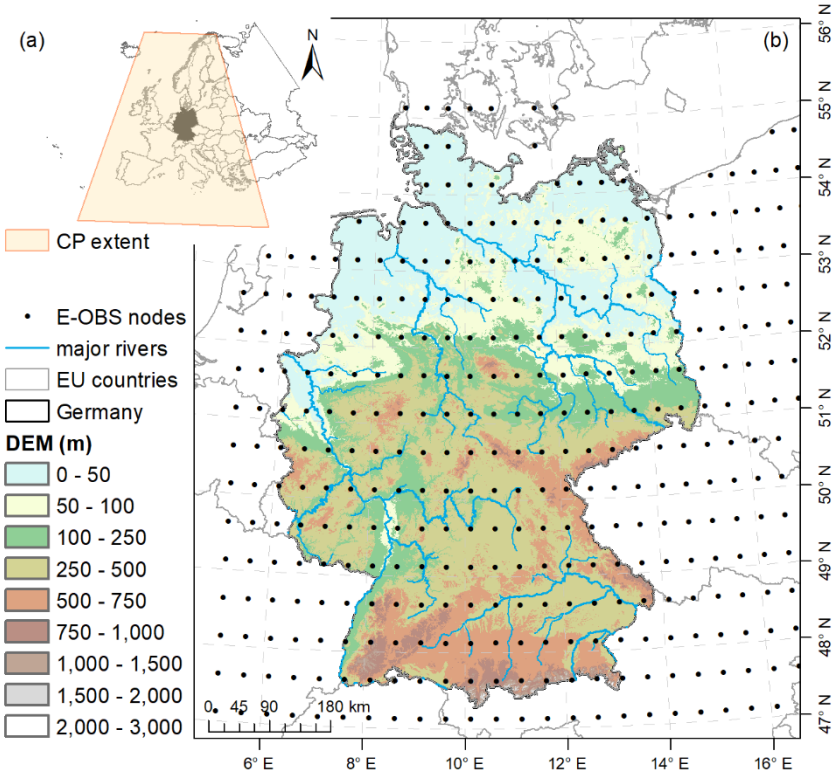


Figure 1: Study area (Germany and adjacent regions) and its topography, major rivers and E-OBS grid nodes. The extent where the mean sea level pressure was extracted to classify the circulation patterns (CP) is shown in subplot (a).

3 Methods

3.1 A stochastic multi-site non-stationary regional weather generator (nsRWG)

We adopt the non-stationary version of the multi-site regional weather generator nsRWG developed by Nguyen et al., (2024). nsRWG represents the spatio-temporal dependence across sites using the first-order multi-variate auto-regressive (MAR-1) model (Bardossy and Plate, 1992). The type-1 extended Generalized Pareto (extGP) distribution is used to model daily non-zero precipitation amounts. This distribution is suitable not only for capturing both the lower bulk of precipitation amounts and extreme values but also for providing a smooth transition along the precipitation range (Naveau et al. 2016; Nguyen et al., 2021).

In nsRWG, the precipitation distribution at each site is conditioned on the large-scale circulation pattern as a latent variable and the regional average daily temperature as a covariate of the extGP distribution scale parameter. In this way, climate variability and climate change due to changes in dynamic and thermodynamic properties of the atmosphere are considered. Atmospheric circulation is classified into 6 circulation patterns based on mean sea level pressure for winter (November 1st - April 30th) and summer (May 1st - October 30th) seasons (12 patterns in total). We use the objective classification algorithm SANDRA (Simulated ANnealing and Diversified RAndomization) based on the k-means clustering approach (Philipp et al., 2007) for circulation pattern classification. Further details about the nsRWG algorithm and configuration can be found in Nguyen et al. (2024).

The cross-scale precipitation performance of nsRWG is evaluated for the E-OBS grid cells in the study area for the period from 1 January 1950 to 31 December 2021. We generate 100 realizations of synthetic precipitation datasets with a time series length of 72 years corresponding to the length of the E-OBS dataset to ensure comparability.

3.2 WEI and xWEI

The weather extremity index (WEI) quantifies the extremity (a product of spatial extent and rarity) of an event, as well as the spatial extent and temporal duration at which the event reached its maximum extremity (Müller and Kaspar, 2014). In this context, the spatial extent of an event is not conceived as an area of spatially contiguous grid cells, but as a number of (potentially disjoint) cells within the study region (here Germany) which exceed a certain return period. The computation of WEI for individual HPE is illustrated in Figure 2. For a given spatial domain (in this case Germany), it starts with the estimation of return periods $P_{t,i}$ at each grid cell i, for durations from 1 to t days. For each duration t, the grid cells in the spatial domain are sorted in decreasing order, based on their return period $P_{t,i}$ (in years), and then aggregated over increasing areas A (in km^2) by using Eq. 1: first, E_{tA} is computed for the most extreme grid cell (n=1). Then, the following grid cells are added to the computation of E_{tA}, increasing the value of n by 1 until the estimated return period from every grid cell is

computed once. For each step, the area A equals the area of one single grid cell times n. In the final step, A corresponds to the size of the entire spatial domain. Finally, this procedure yields, for each duration t, a curve that shows E_{tA} as a function of A. The value of WEI is then defined by the maximum value of E_{tA} for all curves, and the spatial extent and duration at which the event was most extreme corresponds to the values of t and A for which this maximum of E_{tA} is achieved. Note that the E_{tA} curves typically have a well-defined maximum: for low values of A, the steep increase of $\sqrt{A/\pi}$ (the radius of a circle of size A) causes an increase of E_{tA} with A. For larger values of A, the decrease in return periods dominates the behavior and causes the E_{tA} curve to decrease. The corresponding area where the E_{tA} curve reaches its peak (WEI) is denoted here as WEI-area, which represents the spatial scale most severely affected by the HPE. The WEI-area is always much smaller than the area over which the HPE precipitation totals exceed 0 mm. It is rather a weighted measure indicating the area in the domain which is heavily influenced by the HPE and hence prone to HPE-related impacts.

$$E_{tA} = \frac{\sum_{i=1}^{n} ln\left(P_{t,i}\right)}{n} \cdot \sqrt{A/\pi}\,, A = \text{grid size} \times n \tag{1}$$

As each E_{tA} curve represents how the extremity of an event extends across spatial scales, Voit and Heistermann (2022) proposed a cross-scale weather extremity index (xWEI) by integrating E_{tA} over duration (ln(t)) and extent (A). xWEI quantifies how much the extremity of an event extends across both space and time (instead of the event just being extreme at one specific duration and extent). Hence, xWEI corresponds to the volume under the surface which is spanned by the E_{tA} curves (Figure 2.f), placed on a grid:

$$\text{xWEI} = \iint E_{tA} dA d\left(ln(t)\right) \tag{2}$$

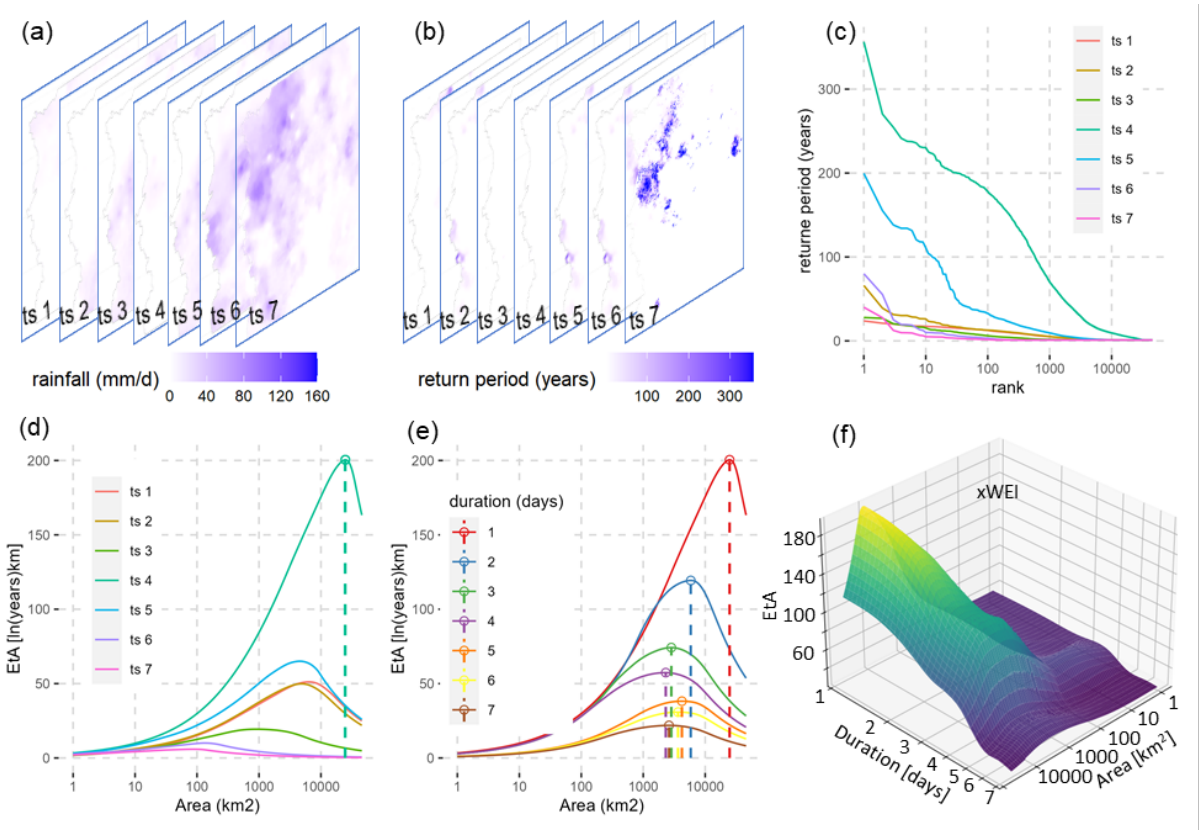


Figure 2. Calculation of the WEI and xWEI for an exemplary HPE: (a) maps of daily rainfall for a precipitation event lasting 7 days; (b) return period at each individual grid cell of each map; (c) sorting of the return period estimates for each duration in decreasing order for each map; (d) calculation of the E_{tA} curves and selection of the curve with the highest peak to represent the extremity pattern of the precipitation for the duration of 1 day; (e) repetition of the procedures (a-d) for other durations (2, 3, …, 7 days) to derive the E_{tA} curves; (f) 3D interpolated surface of E_{tA} for all durations and extent; xWEI is defined as the volume underneath E_{tA} surface.

For the exemplary HPE event shown in Figure 2, the analysis highlights that the daily extremity (in terms of 1-day rainfall intensity) occurs on the 4th day, with the highest E_{tA} curve among the 7 days (Figure 2d). Comparing extremity across durations, the HPE is most extreme for a 1-day duration, affecting an area of over 10,000 km² (WEI-area). For longer durations (≥3 days), the HPE consistently is extreme approximately for the same areas, showing a stabilization in spatial extent for these durations.

To estimate the return periods required to obtain E_{tA}, we use the Generalized Extreme Value (GEV) distribution, which has been found suitable for modeling precipitation extremes (Fowler and Kilsby, 2003) and has also been used in previous WEI studies (Gvoždíková et al., 2019; Minářová et al., 2018; Müller and Kaspar, 2014). The computation of the xWEI requires return periods for multiple durations (1 to 7 days in our case). We use the dGEV method (duration-dependent GEV, Koutsoyiannis et al., 1998) to estimate return periods consistently across durations for each grid cell. Previous studies have shown that this method preserves the tail behaviour of precipitation extremes across durations and modifies the scale and

location parameters of the GEV distribution to explicitly account for dependency on the duration of events. This allows for more accurate modeling of precipitation intensity-duration-frequency relationships and reduces the uncertainty (Ulrich et al., 2020; Fauer et al., 2021). The parameters of the dGEV distribution at each grid cell are obtained by Maximum Likelihood Estimation using the R package IDF (Ulrich et al., 2021) based on the annual maximum series of rainfall across considered durations.

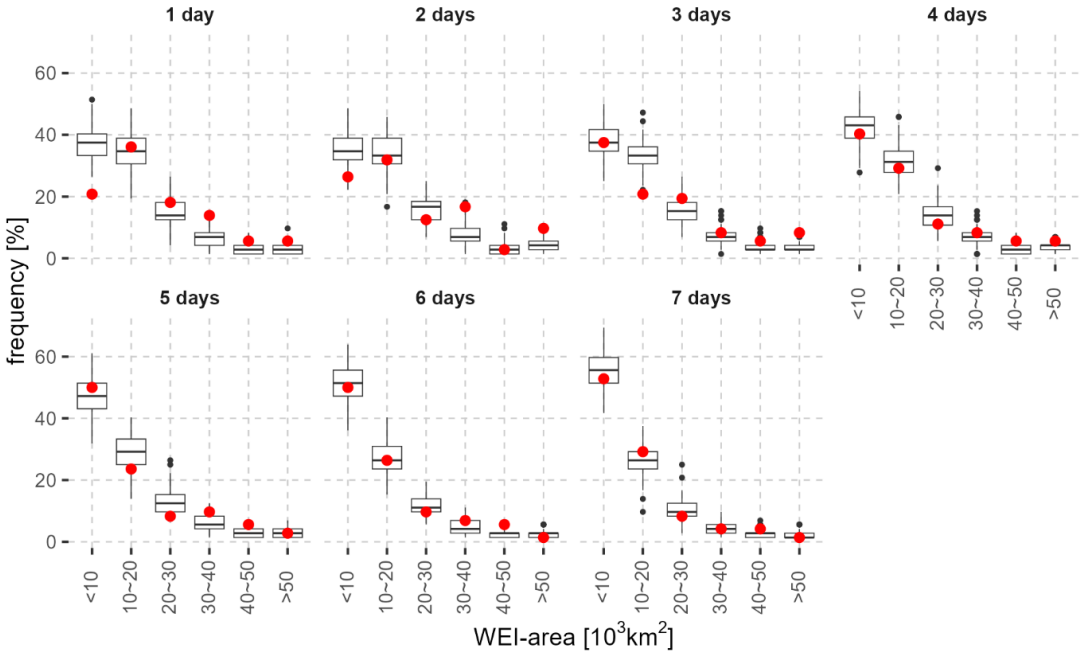
To assess the extremity of HPEs across Germany, durations from 1 to 7 days are selected, as they are sufficient to capture the events responsible for disastrous flood damage in the region (Ganguli and Merz, 2019). While the WEI and xWEI can be extended to sub-daily scales (Voit and Heistermann, 2022), this analysis is conducted at the daily scale due to the limited availability of sub-daily precipitation observations. Additionally, short-duration high-intensity precipitation events usually affect comparatively small areas, whereas multi-day HPEs span larger regions and are more relevant for this study (Lengfeld et al., 2021; Lengfeld et al., 2019; Orlanski, 1975).

The most extreme HPEs for these 7 durations across Germany are identified by extracting annual maximum WEI values for each duration based on E-OBS precipitation data for the historical period 1950-2021. The corresponding WEI-areas (the area where the E_{tA} curve reaches its peak WEI) of the annual most extreme HPEs for each duration are categorized into 6 classes (Figure 3) in order to analyze the spatial properties of HPEs in Germany. The same procedure is carried out for each of the 100 realizations to evaluate the performance of nsRWG in reproducing the cross-scale properties of HPEs in Germany.

4 Results and discussion

4.1 nsRWG performance for WEI

Figure 3 shows the frequency distribution of the observed (E-OBS) and simulated (nsRWG) WEI-areas of the annual maximum HPEs for 7 different durations. The distributions of the annual maximum WEI values are given in Figure 4. The results show that more than half of the annual maximum HPEs in Germany are events with WEI-areas $\leq 20 \times 10^3$ km^2 for all durations. HPEs with larger spatial extent of 20×10^3 km^2 or more have a frequency of less than 20% in the past 72 years. However, these events are the most severe ones regarding their WEI values (Figure 4). The nsRWG is able to reproduce the annual maximum WEI and its corresponding areas for the 7 durations. The boxplots in Figures 3 and 4 show that the simulated distribution patterns are in good agreement with the E-OBS observations. However, for shorter durations (1 day and 2 days), the frequencies of HPEs with WEI-areas $\leq 20 \times 10^3$ km^2 are overestimated, while those with larger WEI-areas are underestimated (Figure 3). For durations of 3 days and longer, the observed and simulated HPE frequencies are more balanced.



205

Figure 3: Frequency distribution of observed and nsRWG-simulated WEI-areas of the annual maximum HPEs in the period 1950-2021 for 7 durations. The titles of the subplots are the durations of interest. The red dots denote the frequency of HPEs in different area categories from E-OBS precipitation data and the boxplots represent the simulated frequency distribution from 100 nsRWG realizations.

210 The relation between the annual maximum WEI values and their return periods derived from the nsRWG data agrees well with the E-OBS observations for the 7 durations (Figure 5). This comparison confirms the ability of the nsRWG to simulate the occurrence probability of HPEs in Germany. The observed probability plots (red dots in Figure 5) are well enclosed by the simulated ranges (shadowed areas) from the nsRWG realizations. This is especially true for high return periods, which demonstrates the good performance in simulating HPEs for different durations. For short durations and return periods 215 between 2 and about 10 years, as well as for durations longer than 4 days with return periods between 10 and 20 years, the nsRWG slightly underestimates the observed WEI. In contrast to traditional intensity-duration-frequency curves, there is no systematic difference between the empirical probability plots of annual maximum WEI series of different durations, as the WEI is based on return periods rather than precipitation totals or averages. This makes the WEI index comparable across temporal scales. For example, for HPEs with the same WEI values, the return period increases with duration. This indicates 220 that the occurrence probability of an HPE with the same extremity becomes smaller with increasing duration.

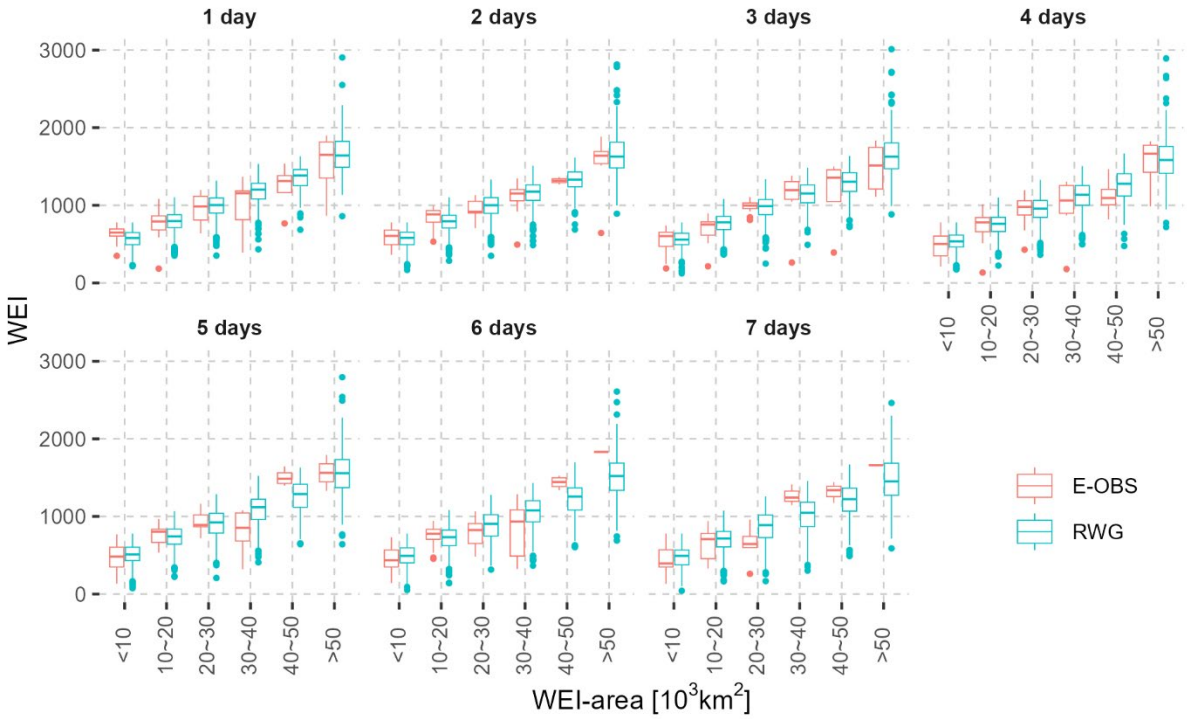


Figure 4: Distribution patterns of observed (red) and nsRWG-simulated (blue) annual maximum WEI series (denoted in y-axis, with the same unit as E_{tA}: ln(years)·km) for 7 durations. The WEI-areas of HPEs are categorized into 6 classes (see x-axis).

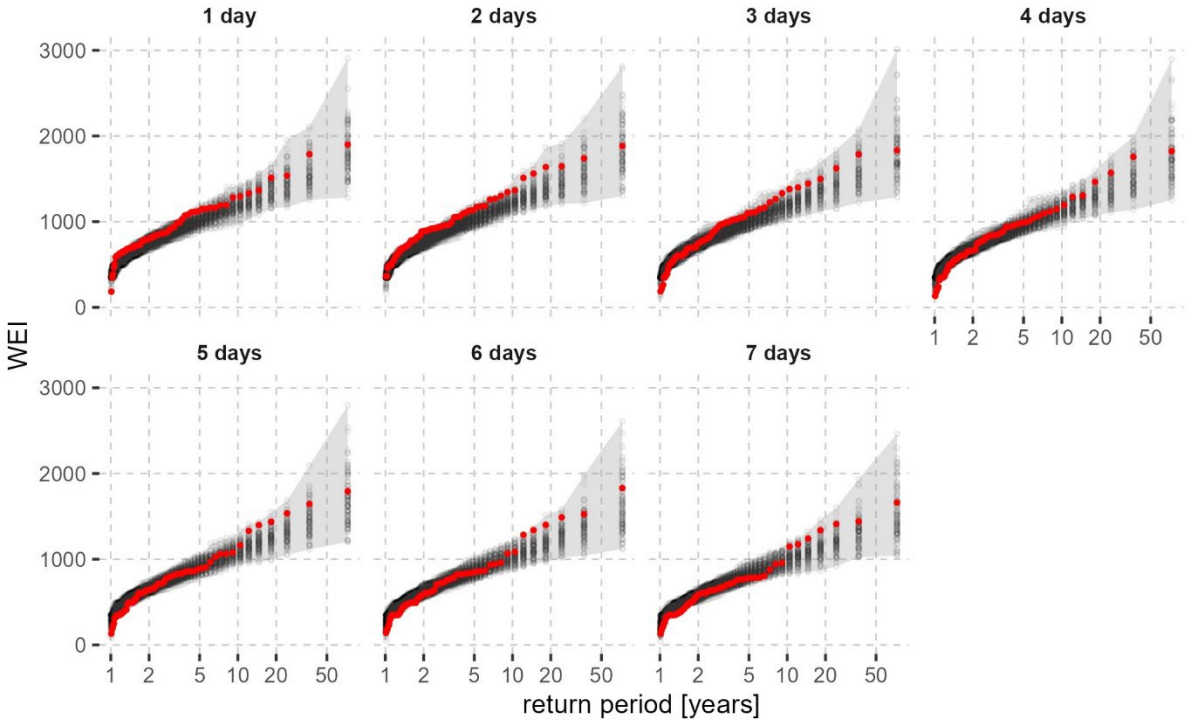


Figure 5: Empirical probability plots of E-OBS observed (red dots) and nsRWG simulated (black dots) annual maximum WEI for 7 durations. The Weibull plotting positions (Makkonen, 2006) are used to estimate the return periods. The grey shaded area indicates the upper and lower boundary of the nsRWG simulations.

4.2 nsRWG performance for xWEI

We compute the xWEI for HPEs in Germany for the period 1950-2021 and extract the annual maximum series from both the E-OBS dataset and the nsRWG realizations. The empirical probability plots of annual maximum xWEI, based on Weibull plotting positions (Makkonen, 2006), agree well for both datasets: the cross-scale extremity index xWEI of the observed data lies within the range of the 100 realizations (Figure 6). However, for return periods between 2 and 5 years, the realizations of nsRWG tend to underestimate the xWEI, similar to the performance with respect to the WEI.

Figure 7 shows the extremity pattern of a real HPE event in August 2002 – one of the most damaging events in Germany. In addition, HPEs from nsRWG realizations with similar xWEI values to the August 2002 HPE are shown. Figure 7 (b) demonstrates that the nsRWG is able to generate HPEs with spatial and temporal extremity patterns very similar to the August 2002 event, characterized by the highest extremity for the durations of 1d and 2d and an affected area of approximately 20 000 km^2. The other two nsRWG-generated HPEs with similar xWEI values as the August 2002 event show different cross-scale extremity patterns (emphasis on longer durations of 3-4 days). The reproducibility of historical HPEs illustrates the ability of nsRWG in representing the cross-scale extremity of HPEs in Germany. The variations in

synthetic events and the respective E_{tA} surfaces further demonstrate how nsRWG can generate synthetic events, which are similar in terms of their cross-scale extremity but have their emphasis at different spatial and temporal scales.

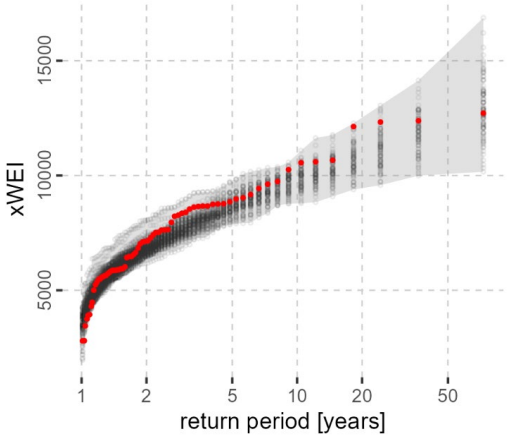


Figure 6: Empirical probability plots of observed (E-OBS, red dots) and simulated (nsRWG, black dots) annual maximum xWEI. The Weibull plotting position is used to estimate the empirical return periods. The grey shadowed area indicates the upper and lower boundary of simulated xWEI from 100 nsRWG realizations.

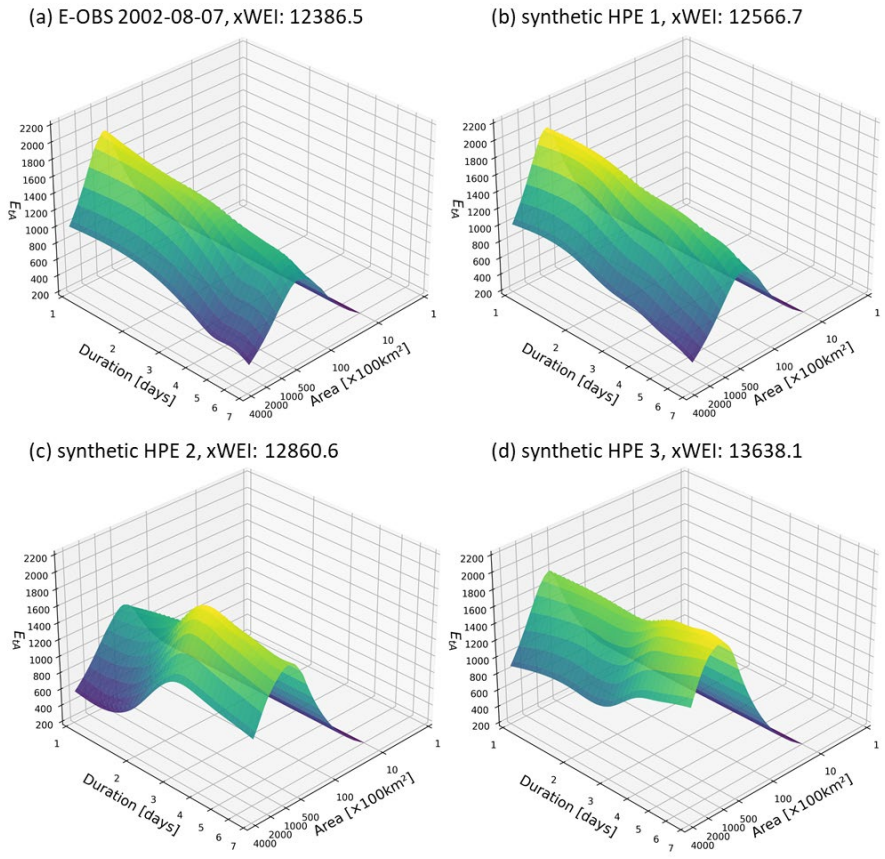


Figure 7: Cross-scale extremity pattern (interpolated E_{tA} curves over duration and area) of HPEs for (a) the August 2002 event and (b-d) 3 HPEs with similar xWEI values generated by nsRWG. The HPE in (b) shows great similarity in the cross-scale extremity pattern with the August 2002 event in (a), while the HPEs in (c) and (d) display different extremity patterns, although their xWEI values are similar to the actual event in August 2002.

5 Conclusions

In this study, we evaluate the ability of a multi-site stochastic regional weather generator to capture the extremity of HPEs across spatial and temporal scales. For this purpose, we set up the nsRWG at a large scale (covering all of Germany and riparian regions) and generate 100 realizations of 72 years of synthetic precipitation data at daily resolution. The performance evaluation of nsRWG in simulating precipitation focuses on the event scale and extreme cases. This focus complements typical proxy statistics (like mean and standard deviation) that tend to only represent the general properties of WG in precipitation generation. Two indices, WEI and xWEI, are used to measure the extremity of observed and synthetic HPEs, both of them based on the spatial aggregation of return periods of precipitation totals for several durations of interest. While WEI quantifies the maximum extremity of an event that occurred at a specific spatial extent and temporal duration,

xWEI integrates extremity across the spatial and temporal scales of interest. The results demonstrate that nsRWG performs well in simulating the extremity patterns across most spatial and temporal scales of HPEs in Germany. However, it tends to overestimate the frequency of events with short durations and relatively small spatial extents. Using the August 2002 event as an example, we illustrate how the nsRWG is able to generate precipitation events with spatio-temporal extremity patterns similar to those of historical disaster-causing HPEs.

With regard to future research, we emphasize that the choice of the spatial domain at which WGs are evaluated (here: all of Germany) is always a trade-off: on the one hand, the impacts of HPEs unfold at the catchment scale, and it would be an obvious next step to evaluate the ability of the WG to reproduce the frequencies of WEI and xWEI occurrence at the scale of specific river catchments (at the cost of computational effort, as this would require a higher spatial resolution). On the other hand, we need to be aware that an HPE that occurred in one catchment may potentially also occur in a neighbouring catchment, so that an analysis at a larger spatial domain (as the present one) is certainly warranted. Such considerations also show the links to "spatial counterfactuals", which have recently gained attention (Merz et al., 2024, Voit and Heistermann, 2024, Vorogushyn et al., 2024). Counterfactuals are scenarios that describe alternative ways of how an event could have unfolded (Woo, 2019; Montanari et al., 2024). These scenarios could involve conditions where specific factors, such as anthropogenic climate change, natural climate variability, or other boundary conditions, are altered or removed (Gauch et al., 2020). Spatial counterfactuals are a special case of counterfactual scenarios that consider transposition of hazard characteristics e.g., precipitation in space. Using spatial counterfactual scenarios, we can investigate the impact of HPEs in the hypothetical case that they had happened elsewhere. Weather generators could be a useful tool to explore how events with similar extremity indices could unfold in different locations of the spatial domain, or with different spatio-temporal signatures, hence supporting the evaluation of counterfactual scenarios.

Code/Data availability

The code and data to exemplify the computation of both WEI and xWEI can be found in the following repository: https://doi.org/10.5281/zenodo.6556463 (Voit, 2022). The gridded precipitation data from E-OBS (version 25.0e; Cornes et al., 2018) is available at the European Climate Assessment & Dataset (ECA&D, https://www.ecad.eu/download/ensembles/download.php). The ERA5 mean sea level pressure and daily air temperature at 2 m height covering Europe can be found at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

Author contributions

XG: data curation, conceptualisation, methodology, formal analysis, visualization, writing (original draft preparation). DVN: data curation, methodology, software. PV: methodology, software, writing – review & editing. BM: supervision, writing – review & editing. MK: writing – review & editing. SV: conceptualisation, supervision, writing (original draft preparation).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This research has been funded by the Federal Ministry of Education and Research of Germany in the framework of the project FLOOD (project number 01LP2324E) as a part of the ClimXtreme Research Network on Climate Change and Extreme Events within the framework program Research for Sustainable Development (FONA3). Xiaoxiang Guan is funded by the China Scholarship Council for his PhD research (Grant #: 202106710029). We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu).

References

Apel, H., Martínez Trepat, O., Hung, N.N., Chinh, D.T., Merz, B. and Dung, N.V. Combined fluvial and pluvial urban flood hazard analysis: concept development and application to Can Tho city, Mekong Delta, Vietnam. Natural Hazards and Earth System Sciences, 16(4), 941-961, doi:10.5194/nhess-16-941-2016, 2016.

Apel, H., Vorogushyn, S., and Merz, B. Brief communication: Impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany. Natural Hazards and Earth System Sciences, 22(9), 3005-3014, doi:10.5194/nhess-22-3005-2022, 2022.

Banerjee, C., Bevere, L., Garbers, H., Grollimund, B., Lechner, R., and Weigel, A. sig-ma. Natural catastrophes in 2023: gearing up for today's and tomorrow's weather risks, 2024.

Bardossy, A. and Plate, E.J. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resources Research, 28(5), 1247-1259, doi:10.1029/91WR02589, 1992.

Beniston, M. and Stephenson, D.B. Extreme climatic events and their evolution under changing climatic conditions. Global and Planetary Change, 44(1), 1-9, doi:10.1016/j.gloplacha.2004.06.001, 2004.

Benoit, L. and Mariethoz, G. Generating synthetic rainfall with geostatistical simulations. Wiley Interdisciplinary Reviews-Water, 4(2), doi:10.1002/wat2.1199, 2017.

Breinl, K., Turkington, T. and Stowasser, M. Stochastic generation of multi-site daily precipitation for applications in risk management. Journal of Hydrology, 498, 23-35, doi:10.1016/j.jhydrol.2013.06.015, 2013.

Caldas-Alvarez, A., Augenstein, M., Ayzel, G., Barfus, K., Cherian, R., Dillenardt, L., Fauer, F., Feldmann, H., Heistermann, M., Karwat, A., Kaspar, F., Kreibich, H., Lucio-Eceiza, E. E., Meredith, E. P., Mohr, S., Niermann, D., Pfahl, S., Ruff, F., Rust, H. W., … Quaas, J. (2022). Meteorological, impact and climate perspectives of the 29 June 2017 heavy precipitation event in the Berlin metropolitan area. *Natural Hazards and Earth System Sciences*, *22*(11), 3701–3724. https://doi.org/10.5194/nhess-22-3701-2022

Chen, J. and Brissette, F.P. Stochastic generation of daily precipitation amounts: review and evaluation of different models. Climate Research, 59(3), 189-U145, doi:10.3354/cr01214, 2014.

Christensen, J.H. and Christensen, O.B. Severe summertime flooding in Europe. Nature, 421(6925), 805-806, doi:10.1038/421805a, 2003.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. and Jones, P.D. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. Journal of Geophysical Research: Atmospheres, 123(17), 9391-9409, doi:10.1029/2017JD028200, 2018.

Falter, D., Schröter, K., Dung, N.V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H. and Merz, B. Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. Journal of Hydrology, 524, 182-193, doi:10.1016/j.jhydrol.2015.02.021, 2015.

Fauer, F.S., Ulrich, J., Jurado, O.E. and Rust, H.W. Flexible and consistent quantile estimation for intensity–duration–frequency curves. Hydrology and Earth System Sciences, 25(12), 6479-6494, doi:10.5194/hess-25-6479-2021, 2021.

Fatichi, S., Ivanov, V. Y., & Caporali, E. (2011). Simulation of future climate scenarios with a weather generator. *Advances in Water Resources*, *34*(4), 448–467. https://doi.org/10.1016/j.advwatres.2010.12.013

Fowler, H.J. and Kilsby, C.G. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. International Journal of Climatology, 23(11), 1313-1334, doi:10.1002/joc.943, 2003.

Ganguli, P., and Merz, B. Extreme Coastal Water Levels Exacerbate Fluvial Flood Hazards in Northwestern Europe. Scientific Reports, 9(1), 1-14, doi:10.1038/s41598-019-49822-6, 2019.

Gao, C., Guan, X.J., Booij, M.J., Meng, Y. and Xu, Y.P. A new framework for a multi-site stochastic daily rainfall model: Coupling a univariate Markov chain model with a multi-site rainfall event model. Journal of Hydrology, 598, doi:10.1016/j.jhydrol.2021.126478, 2021.

Gauch, M. and Klotz, D. and Kratzert, F. and Nearing, S. and Hochreiter, S. and Lin, J. A Machine Learner's Guide to Streamflow Prediction. Workshop on AI for Earth Sciences 34th Conference on Neural Information Processing Systems (NeurIPS 2020) Vancouver, Canada, 2020.

Gvoždíková, B., Müller, M. and Kašpar, M. Spatial patterns and time distribution of central European extreme precipitation events between 1961 and 2013. International Journal of Climatology, 39(7), 3282-3297, doi:10.1002/joc.6019, 2019.

Haberlandt, U., Hundecha, Y., Pahlow, M., Schumann, A.H. Rainfall generators for application in flood studies. In: Schumann, A.H. (Ed.), Flood Risk Assessment and Management. Springer, Netherlands, pp. 117-147, doi:10.1007/978-90-481-9917-4_7, 2011.

Harris, C.N.P., Quinn, A.D. and Bridgeman, J. The use of probabilistic weather generator information for climate change adaptation in the UK water sector. Meteorological Applications, 21(2), 129-140, doi:10.1002/met.1335, 2014.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049, doi:10.1002/qj.3803, 2020.

Hu, G. and Franzke, C.L.E. Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany. Geophysical Research Letters, 47(18), e2020GL089624, doi:10.1029/2020GL089624, 2020.

Hundecha, Y., Pahlow, M. and Schumann, A. Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water Resources Research, 45, doi:10.1029/2008wr007453, 2009.

Jeferson de Medeiros, F., Prestrelo de Oliveira, C. and Avila-Diaz, A. Evaluation of extreme precipitation climate indices and their projected changes for Brazil: From CMIP3 to CMIP6. Weather and Climate Extremes, 38, 100511, doi:10.1016/j.wace.2022.100511, 2022.

Kiem, A. S., Kuczera, G., Kozarovski, P., Zhang, L., & Willgoose, G. (2021). Stochastic generation of future hydroclimate using temperature as a climate change covariate. *Water Resources Research*, *57*(2). https://doi.org/10.1029/2020WR027331

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

Kreibich, H., Petrow, T., Thieken, A. H., Müller, M., and Merz, B. Consequences of the extreme flood event of August 2002 in the city of Dresden, Germany, in: Sustainable Water Management Solutions for Large Cities, edited by: Savic, D. A., Mariño, M. A., Savenije, H. H. G., Bertoni, J. C., IAHS Publication, 293, 164–173, 2005.

Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C. J. H., Apel, H., Aronica, G. T., . . . Merz, B. Adaptation to flood risk: Results of international paired flood event studies. Earth's Future, 5(10), 953-965, doi:10.1002/2017ef000606, 2017.

Lenderink, G. and Fowler, H.J. Understanding rainfall extremes. Nature Climate Change, 7(6), 391-393, doi:10.1038/nclimate3305, 2017.

Lengfeld, K., Walawender, E., Winterrath, T. and Becker, A. CatRaRE: A Catalogue of radar-based heavy rainfall events in Germany derived from 20 years of data. Meteorologische Zeitschrift, 30(6), 469 - 487, doi:10.1127/metz/2021/1088, 2021.

Lengfeld, K., Winterrath, T., Junghänel, T., Hafer, M. and Becker, A. Characteristic spatial extent of hourly and daily precipitation events in Germany derived from 16 years of radar data. Meteorologische Zeitschrift, 28(5), 363-378, doi:10.1127/metz/2019/0964, 2019.

Makkonen, L. Plotting Positions in Extreme Value Analysis. Journal of Applied Meteorology and Climatology, 45(2), 334-340, doi:10.1175/JAM2349.1, 2006.

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themel, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M. and Thiele-Eich, I. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics, 48(3), doi:10.1029/2009RG000314, 2010.

Matte, D., Christensen, J.H. and Ozturk, T. Spatial extent of precipitation events: when big is getting bigger. Climate Dynamics, 58(5), 1861-1875, doi:10.1007/s00382-021-05998-0, 2022.

Mehrotra, R. and Sharma, A. Development and Application of a Multisite Rainfall Stochastic Downscaling Framework for Climate Change Impact Assessment. Water Resources Research, 46(7), doi:10.1029/2009WR008423, 2010.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., . . . Macdonald, E. Causes, impacts and patterns of disastrous river floods. Nature Reviews Earth & Environment, 2(9), 592-609, doi:10.1038/s43017-021-00195-3, 2021.

Merz, B., Nguyen, V.D., Guse, B., Han, L., Guan, X., Rakovec, O., Samaniego, L., Ahrens, B. and Vorogushyn, S. Spatial counterfactuals to explore disastrous flooding. Environmental Research Letters, 19(4), 044022, doi:10.1088/1748-9326/ad22b9, 2024.

Minářová, J., Müller, M., Clappier, A. and Kašpar, M. Comparison of extreme precipitation characteristics between the Ore Mountains and the Vosges Mountains (Europe). Theoretical and Applied Climatology, 133(3), 1249-1268, doi:10.1007/s00704-017-2247-x, 2018.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J.E., Ehmele, F., Feldmann, H., Franca, M.J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J.G., Quinting, J., Schäfer, A.M., Scheibel, M., Seidel, F. and Wisotzky, C. A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe - Part 1: Event description and analysis. Natural Hazards and Earth System Sciences, 23(2), 525-551, doi:10.5194/nhess-23-525-2023, 2023.

Müller, M. and Kaspar, M. Event-adjusted evaluation of weather and climate extremes. Natural Hazards and Earth System Sciences, 14(2), 473-483, doi:10.5194/nhess-14-473-2014, 2014.

Najibi, N., Perez, A. J., Arnold, W., Schwarz, A., Maendly, R., & Steinschneider, S. (2024). A statewide, weather-regime based stochastic weather generator for process-based bottom-up climate risk assessments in California – Part II: Thermodynamic and dynamic climate change scenarios. *Climate Services*, *34*, 100485. https://doi.org/10.1016/j.cliser.2024.100485

NatCatSERVICE. Global Natural Catastrophe and Socioeconomic Impact Database. Munich Re. https://www.munichre.com/en/reinsurance/business/non-life/natcatservice/index.html, 2023

Naveau, P., Huser, R., Ribereau, P. and Hannart, A. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resources Research, 52(4), 2753-2769, doi:10.1002/2015WR018552, 2016.

Nguyen, V.D., Merz, B., Hundecha, Y., Haberlandt, U. and Vorogushyn, S. Comprehensive evaluation of an improved large-scale multi-site weather generator for Germany. International Journal of Climatology, 41(10), 4933-4956, doi:10.1002/joc.7107, 2021.

Nguyen, V. D., Vorogushyn, S., Nissen, K., Brunner, L., & Merz, B. (2024). A non-stationary climate-informed weather generator for assessing future flood risks. *Advances in Statistical Climatology, Meteorology and Oceanography*, *10*(2), 195–216. https://doi.org/10.5194/ascmo-10-195-2024

Orlanski, I. A Rational Subdivision of Scales for Atmospheric Processes. Bulletin of the American Meteorological Society, 56(5), 527-530, 1975.

Philipp, A., Della-Marta, P.M., Jacobeit, J., Fereday, D.R., Jones, P.D., Moberg, A. and Wanner, H. Long-Term Variability of Daily North Atlantic–European Pressure Patterns since 1850 Classified by Simulated Annealing Clustering. Journal of Climate, 20(16), 4065-4095, doi:10.1175/jcli4175.1, 2007.

Qin, X. and Lu, Y. Study of Climate Change Impact on Flood Frequencies: A Combined Weather Generator and Hydrological Modeling Approach. Journal of Hydrometeorology, 15(3), 1205-1219, doi:10.1175/JHM-D-13-0126.1, 2014.

Ramos, A.M., Trigo, R.M. and Liberato, M.L.R. Ranking of multi-day extreme precipitation events over the Iberian Peninsula. International Journal of Climatology, 37(2), 607-620, doi:10.1002/joc.4726, 2017.

Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V.D., Lüdtke, S., Merz, B., Schröter, K., Vorogushyn, S. and Kreibich, H. Process-Based Flood Risk Assessment for Germany. Earth's Future, 9(10), e2021EF002259, doi:10.1029/2021EF002259, 2021.

Serinaldi, F. and Kilsby, C.G. Simulating daily rainfall fields over large areas for collective risk estimation. Journal of Hydrology, 512, 285-302, doi:10.1016/j.jhydrol.2014.02.043, 2014.

Voit, P. and Heistermann, M. A downward-counterfactual analysis of flash floods in Germany. Nat. Hazards Earth Syst. Sci., 24(6), 2147-2164, doi:10.5194/nhess-24-2147-2024, 2024.

Vorogushyn, S., Han, L., Apel, H., Nguyen, V.D., Guse, B., Guan, X., Rakovec, O., Najafi, H., Samaniego, L., Merz, B. It could have been much worse: spatial counterfactuals of the July 2021 flood in the Ahr valley, Germany. Natural Hazards and Earth System Sciences Discussions. [preprint], doi:10.5194/nhess-2024-97, 2024, under review.

Steinschneider, S. and Brown, C. A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments. Water Resources Research, 49(11), 7205-7220, doi:10.1002/wrcr.20528, 2013.

Szönyi M., Roezer V., Deubelli T., Ulrich J., MacClune K., Laurien F., & Norton, R. PERC Flood event review ‘Bernd.’ https://floodresilience.net/zurich-flood-resilience-alliance/, 2022

Thieken, A.H., Mohor, G.S., Kreibich, H. and Müller, M. Compound inland flood events: different pathways, different impacts and different coping options. Natural Hazards and Earth System Sciences, 22(1), 165-185, doi:10.5194/nhess-22-165-2022, 2022.

Tseng, S.C., Chen, C.J. and Senarath, S.U.S. Evaluation of multi-site precipitation generators across scales. International Journal of Climatology, 40(10), 4638-4656, doi:10.1002/joc.6480, 2020.

Ullrich, S.L., Hegnauer, M., Nguyen, D.V., Merz, B., Kwadijk, J. and Vorogushyn, S. Comparative evaluation of two types of stochastic weather generators for synthetic precipitation in the Rhine basin. Journal of Hydrology, 601, 126544, doi:10.1016/j.jhydrol.2021.126544, 2021.

Ulrich, J., Jurado, O.E., Peter, M., Scheibel, M. and Rust, H.W. Estimating IDF Curves Consistently over Durations with Spatial Covariates. Water, 12(11), 3119, doi:10.3390/w12113119, 2020.

Ulrich, J., Ritschel, C., Mack, L., Jurado, O. E., Fauer, F. S., Detring, C., and Joedicke, S.: IDF: Estimation and Plotting of IDF Curves, [code], available at: https://CRAN.R-project.org/package=IDF (last access: 22 July 2022), R package version 2.1.2, 2021

Voit, P. and Heistermann, M. A new index to quantify the extremeness of precipitation across scales. Natural Hazards and Earth System Sciences, 22(8), 2791-2805, doi:10.5194/nhess-22-2791-2022, 2022.

Wilks, D.S. Multisite generalization of a daily stochastic precipitation generation model. Journal of Hydrology, 210(1), 178-191, doi:10.1016/S0022-1694(98)00186-3, 1998.

Winter, B., Schneeberger, K., Dung, N.V., Huttenlau, M., Achleitner, S., Stötter, J., Merz, B. and Vorogushyn, S. A continuous modelling approach for design flood estimation on sub-daily time scale. Hydrological Sciences Journal, 64(5), 539-554. doi: 10.1080/02626667.2019.1593419, 2019.

Woo, G. Downward Counterfactual Search for Extreme Events. Frontiers in Earth Science, 7. doi: 10.3389/feart.2019.00340, 2019.

Yang, L., Franzke, C.L.E. and Duan, W. Evaluation and projections of extreme precipitation using a spatial extremes framework. International Journal of Climatology, 43(7), 3453-3475. https://doi.org/10.1002/joc.8038, 2023.

Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B. and Zwiers, F.W. Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Climate Change, 2(6), 851-870, doi:10.1002/wcc.147, 2011.

Zhou, L., Meng, Y., Lu, C., Yin, S. and Ren, D. A frequency-domain nonstationary multi-site rainfall generator for use in hydrological impact assessment. Journal of Hydrology, 585, doi:10.1016/j.jhydrol.2020.124770, 2020.