



A downward-counterfactual analysis of flash floods in Germany

Paul Voit and Maik Heistermann

Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Correspondence: Paul Voit (voit@uni-potsdam.de)

Received: 18 December 2023 – Discussion started: 2 January 2024

Revised: 19 March 2024 – Accepted: 30 April 2024 – Published: 27 June 2024

Abstract. Counterfactuals are scenarios that describe alternative ways of how an event, in this case an extreme rainfall event, could have unfolded. In this study, we present the results of a counterfactual search for flash flood events in Germany. We used a radar-based precipitation dataset from Germany's national meteorological service (Deutscher Wetterdienst) to identify the 10 most extreme precipitation events in Germany from 2001 to 2022 and then assumed that any of these top 10 events could have happened anywhere in Germany. In other words, the events were shifted around all over Germany. For all resulting positions of the precipitation fields, we simulated the corresponding peak discharge for any affected catchment smaller than 750 km². From all the realizations of this simulation experiment, the maximum peak discharge was identified for each catchment.

In a case study, we first focused on the devastating flood event in July 2021 in western Germany. We found that a moderate shifting of the event in space could change the event peak flow at the Altenahr gauge by a factor of 2. Compared to the peak flow of 1004 m³ s⁻¹ caused by the event in its original position, the worst-case counterfactual of that event led to a peak flow of 1311 m³ s⁻¹. Shifting another event that had occurred just 1 month earlier in eastern Germany over the Ahr River valley even effectuated a simulated peak flow of 1651 m³ s⁻¹.

For all analysed subbasins in Germany, we found that, on average, the highest counterfactual peak exceeded the maximum original peak (between 2001 and 2022) by a factor of 5.3. For 98 % of the basins, the factor was higher than 2.

We discuss various limitations of our analysis, which are important to be aware of, namely, the quantification and selection of candidate rainfall events, the hydrological model, and the design of the counterfactual search experiment. Still, we think that these results might help to expand the view of what could happen in the case that certain extreme events oc-

curred elsewhere and thereby reduce the element of surprise in disaster risk management.

1 Introduction

Flash floods constitute a relevant natural hazard in many regions of the world. In comparison to river floods, the footprint of a flash flood event is small, yet the local impact can be devastating. Flash floods combine low predictability, erratic overflow behaviour, high flow velocities, and often massive debris loads. They are mainly caused by heavy precipitation events (HPEs) with very high rainfall intensities and characterized by a rapid concentration of runoff. Usually, flash floods are defined by a response time of less than 6 h (Borga et al., 2008; Marchi et al., 2010), which mostly confines their occurrence to catchments smaller than 1000 km². The underlying HPEs often are highly variable in space and time (Borga et al., 2008). In addition to the properties of the HPE itself, the geographical context governs the nature of the hydrological response and thus the resulting impact. Hence, both atmospheric and hydrological processes interact across various spatial and temporal scales during flash floods (Georgakakos, 1986).

The management of flash flood risks often requires corresponding extreme value statistics. The robustness of such statistics is contingent upon the length of historical records (Woo, 2019) and might be compromised by the effects of ongoing climate change. Locally, flash floods are rare events; observational data are scarce as the affected catchments are typically small and ungauged (Gaume et al., 2008). This makes it difficult to establish reliable extreme value statistics for many locations. Worst-case flood scenarios and their dependence on spatio-temporal characteristics of precipitation as well as the catchment's hydrological conditions have not

yet been fully understood (Zischg et al., 2018; Marchi et al., 2010). Spatio-temporal patterns of rainfall and their dynamic interaction with topography and land use significantly influence the generation and propagation of flood peaks (Beven and Hornberger, 1982; Singh, 1997; Tarolli et al., 2013; Emmanuel et al., 2015; Zischg et al., 2018). This implies that even slight changes in event realizations could significantly affect the response. Yet, the sample size of the investigated HPEs is often limited.

To enhance our understanding of the flash flood hazard in Germany, we adopt an approach known as “counterfactual thinking” (Roese, 1997; Woo, 2019) which was also proposed recently by Montanari et al. (2023) in the context of flood research. This approach involves considering alternative ways of how events *could* have unfolded. For risk assessment, downward counterfactuals are particularly interesting: they involve thought experiments about past events with outcomes worse than what actually transpired (Roese, 1997). Such thought experiments can provide valuable insights into worst-case scenarios that have not (yet) occurred. This way the level of preparedness could be increased, although the approach typically cannot underpin such worst-case scenarios with occurrence probabilities.

Spatial changes, in particular, play a significant role in counterfactual analysis (Woo, 2019): the coincidence of an HPE with an area characterized by steep slopes, impervious surfaces, and multiple stream intersections can trigger very high flood peaks, which would be less pronounced in less steep and more natural catchments.

Based on 16 years of radar observations, Lengfeld et al. (2019) found that extreme daily precipitation is dependent on the orography but that heavy hourly rainfall can occur anywhere in Germany. Based on the – admittedly strong – assumption that historical HPEs could have happened anywhere in Germany, we propose, in this study, a systematic downward-counterfactual search for flash floods in Germany. To that end, we adopted the following approach.

1. Based on radar-based precipitation estimates from 2001–2022, we created a catalogue of HPEs in Germany and ranked these HPEs using a recently proposed metric to assess the extremity of rainfall across spatial and temporal scales (Voit and Heistermann, 2022).
2. We shifted the 10 most extreme HPEs from our catalogue to each subbasin in Germany and simulated the corresponding quick runoff (QR) response for the whole affected area. This way we created a total of 23 000 counterfactual scenarios for each HPE. Each of these scenarios includes the QR simulations for hundreds of subbasins.
3. Additionally we model, for each subbasin, the QR response to *all* events contained in our catalogue, in their original position. The corresponding results serve as a reference for the maximum historical QR response in

each subbasin, to which we compare the results of the counterfactual search.

Based on this groundwork, we first investigate, in a regional case study, counterfactual scenarios of the devastating July 2021 precipitation event over the Ahr River catchment (see Mohr et al., 2023, for details). We then expand our analysis to all of Germany, explore the potential hydrological response to rare HPEs in case they had happened anywhere in Germany, and search for downward-counterfactual scenarios. Based on this search, we try to answer how close actual historical events (within the last 22 years) have already been to the worst-case scenario and discuss the usefulness of this information for flood risk management.

2 Data

In this section, we will describe the data that were used for the extraction of HPEs as well as the data sources for our hydrological model. The overall study area is Germany. We will also present a case study in which we focus on the catchment of the Ahr River down to the runoff gauge at Altenahr. In our hydrological model, this catchment consists of 37 subbasins (details of this case study are presented in Sect. 4.2). Both the overall study area and the case study area are illustrated in Fig. 1.

2.1 Precipitation data

To allow for a detailed representation of the spatio-temporal variability in rainfall, we used the radar climatology product (Radarklimatologie, RADKLIM v2017.002) provided by Germany’s national meteorological service (Deutscher Wetterdienst; DWD hereafter) between 2001 and 2022. RADKLIM is a reprocessed version of the operational radar-based quantitative precipitation estimation (QPE) product (Radar-Online-Aneichung, RADOLAN; see Winterrath et al., 2012) of the DWD since 2001. To minimize the occurrence of artefacts (Lengfeld et al., 2019) and to allow for heavy rainfall analysis (Kreklow et al., 2019), the radar data are adjusted by additional rainfall data from gauges (hourly and daily), a homogeneous set of algorithms, and advanced climatological corrections (Winterrath et al., 2018b). RADKLIM represents the Germany-wide hourly precipitation at a resolution of 1 km × 1 km. Some parts of Germany (far north, south, and east) have not been covered by radar since 2001, but overall data coverage over Germany is good, with less than 10 % missing hours in most areas (Lengfeld et al., 2019). The RADKLIM dataset is available on the DWD open-data server (Winterrath et al., 2018a). We would like to emphasize the importance of using radar-based precipitation products when dealing with flash floods: compared to rain gauge interpolation, the error in radar-based products has been shown to be considerably smaller (Journée et al., 2023). Zoccatelli et al. (2010) also showed that the errors in rain gauge interpola-

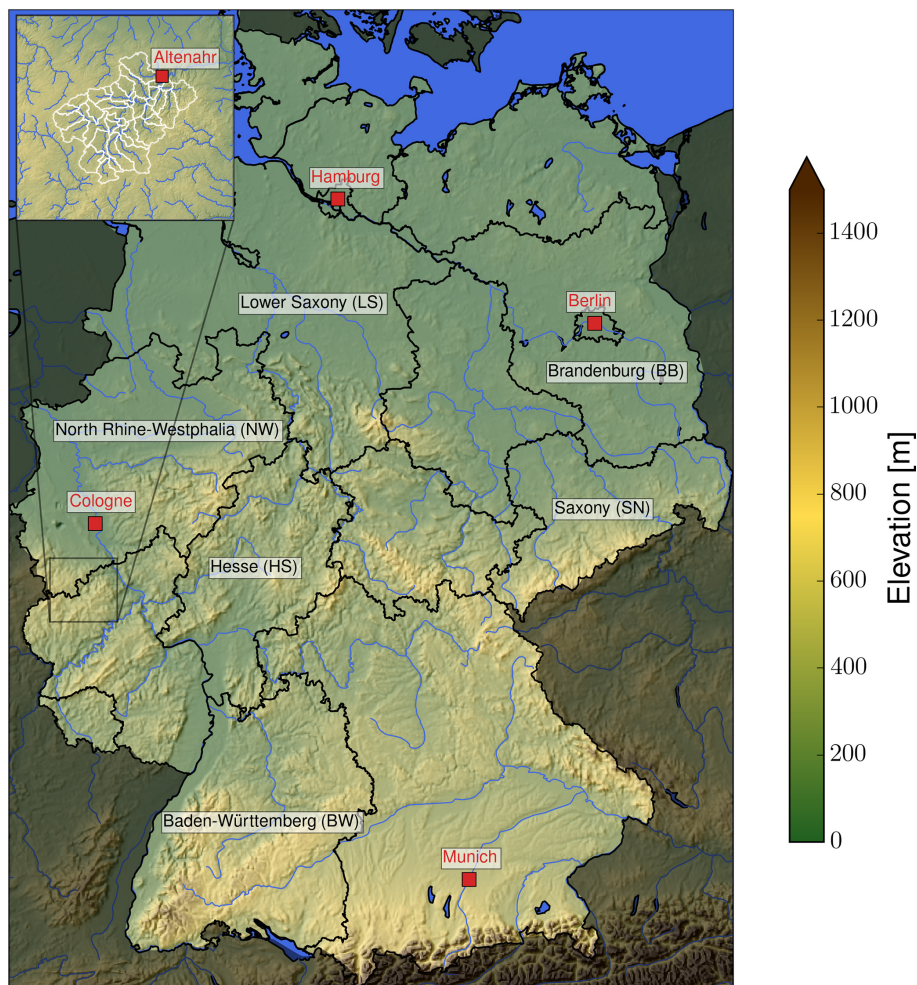


Figure 1. Map of the study region (Germany): topography, major waterbodies (blue), federal states (black), selected cities (red), and subbasins of the Ahr catchment upstream of Altenahr (white; case study region; see Sect. 4.2).

tions for flash-flood-triggering HPEs do not average out at the spatial scales associated with flash floods.

2.2 Digital elevation model

The Digital Elevation Model over Europe (EU-DEM) was used to delineate catchments in Germany and for further analysis of runoff concentration (flow paths and travel time to catchment outlets). For the EU-DEM, SRTM (Shuttle Radar Topography Mission) and ASTER GDEM (Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model) data are fused by a weighted-averaging approach. The dataset has a spatial resolution of 25 m and can be downloaded from the Copernicus Land Monitoring Service (European Commission, 2016).

2.3 Land cover

Information about land cover was derived from CORINE (Coordination of Information on the Environment) CLC5-

2018 (BKG, 2018). The product is based on a classification of high-resolution satellite data into 37 land cover classes (for Germany), according to the nomenclature of the European Environment Agency (EEA). Objects with a minimum size of 5 ha are considered in the classification, and the product is updated every 3 years.

2.4 Soil data

Soil information was derived from the BUEK 200 database (Bodenübersichtskarte, national soil survey at a scale of 1 : 200 000; BGR, 2018), which is compiled from the surveys of each federal state at a scale of 1 : 200 000 by the Federal Institute for Geosciences and Natural Resources (Bundesanstalt für Geowissenschaften und Rohstoffe, BGR) in cooperation with the State Geological Surveys (Staatliche Geologische Dienste, SGD). For each mapping unit, the BUEK 200 database provides areal fractions of dominant soil

types and the corresponding profile information, including texture, bulk density, and much more.

3 Methods

This section describes the methods used to create a catalogue of HPEs, an outline of the hydrological model to model the formation and concentration of quick runoff, and the design of the counterfactual simulation experiment.

3.1 Catalogue of heavy rainfall events in Germany

While the DWD provides a catalogue of HPEs (CatRaRE, Catalogue of Radar-based Heavy Rainfall Events; Lengfeld et al., 2021), we still opted to develop our own catalogue. This decision was motivated by the fact that HPEs which exhibit extreme behaviour across various durations and spatial scales can trigger different flood mechanisms that can intersect and amplify each other. For instance, high-intensity rainfall on a small spatial scale may be embedded within larger events and preceded by periods of low-intensity rainfall that increase soil moisture. Antecedent soil moisture has a significant impact on event runoff coefficients and is essential for flash flood modelling (Marchi et al., 2010). To that end, Voit and Heistermann (2022) have recently proposed a new metric, the cross-scale weather extremity index (xWEI), to detect and assess HPEs that were extreme at various spatial and temporal scales. Both the WEI (as used by CatRaRE) and the xWEI metrics quantify a measure of extremeness along two dimensions: rainfall duration and spatial extent. Hence the variation in extremeness along these dimensions could be illustrated as a surface. While the WEI metric corresponds to the maximum value of that surface, the xWEI metric corresponds to the volume under the surface, meaning that it is high if the extremeness is high across spatial and temporal scales.

The catalogue was created by applying a multi-step procedure. Considering the RADKLIM dataset as a 3-D array (one temporal dimension, two spatial dimensions), we first apply a moving 3-D window (72 h \times 3 km \times 3 km) to the entire dataset. Within this moving window, the rainfall extremeness is computed for each voxel and for various durations. Afterwards, a clustering algorithm is applied to identify spatio-temporal clusters of extreme rainfall. The details of this approach together with an illustration are provided in Appendix A (Fig. A1). The resulting catalogue contains 17 302 events.

3.2 Modelling quick runoff

We used standard GIS (geographic information system) techniques (sink filling, flow accumulation, flow direction, and catchment delineation) implemented in the Python package `PCRaster` (Karssenberget al., 2010) to derive the subbasins. Since our model requires the areal-average precipi-

tation per subbasin as input, the subbasins need to be sufficiently small to represent the effects of spatial rainfall variability on the formation and concentration of quick runoff. For that purpose, we set outlet points for the subbasins at every stream intersection with a Strahler order of 7 or larger. This way we divided the study area into 22 384 subbasins. For the analysis we restricted our modelling to a spatial scale of up to 750 km², which leads to 19 809 remaining basins. The median basin size is 12 km² (25th percentile: 6.9 km², 75th percentile: 20.2 km²). Figure B1 (Appendix B) illustrates the distribution of subbasin sizes as a histogram.

In the case study (Sect. 4.2) we focused on the catchment of Altenahr (Rhineland-Palatinate) as a study region (see Fig. 1). The city of Altenahr was heavily affected by the event named Bernd in July 2021 in western Germany and hit by a flood on 15 July 2021 that caused massive destruction. The catchment upstream of Altenahr, before the inflow of the Vischelbach, has an approximate size of 728.6 km² and is, in our model, split into 37 subbasins. The smallest subbasin has a size of 3 km², the largest is 48 km², and the median size is 17.1 km². The average curve number for the whole catchment is 66 (see Sect. 3.2.1), varying between 61–72 for the individual subbasins (all values for medium soil moisture, soil moisture class 2).

Flash floods are characterized by quick (surface or near-surface) runoff components (Georgakakos, 1986; Marchi et al., 2010; Grimaldi et al., 2010; Borga et al., 2014). Thus, the hydrological model setup can be simplified, as processes like evaporation and groundwater dynamics have a minimal impact on the peak formation. While the formation of quick runoff is mostly controlled by soil conditions and land use, the concentration of quick runoff is primarily driven by topographic relief (Ruiz-Villanueva et al., 2012). Based on these considerations, we adopt the following hypotheses for our model.

- Flash floods peaks are dominated by quick runoff (Marchi et al., 2010; Borga et al., 2014).
- The morphology and topography of the catchment exert the main control on the concentration of quick runoff.
- Flash floods occur predominantly in small to medium-sized catchments with an area smaller than 750 km².
- Evapotranspiration and baseflow dynamics are negligible.
- The objective of the model is not to accurately simulate discharge dynamics. Instead, our focus is primarily on the timing and magnitude of the quick runoff peak flow (QR) and making relative comparisons between different counterfactuals and original events.
- Due to the lack of accurate streamflow data (Gaume et al., 2004; Borga et al., 2014) and the computational

effort to model a large number of counterfactual scenarios, we cannot use a model that requires parameter calibration.

To this end, our model consists of only two components which are described in more detail in the following subsections below.

1. The curve number (CN) method (U.S. Department of Agriculture-Soil Conservation Service, 1972; Natural Resources Conservation Service, 2004; Garen and Moore, 2005) calculates the effective rainfall based on land use, soil characteristics, and antecedent rainfall.
2. The geomorphological instantaneous unit hydrograph (GIUH) method represents the concentration of quick runoff for each subbasin. By superimposing these hydrographs, we can efficiently analyse a large number of counterfactual precipitation scenarios.

With increasing catchment size, the influence of channel mechanics and hydro-engineering on streamflow becomes more important. Due to the limitations of our model, we are unable to incorporate these factors. Consequently, we restrict our QR modelling to subbasins with a spatial scale of up to 750 km². The majority of the 19 809 remaining subbasins are head catchments (13 741) and have an average size of 15 km² and a median size of 11.2 km².

3.2.1 SCS-CN method

We use the established SCS-CN (Soil Conservation Service curve number) method (U.S. Department of Agriculture-Soil Conservation Service, 1972; Ponce and Hawkins, 1996; Natural Resources Conservation Service, 2004) to calculate the effective precipitation depending on soil, land use, and antecedent wetness. For each subbasin, we utilized the BUEK 200 soil database (see Sect. 2.4) to obtain the fractions of four different soil classes (from permeable to non-permeable). This classification was combined with the CORINE CLC5-2018 land use data (see Sect. 2.3). Given that flash flood events primarily occur during summer months (see Sect. 3.3), we made slight adjustments to the CN values for agricultural areas to account for the influence of summer crops (based on Seibert et al., 2020). Ultimately, a single CN value was calculated for each subbasin using a weighted areal average.

Rainfall series for each individual subbasin and event realization were obtained using the zonal-statistics functionality of the Python package `wradlib` (Heistermann et al., 2013), which computes the weighted-average rainfall per subbasin based on the intersection of each RADKLIM pixel with the subbasin. These areal-average rainfall data were then used to calculate the effective rainfall using the SCS-CN method.

3.2.2 GIUH

To route the effective rainfall derived from the SCS-CN method to the subbasin outlet, we utilized the GIUH method. Especially for ungauged basins, this method provides a simple and widely used tool for rainfall-runoff modelling by taking into account the geomorphological features of a basin (Singh et al., 2014; Yi et al., 2022). The GIUH method constructs a hydrograph by estimating the travel time of an instantaneously applied unit of effective rainfall (typically 1 mm) from each grid cell in the catchment to the outlet.

The travel time is determined based on the length of surface flow paths to the outlet and the corresponding flow velocities. Various methods exist to calculate flow velocities. We opted for the spatially distributed travel time model introduced by Maidment et al. (1996) which allows for the use of distributed terrain information in an efficient manner (Bunster et al., 2019). This model demonstrated suitability in a comparative study conducted by Grimaldi et al. (2010). In this method, the flow velocity in a cell is defined as a function of the contributing upstream area A and the local slope s :

$$v = v_m \frac{s^b A^c}{[s^b A^c]_m}, \quad (1)$$

with v as the velocity assigned to a cell with the local slope s and the upstream drainage area A . For b and c , 0.5 has been proven to be a suitable value (Maidment et al., 1996; Grimaldi et al., 2010). v_m describes the average value of the velocity in all cells in the watershed and is set to 0.1 m s⁻¹ based on the study of Grimaldi et al. (2010). $[s^b A^c]_m$ is the watershed average value of the slope-area term. By incorporating the drainage area A into the formula, this method considers the increasing hydraulic radius (Manning's equation) with higher flow volume, thereby capturing the downstream increase in flow velocity without the need to estimate roughness coefficients for individual grid cells. Furthermore, it eliminates the need to differentiate between hill slope and channel grid cells within the catchment. Similarly to previous studies (Sivapalan et al., 2002; Marchi et al., 2010; Creutin et al., 2013), we constrained the resulting velocities within the range of 0.06 to 3 m s⁻¹. By summing the velocities of each grid cell along a flow path, we estimated the travel time for each cell to reach the outlet using the `loddist` function of the Python package `PCRaster` (Karssenberget al., 2010). The hydrograph, representing the QR response of the catchment, is then derived by the probability density function of travel times from all grid cells to the catchment outlet. This method assumes a time- and discharge-invariant velocity field, allowing for a convolution of the GIUHs to model the catchment response to the effective rainfall of an HPE.

In the case that two subcatchments flow together we add the hydrograph (superposition) of the upstream basin to the hydrograph of the downstream basin with a temporal delay. The delay is determined by the travel time from the inlet of the downstream basins to its outlet.

3.3 Design of the downward-counterfactual simulation experiment

For our counterfactual study, we selected the 10 highest-ranking events from our catalogue (Table 1). We then re-located each of these events to each subbasin in Germany. Since the spatial extent of the events is much larger than that of the subbasins, we aligned the pixel with the highest hourly rainfall with the centroid of the corresponding subbasin. We then modelled the QR response for all subbasins within the HPE's bounding box (not just for the subbasin to which we shifted the centroid of the HPE). That way the overall results are not too sensitive to how we actually align an HPE with an individual subbasin. By following this procedure, we generated approximately 230 000 counterfactual QR scenarios across Germany (23 000 subbasins multiplied by 10 HPEs with their centroids shifted across all subbasins). These datasets contain a total of more than 829 million counterfactual QR hydrographs for the individual subbasins, and we refer to them as "cf_germany". Additionally, we filtered the complete cf_germany dataset by limiting the maximum distances over which the HPEs were shifted to 10, 20, 50, and 250 km. We refer to these filtered datasets as cf_10km, cf_20km, cf_50km, and cf_250km.

3.4 Metrics for flash flood response

To compare flood peaks across different basin sizes, we utilized the concept of the unit peak discharge (UPD) (refer to Castellarin, 2007, for a summary of the concept). The UPD ($\text{m}^3 \text{s}^{-1} \text{km}^{-1.2}$) is the ratio between the discharge peak ($\text{m}^3 \text{s}^{-1}$) and the reduced upstream catchment area ($(\text{km}^2)^{0.6}$). To limit the influence of the upstream catchment area, we use an exponent of 0.6 (similarly to Gaume et al., 2008; Emmanuel et al., 2017). Amponsah et al. (2018) used a UPD of $0.5 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$ (which corresponds to $0.66 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-1.2}$) as the lower threshold for the definition of flash floods across a variety of climates and studies in their flash flood catalogue. As an illustration of the unit of the UPD, a UPD of $3 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-1.2}$ could equal an $18 \text{ m}^3 \text{ s}^{-1}$ flood peak in a basin of 20 km^2 size or a peak flow of $72 \text{ m}^3 \text{ s}^{-1}$ in a 200 km^2 basin.

4 Results and discussion

In this section, we present the results of our analysis. Section 4.1 starts by introducing the 10 most severe precipitation events which were identified based on the cross-scale extremity index. By shifting them all over Germany, they form the basis of our spatial counterfactual search experiment. The hydrological simulation results of this experiment are first explored in a case study for the Ahr catchment and put into context of the devastating flood event in July 2021 (Sect. 4.2). Second, we summarize the results of our simulation experiment for all of Germany.

4.1 Top 10 HPEs

In this section, we introduce the 10 most severe precipitation events between 2001 and 2022, based on the DWD's RADKLIM dataset. These events are the basis of our counterfactual simulation experiment.

The 10 most extreme events in our HPE catalogue all occurred during the summer months and are displayed in Fig. 2 and Table 1.

It should be noted that the xWEI metric is sensitive to the spatial extent of an event. Therefore, the top 10 events are generally very large. The catalogue might contain events that are more severe at small spatio-temporal scales, say at the scale of small headwater catchments. The resulting limitations for our analysis will be further discussed in Sect. 5.1. However, events with a large spatial extent and a large xWEI value are likely to include smaller event clusters that are extreme at smaller spatio-temporal scales, which exactly motivated the choice to rank events by the xWEI metric (see also Sect. 3.1). Nonetheless, future applications might choose different catalogues or different metrics and ranking criteria to select candidate events for a counterfactual search.

Very different levels of impacts were reported for these events. In Appendix C, we put each event in the context of other available references (scientific or media) and also attempt to compile estimates of reported damage and loss of life, if available. While all 10 events featured exceptional amounts of rainfall and a corresponding runoff response, only 5 of them caused massive impacts (SN/Aug02; SN/Jun13; BW/May16; BB/Jun17; and, with by far the highest impact, NW/Jul21), while for the remaining events (LS/Jul02, LS/Jul17, HS/May19, BB/Jun20, and BB/Jun21), the impact was apparently not high enough to attract attention beyond the affected regions. The results of the counterfactual scenario analysis, as presented in the following, should help to explain whether the different levels of impacts for these events were mainly caused by their specific geographic position.

4.2 Case study: Altenahr

Before exploring the results for all of Germany, we zoom into the counterfactual scenarios obtained for the Ahr catchment (Fig. 3a). The Ahr was the most severely affected river during the July 2021 floods in western Germany (see Mohr et al., 2023, for more background on the flood event and the Ahr catchment). Typically, a flash flood is characterized by a lag time of up to 6 h between the centroid of the effective rainfall and the hydrograph peak (Borga et al., 2008; Marchi et al., 2010; Morin et al., 2002). So, strictly speaking, the flood event at Altenahr does not qualify as a flash flood: according to our model, the lag time at Altenahr amounted to approximately 8 h. Still, the event at Altenahr is a highly illustrative example for a swift and massive runoff response at the mesoscale which is the result of the temporal superposi-

Table 1. The 10 most extreme HPEs from our catalogue. The ID was constructed from an acronym that specifies the federal state in which the event mainly occurred, the month, and the year (starting from the year 2000). The precipitation (prec.) values in the table (mm) are based on a 10 km × 10 km moving window average; the ranking is based on the xWEI metric.

Rank	ID	Date	xWEI	Max 1 h prec.	Max 24 h prec.	Max 72 h prec.	Location of max 1 h prec.
1	LS/Jul02	15–20 Jul 2002	4148	41	138	154	Plön
2	BB/Jun17	26 Jun–2 Jul 2017	3901	44	149	157	Bautzen
3	SN/Aug02	10–15 Aug 2002	3741	23	224	255	Wittenberg
4	NW/Jul21	11–16 Jul 2021	3542	35	136	150	Dortmund
5	LS/Jul17	22–27 Jul 2017	3327	24	136	233	Göttingen
6	BW/May16	27 May–1 Jun 2016	3304	53	98	106	Erzgebirgskreis
7	BB/Jun21	28 Jun–3 Jul 2021	3235	38	217	219	Uckermark
8	BB/Jun20	11–16 Jun 2020	2964	53	93	104	Ostprignitz-Ruppin
9	HS/May19	18–23 May 2019	2718	26	106	114	Waldeck-Frankenberg
10	SN/Jun13	18–22 Jun 2013	2575	62	121	121	Bautzen

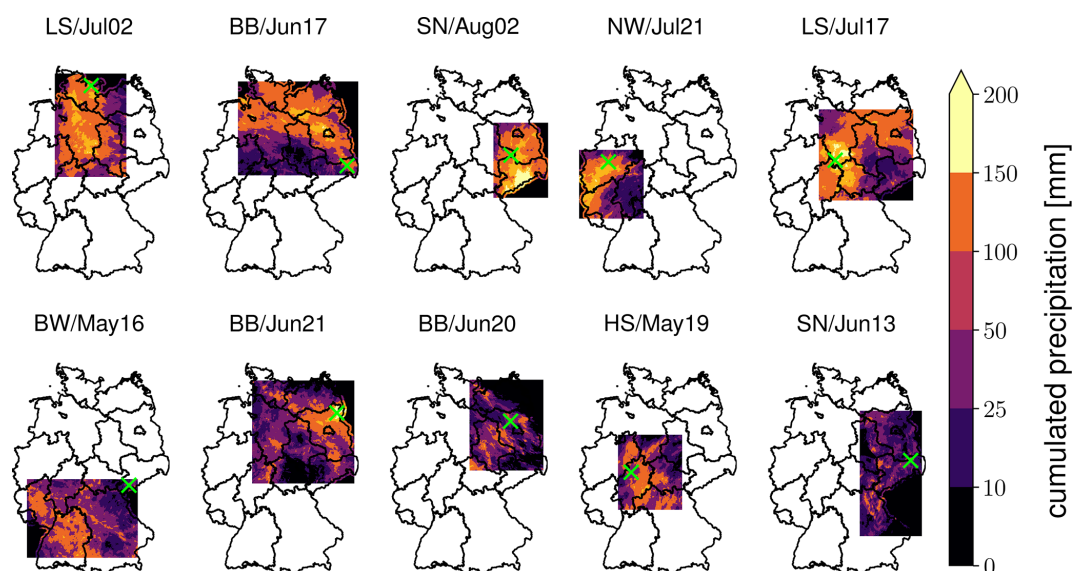


Figure 2. Original position and cumulated precipitation of the 10 most extreme HPEs from the event catalogue. The green cross indicates the location of the highest hourly precipitation during the event which we chose as the centroid when shifting the events to create counterfactuals.

tion of various upstream flash floods. In fact, all 23 subbasins upstream of Altenahr show a lag time of less than 6 h, with 22 of them showing a time lag of even less than 3 h.

By shifting around the top 10 HPEs (as listed in Table 1) over Germany, we created a total of 38 871 counterfactual rainfall scenarios over the Altenahr catchment, representing a large variety of spatial rainfall patterns and average rainfall totals, for all of which we simulated the QR peak flow. In the following, we compare these counterfactual peak flows to the peak simulated for the NW/Jul21 event in its original position. The event label of NW/Jul21_x refers to a spatial counterfactual of the NW/Jul21 event. The same naming convention is adopted for the other events from Table 2.

Figure 4 illustrates the results from the counterfactual study for the Altenahr catchment. The total rainfall for the

catchment for each counterfactual and the resulting highest QR peak is shown. Despite the positive correlation ($r^2 = 0.96$, Fig. 4) between total rainfall and resulting flood peaks, we notice that the same total rainfall amounts can yield markedly different QR peaks.

During the original event (NW/Jul21), the Altenahr catchment received an areal rainfall average of approximately 114 mm, of which 98 mm fell within 12 h on 14 July. The maximum hourly areal average was 12 mm (Fig. 3b). This amount of rainfall results in a modelled QR peak of $1004 \text{ m}^3 \text{ s}^{-1}$. Our model experiment illustrates that, for this specific amount of total areal rainfall ($114 \pm 1 \text{ mm}$), the QR peaks span a range of 536 to $1090 \text{ m}^3 \text{ s}^{-1}$ across all NW/Jul21 counterfactuals (Fig. 4). This signifies that, with

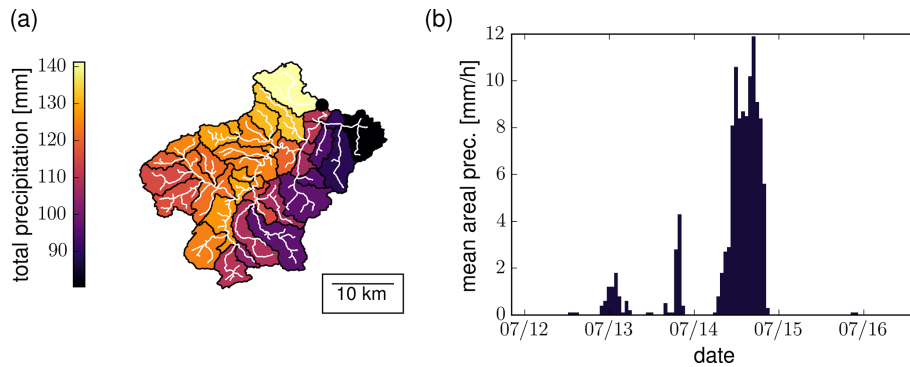


Figure 3. Total rainfall estimates (RADKLIM) for the original NW/Jul21 event for the Altenahr catchment: (a) total rainfall (mm) in the Altenahr subbasins and (b) areal average of precipitation (mm h^{-1}) for the Altenahr catchment. The outlet of the catchment is shown in black, subbasin borders are in black, and streams are in white. Please note that the date format in this and following figures is month/day.

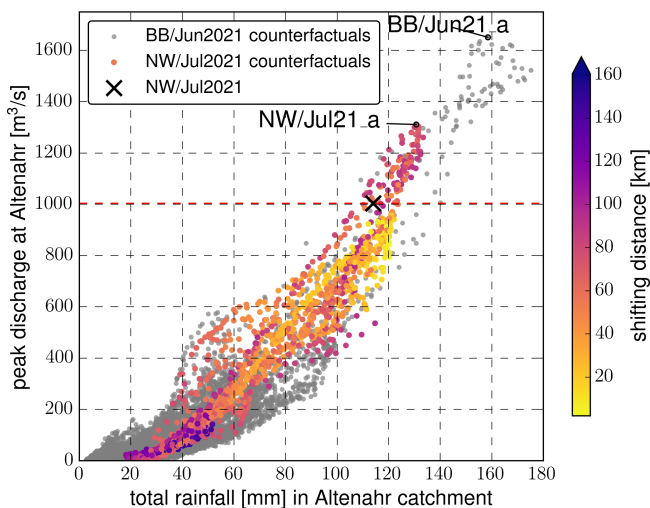


Figure 4. Total rainfall amount and resulting QR peak for counterfactuals of the NW/Jul21 (yellow to blue) and BB/Jun2021 (grey) HPEs for the Altenahr catchment. The black cross represents the areal mean of total rainfall the catchment received during the event, and the resulting runoff for the event is in its original spatial position. The point colour of the NW/Jul21 counterfactuals indicates the distance to the centroid of the original NW/Jul21 event.

an identical total rainfall volume, the QR peak can vary by a factor of 2.

The original event's QR peak is already substantial; however, 6% of the NW/Jul21 counterfactuals would have caused an even higher QR peak. All of these downward counterfactuals were created by a spatial shift in the original event by 45–97 km. The maximum modelled QR is $1311 \text{ m}^3 \text{ s}^{-1}$ (NW/Jul21_a), which is considerably higher than the $1004 \text{ m}^3 \text{ s}^{-1}$ peak resulting from the original event. This outcome would have been achieved if the centroid of NW/Jul21 would have been shifted by only 75 km.

Figure 5a and b illustrate, for the original NW/Jul21 event, the superposition of peaks at the Altenahr gauge from the

discharge of the individual subbasins. The maximum counterfactual rainfall total (130.7 mm for NW/Jul21_a) results in a modelled QR peak of $1311 \text{ m}^3 \text{ s}^{-1}$ (Fig. 5c and d). Altogether, these cases underpin the importance of the spatio-temporal event structure for the peak discharge formation. The mean total precipitation for the whole Altenahr catchment conceals the spatio-temporal distribution of rainfall among its subbasins. In our model, the catchment consists of 37 subbasins (Fig. 3a).

By spatially shifting the other nine HPEs from Table 1 across Germany, we can get an idea of the kind of QR flood peaks that these HPEs could have triggered at Altenahr – had they happened in the region. The BB/Jun21 event is an interesting case: this event happened just 1 month prior to NW/Jul21 in the north-east of Germany (Uckermark). Although rated almost as extreme as the NW/Jul21 event (Table 1), it caused little damage in its original position. However, various spatial positions of this event would have apparently caused even higher QR peaks in Altenahr, up to $1651 \text{ m}^3 \text{ s}^{-1}$ (BB/Jun21_a, Fig. 5e and f). Table 2 displays more information about the three cases shown in Fig. 5. Among all 10 events, the BB/Jun21 counterfactuals lead to the highest modelled QR peaks for the Altenahr gauge.

Out of all counterfactuals, 1% resulted in QR peaks higher than the one from the original event, NW/Jul21. This underlines the rarity of the event. Among these, there are no counterfactuals of the events BW/May16, BB/Jun17, LS/Jul17, HS/May19, and BB/Jun20. Further investigation is needed to understand the differences in the spatio-temporal structure of these events and how these HPEs were different to the other top 10 events to understand why these HPEs did not have the potential to create any maximum counterfactual peaks.

In summary, the analysis of 38 871 QR counterfactuals for the Altenahr catchment has demonstrated that, while the original NW/Jul21 event was exceptional, numerous spatial constellations of the same event and especially of the BB/Jun21 event could have led to higher flood peaks. While the areal-average rainfall total is a key control on peak forma-

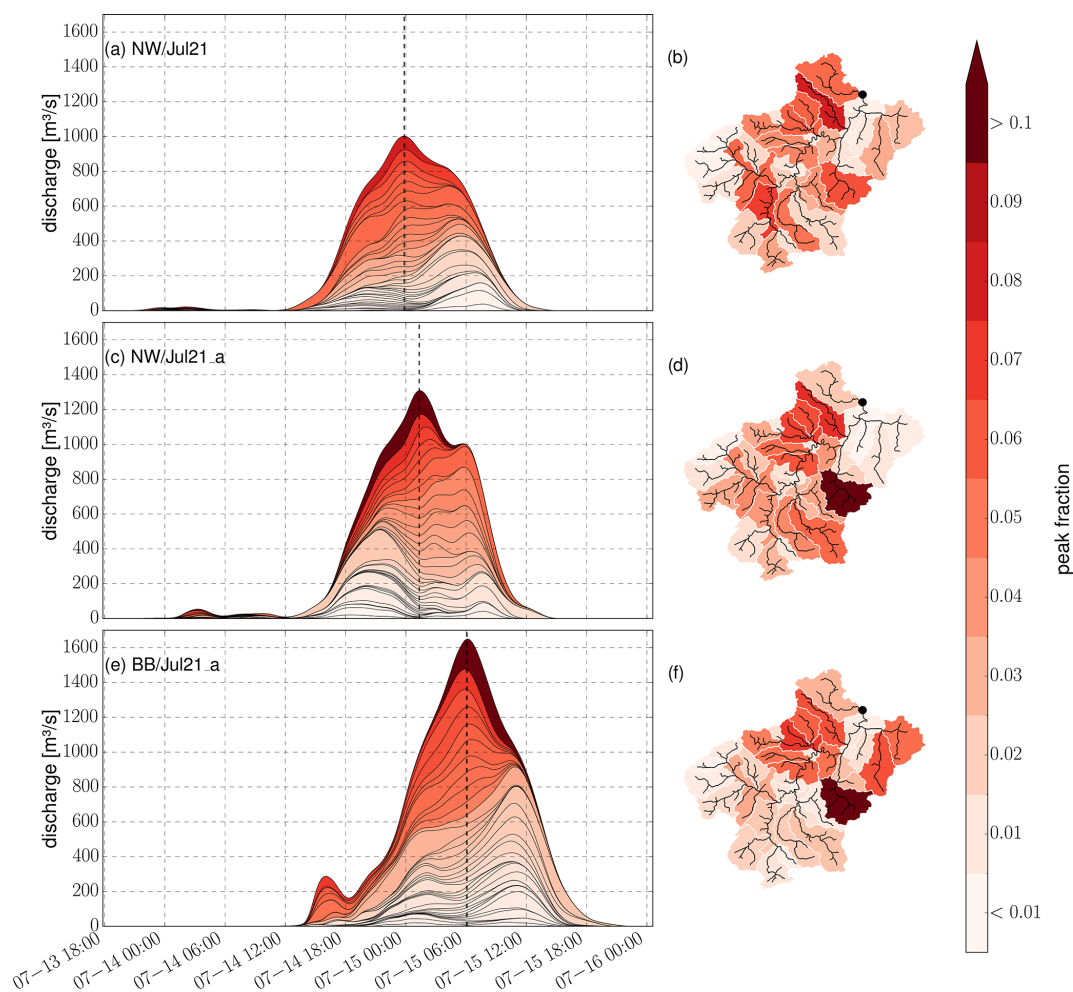


Figure 5. Contributions of individual subbasins to the runoff peak at Altenahr for three scenarios. The left side shows the superposition of runoff from the subbasins. The colour code describes the runoff contribution to the peak flow (white: low, red: high, dotted line: peak position). On the right side, the same colour code is used to display the spatial distribution of the contributions of each subbasin. Streams are shown in black, and the outlet at Altenahr is shown as a black dot. Each row of the plot shows a different precipitation scenario: **(a, b)** original NW/Jul21 event, **(c, d)** NW/Jul21_a counterfactual, and **(e, f)** BB/Jul21_a counterfactual (see also Table 2).

Table 2. Selected counterfactuals for the Altenahr catchment.

ID	QR peak ($\text{m}^3 \text{s}^{-1}$)	Total prec. (mm)	Latitude centroid ($^{\circ}$ N)	Longitude centroid ($^{\circ}$ E)
NW/Jul21	1004	114	50.740	6.965
NW/Jul21_a	1311	131	51.315	7.519
BB/Jul21_a	1651	159	50.437	6.792

tion, the spatio-temporal distribution of this total can moderate flood peak formation substantially.

The discharge and timing of the modelled QR peak for the NW/Jul21 event ($1004 \text{ m}^3 \text{ s}^{-1}$) fits well with recent reconstructions that estimated a peak flow around $1000 \text{ m}^3 \text{ s}^{-1}$ at Altenahr (Mohr et al., 2023). This is surprising given that the RADKLIM product might underestimate the event rainfall (Saadi et al., 2023). In any case, our model confirms that the

NW/Jul21 event triggered a swiftly moving flood wave that by far exceeded the HQ_{100} of $241 \text{ m}^3 \text{ s}^{-1}$ for the Altenahr gauge (Mohr et al., 2023).

4.3 Downward-counterfactual analysis for Germany

In this section we show the results of the downward-counterfactual modelling for all subbasins in Germany. Be-

Table 3. Display of how often events (and their respective downward counterfactuals) caused the highest discharge in a subbasin (column labelled “Count”). The row labelled “Other” describes original events which do not belong to the top 10 events.

Event	Count
SN/Jun13	16 248
BB/Jun21	3033
LS/Jul02	296
SN/Aug02	227
NW/Jul21	2
BB/Jun20	2
Other	1

cause of the large number of individual subbasins, spatial details cannot be shown. However, the results are also illustrated in a web application which allows for zooming into regions of interest (Heistermann and Voit, 2023). Since larger subbasins can generate more runoff than smaller basins, we show the UPD (Sect. 3.4) instead of the absolute peak discharge. On average, there are 41 873 counterfactuals for each subbasin. Figure 6a shows, for each subbasin, the highest UPD derived from original events (2001–2022), while Fig. 6b and c show the maximum UPD and the 99th percentile of all counterfactual scenarios per subbasin.

Looking at historical HPEs and consequent QR peaks that these events triggered, the downward-counterfactual analysis is able to remove the random element of *where* an HPE occurred (Fig. 6b, c). All but one basin showed much higher QR peaks in response to downward-counterfactual events than compared to QR peaks caused by original events (Table 3). Unsurprisingly, the distribution of the UPD in Germany closely follows the topography (Fig. 6b, c, and d). Mountain and low mountain ranges (compare to Fig. 1) display high QR peaks and therefore high UPD in the downward-counterfactual analysis.

For headwater basins, where the QR peak does not depend on the inflow from any upstream basin, the GIUH can give a first idea of a basin’s tendency for quick runoff concentration (Fig. 6d). But contrary to the counterfactual simulations (Fig. 6b and c), this does not give information about potential QR peak flow rates, yet. While GIUHs allow for very efficient hydrological modelling and therefore make a downward-counterfactual analysis possible, they use a uniform precipitation input. As shown in Sect. 4.2 the spatial distribution of rainfall is highly important for the consequent QR peak. For this reason a detailed spatial resolution (a small subbasin size) is desirable to utilize radar rainfall data to its full extent. A small subbasin size consequently leads to a higher number of non-headwater basins whose QR peak characteristics can not be estimated with the GIUH.

Just for one single basin, the highest modelled peak was caused by an original event (which triggered a severe flash flood around Rudolstadt, Thuringia, on 31 May 2008), in

contrast to any counterfactual scenario. For 98 % of the basins, the downward-counterfactual peak would be at least 2 times higher than the highest observed peak in the last 22 years (Fig. 7); for 47 % of the basins, it would be at least 5 times this amount. Figure 7 also shows the corresponding ratios for more “conservative” counterfactual scenarios for which the maximum shifting distance was limited to 10, 20, 50, or 250 km (see Sect. 3.3). For the cf_50km scenario, for instance, 21 % of the discharge peaks from counterfactuals are not higher than the peaks caused by original events. This is due to the fact that a maximum shifting distance of 50 km will leave quite a number of subbasins essentially unaffected by the main footprint of the shifted HPE. Here we need to keep in mind that we only selected 10 out of 17 302 HPEs from the catalogue for the counterfactual search. A better approach for designing such a conservative counterfactual search might be to select, for each subbasin, the most extreme HPE in a specific radius (say 50 km) and then shift this HPE over the corresponding subbasin. But even within the more conservative cf_50km dataset, 51 % of the basins exhibit a ratio of more than 1.5 between the counterfactual and the original peak; more than 30 % have a ratio of more than twice as high as the original peak. Especially in basins which have not yet been affected by severe flash floods in the recent past, the results from the counterfactual analysis could support the preparedness for flood events that might have been unexpected so far, based on observational records.

For the downward-counterfactual study we shifted 10 extreme HPEs across Germany. Additionally, we modelled the runoff that was generated by all the HPEs in our catalogue in their original spatial position. Table 3 shows which events caused the highest discharges for subbasins all across Germany: the counterfactuals of the SN/Jun13 event caused the highest QR peaks in 82 % of the subbasins. Out of the 10 HPEs, this is also the event with the highest hourly precipitation rates (see Table 1). Then again, the BB/Jun21 event also accounts for a substantial proportion of maximum counterfactual peaks, while it only ranks sixth with regard to hourly precipitation levels. Only in two subbasins, the highest QR peaks were caused by NW/Jul21 counterfactuals. In only one case, the worst-case scenario was caused by an original event. While we expect the maximum counterfactual peaks to be governed by the interaction of specific spatio-temporal HPE features and basin properties, the nature of this interaction remains yet to be explained. In other words, it should be subject to future research to better understand which features favour an exceptional runoff response at the flash flood scale. Such research should not be limited to the top 10 events but should aim for a more comprehensive counterfactual search (see Sect. 4.2).

The counterfactual analysis results in a large dataset of potential QR peaks in Germany. Even though these QR peaks might not fully represent all processes involved in discharge generation, they reflect the major runoff processes in small basins and show a range of plausible discharge cases which

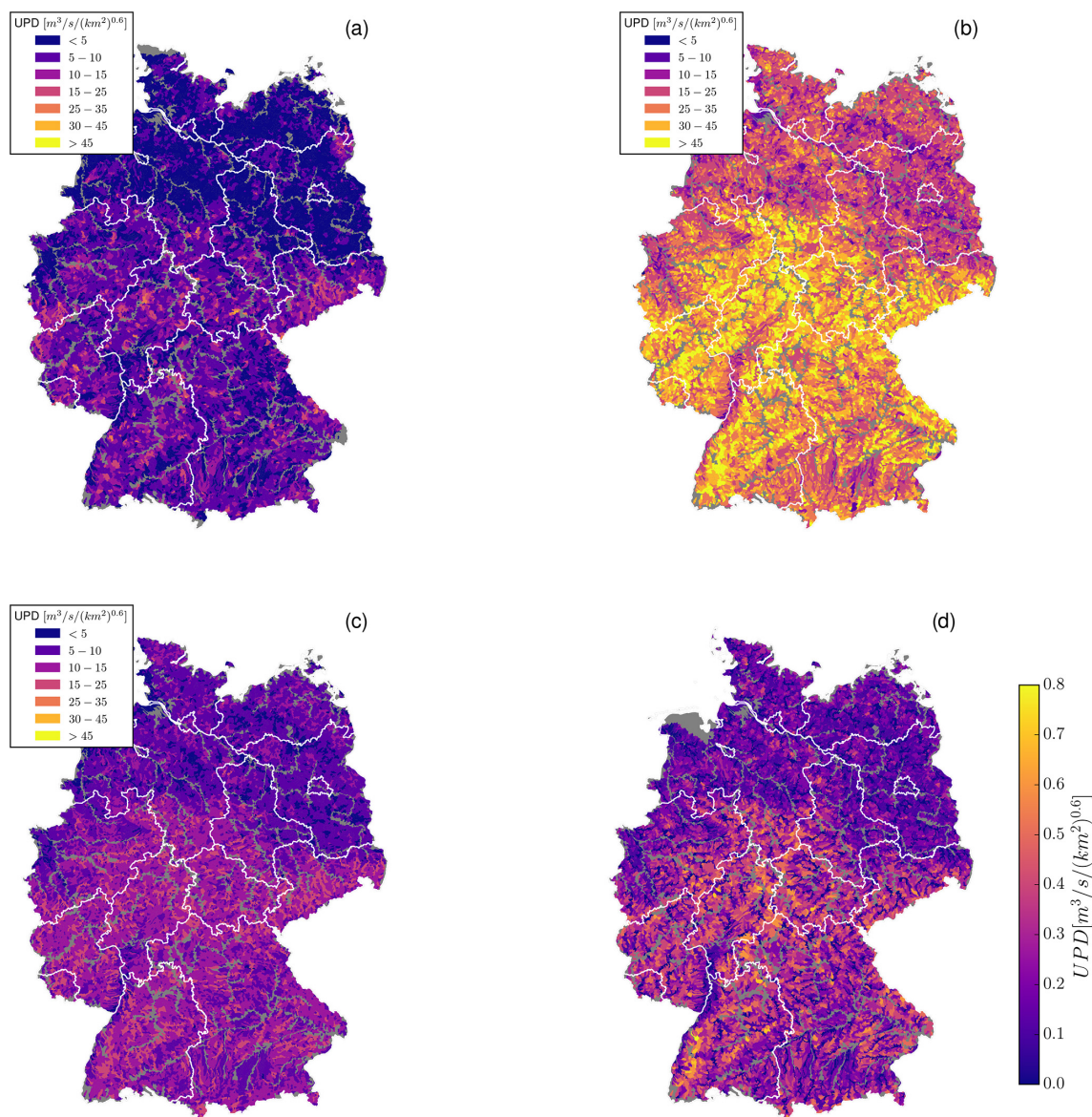


Figure 6. (a) Maximum UPD from original events. (b) Maximum counterfactual UPD. (c) The 99th-percentile UPD derived from downward-counterfactual simulations for Germany. (d) The unit peak discharge derived only from the respective GIUHs. Basins with an area $> 750 \text{ km}^2$ which were not considered in the analysis are shown in grey. Federal-state borders are in white.

can be useful for further analysis. Specifically, the results could be used as a basis to further explore the geographic variation in the flash flood hazard in more detail and to identify subbasins that appear particularly prone to flash floods, mainly as a result of topographical controls.

5 Uncertainties and limitations

In this section, we highlight the uncertainties and limitations that should be kept in mind when interpreting the above results.

5.1 Rainfall data and event catalogue

Journée et al. (2023) showed that errors made by radar-based QPE are smaller than those obtained from rain gauge interpolation. Still, RADKLIM (Winterrath et al., 2018a) might considerably underestimate extreme precipitation. Such underestimation is typically caused by path-integrated attenuation effects (Jacobi and Heistermann, 2016), and it is not too uncommon that these effects are not sufficiently captured and corrected for by the applied rain gauge adjustment methods (see e.g. Saadi et al., 2023, for the NW/Jul21 event or Bronstert et al., 2018, for the BW/May2016 event). Consequently, the resulting peak values of QR might be too low.

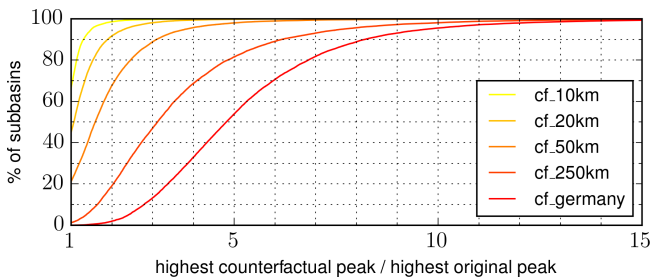


Figure 7. Cumulative distribution of the ratio between the highest counterfactual and the highest original peak for every subbasin is shown in red (cf_germany). From yellow to orange we show the same ratio but for counterfactuals with a limited shifting distance (10, 20, 50, 250 km; see Sect. 3.3). QR peaks resulting from counterfactual simulations are much higher than the QR peaks caused by original events. As the shifting distance increases, more counterfactuals are considered for each subbasin. As a result, it becomes more likely that the counterfactual peaks are substantially higher than the highest original peak.

The same follows from the fact that the rainfall dataset, RADKLIM, is quite short from the perspective of extreme value statistics. While we argue that shifting HPEs across Germany might, to some extent, make up for this shortcoming, we have to prepare for that fact that other events are yet to be observed that might dwarf the top 10 events from our catalogue.

And, finally, the top 10 events from our catalogue might not yet represent the worst case in terms of the QR response at the “flash flood scale”. Particularly for very small headwater catchments, other events from the catalogue could trigger higher runoff peaks even if their xWEI values were smaller. For prospective research, other severity indices, ranking criteria, or catalogues might still be considered or developed which could provide a more explicit focus on flash floods and might hence serve to produce an even more exhaustive counterfactual search.

Then again, the potential underestimation of rainfall also applies to the historical (original) events to which we compare the counterfactual events. Hence, the ratio between the historical and the maximum counterfactual peak flows might be more robust against any rainfall estimation bias – although we need to keep in mind the non-linear transformation of rainfall to runoff (see next section).

Some HPEs, e.g. the SN/Aug2002 or the NW/Jul21 events, are not completely captured by the DWD’s weather radar network, as they extended across the borders of Germany. For these events, the extremeness is necessarily underestimated. We still decided to use these HPEs in our counterfactual simulation experiment because they are, even while being incompletely captured, among the 10 most extreme HPEs observed in Germany within the last 22 years.

For the DWD’s operational radar-based precipitation product (RADOLAN), Saadi et al. (2023) reported an underesti-

mation of 18 % compared to rain gauges for the NW/Jul21 event; Bronstert et al. (2018) found an underestimation of about 30 % for the BW/May2016 event. For the RADKLIM product, the uncertainty is expected to be lower than for the RADOLAN product, e.g. due to the usage of additional data for the rain gauge adjustment. Yet, a systematic assessment of biases in RADKLIM is not yet available. In any case, the level of underestimation is expected to vary dramatically from event to event, as different sources of error govern the overall uncertainty in space and time (Heistermann et al., 2015).

5.2 Hydrological model

The applied hydrological model has, as any model, a number of limitations which we would like to discuss in more detail.

The unit hydrograph method assumes a linear and time-invariant response of a watershed to a spatially homogeneous pulse of effective rainfall (Yi et al., 2022). This assumption is a simplification.

The SCS-CN method implements antecedent soil moisture by considering the total rainfall amount within the last 5 d. Although we added a temporal buffer around our events, we always started the calculations assuming previously dry soils. While the modelled soil moisture class will change as the event unfolds, this assumption decreases runoff generation in the beginning of the event. The worst-case scenario, in terms of QR peaks, would be saturated soils at the beginning of an event.

Since our model does not include baseflow, there is certainly a small fraction of the total runoff missing in the QR peaks. Additionally, we know that the clogging of bridges with uprooted trees and debris can play a major role in the formation of flood peaks (Borga et al., 2014). Our model does not account for such effects, nor does it include a hydrodynamic channel model. Together with the expected underestimation of rainfall (see Sect. 5.1) our results are likely to underestimate discharge peaks.

Utilizing a smaller subbasin size would be advantageous, particularly in the context of investigating flash floods. For example, within our chosen spatial discretization, we were unable to reproduce the extraordinary discharge peak during the Braunsbach flooding in May 2016 (Bronstert et al., 2018), which was generated in a subbasin of 6 km². However, computational cost increases exponentially with spatial resolution, so we have not yet implemented smaller subbasins.

Furthermore, our study relies on an uncalibrated model. The main reason for this is the lack of stream gauge records for small catchments. In addition, stream gauges are often unable to effectively observe extreme flash floods due to being damaged by the actual flood wave (Amponsah et al., 2018). Marchi et al. (2010) showed that only 20 % of flash flood events in small catchments were gauged by a stream gauge section. For these reasons, flash flood events are usually underrepresented in streamflow records (Borga

et al., 2014). However, both model components, the SCS-CN model for QR formation and the GIUH for QR concentration, are widely used, and their applicability was validated in numerous contexts.

Taking all these aspects into account, we would like to emphasize once more that our model is not designed for precise discharge predictions. Instead, it serves as a tool to consistently representing the effects of rainfall, topography, soils, and land use while enabling us to simulate a substantial number of counterfactual scenarios. This large number of simulations is a key feature of this study as it allows for comprehensively exploring possible realizations of counterfactual rainfall events and their effect on peak discharge.

5.3 Spatial shifting of events

In our counterfactual analysis, we assumed that any of the analysed HPEs could have occurred anywhere in Germany. This is a very strong assumption, and it should be emphasized that the validity of this assumption remains an open question. Certainly, an HPE results from the interaction of large- and regional-scale circulation patterns with regional and local features of the earth's surface. For example, orographic effects can augment precipitation and lead to anchoring convection (Marchi et al., 2010; Tarolli et al., 2013). Our study does not consider such effects, which could lead to unrealistic counterfactuals. For this reason we also carried out a more conservative analysis in which we restricted the spatial shifting of HPEs to a radius of 10, 20, 50, and 250 km around their original centroid. These results are displayed in Fig. 7. It would also be very helpful for future research if the atmospheric modelling community further explored how exceptional HPEs could have unfolded under disturbed initial and boundary conditions or under a warmer climate (see e.g. Ludwig et al., 2023, for a pseudo global warming analysis of the July 2021 event) and thereby provide a better basis to evaluate the assumptions behind our counterfactual search.

6 Conclusions

In this study, we presented a downward-counterfactual scenario analysis to assess the flash flood hazard in small to medium-sized basins in Germany. Instead of relying on local observational records of limited length, we identified the most severe heavy precipitation events from 2001 until 2022 and assumed that these events could have occurred anywhere in Germany. The quick runoff response to the resulting counterfactual rainfall scenarios was simulated using a parsimonious and computationally efficient rainfall-runoff model and compared to the quick runoff response of historical events that actually took place in the corresponding subbasins.

Using a radar-based precipitation product, we were able to account in detail for the effects of different spatio-temporal event realizations on the quick runoff response. These effects

can substantially moderate the role of the total accumulation of areal-mean rainfall. This was first demonstrated in a case study of the July 2021 flood event (NW/Jul21) for the Ahr River catchment, down to the Altenahr runoff gauge. Shifting the NW/Jul21 rainfall event in space resulted in a wide range of quick runoff peak values of which 6 % exceeded the response to the original event. Furthermore, shifting another event (BB/Jun21), which had occurred 1 month earlier in eastern Germany, to the Ahr catchment effectuated a peak that exceeded the worst-case downward-counterfactual peak of the NW/Jul21 event by another 26 %.

We then expanded the analysis to all of Germany and found that, on average, the worst-case downward counterfactual exceeded the maximum original quick runoff peak by a factor of 5.3. In general, the quick runoff response is dominated by topography. It turned out that the SN/Jun13 event (see Table 1) caused the maximum counterfactual peak in the majority of basins. Still, readers should be aware of various limitations of our approach, some of which might lead to a considerable underestimation of counterfactual quick runoff peaks.

To make our results easily accessible, we created a web viewer where interested users can explore the results for each subbasin in Germany (Heistermann and Voit, 2023). Still, our results leave various open questions: the most obvious, of course, is about the validity of shifting events all over Germany. Furthermore, focusing on the top 10 events as ranked by the xWEI metric might hide events that were more severe at the flash flood scale. So we should further explore the event catalogue to understand which spatio-temporal structure makes an event particularly hazardous. Besides, it would be interesting to see how the counterfactual peaks compare to the values which are currently used for risk management. Furthermore, we just looked at the worst-case scenario for individual basins. However, large precipitation events can trigger flash floods in multiple basins simultaneously. The identification of regional flash flood clusters caused by one event is relevant in the context of disaster response. It should be clear that our design of counterfactual scenarios only addresses one single aspect: the spatial position of the precipitation field and its effect on the hydrological hazard intensity. A more comprehensive counterfactual search would require accounting for impact-related aspects and processes. Such aspects could e.g. be the daytime or weekday at which an event occurs, the effectiveness of an early warning chain, or cascading effects of damage to infrastructure.

We would like to emphasize that the presented approach should be considered a framework rather than a fixed method with fixed results: users could employ different catalogues, make different assumptions about spatial shifting of heavy precipitation events, use a different hydrological model, and define different metrics to assess the impact relevance of the hydrological response. The key message here is that the presented framework for counterfactual scenario analysis provides a different view on flash flood hazards which should

be helpful in reducing the element of surprise in disaster risk management.

Appendix A: Creation of an HPE catalogue

The catalogue was created as follows (see also Fig. A1 for illustration). For simplification we just used two durations in Fig. A1 (1 and 72 h), while in our actual study we used eight durations (1, 2, 4, 6, 12, 24, 48, 72 h).

1. We applied a $72\text{ h} \times 3\text{ km} \times 3\text{ km}$ moving window for each pixel in the RADKLIM dataset. In Fig. A1a and b the pixel is surrounded by a red box. In this moving window we aggregate the rainfall to the durations of respective durations (Fig. A1c and d). For each duration we calculate the return periods for every pixel in the moving window (Fig. A1e and f). Now we can compute the xWEI metric. The return periods get sorted by decreasing order (Fig. A1g and h). We then compute the extremeness E_{tA} based on Müller and Kaspar (2014):

$$E_{tA} = \frac{\sum_{i=1}^n \ln(p_{t,i})}{n} \cdot \frac{\sqrt{A}}{\sqrt{\pi}} [\ln(\text{year})\text{km}]. \quad (\text{A1})$$

The process is explained in more detail in Voit and Heistermann (2022).

Following this procedure, we get an E_{tA} curve for every duration (Fig. A1i and j). The E_{tA} curves are placed on a grid (Fig. A1k). The E_{tA} curves span a surface. The volume underneath that surface is the xWEI value for the pixel (Fig. A1l), which is high if the rainfall in the $3\text{ km} \times 3\text{ km}$ neighbourhood was extreme at multiple durations (between 1 and 72 h).

2. This way the xWEI moving window works as a filter for the rainfall data. The result is a dataset of xWEI values with the same dimensions (x , y , time) as the RADKLIM dataset. An xWEI value of 10 is approximately equal to an event that had a return period of around 10 years for one duration and at a spatial scale of 9 km^2 .
3. All cells with an xWEI < 10 were discarded (set to NaN, not a number) to ensure that there are just cells remaining which signify extreme rainfall. The remaining adjacent cells were clustered based on their neighbourhood (pixels within 10 km). This way we obtained distinct clusters where the rainfall must have been exceptionally high.
4. Finally, we determined the bounding box and computed the xWEI value for the entire bounding box, for each identified cluster.

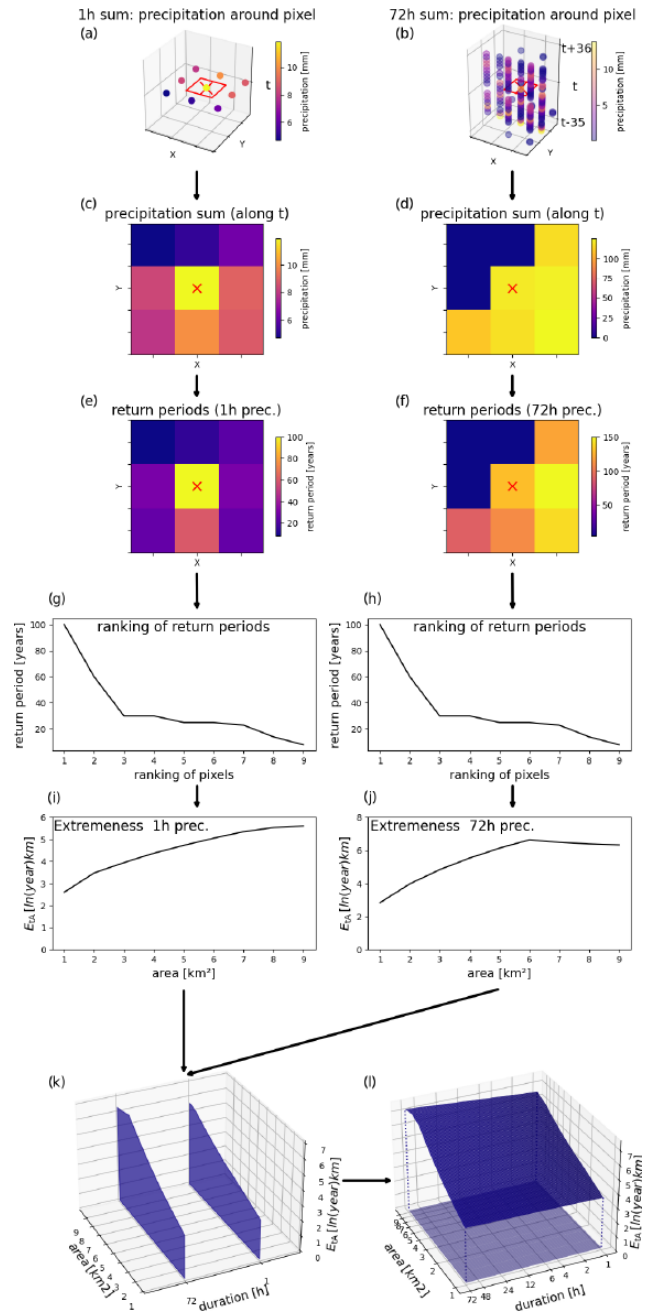


Figure A1. Pixel-wise computation of the xWEI metric. (a, b) The rainfall data in a $3\text{ km} \times 3\text{ km}$ neighbourhood for the respective duration: (a) 1 h precipitation and (b) 72 h precipitation. (c, d) Precipitation sums for the respective durations. (e, f) Return periods of the precipitation sums. (g, h) Ranked return periods. (i, j) E_{tA} curves computed from the ranked return periods. (k) E_{tA} curves placed on a grid. (l) A surface spanned across the curves. The volume under this surface is the xWEI value of the pixel.

Appendix B: Distribution of subbasin sizes

Figure B1 shows the distribution of subbasin sizes for the study area comprising Germany.

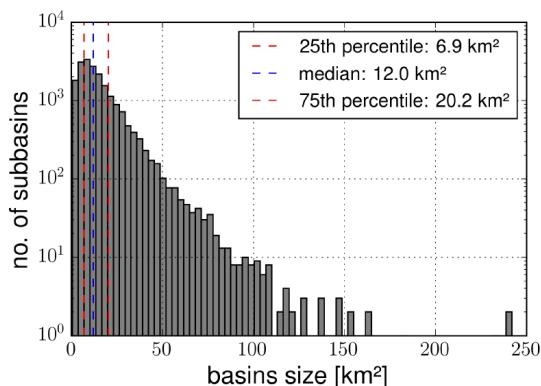


Figure B1. Distribution of subbasin sizes in the study area. The blue line indicates the median size; the red lines are the 25th and 75th percentiles.

Appendix C: Description of the top 10 events

We supply further detail about the top 10 events between 2001–2022 (Sect. 4.1) which were identified using the procedure described in Sect. 3.1 and Appendix A.

- LS/Jul02 hit the Harz Mountains in the centre of Germany with high rainfall sums and led to the flooding of some cities (e.g. Braunschweig). Apparently this HPE did not cause extensive damage as there is not much literature about this event, apart from local newspapers. Furthermore, this event was overshadowed by one of the largest flood catastrophes in Germany just 1 month later (SN/Aug2002). We can just hypothesize that the event would have caused more damage, had it not happened in the Harz area, which is a watershed. Additionally, there are large reservoirs in this area which regulate streamflow and might have prevented the formation of a larger flood wave.
- BB/Jun17 caused massive urban flooding in Berlin. This HPE caused the largest amount of insured losses in the period 2002 to 2017 (EUR 60 million) in the greater Berlin area (Caldas-Alvarez et al., 2022).
- The SN/Aug02 HPE caused extensive flooding in central Europe (Germany, Austria, Czech Republic, and Slovakia). The flooding occurred in the catchments of the Danube and the Elbe. In Germany alone the flood caused 21 casualties and record-breaking damages of EUR 11.6 billion (Thieken et al., 2007; CRED/UCLouvain, 2023).

- Regarding damages, the NW/Jul21 HPE exceeded all previously recorded events, even though the rainfall sums were not the most extreme, compared to other historic events (Ludwig et al., 2023). The HPE affected mainly Belgium, the Netherlands, and western Germany. Damages totalling EUR 40 billion and 191 casualties (CRED/UCLouvain, 2023) are the consequences of this HPE.
- The flood following LS/Jul17 caused damage in the districts surrounding the Harz Mountains and the city of Hildesheim (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, NLWKN). According to the DWD the meteorological extremeness of this HPE was similar to the infamous SN/Aug02 event, but due to the location the consequences were not as serious (Becker et al., 2017).
- BW/May16 was a large HPE across central Europe which affected southern Germany. The event included episodes of intense small-scale precipitation which caused e.g. the flash flood that partly destroyed the city of Braunsbach (Bronstert et al., 2018). This caused damages of EUR 2 billion and seven deaths (CRED/UCLouvain, 2023).
- Even though BB/Jun21 displayed the highest daily rainfall sum in Germany in 2021 (198.7 mm, Becker et al., 2017), the event did not cause a lot of damage.
- The BB/Jun20 HPE showed heavy rainfall, especially in shorter durations, in the Brandenburg area and caused smaller floods but did not cause extensive damage.
- Even though the precipitation sums during HS/May19 exceeded a 100-year return period in many locations, this HPEs did not cause large amounts of damage.
- The SN/Jun13 event occurred from 18 until 22 June 2013. Although it was rated as very extreme with regard to the xWEI, it did not appear to cause much damage, except for some more local flash floods in Saxony. We would like to emphasize that this event must not be confused with the event that hit central Europe from 30 May to 4 June 2013 and caused large-scale flooding of many rivers, specifically the Danube and Elbe (Schröter et al., 2015). Despite its large impact, the latter event does not appear among the top 10 HPEs with regard to the xWEI (it is ranked 11th, however, with regard to the WEI index). While it might appear surprising that the event from 18 to 22 June 2013 did not cause more damage, it could also have received less attention in the aftermath of the earlier flood disaster in June 2013.

Code and data availability. We have published code and data to exemplify the computation of both WEI and xWEI in the fol-

lowing repository: <https://doi.org/10.5281/zenodo.6556446> (Voit, 2022). We have published notebooks and code which demonstrate our whole workflow for this study for a small, exemplary region (Altenahr basin; see Sect. 4.2): the derivation of GIUHs from a digital elevation model, the extraction of rainfall data and effective rainfall for the subbasins from RADKLIM data, and the modelling of quick runoff. The code is published at <https://doi.org/10.5281/zenodo.10473424> (Voit, 2024).

All data used in this study are accessible in the open-data repository of the DWD. The RADKLIM v2017.002 dataset is available at https://doi.org/10.5676/DWD/RADKLIM_RW_V2017.002 (Winterrath et al., 2018a). The EU-DEM is available at <https://ec.europa.eu/eurostat/de/web/gisco/geodata/digital-elevation-model/eu-dem#DD> (European Commission, 2016). The CLC5-2018 land cover data are available at <https://gdz.bkg.bund.de/index.php/default/open-data/corine-land-cover-5-ha-stand-2018-clc5-2018.html> (BKG, 2018). The soil data are available at https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK200/buek200_node.html (BGR, 2018).

Author contributions. PV and MH conceptualized this study. PV developed the software and carried out the analysis; MH contributed to the analysis and developed the web viewer. PV prepared the manuscript with contributions from MH.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We would like to thank the open-source community without whose software and data this study would have not been possible. Some small parts of the text were improved in exchange with a language model (<https://chat.openai.com/chat> (last access: 19 June 2024) in a previous version of this paper. We thank Boha Shehu for giving further insight into the reconstruction of discharge at the Altenahr gauge.

Financial support. This research has been supported by the Deutsche Forschungsgemeinschaft (grant no. GRK 2043).

Review statement. This paper was edited by Kai Schröter and reviewed by two anonymous referees.

References

- Amponsah, W., Ayrál, P.-A., Boudevillain, B., Bouvier, C., Braud, I., Brunet, P., Delrieu, G., Didon-Lescot, J.-F., Gaume, E., Lebouc, L., Marchi, L., Marra, F., Morin, E., Nord, G., Payrastra, O., Zoccatelli, D., and Borga, M.: Integrated high-resolution dataset of high-intensity European and Mediterranean flash floods, *Earth Syst. Sci. Data*, 10, 1783–1794, <https://doi.org/10.5194/essd-10-1783-2018>, 2018.
- Becker, A., Junghänel, T., Hafer, M., Köcher, A., Rustemeier, E., Weigl, E., and Wittich, K.-P.: 2017, Juli: Einordnung der Stark- und Dauerregen in Deutschland zum Ende eines sehr nassen Juli 2017, https://www.dwd.de/DE/klimaumwelt/klimawandel/_functions/aktuellemeldungen/170731_starkniederschlaege_einordnung.html (last access: 20 June 2024), 2017.
- Beven, K. J. and Hornberger, G. M.: Assessing the effect of spatial pattern of precipitation in modeling stream flow hydrographs, *JAWRA J. Am. Water Resour. As.*, 18, 823–829, <https://doi.org/10.1111/j.1752-1688.1982.tb00078.x>, 1982.
- BGR: BÜK200 V5.5, BGR, https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK200/buek200_node.html (last access: 20 June 2024), 2018.
- BKG: CORINE CLC5-2018, <https://gdz.bkg.bund.de/index.php/default/open-data/corine-land-cover-5-ha-stand-2018-clc5-2018.html> (last access: 20 June 2024), 2018.
- Borga, M., Gaume, E., Creutin, J. D., and Marchi, L.: Surveying flash floods: gauging the ungauged extremes, *Hydrol. Process.*, 22, 3883, <https://doi.org/10.1002/hyp.7111>, 2008.
- Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M.: Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows, *J. Hydrol.*, 518, 194–205, <https://doi.org/10.1016/j.jhydrol.2014.05.022>, 2014.
- Bronstert, A., Agarwal, A., Boessenkool, B., Crisologo, I., Fischer, M., Heistermann, M., Köhn-Reich, L., López-Tarazón, J. A., Moran, T., Ozturk, U., Reinhardt-Imjela, C., and Wendi, D.: Forensic hydro-meteorological analysis of an extreme flash flood: The 2016-05-29 event in Braunsbach, SW Germany, *Sci. Total Environ.*, 630, 977–991, <https://doi.org/10.1016/j.scitotenv.2018.02.241>, 2018.
- Bunster, T., Gironás, J., and Niemann, J. D.: On the influence of upstream flow contributions on the basin response function for hydrograph prediction, *Water Resour. Res.*, 55, 4915–4935, <https://doi.org/10.1029/2018WR024510>, 2019.
- 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., Schoppa, L., Schwitalla, T., Steidl, S., Thielen, A. H., Tradowsky, J. S., Wulfmeyer, V., and Quaas, J.: Meteorological, impact and climate perspectives of the 29 June 2017 heavy precipitation event in the Berlin metropolitan area, *Nat. Hazards Earth Syst. Sci.*, 22, 3701–3724, <https://doi.org/10.5194/nhess-22-3701-2022>, 2022.
- Castellari, A.: Probabilistic envelope curves for design flood estimation at ungauged sites, *Water Resour. Res.*, 43, W04406, <https://doi.org/10.1029/2005WR004384>, 2007.
- CRED/UCLouvain: EM-DAT International Disaster Database, <https://www.emdat.be/> (last access: 20 June 2024), 2023.

- Creutin, J. D., Borga, M., Grunfest, E., Lutoff, C., Zoccatelli, D., and Ruin, I.: A space and time framework for analyzing human anticipation of flash floods, *J. Hydrol.*, 482, 14–24, <https://doi.org/10.1016/j.jhydrol.2012.11.009>, 2013.
- Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., and Payrastré, O.: Influence of rainfall spatial variability on rainfall–runoff modelling: benefit of a simulation approach?, *J. Hydrol.*, 531, 337–348, <https://doi.org/10.1016/j.jhydrol.2015.04.058>, 2015.
- Emmanuel, I., Payrastré, O., Andrieu, H., and Zuber, F.: A method for assessing the influence of rainfall spatial variability on hydrograph modeling. First case study in the Cevennes Region, southern France, *J. Hydrol.*, 555, 314–322, <https://doi.org/10.1016/j.jhydrol.2017.10.011>, 2017.
- European Commission: Digital Elevation Model over Europe (EU-DEM), Eurostat, <https://ec.europa.eu/eurostat/de/web/gisco/geodata/digital-elevation-model/eu-dem#DD> (last access: 19 June 2024), 2016.
- Garen, D. C. and Moore, D. S.: Curve number hydrology in water quality modeling: uses, abuses, and future directions 1, *JAWRA J. Am. Water Resour. As.*, 41, 377–388, 2005.
- Gaume, E., Livet, M., Desbordes, M., and Villeneuve, J.-P.: Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999, *J. Hydrol.*, 286, 135–154, <https://doi.org/10.1016/j.jhydrol.2003.09.015>, 2004.
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., Daliakopoulos, I., Garcia, J., Irimescu, A., Kohnova, S., Koutroulis, A., Marchi, L., Matreata, S., Medina, V., Preciso, E., Sempere-Torres, D., and Viglione, A.: A compilation of data on European flash floods, *J. Hydrol.*, 367, 70–78, <https://doi.org/10.1016/j.jhydrol.2008.12.028>, 2008.
- Georgakakos, K. P.: On the design of national, real-time warning systems with capability for site-specific, flash-flood forecasts, *B. Am. Meteorol. Soc.*, 67, 1233–1239, [https://doi.org/10.1175/1520-0477\(1986\)067<1233:OTDONR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1986)067<1233:OTDONR>2.0.CO;2), 1986.
- Grimaldi, S., Petroselli, A., Alonso, G., and Nardi, F.: Flow time estimation with spatially variable hillslope velocity in ungauged basins, *Adv. Water Resour.*, 33, 1216–1223, <https://doi.org/10.1016/j.advwatres.2010.06.003>, 2010.
- Heistermann, M. and Voit, P.: Counterfactual flash flood analysis for Germany, GitHub [data set], <https://hykli-up.github.io/counterfactual/> (last access: 20 June 2024), 2023.
- Heistermann, M., Jacobi, S., and Pfaff, T.: Technical Note: An open source library for processing weather radar data (wradlib), *Hydrol. Earth Syst. Sci.*, 17, 863–871, <https://doi.org/10.5194/hess-17-863-2013>, 2013.
- Heistermann, M., Collis, S., Dixon, M. J., Giangrande, S., Helmus, J. J., Kelley, B., Koistinen, J., Michelson, D. B., Peura, M., Pfaff, T., and Wolff, D. B.: The emergence of open-source software for the weather radar community, *B. Am. Meteorol. Soc.*, 96, 117–128, <https://doi.org/10.1175/BAMS-D-13-00240.1>, 2015.
- Jacobi, S. and Heistermann, M.: Benchmarking attenuation correction procedures for six years of single-polarized C-band weather radar observations in South-West Germany, *Geomat. Nat. Haz. Risk*, 7, 1785–1799, <https://doi.org/10.1080/19475705.2016.1155080>, 2016.
- Journée, M., Goudenhoofdt, E., Vannitsem, S., and Delobbe, L.: Quantitative rainfall analysis of the 2021 mid-July flood event in Belgium, *Hydrol. Earth Syst. Sci.*, 27, 3169–3189, <https://doi.org/10.5194/hess-27-3169-2023>, 2023.
- Karssenberg, D., Schmitz, O., Salamon, P., De Jong, K., and Bierkens, M. F.: A software framework for cYionstruction of process-based stochastic spatio-temporal models and data assimilation, *Environ. Modell. Softw.*, 25, 489–502, <https://doi.org/10.1016/j.envsoft.2009.10.004>, 2010.
- Kreklow, J., Tetzlaff, B., Kuhnt, G., and Burkhard, B.: A rainfall data intercomparison dataset of RADKLIM, RADOLAN, and rain gauge data for Germany, *Data*, 4, 118, <https://doi.org/10.3390/data4030118>, 2019.
- 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, *Meteorol. Z.*, 28, 363–378, <https://doi.org/10.1127/metz/2019/0964>, 2019.
- 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, *Meteorol. Z.*, 469–487, <https://doi.org/10.1127/metz/2021/1088>, 2021.
- Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., Ehret, U., Feldmann, H., Hundhausen, M., Knippertz, P., Küpfer, K., Kunz, M., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Seidel, F., and Wisotzky, C.: A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 2: Historical context and relation to climate change, *Nat. Hazards Earth Syst. Sci.*, 23, 1287–1311, <https://doi.org/10.5194/nhess-23-1287-2023>, 2023.
- Maidment, D., Olivera, F., Calver, A., Eatherall, A., and Fraczek, W.: Unit hydrograph derived from a spatially distributed velocity field, *Hydrol. Process.*, 10, 831–844, [https://doi.org/10.1002/\(SICI\)1099-1085\(199606\)10:6<831::AID-HYP374>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1099-1085(199606)10:6<831::AID-HYP374>3.0.CO;2-N), 1996.
- Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, *J. Hydrol.*, 394, 118–133, <https://doi.org/10.1016/j.jhydrol.2010.07.017>, 2010.
- Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gatzke, 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, *Nat. Hazards Earth Syst. Sci.*, 23, 525–551, <https://doi.org/10.5194/nhess-23-525-2023>, 2023.
- Montanari, A., Merz, B., and Blöschl, G.: HESS Opinions: The Sword of Damocles of the Impossible Flood, *EGU Sphere [preprint]*, <https://doi.org/10.5194/egusphere-2023-2420>, 2023.
- Morin, E., Georgakakos, K. P., Shamir, U., Garti, R., and Enzel, Y.: Objective, observations-based, automatic estimation of the catchment response timescale, *Water Resour. Res.*, 38, 30–1, <https://doi.org/10.1029/2001WR000808>, 2002.
- Müller, M. and Kaspar, M.: Event-adjusted evaluation of weather and climate extremes, *Nat. Hazards Earth Syst. Sci.*, 14, 473–483, <https://doi.org/10.5194/nhess-14-473-2014>, 2014.
- Natural Resources Conservation Service: Estimation of direct runoff from storm rainfall, *National Engineering Handbook, Part 630 Hydrology*, Chap. 10, https://irrigationtoolbox.com/NEH/Part630_Hydrology/H_210_630_10.pdf (last access: 20 June 2024), 2004.

- Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN): Das Juli-Hochwasser 2017 im südlichen Niedersachsen, <https://www.nlwkn.niedersachsen.de/download/124949> (last access: 20 June 2024), 2021.
- Ponce, V. M. and Hawkins, R. H.: Runoff curve number: Has it reached maturity?, *J. Hydrol. Eng.*, 1, 11–19, [https://doi.org/10.1061/\(ASCE\)1084-0699\(1996\)1:1\(11\)](https://doi.org/10.1061/(ASCE)1084-0699(1996)1:1(11)), 1996.
- Roose, N. J.: Counterfactual thinking, *Psychol. Bull.*, 121, 133, <https://doi.org/10.1037/0033-2909.121.1.133>, 1997.
- Ruiz-Villanueva, V., Borga, M., Zoccatelli, D., Marchi, L., Gaume, E., and Ehret, U.: Extreme flood response to short-duration convective rainfall in South-West Germany, *Hydrol. Earth Syst. Sci.*, 16, 1543–1559, <https://doi.org/10.5194/hess-16-1543-2012>, 2012.
- Saadi, M., Furusho-Percot, C., Belleflamme, A., Chen, J.-Y., Trömel, S., and Kollet, S.: How uncertain are precipitation and peak flow estimates for the July 2021 flooding event?, *Natural Hazards and Earth System Sciences*, 23, 159–177, <https://doi.org/10.5194/nhess-23-159-2023>, 2023.
- Schröter, K., Kunz, M., Elmer, F., Mühr, B., and Merz, B.: What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation, *Hydrol. Earth Syst. Sci.*, 19, 309–327, <https://doi.org/10.5194/hess-19-309-2015>, 2015.
- Seibert, S. P., Auerswald, K., Seibert, S. P., and Auerswald, K.: Abflussentstehung–wie aus Niederschlag Abfluss wird, Hochwasserminderung im ländlichen Raum: Ein Handbuch zur quantitativen Planung, Springer, 61–93, https://doi.org/10.1007/978-3-662-61033-6_4, 2020.
- Singh, P., Mishra, S. K., and Jain, M. K.: A review of the synthetic unit hydrograph: from the empirical UH to advanced geomorphological methods, *Hydrolog. Sci. J.*, 59, 239–261, <https://doi.org/10.1080/02626667.2013.870664>, 2014.
- Singh, V.: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrol. Process.*, 11, 1649–1669, 1997.
- Sivapalan, M., Jothityangkoon, C., and Menabde, M.: Linearity and nonlinearity of basin response as a function of scale: Discussion of alternative definitions, *Water Resour. Res.*, 38, 4–1, <https://doi.org/10.1029/2001WR000482>, 2002.
- Tarolli, M., Borga, M., Zoccatelli, D., Bernhofer, C., Jatho, N., and Janabi, F. a.: Rainfall space-time organization and orographic control on flash flood response: the Weisseritz event of August 13, 2002, *J. Hydrol. Eng.*, 18, 183–193, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000569](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000569), 2013.
- Thielen, A. H., Kreibich, H., Müller, M., and Merz, B.: Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002, *Hydrolog. Sci. J.*, 52, 1016–1037, 2007.
- U.S. Department of Agriculture-Soil Conservation Service: Estimation of Direct Runoff From Storm Rainfall, SCS National Engineering Handbook, Section 4, Hydrology, Chap. 10, <https://irrigationtoolbox.com/NEH/Part630hydrology/neh630-ch21.pdf> (last access: 20 June 2024), 1972.
- Voit, P.: plvoit/xWEI-Quantifying-the-extremeness-of-precipitation-across-scales: xWEI (v.1.0.0), Zenodo [code and data set], <https://doi.org/10.5281/zenodo.6556446>, 2022.
- Voit, P.: A downward counterfactual analysis of flash floods in Germany – Code repository (v0.1), Zenodo [code], <https://doi.org/10.5281/zenodo.10473424>, 2024.
- Voit, P. and Heistermann, M.: A new index to quantify the extremeness of precipitation across scales, *Natural Hazards and Earth System Sciences*, 22, 2791–2805, <https://doi.org/10.5194/nhess-22-2791-2022>, 2022.
- Winterrath, T., Rosenow, W., and Weigl, E.: On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management, *Weather Radar and Hydrology, IAHS-Aish P.*, 351, 323–329, 2012.
- Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: Gauge-adjusted one-hour precipitation sum (RW): RAD-KLIM Version 2017.002: Reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW), DWD [data set], https://doi.org/10.5676/DWD/RADKLIM_RW_V2017.002, 2018a.
- Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Walawender, E., Weigl, E., and Becker, A.: Erstellung einer radargestützten hochaufgelösten Nieder-schlagsklimatologie für Deutschland zur Auswertung der rezenten Änderungen des Extremverhaltens von Niederschlag, Freie Universität Berlin, Berlin, <https://doi.org/10.17169/refubium-25153>, 2018b.
- Woo, G.: Downward counterfactual search for extreme events, *Front. Earth Sci.*, 7, 340, <https://doi.org/10.3389/feart.2019.00340>, 2019.
- Yi, B., Chen, L., Zhang, H., Singh, V. P., Jiang, P., Liu, Y., Guo, H., and Qiu, H.: A time-varying distributed unit hydrograph method considering soil moisture, *Hydrology and Earth System Sciences*, 26, 5269–5289, <https://doi.org/10.5194/hess-26-5269-2022>, 2022.
- Zischg, A. P., Felder, G., Weingartner, R., Quinn, N., Coxon, G., Neal, J., Freer, J., and Bates, P.: Effects of variability in probable maximum precipitation patterns on flood losses, *Hydrol. Earth Syst. Sci.*, 22, 2759–2773, <https://doi.org/10.5194/hess-22-2759-2018>, 2018.
- Zoccatelli, D., Borga, M., Zanon, F., Antonescu, B., and Stancalie, G.: Which rainfall spatial information for flash flood response modelling? A numerical investigation based on data from the Carpathian range, Romania, *J. Hydrol.*, 394, 148–161, <https://doi.org/10.1016/j.jhydrol.2010.07.019>, 2010.