



Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 large ensemble

Florian Willkofer¹, Raul R. Wood^{1,2,3}, and Ralf Ludwig¹

¹Department of Geography, Ludwig-Maximilians-Universität München, 80333 Munich, Germany

²WSL Institute for Snow and Avalanche Research SLF, 7260 Davos Dorf, Switzerland

³Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, 7260 Davos Dorf, Switzerland

Correspondence: Florian Willkofer (florian.willkofer@lmu.de, florian.willkofer@campus.lmu.de)

Received: 3 September 2023 – Discussion started: 6 September 2023

Revised: 15 April 2024 – Accepted: 18 April 2024 – Published: 8 July 2024

Abstract. Severe floods with extreme return periods of 100 years and beyond have been observed in several large rivers in Bavaria in the last 3 decades. Flood protection structures are typically designed based on a 100-year event, relying on statistical extrapolations of relatively short observation time series while ignoring potential temporal non-stationarity. However, future precipitation projections indicate an increase in the frequency and intensity of extreme rainfall events, as well as a shift in seasonality. This study aims to examine the impact of climate change on the 100-year flood (HF_{100}) events of 98 hydrometric gauges within hydrological Bavaria. A hydrological climate change impact (CCI) modeling chain consisting of a regional Single Model Initial-condition Large Ensemble (SMILE) and a single hydrological model was created. The 50 equally probable members of the Canadian Regional Climate Model version 5 large ensemble (CRCM5-LE) were used to drive the hydrological model WaSiM (Water balance Simulation Model) to create a hydro-SMILE. As a result, a database of 1500 model years (50 members \times 30 years) per investigated time period was established for extreme value analysis (EVA) to illustrate the benefit of the hydro-SMILE approach for a robust estimation of HF_{100} based on annual maxima (AM) and to examine the CCI on the frequency and magnitude of HF_{100} in different discharge regimes under a strong-emission scenario (RCP8.5). The results demonstrate that the hydro-SMILE approach provides a clear advantage for a robust estimation of HF_{100} using the empirical probability of 1500 AM compared to its estimation using the generalized extreme value (GEV) distribution of 1000 samples of typically available time series sizes of 30, 100, and 200 years. Thereby, by applying

the hydro-SMILE framework, the uncertainty from statistical estimation can be reduced. The study highlights the added value of using hydrological SMILES to project future flood return levels. The CCI of HF_{100} varies for different flow regimes, with snowmelt-driven catchments experiencing severe increases in frequency and magnitude, leading to unseen extremes that impact the distribution. Pluvial regimes show a lower intensification or even decline. The dynamics of HF_{100} driving mechanisms depict a decline in snowmelt-driven events in favor of rainfall-driven events, an increase in events driven by convective rainfall, and almost no change in the ratio between single-driver and compound events towards the end of the century.

1 Introduction

The devastating force of floods poses a significant threat to infrastructure, livestock, and human life. In Germany, two of the most severe floods in the last 3 decades were the 2002 and 2013 flood events (along with other major events in 1999, 2005, and 2016) (Thieken et al., 2016; Blöschl et al., 2013). The 2002 and 2013 events caused a total of about EUR 17 billion in economic damage due to their large spatial extent and high water levels, with the 2013 flood considered to be the most extreme event in the last 60 years (Thieken et al., 2016). However, different climatic and catchment conditions caused these events, with the 2002 event resulting from intense rainfall and leading to flash floods across multiple small catchments, while the 2013 event was due to high antecedent soil moisture from long-lasting precipi-

tation followed by more moderate but spatially widespread rainfall (Thielen et al., 2016). In addition to precipitation magnitude, other flood drivers such as antecedent soil moisture conditions, snowmelt, and flood-driving processes determined by catchment and river characteristics contribute to the non-linearity of the hydrological response to extreme-precipitation events (Blöschl et al., 2015). Recent studies analyzing European flood events over the last 5 decades suggest an increase in the magnitude and frequency of high flows and flood events depending on the event type and region (Blöschl et al., 2019; Bertola et al., 2020; Blöschl et al., 2015). However, this trend depends on the time frame considered for the analysis, and the evaluation period remains crucial for either the estimation or the development of high return periods (Blöschl et al., 2015; Schulz and Bernhardt, 2016). Precipitation (heavy precipitation and long-lasting rainfall) and snowmelt (in regions with snowmelt-governed regimes) remain the primary natural causes of flooding, with other influences (e.g., catchment characteristics, antecedent catchment conditions, compound events with snow- or glacier melt) and snowmelt becoming less important once a certain threshold of extreme precipitation is exceeded (Brunner et al., 2021b).

According to the sixth Intergovernmental Panel on Climate Change (IPCC) Assessment Report, there is high confidence that a warmer climate will intensify wet weather and climate conditions affecting flooding (IPCC, 2021). Even with a 1.5 °C warming limit under the Paris Agreement, heavy precipitation, along with extreme-discharge events, is likely to intensify in Europe, with increasing confidence above 2 °C warming (IPCC, 2021).

For most discharge gauges, observational records begin in the 19th century or even later (Blöschl et al., 2015). Although most of these observations offer sufficiently long time series of data for estimating peak flows of moderate return periods, they still hinder a robust statistical estimation of extreme return periods, such as the 100-year flood and above. These types of extreme hydrological events are required for structural flood protection and risk management (Wilhelm et al., 2022; Brunner et al., 2021a; Blöschl et al., 2019). Brunner et al. (2021a) illustrate the challenges in modeling and predicting high flows due to data availability, process representation, and human influences.

Recently, single-model initial-condition large ensembles (SMILEs) have emerged as a powerful tool to enhance the statistical analysis of extremes in climatological behavior (von Trentini et al., 2020; Wood and Ludwig, 2020; Wood et al., 2021; Aalbers et al., 2018; Martel et al., 2020). Unlike other common ensembles of different global or regional climate model (GCM/RCM) combinations, SMILEs comprise multiple equiprobable realizations (members) of a single GCM or a GCM–RCM combination that differ only in their initial conditions, representing the chaotic nature of the climate system (Arora et al., 2011; Fyfe et al., 2017; Kirchmeier-Young et al., 2017; Sigmond et al., 2018; Leduc et al., 2019). The actual model structure, physics, parameter-

ization, and external forcings are preserved. Thus, SMILEs offer a profound database for analyzing internal (or natural) climate variability (Wood and Ludwig, 2020; Martel et al., 2018), separating natural variability from an actual change signal (Aalbers et al., 2018; Wood and Ludwig, 2020) and extreme events (Wood et al., 2021; Martel et al., 2018). Applying SMILEs for hydrological modeling allows for the creation of a so-called hydro-SMILE, which in turn allows for the exploitation of vast data for the analysis of the hydrological response of catchments to extreme-precipitation events.

Due to the high spatiotemporal resolution, this ensemble-based climate and hydrological modeling approach is computationally demanding. However, the high spatiotemporal resolution of a hydro-SMILE is particularly valuable for an enhanced representation of extreme values in models as it allows for spatially refined catchment features (e.g., slopes, soil characteristics, land use) and more precise values (e.g., discharge) due to a higher temporal resolution. Thus, this study focuses on only a single region comprised of the major Bavarian river basins (upper Danube, Main, Inn) with all their tributaries to account for the computational demand, as well as the advantages gained by the high resolution.

In this study, a climatological SMILE is employed to drive a physically based hydrological model with high spatiotemporal resolution for the major Bavarian river catchments. The resulting hydro-SMILE is used to answer the following questions.

- Is there a benefit in applying a SMILE for hydrological impact modeling regarding the estimation of high flows of large return periods?
- How does climate change affect the dynamics in terms of the frequency and magnitude of extreme discharges?
- How are the driving mechanisms of these extreme discharges changing?

Although the data presented in this study would allow for an analysis of events beyond the 100-year flood, we focus on this extreme event to answer these questions as this event is widely used in the literature, higher return periods are prone to increased uncertainties (e.g., HF₁₀₀₀), and HF₁₀₀ serves as design criterion for water management infrastructure in this region and elsewhere. The study area is first introduced in Sect. 2.1, followed by an overview of the climatological SMILE post-processing in Sect. 2.2.1. The hydrological model setup used to produce the hydro-SMILE, along with an evaluation of its performance, is then presented in Sect. 2.2.2. The subsequent sections describe the methods to illustrate the benefit of a hydro-SMILE for the estimation of peak flow with high return periods (Sect. 2.2.3), to assess the influence of climate change on the change in the magnitude and frequency of the 100-year flood (Sect. 2.2.4), and to determine the changes in drivers of events with magnitudes of at least that of the 100-year flood (Sect. 2.2.5). Finally, the

results of the analysis are then presented in Sect. 3.1 to 3.3 and are later discussed in Sect. 4, followed by concluding remarks in Sect. 5.

2 Study area, data, and methods

2.1 Study area

This study focuses on the major Bavarian rivers, including the upper Danube upstream of Achleiten, the Main, the Inn, and the upstream tributaries of the Elbe, as well as their smaller and larger tributaries originating from adjacent states (Baden-Württemberg, Hesse, Thuringia) and countries (Austria, Switzerland, Italy, Czech Republic). The catchments of these rivers extend beyond the political borders of Bavaria (Fig. 1). The entirety of these catchments is referred to as hydrological Bavaria in this study.

Hydrological Bavaria covers approximately 100 000 km² and features a diverse landscape ranging from the Alps (with the highest point being Piz Bernina at 4049 m above sea level (m a.s.l.)) and the alpine foreland in the south to the southern German escarpment in the north of the study area (with the lowest point being 90 m a.s.l. at Frankfurt-Osthafen) and the eastern mountain ranges to the east (Willkofer et al., 2020; Poschlod et al., 2020). The complexity of these landscapes and different climatological conditions (up to 1100 mm annual total precipitation in the north and 2500 mm in the south; an mean annual temperature of 10 °C in the north down to 5 °C (−8 °C on alpine summits; Poschlod et al., 2020) in the south) result in a variety of runoff regimes (Poschlod et al., 2020).

The discharge of many rivers within hydrological Bavaria is influenced by artificial retention structures (i.e., dams, retention basins), naturally formed lakes or transfer systems (drinking-water supply, low-flow elevation) (Willkofer et al., 2020). The major river catchments were divided into a total of 98 smaller sub-catchments to better represent the various flow regime types of the respective gauges, which are, furthermore, of common interest for flood protection (Willkofer et al., 2020).

2.2 Data and methods

To assess the impact of climate change on extreme return periods of peak flows, the hydroclimatic modeling chain illustrated in Fig. 2 was introduced within the scope of the ClimEx project (Climate Change and Hydrological Extreme Events, <https://www.climex-project.org>, last access: 19 July 2023). This common chain is divided into a climate section and a hydrological impact section and covers three spatial scales (GCM scale, RCM scale, hydrological model scale), with increasing resolutions along the chain.

Since the introduced model chain requires a vast number of computational resources, the ClimEx project employed the high-performance computing systems of the Leibniz Super-

computing Centre (LRZ), as well as its technical and consultative support to migrate and adapt software and data to its systems, to facilitate calculations, and to provide an extensive amount of storage to archive the data and make them available to the scientific community (data available at <https://www.climex-project.org>).

2.2.1 Climate data

A SMILE composed of 50 independent members of the Canadian Earth System Model version 2 (CanESM2) large ensemble (LE) was used as a base for all further analysis. The CanESM2-LE was produced by the Canadian Centre for Climate Modeling and Analysis (CCCma) and has been described in previous publications (Fyfe et al., 2017; Kirchmeier-Young et al., 2017; Arora et al., 2011; Leduc et al., 2019). All members of the CanESM2-LE used natural and anthropogenic forcings for the historical period from 1950 to 2005 and the representative concentration pathway 8.5 (RCP8.5; van Vuuren et al., 2011) emission scenario from 2006 to 2099 (Kirchmeier-Young et al., 2017; Leduc et al., 2019; Fyfe et al., 2017; Sigmond et al., 2018). The individual members differ only in terms of their initial conditions rather than in terms of changes in model structure, physics, or parameters; therefore, they offer a range of internal or natural variability of the climate system at a global scale.

These 50 members were dynamically downscaled from $\sim 2.85^\circ$ (≈ 310 km) to 0.11° (≈ 12 km) using the Canadian Regional Climate Model version 5 (CRCM5; Martynov et al., 2013; Šeparović et al., 2013) over two spatial domains, the European and the northeastern North American domains (Leduc et al., 2019). As with the CanESM2-LE, variations between the individual members were obtained by unique initial conditions for each member, thus providing a range of internal or natural variability on a regional scale. The resulting CRCM5 large ensemble (CRCM5-LE; Leduc et al., 2019) of 50 transient members provides the basis for assessing the impact of climate change on hydro-meteorological extreme events for hydrological Bavaria. We focus on the model years 1961 to 2099 as opposed to 1950 to 2099 to account for the time it takes for the RCM to produce fully independent realizations due to the inertia of the ocean model (Leduc et al., 2019). A comparison between the CRCM5-LE and the E-OBS observational gridded dataset (Haylock et al., 2008) on the CRCM5 grid revealed biases for a historical period between 1980 and 2012, showing regional and seasonal variations in the magnitude of temperature and precipitation over Europe (Leduc et al., 2019). Since the creation of RCM-LEs is challenging in terms of computational demand (performance and storage), only a few are available to date (Addor and Fischer, 2015; Leduc et al., 2019; Aalbers et al., 2018; Brönnimann et al., 2018). However, all of the RCM-LEs differ in terms of their domain size, spatial resolution, and ensemble size. All RCM-LEs are ensembles of opportunity and are dependent on the availability of the driv-

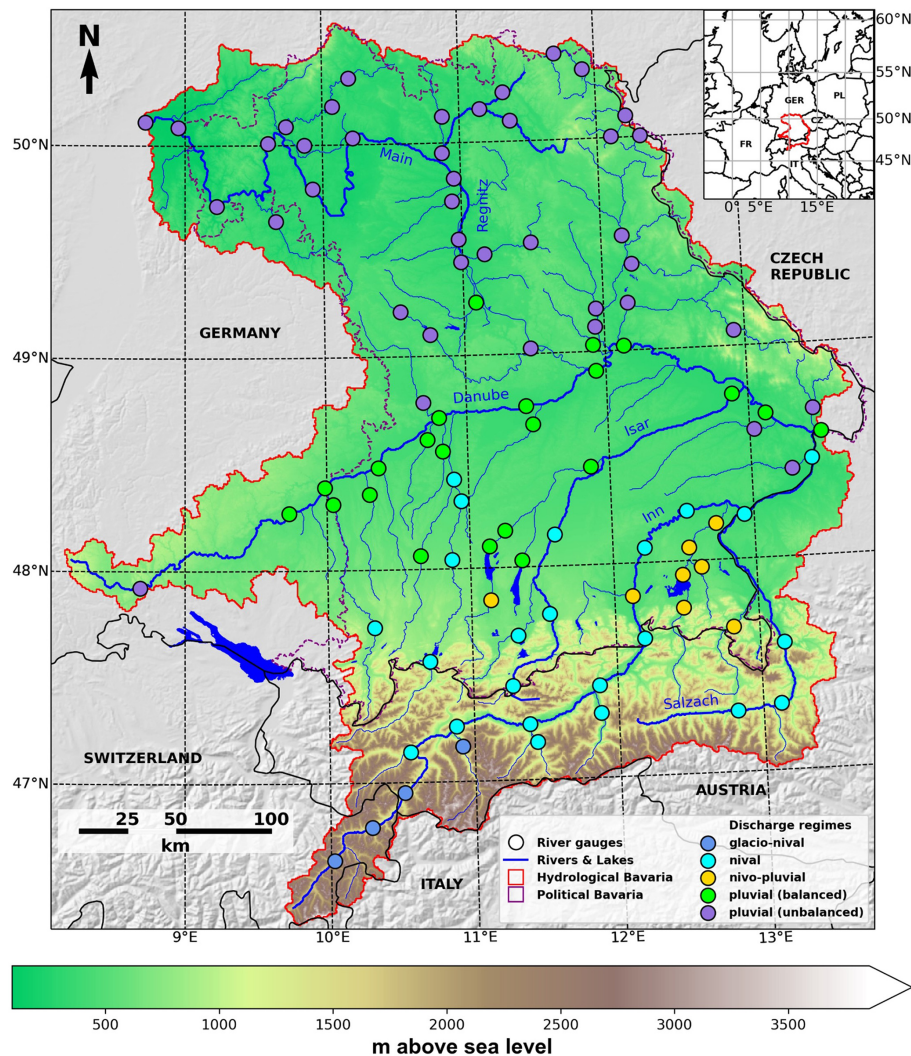


Figure 1. Map showing the elevation of hydrological Bavaria (red line; European Environment Agency, 2013), which comprises political Bavaria (dashed purple line) and the 98 hydrometric gauges used in this study, as well as their respective discharge regime type (colored dots) at their respective rivers (blue lines).

ing GCM-LEs, which, in the CMIP5 phase, are all based on RCP8.5. Thus, in this study, only a single RCM-LE, as well as a single scenario, was employed.

Since this bias was considered to affect the behavior of the outputs of the hydrological model due to shifts in seasonality and magnitude, a bias correction was applied. The required meteorological data of precipitation, air temperature, relative air humidity, incoming shortwave radiation, and wind speed were adjusted to match a meteorological reference of interpolated 3-hourly station data (Sub-Daily Climate REFERENCE, SDCLIREF; Ludwig et al., 2019) on the RCM grid using an adaptation of the quantile-mapping approach of Mpelasoka and Chiew (2009). This approach, as described in Willkofer et al. (2018), involved using multiplicative or additive correction factors and was further adapted for using 3-hourly correction factors for every quantile and month (for further

details, see Sect. S3 in the Supplement). To preserve an internal spread between the members, a single set of factors was deduced from a combination of all 50 members. Despite the numerous benefits (increased reliability of climate change projections of the hydrological impact model, reduced bias in mean annual discharge) and shortcomings (disrupting feedbacks between fluxes, modification of change signals, assumption of a stationary bias) of bias correction (e.g., Teutschbein and Seibert, 2012; Maraun, 2016; Ehret et al., 2012; Dettinger et al., 2004; Chen et al., 2021; Huang et al., 2014), bias correction is often inevitable for climate change impact studies (Gampe et al., 2019).

Subsequently, the bias-corrected data were statistically downscaled to the hydrological model scale (500 m × 500 m) using a mass-preserving approach (Marke, 2008; Ludwig et al., 2019). This approach involved the spatial interpola-

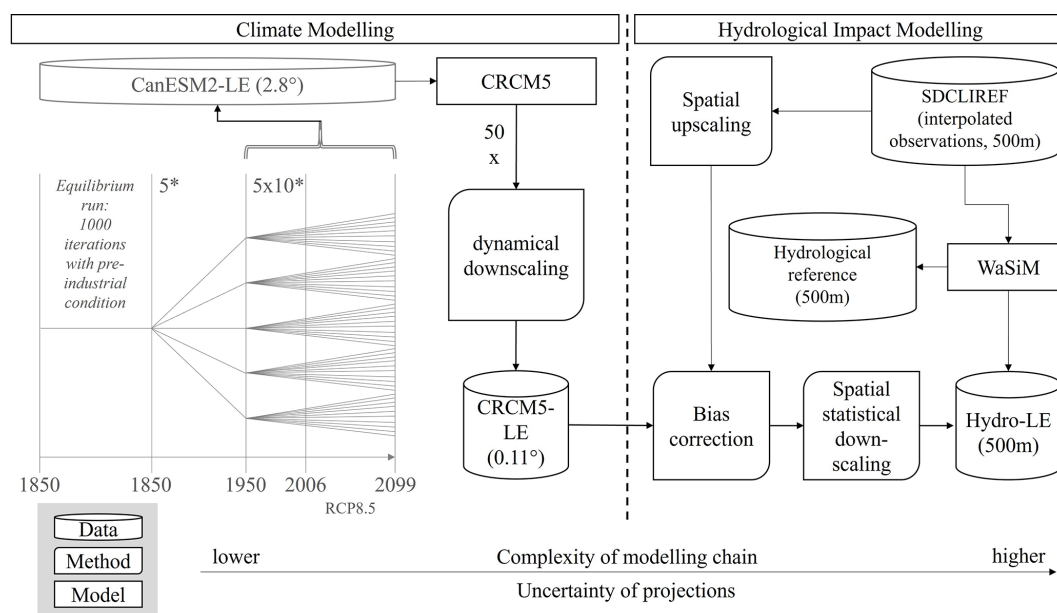


Figure 2. The ClimEx modeling chain uses the CanESM2 large ensemble (LE, gray, not created within the ClimEx project) to generate the CRCM5-LE. The CRCM5-LE is then used to explore the impacts of climate change on the hydrology of hydrological Bavaria through a hydrological large ensemble (hydro-LE) created using the hydrological model WaSiM. The SDCLIREF dataset of interpolated meteorological observations was employed to calibrate and validate the hydrological model, as well as for the bias correction. The CRCM5-LE represents a SMILE, consisting of a single model that downscales output from the employed ESM using slight differences in the initialization.

tion (inverse distance weighting) of anomalies for each time step from the monthly mean reference state (1981–2010) at the CRCM5-LE cell center points to the hydrological model scale (Brunner et al., 2021b). The interpolated time step anomalies at the hydrological scale were then applied (multiplied or added) to the respective gridded monthly climatological reference fields of the SDCLIREF (Brunner et al., 2021b). In order to ensure the mass conservation, the down-scaled RCM data were upscaled to the original RCM grid scale (mass conservative remapping) and compared to the RCM time step values to determine any correction factors necessary, which were then applied to the down-scaled grid cells to close the mass balance.

For further details, readers are referred to a comprehensive summary in the Supplement for the CanESM2-LE (Sect. S1), the CRCM5-LE (Sect. S2), and the bias correction (Sect. S3).

2.2.2 Hydrological model WaSiM

The Water balance Simulation Model (WaSiM; Schulla, 2021) was employed to perform the hydrological simulations driven by the CRCM5-LE, resulting in a hydro-SMILE (the WaSiM-LE). WaSiM is a distributed, mostly physically based, and deterministic model for simulations on various spatial (1 m to 10 km) and temporal (minute to daily) scales, with a constant time step. It includes routines for evapotranspiration, snow accumulation and snowmelt, glaciers, soil water transfer, groundwater, and discharge generation and routing (Schulla, 2021). The model is frequently used for hy-

drological climate change impact studies on various topics, such as glaciers, groundwater, and discharge, for small-scale to mesoscale catchments (Jacob et al., 2017; Neukum and Azzam, 2012; Jónsdóttir, 2008).

The model was set up at a high spatiotemporal resolution (500 m and 3 h) for 98 catchments of hydrological Bavaria, with a focus on high-flow representation using distributed data derived from the European DEM (EU-DEM; European Environment Agency, 2013), land use data provided by the CORINE land cover dataset (European Environment Agency, 2019), distributed soil information from the European Soil Database (ESDBv2.0; Panagos, 2006; European Commission and the European Soil Bureau Network, 2004), and groundwater information provided by the Hydrogeologische Übersichtskarte (HÜK; Dörhöfer et al., 2001) and International Hydrogeological Map of Europe (IHME; BGR, 2014). A single set of parameters for spatially distributed modules (i.e., evapotranspiration, soil properties) was defined globally for the entire modeling domain (Willkofer et al., 2020). Although there are abundant in situ data available for the study region, these are mainly provided as point measurements which are often representative of the entire catchment area and therefore require interpolation to the hydrological model resolution, which can introduce large uncertainties. Furthermore, some approaches of the hydrological model offer free parameters which cannot be measured. Hence, a calibration of the model for a limited number of free and usually locally measurable parameters was deemed to be neces-

sary. For further details about the model calibration procedure, we would like to refer to Willkofer et al. (2020). Local parameters for discharge storage components (i.e., interflow, direct flow) were calibrated using an automated algorithm (dynamically dimensioned search (Tolson and Shoemaker, 2007) and simulated annealing with progressing iterations (Černý, 1985; Kirkpatrick et al., 1983)), minimizing a weighted combination of performance metrics (overall metric – OM; Eq. 1), including the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), the Kling–Gupta efficiency (KGE; Gupta et al., 2009), the logarithmic NSE, and the ratio of root mean squared error to standard deviation (RSR; Moriasi et al., 2007; Willkofer et al., 2020). A best fit would result in $OM = 0$, with larger deviations from 0 indicating a worse model fit. Due to the focus on high-flow representation, more emphasis was placed on the respective measures (i.e., NSE and KGE). For further details about the model setup, the reader is referred to Willkofer et al. (2020).

$$OM = 0.5 \times (1 - NSE) + 0.25 \times (1 - KGE) + 0.15 \times (1 - \log NSE) + 0.1 \times RSR \quad (1)$$

The simulations of a single parameter set for various catchments within a heterogeneous landscape revealed satisfactory to very good results for most of the 98 gauges during the 30-year reference period of 1981 to 2010. However, for a few gauges, the model was not able to reproduce the observed discharge satisfactorily (16 (5) gauges showing NSE (KGE) values below 0.5; see also Fig. S1a and b in Sect. S4 in the Supplement) (Willkofer et al., 2020; Poschlod et al., 2020). Furthermore, the simulations reproduced the mean high flow sufficiently well, with over 60 % of the gauges showing absolute deviations from observed values below 20 %. Nonetheless, gauges in alpine or pre-alpine catchments exhibited a deficit in mean high-flow values due to the lack of observed precipitation resulting from an undercatch of precipitation for that region (Poschlod et al., 2020). Consequently, the level of trust (LOT) for the peak flows of flood events with return periods of 5, 10, and 20 years, introduced in Willkofer et al. (2020), showed a moderate to high confidence for most catchments. The LOT further depends on the model performance to a certain degree, where gauges depicting a lower model performance often exhibit a lower LOT as well. LOT values were not provided for extreme flood events (i.e., 100-year flood events) since they are subject to significant epistemic uncertainty due to the restricted availability of simulated data (30 years). In Brunner et al. (2021a), the same hydrological model simulations were evaluated on a daily timescale – in contrast to the 3-hourly timescale used here – in terms of general evaluation metrics (i.e., NSE, KGE, volume efficiency, and mean absolute error), as well as for flood-specific characteristics (i.e., number of events, mean timing of the event, mean volume, mean duration). The evaluation of flood characteristics showed a good agreement on the number of events, showing only a slight underestimation

of events; a good agreement on the timing of events, with only a slight delay in flood occurrence; and an overestimation of flood volume and duration.

Due to the holistic calibration approach employing a single set of parameters over several heterogeneous catchments, the at times poor performance of individual catchments can be expected as catchment-specific characteristics can only be considered to a certain extent (e.g., karstic soils, transfer systems, artificial reservoirs). While calibrating each catchment individually might lead to a higher performance at the respective catchment scale, it also increases the likelihood of overfitting. Furthermore, since the hydrological model serves climate change impact analysis, relative change values are of more interest than changes in absolute values. Nonetheless, the performance must be considered for interpretation. A brief overview of the model's performance for each gauge is given in the Supplement (Sect. S4).

The resulting hydro-SMILE comprises 50 members of transient simulated data from 1961 to 2099, providing a total of 6950 model years to be exploited to analyze extreme values.

2.2.3 Benefit of a hydro-SMILE for the estimation of extreme peak flows

This study used the simulated discharge for the reference period of 1981 to 2010, taken out of the entire dataset, to assess the benefits of the hydro-SMILE in estimating return levels. Like the individual members of the CRCM5-LE, the members of the WaSiM-LE are equally probable and, therefore, provide a comprehensive database to facilitate the analysis of extreme values.

Figure 3 illustrates the approach taken to emphasize the benefits of the hydro-SMILE in analyzing peak flows of high return periods for the reference period. The 30-year reference period (ref) was selected for all 50 members, resulting in 1500 model years (50 members \times 30 years) of discharge data for each of the 98 gauges. First, the annual maximum of each model year (hydrological year) was extracted for the analysis. Since the database consists of 1500 model years, this number is considered to be sufficient to employ empirical non-exceedance probabilities (Martel et al., 2020). However, to demonstrate the benefit of the hydro-SMILE database a statistical analysis using the stationary GEV distribution was also conducted for comparison purposes. A bootstrapping approach with resampling was used to create 1000 samples (n) of different sizes m (30, 100, and 200 years) (each sample without replacement). Using 1000 samples ensures that one value of the 1500 AM has a chance of > 99 % of being selected by chance for $m = 30$. The GEV was employed to estimate the return periods and corresponding confidence intervals. The parameters of the GEV distribution were estimated using L-moments (LMs). The GEV distribution was selected as it is among the better-performing methods relying on AM (Bezák et al., 2014) and is the recommended choice

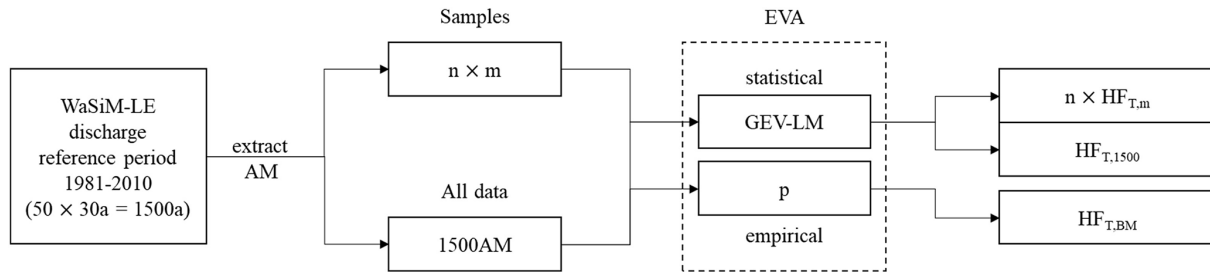


Figure 3. Process chain illustrating the benefit of a hydro-SMILE for climate change impact studies on peak flows of extreme return periods. The process includes extreme value analysis (EVA) based on annual maximum (AM), with bootstrapping resampling to create n different samples of sample size m . The probability of non-exceedance (p) and the generalized extreme value (GEV) distribution with the L-moment (LM) estimators are used to derive estimates for high-flow values of the return period T (HF_T) for the samples (m), all data (1500), and the benchmark (BM). The statistical analysis was performed using the extRemes package (v2.0) for R (Gilleland and Katz, 2016).

for German gauges (Salinas et al., 2014; Fischer and Schumann, 2016).

Although the sample size of 30 and 100 AM may be small for estimating peak flows of high return periods, they were selected along with a size of 200 AM as they represent an average (30 years) to rare (100 and 200 years) data availability of observed discharge values at different gauges (GRDC, 2021). The resulting 1000 estimates for return levels of peak flows offer a comprehensive database to demonstrate the benefit of the hydro-SMILE. Additionally, the GEV was calculated using the entire 1500-AM database for each gauge to allow for a comparison with a benchmark value. This benchmark for the return levels of peak discharge was deduced by applying the quantile based on the empirical probability of non-exceedance p (Eq. 2) to all 1500 AM values for each gauge, and it is considered to represent a robust estimate. This analysis focused on the 100-year flood, which is an event of the 100-year return period T (HF_{100} ; $T = 100$), and the corresponding 99th percentile p of the distribution of the 1500 AM values as a benchmark.

$$p = 1 - \frac{1}{T} \quad (2)$$

Values for the benchmark derived by the empirical probability, as well as the HF_{100} values estimated using the GEV, are further normalized to the benchmark to allow for a better comparison.

2.2.4 Projection of changes in frequency and magnitude

This study further investigates the dynamics of the magnitude and frequency of HF_{100} for three future periods (near future: 2020–2049, mid-future: 2040–2069, far future: 2070–2099) frequently used in similar CCI studies (e.g., Hattermann et al., 2018), providing the same database of 1500 AM values as for the benefit analysis. Therefore, the robust estimates of extreme return levels of peak flows derived by the empirical probabilities are used for the assessment of climate change impacts on their magnitude (C_M , Eq. 3) and frequency (C_F ,

Eq. 4a to c) in the three future periods.

$$C_M = \left(\frac{HF_{T_{fut}} - HF_{T_{ref}}}{HF_{T_{ref}}} \right) \times 100\% \quad (3)$$

$$C_F = \frac{1}{1 - f(HF_{T_{ref}})} \quad (4a)$$

$$f = F(HF_{T_{fut}}) \quad (4b)$$

$$F(x) = \sum_{i=1}^j h_i = \sum_{i=1}^j \frac{h(x_i)}{n} \quad (4c)$$

The change in magnitude is given as the difference between the future ($HF_{T_{fut}}$) and reference value ($HF_{T_{ref}}$) relative to the reference value in percent. The change in frequency is expressed as the return period value T and is calculated by applying the empirical cumulative distribution function F (with a frequency for an event h_i described as the ratio between the frequency for the specific event $h(x_i)$ and the number of all values n , Eq. 4c) for the respective future period (f , Eq. 4b) to the value of the 100-year flood of the reference period (Eq. 4a). The quantile value of f for the reference 100-year flood value is then used to deduce the future return period by solving the empirical probability of non-exceedance for the return period T (Eq. 4a). The change signals are calculated for each of the above-mentioned 30-year future periods. However, this analysis requires stationarity for the underlying data. Since we use all the 1500 model years provided by the 50 members, we determine stationarity if fewer than 5 % of the members exhibit a significant trend for each individual gauge. A Mann–Kendall (MK) test (Mann, 1945; Kendall, 1955) for stationarity conducted on each individual member and gauge revealed no significant trend for the reference period (with significance level $\alpha = 0.01$) for more than 95 % of the members along all gauges. However, for the future periods, the MK test exhibits significant trends for more than 5 % of the members in 6 of the 98 gauges. Limiting the evaluation periods to 20 years instead of 30 years lead to similar results for the MK test, showing no apparent trend for all gauges

Table 1. Thresholds for the identification of the driving mechanisms (drivers) of extreme-discharge events above HF_{100} . P represents the precipitation events of heavy rain (hr) and steady rain (sr), melt represents melting water from snow and glaciers (in mm snow water equivalent), and SWC_{root} represents the soil water content of the soil's root zone.

Driver	Sub-category	Volume	Accumulation period
P	hr	15 mm	3 h
	hr	20 mm	6 h
	sr	25 mm	12 h
	sr	30 mm	24 h
	sr	40 mm	48 h
	sr	60 mm	72 h
Melt	snow	15 mm	2 weeks
	glacier	15 mm	2 weeks
SWC_{root}		$110\% \times \mu(SWC_{root})_{REF}$	2 weeks

in the reference period but showing a significant trend (more than 5 % of members with a trend) for at least one gauge in the future periods. Poschlod et al. (2020) and Brunner et al. (2021b) conducted their analysis on the same database using time slices of at least 30 years as well. Thus, we choose to use 30-year periods since stationarity criteria are met in most catchments, and we opt for the larger database, as well as maintaining consistency with these studies.

2.2.5 Dynamics of driving mechanisms

The employed process-based hydrological model allows for a more detailed investigation of the dynamics of the driving mechanisms of extreme discharges of the 100-year flood and beyond. First, extreme events of magnitudes of at least HF_{100} are extracted for each of the 98 gauges and 50 members of the hydro-LE. To avoid sampling a single event multiple times, a 5 d period was used to separate individual events, as suggested by Svensson et al. (2005). The starting date of the events served as entry to extract data from potential flood drivers, which include precipitation, melt from snow and glaciers, and soil water content prior to the event. Precipitation events were further separated into heavy-rain events (hr) and steady-rain events (sr), while there was no distinction between liquid and solid precipitation. The respective thresholds to identify and separate the different precipitation event types (see Table 1) were adopted from the German Weather Service (DWD, Deutscher Wetterdienst, 2024).

A melt-driven event (snow and/or glacier) was identified if the snow water equivalent from melt exceeded 15 mm 2 weeks before the event (adapted from Brunner et al., 2021b). Extreme-discharge events may also be caused by a superposition of these driving mechanisms and are often referred to as compound events (e.g., rain on snow, rain on saturated soils). Thus, if more than one driver is identified, we

ascribe it to the compound event type. Furthermore, since an elevated soil water content cannot be responsible for an extreme-discharge event alone, it is only considered to be a contributing factor and is always part of the compound event type. In this case, the contribution of an elevated soil water content is considered when the soil water content is 10 % higher than the long-term mean soil water content for the entire reference period. All possible combinations (single drivers and compound events) result in 32 different or superimposed mechanisms. To reduce complexity and to focus on specific aspects of the different drivers, we aggregate the 32 possible combinations to six different major contributions listed in Table 2.

First, we analyze the dynamics of melt vs. rainfall events; second, the dynamics of heavy rainfall (convective events) vs. steady rainfall (advective events) are investigated; and third, we illustrate the dynamics of single drivers (any type) vs. compound events (any type).

3 Results

3.1 Benefits of hydro-SMILEs for the estimation of extreme return periods of peak flows

Large ensembles provide a vast amount of data; therefore, they are considered to be beneficial for extreme value analysis (Kendon et al., 2008; Kjellström et al., 2013; Wood and Ludwig, 2020). The benefit of a hydro-SMILE in determining robust extreme hydrological discharge values for hydrological Bavaria is analyzed, specifically for the 100-year flood. The robust values for the discharge gauges, derived using the empirical probability of non-exceedance for a 100-year event, serve as a benchmark for comparison with values derived by the GEV distribution using three different sample sizes (30, 100, 200) of AM values (Fig. 4).

The results shown in Fig. 4a, b, and c illustrate that the estimates of HF_{100} are more robust with an increasing number of AM values used for the GEV, as indicated by the decreasing spread of the blue markers around the black benchmark line with increasing sample size. Table 3 summarizes the statistical characteristics of the deviation of the estimates from the benchmark across all 98 gauges. While the range of the relative deviation of the 1000 samples of HF_{100} estimates from the benchmark is between 0.33 and 2.71 when calculated with a sample size of 30 AM values (Fig. 4a), this range diminishes to 0.49 and 1.91 for 100 AM values (Fig. 4b) and 0.56 and 1.60 for 200 AM values (Fig. 4c). Therefore, the range of the 1000 estimates diminishes with an increase in sample size, and the values cluster more densely around the benchmark. However, despite the remaining non-negligible range of deviations from the benchmark, the mean (1.01) and the median (0.98 to 1.0) across all values for all gauges are close to the benchmark value for different sample sizes. The inner 50 % of the 1000 samples across all 98 gauges exhibits

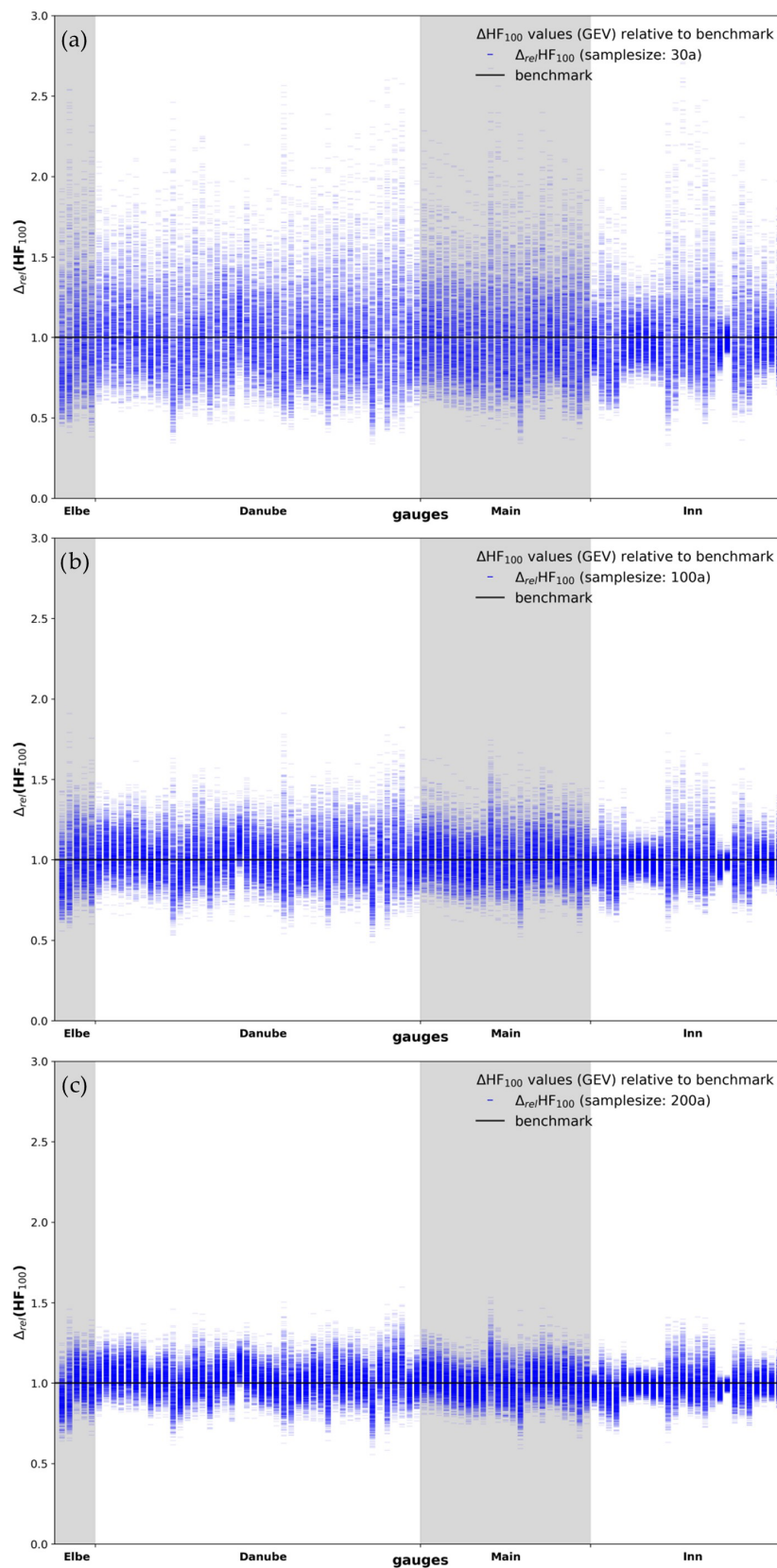


Figure 4. Comparison of HF₁₀₀ estimates calculated using the GEV distribution with 1000 AM samples of (a) 30, (b) 100, and (c) 200 years per gauge (blue markers) with the respective benchmark value (solid black line) for 98 gauges.

Table 2. Composition of the different drivers for the analysis of the dynamics of the HF₁₀₀ driving mechanisms.

Melt vs. rainfall	
Melt	Only melt events and associated compounds – excluding rainfall compounds
Rainfall	All rainfall-driven events and associated compounds – including compounds with melt events
Heavy rainfall vs. steady rainfall	
Heavy rainfall	All event types which include heavy rain – including compounds with steady rainfall
steady rainfall	All event types which include steady rain – excluding any compounds with heavy rainfall
Single vs. compound event	
Single event	All events caused by a single driver
compound event	All events caused by multiple drivers

Table 3. Summary of overall statistics of the relative deviation of the HF₁₀₀ estimates from the benchmark value across all gauges. The table includes the number of samples (n), the sample size (m) given in annual maximum (AM) values, and the 0.25 and 0.75 quantile (Q_{25} , Q_{75}) values.

N	m	Minimum	Q_{25}	Mean	Median	Q_{75}	Maximum
1000	30	0.33	0.84	1.01	0.98	1.15	2.71
1000	100	0.49	0.92	1.01	1.00	1.08	1.91
1000	200	0.56	0.94	1.01	1.00	1.07	1.60
1	1500	0.98	1.00	1.02	1.01	1.03	1.09

the largest deviation with a sample size of 30 AM (between 0.84 and 1.15) and the lowest deviation for 200 AM (0.94 to 1.07). Therefore, only 25 % of the samples show underestimations below 0.84 (0.92, 0.94), and only 75 % exhibit overestimations larger than 1.15 (1.08, 1.07) with a sample size of 30 AM (100 AM, 200 AM). Thus, with deviations larger than 15 % for 50 % of the estimates calculated using a sample size of 30 AM, only half of the estimated HF₁₀₀ values are within an acceptable range (± 15 %, considering model parameter uncertainty and errors in observations affecting the model quality regarding high flows) compared to the benchmark. This number increases with a larger sample size.

While the majority of gauges show estimates that are evenly distributed around the benchmark, some gauges exhibit a tendency towards over- or underestimation of the HF₁₀₀ estimates, with more values falling above or below the benchmark line. This behavior may be different when using more than 1000 samples to conduct the analysis. The difference between the benchmark value obtained from empirical probability and the estimates obtained from the GEV distribution can vary greatly depending on the samples selected from 1500 AM values.

A comparison between the HF₁₀₀ estimates derived using the empirical probability of non-exceedance and those obtained using the GEV distribution is shown in the Supplement (Sect. S5). The values gained from the GEV distribution still exhibit deviations from the benchmark, although they are only marginally different from it.

3.2 Changing dynamics of the 100-year peak flows in future projections

The changes in HF₁₀₀ for the investigated gauges in hydrological Bavaria in the 21st century are summarized for five distinct discharge regimes (defined by the Pardé coefficient) which were adapted from Poschlod et al. (2020) (Fig. 1). The regimes comprise the glacio-nival regime of four high-alpine catchments, a nival regime of mostly alpine to pre-alpine catchments, a nivo-pluvial regime of pre-alpine catchments, a balanced pluvial regime (little variation in mean monthly discharges) along the Danube and its tributaries in the Alpine Foreland, and the unbalanced pluvial regime (more pronounced peak in monthly discharge from January to March) (Poschlod et al., 2020). One gauge that was originally assigned to its own regime in Poschlod et al. (2020) has been re-allocated to the pluvial (unbalanced) regime as it exhibits a similar mean discharge behavior.

Within the study area, the flood protection structures are typically designed based on a stipulated estimation of HF₁₀₀ from observations, which represent a stationary condition in the past. Any future increase in the magnitude and frequency of these extreme values poses a threat to these structures.

Figure 5 displays violin plots that illustrate the range of changes in the magnitude of HF₁₀₀ events for the different discharge regimes, as well as the distribution of changes across the respective clusters of gauges for the near- (horizon 2035), mid- (horizon 2055), and far-future (horizon 2085) periods. Overall, 78 % of all gauges (76/98) show an increase

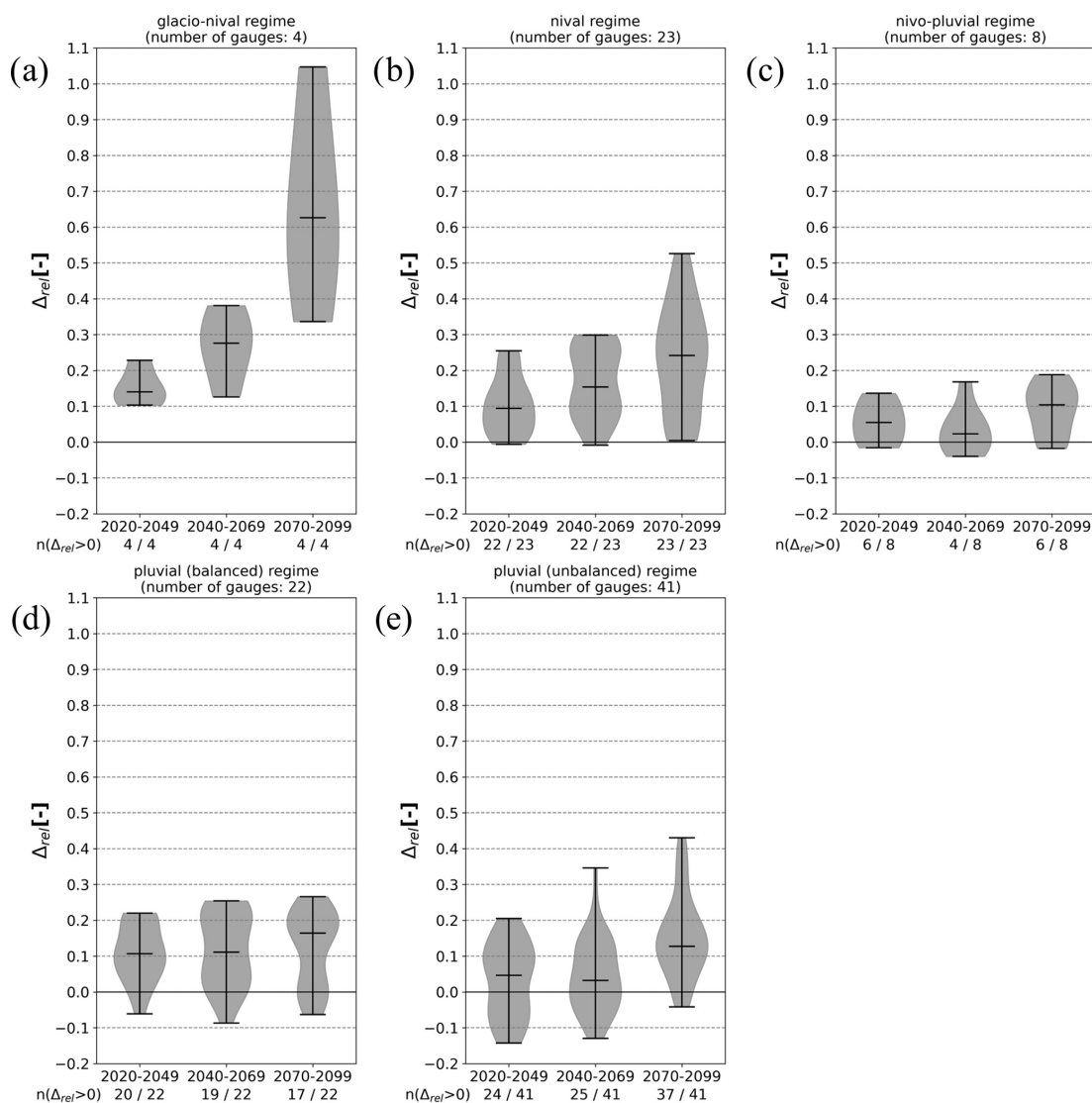


Figure 5. Violin plots indicating the changes in the magnitude of HF_{100} for the three future periods (near, middle, far) compared to the reference period, with changes presented as the relative difference (Δ_{rel}) between the reference and the future HF_{100} value for each gauge. Results of the 98 gauges are aggregated for the five discharge regimes: (a) glacio-nival, (b) nival, (c) nivo-pluvial, (d) pluvial (balanced), and (e) pluvial (unbalanced). The figures display the total number of gauges per regime, as well as the number of gauges depicting an increase in magnitude.

in magnitude for the 2035 horizon, 76 % (74/98) show an increase in magnitude for the 2055 horizon, and 89 % (87/98) show an increase in magnitude for the 2085 horizon.

The CCI values are most severe for the glacio-nival regime (Fig. 6a) as all three future periods exhibit an increase in the magnitude of the HF_{100} events of at least 10 % compared to the reference period. The nivo-pluvial regime (Fig. 6c) shows the smallest spread and the lowest increase in HF_{100} magnitude across all future periods compared to the reference period. As the distance from the Alps increases and the discharge regimes shift from snowmelt influenced to more precipitation driven, the number of gauges projecting a decrease in HF_{100} intensities increases. However, the majority

of gauges still exhibit an increase in intensities, with up to 18.8 % for the nivo-pluvial regime (Fig. 5c), 26.6 % for the balanced pluvial regime (Fig. 5d), and 43 % for the unbalanced pluvial regime (Fig. 5e) in the far future. The gradient of an increase in magnitude over all three projection periods is small for the nivo-pluvial and balanced pluvial regimes, which show the least intensification in terms of HF_{100} values for the respective periods. However, the gradient of increase is more distinct for the remaining regimes, with the largest increase in the glacio-nival regime (Fig. 5a). The gauges in this regime depict the strongest increase in HF_{100} intensities for the 2085 horizon, with an increase of 36.6 % to 104.7 %.

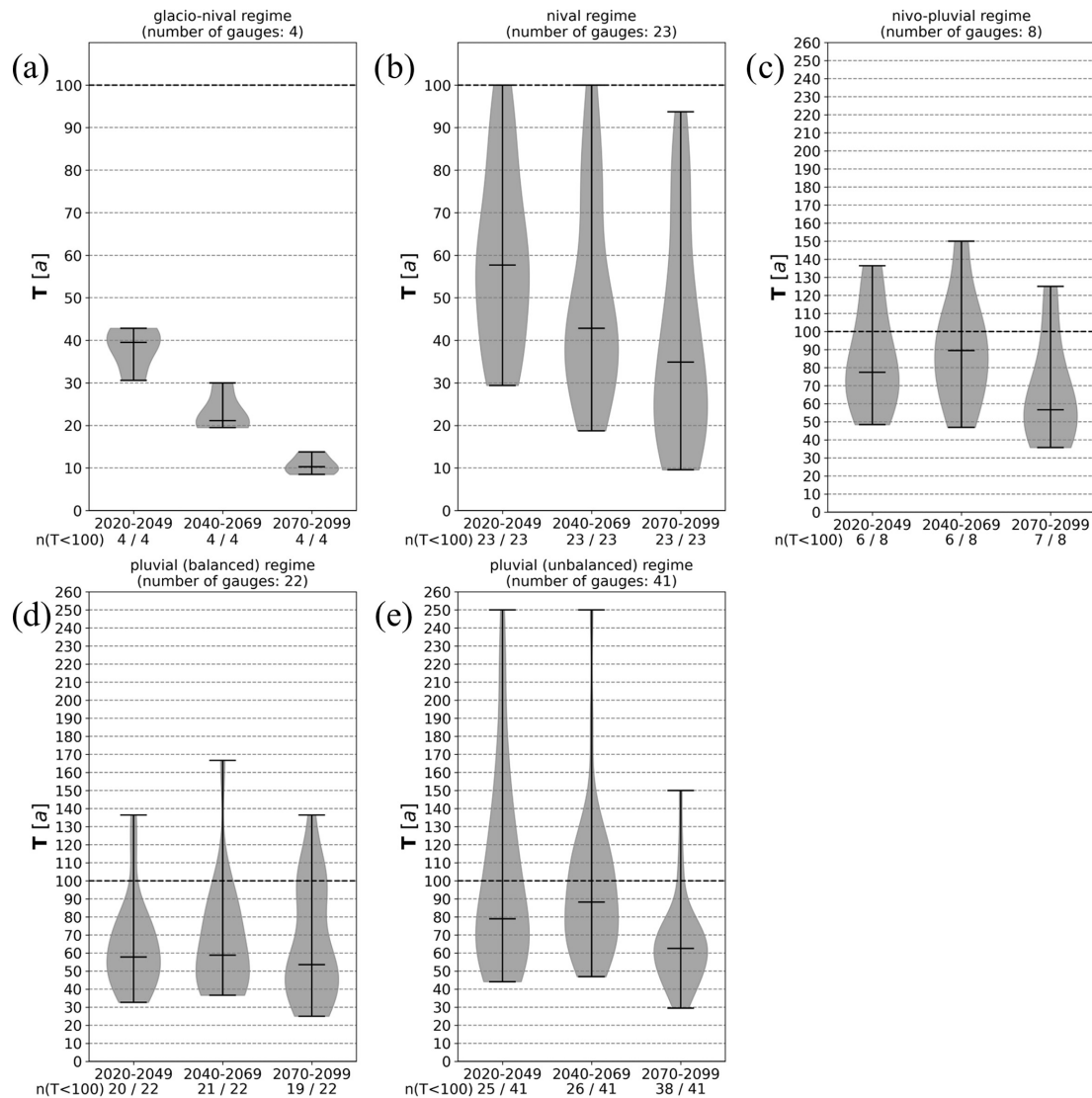


Figure 6. Violin plots indicating the changes in the frequency of HF₁₀₀ for the three future periods (near, middle, far) compared to the reference period, with changes presented as absolute values of return periods (T [a]) of the respective future period compared to the 100-year return period for each gauge. Results of the 98 gauges are aggregated for the five discharge regimes: (a) glacio-nival, (b) nival, (c) nivo-pluvial, (d) pluvial (balanced), and (e) pluvial (unbalanced). The figures display the total number of gauges per regime, as well as the number of gauges depicting an increased frequency.

Based on the future projections of the hydro-SMILE, the discharge values of HF₁₀₀ are likely to increase for most of the gauges of hydrological Bavaria. Consequently, the frequency of the HF₁₀₀ discharge for the reference period also increases. Figure 6 shows the change in frequency between the future and the reference period for the different regimes.

Values indicate the new return period associated with the HF₁₀₀ discharge from the reference period. This means values below 100 indicate an increase in frequency. The glacio-nival regime (Fig. 6a) also exhibits the strongest increase in frequency among all regimes, with HF₁₀₀ of the past becoming equivalent to a 31- to 43-year event in the near future,

thus becoming roughly 2 to 3 times more frequent. For the 2085 horizon, the same HF₁₀₀ event becomes an 8- to 14-year event, showing a 7- to 12-fold increase in frequency. A similar development is visible for some gauges in the nival regime (Fig. 6b). While the violin plot for this regime indicates that the reference 100-year event will become a 70-year event for more than 50 % of gauges, some gauges show no or only a minor increase in frequency as well. The changes for the remaining regimes are less severe but still indicate an increase in frequency for up to 50 % of the respective gauges until the middle of the century and for more than 50 % in the far future. The changes for the nivo-pluvial regime (Fig. 6c)

and the unbalanced pluvial (Fig. 6d) regime show that the frequency declines for less than 50 % of the gauges in the near- and mid-future periods. Therefore, the 100-year event becomes more frequent for more than 50 % of the gauges with varying extent. While the magnitude of changes is similarly moderate (except for the far future) for Fig. 6c and e, projected future return periods for the HF₁₀₀ event for Fig. 6d depict stronger change signals towards higher frequencies with more than 50 % of gauges showing values smaller than 60 years. Furthermore, the nivo-pluvial, as well as the balanced and unbalanced pluvial regimes, exhibits a slight decrease in frequency in the middle future compared to the remaining projection periods, while the magnitude does not show this behavior. However, this circumstance may be explained by the change in the driving agent from snowmelt-driven events in the near future to rainfall-induced events at the end of the century. Thus, at the 2055 horizon, the shift in the ratio of both event types contributes to this slight decline in frequency.

Some gauges within the nivo-pluvial and both pluvial regimes depict an, in part, large decrease in frequency and/or magnitude. These gauges usually exhibit natural or artificial influences, such as the retention effect of natural lakes, reservoirs, or diversions or gauges of small catchments which might experience fewer dynamics in the changing of flood drivers or even a reduction.

Overall, the changes in frequency and magnitude due to the projected changes in climate according to the CRCM5-LE become less severe with increasing distance from the Alps. Furthermore, the increase in frequency and magnitude for alpine catchments is seemingly high but is in line with the results of Hattermann et al. (2018), which showed comparable results for the near-future period (100-year event frequency between 20 and 40 years). The influencing factors for these sometimes severe changes are manifold. However, Brunner et al. (2021b) analyzed the relation between the extremeness of precipitation and discharge for 78 out of the 98 gauges within hydrological Bavaria and concluded that an increase in extreme-precipitation magnitude is of higher importance for extreme return levels of discharge than land surface processes, such as antecedent soil moisture or changes in snowpack due to warmer temperatures. If precipitation volumes are sufficiently large, they quickly saturate the soil or yield an excessive amount of direct runoff due to infiltration excess (Brunner et al., 2021b).

The mean magnitude of the annual maximum precipitation is projected to change for different temporal aggregation levels (3-hourly to 5-daily) in the CRCM5-LE (Wood and Ludwig, 2020); in addition, the magnitude of the 100-year return period rainfall increases by 10 %–20 %, and the frequency increases by 2 to 4 times (Martel et al., 2020) for hydrological Bavaria. The changes are associated with seasonal shifts from summer to winter events and are particularly pronounced in the alpine region (Martel et al., 2020; Wood and Ludwig, 2020). Severe floods that occur simultaneously in

different catchments of the study area are usually associated with a cutoff-low Vb cyclone that results in prolonged precipitation events lasting up to 15 d over the same region (Stahl and Hofstätter, 2018; Mittermeier et al., 2019). Under the changing climate conditions projected by the CRCM5-LE by the end of the 21st century employing the RCP8.5 scenario, these events are likely to intensify in volume and frequency during winter and spring and are likely to occur less frequently during the summer months but with an increased precipitation volume (Mittermeier et al., 2019).

The spatial distribution of the dynamics in HF₁₀₀ frequency and magnitude is shown in the Supplement (Sect. S6).

3.3 Changes in driving mechanisms

Figure 7 shows the dynamics of three different combinations of driving mechanisms (columns of panels) listed in Table 1 for extreme-discharge events equal to and above HF₁₀₀ between the reference period (REF) and the three future periods (FUT1 to FUT3) for the five different discharge regime types (rows of panels). As mentioned in the “Data and methods” section, the 32 possible combinations were aggregated to six groups, as listed in Table 2.

The first column of panels in Fig. 7 shows the changes in the ratio between snowmelt-driven events (excluding rainfall compounds) and rainfall-driven events (including melt compounds). Figure 7f–j show the changes in the ratio between heavy-rainfall and steady-rain events (excluding snowmelt events from Fig. 7a–e), and Fig. 7k–o depict the changes in the ratio of events that can be attributed to a single cause and to a compound of drivers.

The glacio-nival and the nival regimes show the highest ratio of snowmelt-driven events (Fig. 7a and b). For the reference period, this event type is the major driver for the extreme discharges for the glacio-nival regime (67.7 %), while, for all other regimes and periods, rainfall and its compounds dominate the ratio. Figure 7a to e also show a seventh category of “other”, which, in this case, comprises events which could not be ascribed to any of the investigated drivers and are likely events that originate from an upstream flood. All regimes indicate a decrease in snowmelt-driven event types from the REF period to the far-future (FUT3) period, with the largest decrease being visible for the glacio-nival regime (67.7 % to 20.6 %). In the nivo-pluvial and pluvial regimes (Fig. 7c to e), the ratio of snowmelt event types becomes negligible in the future, with values of only slightly above 0 %. Furthermore, the ratio of snowmelt events diminishes from the glacio-nival and nival regime to the pluvial regimes (Fig. 7c to e) with increasing distance to the Alps. The pluvial regimes (Fig. 7d and e) show only minor changes in the ratio of rainfall events from REF to FUT3 (89.3 % to 80.6 % for the unbalanced regime and 84.6 % to 87.5 % for the balanced regime). The disappearance of snowmelt-driven events in future periods indicates that more events are driven fac-

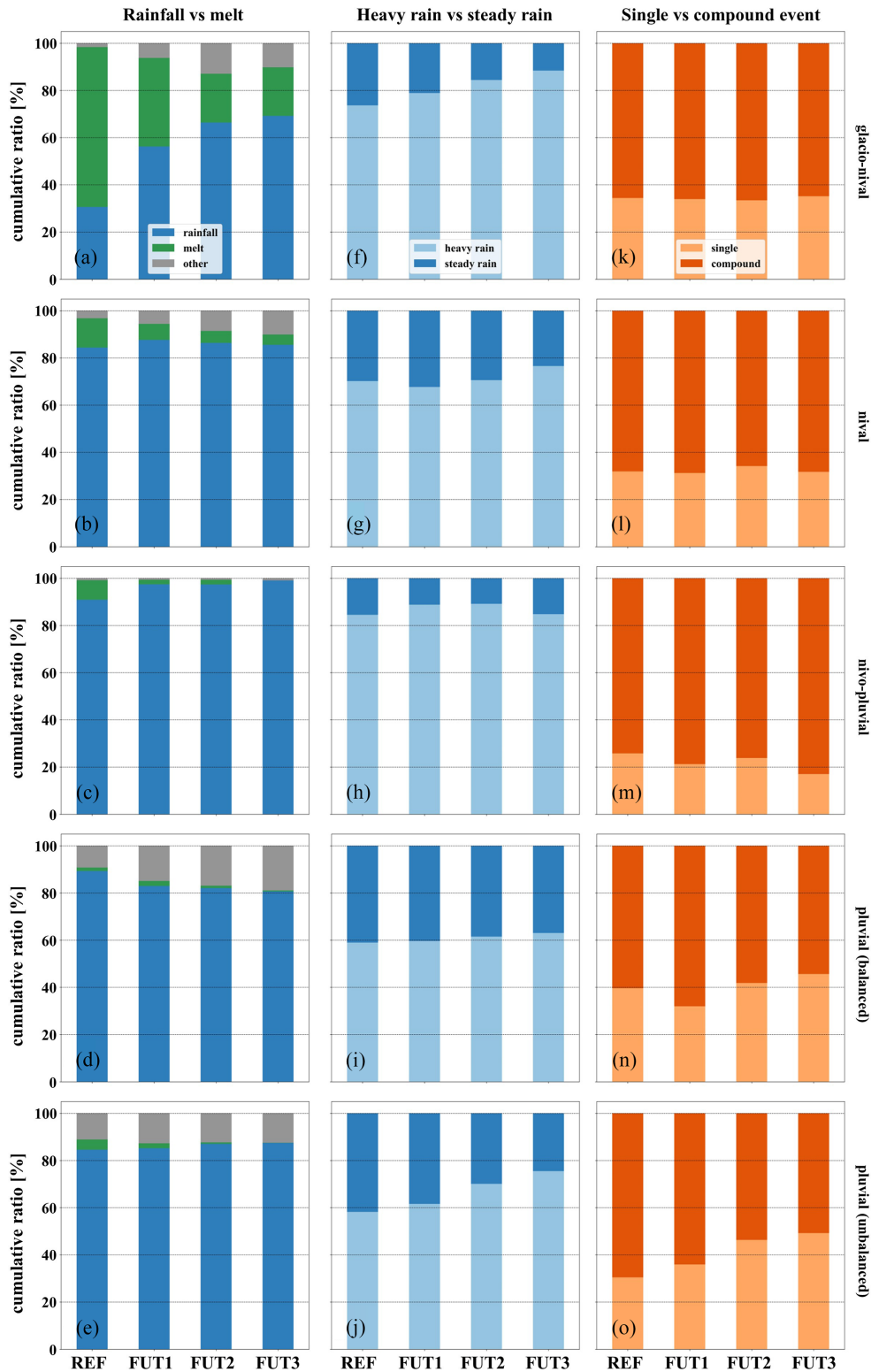


Figure 7. Dynamics in driving mechanisms of floods equal to or larger than HF₁₀₀ for the five different discharge regimes and within the four different periods. The columns of panels show the composition of melt- vs. rainfall-driven events (left), heavy-rainfall vs. steady-rainfall events (center), and single vs. compound events (right). The rows of panels correspond to the five different discharge regimes, as indicated by the regime type on the right y axis. The bars in each individual panel show the cumulative ratio of the different mechanisms for the reference period (REF, 1981–2010) in the first bar, and the following bars represent the different future periods (FUT1: 2021–2040; FUT2: 2041–2070; FUT3: 2070–2099).

tors falling in the category of other, with an increase in this category being seen especially in the nival regime (3.2 % to 10.1 %) and the unbalanced pluvial regime (9.2 % to 18.9 %). Overall, these results indicate a reduction in snow accumulation during the winter due to an increase in wintery temperatures of between 3 and 5 °C, as projected by the driving CRCM5-LE for the end of the century (FUT3) for middle Europe (von Trentini et al., 2020). For the glacio-nival and nival regimes, the large decrease in the snowmelt ratio also indicates a reduced contribution of melt from glaciers due to severe loss of mass towards the far future.

Figure 7f–j illustrate the dynamics in the ratio of hr and sr event types, therefore representing only the rainfall part of Fig. 7a to e. In all five regime types, the hr event type and its compounds (including compounds with sr) are the dominant driver, with a ratio of 58.8 % in the pluvial unbalanced regime (REF) to 88.4 % in the glacio-nival regime (FUT3). The percentage of hr events increases towards the end of the century for all regimes by 4 (pluvial balanced) to 17.2 (pluvial unbalanced) percent points, except for the pluvio-nival regime, where the percentages first increase from REF in FUT1 (88.9 %) and FUT2 (89.2 %) and then decrease again for FUT3 to the level of REF (REF: 84.4 %, FUT3: 84.8 %). The glacio-nival and pluvial unbalanced regimes show the strongest increase in hr event types from REF to FUT3, with 14.7 and 17.2 percent points, respectively. The dynamics for these regimes may be caused by an increase in summer temperatures (between 5 and 6.5 °C) but also in spring and fall temperatures towards the end of the century, as projected by the CRCM5-LE (von Trentini et al., 2020). The higher temperatures result in more available water vapor and, thus, precipitable water in the atmosphere and a higher potential for convective (hr) events, especially over the Alps (Giorgi et al., 2016). Furthermore, the strong increase for the alpine gauges of the glacio-nival regime may be related to a stronger increase in heavy-precipitation events over the Alps compared to regimes outside the Alps (Wood and Ludwig, 2020). In general, the balanced and unbalanced pluvial regimes show the lowest contribution of hr events (59 % and 58.3 % in REF); therefore, any change in the number of hr events or a general increase in the intensity of short-duration rainfall might lead to more events being classified as hr compounds. On the other hand, in other regimes, the hr compound is already large, and, hence, any changes in the rainfall dynamics will only yield a limited increase in the event classification.

Figure 7k–o illustrate the dynamics in the ratio between single-driver event types and compound event types for the five different regimes. Here, the compound class comprises the snowmelt and rainfall event types in Fig. 7a–e, neglecting events classified as other. Therefore, single-driver event types depict events caused by only one of the driving mechanisms listed in Table 1, whereas compound event types comprise all other possible combinations. In all five regimes and time periods, compound drivers are attributed to at least 50.7 % (FUT3, pluvial unbalanced) and up to 82.9 % (FUT3, nivo-

pluvial) of events. Except for the two pluvial regimes (balanced and unbalanced), there is very little change in the ratio. For these regimes, the number of events caused by a single driver increases from REF to FUT3 by 6 and 18.8 percent points for the balanced and unbalanced regimes, respectively. This strong signal in the dynamics for the unbalanced pluvial regime indicates an increase in short events of high intensity, which, in turn, may lead to a higher risk for flash floods. The nivo-pluvial regime further depicts a slight decrease of 8.8 percent points in events caused by a single driver from REF to FUT3.

4 Discussion

The variability of statistical characteristics within a time series can affect the estimation of extreme values due to extraordinary events (Fischer and Schumann, 2016). The results of this study emphasize the benefit of using data provided by a climatological SMILE for hydrological impact studies as this provides a profound basis for extreme value statistics and allows for more accurate estimation of extreme values, as also shown by other studies (van der Wiel et al., 2019; Champagne et al., 2020; Ehmele et al., 2020; Maher et al., 2021). In particular, van der Wiel et al. (2019) follow a similar approach for comparing statistical distributions and use an empirical approach to derive extreme-discharge values by employing a different approach for the creation of a 2000-year large ensemble of climate data forcing a global hydrological model. Our study shows similar results for the comparison between statistical estimates and empirically derived HF₁₀₀ values, favoring the empirical over the statistical approach when using a large ensemble of hydrological model data. However, the sometimes large deviations between the benchmark (robust estimate derived from the empirical probability for a 100-year flood event using 1500 AM values) and the estimates derived using a GEV based on different sample sizes (30, 100, 200) might be reduced when using an extreme value distribution (EVD), which is better suited to the respective sample when enough data are available (i.e., when using a hydro-SMILE, as shown here). In some cases, the GEV might not be the best distribution for the samples of the respective gauge, which might affect the differences from the benchmark since higher quantiles heavily depend on the distribution (Schulz and Bernhardt, 2016). Hence, a large ensemble of hydrological model data further reduces the uncertainties originating from different distributions, which may be considerable, as shown by Lawrence (2020). However, the approach presented in this study illustrates the benefit of a hydro-SMILE as it provides a more robust estimate by employing empirical probabilities for the deduction of extreme values. Therefore, these robust estimates allow for a more robust assessment of future dynamics of extremely high flows and their driving mechanisms, as also shown in Brunner et al. (2021b).

The results of this study are subject to uncertainties (parameters, process description) as they are produced by data created at the end of a cascade of modeling steps usually applied for climate change impact studies, as displayed in Fig. 2. Different components (e.g., climate model, hydrological model, bias correction) affect different discharge characteristics or indicators (e.g., extreme indicators, mean discharge) (Gampe et al., 2019; Muerth et al., 2012, 2013; Velázquez et al., 2013; Willkofer et al., 2018). A thorough assessment of the contribution of the chain compartments to the overall uncertainty would require an ensemble of multiple climate and hydrological models.

The overall strong increase in the frequency and magnitude of HF₁₀₀ in the future may be driven by deficiencies in the employed hydrological model, such as the generalized glacier model among affected catchments, or the single snowmelt approach used for the entirety of hydrological Bavaria (as described in Willkofer et al., 2020). However, as stated in the previous section, this scale of change was also found by Hattermann et al. (2018) for the upper Danube basin using the same emission scenario projections but a different hydrological and climate model, which might indicate that the change signals are likely independent of the chosen hydrological or climate model. However, several gauges north of the Alps exhibit a decrease in the frequency and magnitude of HF₁₀₀ over the three different future periods compared to the reference period. As mentioned, these gauges are, in parts, affected by artificial or natural retention (e.g., reservoirs) or transfer systems which are implemented in the model, and this may influence the results. Additionally, despite the projected increase in extreme rainfall events of the CRCM5-LE, even north of the Alps (Wood and Ludwig, 2020), the non-linear behavior of the processes involved in runoff generation may not translate this increase into extreme-discharge events (Brunner et al., 2021a). Furthermore, this increase in extreme rainfall events is less severe north of the Alps (Wood and Ludwig, 2020), which may further contribute to the decline or minor increase in the frequency and magnitude of the HF₁₀₀ events.

The results of the CCI in relation to the frequency and magnitude also depend on the performance of the hydrological model. Since the performance is influenced by observations used for parameter calibration, the quality of these data is crucial, especially for extreme values. For the most extreme events (e.g., HF₁₀₀ and above), the river may inundate the surrounding area, and the water level–discharge relationship at the gauging station used to determine discharge values may not be valid anymore and is likely to underestimate the peak discharge. Therefore, the actual observed discharge – and, thus, the calibrated model – is prone to these measurement uncertainties. This is a general limitation in hydrological modeling. Furthermore, the discharge of rivers within hydrological Bavaria is heavily impacted by management structures for flood protection or hydro-power generation; in particular, the southern tributaries of the Danube in the Alpine

Foreland and within the Alps are heavily regulated. Since the management follows somewhat fuzzy rules and because actual data are restricted by private companies in most cases, the management rules for these structures must be deduced from publicly available data and implemented in the hydrological model. These rules are susceptible to extreme conditions as they do not allow for adaptations during model runtime (e.g., flushing a reservoir prior to an anticipated heavy-precipitation event).

The projected future changes in extreme discharges may be attributed, in part, to the climatological reference dataset as it affects the performance of the hydrological model, as well as the climate change signal, through bias adjustment (Gampe et al., 2019; Meyer et al., 2019; Willkofer et al., 2018). Precipitation in high altitudes (e.g., the Alps) may be under-captured (Westra et al., 2014; Poschlod, 2021; Prein and Gobiet, 2017; Rauthe et al., 2013; Poschlod et al., 2020; Willkofer et al., 2020), resulting in an underestimation of observed precipitation in these regions, especially when it comes to extreme values. Assuming a temporally stationary bias, changes in the extremes might be overestimated due to an over-adjustment of the distribution of the reference period towards underestimated observations compared to the future periods. Furthermore, the variables are adjusted individually, and, thus, physical coherency, as for a multivariate approach proposed by Meyer et al. (2019), is not guaranteed. This specifically affects discharges governed by snowmelt or glacier melt of higher elevation within the Alps (Meyer et al., 2019).

The analysis of the dynamics in driving mechanisms of extreme discharges of HF₁₀₀ and above involved a set of thresholds for several parameters (rainfall, snowmelt and glacier melt, soil water content), as well as their combinations. Thresholds other than those selected according to the DWD (Deutscher Wetterdienst, 2024) may yield different ratios of the illustrated dynamics. The different drivers and their different combinations considered for this analysis could have been aggregated to other overarching categories (e.g., showing the contribution of soil moisture). However, we opted for the illustrated aggregations as changes in the extremeness of these events directly translate into changes in the extremeness of the discharge (Brunner et al., 2021a). Furthermore, the analysis only focuses on discharge events which are above the benchmark 100-year flood event calculated for the reference period. Hence, for gauges depicting a decrease in HF₁₀₀ frequency and magnitude in future periods, events resulting from the changes in future return levels are not considered here. However, it is unlikely that the overall dynamics gained from this approach might considerably change by applying the future HF₁₀₀ values as a threshold to extract the events.

Since the presented modeling approach only comprises one GCM–RCM combination forced by the more extreme RCP8.5 emission scenario, as well as one hydrological model, the significance of the findings regarding the vari-

ance of change effects in the future in the development of extreme peak flows is limited. Furthermore, the projected climate change signals of the CRCM5-LE were found to depict a stronger warming and drying compared to other large ensembles (von Trentini et al., 2020), which might result in these partly extreme increases in the frequency and magnitude of the HF₁₀₀ values among many gauges of hydrological Bavaria.

Projected discharge extremes at the upper end of the distribution that have not been observed to date might be created by unrealistic compound events due to flaws in the bias correction approach (Kelder et al., 2022). Thus, these events directly influence the EVD, producing higher return values and, consequently, a larger change signal. However, as extreme-precipitation events of various durations are expected to intensify within the studied region, the probability of yet unseen floods due to compounding events may also increase in the future.

5 Conclusion

This study emphasizes the benefit of employing a climatological SMILE with a hydrological model to create a hydro-SMILE to foster extreme value statistics and to analyze the impacts of climate change on hydrological extreme values such as HF₁₀₀ due to the provision of a very large database. This database allows for the application of empirical exceedance probabilities to estimate robust discharge values of high return periods rather than statistical extrapolation based on extreme value distributions. The results show that the performance of statistical estimates largely depends on the available length of the time series and their values when compared to the empirical benchmark. However, even with a length of 200 AM, the variance of the scatter of the HF₁₀₀ estimates of the 1000 samples was rather large.

As mentioned by Willkofer et al. (2020), the performance of the hydrological model allows for CCI studies – in this case, using the CRCM5-LE to elaborate on the effects of climate change on the development of HF₁₀₀. The projections reveal a strong increase in the magnitude and frequency of HF₁₀₀ events for alpine and pre-alpine catchments exhibiting a snowmelt-driven discharge regime within the reference period. This strong increase in the magnitude and frequency is considerably smaller for catchments north of the Alps and of a more pluvial discharge regime. The sometimes tremendous changes in HF₁₀₀ intensities and frequencies may be ascribed to the emission scenario (RCP8.5). Thus, the addition of different SMILEs and hydrological models may foster the significance of the findings due to different climate projections and simulated climatological and hydrological processes along the model chain. However, the establishment of such extensive model chains requires vast computational resources. Nevertheless, this effort should be considered in light of the benefits this profound database offers for extreme

value statistics, fostering knowledge about the propagation of the natural variability of the climate system to the hydrological response (Brunner et al., 2021b) or allowing us to distinguish climate change signals (or forced responses) from natural variability for extreme values (Wood and Ludwig, 2020; Aalbers et al., 2018).

This study further shows the benefit of a hydro-SMILE when driven by a processed-based hydrological model as it allows for a more detailed analysis on the processes responsible for the genesis of such extreme discharges. Within the study area, extreme-discharge events larger than HF₁₀₀ are less likely to be caused by snowmelt events in the future as higher wintery temperatures will result in less snow accumulation. Hence, rainfall becomes the dominant driver in the future. Further, those events are more likely to be caused by heavy rainfall compared to steady rainfall in the future, although the degree in dynamics may vary for the different regimes. While compound events of superimposing drivers might remain the major cause for discharges equal to and greater than the 100-year flood, the number of events caused by a single driver such as heavy rainfall is likely to increase in the future, at least for the two pluvial regimes.

Furthermore, the results highlight the need to incorporate climate projections in the design of new flood protection infrastructures or to adapt existing structures to reduce future flood risk, not only in hydrological Bavaria but everywhere in general. Further studies that focus on flood inundation are necessary to fully analyze the extent of the increase and frequency of this event for the design of flood protection infrastructures.

Code availability. The source code of the hydrological model WaSiM is not publicly available. An executable for general purposes can be downloaded at http://wasim.ch/en/products/wasim_richards.htm (Schulla, 2023, 2021). The code used to process the data and to produce the figures can be requested from the first author.

Data availability. The CRCM5-LE data for the historical and RCP8.5 simulations are available at <https://www.climex-project.org/data-access/> (Ouranos, 2020). The European Digital Elevation Model (EU-DEM) dataset is available at European Environment Agency (2013, <https://www.eea.europa.eu/data-and-maps/data/eu-dem>). The CORINE Land Cover 2006 dataset is available at European Environment Agency (2019, <https://doi.org/10.2909/08560441-2fd5-4eb9-bf4c-9ef16725726a>). The CRCM5-LE pre-industrial control simulations as well as the data of the WaSiM-LE are available upon reasonable request.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/hess-28-2969-2024-supplement>.

Author contributions. FW developed the concept of this study, including the methods, investigation, and visualization; performed the formal analysis and validation; and prepared the original draft. FW and RRW were responsible for data curation and software development. RL acquired the funding for the presented research, provided the required resources, was responsible for the project administration, and supervised the presented research.

Competing interests. The contact author has declared that none 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. The authors acknowledge our colleagues, Gilbert Brietzke, André Kurzmann, and Jens Weismüller, from the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities for their technical support during the development of the hydrological large ensemble. We also thank the Leibniz Supercomputing Centre for providing the HPC infrastructure and computation time.

The production of ClimEx was funded within the ClimEx project by the Bavarian State Ministry for the Environment and Consumer Protection. The CRCM5 was developed by the ESCER center of the Université du Québec à Montréal (UQAM) in collaboration with Environment and Climate Change Canada. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis for executing and making available the CanESM2 Large Ensemble simulations used in this study, and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. Computations with the CRCM5 for the ClimEx project were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer is funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science and the Arts.

Financial support. This research has been supported by the Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (grant no. 81-0270-024570/2015).

Review statement. This paper was edited by Rohini Kumar and reviewed by three anonymous referees.

References

- Aalbers, E. E., Lenderink, G., van Meijgaard, E., and van den Hurk, B. J. J. M.: Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?, *Clim. Dynam.*, 50, 4745–4766, <https://doi.org/10.1007/s00382-017-3901-9>, 2018.
- Addor, N. and Fischer, E. M.: The influence of natural variability and interpolation errors on bias characterization in RCM simulations, *J. Geophys. Res.*, 120, 10180–10195, <https://doi.org/10.1002/2014JD022824>, 2015.
- Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, *Geophys. Res. Lett.*, 38, L05805, <https://doi.org/10.1029/2010GL046270>, 2011.
- Bertola, M., Viglione, A., Lun, D., Hall, J., and Blöschl, G.: Flood trends in Europe: are changes in small and big floods different?, *Hydrol. Earth Syst. Sci.*, 24, 1805–1822, <https://doi.org/10.5194/hess-24-1805-2020>, 2020.
- Bezák, N., Brilly, M., and Šraj, M.: Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis, *Hydrolog. Sci. J.*, 59, 959–977, <https://doi.org/10.1080/02626667.2013.831174>, 2014.
- BGR: International Hydrogeological Map of Europe 1 : 1 500 000 (IHME1500 v1.1), Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover, Germany, Paris, France, https://download.bgr.de/bgr/grundwasser/IHME1500/v12/shp/IHME1500_v12.zip (last access: 17 August 2023), 2014.
- Blöschl, G., Nester, T., Komma, J., Parajka, J., and Perdigo, R. A. P.: The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods, *Hydrol. Earth Syst. Sci.*, 17, 5197–5212, <https://doi.org/10.5194/hess-17-5197-2013>, 2013.
- Blöschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigo, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., and Viglione, A.: Increasing river floods: fiction or reality?, *WIREs Water*, 2, 329–344, <https://doi.org/10.1002/wat2.1079>, 2015.
- Blöschl, G., Hall, J., Viglione, A., Perdigo, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Bohâe, M., Bonacci, O., Borga, M., Èanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živkoviæ, N.: Changing climate both increases and decreases European river floods, *Nature*, 573, 108–111, <https://doi.org/10.1038/s41586-019-1495-6>, 2019.
- Brönnimann, S., Rajczak, J., Fischer, E. M., Raible, C. C., Rohrer, M., and Schär, C.: Changing seasonality of moderate and extreme precipitation events in the Alps, *Nat. Hazards Earth Syst. Sci.*, 18, 2047–2056, <https://doi.org/10.5194/nhess-18-2047-2018>, 2018.
- Brunner, M. I., Slater, L., Tallaksen, L. M., and Clark, M.: Challenges in modeling and predicting floods and droughts: A review,

- WIREs Water, 8, e1520, <https://doi.org/10.1002/wat2.1520>, 2021a.
- Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E., and Ludwig, R.: An extremeness threshold determines the regional response of floods to changes in rainfall extremes, *Commun. Earth Environ.*, 2, 173, <https://doi.org/10.1038/s43247-021-00248-x>, 2021b.
- Černý, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *J. Optim. Theory Appl.*, 45, 41–51, <https://doi.org/10.1007/BF00940812>, 1985.
- Champagne, O., Leduc, M., Coulibaly, P., and Arain, M. A.: Winter hydrometeorological extreme events modulated by large-scale atmospheric circulation in southern Ontario, *Earth Syst. Dynam.*, 11, 301–318, <https://doi.org/10.5194/esd-11-301-2020>, 2020.
- Chen, J., Arsenault, R., Brissette, F. P., and Zhang, S.: Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs?, *Water Resour. Res.*, 57, e2020WR028638, <https://doi.org/10.1029/2020WR028638>, 2021.
- Dettinger, M. D., Cayan, D. R., Meyer, M. K., and Jeton, A. E.: Simulated Hydrologic Responses to Climate Variations and Change in the Merced, Carson, and American River Basins, Sierra Nevada, California, 1900–2099, *Climatic Change*, 62, 283–317, <https://doi.org/10.1023/B:CLIM.0000013683.13346.4f>, 2004.
- Deutscher Wetterdienst: Wetter und Klima – Deutscher Wetterdienst – Warnungen aktuell – Warnkriterien, https://www.dwd.de/DE/wetter/warnungen_aktuell/kriterien/warnkriterien.html (last access: 15 June 2023), 2024.
- Dörhöfer, G., Hannappel, S., and Voigt, H.-J.: Die Hydrogeologische Übersichtskarte von Deutschland HÜK200, *Z. Angew. Geol.*, 47, 153–159, 2001.
- Ehmele, F., Kautz, L.-A., Feldmann, H., and Pinto, J. G.: Long-term variance of heavy precipitation across central Europe using a large ensemble of regional climate model simulations, *Earth Syst. Dynam.*, 11, 469–490, <https://doi.org/10.5194/esd-11-469-2020>, 2020.
- Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions “Should we apply bias correction to global and regional climate model data?”, *Hydrol. Earth Syst. Sci.*, 16, 3391–3404, <https://doi.org/10.5194/hess-16-3391-2012>, 2012.
- European Commission and the European Soil Bureau Network: The European Soil Database distribution version 2.0, CD-ROM, EUR 19945 EN, 2004.
- European Environment Agency: European Digital Elevation Model (EU-DEM), European Environment Agency [data set], <https://www.eea.europa.eu/data-and-maps/data/eu-dem> (last access: 18 July 2023), 2013.
- European Environment Agency: CORINE Land Cover 2006 (raster 100m), Europe, 6-yearly – version 2020_20u1, European Environment Agency [data set], <https://doi.org/10.2909/08560441-2fd5-4eb9-bf4c-9ef16725726a>, 2019.
- Fischer, S. and Schumann, A.: Robust flood statistics: comparison of peak over threshold approaches based on monthly maxima and TL-moments, *Hydrolog. Sci. J.*, 61, 457–470, <https://doi.org/10.1080/02626667.2015.1054391>, 2016.
- Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J., and Jiao, Y.: Large near-term projected snowpack loss over the western United States, *Nat. Commun.*, 8, 14996, <https://doi.org/10.1038/ncomms14996>, 2017.
- Gamepe, D., Schmid, J., and Ludwig, R.: Impact of Reference Dataset Selection on RCM Evaluation, Bias Correction, and Resulting Climate Change Signals of Precipitation, *J. Hydrometeorol.*, 20, 1813–1828, <https://doi.org/10.1175/JHM-D-18-0108.1>, 2019.
- Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, *J. Stat. Soft.*, 72, 1–39, <https://doi.org/10.18637/jss.v072.i08>, 2016.
- Giorgi, F., Torma, C., Coppola, E., Ban, N., Schär, C., and Somot, S.: Enhanced summer convective rainfall at Alpine high elevations in response to climate warming, *Nat. Geosci.*, 9, 584–589, <https://doi.org/10.1038/ngeo2761>, 2016.
- GRDC: Global Runoff Data Centre, Summary Statistics by Country, Bundesanstalt für Gewässerkunde, https://grdc.bafg.de/SharedDocs/ExterneLinks/GRDC/summary_stat_cc_pdf.pdf?__blob=publicationFile (last access: 13 May 2023), 2021.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- Hattermann, F. F., Wortmann, M., Liersch, S., Toumi, R., Sparks, N., Genillard, C., Schröter, K., Steinhausen, M., Gyalai-Korpos, M., Máté, K., Hayes, B., del Rocio Rivas López, M., Rácz, T., Nielsen, M. R., Kaspersen, P. S., and Drews, M.: Simulation of flood hazard and risk in the Danube basin with the Future Danube Model, *Climate Services*, 12, 14–26, <https://doi.org/10.1016/j.cliser.2018.07.001>, 2018.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, D20119, <https://doi.org/10.1029/2008JD010201>, 2008.
- Huang, S., Krysanova, V., and Hattermann, F. F.: Does bias correction increase reliability of flood projections under climate change? A case study of large rivers in Germany, *Int. J. Climatol.*, 34, 3780–3800, <https://doi.org/10.1002/joc.3945>, 2014.
- Iacob, O., Brown, I., and Rowan, J.: Natural flood management, land use and climate change trade-offs: the case of Tarland catchment, Scotland, *Hydrolog. Sci. J.*, 62, 1931–1948, <https://doi.org/10.1080/02626667.2017.1366657>, 2017.
- IPCC: Summary for Policymakers, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3–32, <https://doi.org/10.1017/9781009157896.001>, 2021.
- Jónsdóttir, J. F.: A runoff map based on numerically simulated precipitation and a projection of future runoff in Iceland/Une carte d’écoulement basée sur la précipitation numériquement simulée et un scénario du futur écoulement en Islande, *Hydrolog. Sci. J.*, 53, 100–111, <https://doi.org/10.1623/hysj.53.1.100>, 2008.

- Kelder, T., Wanders, N., van der Wiel, K., Marjoribanks, T. I., Slater, L. J., Wilby, R. I., and Prudhomme, C.: Interpreting extreme climate impacts from large ensemble simulations – are they unseen or unrealistic?, *Environ. Res. Lett.*, 17, 44052, <https://doi.org/10.1088/1748-9326/ac5cf4>, 2022.
- Kendall, M. G.: Rank correlation methods, 2nd edn., Griffin, London, 1955.
- Kendon, E. J., Rowell, D. P., Jones, R. G., and Buonomo, E.: Robustness of Future Changes in Local Precipitation Extremes, *J. Climate*, 21, 4280–4297, <https://doi.org/10.1175/2008JCLI2082.1>, 2008.
- Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of Extreme Events in Arctic Sea Ice Extent, *J. Climate*, 30, 553–571, <https://doi.org/10.1175/JCLI-D-16-0412.1>, 2017.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by simulated annealing, *Science*, 220, 671–680, <https://doi.org/10.1126/science.220.4598.671>, 1983.
- Kjellström, E., Thejll, P., Rummukainen, M., Christensen, J. H., Boberg, F., Christensen, O. B., and Fox Maule, C.: Emerging regional climate change signals for Europe under varying large-scale circulation conditions, *Clim. Res.*, 56, 103–119, <https://doi.org/10.3354/cr01146>, 2013.
- Lawrence, D.: Uncertainty introduced by flood frequency analysis in projections for changes in flood magnitudes under a future climate in Norway, *J. Hydrol.-Regional Studies*, 28, 100675, <https://doi.org/10.1016/j.ejrh.2020.100675>, 2020.
- Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Britzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), *J. Appl. Meteorol. Climatol.*, 58, 663–693, <https://doi.org/10.1175/JAMC-D-18-0021.1>, 2019.
- Ludwig, R., Wood, Raul, R., Willkofer, F., von Trentini, F., Mittermeier, M., Böhnisch, A., and Poschlod, B.: *ClimEx – Klimawandel und Extremereignisse: Risiken und Perspektiven für die bayerische Wasserwirtschaft, Abschlussbericht, Ludwig-Maximilians-Universität*, 190 pp., 2019.
- Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, *Earth Syst. Dynam.*, 12, 401–418, <https://doi.org/10.5194/esd-12-401-2021>, 2021.
- Mann, H. B.: Nonparametric Tests Against Trend, *Econometrica*, 13, 245–259, <https://doi.org/10.2307/1907187>, 1945.
- Maraun, D.: Bias Correcting Climate Change Simulations – a Critical Review, *Curr. Clim. Change Rep.*, 2, 211–220, <https://doi.org/10.1007/s40641-016-0050-x>, 2016.
- Marke, T.: Development and Application of a Model Interface to couple Land Surface Models with Regional Climate Models for Climate Change Risk Assessment in the Upper Danube Watershed, Ludwig-Maximilians-Universität München, <https://doi.org/10.5282/edoc.9162>, 2008.
- Martel, J.-L., Mailhot, A., Brissette, F., and Caya, D.: Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales, *J. Climate*, 31, 4241–4263, <https://doi.org/10.1175/JCLI-D-17-0282.1>, 2018.
- Martel, J.-L., Mailhot, A., and Brissette, F.: Global and Regional Projected Changes in 100-yr Subdaily, Daily, and Multiday Precipitation Extremes Estimated from Three Large Ensembles of Climate Simulations, *J. Climate*, 33, 1089–1103, <https://doi.org/10.1175/JCLI-D-18-0764.1>, 2020.
- Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., and Dugas, B.: Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation, *Clim. Dynam.*, 41, 2973–3005, <https://doi.org/10.1007/s00382-013-1778-9>, 2013.
- Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, *Hydrol. Earth Syst. Sci.*, 23, 1339–1354, <https://doi.org/10.5194/hess-23-1339-2019>, 2019.
- Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y., and Ludwig, R.: Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning, *Geophys. Res. Lett.*, 46, 14653–14661, <https://doi.org/10.1029/2019GL084969>, 2019.
- Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, *T. ASABE*, 50, 885–900, <https://doi.org/10.13031/2013.23153>, 2007.
- Mpelasoka, F. S. and Chiew, F. H. S.: Influence of Rainfall Scenario Construction Methods on Runoff Projections, *J. Hydrometeorol.*, 10, 1168–1183, <https://doi.org/10.1175/2009JHM1045.1>, 2009.
- Muerth, M., Gauvin St-Denis, B., Ludwig, R., and Caya, D.: Evaluation of different sources of uncertainty in climate change impact research using a hydro-climatic model ensemble, International Congress on Environmental Modelling and Software, 1–5 July 2012, Leipzig, Germany, Universitätsbibliothek der Ludwig-Maximilians-Universität München, <https://doi.org/10.5282/ubm/epub.14094>, 2012.
- Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velázquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R.: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, *Hydrol. Earth Syst. Sci.*, 17, 1189–1204, <https://doi.org/10.5194/hess-17-1189-2013>, 2013.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Neukum, C. and Azzam, R.: Impact of climate change on groundwater recharge in a small catchment in the Black Forest, Germany, *Hydrogeol. J.*, 20, 547–560, <https://doi.org/10.1007/s10040-011-0827-x>, 2012.
- Ouranos: CRCM5-LE ClimEx [data set], <https://www.climex-project.org/data-access/> (last access: 14 August 2023), 2020.
- Panagos, P.: The European soil database, in: GEO: connexion, 5, 32–33, <https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data> (last access: 17 August 2023), 2006.
- Poschlod, B.: Using high-resolution regional climate models to estimate return levels of daily extreme precipitation over

- Bavaria, *Nat. Hazards Earth Syst. Sci.*, 21, 3573–3598, <https://doi.org/10.5194/nhess-21-3573-2021>, 2021.
- Poschlod, B., Willkofer, F., and Ludwig, R.: Impact of Climate Change on the Hydrological Regimes in Bavaria, *Water*, 12, 1599, <https://doi.org/10.3390/w12061599>, 2020.
- Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, *Int. J. Climatol.*, 37, 305–327, <https://doi.org/10.1002/joc.4706>, 2017.
- Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), *Meteorol. Z.*, 22, 235–256, <https://doi.org/10.1127/0941-2948/2013/0436>, 2013.
- Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R.: Regional parent flood frequency distributions in Europe – Part I: Is the GEV model suitable as a pan-European parent?, *Hydrol. Earth Syst. Sci.*, 18, 4381–4389, <https://doi.org/10.5194/hess-18-4381-2014>, 2014.
- Schulla, J.: Model Description WaSiM (Water balance Simulation Model), Zurich, http://wasim.ch/downloads/doku/wasim/wasim_2021_en.pdf (last access: 19 August 2023), 2021.
- Schulla, J.: WaSiM (Richards) [code], http://wasim.ch/en/products/wasim_richards.htm, last access: 15 July 2023.
- Schulz, K. and Bernhardt, M.: The end of trend estimation for extreme floods under climate change?, *Hydrol. Process.*, 30, 1804–1808, <https://doi.org/10.1002/hyp.10816>, 2016.
- Šeparovič, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., Tete, K., and Valin, M.: Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model, *Clim. Dynam.*, 41, 3167–3201, <https://doi.org/10.1007/s00382-013-1737-5>, 2013.
- Sigmond, M., Fyfe, J. C., and Swart, N. C.: Ice-free Arctic projections under the Paris Agreement, *Nat. Clim. Change*, 8, 404–408, <https://doi.org/10.1038/s41558-018-0124-y>, 2018.
- Stahl, N. and Hofstätter, M.: Vb-Zugbahnen und deren Auftreten als Serie mit Bezug zu den resultierenden Hochwassern in Bayern und Auswirkungen auf Rückhalteräume im Isareinzugsgebiet, *Hydrol. Wasserbewirts.*, 62, 77–97, https://doi.org/10.5675/HyWa_2018_2_2, 2018.
- Svensson, C., Kundzewicz, W. Z., and Maurer, T.: Trend detection in river flow series: 2. Flood and low-flow index series/Détection de tendance dans des séries de débit fluvial: 2. Séries d'indices de crue et d'étiage, *Hydrolog. Sci. J.*, 50, 811–824, <https://doi.org/10.1623/hysj.2005.50.5.811>, 2005.
- Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, 456–457, 12–29, <https://doi.org/10.1016/j.jhydrol.2012.05.052>, 2012.
- Thielen, A. H., Kienzler, S., Kreibich, H., Kuhlicke, C., Kunz, M., Mühr, B., Müller, M., Otto, A., Petrow, T., Pisi, S., and Schröter, K.: Review of the flood risk management system in Germany after the major flood in 2013, *Ecol. Soc.*, 21, 51, <https://doi.org/10.5751/ES-08547-210251>, 2016.
- Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, <https://doi.org/10.1029/2005WR004723>, 2007.
- van der Wiel, K., Wanders, N., Selten, F. M., and Bierkens, M. F. P.: Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World, *Geophys. Res. Lett.*, 46, 2093–2102, <https://doi.org/10.1029/2019GL081967>, 2019.
- van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, *Climatic Change*, 109, 5–31, <https://doi.org/10.1007/s10584-011-0148-z>, 2011.
- Velázquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R., and Turcotte, R.: An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources, *Hydrol. Earth Syst. Sci.*, 17, 565–578, <https://doi.org/10.5194/hess-17-565-2013>, 2013.
- von Trentini, F., Aalbers, E. E., Fischer, E. M., and Ludwig, R.: Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe, *Earth Syst. Dynam.*, 11, 1013–1031, <https://doi.org/10.5194/esd-11-1013-2020>, 2020.
- Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., and Roberts, N. M.: Future changes to the intensity and frequency of short-duration extreme rainfall, *Rev. Geophys.*, 52, 522–555, <https://doi.org/10.1002/2014RG000464>, 2014.
- Wilhelm, B., Rapuc, W., Amann, B., Anselmetti, F. S., Arnaud, F., Blanchet, J., Brauer, A., Czymzik, M., Giguet-Covex, C., Gilli, A., Glur, L., Grosjean, M., Irmeler, R., Nicolle, M., Sabatier, P., Swierczynski, T., and Wirth, S. B.: Impact of warmer climate periods on flood hazard in the European Alps, *Nat. Geosci.*, 15, 118–123, <https://doi.org/10.1038/s41561-021-00878-y>, 2022.
- Willkofer, F., Schmid, F.-J., Komischke, H., Korck, J., Braun, M., and Ludwig, R.: The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments, *J. Hydrol.-Regional Studies*, 19, 25–41, <https://doi.org/10.1016/j.ejrh.2018.06.010>, 2018.
- Willkofer, F., Wood, R. R., Trentini, F. von, Weismüller, J., Poschlod, B., and Ludwig, R.: A Holistic Modelling Approach for the Estimation of Return Levels of Peak Flows in Bavaria, *Water*, 12, 2349, <https://doi.org/10.3390/w12092349>, 2020.
- Wood, R. R. and Ludwig, R.: Analyzing Internal Variability and Forced Response of Subdaily and Daily Extreme Precipitation Over Europe, *Geophys. Res. Lett.*, 47, e2020GL089300, <https://doi.org/10.1029/2020GL089300>, 2020.
- Wood, R. R., Lehner, F., Pendergrass, A. G., and Schlunegger, S.: Changes in precipitation variability across time scales in multiple global climate model large ensembles, *Environ. Res. Lett.*, 16, 084022, <https://doi.org/10.1088/1748-9326/ac10dd>, 2021.