








## An extremeness threshold determines the regional response of floods to changes in rainfall extremes

Manuela I. Brunner<sup>1,2</sup>, Daniel L. Swain<sup>1,3,4</sup>, Raul R. Wood<sup>5</sup>, Florian Willkofer<sup>5</sup>, James M. Done<sup>1</sup>, Eric Gilleland<sup>1</sup> & Ralf Ludwig<sup>5</sup>

Precipitation extremes will increase in a warming climate, but the response of flood magnitudes to heavier precipitation events is less clear. Historically, there is little evidence for systematic increases in flood magnitude despite observed increases in precipitation extremes. Here we investigate how flood magnitudes change in response to warming, using a large initial-condition ensemble of simulations with a single climate model, coupled to a hydrological model. The model chain was applied to historical (1961–2000) and warmer future (2060–2099) climate conditions for 78 watersheds in hydrological Bavaria, a region comprising the headwater catchments of the Inn, Danube and Main River, thus representing an area of expressed hydrological heterogeneity. For the majority of the catchments, we identify a ‘return interval threshold’ in the relationship between precipitation and flood increases: at return intervals above this threshold, further increases in extreme precipitation frequency and magnitude clearly yield increased flood magnitudes; below the threshold, flood magnitude is modulated by land surface processes. We suggest that this threshold behaviour can reconcile climatological and hydrological perspectives on changing flood risk in a warming climate.

<sup>1</sup>National Center for Atmospheric Research, Boulder, CO, USA. <sup>2</sup>Institute of Earth and Environmental Sciences, University of Freiburg, Freiburg, Germany. <sup>3</sup>Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA. <sup>4</sup>The Nature Conservancy of California, San Francisco, CA, USA. <sup>5</sup>Department of Geography, Ludwig-Maximilians University Munich, Munich, Germany. ✉email: [manuela.brunner@hydrology.uni-freiburg.de](mailto:manuela.brunner@hydrology.uni-freiburg.de); [r.ludwig@lmu.de](mailto:r.ludwig@lmu.de)

There is clear theoretical, model-based, and empirical evidence that global precipitation extremes, i.e. precipitation exceeding a high threshold, will increase in a warming climate<sup>1–4</sup>. However, there is greatly more uncertainty regarding the hydrologic response regarding flooding and there is not yet clear evidence for widespread increases in flood occurrence either in observations<sup>5–10</sup> or in model simulations<sup>11–13</sup>. While there is still a theoretical expectation that flood events will increase in a warming climate<sup>14–17</sup>, and while such flood increases have been documented regionally<sup>18,19</sup>, the absence of broader observational trends supporting this hypothesis is conspicuous.

In the literature on hydrological processes, the lack of such trends is often attributed to changes in non-precipitation-flood drivers, such as temperature-driven decreases in snow accumulation and increases in evaporation that yield decreases in soil moisture<sup>9,20–23</sup>. Because of the compounding nature of different flood drivers, establishing a direct link between increases in extreme precipitation and increases in flooding is challenging<sup>24–26</sup>. Indeed, previous studies suggest that the strength of the relationship between precipitation and discharge may depend on a range of factors including catchment size, event magnitude<sup>25,27</sup>, and season<sup>28</sup> though the details of these complex relationships remain largely unknown and are hard to generalize.

Further complicating such investigations is the rarity of extreme events with long return intervals and their sparseness in observed precipitation and streamflow records. Several approaches have been proposed to address this data scarcity problem, including: pooling observations across different catchments<sup>29</sup> or seasonal predictive ensemble members<sup>30,31</sup>; tree-ring and historic reconstructions<sup>32,33</sup>; stochastic streamflow generation<sup>34,35</sup>; and ensemble modeling using Single Model Initial-condition Large Ensembles SMILEs<sup>36</sup>. To date, however, few studies have combined atmospheric SMILEs with hydrological models to obtain a SMILE of streamflow time series, i.e. a ‘hydro-SMILE’<sup>37–39</sup>. The availability of such a hydro-SMILE is crucial in assessing the relationship between future changes in extreme precipitation and flooding – particularly high-end extreme events (i.e., those occurring twice or fewer times per century), which are rare to nonexistent in observed time series.

Here, we seek to reconcile the extreme precipitation-flood paradox in a warming climate: is there a precipitation threshold beyond which increasing precipitation extremes directly translate into increasing flood risk? We hypothesize that such a threshold should exist because moderately extreme events may be buffered by decreased soil moisture (due to warming) while very extreme events may quickly lead to soil saturation and subsequently to direct translation of precipitation to runoff. Using a hydro-SMILE approach, we consider precipitation and flood characteristics from historical (1961–2000) and warmer future (2060–2099) climates for 78 catchments in major Bavarian river basins (Main, Danube, and the Inn river with their major tributaries; henceforth Hydrological Bavaria) characterized by a wide variety of hydroclimates, soil types, land uses, and streamflow regimes<sup>39,40</sup>. We find that there does indeed exist a catchment-specific extremeness threshold (i.e. return interval threshold) above which precipitation increases clearly yield increased flood magnitudes, and below which flood magnitude is strongly modulated by land surface processes such as soil moisture availability. Ultimately, this finding may help reconcile seemingly conflicting climatological and hydrological perspectives on changing flood risk in a warming climate.

Addressing the precipitation-flood paradox is simply not possible using observations alone, as the high-end extreme events of interest are rare to nonexistent in temporally limited observational records. This real-world data limitation effectively precludes statistical analyses of extreme events with return periods exceeding ~50 years. To overcome this problem, we use a hydro-SMILE to obtain a large number of extreme precipitation–streamflow pairs. The hydro-SMILE consists of hydrological simulations obtained by driving a

hydrological model with climate simulations from a single model initial-condition large ensemble (SMILE) climate model. The underlying model simulations were originally generated by Willkofer et al.<sup>40</sup> as part of the ClimEx project<sup>41</sup>. The hydro-SMILE simulations consist of daily streamflow ( $\text{mm d}^{-1}$ ), snow-water-equivalents (SWE, mm), and soil moisture (%) – all of which were obtained by driving the hydrological model WaSiM-ETH<sup>42</sup> with a 50-member ensemble of high-resolution climate input (spatial:  $500 \times 500 \text{ m}^2$ , temporal: 3 h) (for further information on the hydro-SMILE see Section “Hydro-SMILE”).

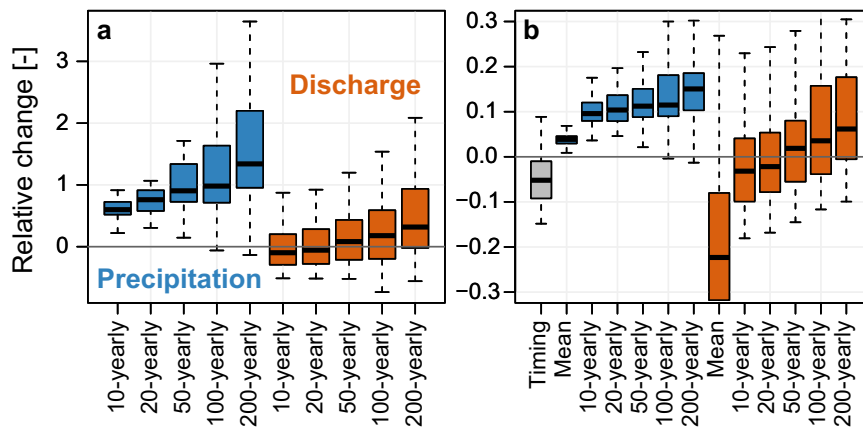
While such a large ensemble approach resolves the small or zero size problem for very extreme events, new sources of uncertainty do also arise. We acknowledge that the hydro-SMILE modeling chain is affected by uncertainties introduced through both the underlying climate and hydrological models. Climate model uncertainties include those relating to precipitation process-representation, downscaling, and bias-correction procedures, hydrological model uncertainties comprise model and parameter uncertainties. These latter uncertainties may be particularly relevant for the very extreme events under consideration in the present study because model calibration and evaluation rely upon observed events – and (as previously noted) modern observational records simply don’t exist for events of the extreme magnitudes considered here. However, we point out that this particular element of the overall uncertainty is essentially irreducible, and will likely remain so until the length of the observed record increases substantially some decades in the future. As such, the use of a hydro-SMILE is an appropriate method – and arguably the singular method available, at present – to comprehensively and quantitatively address the extreme precipitation-flood paradox.

## Results

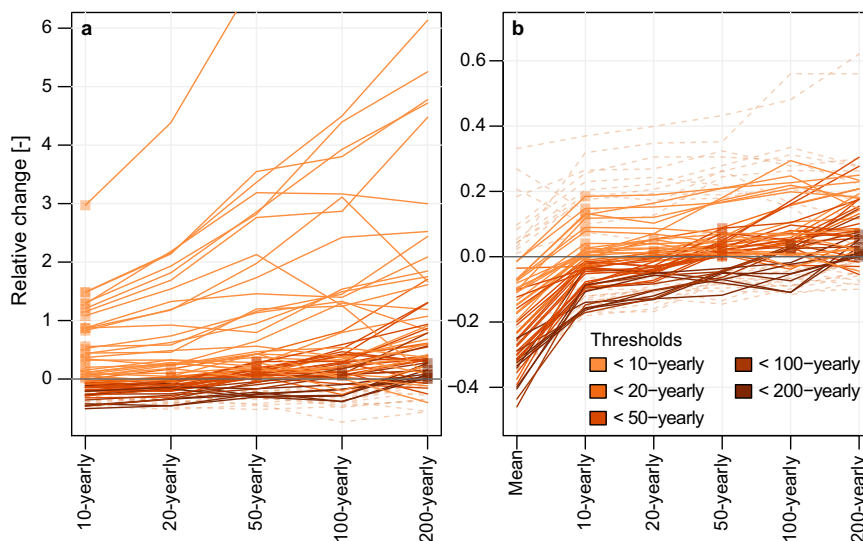
### Threshold behavior in flow response to extreme precipitation.

We first seek to assess whether there exists a return interval threshold beyond which precipitation ( $P$ ) increases consistently translate into streamflow ( $Q$ ) increases, and thereby to increases in flood magnitude. To do so, we use a hydro-SMILE consisting of a 50-member ensemble of 3-hourly precipitation and streamflow time series for Hydrological Bavaria (see Methods section “Study region” and Supplementary Figure 1), which we aggregated to daily resolution. The hydro-SMILE was derived for the period 1961–2099 by combining the Canadian Regional Climate Model large ensemble CRCM5-LE<sup>41</sup> with the hydrological model WaSiM-ETH<sup>40,42</sup> (see Methods sections “Hydro-SMILE” and “Hydrological model evaluation”). From this ensemble, we extract precipitation–discharge ( $P - Q$ ) pairs for a historical (1961–2000) and future time period (2060–2099) by first applying a peak-over-threshold approach on precipitation and then identifying corresponding peak discharges (see Methods section “Event identification”). We then empirically compute  $P$  and  $Q$  magnitudes for different levels of extremeness, i.e. mean events and progressively more extreme events with 10, 20, 50, 100, and 200 year return intervals, by pooling events extracted from the 50 ensemble members. Finally, we derive future relative changes in extreme event magnitudes by comparing magnitudes for a future period (2060–2099) with magnitudes of a historic period (1961–2000) (see Methods section “Changes in event magnitudes and  $P - Q$  relationship”).

We find that median future changes in daily precipitation and corresponding discharge extremes overall catchments depend on their respective level of extremeness (here defined as their return interval, RI; Fig. 1). Precipitation frequency and magnitude are found to increase for all levels of extremeness, with the largest median increases corresponding to the most extreme events which is consistent with prior findings<sup>43–45</sup>. 50-year RI precipitation events (i.e. events of a magnitude occurring



**Fig. 1 Future changes in precipitation (P) and streamflow (Q) magnitudes for different levels of extremeness overall 78 catchments.** Relative changes [–] in (a) event frequency and (b) peak magnitude for mean and progressively more extreme events (those with 10, 20, 50, 100, and 200 year empirical return intervals, respectively). Relative changes are computed by comparing event characteristics of a future period (2060–2099) to characteristics of a historical period (1961–2000). The gray bar in (b) shows the relative change in event timing (day of the year, negative values indicate earlier extreme event occurrence overall events). Meaning of boxplot elements: central line: median, box limits: upper and lower quartiles, upper whisker:  $\min(\max(x), Q_3 + 1.5 \times IQR)$ , lower whisker:  $\max(\min(x), Q_1 - 1.5 \times IQR)$ , no outliers displayed.



**Fig. 2 Catchment-specific return interval thresholds above which precipitation increases result in discharge increases.** Relative changes (1 corresponds to 100% increase) in (a) event frequency and (b) peak magnitude for mean and progressively more extreme events (those with 10, 20, 50, 100, and 200 year empirical return intervals, respectively) for each of the 78 catchments (1 line = 1 catchment). Dashed lines denote catchments without a distinct threshold. Relative changes are computed by comparing event characteristics of a future period (2060–2099) to characteristics of a historical period (1961–2000). The approximate location of the return interval threshold is indicated using different line colors with darker colors representing higher return interval thresholds.

approximately twice per century), occur twice as often (a 100% increase) in the future period vs. the historical period, while the frequency of 200 years RI events increases by up to 200%. Median increases in precipitation magnitudes corresponding to these frequency increases range from an increase <10% for 50 year RI events to an up to 15% increase for 200 year events.

In notable contrast to precipitation changes, changes in flood frequency and magnitude exhibit a more complex response as a function of flood event extremeness. We find that there exists a return interval threshold below which flood frequency and magnitude decrease, and above which they increase. The mean location of this threshold across all catchments lies between event RIs of 20–50 years for both frequency and magnitude (Fig. 1). However, the exact location of this threshold is catchment-dependent (Fig. 2). Some catchments already show increases in magnitude/frequency at very low thresholds (<10 years, lightly colored catchments), while in

other catchments a threshold only emerges at very long return intervals (100 or 200 years, darkly colored catchments). A few catchments (20%) don't show any threshold behavior at all as they either exhibit uniformly increasing or decreasing discharges independent of the return interval. However, even in catchments without a distinct threshold, the discharge response becomes increasingly positive for increasing event magnitudes.

This finding of a catchment-specific return interval threshold in a great majority of instances suggests that the extreme streamflow response in a warming climate changes sign, from negative to positive, when comparing more 'common' flood events (i.e. those occurring 5 or more times per century) to more 'rare' flood events (i.e. those occurring two or fewer times per century). This finding has major implications for the interpretation of time series of observed streamflow, as the historical record is often too short to robustly characterize changes in high-

magnitude events occurring only several times per century, and any such threshold behavior might go undetected as a result. Still, the results corroborate findings by earlier studies suggesting that historical changes in flooding do, to some degree, depend on event extremeness<sup>25,27</sup>.

Next, we assess which meteorological factors and catchment characteristics influence the location of the overall flood response threshold along the extremeness spectrum when considering median changes in extremes overall catchments. For this assessment, we compare historical and future precipitation and discharge extremes for (a) small (<1000 km<sup>2</sup>) and large (>1000 km<sup>2</sup>) catchments, (b) low- (<1000 m.a.s.l.) and high-elevation catchments (>1000 m.a.s.l.), (c) winter (Oct–Mar) and summer (April–Sept) events, (d) snow-influenced (>10 mm stored SWE) and rainfall-driven events (<10 mm stored SWE), and (e) events extracted using different precipitation temporal aggregation levels (1-day, 3-day, and 5-day accumulated precipitation) (see Methods section “Changes in event magnitudes and  $P - Q$  relationship”).

Our results show that the threshold above which precipitation increases translate into increases in flood frequency and magnitude is strongly modulated by elevation, season, and event type (Figs. 3, 4), but does not meaningfully depend upon the precipitation temporal aggregation level (Supplementary Figure 2) or upon catchment size (Supplementary Figure 3). This result may change if studying a dataset with a wider range of catchment sizes. However, when studying larger catchments, interactions of flood waves from different tributaries will have to be considered. The return interval threshold does not exist at all or occurs at a much lower extremeness level in high-elevation catchments (<10 years RI) versus low-elevation catchments (~50 years RI). In other words, precipitation frequency and magnitude increases in high-elevation catchments are more directly translated into flood frequency and magnitude increases than in low-elevation catchments for any given event extremeness level (Figs. 3c, 4c). In addition to elevation, this threshold also depends on the season. In high-elevation catchments, discharge frequency and magnitude increases are stronger in winter than in summer. In contrast, flood frequency and magnitude mostly decrease in low-elevation catchments in winter while they increase in summer for high-magnitude events (Figs. 3b, 4b). A substantial portion of this elevational separation in flood response may be explained by differences in extreme precipitation event type, i.e. whether an event is snow-influenced or rainfall-driven (Figs. 3d–f, 4d–f). In low-elevation catchments, flood frequency and magnitude decrease for snow-influenced events caused by a decrease in extreme precipitation during such events while they increase for very extreme rainfall-driven events (return intervals >50 years) (Figs. 3e, 4e). In contrast, high-elevation catchments show flood frequency and magnitude increases for both snow-influenced and moderately extreme rainfall-driven events (Figs. 3f, 4f). This behavior would be consistent with a simultaneous decrease in mean snowpack accumulation and the number of rain-on-snow events<sup>39,46–50</sup>, which in some cases have lower peaks than solely rainfall-driven events<sup>23</sup>.

**Flood-precipitation dependence strengthens.** In addition to assessing changes in precipitation and flood magnitude, we consider the (non-)stationarity of the relationship between the two variables over time in a warming climate. We compare different measures of dependence including correlation and extremal (i.e. tail) dependence<sup>51</sup> for progressively more extreme events for the historical and future period (see Methods section “Changes in event magnitudes and  $P - Q$  relationship”). Similar to changes in flood frequency and magnitude, we find that changes in the strength of the  $P - Q$  relationship overall catchments are generally positive above a certain return interval threshold and depend on event magnitude, season,

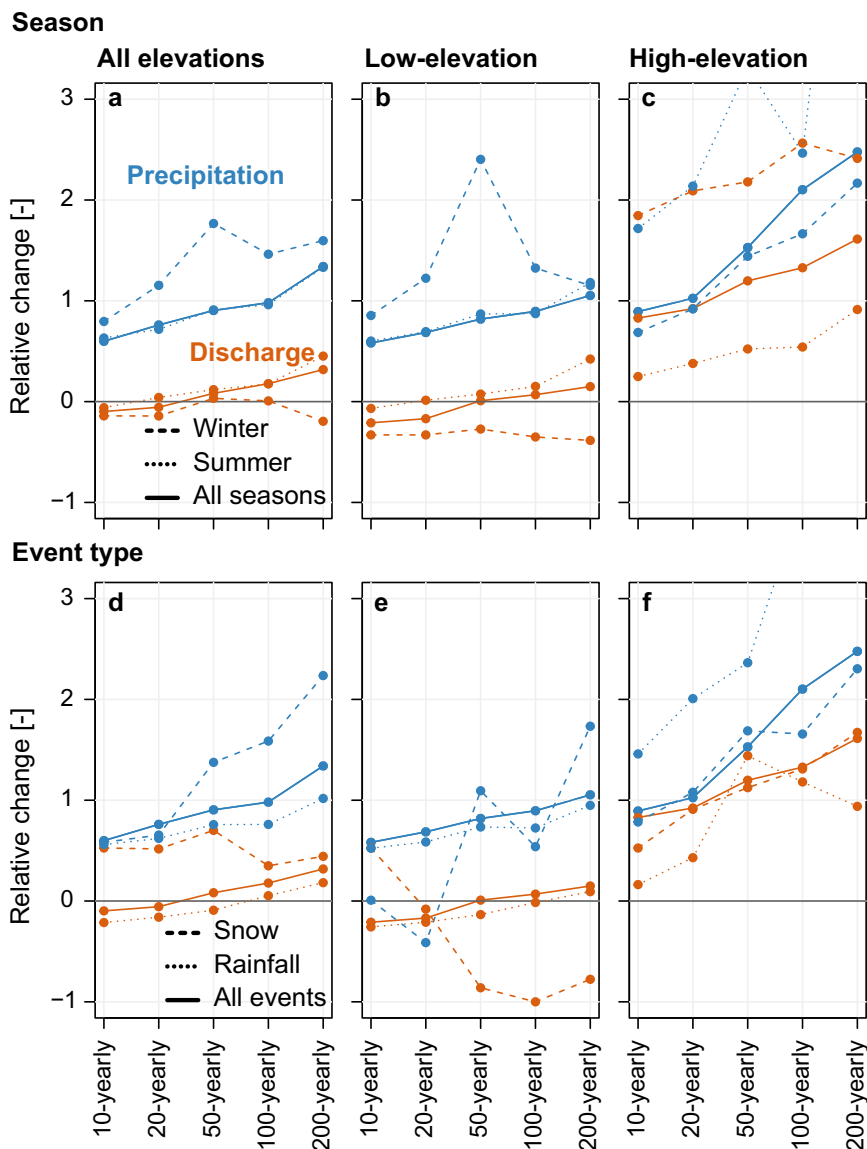
and in particular elevation (Fig. 5). The median  $P - Q$  relationship changes overall 78 catchments are generally stronger in high- versus low-elevation catchments, and are also stronger in winter than in summer. In low-elevation catchments, the relationship weakens for moderate extreme events and intensifies only for very extreme events, particularly in summer. In high-elevation catchments, the relationship intensifies for both moderate and severe extremes. In these catchments, however, the strengthening of the relationship in winter decreases as events become more extreme, while it intensifies more strongly for the more extreme events in summer. These findings suggest that influences on the threshold above which the  $P - Q$  relationship strengthens are complex, and likely vary widely across hydroclimates as suggested by variations by season and event type. They are also suggestive of a potentially important role for antecedent land surface conditions in modulating the underlying relationship – a topic we explore further in the next section.

**Role of antecedent conditions in flood response.** We also assess the extent to which land surface and hydro-meteorological drivers beyond precipitation govern flood magnitudes at different levels of extremeness. For this assessment, we construct a multiple linear regression model that predicts flood magnitude (mean and 100 year RI) using a set of predictors: mean event precipitation, mean event temperature, mean event SWE, and mean event soil moisture anomalies, which are only weakly collinear according to the variable inflation factor (VIF does not exceed 10 for any pair and only exceeds 4 for very few pairs; see Methods section “Importance of hydro-meteorological drivers”). We consider the sign and magnitude of the associated regression coefficients, and their change between the two time periods of interest (historical: 1961–2000, future: 2060–2099).

The regression analysis shows that flood magnitude is driven by different meteorological conditions and land surface processes whose importance varies widely by the level of extremeness, elevation, and season (Fig. 6 upper panel). For moderate and severe extremes at both low and high elevations, precipitation is positively related to discharge magnitude (i.e. for sufficiently extreme events, precipitation increases almost always lead to discharge increases). In contrast, the role of all the other drivers particularly that of temperature strongly depends on the level of extremeness, elevation, and season and is not statistically significant in all cases.

In low-elevation catchments, temperature increases are associated with discharge decreases, particularly for moderate extremes (negative regression coefficients) (Fig. 6a). In summer, higher temperatures mean higher evapotranspiration and therefore lower soil moisture, which means higher soil water storage capacity and therefore less direct runoff resulting from a given amount of precipitation. In winter, higher temperatures are associated with less snow accumulation and therefore less rain-on-snow events<sup>46,47,49,50</sup>, which can lead to smaller flood peaks because solely rainfall-driven events may not be as severe as rain-on-snow events<sup>23</sup>. While these temperature effects are strong for moderate floods, temperature loses importance moving toward more extreme events. This effect is particularly pronounced in summer, where the negative effect of temperature weakens while the positive relation between event magnitude and precipitation intensifies. In winter, temperature effects are still important, however, also to a smaller degree (Fig. 6b).

In low-elevation catchments during winter, soil moisture and snow accumulation are indeed important drivers of flood magnitude. Increases in soil moisture lead to increases in flood magnitudes, as precipitation can more directly be converted into a runoff. In contrast, more snow accumulation is related to smaller floods because water is temporarily stored in the



**Fig. 3 Factors influencing future frequency changes in extreme precipitation and discharge magnitudes for different levels of extremeness.** Median relative change [-] in *P* and *Q* frequency per season (a–c) and event type (d–f) across all, low-elevation, and high-elevation catchments for mean and progressively more extreme events (those with 10, 20, 50, 100, and 200 year return intervals, respectively). Relative changes are computed by comparing event characteristics of a future period (2060–2099) to characteristics of a historical period (1961–2000).

snowpack, and does not form runoff until melting at some later point. While the soil-drying effects of increasing temperatures may lead to flood decreases in low-elevation catchments, they can also lead to flood increases in high-elevation catchments (particularly in winter). This effect arises largely from the phase change of precipitation, which falls increasingly as rain rather than snow in a warming climate<sup>47</sup>, and which has been directly linked with an increase in flood magnitude in such regions<sup>23</sup>. Interestingly, the positive association between temperature and flood magnitude at high elevations exists not only for moderate events, but also for very extreme events.

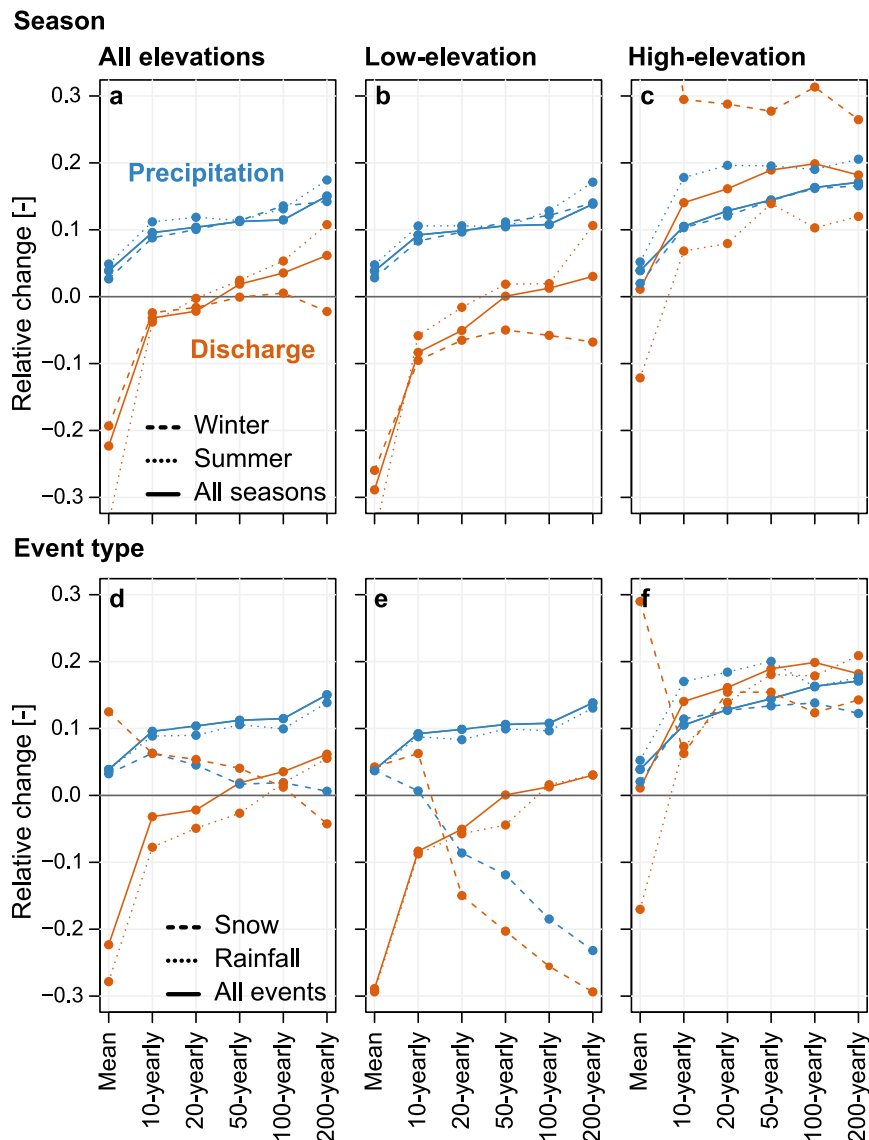
Our analysis of future changes in flood driver importance further shows that the future relevance of precipitation as a flood driver increases for severe events while the importance of temperature increases for moderate but decreases for severe extremes (Fig. 6c–d). This may potentially be understood in the context of soil saturation as a modulating factor: for typical and even moderate events, antecedent soil-drying and snowpack

losses resulting from warming temperatures oppose the effect of increasingly extreme precipitation volume; but for sufficiently severe precipitation events, the extremely large volume of water entering the system may be able to quickly saturate the soil column and overcome even a substantial degree of antecedent soil-drying. In addition, increasingly extreme precipitation may lead to infiltration excess even in the case when soils are not yet saturated. Collectively, these findings support the following generalization: the more extreme a flood event, the more important precipitation becomes as a singular driver – particularly in a warmer future climate.

**Discussion**

In this work, we demonstrate for hydrological Bavaria that there is an extremeness or return interval threshold, which varies by catchment, season, and event type, above which extreme precipitation increases outweigh the soil-drying effects of warming temperatures. This result suggests that in other regions around



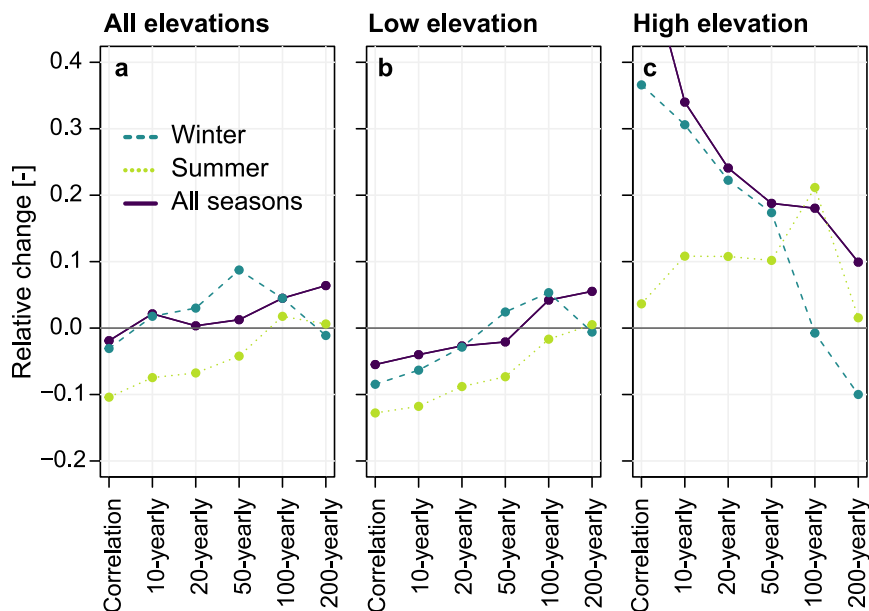


**Fig. 4** Factors influencing future magnitude changes in extreme precipitation and discharge for different levels of extremeness. Median relative change [–] in  $P$  and  $Q$  magnitude per season (a–c) and event type (d–f) across all, low-elevation, and high-elevation catchments for mean and progressively more extreme events (those with 10, 20, 50, 100, and 200 year return intervals, respectively). Same as Fig. 3, but here for future magnitude changes. Relative changes are computed by comparing event characteristics of a future period (2060–2099) to characteristics of a historical period (1961–2000).

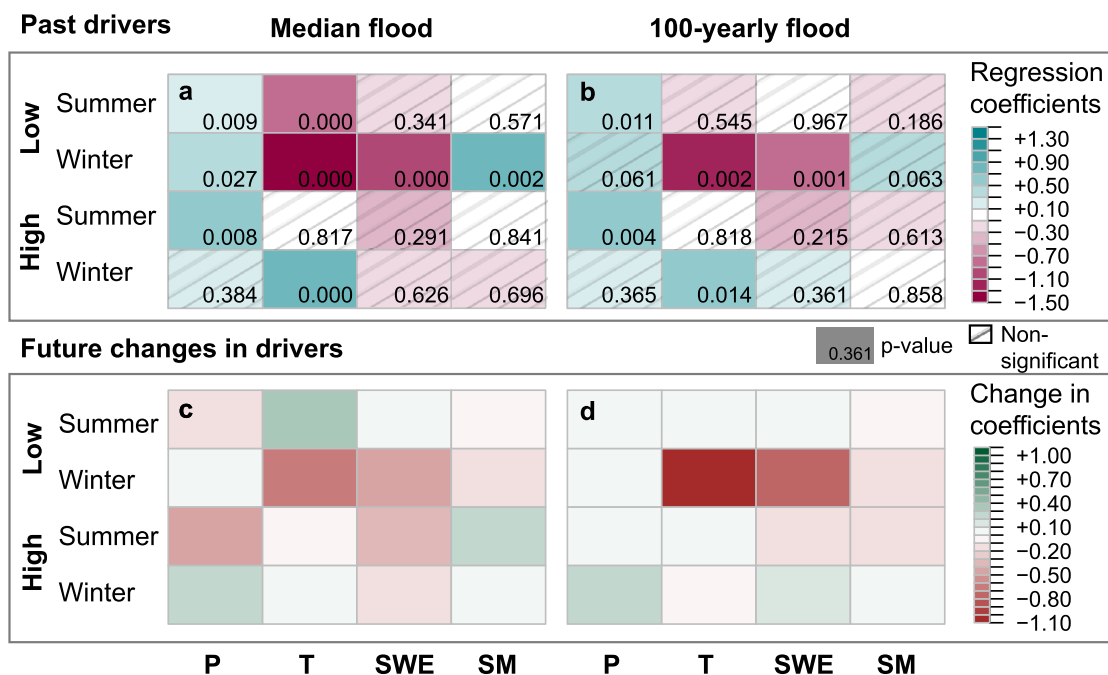
the globe with similar hydro-climates, i.e. temperate climates with pluvial or nival flow regimes, flood risk in a warming climate may also exhibit divergent changes above and below some locally-defined extremeness or return interval threshold. We further find that the hydrologic response to extreme precipitation varies predictably as a function of event magnitude in a warming climate, with streamflow responses becoming increasingly positive even in the few study catchments which do not exhibit distinct threshold behavior. This, when viewed in the context of prior research, may offer evidence for the broader geographic generalizability of our findings. We find that increases in precipitation yield larger and more consistent increases in flood magnitude for more extreme versus more moderate events which is supported by previous observational studies showing only weak dependence between extreme precipitation and moderate flood occurrence in the United States<sup>10</sup>, stronger increasing flood trends for extreme than moderate floods in Central Europe<sup>27</sup>, and trends in extreme discharge that only align with trends in floods for the rarest events in Australian catchments<sup>25</sup>. Thus, there does

appear to be a growing body of real-world evidence suggestive of the existence of a precipitation-flood response threshold across a wider range of hydroclimatic and hydrologic regimes than explicitly considered in the present study.

The complex influences of elevation, season, and event type upon the return interval threshold suggest that the location of this critical cross-over point may vary somewhat widely across regions of the world with varying topography and background climate. Substantial modulation of this threshold would likely occur depending on climatic factors such as aridity and the local relevance of snowmelt, catchment size, and land use and management. Consider, for example, a semi-arid or subtropical regime (as opposed to the moist mid-latitude regime that characterizes the catchments in the present study). In such a location, the return interval threshold might be higher due to drier antecedent soil conditions a temperature-related phenomenon we also see when comparing seasonally-varying summer with winter thresholds (Figs. 3, 4). The existence of a high return interval threshold in drier Mediterranean regions is supported by



**Fig. 5 Factors influencing future changes in the  $P - Q$  relationship for different levels of extremeness.** Median relative change [–] in  $P - Q$  dependence (areal precipitation sum and peak discharge) per season across (a) all, (b) low-elevation, and (c) high-elevation catchments for correlation and tail dependence for progressively more extreme events (those with 10, 20, 50, 100, and 200 year return intervals, respectively). Relative changes are computed by comparing event characteristics of a future period (2060–2099) to characteristics of a historical period (1961–2000).



**Fig. 6 Importance of flood drivers in the past and its future changes.** Importance of precipitation (P), temperature (T), snow-water-equivalent (SWE), and soil moisture (SM) as drivers of historical (a) moderate floods (median overall events identified in a catchment) and (b) extreme floods (100-yearly flood). Future changes in driver importance for (c) moderate floods and (d) extreme floods. All panels are divided into low- vs. high-elevation catchments and distinguish between summer and winter events. Turquoise and pink colors (a, b) indicate positive and negative correlation coefficients, respectively (coefficients not statistically significant at  $p < 0.05$  level are hatched), and green and red colors (c, d) indicate increases and decreases in driver importance, respectively.

observation-based studies that have demonstrated a stronger relationship between precipitation and discharge for larger versus smaller flood events in Spain<sup>52</sup>, and have shown decreases in the occurrence of moderate floods in southern Europe<sup>9,27</sup>. In contrast, if we consider cold high-latitude regions and/or high altitude regions with a snow-dominant precipitation regime, the return interval threshold might be expected to be much lower.

Indeed, this relationship is apparent from our threshold analysis for snow-influenced events in high-elevation regions (Figs. 3, 4). Additionally, and as suggested by our results (Supplementary Figure 3) the return interval threshold may also be modulated by catchment area (generally increasing with catchment size). For larger river basins than the ones included in our Bavarian selection, this finding would imply higher return level thresholds than

20–50 years. Furthermore, direct human influence on streamflow such as dynamic reservoir operations and/or flood management interventions might lead to higher return interval thresholds because smaller floods can be buffered by temporary water storage<sup>53</sup>. In contrast, urbanized catchments (characterized by a high fraction of water-impervious surfaces) might have lower return interval thresholds than catchments with unsealed surfaces because of a more direct relationship between extreme precipitation and flood response<sup>54,55</sup>. Such a return interval threshold might even vary from year to year in a single location-occurring at a higher level of event extremeness during drought versus pluvial periods.

How exactly such a return interval threshold varies for different hydro-climates remains to be investigated using a global hydro-SMILE. Creating such a global hydro-SMILE for flood analyses requires the combination of a globally downscaled and bias-corrected atmospheric SMILE with a global hydrological model specifically calibrated for flood peaks. Satisfactory calibration for far-from-mean state conditions is challenging using calibration metrics commonly used for large-scale model calibration<sup>56</sup> and data storage and computational costs are high at a global scale when a large spatial domain is combined with a large ensemble size. In addition, global-scale models may not as accurately represent complex land surface processes as smaller-scale models and appropriate reference datasets for meteorology, soils, and hydrogeology are harder to obtain. Creating such a global hydro-SMILE therefore remains a considerable research effort, but one of substantial importance in a warming climate.

There are two important implications arising from the existence of a return interval threshold above which increases in precipitation directly translate to increases in flood occurrence. First, this threshold existence suggests that previous studies that focused on less extreme floods, which have shown little change or even decreases in annual streamflow maxima or events with return intervals of less than ~20 years<sup>57,58</sup>, will likely be unrepresentative of changes in higher-magnitude events. A robust statistical signal is unlikely to arise in most historical datasets shorter than 100 years because the strongest link between increasing extreme precipitation and flood magnitude occurs for rare, high-magnitude events with return intervals exceeding 20–50 years. This result points to an important limitation of observation-only studies, as well as to the critical importance of large modeling ensembles that can yield larger sample sizes for rare, high-magnitude events. Second, our analysis suggests that despite historical uncertainties, large increases in flood magnitude are likely in a warming climate for the very largest events—potentially including those unprecedented in the modern historical record (i.e., events with 200-year RI, Fig. 1). The fact that climate warming may act to decrease the magnitude of more moderate flood events while simultaneously increasing the magnitude of the most extreme events, however, highlights the considerable risk of developing a “false sense of security” based on recent historical experience. These findings therefore have major implications for climate adaptation and flood risk mitigation activities, as well as infrastructure design, in a warming climate.

Ultimately, we suggest that this analysis may help reconcile seemingly conflicting perspectives in the climatological and hydrological literature on flood risk in a warming climate. The apparent “precipitation-flood” paradox – whereby precipitation extremes have increased, but floods have not<sup>5,24</sup> – may in fact be fully resolved by separating flood events by their extremeness. In this sense, both perspectives may ultimately be correct: hydrologic evidence suggesting no consistent increase in recent flood magnitude because of land surface drying and the changing role of snow using observational records of limited length<sup>9,59,60</sup> is physically consistent with climatological arguments pointing to a large increase in the magnitude and

frequency of historically rare or unprecedented precipitation events and subsequent flood risk<sup>61,62</sup>.

Future research aimed at expanding the coverage of the regional hydro-SMILE approach to a wider range of hydrologic and climatological regimes will be critical in confirming the broader generalizability of our findings in the present study, but emerging observational evidence does suggest that threshold behavior in precipitation-flood response is plausible across a wide range of regimes in a warming climate<sup>9,10,25,52</sup>. In this work, we confirm that antecedent land surface conditions are indeed critical in modulating more common or moderate flood events, but that precipitation becomes the dominant driver for very extreme events and ultimately overwhelms the effects of soil moisture or snowpack. Finally, we emphasize that the inherent limitations of the historical observational record can be obviated through the use of a climate model large ensemble approach in combination with an advanced hydrological model—a framework that might be useful for more broadly assessing complex and possibly non-linear changes in extreme events in the warming earth system.

## Methods

**Study region.** We study the relationship between extreme precipitation and flood events and its influencing factors in a warming climate for a set of 78 catchments with nearly natural flow conditions in Hydrological Bavaria (Supplementary Figure 1). This region comprises the Main, Danube, and Inn rivers with their major tributaries. This study region is particularly well suited to analyze variations in the precipitation–discharge ( $P - Q$ ) relationship because the constituent catchments are characterized by diverse topographic and climatic conditions, ranging from a wet alpine region in the south ( $1700 \text{ mm y}^{-1}$ ) to a relatively flat and dry foreland to the north ( $700 \text{ mm y}^{-1}$ ), and diverse soil types and land uses. The variations in these conditions lead to a wide range of hydrologic regimes, including snow-influenced regimes with flood peaks in spring and summer to primarily rainfall-influenced regimes with the main flood season in winter. While these regime types can be considered representative of the temperate climate zone with similar runoff regimes (pluvial to nival), our catchment selection does not cover other climate zones such as cold-climates, semi-arid to arid regions, and the tropics.

**Hydro-SMILE.** For this analysis, we use a hydro-SMILE, i.e. hydrological simulations obtained by driving a hydrological model with a Single Model Initial-Condition Large Ensemble (SMILE) climate model. The underlying simulations were originally generated by Willkofer et al.<sup>40</sup> as part of the ClimEx project<sup>41</sup>. The simulations consist of daily streamflow ( $\text{mm d}^{-1}$ ), snow-water-equivalents (SWE, mm), and soil moisture (%) – all of which were obtained by driving the hydrological model WaSiM-ETH<sup>42</sup> with a 50-member ensemble of high-resolution climate input (spatial:  $500 \times 500 \text{ m}^2$ , temporal: 3 h). The climate input consists of an ensemble provided through the Canadian Regional Climate Model version 5 nested with the Canadian Earth System Model<sup>63</sup> under RCP 8.5<sup>64</sup> – a ‘high-warming’ climate scenario. WaSiM-ETH is a distributed, mainly physically-based hydrological model comprising modules for evapotranspiration, interception, snow accumulation, and melt, glaciers, runoff generation, soil water storage, and discharge routing<sup>42</sup>. The model was set up for 98 catchments in Hydrological Bavaria by Willkofer et al.<sup>40</sup> using spatial information on elevation, slope, and exposition derived from a digital elevation model for Europe (EU-DEM<sup>65</sup>), land-use derived from the CORINE land cover dataset<sup>66</sup>, soil characteristics derived from the European soil database (ESDB v2.0<sup>67</sup>), and hydro-geology (hydraulic conductivity) derived from the Bavarian hydrogeology map<sup>68</sup> and the international hydrogeological map of Europe (IHME1500 v1.1.<sup>69</sup>) to define global model parameters (i.e. parameters applied to the 98 catchments) describing evapotranspiration rates, infiltration rates, groundwater fluxes, snowmelt, and glacier dynamics and by calibrating four parameters, i.e. those related to recession and direct flow. These local parameters were calibrated for the period 2004–2010 using the dynamically dimensioned search algorithm<sup>70</sup> on the observed 3 h discharge of the 98 catchments provided by the Bavarian Environment Agency (Bayerisches Landesamt für Umwelt - LfU<sup>71</sup>) and sub-daily observed interpolated meteorological input (i.e. precipitation, temperature, relative humidity, incoming shortwave radiation, and wind speed).

The meteorological Sub-Daily Climatological REference dataset (SDCLIREF) created in the ClimEx-project is based on a combination of hourly and disaggregated daily station data. To obtain the disaggregated daily station data, the method of fragments<sup>72</sup> was used to extend the sub-daily record to 1981–2010 and to densify the station network. The station data were then interpolated to a  $500 \times 500 \text{ m}^2$  grid using a combination of multiple linear regression, considering elevation, exposition, latitude, and longitude, and inverse distance weighting similar to Rauthe et al.<sup>73</sup>. The dynamically dimensioned search algorithm used a multi-objective function targeted at optimizing flood characteristics composed of the Nash–Sutcliffe efficiency ( $E_{NS}$ <sup>74</sup>) and the Kling–Gupta efficiency ( $E_{KG}$ <sup>75</sup>), which both focus on high flows<sup>76</sup>, the  $\log(E_{NS})$ , which emphasizes low flows, and the root-



mean-squared error to standard deviation ratio ( $R_{SR}$ ), which quantifies volume errors. The overall objective function assigns a lot of weight to the metrics  $E_{NS}$  and  $E_{KG}$  because our study focuses on flood events:

$$M = 0.5 \times (1 - E_{NS}) + 0.25 \times (1 - E_{KG}) + 0.15 \times (1 - \log(E_{NS})) + 0.1 \times R_{SR}. \quad (1)$$

The calibrated model was first run for a reference period 1981–2010 with the sub-daily (3 h) observed interpolated meteorological input also used for model calibration. After running the model for the reference period, it was run for a simulation period 1961–2099 with meteorological data derived from the fifth-generation Canadian Regional Climate Model large ensemble (CRCM5-LE) 50 members<sup>41</sup> consisting of a dynamically downscaled version (0.11°; 12 km) of the second generation Canadian Earth System Model large ensemble (CanESM2-LE)<sup>77</sup>. The CRCM5-LE data were further bias-corrected using a quantile mapping approach<sup>78,79</sup> adjusted to sub-daily time steps and the SDCLIREF as the reference climatology (1981–2010). Correction factors were determined for each quantile bin for each month and sub-daily time step. To preserve the ensemble spread, all members were pooled to obtain the correction factors and these factors were subsequently applied to each ensemble member separately. The bias-corrected data were then further downscaled to  $500 \times 500 \text{ m}^2$  spatial resolution. The center point of each  $0.11^\circ$  CRCM5-LE grid cell was treated as a virtual meteorological station and for each time step the anomaly from the mean state was interpolated to the  $500 \times 500 \text{ m}^2$  grid using inverse distance weighting. The interpolated anomalies were then multiplied/added to the climatological reference fields from the SDCLIREF. Afterwards, the downscaled data were corrected in order to ensure the conservation of mass for each downscaled  $0.11^\circ$  grid cell.

Previous studies have demonstrated that CRCM5-LE (1) shows realistic patterns of daily and sub-daily extreme precipitation<sup>80</sup> and of the timing of annual maximum precipitation over Central Europe<sup>81</sup>; (2) that its high-resolution allows for a realistic representation of local precipitation extremes, especially over coastal and mountainous regions<sup>41</sup>; (3) that it is consistent with the EURO-Cordex ensemble<sup>82</sup>, and (4) that it compares well to other large ensembles with respect to regional precipitation pattern changes<sup>81</sup>.

For the subsequent analyses of extreme precipitation and flood events, the 3 h meteorological and streamflow time series were aggregated to a daily scale and averaged over each catchment.

**Hydrological model evaluation.** We here evaluate the hydrological model for the 78 catchments used in this study for the reference period 1981–2010 using observed daily streamflow from the hydrological services of Bavaria and Baden-Württemberg (both in Germany), Austria, and Switzerland with respect to a set of measures including visual inspection, general efficiency metrics, and flood characteristics of events determined using a peak-over-threshold approach with the 98th flow percentile as a threshold and a minimum time lag of 10 days between successive events to ensure independence. The general efficiency metrics considered are the Kling–Gupta efficiency<sup>75</sup>, Nash–Sutcliffe efficiency<sup>74</sup>, volumetric efficiency, and mean absolute error, four metrics often used in flood simulation studies. The flood characteristics considered are the number of events, mean timing (day of the year), mean peak magnitude ( $\text{mm d}^{-1}$ ), mean volume ( $\text{mm event}^{-1}$ ), mean duration (days), and  $P-Q$  dependence. The start and end of events are determined as the time when discharge rises and falls below the threshold, respectively, event duration is defined as the time elapsing between the start and end of an event, and the volume as the cumulative flow exceeding the threshold over the whole event duration.

The model shows a satisfactory performance qualitatively and quantitatively using general and flood-specific evaluation metrics (Supplementary Figure 4). Kling–Gupta efficiencies ranged from the first quartile of 0.67 to the third quartile of 0.85, Nash–Sutcliffe efficiencies from the first quartile of 0.56 to the third quartile of 0.8, and volumetric efficiencies from the first quartile of 0.68 to the third quartile of 0.8. The mean absolute error lay at  $0.35 \text{ mm d}^{-1}$  (Supplementary Figure 4a). The flood-specific performance evaluation showed a slight underestimation of the number of events (relative error: 1st quartile:  $-0.14$ , median:  $-0.06$ , 3rd quartile: 0.07), a slight delay of the timing of flood occurrence (relative error: 1st quartile:  $-0.01$ , median: 0.05, 3rd quartile: 0.11), a slight overestimation of flood peaks (relative error: 1st quartile:  $-0.02$ , median: 0.07, 3rd quartile: 0.22), an overestimation of both flood volume (relative error: 1st quartile: 0.08, median: 0.32, 3rd quartile: 0.54) and duration (relative error: 1st quartile:  $-0.02$ , median:  $-0.14$ , 3rd quartile: 0.39), and an underestimation of  $P-Q$  dependence (relative error: 1st quartile:  $-0.35$ , median:  $-0.24$ , 3rd quartile:  $-0.04$ ) (Supplementary Figure 4b). Overall, the model performance with respect to high flows and flooding is satisfactory. In addition, the results of our change impact assessment are less affected by inconsistencies between observed and simulated flow because we assess relative rather than absolute changes in precipitation and flood magnitudes.

**Event identification.** Using the daily streamflow simulations from the 50 members of the hydro-SMILE, we identify pairs of extreme precipitation (i.e. areal sum over catchment) and corresponding streamflow for two non-overlapping periods of 40 years, a historical (1961–2000) and a warmer future period (2060–2099). Periods of 40 years were chosen to maximize the sample size while ensuring that the two periods are as distinct as possible. To identify these  $P-Q$  pairs, we first define daily extreme

precipitation events ( $\text{mm d}^{-1}$ ) using the 99th percentile (determined on all days (including 0 precipitation days) using the full-time series 1961–2099) as a threshold and by prescribing a minimum time lag of 10 days between events in order to ensure independence (i.e. to enable declustering). This event extraction procedure results in roughly 2–2.5 events chosen per year on average depending on the catchment. Over the 2000 model years of data per time period (40 years across 50 ensemble members), we select approximately 5000 extreme events per catchment. The start of each precipitation event is defined as the day when precipitation exceeds 1 mm prior to the first threshold exceedance and the end of each precipitation event is defined as the time when precipitation falls below 1 mm after the final threshold exceedance (for an illustration of the event identification procedure see Supplementary Figure 5). Next, for each precipitation event, we identify the corresponding streamflow peak ( $\text{mm d}^{-1}$ ) within a time window from the start of the precipitation event to 5 days after the end of the precipitation event. Finally, for each event, we determine temperature ( $^\circ\text{C}$ ) on the day of peak precipitation and snow-water-equivalent (mm) and soil moisture anomalies (deviation from the mean, percentage) on the day prior to the occurrence of the precipitation extreme. We repeat this event extraction procedure for two additional temporal aggregation levels (3-day and 5-day mean precipitation accumulations) in order to assess the effect of precipitation aggregation on future precipitation and discharge changes because event identification using different aggregation levels results in the extraction of different event sets.

**Changes in event magnitudes and  $P-Q$  relationship.** In the first part of our analysis, we use the  $P-Q$  event pairs identified to analyze how precipitation and corresponding flood magnitudes as well as the relationship between the two variables may change in the future. To do so, we compare the statistical characteristics of these variables for the future period (2060–2099) to the characteristics of the historical period (1961–2000).  $P$  and  $Q$  magnitudes are determined empirically by pooling events extracted from the 50 ensemble members for different levels of extremeness, i.e. ‘mean’ events (those which occur, on average, once or twice per year) and progressively more extreme events with 10, 20, 50, 100, and 200 year return intervals, respectively. Sample quantiles are computed for probabilities corresponding to different return periods  $T$  using:

$$p = 1 - (\mu/T), \quad (2)$$

where  $\mu$  is the mean inter-arrival time between events. The  $P-Q$  relationship is characterized for different dependence measures including Pearson’s correlation coefficient and the tail dependence coefficient  $\chi^{51}$ , which provides a simple measure of extremal dependence, at different levels of extremeness (i.e. probabilities corresponding to return intervals of 10, 20, 50, 100, and 200 years). Future changes are expressed as relative changes with respect to the characteristics of the historical period.

We identify factors potentially influencing the nature of change in  $P-Q$  magnitudes and relationship by looking at different levels of extremeness, i.e. return intervals, small and large catchments, high-elevation, and low-elevation catchments, winter and summer events, and snow-influenced and rainfall-driven events. The levels of extremeness considered for both  $P$  and  $Q$  are the mean and quantiles corresponding to return intervals of 10, 20, 50, 100, and 200 years. Within the 2000 model years available for analysis, roughly 10 events have a return interval of 200 years while roughly 200 events have a return interval of 10 years in each catchment. Small to medium-size catchments are distinguished from large catchments by setting an area threshold of  $1000 \text{ km}^2$ <sup>83</sup>, which results in 21 small and 57 large catchments. Similarly, low-elevation catchments are separated from high-elevation catchments using an elevation threshold of 1000 m above sea level<sup>84</sup>, which results in 55 low-elevation catchments and 23 high-elevation catchments. Winter events are defined as those events happening between October and March and summer events as those events occurring between April and September. Our results are not sensitive to the use of an alternative seasonal definition aligning with the start of the hydrological year (Nov–April, May–Oct). Throughout the analysis, snow-influenced events are defined as those events during which there was at least 10 mm of SWE while rainfall-driven events are those with less than 10 mm of SWE<sup>47</sup>.

**Importance of hydro-meteorological drivers.** In the second part of the analysis, we identify potential hydro-meteorological drivers influencing extreme precipitation and flood magnitudes and their statistical relationships. A comparison of driver importance for the two periods (historical and future) allows us to identify drivers losing or gaining importance in the future. For both periods, we fit multiple linear models to flood magnitudes (mean or quantiles for the 78 catchments) using four explanatory variables, all of which exhibit only weak collinearity according to the variable inflation factor, which lies around 1–2 for most variables and does not exceed 4 in most cases. The explanatory variables include mean event precipitation for each catchment (i.e. mean precipitation for the extreme events identified), mean event temperature, mean event SWE, and mean event soil moisture anomaly. Both flood magnitudes and the explanatory variables are standardized prior to model fitting by subtracting the mean and dividing by the standard deviation ( $z$ -scores) in order to make the resulting regression coefficients inter-comparable and easily interpretable. Comparing regression coefficients of the future model to the coefficients of the historical model (absolute changes) enables quantification of changes in future driver importance. Similar to the change analysis, we also distinguish between different levels of extremeness to determine how driver importance varies for events with different return intervals (mean and 100 year event), between low-

and high-elevation catchments to define to which degree driver importance depends on catchment elevation, and between winter and summer events to shed light on how driver dependence varies by season.

### Data availability

The raw data of the CRCM5-LE is publicly available to the scientific community ([www.climex-project.org](http://www.climex-project.org)). The extreme precipitation-discharge pairs generated with the hydro-SMILE and analyzed in this study are available through HydroShare: <http://www.hydroshare.org/resource/945d7b4f61d145d789eb090fbf51cb5>.

### Code availability

The code used to process the data and to produce the figures can be requested from the first author.

Received: 26 February 2021; Accepted: 29 July 2021;

Published online: 26 August 2021

### References

- Prein, A. F. et al. Increased rainfall volume from future convective storms in the US. *Nat. Clim. Change* **7**, 880–884 (2017).
- Swain, D. L., Langenbrunner, B., Neelin, J. D. & Hall, A. Increasing precipitation volatility in twenty-first-century California. *Nat. Clim. Change* **8**, 427–433 (2018).
- Pendergrass, A. G. et al. Nonlinear response of extreme precipitation to warming in CESM1. *Geophys. Res. Lett.* **46**, 10551–10560 (2019).
- Na, Y., Fu, Q. & Kodama, C. Precipitation probability and its future changes from a global cloud-resolving model and CMIP6 simulations. *J. Geophys. Res. Atmos.* **125**, 1–23 (2020).
- Madsen, H., Lawrence, D., Lang, M., Martinkova, M. & Kjeldsen, T. R. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *J. Hydrol.* **519**, 3634–3650 (2014).
- Mediero, L. et al. Identification of coherent flood regions across Europe by using the longest streamflow records. *J. Hydrol.* **528**, 341–360 (2015).
- Hodgkins, G. A. et al. Climate-driven variability in the occurrence of major floods across North America and Europe. *J. Hydrol.* **552**, 704–717 (2017).
- Dickinson, J. E., Harden, T. M. & McCabe, G. J. Seasonality of climatic drivers of flood variability in the conterminous United States. *Sci. Rep.* **9**, 1–10 (2019).
- Blöschl, G. et al. Changing climate both increases and decreases European floods. *Nature* **573**, 108–111 (2019).
- Do, H. X., Mei, Y. & Gronewold, A. D. To what extent are changes in flood magnitude related to changes in precipitation extremes? *Geophys. Res. Lett.* **47**, 1–10 (2020).
- Hirabayashi, Y. et al. Global flood risk under climate change. *Nat. Clim. Change* **3**, 816–821 (2013).
- Thober, S. et al. Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming. *Environ. Res. Lett.* **13**, 014003 (2018).
- Brunner, M. I., Hingray, B., Zappa, M. & Favre, A. C. Future trends in the interdependence between flood peaks and volumes: hydro-climatological drivers and uncertainty. *Water Resour. Res.* **55**, 1–15 (2019).
- Maraun, D. When will trends in European mean and heavy daily precipitation emerge? *Environ. Res. Lett.* **8**, 014004 (2013).
- Donat, M. G., Lowry, A. L., Alexander, L. V., O’Gorman, P. A. & Maher, N. More extreme precipitation in the world’s dry and wet regions. *Nat. Clim. Change* **6**, 508–513 (2016).
- Lopez-Cantu, T., Prein, A. F. & Samaras, C. Uncertainties in future U.S. extreme precipitation from downscaled climate projections. *Geophys. Res. Lett.* **47**, 1–11 (2020).
- Rastogi, D., Touma, D., Evans, K. J. & Ashfaq, M. Shift towards intense and widespread precipitation events over the United States by mid 21st century. *Geophys. Res. Lett.* **47**, e2020GL089899 (2020).
- Dankers, R. & Feyen, L. Climate change impact on flood hazard in Europe: an assessment based on high-resolution climate simulations. *J. Geophys. Res. Atmos.* **113**, 1–17 (2008).
- Hodgkins, G. A., Dudley, R. W., Archfield, S. A. & Renard, B. Effects of climate, regulation, and urbanization on historical flood trends in the United States. *J. Hydrol.* **573**, 697–709 (2019).
- Whitfield, P. H. Floods in future climates: a review. *J. Flood Risk Manag.* **5**, 336–365 (2012).
- Andrés-Doménech, I., García-Bartual, R., Montanari, A. & Marco, J. B. Climate and hydrological variability: the catchment filtering role. *Hydrol. Earth System Sci.* **19**, 379–387 (2015).
- Blöschl, G. et al. Increasing river floods: fiction or reality? *Wiley Interdiscip. Rev. Water* **2**, 329–344 (2015).
- Davenport, F. V., Herrera-Estrada, J. E., Burke, M. & Noah, S. Flood size increases nonlinearly across the western United States in response to lower snow-precipitation ratios. *Water Resour. Res.* **56**, e2019WR025571 (2020).
- Sharma, A., Wasko, C. & Lettenmaier, D. P. If precipitation extremes are increasing, why aren’t floods? *Water Resour. Res.* **54**, 8545–8551 (2018).
- Wasko, C. & Nathan, R. Influence of changes in rainfall and soil moisture on trends in flooding. *J. Hydrol.* **575**, 432–441 (2019).
- Van Der Wiel, K., Selten, F. M., Bintanja, R., Blackport, R. & Screen, J. A. Ensemble climate-impact modelling: extreme impacts from moderate meteorological conditions. *Environ. Res. Lett.* **15**, 34050 (2020).
- Bertola, M., Viglione, A., Hall, J. & Blöschl, G. Flood trends in Europe: are changes in small and big floods different? *Hydrol. Earth System Sci.* **24**, 1805–1822 (2020).
- Dobler, C., Bürger, G. & Stötter, J. Assessment of climate change impacts on flood hazard potential in the Alpine Lech watershed. *J. Hydrol.* **460–461**, 29–39 (2012).
- Prosdocimi, I. et al. Areal models for spatially coherent trend detection: the case of British peak river flows. *Geophys. Res. Lett.* **46**, 13054–13061 (2019).
- Thompson, V. et al. High risk of unprecedented UK rainfall in the current climate. *Nat. Commun.* **8**, 1–6 (2017).
- Kelder, T. et al. Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *npj Clim. Atmos. Sci.* **3**, 1–13 (2020).
- Meko, D. M. et al. Medieval drought in the upper Colorado river basin. *Geophys. Res. Lett.* **34**, 1–5 (2007).
- Blöschl, G., Kiss, A., Viglione, A., Barriendos, M. & Böhm, O. Current European flood-rich period exceptional compared with past 500 years. *Nature* **583**, 560–566 (2020).
- Hundecha, Y. & Merz, B. Exploring the relationship between changes in climate and floods using a model-based analysis. *Water Resour. Res.* **48**, W04512 (2012).
- Brunner, M. I., Papalexio, S., Clark, M. P. & Gilleland, E. How probable is widespread flooding in the United States? *Water Resour. Res.* **56**, e2020WR028096 (2020).
- Deser, C. et al. Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).
- van der Wiel, K., Wanders, N., Selten, F. M. & 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 (2019).
- Kempen, G. V., Wiel, K. V. D. & Melsen, L. A. The impact of hydrological model structure on the simulation of extreme runoff events. *Nat. Hazards Earth Sys. Sci.* **21**, 961–976 (2021).
- Poschlod, B., Willkofer, F. & Ludwig, R. Impact of climate change on the hydrological regimes in Bavaria. *Water* **12**, 1–23 (2020).
- Willkofer, F. et al. A holistic modelling approach for the estimation of return levels of peak flows in Bavaria. *Water* **12**, 2349 (2020).
- Leduc, M. et al. 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 (2019).
- Schulla, J. Model description WaSiM (Water balance Simulation Model). Tech. Rep., Zurich (2019). [http://www.wasim.ch/en/products/wasim\\_description.htm](http://www.wasim.ch/en/products/wasim_description.htm).
- Pendergrass, A. G. & Knutti, R. The uneven nature of daily precipitation and its change. *Geophys. Res. Lett.* **45**, 911–980 (2018).
- Myhre, G. et al. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Sci. Rep.* **9**, 1–10 (2019).
- Martel, J. L., Mailhot, A. & 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. Clim.* **33**, 1089–1103 (2020).
- McCabe, G. J., Clark, M. P. & Hay, L. E. Rain-on-snow events in the western United States. *Bull. Am. Meteorol. Soc.* **88**, 319–328 (2007).
- Li, D., Lettenmaier, D. P., Margulis, S. A. & Andreadis, K. The role of rain-on-snow in flooding over the conterminous United States. *Water Resour. Res.* **55**, 8492–8513 (2019).
- Musselman, K. N. et al. Projected increases and shifts in rain-on-snow flood risk over western North America. *Nat. Clim. Change* **8**, 808–812 (2018).
- Chegwidden, O., Rupp, D. E. & Nijssen, B. Climate change alters flood magnitudes and mechanisms in climatically diverse headwaters across the northwestern United States. *Environ. Res. Lett.* **15**, 09404 (2020).
- Sikorska-Senoner, A. E. & Seibert, J. Flood-type trend analyses for apline catchments. *Hydrol. Sci. J.* **65**, 1281–1299 (2020).
- Coles, S. An introduction to statistical modeling of extreme values. Springer Series in Statistics (Springer London, London, 2001). <http://link.springer.com/10.1007/978-1-4471-3675-0>.

52. Bracken, L. J., Cox, N. J. & Shannon, J. The relationship between rainfall inputs and flood generation in south-east Spain. *Hydrol. Processes* **22**, 683–696 (2008).
53. Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A. & Fiori, A. Reservoir effects on flood peak discharge at the catchment scale. *Water Resour. Res.* **54**, 9623–9636 (2018).
54. Vogel, R. M., Yaindl, C. & Walter, M. Nonstationarity: Flood magnification and recurrence reduction factors in the united states. *J. Am. Water Resour. Assoc.* **47**, 464–474 (2011).
55. Blum, A. G., Ferraro, P. J., Archfield, S. A. & Ryberg, K. R. Causal effect of impervious cover on annual flood magnitude for the United States. *Geophys. Res. Lett.* **47**, e2019GL086480 (2020).
56. Brunner, M. I. et al. Flood spatial coherence, triggers and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated models. *Hydrol. Earth System Sci.* **25**, 105–119 (2021).
57. Bormann, H., Pinter, N. & Elfert, S. Hydrological signatures of flood trends on German rivers: flood frequencies, flood heights and specific stages. *J. Hydrol.* **404**, 50–66 (2011).
58. Do, H. X., Westra, S. & Leonard, M. A global-scale investigation of trends in annual maximum streamflow. *J. Hydrol.* **552**, 28–43 (2017).
59. Hall, J. et al. Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrol. Earth System Sci.* **18**, 2735–2772 (2014).
60. Archfield, S. A., Hirsch, R. M., Viglione, A. & Blöschl, G. Fragmented patterns of flood change across the United States. *Geophys. Res. Lett.* **43**, 232–239 (2016).
61. Uhe, P. F. et al. Enhanced flood risk with 1.5 °C global warming in the Ganges-Brahmaputra-Meghna basin. *Environ. Res. Lett.* **14**, 074031 (2019).
62. Swain, D. L. et al. Increased flood exposure due to climate change and population growth in the United States. *Earth's Future* **8**, e2020EF001 (2020).
63. Martynov, A. et al. Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Clim. Dyn.* **41**, 2973–3005 (2013).
64. Moss, R. H. et al. The next generation of scenarios for climate change research and assessment. *Nat.* **463**, 747–756 (2010).
65. European Environment Agency. Copernicus Land Monitoring Service - EU-DEM (2017). <https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem>.
66. European Environment Agency. Corine Land Cover 2006 raster data (2016). <https://www.eea.europa.eu/data-and-maps/data/clc-2006-raster-4>.
67. European Soil Data Centre (ESDAC). European Soil Database v2.0 (vector and attribute data) (2001). <https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data>.
68. Bayerisches Landesamt für Umwelt. Hydrogeologische Karte 1:200.000 (2000). [https://www.lfu.bayern.de/geologie/hydrogeologie\\_karten\\_daten/hk200/index.htm](https://www.lfu.bayern.de/geologie/hydrogeologie_karten_daten/hk200/index.htm).
69. Bundesanstalt für Geowissenschaften und Rohstoffe (BGR). IHME1500 - International Hydrogeological Map of Europe 1:1500000 (2013). [https://www.bgr.bund.de/EN/Themen/Wasser/Projekte/laufend/Beratung/Ihme1500/ihme1500\\_projektbeschr\\_en.html](https://www.bgr.bund.de/EN/Themen/Wasser/Projekte/laufend/Beratung/Ihme1500/ihme1500_projektbeschr_en.html).
70. Tolson, B. A. & Shoemaker, C. A. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* **43**, 1–16 (2007).
71. Bayerisches Landesamt für Umwelt. Wasserstand und Abfluss (2021). [https://www.lfu.bayern.de/wasser/wasserstand\\_abfluss/index.htm](https://www.lfu.bayern.de/wasser/wasserstand_abfluss/index.htm).
72. Westra, S., Mehrotra, R., Sharma, A. & Srikanthan, R. Continuous rainfall simulation: 1. A regionalized subdaily disaggregation approach. *Water Resour. Res.* **48**, 1–16 (2012).
73. Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A. & Gratzki, A. A central European precipitation climatology - part I: generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift* **22**, 235–256 (2013).
74. Nash, J. E. & Sutcliffe, I. V. River flow forecasting through conceptual models part I - a discussion of principles. *J. Hydrol.* **10**, 282–290 (1970).
75. Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **377**, 80–91 (2009).
76. Mizukami, N. et al. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrol. Earth System Sci.* **23**, 2601–2614 (2019).
77. Arora, V. K. et al. Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.* **38**, 3–8 (2011).
78. Mpelasoka, F. S. & Chiew, F. H. Influence of rainfall scenario construction methods on runoff projections. *J. Hydrometeorol.* **10**, 1168–1183 (2009).
79. Willkofer, F. et al. The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments. *J. Hydrol. Reg. Stud.* **19**, 25–41 (2018).
80. Poschlod, B., Ludwig, R. & Sillmann, J. Ten-year return levels of sub-daily extreme precipitation over Europe. *Earth Syst. Sci. Data* **13**, 983–1003 (2021).
81. Wood, R. R. & Ludwig, R. Analyzing internal variability and forced response of subdaily and daily extreme precipitation over Europe. *Geophys. Res. Lett.* **47**, e2020GL089300 (2020).
82. von Trentini, F., Aalbers, E., Fischer, E. & Ludwig, R. Comparing internal variabilities in three regional single-model initial-condition large ensembles (SMILE) over Europe. *Earth Syst. Dyn.* **11**, 1013–1031 (2020).
83. Reszler, C., Switanek, M. B. & Truhetz, H. Convection-permitting regional climate simulations for representing floods in small-and medium-sized catchments in the Eastern Alps. *Nat. Hazards Earth Syst. Sci.* **18**, 2653–2674 (2018).
84. Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M. & Weingartner, R. Mountains of the world, water towers for humanity: typology, mapping, and global significance. *Water Resour. Res.* **43**, 1–13 (2007).

## Acknowledgements

The hydrological simulations and its driving climate simulations (CRCM5-LE) have been produced within the ClimEx project ([www.climex-project.org](http://www.climex-project.org)) funded by the Bavarian Ministry for the Environment and Consumer Protection. We acknowledge the substantial efforts from all involved ClimEx members, that went into setting up and running the hydrological model as well as creating the meteorological reference dataset and the bias-corrected and downscaled climate simulations. The CRCM5-LE and the hydrological simulations have been computed 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. MIB was supported by the Swiss National Science Foundation via a PostDoc.Mobility grant (Number: P400P2\_183844). DLS was supported by a joint collaboration between the Institute of the Environment and Sustainability at the University of California, Los Angeles; the Center for Climate and Weather Extremes at the National Center for Atmospheric Research; and the Nature Conservancy of California as well as NSF PRE EVENTS award 1854940. JMD was supported by the NSF PRE EVENTS award 1854940.

## Author contributions

M.I.B., D.L.S. and J.D. developed the study concept and RL developed the concept for the hydro-SMILE. M.I.B. developed the methodology with inputs from D.L.S., R.R.W., E.G.F.W. and R.R.W. developed and computed the hydrological simulations for the 50 SMILE ensembles and R.R.W. designed Supplementary Figure 1. M.I.B. performed all the analyses. M.I.B. and D.L.S. wrote the first draft of the paper, which was revised and edited by all co-authors.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43247-021-00248-x>.

**Correspondence** and requests for materials should be addressed to M.I.B. or R.L.

**Peer review information** *Communications Earth & Environment* thanks Ridwan Siddique and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Heike Langenberg. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021