

Interactive Discussion: Author Response to Referee #2

Southern Germany's 100-year flash flood discharge expected to increase by 30 % under an RCP8.5 climate

Paul Voit, Marie Hundhausen, Larisa Seregina, Hendrik Feldmann, Bodo Ahrens, and Maik Heistermann
NHESS Discussions, doi:10.5194/egusphere-2026-1229

RC: Reviewer Comment, AR: Author Response, ☐ Manuscript text

Dear referee,

thank you for taking the time and effort to review this manuscript, and for your open and critical assessment.

Please find our responses to your comments below. These should be considered as preliminary (part of the interactive discussion) since the actual implementation of changes depends on the editorial decision.

Thanks again for your efforts!

Kind regards,
Paul Voit on behalf of the authors

1. Comments and responses to the Referee 2

RC: *Emphasising the RCP8.5 scenario at the end of the century when reporting projected changes could be misleading. The study uses model runs based on this scenario. While CPM runs with 30-year time periods are only available with CMIP5 projections, I believe that emphasising the RCP8.5 scenario at the end of the century is problematic in terms of communication. This is because it has been shown to be a highly unlikely scenario (Hausfather et al., 2020). Hundhausen et al. (2023), who used four of the six model runs, based their analyses on global warming levels 2 and 3. It seems that the choice of the end of the century was driven by one of the other two CPM runs, which is not transient. However, given that this study cannot explore the role of climate sensitivity comprehensively due to the number of GCM-CPM combinations used and cannot disentangle the climate change signal from internal climate variability (see next comment), there is no need to attempt to provide an overall percentage change in flash flood hazard in the region as 'expected'. Also, the methodology used to derive the 'global delta T' values reported in Table 1 is not explained. I could not find this information in Hundhausen et al. (2023). How were these values derived? I may have overlooked some explanations in Hundhausen et al. (2023), but I believe this is a significant issue, particularly in light of the challenges associated with RCP8.5.*

AR: We thank the reviewer for raising this important point. We agree that the wording of the original manuscript could give the impression that the reported 30% increase represents an expected future change under a likely

emissions pathway. This was not our intention.

Our analysis is based on a specific ensemble of CPM simulations forced by CMIP5 GCM simulations under RCP8.5. Our main aim was to compile an ensemble of CPM runs which - although small - would allow us to account for the large uncertainties. RCP8.5 was the only scenario available to allow for such a mini-ensemble. The choice of the 2071–2100 period follows the temporal coverage of the available CPM simulations and should therefore be understood as a late-century climate projection under RCP8.5, rather than as a prediction of the most likely future climate. We will revise the title, abstract, results, and conclusions accordingly and no longer use the term "expected" when referring to the 30% change.

We also agree that the use of RCP8.5 requires a more careful framing. We now explicitly state that RCP8.5 is used here as the forcing scenario available for the CPM ensemble considered in this study and as a high-end warming scenario, rather than as a statement about the likelihood of future emissions, and that the corresponding level of warming might not be expected by the end of the century.

Regarding the global temperature differences listed in Table 1, we agree that their derivation was insufficiently documented and apologize for that. ΔT was computed by comparing the global average temperature from the first simulation period (1971-2000) to the global average temperature of the last simulation period (2071-2100). We will add this information to the manuscript. Internally, we have also discussed to use regional warming rates but defining the size of the region remains arbitrary as precipitation in southern Germany can be caused by larger scale circulations. We hence stuck to the commonly used global average ΔT to characterize the individual ensemble members because the precise comparison of regional warming rates is not subject of this study.

Finally, we will remove wording suggesting that the analysis provides a comprehensive assessment of climate sensitivity. The six-member ensemble represents a limited sample of possible climate responses, and the revised manuscript makes this limitation explicit.

RC: *In light of the climate ensemble used, investigating changes in the 100-year flash flood is not adequate. While I acknowledge that the authors discuss this issue several times in the paper, I believe that the model ensemble used in this study does not allow for an analysis of changes in 100-yearly flash floods between the historical and future periods. Firstly, it is not reasonable to assume that enough extremes are sampled within 30-year periods. There is a 26% chance that at least one 100-year (or rarer) event will occur in a 30-year time series. In my opinion, looking at a change signal between two 30-year periods, even with a statistical fit, is therefore not reasonable. Pooling all climate members to improve this would mix the questions of climate sensitivity and sampling, which, in my opinion, contradicts the initial aim of providing insights into projections, as stated in the title. Secondly, several studies have shown that, in order to study extreme events such as floods and their changes due to climate warming, natural variability needs to be quantified (e.g. Aalbers et al., 2017; Brunner et al., 2021; Willkofer et al., 2024). One way to sample more extremes is to perform a spatial aggregation across the different basins by emphasising the median results, as done by the authors. However, using this method to examine 100-year floods would require proper quantification of the spatial independence of events. In my opinion, the results of this study would be more reliable if the authors had looked at return periods below 10 years.*

AR: We agree with the reviewer that 30-year samples do not provide a robust basis for estimating absolute 100-year return levels at individual catchments. This limitation was acknowledged in the original manuscript, but we agree that its implications for the interpretation of our results were not sufficiently emphasized.

We would like to clarify the specific question addressed by our analysis. We do not intend to provide an observationally constrained estimate of the absolute 100-year flood magnitude at individual catchments.

Rather, we investigate whether a spatially aggregated change signal in the upper tail of simulated discharge distributions emerges consistently between the historical and future climate periods across a large number of catchments and across multiple CPM simulations. In this context, spatial independence is not a required criterion in terms of the triggering rainfall (where spatial independence is in fact hard to define at all) while it is guaranteed in the hydrological sense due to the spatial uniqueness of each catchment. The combination of fitting GEV parameters for each catchment (for extrapolation) *and* aggregation of changes across the entire study area may not seem ideal, but is a pragmatic choice to get a hold on events with a magnitude that is actually impact-relevant (which is typically not the case for events with return periods lower than 10 years). The paramount importance of the 100-year-flood in flood risk management justifies this approach, given that the limitations are sufficiently discussed (which we will expand in the revised manuscript). To complete the analysis we will also add the analysis for the 10-year flood.

As for the question of climate sensitivity, we should emphasize that is not a prime subject of our study (see also below responses), and we will clarify this in the revised manuscript and a revised title ("Changes in 100-year flash-flood discharge in southern Germany based on a convection-permitting climate model ensemble").

RC: *The hydrological assumptions made in this study are not fit for purpose, and they lack evaluation and references to support them. The authors begin section 3.2 with the following statement: 'Flash floods are governed by the formation and concentration of quick runoff components. Therefore, groundwater and evapotranspiration processes can be neglected which immensely simplifies the structure of the required hydrological model'. Firstly, such a statement should be referenced. Secondly, I believe that it may be challenging to find suitable references to support this statement, as it does not align with studies on runoff generation processes. Although the SCS method coupled with a GIUH has been used to study flash floods in ungauged basins, several studies have shown that surface runoff does not necessarily control the entire flood hydrograph, but that instead subsurface processes are important even in flash flood prone areas (Roux et al., 2011; Garambois et al., 2014; Zhang et al., 2021). Furthermore, the explanations in the method section do not clarify how antecedent wetness conditions are considered, especially given that evapotranspiration and groundwater are entirely neglected. Does the model run for only five days before each storm, or continuously between May and October? Thirdly, this study does not rely on any hydrological modelling evaluation. In my opinion, the provided references do not demonstrate that the used hydrological model is fit for purpose; they only show that counterfactual simulations are useful for studying flash floods, which is not the current paper's focus. Although, flash flood observations are scarce (as mentioned by the authors), the model should be evaluated against streamflow data where available, for a few events (not necessarily extreme) and locations at least. Using relative change signals, i.e. comparing simulations between the historical and future periods, does not address this issue because the bias is likely to be non-stationary, especially if intense precipitation events trigger flash floods under conditions not observed in the historical records.*

AR: We agree that our original formulation was too categorical. The statement that groundwater and evapotranspiration can be neglected could be interpreted as implying that subsurface processes are unimportant for flash-flood generation. This is not what we intend to claim.

The hydrological model is deliberately formulated as a parsimonious representation of the dominant processes controlling the magnitude and timing of quick runoff at the catchment outlet. Groundwater and evapotranspiration are not represented as explicit dynamic components. This does not imply that subsurface processes are absent from the physical system. Rather, their effects are implicitly represented through the antecedent wetness conditions used in the SCS-CN runoff-generation component. The quick runoff represented by the SCS-CN method includes also subsurface runoff as stated in the (USDA-NRCS, 2004, page 10-2).

In flood hydrology baseflow is generally dealt with separately, and all other types are combined into direct runoff, which consists of channel runoff, surface runoff, and subsurface flow in unknown proportions. The curve number method estimates this combined direct runoff.

We will therefore revise Section 3.2 to make this distinction explicit. We will also discuss the resulting limitation of the model more explicitly.

The reviewer also correctly points out that the original manuscript did not describe sufficiently clearly how antecedent wetness is represented. We have added a detailed description of the model spin-up and precipitation history preceding each event. The model is running continuously between March and October and the soil moisture is calculated based on the precipitation in the previous five days.

Finally, we agree with the reviewer that a relative future/historical change signal does not eliminate hydrological model uncertainty or guarantee stationarity of model errors. We will consequently remove wording suggesting that the relative-change approach "addresses" this uncertainty. Instead, we will state that using relative changes reduces sensitivity to certain systematic biases, but does not remove structural or potentially non-stationary model biases.

Flash floods are characterized by quick (surface or near-surface) runoff components (Georgakakos, 1986; Marchi et al., 2010; Grimaldi et al., 2010; Borga et al., 2014). Having said all that, we would like to maintain that the model setup used in our study is widely used across a large number of flash-flood-related studies. This does not mean superior model approaches are unavailable. But clearly, the aim of our study is not to provide a hydrological modelling benchmark experiment (and only within a benchmark experiment with competing models can performance metrics could be reasonably evaluated), but to use an established and efficient modelling approach for a consistent comparison of flood events. While it might be possible that modeling biases might not be stationary, this limitation applies to any model, as does the difficulty in quantifying this non-stationarity in the case of extreme events. While we agree to discuss the limitations of the model in further detail, we do not agree to expand the study's scope towards the assessment of hydrological models under extreme event conditions.

RC: *The hydrological modelling assumptions may implicitly enforce flash flood scaling in line with precipitation intensity scaling. Due to the assumptions made within the hydrological modelling framework (i.e. no evapotranspiration, a limited antecedent wetness window and no subsurface processes influencing runoff), the response of the basins to increasing precipitation may be linear by construction. The authors discuss how the flash flood scaling observed in their study aligns with the widely reported C-C scaling for precipitation intensity. Apart from the fact that such scaling should be assessed using local temperature changes rather than global ones, I think the scaling may be linear due to the modelling assumptions made. While this is a reasonable assumption, I do not think it is supported in this study. Assuming that flash floods are not modulated by any subsurface processes could lead to misunderstandings.*

AR: We agree with the reviewer that the structure of the hydrological model may constrain the relationship between changes in precipitation and changes in simulated flood peaks. We also agree that our original comparison with Clausius–Clapeyron scaling could be interpreted too strongly as evidence for a physical scaling relationship between precipitation and flood extremes.

The purpose of the comparison with Clausius–Clapeyron scaling was to provide context for the magnitude of the simulated change, rather than to demonstrate that flash-flood extremes obey Clausius–Clapeyron scaling. We will revise this section accordingly.

In particular, we will explicitly state that the hydrological model does not allow us to disentangle the contributions of precipitation changes, antecedent wetness, runoff generation, and runoff concentration to the resulting flood scaling. We therefore no longer interpret the similarity between the approximately 30% increase in flood return levels and the corresponding precipitation scaling as evidence that flash floods physically follow Clausius–Clapeyron scaling.

At the same time, it is obvious and well established in the literature that the relationship between precipitation and runoff generation with the CB-approach is non-linear and modulated by antecedent wetness (while the temporal concentration of runoff is represented independently through the GIUH). The resulting discharge response therefore depends on the rainfall, its temporal structure of rainfall, static catchment characteristics and dynamic antecedent wetness.

RC: *No evaluation is carried out to assess the ability of the CPMs to reproduce precipitation observations for the different sub-basins, and no bias correction is applied. The authors state that using a ratio (i.e. a relative change signal) accounts for potential biases in the CPM simulations implicitly. However, biases can be non-stationary, and as the authors also state, the rainfall-runoff relationship is non-linear. Therefore, biases in the CPM simulations could be translated non-linearly into flash flood projections. An evaluation of the CPM performance is required, along with clear explanations showing how this assumption is valid in the context of flash flood projections.*

AR: We agree with the reviewer that the original statement that the use of relative changes "implicitly accounts for" CPM precipitation biases was too strong. A relative change does not in general remove systematic model biases, particularly if those biases are non-stationary or if the rainfall–runoff relationship is nonlinear.

We will therefore revise this statement throughout the manuscript. We now describe the future/historical ratio as a way of reducing sensitivity to certain systematic biases that affect both periods similarly, while explicitly acknowledging that this does not eliminate precipitation biases or their propagation through the hydrological model.

However, the performance of extreme precipitation for the ensemble members KIT-MPI, KIT-CNRM, KIT-ECE, KIT-HAD where thoroughly evaluated in Hundhausen et al. (2024) for DWD-MIROC5 it was evaluated in Rybka et al. (2023), and the GUF-ECE member in Purr et al. (2021). Except for GUF-ECE this publications are also cited in the table summarizing the ensemble members. We suggest to add Purr et al. (2021) to the table to avoid confusion.

RC: *The definition of a flash flood is not explicitly given. From the paper, I understand that any flood occurring between May and October in any of the sub-basins is considered a flash flood. This is a strong assumption that should at least be made explicit and justified based on data or suitable references.*

AR: We thank the reviewer for pointing out that our definition of flash floods was not sufficiently explicit.

We agree that restricting the simulations to March–October does not, by itself, constitute a definition of a flash flood. The seasonal restriction was introduced because the study focuses on convective heavy-rainfall events and because the hydrological model does not include a snow routine. It was not intended to imply that every flood occurring during this period is necessarily a flash flood. Another important component of our study is, however, that we restrict our analysis to basins with a size of up to 750 km². Since the overall study area is characterized by mountainous conditions, these basins would qualify as flash-flood prone, mostly showing response times of less than six hours. We will clarify the corresponding terminology throughout the revised manuscript.

RC: *There is only one result plot. While I appreciate short, concise papers, I believe this falls short of the*

requirements of a journal such as NHESS. Many other analyses could be performed to provide further insight into the simulations conducted in this study. For example, how do the different change signals relate to the hydrological variability between large numbers of sub-basins? How do changes in precipitation intensity drive flash flood responses? Could changes in precipitation characteristics cause a shift in flash flood seasonality? Some more explicit analyses of the variability between different CPMs in generating flash floods in future could be performed. The maps in the supplementary materials could be used, at least partially, in the main manuscript to illustrate the different change ratios.

AR: In our opinion, the number of result plots is not a criterion to evaluate the quality of the manuscript, and we do not agree with the idea to inflate the content if we have a clear scope. In order to address the referee's concern, we suggest, however, to further condense the paper and to make it a "brief communication" (see https://www.natural-hazards-and-earth-system-sciences.net/about/manuscript_types.html), a manuscript type for which the number of plots and figures fits exactly.

References

- Aalbers, E. E., Lenderink, G., van Meijgaard, E., van den Hurk, B. J. (2017, September). Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? *Climate Dynamics*, 50, 4745–4766. doi:10.1007/s00382-017-3901-9
- Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E., Ludwig, R. (2021, August). An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Communications Earth and Environment*, 2. doi:10.1038/s43247-021-00248-x
- Garambois, P. A., Larnier, K., Roux, H., Labat, D., Dartus, D. (2014, February). Analysis of flash flood-triggering rainfall for a process-oriented hydrological model. *Atmospheric Research*, 137, 14–24. doi:10.1016/j.atmosres.2013.09.016
- Hausfather, Z., Peters, G. P. (2020, January). Emissions – the ‘business as usual’ story is misleading. *Nature*, 577, 618–620. doi:10.1038/d41586-020-00177-3
- Roux, H., Labat, D., Garambois, P.-A., Maubourguet, M.-M., Chorda, J., Dartus, D. (2011, September). A physically-based parsimonious hydrological model for flash floods in Mediterranean catchments. *Natural Hazards and Earth System Sciences*, 11, 2567–2582. doi:10.5194/nhess-11-2567-2011
- Willkofer, F., Wood, R. R., Ludwig, R. (2024, July). Assessing the impact of climate change on high return levels of peak flows in Bavaria applying the CRCM5 large ensemble. *Hydrology and Earth System Sciences*, 28, 2969–2989. doi:10.5194/hess-28-2969-2024
- Zhang, G., Cui, P., Gualtieri, C., Zhang, J., Ahmed Bazai, N., Zhang, Z., . . . Lei, M. (2021, December). Stormflow generation in a humid forest watershed controlled by antecedent wetness and rainfall amounts. *Journal of Hydrology*, 603, 127107. doi:10.1016/j.jhydrol.2021.127107

References

- Borga, M., Stoffel, M., Marchi, L., Marra, F., and Jakob, M.: Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows, *Journal of Hydrology*, 518, 194–205, 10.5194/egusphere-2026-122910.1016/j.jhydrol.2014.05.022, 2014.
- Georgakakos, K. P.: On the design of national, real-time warning systems with capability for site-specific,

- flash-flood forecasts, *Bulletin of the American Meteorological Society*, 67, 1233–1239, 1986.
- Grimaldi, S., Petroselli, A., Alonso, G., and Nardi, F.: Flow time estimation with spatially variable hillslope velocity in ungauged basins, *Advances in Water Resources*, 33, 1216–1223, 10.5194/egusphere-2026-122910.1016/j.advwatres.2010.06.003, 2010.
- Hundhausen, M., Feldmann, H., Kohlhepp, R., and Pinto, J. G.: Climate change signals of extreme precipitation return levels for Germany in a transient convection-permitting simulation ensemble, *International Journal of Climatology*, 44, 1454–1471, 10.5194/egusphere-2026-122910.1002/joc.8393, 2024.
- Marchi, L., Borga, M., Preciso, E., and Gaume, E.: Characterisation of selected extreme flash floods in Europe and implications for flood risk management, *Journal of Hydrology*, 394, 118–133, 10.5194/egusphere-2026-122910.1016/j.jhydrol.2010.07.017, 2010.
- Purr, C., Brisson, E., and Ahrens, B.: Convective rain cell characteristics and scaling in climate projections for Germany, *International Journal of Climatology*, 41, 3174–3185, 10.5194/egusphere-2026-122910.1002/joc.7012, 2021.
- Rybka, H., Haller, M., Brienien, S., Brauch, J., Früh, B., Junghaenel, T., Lengfeld, K., Walter, A., and Winterrath, T.: Convection-permitting climate simulations with COSMO-CLM for Germany: Analysis of present and future daily and sub-daily extreme precipitation, *Meteorologische Zeitschrift*, 32, 91–111, 10.5194/egusphere-2026-122910.1127/metz/2022/1147, 2023.
- USDA-NRCS: Chapter 10: Estimation of Direct Runoff from Storm Rainfall, *National Engineering Handbook, Part 630 Hydrology 210-VI-NEH*, U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, DC, <https://directives.nrcs.usda.gov/sites/default/files2/1712930608/7300.pdf>, accessed 22 September 2026, 2004.