Interactive Discussion: Author Response to Referee #3

From Worst-Case Scenarios to Extreme Value Statistics: Local Counterfactuals in Flood Frequency Analysis

Paul Voit, Felix Fauer, and Maik Heistermann
NHESS Discussions, doi:10.5194/egusphere-2025-4951

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

Dear Referee,

thank you for serving as referee for this manuscript and for your comments and suggestions. The quality of the manuscript will certainly improve on the basis of this review.

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, Felix Fauer, and Maik Heistermann

RC: ***The paper would benefit from clearer definitions and a distinct conceptual separation between several key ideas that are currently addressed somewhat implicitly. In particular, the notion of "local counterfactuals" should be more rigorously defined early in the manuscript, including its hydrological interpretation and how it differs from related concepts such as storm transposition, spatial counterfactuals, and regionalisation. While the background section is strong, readers unfamiliar with these concepts may find it difficult to immediately understand what is novel versus what is adapted from existing approaches.***

AR: We suggest to add the following description of existing concepts to the introduction.

- **Regionalization**: Data from hydrologically similar catchments are incorporated into the estimation of distribution parameters to enhance the robustness of extreme value analysis (EVA) (e.g., Gaume et al., 2010; Guse et al., 2010; Nguyen et al., 2014; Halbert et al., 2016).
- **Probable maximum precipitation** (PMP): rainfall events from a "meteorological homogeneous" transposition domain are included in the analysis to increase the robustness (Fuller, 1914; District and Morgan, 1916) and to estimate the PMP (Hansen, 1987; WMO, 2009). Instead of exceedance probabilities this method only yields upper and lower bounds of precipitation. PMP can be used to estimate the upper bounds of a probable maximum flood (PMF), if used as forcing of a hydrological model. While PMP is widely applied in North America and Australia for designing high-risk infrastructure (e.g., dams and nuclear power plants), it is not prominently used in Europe. However, in recent years various studies regarding flood risk management have proposed and investigated different concepts of storm transposition, referring to the idea as "spatial counterfactuals" (Montanari et al., 2023; Merz et al., 2024; Voit and Heistermann, 2024; Vorogushyn et al., 2024; Thompson et al., 2025).
- **Stochastic storm transposition**: Building on the PMP/PMF concept, historical HPEs from a transposition domain are sampled using a Poisson distribution and randomly assigned (uniform distribution) within the domain, potentially affecting the catchment of interest (CoI). For flood frequency analysis, the resulting runoff in the CoI is simulated (e.g., Wright et al., 2014). This approach allows for the calculation of occurrence probabilities. For a detailed description see Wright et al. (2017). Globally, stochastic storm transposition (SST) remains rarely applied in practice (Wright et al., 2020) but it will form the core of the U.S. Federal Emergency Management Agency's "Future of Flood Risk Data" initiative, aimed at remapping the nation's floodplains (Abbasian et al., 2025).
- **Stochastic weather generators** are statistical models that simulate sequences of weather variables, such as temperature and precipitation, by randomly generating data based on observed patterns. They can be used to generate very long time series of meteorological forcings for a hydrological model (e.g. Falter et al., 2015; Apel et al., 2016).

RC: ***The selection of a 30 km radius and ten neighboring catchments seems reasonable but is largely based on empirical judgment. The manuscript should offer a clearer rationale for these choices, whether based on meteorological homogeneity, hydrological similarity scales, or sensitivity analysis.***

AR: We agree that the choice and combination of similarity metrics is a pragmatic one – an expert guess, if you will. The only way to actually *assess* the validity of our similarity metrics is to compare them against others in a kind of benchmark experiment: considering our KDTree-approach as a filter, the question behind such an experiment would be which filter could provide the best results in terms of improvement of our performance metric (i.e. the QSS). In fact, we did this when we analysed the sensitivity of the QSS to different neighborhood radii around our CoI. But while it would be highly interesting to expand such an analysis to the similarity metrics, we think that this is beyond the scope of the present study in which we rather aim to introduce a *framework* and provide proof-of-concept. We will, however, expand our discussion of the limitations of similarity metrics, and also outline perspectives for future research to assess the validity of similarity metrics.

RC: ***The criteria used to define catchment similarity via the KDTree deserve further explanation.***

AR: Yes, and your comment is in line with the other referees. The catchment similarity is a critical point. We added more information (l. 123 ff) to further describe the process.

We based similarity mostly on descriptors of topography, land use and soil which should i) strongly govern the formation and concentration of surface runoff and ii) ensure that potential orographic effects could occur both in the CoI and the NCs. Following descriptors were chosen:

- Peak [m³/s], time to peak [s] and standard deviation [m³/s] of the unit hydrograph: The unit hydrograph is derived directly from the DEM, similar hydrographs imply, to a certain degree, similar topography.
- Upstream catchment area
- Curve number (soil moisture class 2): The curve number represents soils and land use in our model. A similar curve number would lead to a similar runoff generation in our model.
- Mean and standard elevation of the DEM and mean slope. With this descriptor we try to avoid sampling rainfall events from catchments which are e.g. situated at a substantially different elevation. If the CoI was e.g. close to a mountain range, rainfall events should not be sampled from this mountainous area, because they might not be representative for the rainfall events occuring in the CoI.
- Unit Peak Discharge: The peak of the unit hydrograph divided by the catchment area is yet another descriptor of the hydrological character of the catchment.

We used the KDTree-algorithm from the Python library "SciKit-Learn" and scaled all catchment descriptors with the "StandardScaler" from this library to ensure that none of this descriptors dominates the decision for similarity. However, we acknowledge that some descriptors are correlated.

RC: ***Applying an uncalibrated SCS-CN and GIUH-based lumped model across thousands of catchments introduces significant uncertainty. The model mainly captures fast runoff generation and overlooks Hortonian runoff, slow flow components, and spatial variability in precipitation. This is especially important for catchments with winter flood regimes, where soil saturation and slower processes may prevail. The authors should discuss how these simplifications could impact both annual maxima and GEV tail behaviour.***

AR: The referee is right. Uncertainties introduced by the hydrological model were already discussed in section 5 of the preprint, and we will expand this discussion based on the referee's comment. Here, we would like to maintain that for extreme events, we assume slow flow components to be negligible at the scale of small catchments. It is correct that high flows may occur during winter subject to saturated soils and low evapotranspiration or in spring subject to snow melt events, and that these high flows may constitute some of the annual maxima and hence govern return levels for small return periods. The tail behaviour of small catchments, however, is clearly dominated by convective heavy rainfall that occur in the summer. As for the effect of Hortonian surface runoff (infiltration excess), the referee is correct that this process is inadequately represented in the SCS-CN framework. We assume that this would lead to an underestimation of runoff generation for extreme events at very short durations, or, in other words, that a model that represents infiltration excess would lead to heavier tails of the GEV distribution.

RC: ***The conclusions would benefit from clearer guidance on when the proposed method may be unsuitable.***

AR: Thank you, this was also mentioned by another referee. We added following sentence:

In regions with high orographic gradients or highly heterogeneous rainfall patterns the proper size of the TD might have to be reduced or optimized in benchmark experiments similar to the one carried out in this study.

RC: ***The manuscript frequently refers to "worst-case scenarios", but this term is not clearly defined.***

AR: The manuscript mentions the term "worst-case flood" (or scenario) two times: in the title and in the conclusions. But the referee is completely right that it lacks clarity, an issue that was also noted by the other referees.

In fact, we think that the term "worst-case" is not required in the context of our study. Based also on other comments, we changed the manuscript title to "Considering rainfall events from a neighborhood improves local flood frequency analysis". We will remove the sentence in which the term occurred in the manuscript (ll. 305 ff. of the preprint) as it does not essentially relate to the subject of our study.

RC: ***The appropriateness of using annual maxima with maximum likelihood estimation for such short samples could be briefly discussed in relation to alternative approaches such as POT or L-moments.***

AR: We agree. We have started using the ML method in our first publications but have also realized that L-Moments is considered to be the more stable method. We also agree in regard to POT but this method is also not widely used by practitioners due to its more complex application, e.g. for the definition of the threshold. We are currently involved in a study in which we compare extreme rainfall events in southern India which are evaluated by an index based either on annual maxima or POT. Based on the results of that study, we might move to the POT method in the future to increase the robustness of our assessment. With regard to this manuscript, we suggest to add a brief discussion to the section "Limitations" in which we should point out that the use of a POT approach together with the Generalized Pareto distribution (GPD) might be preferable in case of limited time series lengths, and certainly compatible with the framework presented in our manuscript.

References

Abbasian, M., Wright, D. B., Notaro, M., Vavrus, S., and Vimont, D. J.: Flood frequency sampling error: insights from regional analysis, stochastic storm transposition, and physics-based modeling, Journal of Hydrology, p. 133802, 2025.

Apel, H., Martínez Trepat, O., Hung, N. N., Chinh, D. T., Merz, B., and Dung, N. V.: Combined fluvial and pluvial urban flood hazard analysis: concept development and application to Can Tho city, Mekong Delta, Vietnam, Natural Hazards and Earth System Sciences, 16, 941–961, 10.5194/egusphere-2025-495110.5194/nhess-16-941-2016, 2016.

District, M. C. and Morgan, A. E.: Exhibits to Accompany Report of the Chief Engineer, Arthur E. Morgan: Submitting a Plan for the Protection of the District from Flood Damage, Miami Conservancy District, 1916.

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H., and Merz, B.: Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain, Journal of Hydrology, 524, 182–193, 10.5194/egusphere-2025-495110.1016/j.jhydrol.2015.02.021, 2015.

Fuller, W. E.: Flood flows, Transactions of the American Society of Civil Engineers, 77, 564–617, 1914.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, Journal of hydrology, 394, 101–117, 2010.

Guse, B., Hofherr, T., and Merz, B.: Introducing empirical and probabilistic regional envelope curves into a mixed bounded distribution function, Hydrology and Earth System Sciences, 14, 2465–2478, 10.5194/egusphere-2025-495110.5194/hess-14-2465-2010, 2010.

Halbert, K., Nguyen, C. C., Payrastre, O., and Gaume, E.: Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods, Journal of Hydrology, 541, 90–98, 10.5194/egusphere-2025-495110.1016/j.jhydrol.2016.01.017, 2016.

Hansen, E. M.: Probable maximum precipitation for design floods in the United States, Journal of Hydrology, 96, 267–278, 10.5194/egusphere-2025-495110.1016/0022-1694(87)90158-2, 1987.

Merz, B., Nguyen, V. D., Guse, B., Han, L., Guan, X., Rakovec, O., Samaniego, L., Ahrens, B., and Vorogushyn, S.: Spatial counterfactuals to explore disastrous flooding, Environmental Research Letters, 10.5194/egusphere-2025-495110.1088/1748-9326/ad22b9, 2024.

Montanari, A., Merz, B., and Blöschl, G.: HESS Opinions: The Sword of Damocles of the Impossible Flood, EGUsphere, 2023, 1–20, 10.5194/egusphere-2025-495110.5194/egusphere-2023-2420, 2023.

Nguyen, C. C., Gaume, E., and Payrastre, O.: Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations, Journal of Hydrology, 508, 385–396, 10.5194/egusphere-2025-495110.1016/j.jhydrol.2013.09.058, 2014.

Thompson, V., Coumou, D., Beyerle, U., Ommer, J., Cloke, H. L., and Fischer, E.: Alternative rainfall storylines for the Western European July 2021 floods from ensemble boosting, Communications Earth & Environment, 6, 427, 2025.

Voit, P. and Heistermann, M.: A downward-counterfactual analysis of flash floods in Germany, Natural Hazards and Earth System Sciences Discussions, 2024, 1–23, 10.5194/egusphere-2025-495110.5194/nhess-2023-224, 2024.

Vorogushyn, S., Han, L., Apel, H., Nguyen, V. D., Guse, B., Guan, X., Rakovec, O., Najafi, H., Samaniego, L., and Merz, B.: It could have been much worse: spatial counterfactuals of the July 2021 flood in the Ahr valley, Germany, Natural Hazards and Earth System Sciences Discussions, 2024, 1–39, 10.5194/egusphere-2025-495110.5194/nhess-2024-97, 2024.

WMO: Manual on estimation of probable maximum precipitation (PMP), `https://library.wmo.int/viewer/35708/?offset=#page=1&viewer=picture&o=bookmarks&n=0&q=`, last accessed: 18 September 2024, 2009.

Wright, D. B., Smith, J. A., and Baeck, M. L.: Flood frequency analysis using radar rainfall fields and stochastic storm transposition, Water Resources Research, 50, 1592–1615, 2014.

Wright, D. B., Mantilla, R., and Peters-Lidard, C. D.: A remote sensing-based tool for assessing rainfall-driven hazards, Environmental modelling & software, 90, 34–54, 2017.

Wright, D. B., Yu, G., and England, J. F.: Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects, Journal of Hydrology, 585, 10.5194/egusphere-2025-495110.1016/j.jhydrol.2020.124816, 2020.