Novel extensions to the Fisher copula to model flood spatial dependence over North America

Duy Anh Alexandre, Chiranjib Chaudhuri, and Jasmin Gill-Fortin

Responses to referee #2

1. **Are all the gauging stations are located in natural river basin? What if the gauged flood data is affected by reservoir operation? How to consider the influence of human intervention in floods.**

   In our study, we did not make an explicit distinction between stations in natural river basin and affected by reservoir operation. Nonetheless, during our preprocessing step to select high-quality stations, we tested the riverflow stationarity with a Mann-Kendall test and discarded stations where the trend is considered significant. This has a side effect of eliminating stations where human intervention can cause a disruptive change in the discharge time series. Furthermore, among the 130 static covariates used to predict fluvial flooding in our model, two are linked to human intervention and may capture the specific effect linked to stations with significant human intervention. They are percentage of urbanized land (LULC7) and the total upstream area protected by dams based on the GOODD dataset (log_dam_area). The GOODD dataset contains more than 38,000 dams as well as their associated catchments, allowing the analysis of their impacts in hydrological studies [1].

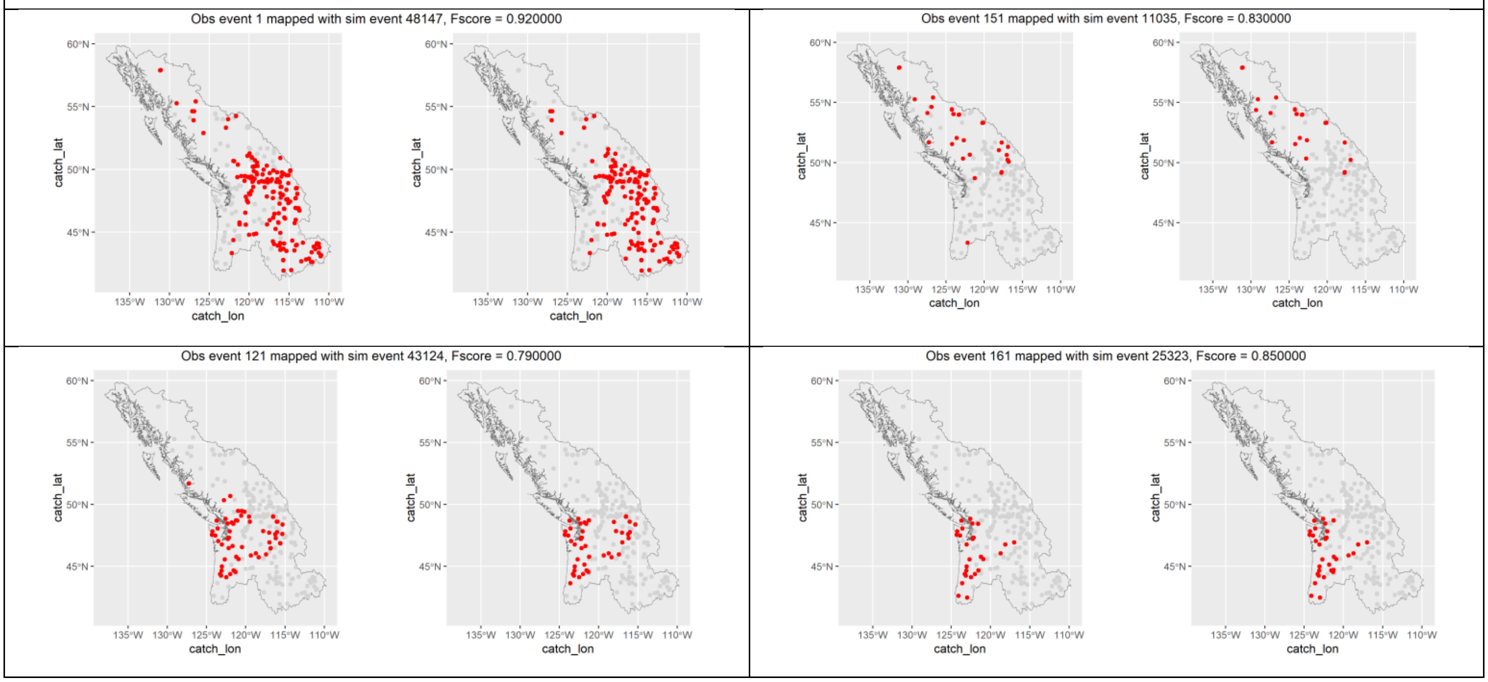
2. **The research mainly focus on the analyzing peak magnitude and total quantity of floods. Actually, the entire processes of flood hydrograph are worth more attention. Is it possible to show the simulation results over some specific flood events considering the entire time horizon of a flood.**

   Our study of fluvial flood spatial dependence focuses on the extreme floods and as such, the peak magnitude for the whole duration of each flood event is the main quantity of interest. As such, each flood event is summarized by its peak flow for every impacted location, so our simulated floods do not describe the whole time horizon of the flood. Other simulating components can be integrated to simulate the flood duration, hydrograph curvature, time of peak flow, etc... but we consider these developments to be outside the scope of the current study, which focuses on the spatial dependence of extreme fluvial floods (as summarised by the flood peak magnitude).

3. **How to verify the model results of the simulated flood footprint, as shown in Figure 11 d. Can the model compare the simulated footprint result with some actual floods?**

The simulated flood footprint was validated using various aggregated metrics to be able to compare them to the observed floods. For example, the size of each event footprint (defined as the number of gauges impacted by a significant flow during a flood) is validated against the observed event set (figure 7 and 8, manuscript). As such, it is not straightforward how to compare a simulated event directly with another historical flood event. However, for each historical flood event, we found that a simulated event with a similar footprint is present in the simulated event set (this matching is done using the F1 score to calculate similarity between event footprints). We present a non-exhaustive example of 4 historical floods with distinctive patterns, and their best matching simulated flood events.

Historical and simulated event matching



4. **Is there any specific techniques in generating multi-site footprint to reduce computation effort? Would the computation of river-basin wide dependent flood cause trouble?**

The conditional simulation technique described in section 4.2 of the manuscript is precisely developed with the aim of generating multi-site footprint events in a computationally efficient way, without having to calculate and invert a complete correlation matrix for all unit catchments, which would be a substantial computational burden. Using this approach, for a given simulated flood event, realized flood values for all gauged stations are used to sequentially simulate values at each ungauged catchment. Each region in our model (following the HydroBASINS level 2 delimitation) contains $O(10^2)$ gauged stations and $O(10^5)$ ungauged catchments. This computation step is very fast and when run on 20 CPUs in parallel on a personal computer, takes less than 10 minutes to be completed for each region.

5. **390-395, what is number of catchments with flow > half the return period of event magnitude. The two paragraphs read confusing. Can you explain more over Figures 12 and 13.**

Figure 13 represents the proportion of catchments in a given event with flow return period higher than the flow corresponding to half the maximum flow return period (RP). This is more difficult to articulate in words than to understand. Suppose a flood event has a maximum flow (in RP scale) corresponding to RP100. Then we would count the number of catchments exceeding their respective RP50 flows (50 = 100/2), and divide by the number of catchments affected by the flood event (defined as having greater than RP5 flow). If another event had a maximum peak flow corresponding to RP20, we would count the number of catchments exceeding their respective RP10 flows (10 = 20/2), etc. The idea is to quantify for each flood event, the extent of the most impactful region. If this proportion is low, this means that the most extreme floods are localized at a few catchments, even if the flood event footprint can be large.
Figure 13 then plots the histogram of that proportion for all events (1), for events with maximum flow in the range RP5-RP20 (2), for events with maximum flow in the range RP20-RP100 (3), and for events with maximum flow greater than RP100 (4). Please note that this same metric was used in [2] for evaluation of their simulated event set, so we chose to follow the same approach as to produce a result plot comparable to figure 11 in [2].

References

[1] Mulligan, M., van Soesbergen, A. & Sáenz, L. GOODD, a global dataset of more than 38,000 georeferenced dams. *Sci Data* **7**, 31 (2020). https://doi.org/10.1038/s41597-020-0362-5

[2] Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., et al. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55, 1890–1911. https://doi.org/10.1029/2018WR024205