Response to the interactive comments to the manuscript **hess-2020-396**

"Do small and large floods have the same drivers of change? A regional attribution analysis in Europe"

by M. Bertola, A. Viglione, S. Vorogushyn, D. Lun, B. Merz and G. Blöschl

We reproduce and number in the following document all the comments of the Referee *in italic characters*, followed by our answers.

Anonymous Referee #2

The manuscript addresses an important and popular topic in statistical hydrology: how and why are floods changing? General: The paper is well written and rooted in the literature. Flood attribution in this study is limited to three drivers, and the focus is on which physical process is relatively more important (as compared with other studies which have attempted to attribute floods to many different processes, e.g. Schlef et al 2019). I think the paper presents an interesting analysis, though I think its main conclusions and results are not well interpreted for the general scientific community, limiting the applicability and generalizability of the work. I give specific details in this review.

> We thank the Anonymous Referee #2 for the time she/he spent on our manuscript and for the useful and constructive comments that will help to improve the quality of the manuscript. We have carefully considered and addressed all her/his comments in the following.
>
> For the sake of clarity, the choice of the three drivers was driven by the results of recent studies (i.e. Blöschl et al. 2017, 2019; Berghuijs et al., 2019) that pointed out potential correlations between timing and magnitude of floods and extreme precipitation, soil moisture and snowmelt, across Europe. This study aims at formally quantifying the contribution of these drivers to flood changes, i.e. 'flood change attribution' and not 'flood attribution', as done in Schlef et al. (2019).

1. The manuscript's results do not convince me of the conclusions - maybe it is in the presentation - but I am not really convinced that the conclusions about the observed flood changes are valid. Without sufficient validation of the approach, it's unclear if the method performed as expected. For example, in choosing out a case from the dataset, can we validate that in fact within a region, extreme precip increased and floods increased for a q2 or q100 return period (not from spatial difference / relative contribution plots, but from actual time series and data within that region?

> We thank the reviewer for this comment; we understand that our approach should be further clarified in the manuscript.
>
> In the revised manuscript we will show actual time series and average flood and driver changes in one additional figure, to support our results for the three example regions analyzed in Sect. 3.3 (in line with the request of anonymous reviewer #1, see reply to SC1, nr. 10). Based on this additional figure we will indeed demonstrate that, within a region (e.g., Northwestern Europe), extreme precipitation increased, and floods

increased for a q_2 and a q_{100} return period. This study more generally suggests that the changes in flood quantiles potentially caused by the three considered drivers are overall compatible, in terms of patterns and magnitude, with the flood changes observed in previous studies (e.g. Blöschl et al., 2019; Bertola et al., 2020). Some discrepancies are nevertheless observed, for instance, in Scandinavia, where the contributions of the drivers are all positive or close to zero, while mostly moderate negative flood trends were observed in previous studies (see Sect. 4.2). In Sect. 4.2 we commented on possible reasons for this discrepancy (e.g. other potential drivers not accounted for in this study). We will revise the terminology used in the manuscript to clarify this point.

On the other hand, we did not cross-validate the model against data from additional stations or for other periods of time, because we do not aim at estimating driver contributions locally, in ungauged basins, nor at extrapolating the results of the model to the future. We are instead interested in the average driver contributions to changes in flood quantiles over the five analysed decades, and the results should be interpreted at the European scale.

Additionally, in order to avoid spurious correlations and to make sure that hydrologically meaningful contributions are identified, in the Bayesian framework we adopted informative prior distributions of the elasticity parameters (i.e. the parameters controlling the relationship between flood and driver changes), based on expert judgement and qualitative reasoning (Sect. 2.5). In practice, the informative prior distributions reflect the fact that flood and driver changes are expected to have the same sign (e.g. floods increased because precipitation increased, and positive flood changes cannot be attributed to negative precipitation changes).

We will clarify these points in the revised manuscript.

2. Introduction: While I agree in general that focusing on the mean/median can mask changes in the various return periods of a flow distribution, the mean is also traditionally an indicator of changes within the distribution and thus is an important piece of the story on how nonstationarity may be impacting a particular basin. Also after reading the paper, I am not certainly convinced that extreme precipitation is well aligned with the 100-yr event and would like to see more on the bounds of the 2 and 100 yr return periods. Speaking of return periods: in the spirit of helping to change the conversation from return periods to a more meaningful statistic, like reliability, I would recommend reframing the need to examine changes in floods from 'return period based' to something more robust. At the very least, return period must be well defined at the start of this paper: When the authors refer to return period in the manuscript, I think they mean "average return period" (e.g. Read and Vogel, 2015). Additional issues with the use of return period here: Please describe how the formulation of 2.1 holds true when $p = 1/T$ is no longer valid..). Can the Gumbel parameters be inferred from the 2- and 100-yr floods if the distribution is changing? I do not follow why the method for extreme precipitation was used. I am assuming there is a reason that this was made more complicated than pairing the flood data with the rainfall data in a more straightforward way. In using the average occurrence day, is there a chance that the actual highest precip/flood days are left out of the analysis (for example if they do not occur within the average window)?

Introduction: We will add the points mentioned by the reviewer in the introduction of the revised manuscript.

Return periods: In this study, we analyze changes in time of selected flood quantiles q, which are associated with fixed annual exceedance probability 1-p (in the notation used in the manuscript) through the quantile function q(p, $\xi(t), \sigma(t)$). In a non-stationary context, the pdf is a function of time and, consequently, also the flood quantiles (associated with fixed annual exceedance probabilities) change with time. The Gumbel parameters can be inferred from (time dependent) flood quantiles, associated with fixed exceedance probabilities. In the manuscript we refer to the return periods, rather than the annual exceedance probabilities, because they are widely used and understood in the engineering practice. Therefore, for ease of interpretation, the return period T is obtained from the annual exceedance probability 1-p through the relationship p=1-1/T, although other formulations are available under non-stationarity conditions. We do refer to the average return period as defined, for example, in Read and Vogel (2015). In this study, we directly model the changes in flood quantiles because, in a Bayesian framework it is typically easier for experts to formulate prior beliefs in terms of flood quantiles, which they are familiar with, rather than in terms of distribution parameters (see, e.g., the causal information expansion based on expert judgement in Viglione et al., 2013). Examples of return period terminology used in a similar non-stationary context in the literature are Renard et al. (2006), Machado et al. (2015), Šraj et al. (2016). For these reasons we prefer to maintain the return period terminology in the manuscript. However, we will clarify the terminology used in the method section 2.1. Using the reliability (as defined in Read and Vogel, 2015) instead of the return period, is not applicable to this context, because it requires the additional definition of the lifetime of a system/project. However, in the revised manuscript we will mention the existence of alternative ways of communicating event likelihood in stationary and non-stationary contexts, such as the reliability.

Extreme precipitation: We did not pair floods with the corresponding event precipitation because we do not aim at doing event attribution, but at attributing flood changes to the long-term evolution of the drivers in the average season of occurrence of floods. In other words, we use flood seasonality to identify drivers that are typically relevant for the generation of the annual peaks. The variability of flood seasonality in each station is taken into account by the width of the time window that is used to extract the 7-day maximum precipitation and snowmelt (i.e. if floods occur evenly distributed throughout the year, the width of the window is 12 months, and if floods occur always on the same date, this window is reduced to 3 months). We will clarify it in the revised manuscript.

3. Lines 270-272: With regard to elasticities specifically, why was a decadal % used to identify drivers? In the results generally, the interpretations of the individual elements are limited. For example, 273 "Extreme precipitation contributes positively to flood changes in northwestern and central Europe, and negatively in southern and eastern Europe". Also 276-77: "The contributions of snowmelt to changes in q2 and q100 are predominantly negative and marked in Eastern Europe, with small differences towards smaller contributions in absolute values with return period". It's a bit of work for the reader to translate this, using Fig 5. Put this in terms that are clearly translatable. This issue persists throughout the results, and clarification could especially be helpful in regard to the elasticities discussion. Translate elasticity into the meaningful metric that it is (e.g. in lines 286-88, an elasticity of zero indicates XXX).

The elasticities (Fig. 4, lines 256-269) are measured in %/% and represent the change in flood quantiles due to 1% change in one driver. Lines 270-272 and Fig. 5 refer to the contributions of the drivers to changes in q2 and q100. These contributions are obtained by multiplying the elasticity of flood quantiles to the drivers by the average trend of the corresponding driver over the region (in %/decade). The contributions are measured in %/decade because they represent the fraction of the trend in flood quantiles (which is measured in %/decade) that is explained by a specific driver. We will clarify this in the revised manuscript.

We will also rephrase the sentences pointed out by the reviewer in the results section.

4. Again, in 4.2, the results are restated, but are not communicated in a way that is useful for an audience. Take it a step further to explain how the sign/magnitude of the changes/relative contributions are meaningful in the context of the problem this manuscript is addressing. The Conclusions (4.4) really offer no further insight for the reader than the results. Suggest revising this to focus on implications.

We will rephrase section 4.2 and revise section 4.4 according to the suggestions of the reviewer.

Minor Comments:

5. Fig 8: why use a hypothetical catchment instead of selecting from those available within the study region?

The aim of this study is to estimate the average contributions of drivers to changes in flood quantiles within each region, and not for a single catchment in the region. For this reason, we do not represent the results for a catchment area corresponding to one specific existing catchment in the region. Instead, we show the average regional driver contributions for one hypothetical medium-sized catchment (i.e. catchment area of 1000 km^2). Consequently, the results should be read and interpreted at the large continental scale, rather than at the local (i.e. catchment) scale. We will clarify this in the revised manuscript.

6. lines 345, 348 spell out numbers below ten.

They will be spelled out in the revised manuscript

References:

Berghuijs, W. R. *et al.* (2019) ‘The Relative Importance of Different Flood-Generating Mechanisms Across Europe’, *Water Resources Research*, pp. 1–12. doi: 10.1029/2019WR024841.

Bertola, M. *et al.* (2020) ‘Flood trends in Europe: are changes in small and big floods different?’, *Hydrology and Earth System Sciences*, 24(4), pp. 1805–1822. doi: 10.5194/hess-24-1805-2020.

Blöschl, G. *et al.* (2017) ‘Changing climate shifts timing of European floods’, *Science*, 357(6351), pp. 588–590. doi: 10.1126/science.aan2506.

Blöschl, G. *et al.* (2019) 'Changing climate both increases and decreases European river floods', *Nature*, 573(7772), pp. 108–111. doi: 10.1038/s41586-019-1495-6.

Machado, M. J. *et al.* (2015) 'Flood frequency analysis of historical flood data under stationary and non-stationary modelling', *Hydrology and Earth System Sciences*, 19(6), pp. 2561–2576. doi: 10.5194/hess-19-2561-2015.

Renard, B., Lang, M. and Bois, P. (2006) 'Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: case study with peak-over-threshold data', Stochastic Environmental Research and Risk Assessment, 21(2), pp. 97–112. doi: 10.1007/s00477-006-0047-4.

Schlef, K.E., Moradkhani, H. & Lall, U. (2019) 'Atmospheric Circulation Patterns Associated with Extreme United States Floods Identified via Machine Learning', *Sci Rep* **9,** 7171 https://doi.org/10.1038/s41598-019-43496-w

Šraj, M. *et al.* (2016) 'The influence of non-stationarity in extreme hydrological events on flood frequency estimation', *Journal of Hydrology and Hydromechanics*, 64(4), pp. 426–437. doi: 10.1515/johh-2016-0032.

Viglione, A. *et al.* (2013) 'Flood frequency hydrology: 3. A Bayesian analysis', *Water Resources Research*, 49(2), pp. 675–692. doi: 10.1029/2011WR010782.