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 #1

The manuscript "Do small and large floods have the same drivers of change? A regional attribution analysis in Europe" by Bertola et al is the natural sequel of the previous HESS paper by some of the same authors (https://doi.org/10.5194/hess-24-1805-2020) taking the investigation from the detection to the attribution of changes in high flows of different frequencies. The manuscript is well organised and deals with a very interesting topic which I imagine will attract many readers. It is highly relevant for a European readership and presents an investigation of which physical variables appear to drive the magnitude of high flows in Europe differentiation between the common and the extreme high flows.

> We thank the Anonymous Referee #1 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.

1. In the introduction the authors frame their study within the current literature giving a nice excursus of what the current state of modelling change is. I have some disagreement on some of the language they use, though. They mention several papers saying that most studies focus on the change in the mean annual flood, which they then contrast to their interesting new approach. On the other hand though most studies I have seen in the literature (including those cited) focus on explaining the change in the location parameter (or sometimes the scale parameter) - but typically the mean flood would be a combination of all distribution parameters. So modelling a change in location typically reflect on a change in the mean flood, but the model aims at modelling some slightly different quantity. More importantly, when location and scale are both allowed to change the mean flood would change as a function of both parameters, so the model for the mean flood would be rather complex.

> In the introduction we meant that most of the studies about flood changes in the literature focus on changes in mean flood behaviour (not necessarily the mean annual flood), and they do not explicitly account for changes in flood quantiles with large return periods. For example, at lines 27-29 we refer to the trend detection studies that use tests (e.g. the Mann-Kendall test or the Petitt test) to detect changes in the mean flood behaviour.
>
> As the referee says, most studies in the literature about non-stationary frequency analyses allow the location parameter (and, more rarely, the scale parameter) to vary

in time using time-varying covariates, as described in the manuscript at lines 92-97. This translates to changes in the mean annual flood (or in flood quantiles), which is a function of the distribution parameters, although changes in model parameters are modelled. In our approach we focus on flood quantiles, rather than distribution parameters, in order to explicitly model the relationships between small and large floods (i.e. q_2 and q_{100}) and the potential drivers of flood change, and to separate the effects of drivers on floods associated with different return periods.

We will clarify this terminology in the introduction of the revised manuscript.

2. In equation 4 it is not very clear to me how the model is regional and each station contribute information to the model. I understand that all station-years contribute to the likelihood and things are then corrected using the likelihood inflation? I mean this is not a multilevel model in which station-specific parameters are allowed, is that right?

The referee is right; all station-years contribute to the likelihood and the likelihood is corrected using the magnitude adjustment to account for spatial cross-correlation between sites. This is not a multilevel model and the only station-specific parameter is the error term ε_g which accounts for the additional local variability not explained by catchment area and the covariates. We will clarify it in the revised manuscript.

3. Further, I understand that the model for the two quantities is estimated at the same time, so the q2 is "hidden" in the x100 model: to make this maybe more obvious I would use a bracket before to "connect" equation 4a and 4b.

It will be done in the revised manuscript.

4. I am also not entirely sure why no ε_g was allowed in the growth factor model. For those who might want to code this up themselves it might be helpful to have the formulae translating parameters to quantile and even more, to be able to read the Stan code - I would recommend that the authors share their code either via GitHub or via some more academic-oriented repository such as Zenodo.

The error term is not included in the growth factor x'_{100} because we make the assumption that the growth curve is the same across all sites within the region, while the median flood is allowed to vary between sites. This is similar to the index flood method of Dalrymple (1960) and Hosking and Wallis (1997). We will better explain it in the revised manuscript. In sect 2.1 of the revised manuscript we will also add the relationships linking the Gumbel parameters ξ and σ to q_2 and x'_{100} (as in Bertola et al., 2020), i.e.:

$$\begin{aligned} q_2 &= \xi + \sigma \cdot y_2 \\ x'_{100} &= \sigma(y_{100} - y_2)/(\xi + \sigma \cdot y_2) \end{aligned} \qquad (A1)$$

where $y_2 = -\ln(-\ln(0.5))$ and $(y_{100} - y_2) = -\ln(-\ln(0.99)) + \ln(-\ln(0.5))$.

We will share the stan code in GitHub with the revised manuscript.

5. To summarise: I think the model could be described with more details, especially for those who have not read the first paper on which this builds.

We thank the referee for her/his suggestions, we will improve the description of the model as detailed by our answers to the specific questions above (see answers to points 2 to 4).

6. Finally this is more of a curiosity, I was wandering what forms do the parameters functions take when one re-transforms the quantiles back to parameters. Can these shapes tell us something interesting about what types of functional relationship exist between the physical variables and the distribution parameters?

The relationships linking q_2 and x'_{100} to the Gumbel parameters can be obtained inverting eq. A1:

$$\begin{aligned} \xi &= q_2\left(1 - x'_{100}\frac{y_2}{y_{100} - y_2}\right) \\ \sigma &= \frac{q_2 x'_{100}}{y_{100} - y_2} \end{aligned} \tag{A2}$$

The regional change model adopted is:

$$\begin{aligned} \ln q_2 &= \ln \alpha_{2_0} + \gamma_{2_0} \ln S + \alpha_{2_1} \ln X_1 + \alpha_{2_2} \ln X_2 + \alpha_{2_3} \ln X_3 + \varepsilon \\ \ln x'_{100} &= \ln \alpha_{g_0} + \gamma_{g_0} \ln S + \alpha_{g_1} \ln X_1 + \alpha_g \ln X_2 + \alpha_{g_3} \ln X_3 \end{aligned} \tag{A3}$$

By substituting Eq. A3 into Eq. A2, we get:

$$\xi = \exp\left(\ln \alpha_{2_0} + \gamma_{2_0} \ln S + \textstyle\sum_{i=1}^{3} \alpha_{2_i} \ln X_i + \varepsilon\right) - \left(\frac{y_2}{y_{100}-y_2}\right) \cdot \exp(\ln \alpha_{2_0} + \ln \alpha_{g_0} + (\gamma_{2_0} + \gamma_{g_0}) \ln S + \textstyle\sum_{i=1}^{3}(\alpha_{2_i} + \alpha_{g_i}) \ln X_i + \varepsilon)$$

$$\sigma = \frac{1}{y_{100}-y_2} \cdot \exp(\ln \alpha_{2_0} + \ln \alpha_{g_0} + (\gamma_{2_0} + \gamma_{g_0}) \ln S + \textstyle\sum_{i=1}^{3}(\alpha_{2_i} + \alpha_{g_i}) \ln X_i + \varepsilon)$$

As a result, the functional relationships between the drivers and Gumbel parameters have a similar structure to those expressed in Eq. A3 between the alternative parameters and the drivers. Additionally, we observe that the second term in the location parameter has little influence, given that $\frac{y_2}{y_{100}-y_2} \cong 0.086$.

7. I find the modelling strategy of the authors quite interesting because they effectively model two quantiles which are indeed of interest rather than the parameters: should we then ditch the standard parametrisation of the Gumbel distribution or are the parameters still useful?

From a practical perspective, flood quantiles are clearly attractive, since flood risk managers are indeed interested into these quantities (e.g. the 100-year flood) and their changes in time. 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 associated with large return periods, 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). The distribution parameters are nevertheless fundamental as they determine the location and spread of the distribution.

8. Regarding the choice of the priors: the authors choose to set a hard bound on the elasticity parameters: did this create any problem in the estimation? I mean: is the posterior distribution very concentrated on the lower bound or does it spread nicely?

The introduction of these hard bounds in the priors is done in order to hydrologically ‘inform’ the attribution analysis. The elasticity parameters are, in fact, reasonably expected to be positive, given the selected drivers (corresponding to changes of the same sign in drivers and floods). In most of the cases/regions that we considered across Europe, the posterior distributions of the model parameters look nicely spread.

The referee is kindly referred to the example figure below (Fig. A1), where the posterior distribution of the elasticity of q_2 and q_{100} to the three drivers are shown for the three regions analysed in Sec. 3.3, located in northwestern, southern and eastern Europe (see Fig. 1 of the manuscript for the location of these three regions), respectively. In few cases, when the covariate change and the flood change have different signs, the posterior distribution of the related elasticity parameter is concentrated on the lower bound. This can be observed for example in Fig. A1c in the case of the posterior distribution of the elasticity of q_T to antecedent precipitation, which slightly increases over time, while flood magnitude decreases for both T=2 and 100 years. We will clarify it in the revised manuscript

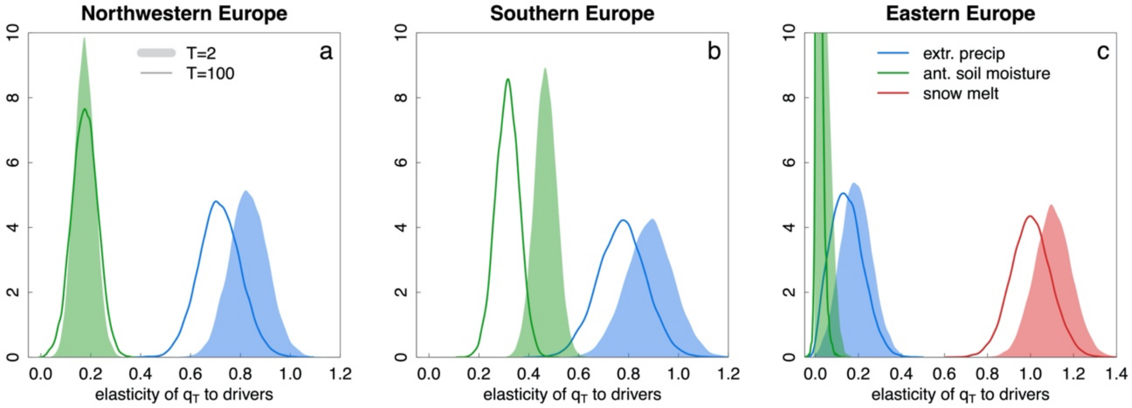

Figure A1: Posterior distribution of the elasticity of q_T to the drivers in three regions respectively located in northwestern (a), southern (b) and eastern (c) Europe.

9. I am somewhat dubious about the pooling of stations done by the authors and the use of averaged quantities across the rather large 200km x 200 km squares. To begin with the pooling will necessarily pool together information on small basins and large basins: this might not be problematic but I am more worried that with such large squares the pooling will put together very different types of basins (for example, alpine small basins and lowland larger basins): the response these basins have to drivers might be very different. Since from my understanding there aren't station specific parameters in the model, there might be some issues with the homogeneity of the groups and the ability of the model to identify the effect of the drivers on high flows. On the other hand, the average value of such large square might be not very useful to explain the variability of high flows for small basins and possibly inflate the variability of the results. I don't really see a way of out of this - I think the authors made some pragmatic decisions to be able to perform their study, but I wander whether we can fully trust their findings. In a similar vein: some areas are much more densely gauged than others, allowing possibly for a more precise estimation. This is not mentioned at all in the current manuscript.

In this study we are interested in the average regional behaviour and flood attribution at the large scale. The results of the study should therefore be interpreted at the continental scale as average contributions of the drivers to flood changes in the regions, rather than at the catchment scale.

As in Bertola et al. (2020), flood data of multiple sites are pooled in this study within spatial windows of size of 600km×600km, with an overlapping length of 200km in both directions. The size and overlapping length of the windows were chosen in Bertola et al. (2020) after several preliminary tests, in order to ensure a sufficient number of gauges within each window and an appropriate spatial resolution at which to present the regional trends at the continental scale. Significant differences in spatial change pattern were not observed when changing the window size. The rationale behind the homogeneity assumption is that the spatial windows are characterized by comparatively homogeneous climatic conditions, flood generation processes and processes driving flood changes. The attribution analysis is thereby performed at the regional scale, where average regional contributions of the decadal changes in the drivers to average regional trends in flood quantiles are estimated. We have not assessed the statistical homogeneity of the regions in terms of the flood change model used here. One reason is that formal procedures to assess the regional homogeneity, such as those used in regional flood frequency analysis (e.g. Hosking and Wallis, 1993; Viglione et al., 2007), are not available in the context of the present model. Also, while deviation from regional homogeneity would probably invalidate estimates of local flood change statistics from the regional information (e.g. in the prediction in ungauged basins; see Blöschl et al., 2013), we expect its effect on the average regional behaviour to be less relevant.

We will acknowledge and clarify this assumption in the revised manuscript.

Catchment area (S) and the drivers (X1, X2, X3) are indeed station specific. The average regional values shown in Fig. 4-7 are obtained for hypothetical catchment area of 1000 km^2 and for average changes in the drivers in each region over the period 1960-2010.

The different density of stations across Europe clearly influences the precision of the estimation and it is taken into account by the width of the credible bounds, represented for each region by white circles in Fig. 4-7. We will mention this in the revised manuscript.

10. Figure 8 is very interesting, but maybe I would complement it with two other visuals which would be relevant: the changes in the precipitation, soil moisture and snowmelt in each of the regions (to make more sense of how the curves morph from row to row in Figure 8) and final change in the different quantiles between the beginning and the end of the recording period (or any two moments in time).

We thank the referee for her/his suggestion. We will add this information to Figure 8.

References:

Bertola, M., Viglione, A., Lun, D., Hall, J., and Blöschl, G.: Flood trends in Europe: are changes in small and big floods different?, Hydrology and Earth System Sciences, 24, 1805–1822, https://doi.org/10.5194/hess-24-1805-2020, 2020.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales, Cambridge University Press, https://doi.org/10.1017/CBO9781139235761, 2013.

Dalrymple, T.: Flood frequency methods, US Geological Survey, water supply paper A, 1543, 11–51, 1960.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resour. Res., 29, 271–281, https://doi.org/10.1029/92WR01980, 1993.

Hosking, J. R. M. and Wallis, J. R.: Regional Frequency Anal- ysis, Cambridge University Press, Cambridge, UK, p. 240, ISBN 0521430453, 1997.

Viglione, A., Laio, F., and Claps, P.: A comparison of homogeneity tests for regional frequency analysis, Water Resour. Res., 43, 1– 10, https://doi.org/10.1029/2006WR005095, 2007.

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