**Reply to Referee #1 Alberto Viglione:**

This paper presents a European data-based analysis of the correlation between a number of atmospheric indices and flood exceedance probabilities at the sub-annual timescale. The novelty of the paper is related to the extent of the study region, i.e., all of Europe. The outcomes are interesting because of the coherent spatial patterns of the identified correlations in climatically different parts of Europe. The paper is well written, properly concise and clear. I believe it can become a very valuable entry for HESS. However, as always, some improvements are possible. My main comments/criticisms/suggestions are the following:

**Response:** We would like to thank Alberto Viglione for his comments. The points he raises are relevant and addressing them will definitely improve our manuscript. We hope that after this discussion (as well as after we revise our manuscript) all issues raised can be clarified.

**General comment 1**

- Title: I do not think that it is possible to easily answer this question, it never is when dealing with extreme value statistics in the real world. Actually, while I consider very interesting the analysis of the correlation between the atmospheric indices and the parameters of flood exceedance probabilities, I am less convinced about the accuracy of flood frequency estimation provided here. The reason is that, in engineering hydrology, I think nobody would fit locally a GEV distribution using a likelihood-based method with no information on its shape parameter. Regional analysis is normally used to improve quantile estimation for high return periods (say 100-years) which is not performed here. I would agree that the paper provides an indication that there is potential for improving flood frequency estimation by including atmospheric dynamics in our models, but I guess there is much more to do to actually improve the existing regional models in use. Maybe this is what the Authors meant but, to me, the title is a bit misleading. In my opinion, a title that focuses on the identified correlations between atmospheric indices and local floods would be better.

**Response:** We found the comment very important and will adopt the reviewer's recommendation concerning the title of our manuscript. We agree that a title focusing on the identified links between circulation indices and local floods better describes our methodology and findings. We will thus change the title accordingly and we will focus on the streamflow-climate interactions. In addition, we will reduce the extrapolation to high return periods and we will calculate streamflow quantiles for a probability of exceedance 0.02 (50-year return period). The common time period of streamflow data and circulation indices is between 50-70 years, so the extrapolation and possible uncertainty from the absence of a regionalization framework is considerably reduced. A comment will be added in the discussion about the possibility of improving quantile estimation by using a regionalization framework.
Finally, an informative prior distribution will be used for the shape parameter, in order to constrain the shape parameter from adopting unreasonably high or low values and to improve the GEV fits (see also our reply to general comment 2).

**General comment 2**

- It is always strange, to me, to see studies that use Bayesian inference without using prior information, especially when some very useful prior information is out there. For example, for the GEV shape parameter of the stationary model (but also for the nonstationary one) I recommend using (at least) the "geophysical" prior in Martins and Stedinger (2000). They demonstrate that without a prior on the shape parameter of the GEV, maximum-likelihood estimates (and therefore presumably also Bayesian estimates) for hydrological samples are much less accurate than those obtained with other methods (e.g., L-moments). The "geophysical" prior in Martins and Stedinger (2000) provides "common sense" limits, but even better constraints could be obtained through regional analysis, of course.

**Response:** We found this comment very useful and we repeated the analysis with an informative prior for the shape parameter. The "geophysical" prior of Martins and Stedinger (2000) is bounded in the interval (-0.5, 0.5). There are studies, however, that have found shape parameters of hydro-climatic data higher than 0.5 (e.g. Papalexiou and Koutsoyiannis, 2013). For this reason we considered the restriction of the "geophysical" prior a bit strict and used instead a normal distribution with similar characteristics with those of the "geophysical" prior. The normal distribution allows more extreme shape values but with small probability. The prior we chose is the empirical distribution of the shape parameter found by Papalexiou and Koutsoyiannis (2013) when they fit the GEV to annual precipitation time series worldwide. To our knowledge, it is the study investigating the highest number of hydro-climatic data worldwide. Of course, streamflow may be characterized by slightly higher shape parameter than precipitation. We will comment on this issue in the discussion and conclusions section.

**General comment 3**

- The motivation for assuming that only the location parameter varies in time (through its relationship to the covariates) should be discussed more in detail. Considering the proposed model, with the scale and shape parameters fixed, implies that the variance of the flood series does not change over time (e.g., if the mean annual flood peak increases of 5 m3/s, also the 100-yr flood increases of 5 m3/s, and so all other quantiles).
Is this a reasonable assumption? For example, Serago and Vogel (2018) strongly criticize it and propose to use models with the coefficient of variation of the flood series constant in time, since that is consistent with observations in many studies (see the cited literature there). Using a model where CV is constant would be as parsimonious as the one used here and, according to Serago and Vogel (2018), more justified. I suspect that using this other assumption would not invalidate the spatial patterns that are shown in Figures 1 to 4, but would result in very different values in Figures 5 to 7.

**Response:** The reviewer is right, indeed both the location and scale parameter are expected to change based on the climate state. However, we tested for significance of varying scale parameter by running the model with both location and scale variable. This preliminary study showed only very few cases with significant slopes of the scale parameter. For this reason and for reasons of parsimony, we decided to keep the scale parameter stable and to condition only the location parameter on the climate indices.
The model with constant coefficient of variation (CV) is an interesting alternative to the model that we present in our manuscript. However, investigating additionally this model would lead to a different and considerably extended manuscript. We feel that such a change is beyond the scope of this paper. We will comment on the possibility of a constant coefficient of variation (CV) in the discussion and conclusions section.

**General comment 4**

- One issue I would also suggest to discuss is the uncertainty in the covariates. The model used here assumes that the covariates are exactly known. If the uncertainty in their knowledge would also be included, would the flood quantile estimates still be more precise than for the classical GEV model?

**Response:** Thank you for this comment. In our manuscript we investigate only contemporaneous relationships between climate indices and flood peaks and do not focus on prediction. Our goal is mainly to identify spatial patterns of these relationships. For this reason we assume that covariates are exactly known. Of course if one wants to use the current model in a predictive mode, the uncertainty in the covariates must be additionally considered. We will add a comment on this issue in the discussion.

Additional detailed comments:
Line 20: I would expect that the improvement of estimation of flood probabilities is conditional on how well the covariates can be predicted.

**Response:** The reviewer is right. We will rephrase this sentence in analogy to our detailed response to general comment 4.

Line 71: I am a little confused by the positive-negative anomalies vs. Northernsouthern Europe because the sentence terminates with “during its positive state”.
Maybe a rephrasing could help.

**Response:** In the new version a change will be made from “Particularly NAO has been shown to significantly influence the European winter climate with positive (negative) anomalies of moisture fluxes, cyclone passages and precipitation over northern (southern) Europe during its positive state” to “Particularly NAO has been shown to significantly influence the European winter climate: its positive state has been linked to positive (negative) anomalies of moisture fluxes, cyclone passages and precipitation over northern (southern) Europe”.

Line 107: The motivation of using Bayesian inference because of the quantification of uncertainty sounds a bit weak. The quantification of uncertainties is possible also with other methods than Bayesian, which is instead usually selected when subjective preferences or prior information is available (at least by us... statisticians have more profound reasons).

**Response:** We will add a sentence stating that the Bayesian framework is justified since prior information on the shape parameter exists in literature.

Line 150: The climate covariates are assumed exactly known in the method. Would it be possible to account for the fact that they are stochastic variables as well? I do not ask to change the method but maybe some discussion could be dedicated to this issue (see main comments).

**Response:** See our reply to general comment 4.

Line 158: The motivation for assuming that only the location parameter vary in time, i.e., the brevity of records, is not very convincing. The Authors should discuss it more (see main comments).

**Response:** See our reply to general comment 3.

Line 174: Since Bayesian inference is done here, there is no reason why priors should not be used. For the GEV shape parameter of the stationary model I recommend to use (at least) the “geophysical” prior in Martins and Stedinger (2000) (see main comments).

**Response:** See our reply to general comment 2.

Line 174: Which non-informative priors are used? Not all of them would result in the same inference. For repeatability, they should be stated.

**Response:** We will add a description of the prior distributions used: uniform priors for the location and scale parameters and a normal informative prior for the shape parameter (see also our reply to general comment 2.

Line 184: I would also look at the posterior distribution of the slope parameter and do the same as the Authors do here. I would just add a sentence to state that this is not a significance test (which has no meaning in Bayesian statistics).

**Response:** We will add a comment on the fact that this is not a significance test.

Line 216: I worry that, for engineering purposes, the estimates of 100-yr floods through GEV without accounting for regional information is not to be recommended.

**Response:** See our reply to general comment 1.

Line 221: Starting the sentence with “Since a Bayesian framework is used” is confusing because it sounds like saying that uncertainties cannot be quantified with other methods too.

**Response:** We will rephrase this sentence.

Line 224 (and elsewhere): I would use the wording “posterior mode” instead of “maximum likelihood” because they may not be the same (it depends on the type of noninformative priors that are used). Bayesian posterior predictive distribution of flood peak quantiles or their posterior mean could have also been used. Is there a reason for choosing the posterior mode?

**Response:** The posterior mode was used in order to make results comparable with those of frequentist approaches. In the revised manuscript we will use the posterior median of flood peak quantiles, because it is more representative of the posterior distribution.

Line 258: Is there any (even speculative) reason for the contradicting patterns for Scandinavia?

**Response:** We will shortly discuss potential reasons for the deviant behavior in the revised manuscript version. In this regard, we will particularly refer to the special catchment characteristics in Scandinavia.
The fact, that the integration of seasonal mean climate indices in the flood frequency analysis improves the extreme value distributions for most catchments in Central Europe indicates, that catchment wetness (due to variations of seasonal precipitation sums) might play an important role for flood generation in those regions. In contrast, Scandinavian rivers usually have small catchments and are particularly fed by snow melt in spring and both, temperature and precipitation, may be important for runoff generation. A positive state of SCA is associated with negative precipitation anomalies (Supplement 2), but also with positive anomalies of temperature and incoming solar radiation (not shown) and vice versa. Thus intense snow melt events might be more likely during dry conditions associated with a positive SCA index.

Line 265: I also believe that the coherence in space is indicative of a real signal, however spatial correlation of the flood time-series could be a nuisance here, meaning that one sees the same dynamics in many sites because the same floods are occurring there (and therefore they should count as one site only). Since the spatial patterns are here over very large regions, the spatial correlation of the flood time-series cannot alone be responsible for it. However, I would suggest mentioning the problem.

**Response:** We agree that spatial correlation of floods plays a role for the detected coherence particularly for smaller regions, i.e. nearby gauges. We will add a comment on the spatial correlation of the flood time-series.

Line 299: One curiosity. Since the proposed model has constant variance (and the dependence of small and large floods on the covariate is the same in terms of the difference in m3/s) I suspect the relative difference to be affected by catchment area (meaning by the average flow in the river). Is it the case? Of course, since the model is fitted independently to every site, the differences in fitted shape parameters will make this relationship noisier.

**Response:** Since the difference between streamflow quantiles for high and medium covariate is normalized by the streamflow quantile for medium covariate we were not expecting that the catchment size plays a role in the percent relative differences. We assume that the high relative differences are due to a stronger influence of the climatic indices.

Section 3.2: I wonder how much the relative differences calculated here are due to the slope of the regression for the location parameter vs. the estimated shape parameters. Since no priors are used, I suspect that the posterior distributions of the shape parameter can be wide (spanning unreasonable values) and widely different between sites. Maybe a figure/table that also informs the reader about the obtained shape parameters would be useful.

**Response:** We will add a table with summary statistics of the shape parameter.

Figure 7: Shouldn't be the classical GEV the same within each column? The credible bounds look different. Have I missed something?

**Response:** We are sorry, this was a typo error noticed by the reviewer and was corrected.

Lines 335: The asymmetry of the credible bounds around the posterior mode is very well expected. If the posterior predictive distribution (or posterior mean) would have been used, that would have lied much more in the center of the bounds.

**Response:** The posterior median is used now. Indeed credible bounds are less asymmetrical.

Line 352: I would add here a brief discussion on the predictability of the covariates since that is needed to make use of the model for prediction.

**Response:** See our reply to general comment 4.

Line 357: I don't get the meaning of "...leads to highly varying flood quantile estimations for different probabilities of exceedance". Is the sentence referring to variations in time? Or space? Or between models with different covariates? And, finally, the variability for "different probabilities of exceedance" of flood peaks at one site exists in terms of relative differences. In term of difference in m3/s, there is no variability at all, since in the models only the location parameter can vary. Maybe I just misunderstood. A rephrasing could help.

**Response:** Indeed the variability concerns the relative differences. The sentence is referring to the comparison between the classical GEV and the climate-informed models. We will rephrase it to make this clearer.

Line 358: Related to my previous comment on line 299: is it because the catchments in North-West Scandinavia and Britain are smaller than the others? Or is it because of unreasonably large shape parameters of the GEV?

**Response:** The highly varying results in this area are in our opinion the result of a more important influence of the circulation indices. No influence of the catchment size was found.

Line 363: It is for me hard to see the decadal-scale variability in Figure 7. Maybe that could be shown in the figure.

**Response:** In the revised manuscript we will add a figure with the evolution in time of the climate indices, showing also their decadal-scale variability. We think that this will help the readers better interpret figure 7.

Line 402: As an additional challenge (on top of the three that the Authors have listed) I would add the fact that now covariates are assumed perfectly known and should be instead treated as stochastic variables, I think.

**Response:** See our reply to general comment 4.

**References:**

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res., 49(1), 187–201, doi:10.1029/2012WR012557, 2013.