Reply to Referee #3 Francesco Marra:

This study presents a methodology to assess and quantify the impact of climatic covariates in the estimation of time-dependent flood probabilities. The method is tested on a wide sample of catchments in Europe. The paper is clearly and concisely written and the topic is of interest for the readers of this journal.
In my opinion, the study should be seen as an additional step in the efforts of the hydrological community towards a better understanding and quantification of flood risk and, as well underlined by the authors, the spatial consistency of the results indicates some degree of significance in the adopted model. However, further steps are required before the suggested method can be effectively applied in practice. Both the reviewers before me pointed out very interesting comments, many of which I happened to share. I come last so I'll try not to overlap.

Response: We would like to thank Francesco Marra for his comments. The referee has shared very valid concerns that we hope to address in a revised version of our manuscript.

General comment 1
In general, my main concern derives from the GEV approach that requires a number of hypotheses and, if not integrated within a regionalization framework, is prone to extremely large uncertainties.

Response: The concern of the reviewer is valid. Indeed, a regional framework is commonly used in order to extrapolate inference to higher return periods. Here, focusing more on identifying coherent patterns in space, we use a local framework, which is able to recognise significant influence of certain indices to the extreme streamflow quantiles in certain regions of Europe. In order to reduce uncertainty, we will constrain our analysis to the 50-year return period which covers the data length: in our study the overlapping period between climate indices and streamflow time series is between 50 and 70 years. Furthermore, in order to improve GEV fits in the revised manuscript, prior information for the shape parameter will be included in the analysis (see also our response to main comment 2 of Alberto Viglione).

In addition to the references recommended by Elena Volpi, I may suggest the reading of Marani and Ignaccolo (2015), that provide a different perspective on extreme value analysis and the GEV approach (potentially for nonstationary extremes) that deserves attention.

Response: Thanks for this very interesting paper. We will consider it for the discussion of alternative models.

To conclude, I think the paper definitely deserves publication, but some more discussion and comments on the adopted methods are required. Potentially, some additional analyses could be of help. Below my detailed comments.

• Is the use of climate-informed models contradicting the identical-distribution assumption behind the use of GEV? This perhaps needs to be discussed.

Response: The condition of independent and identically distributed observations of GEV can be relaxed to include parameters conditioned on time-varying covariates. This can be achieved by converting the original data to generalized "residuals" that are identically distributed (Katz et al., 2002). We will add a comment on this issue in the revised manuscript.

• The inclusion of climate information in the model raises the number of parameters to be estimated to 4. Is there a risk of overfitting?

Response: Our climate-informed models have one more parameter compared to the classical GEV, derived by physical reasoning (the climate indices influence the climate and hydrology of the European area). According to e.g. Katz et al. (2002), such models are reasonable. Furthermore, our conditional models are compared to the classical GEV by means of the DIC, which penalizes model complexity. We are thus confident that the possibility of overfitting is minimized.

• How the authors explain that the linear model applied to the scale parameters (rather than location) provides similar results? Shouldn't the two parameters be related one another since the location is related the mean and the scale to the variance of the annual maxima? Is it correct to change one of them and keep the other fixed?

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 GEV approach is highly sensitive to the shape parameter, which is prone to large estimation uncertainty when derived from short data records (50 years), particularly when using the maximum likelihood method. Why not using an Lmoments estimation method? Could the inclusion of prior information on the GEV shape parameter improve the accuracy of the results? Perhaps this aspect should be addressed in the study to check consistency in the significant indices (in the end shape and scale are then used as prior information in the estimation of the climate-informed model parameters).

Response: Thanks for this interesting comment. The Bayesian approach was chosen because of its advantages concerning quantification of uncertainty. The choice of a likelihood-based method offers additionally a straightforward way of including covariates in the frequency analysis. The reviewer is right, inclusion of prior information on the GEV shape parameter improves the accuracy of the results. In our revised manuscript we will use an informative prior for our classical and climate-informed GEV. Please see our reply to general comment 2 of Alberto Viglione for more details on the prior distribution chosen. One more detail that we would like to highlight is that the posterior distributions of the parameters of the classical GEV are not currently used as priors for the climate-informed case. Each model is fitted independently.

• A linear model to relate climatic indices and GEV location parameter is chosen. Clearly, more complex models are not recommended due to the limited data sample and overfitting problems, but this represents a simplification of reality. How can this affect the results? This should be discussed.

Response: Of course, a linear univariate model is always a simplification and non-linear relationships between the state of climate indices and the European surface climate might exist, which also result in a non-linear influence on streamflow extremes. The choice of a linear model can affect results, since the effect of the covariate on the dependent variable may be overestimated or underestimated for certain covariate values. However, linear methods are frequently applied, particularly if the data sets do not allow to fit more complex models.
In our discussion and conclusions section we mention the possibility of an asymmetrical model for the positive and negative phase of the indices. In Lines 389-392 we write:
"A symmetrical influence of the positive and negative phases of the climate indices on the extreme value distribution has been assumed in this study. However, an asymmetrical relation may better describe the effect of certain climate modes on streamflow extremes. For example, Sun et al. (2014) used an asymmetric piecewise-linear regression to account for the different effects of El Niño and La Niña on rainfall extremes in Southeast Queensland, Australia".

In our revised manuscript we will extend this discussion to point out more clearly the limitations because of choosing a linear model as discussed above.

• What do the authors recommend for situations in which more than one climatic index is significant?

Response: A multi-linear model for such cases is also possible. For example, in our study there seems to be a significant influence of both NAO and SCA in winter in Central Europe. We comment on this issue in our discussion (Lines 383-387):
“Single covariate models were developed, focusing on the separate effect of each individual climate mode. The methodology can be extended to a model considering several covariates at the same time. In that case, dependencies between the covariates, if existent, should be taken into consideration. López and Francés (2013) overcame this problem by using the principal components of climatic indices as covariates for the flood frequency analysis”.

Minor comments:
- Lines 24-28 in the abstract are not easy to read, I suggest to rephrase them;

Response: We will rephrase them according to the revised manuscript.

- Introduction: the proposed method is of interest for (re-)insurance applications and for flood risk management. I think the design applications are not interested since year-by-year variability is not relevant

Response: Indeed the proposed framework makes more sense when the year-to-year variability is investigated. We will omit “engineering design” in line 51.

- 169-172: please provide more details for readers not familiar with the technique;

Response: We will add more details on the No-U-Turn Sampler-Hamiltonian Monte Carlo approach.

References:

Katz, R. W., Parlange. M. B. and Naveau, P.: Statistics of extremes in hydrology. Advances in Water Resources 25(8-12): 1287 – 1304, 2002.