Reply to Referee #2 Elena Volpi:

The manuscript investigates the effectiveness of performing climate-informed extreme value analysis for flood probability estimation at the European scale. More specifically, the Authors analyze the effects of large-scale circulation patterns on seasonal extreme distributions by accounting for the relationship between extreme probabilities and climatic indices. As stated by the Authors, climatic indices are considered in recent literature works to justify or explain a non-stationary behavior depicted by extreme events.
In this regard, the innovative contribution of this paper is to perform a large-scale analysis, at a spatial scale that is "comparable" to that of the climatic indices considered in the work aiming at defining the conditional probability distribution of extreme flood events and proving coherent spatial patterns.
The manuscript is well written and organized; the methodology is almost well described, even if additional details could be included to help for reader understanding, and conclusions are well supported by results. Finally, within the Conclusion Section a detailed list of the limitations of the study is provided. Summarizing, the topic is of interest for the scientific community and the manuscript deserves to be considered for publication in this Journal. I have some comments about the work that are listed in the following paragraph; I hope that they will be helpful for manuscript improvement.

Response: We would like to thank Elena Volpi for her comments. In the revised manuscript we will follow most of the reviewer's recommendations, since this will definitely improve our study. Below, we provide justification for some suggestions that we do not follow. We see from the comments of the reviewer that some parts of our study need a more detailed explanation. In our revised manuscript we will provide these additional details.

General comment 1

1. The Authors hint in the Introduction Section at the nonstationary framework incorporating climatic indices into flood frequency analysis, but they do not make a clear distinction between periodicity (or cyclo-stationarity) and trends (in the mean or variance). For the sake of clarity, this could be discussed from the very beginning of the manuscript (e.g. at line 49). Are the Authors assuming stationarity which is a "prerequisite to make inference from data", as discussed in detail by the cited papers by Koutsoyiannis and Montanari (2015) and Serinaldi and Kilsby (2015)?

General comment 2

2. At line 136 the Authors define the model driven by climatic indices as "climatic informed model", justifying this choice based on the fact that "if covariates have a stochastic structure and no deterministic component, the resulting distribution is not truly nonstationary". I do agree on this, as the Authors states at line 135 that the climatic indices are stochastic process not showing clear trends. But, it is expected they are characterized by persistence and/or periodicity. A detailed description of the stochastic behavior of the climatic indices is missing in the manuscript, while they are clearly described from a physical point of view (lines 59-90). E.g. which are the relevant time-period and is the period covered by observations long enough to catch climatic indices periodicities?

Response to general comments 1 and 2:

We want to thank the reviewer for these interesting comments. In our manuscript we acknowledge the issue that models conditional on time-varying covariates with a stochastic structure can be stationary, even if the probability density function changes in consequent years. However, we feel that addressing the issue of stationarity/nonstationarity (and thus ergodicity) and the stochastic structure of the covariates in adequate detail would change considerably the focus of our manuscript and we prefer not to make this addition. For this reason we chose the term "classical" and "climate-informed models" and we do not refer to stationary/nonstationary models. We will consider going more in this direction in our future work.

In the revised manuscript we will add a more detailed comment stating that “a detailed description of the stochastic behaviour of the circulation indices would be needed in order to argue in the direction of stationarity or nonstationarity”. Furthermore, we will add a comment in the introduction in order to make a more clear distinction between periodicity and trends in the description on nonstationary models. Finally, we will add a figure illustrating the evolution in time of seasonal climate indices.

General comment 3

3. Even if the aim of the work is to find results at the European scale, I would suggest the Authors to add a figure showing results for a single station, as an illustrative example to explain the methodology and the rationale behind it (e.g. the structure of the climatic informed GEV). Similar to figure 7, it could be of interest to show the evolution in time of the climatic indices (see comments 2) and the performance of classical GEV and climatic-informed GEV, especially for quantile extrapolation, with uncertainty bounds.

Response: We will adopt the recommendation of the reviewer and we will add an example figure illustrating the performance of classical GEV and climatic-informed GEV with uncertainty bounds for possible covariate values (additionally to the time-varying uncertainty bounds shown in figure 7). A figure with the evolution in time of the climate indices will also be added (see also our response to general comments 1-2). However, we do not plan to add a figure explaining the climate-informed model. As we also state in the introduction, during the last years there have been many studies applying such a conditional framework to single or a few stations. We feel there is enough published material explaining this methodology.

4. If I understand correctly, conditional models preferred o classical GEV in Table 1 are those respecting both criteria (minimum value of DIC and significantly different from zero coefficient of linear variation with the climatic indices); this could be highlighted in the result section from the beginning of the section. The number of times (stations) each conditional model is preferred with respect to classical GEV is not so high, being in the best case the 44% and on average at about 20%. The use of two criteria does not seem to affect this result much (as in lines 276-280); hence, the evidence of the climatic informed model does not appear to be very strong, even if clear spatial patterns emerge. The latter is the more relevant result, based on my opinion, and this should be stressed in the abstract and conclusion sections.

Response: We will adopt this suggestion and in the revised manuscript we will highlight the selection criteria in the result section. We will stress more clearly that the effect of each index independently is not always high (in many cases it affects a 20% of the database). However, the number of stations affected by at least one index significantly is much higher, especially in winter. We feel that this is a result that indicates a real influence of the circulation indices to the streamflow extremes.

5. Since spatial patterns are influenced by correlations among climatic indices (that are illustrated in the supplementary material as spatial maps), I suggest the Authors to report in the manuscript a table summarizing cross-correlations among the indices (even if they are not an exhaustive measure of the underling complex physical phenomena).

Response: We will adopt this suggestion and we will add a Table in the Supplementary material summarising linear correlations between the seasonal indices.

6. Lines 276-279. DIC is a measure of model evidence; even if the climatic informed model has a smaller value of DIC with respect to classical GEV, the difference among the two values is probably not enough to results in a “strong evidence” of the first model compared to the second one. See, e.g., Kass and Raftery (1995) where two different interpretations of the Bayes factor are provided.

• Kass, R. E., Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.

Response: Since we are using two criteria, the DIC and slope significance, we feel that the evidence is enough in the case of pairwise comparison with the classical model. In the case of comparison between the climate-informed models (for example Fig. 3-4) a comment will be added highlighting that more evidence may be required for a formal decision. In the discussion and conclusions section we already mention some of the limitations of DIC. We will discuss further the issue of "strong evidence" for model selection.

7. Figure 7 compares conditional (climate informed) and unconditional quantiles considering p=0.01 for three stations. It should be clearly stated that conditional quantiles are computed in this case based on the observed values of the climatic indices year by year.

Response: We will add a comment stating this year by year quantile calculation.

8. As the climate informed models have a larger number of parameters (one more in this case) to be estimated based on data, it is expected that their uncertainty bounds are larger than those provided by classical GEV. In other words, nonstationarity flood frequency analysis adds an additional component of uncertainty if the model between parameters and covariates is estimated from data and not fully a-priori defined based on additional physical information (Serinaldi and Kilsby, 2015). However, this is not what emerges from figure 7. This issue should be clarified.

Response: This was a very helpful comment. We realise that figure 7 and the conditional/unconditional uncertainty bounds need to be further discussed. Our figure does not contradict the findings and discussion of Serinaldi and Kilsby (2015).
The range of the uncertainty bounds is an interplay between the model complexity and the additional information provided by the more complex models. The relation between the two is not always trivial. In general, more complex models not providing extra information are expected to lead to an increase in uncertainty. More complex models providing "adequate" additional information are expected to lead to decreased uncertainty.
In figure 7, uncertainty bounds are narrower in the case of the "best" conditional models (e.g. plot A1). The bounds are also narrower for common values of the climate indices, while they can be larger than the classical case when extrapolations are made to uncommon (high and low) index values. In this case not enough extra information is available and uncertainty increases.
We will discuss this issue in more detail in the manuscript. We will furthermore add one more explanatory figure showing the uncertainty bounds versus the covariate values (see also our reply to general comment 3). This will better clarify our point and findings.

9. Lines 329-330. This should be true if the climate indices can be accurately predicted. The issue should be discussed further since it is closely related to the implications of the results presented in the paper for practical applications. Furthermore, I'm asking myself if the improvement in flood quantile estimate at the local scale thanks to climate indices is really significant from a practical point of view given the large uncertainty that characterizes all the estimates (fig. 7); I would like to read a comment on this from the Authors.

Response: A comment will be added in the discussion explaining the effect of the index uncertainty in the limitations of the study. Indeed if one wants to predict the indices for the next season in order to use them for the estimation of streamflow quantiles, uncertainties will be higher.

10. Line 58. The Authors could also consider the recent paper from Serinaldi et al. (2018) discussing limitations of nonstationary detection based on trend tests.
• Serinaldi, F., Kilsby, C. G., Lombardo, F. (2018). Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology.

Response: Thanks for this very interesting paper. We will consider it for the discussion of conditional / nonstationary model.

11. Line 71. A reference is needed.

Response: We will add it.

12. Please definite t after eq. (3).

Response: We will correct that.

13. Line 173. Please define what is meant by non-informative priors in this case. If the non-informative prior is a uniform distribution, its support (range of variability of the random parameter) could have effects of posterior distribution and evidence estimation.

Response: We will add a description of the prior distributions. In the revised manuscript we will use uniform priors for the location and scale parameters and a normal informative prior for the shape parameter.

14. Eq. (8). y or Y ?

Response: We will replace y with Y.

15. Line 194. Please define $\bar{\theta}$.

Response: We will add this definition.

16. Line 219. Are the Authors assuming a Gaussian (marginal) distribution for climatic indices? The assumptions on those variables and their stochastic behaviour are not clear (see also general comments 2 and 3).

Response: We are currently not making an assumption about the marginal distribution of climate indices. We will add a figure, however, in the supplementary material with the histograms of all seasonal indices so that the readers can get a better idea about their distribution.