Referee #2

Thank you for your comments and valuable suggestions. Below are our point-by-point responses to the referee's comments (in italic). We hope they will find them to be comprehensive and satisfactory. In the responses, we refer to specific figures or lines in the main text, to allow the referee to follow the changes implemented in the revised manuscript.

The topic of the manuscript is certainly within the scope of the journal and the use of a Bayesian Bayesian hierarchical framework for modeling seasonal extremes is groundbreaking; there are no other papers that I am aware of that utilize this approach for this purpose.

Response: Thank you very much for acknowledging the value of our work.

My few comments are mainly related to terminology, the underlying dataset, and a question about the title and future application of the framework (which, of course, would be beyond the scope of this study).

1) *L54 and L58: Use of the word "nonstationarity"*
   In L54, it appears that the term "nonstationarity" is meant in terms of things such as climate change, more akin to what we might think of as long term changes to the system as what non-statisticians think of as only nonstationarity; however, nonstationarity also refers simply to the seasonal signal in the streamflow time series. In L58, the study is asking whether the "representation of nonstationarity through suitable covariates improves season predictions…" Here it appears that you are referring to nonstationarity in the more precise statistical term of nonstationarity. It may be helpful to add a sentence or phrase in the introduction to define nonstationarity in the strict statistical terms so non-statisticians reading the text will not be confused. (I hope I did not confuse things!)
   Response:
   Thanks for the comment. In the paper, we define nonstationarity as temporal variability, i.e., year-to-year variability. However, the framework can also be applied in the climate change context if one has snow projections.
   Based on your suggestion, we modified line 58 to "How does the representation of nonstationarity (inter-annual-variability) through suitable covariates improve seasonal predictions?" **See line 61** in the revised manuscript.

2) *The study uses 7 streamgages to complete the testing of this framework. Looking at Table 1, UCRB7 is an outlier in drainage area, mean streamflow, and mean seasonal streamflow from the other streamgages. It is also located substantially further away from the other streamgages.*
   a. *How does the framework account for streamgages that are outliers. In Figure A4, the model certainly has a different behavior for CHRB7 for the cross-validation in only high flow years. Does this affect the*
      Response:
      As shown in Figure 3, the correlation between the stream gauges is similar for all the gauges, and station 7 does not behave as an outlier. We think that the issue that the reviewer noticed in Figure A4 is related to the skill of the covariate for a 1-month lead time rather than the use of the Gaussian copula. This is because all the models in Figure 4 consider a Gaussian copula. The only difference between them is the covariates (station 7's skill for other lead times is similar to those from the other stations). In addition, this decrease in skill

is more related to years with low flows as shown in Figure 4A of the supplementary material which shows higher CRPSS for a year with high flows only.

b. *How robust is your understanding of the spatial dependencies on performance skill when only these 7 streamgages are used? This is a key question you had planned to examine (L59)?*

Response:

Figures 8 (7) and 10 (9) in the revised manuscript (manuscript) show that the performance skill can be improved when a copula is added to the model. This means that the model can capture observed values of average maximum specific spring flow inside the ensemble spread of the projected flow (whisker of Figure 10b), which does not happen for the model without a copula (Figure 10a). Therefore, there is an increase in the Energy Skill Score when using the spatial model (Figure 8). All these metrics are multivariate. This model can be implemented for a larger spatial sample without reducing skill but at the cost of increasing computation time.

c. *Could you comment on why a much larger study area or set of streamgages was not used? If a limitation of this framework is that it cannot be applied to large streamgage networks, I wonder what implications this has for its practical applicability.*

Response:

In this application, we restricted the study region to the Upper Colorado River Basin (UCRB) since we wanted to ensure that the gauges considered were sufficiently correlated in space, requiring a multivariate modeling framework. There are more stream gauges in the UCRB, but they have missing values for the period 1965-2018. Although this framework can be applied to a more extensive stream gauges network, we do not recommend that since clusters of different streamflow behavior will develop as the region of interest increases. In that case, it is more efficient to fit a model for each cluster than fit a model for the entire region, which will be more computationally expensive. Fitting a model for each cluster allows using different covariates (more skillful) for each cluster too. We mentioned this in the Discussion section. **See lines 443-448** in the revised manuscript.

3) *Title and future applications*

a. *Of particular interest to the Upper Colorado is also the situation of drought prediction. Could this approach be useful for that situation as well? This is beyond the scope of this study but what would be some of the difference in applying the framework to low-flow extremes.*

Response:

Thanks for the question. Yes, the framework can be easily applied to predict low-flow extremes using the same marginal distribution (GEV) or another suitable distribution, and identifying suitable covariates. We mentioned this in the discussion. **See lines 432-433** in the revised manuscript.

b. *Along these lines, the title implies that "seasonal streamflow extremes" would mean both tails of the distribution (high and low streamflows); however, flooding is only examined here. Consider changing the title to reflect this.*

Response:

Thanks much for the suggestion. We changed the title to "**A space-time Bayesian hierarchical modeling framework for projection of seasonal maximum streamflow**."