Answers to reviewer 2 (Anonymous)

I would like to congratulate authors on the thorough and thoughtful response to the review comments. The level of clarity, completeness, and care with which you have addressed the points raised truly sets a high standard, and I very much appreciate the effort that has gone into improving the manuscript.

I have only one minor comment that I believe could further enhance clarity, although it does not affect my overall positive assessment.

In your response, you clarify that the objective is not to introduce a novel statistical downscaling methodology, but rather to apply and evaluate an established PCA feed multilinear regression based SDM framework across a pan-European domain. This is well justified. However, in the manuscript you note that coefficient level significance filtering is not applied, and robustness is instead assessed through testing a large set of configurations.

In this context, I was wondering whether an additional step, i.e. testing the significance of the PCs included in each model configuration (T1–T7), could help reduce unnecessary complexity. My expectation is that some of the more complex models (in terms of atmospheric predictors) might achieve similar skill with fewer statistically significant covariates. This would also be consistent with the multiple linear regression approach adopted in Tadese et al. (2020).
I believe this step is particularly relevant given the choice to apply PCA individually to each variable. While PCA ensures orthogonality among EOF–PC pairs within each variable, this orthogonality does not extend across variables. As a result, some degree of redundancy or interdependence between predictors may remain. An alternative could be to apply PCA jointly across all fields to enforce orthogonality between components; however, I understand that this may introduce practical challenges, as the resulting covariance matrix could become close to singular due to dependencies among variables.

➔We thank the reviewer for this comment. We would like to clarify that PCA is applied jointly across all (standardized) atmospheric variables, not individually to each variable. Specifically, all predictors (consisting of the different atmospheric variables, the different locations considered, and the different time lags considered) are concatenated into a single multivariate dataset, to which a single PCA is applied (as in Tadese et al. 2020). The orthogonality is therefore ensured across the full predictor space, which was illustrated by the variance inflation factor being close to 1 for the leading 10 principal components (which contain information across variables) across Europe in the response to the first round of reviews. For further clarity, we added the following comment in lines 172-173:

"*The procedure is applied once to a single matrix in which all predictors are concatenated, including all variables, spatial locations, and time lags*."

That said, it is totally possible that the complexity of the resulting configuration could be further reduced by retaining only principal components with statistically significant regression coefficients as in Tadesse et al. 2020. This was already evocated in lines 214-216 "*Coefficient-level significance filtering is not applied; robustness is instead sought through the assessment of a large set of configurations, acknowledging that some residual unnecessary complexity may remain.".*

For completeness, and as a possible avenue for improvement of the model, we have added the following comment on the discussion sections highlighting this possible additional step (lines 616-618):

Regarding possible improvements in the statistical method underlying our SDM, filtering principal components based on the statistical significance of their regression coefficients could reduce noise and improve generalization of the SDM (Tadesse et al., 2020).

Related to this, Table 2 could benefit from indicating how many principal components are required to explain 99% of the variance for each predictor variable. If I understand correctly, this effectively corresponds to the number of covariates entering each model (with PCA applied separately per variable), and including this information would improve transparency on model complexity.

➔Since the PCA is not applied per variable but on the matrix combining all variables x locations x time-lags, the number of principal components after dimensionality reduction depends on the combinations chosen across columns in Table 2 (i.e. on the resulting configuration), not only on the first column (atmospheric variables). This is why the resulting number of principal components was shown in Figure 3d,h for each configuration possible.

Additionally, a few very minor points:
In Table 2 (T6), I believe "Dmax" is intended for the squared wind components.

➔ Thank you, it has been corrected.

In Figure 2, one of the summations starts at i = 0 and the other at i = 1.

➔In the first round of reviews, Figure 2 was moved to Figure 4, and the notation in the Figure was changed to algebraic so the summations don't appear anymore.

The figure also highlights that PCA is applied per individual component. While this is perfectly valid, it might be helpful to clarify that index j refers to the EOF (I guess ordered by explained variance) of each variable (indexed by i).
➔ See previous response. Note that even if PCA is applied to a single matrix integrating all variables, the resulting principal components can still be separated per atmospheric variable (this was the meaning of the previously used notation). The current algebraic notation doesn't suggest any separation per variable.
Overall, I find the manuscript to be significantly improved and very clear in its scope and contribution.

We thank the reviewer for their feedback and positive evaluation.

Answers to reviewer 1 (Marta Marcos)

I appreciate the authors' efforts to respond and incorporate my comments. My main concern about the accuracy of the statistical model to reproduce extremes from the dynamical runs has been well addressed. I think the validation provided now, including qq-plots and comparisons of return periods, improves the manuscript by quantifying these uncertainties. The statistical model is surprisingly good. I did not expect such a good performance, after having seen some earlier applications of similar models to tide gauge observations. I think this is a strong point of the manuscript, that shows that the statistical model is able to simulate hydrodynamic outputs even if it may be not that accurate when applied to observations. The authors may want to highlight this point.
I find the discussion of statistical vs dynamical projections much more insightful in the current version. I think it is important to highlight the strengths of the approach but also, where the results of the statistical projections should be taken with caution.
I am fine with the changes in the structure that have resulted from the other reviewer's request.

➔We thank the reviewer for her comments and overall assessment. As suggested, we have added the following comment in the discussion to highlight the possible lower performance when targeting observed extremes (lines 629-631):

In fact, this underprediction tendency may partly explain the improved skill of the presented SDM in reproducing extremes in the dynamical model compared to previous studies benchmarking against observations, and may imply lower skill when targeting observed extremes.