COMMENTS FOR THE AUTHOR

Response to Reviewer 1

We would like to thank the reviewer for his/her helpful comments. Thank you. All of your comments have been taken into consideration, and the paper was modified accordingly. Please find below our responses.

Comment 1.

Editor: Clarify the procedure used to remove collinear predictors, particularly the rationale for excluding variables such as NAO, PDO, TNI, and QBO.

Referee 1: It remains unclear how you removed the collinear variables based on the VIF values. For instance, in the updated table (Table 2, page 17), I couldn't understand the rationale for removing NAO, PDO, TNI, and QBO, and not sure you followed the stepwise removal of collinear predictors (one predictor at a time), starting from the one with the largest VIF. For instance, unless there is a concrete reason, I would suggest removing the Variables with VIF values of Infinity. Then recalculate the VIF. If multicollinearity still exists, remove the Predictor with the highest VIF. Again, recalculate VIF, and so on. The Predictors NAO, PDO, TNI, and QBO already have low VIF when you calculate VIF using all variables, but you removed them. Please explain the rationale behind this and put all your steps into the supplementary material.

Answer: We thank the reviewer for this comment. In the revised manuscript, we clarified the predictor-screening procedure in both the Methods and Results sections. In Section 2.4.3 (*Multicollinearity diagnosis and variable selection*), we now explain more explicitly that predictor selection followed a screening procedure based on (i) the identification of redundant groups of predictors through pairwise correlation analysis and their correlation with streamflow, and (ii) the iterative assessment of multicollinearity using VIF, with recalculation after each elimination step.

In Section 3.2.3 (*Assessing multicollinearity and variable selection for improved time series forecasting*), we further clarified how this criterion was applied to the candidate variables. The revised text now highlights that, in addition to the strongly collinear ENSO and Atlantic SST indices identified through the correlation heatmap and VIF analysis, variables such as NAO, PDO, TNI, and QBO were not retained, not because they exhibited severe multicollinearity, but because they showed weaker absolute correlations with streamflow and provided less independent predictive value than the predictors included in the final subset. This clarification better distinguishes between variables excluded due to severe multicollinearity and those discarded because of their lower hydrological relevance for forecasting.

Changes introduced in the revised manuscript:

"2.4.3 Multicollinearity diagnosis and variable selection

A two-step screening procedure was applied to select the exogenous predictors used in the forecasting models. First, pairwise correlation matrices visualized as heatmaps were used to identify groups of highly correlated MVs, thereby detecting potential redundancy among predictors and identifying those with the strongest association with streamflow (Dormann et al., 2013). Second, the Variance Inflation

Factor (VIF) was computed to quantify multivariate collinearity among the candidate predictors (Katrutsa and Strijov, 2017).

Predictors with VIF values greater than 10 were considered excessively collinear and were iteratively removed. When several variables conveyed similar large-scale climatic information, selection was guided not only by the reduction of VIF values but also by the strength of their absolute correlation with streamflow, in order to retain the predictor showing the most hydrologically relevant signal while preserving model parsimony. After each elimination step, VIF values were recalculated until the reduced predictor set showed acceptable multicollinearity levels.

3.2.3 Assessing multicollinearity and variable selection for improved time series forecasting

…

Similarly, the high correlations observed among the TNA, NTA, and TSA indices indicate overlapping information content; therefore, only one of these indices was retained in the final predictor set to mitigate multicollinearity. In contrast, variables such as NAO, PDO, TNI, and QBO showed weaker absolute correlations with streamflow and, therefore, were considered less informative for forecasting purposes than the predictors retained in the final subset.

…"

Comment 2.

Editor: Better justify why the model improves peak flow prediction but performs similarly to or worse than the climatological benchmark in other cases.

Referee 1: In your modeling chain, only the peak flows are forecasted with better performance; most of the other forecasts are worse than the climatological benchmark. Please justify this and provide an aligned report on the abstract, discussion, and conclusions of your manuscript.

Answer: We have clarified that the climatological benchmark performs well under average monthly conditions because it effectively reproduces the dominant seasonal cycle. In contrast, the LOG-NNSARIMAX-MVNL model provides added value during peak-flow conditions by incorporating nonlinear precipitation signals and macroclimatic variability, which help capture departures from the historical monthly mean.

The following changes were made to the abstract:

"...These results underscore the effectiveness of integrating nonlinear precipitation patterns, macroclimatic variability, and temporal dependence to improve forecast accuracy under data-scarce conditions, particularly for peak flows and anomalously wet periods that are not adequately captured by the climatological benchmark…"

Accordingly, the third paragraph of Section 3.3 ("Forecasted streamflow series") was revised as follows:

"Across the validation and test periods, the monthly climatology model yielded the lowest overall RMSE (33.9) and the highest R^2 (0.73), thus outperforming the SARIMAX-based models in terms of average predictive accuracy. This result is expected because climatology effectively reproduces the dominant seasonal cycle of monthly streamflow, which explains a large proportion of the variance under typical conditions. However, when model performance was specifically evaluated under high-flow conditions, the LOG-NNSARIMAX-MVNL configuration achieved the lowest RMSE (60.8), indicating superior skill in reproducing peak flows. This improved performance likely arises from the model's ability to incorporate nonlinear precipitation patterns, macroclimatic variability, and temporal dependence, which provide additional predictive information during months that deviate markedly from the historical seasonal mean. In contrast, the climatological benchmark, by construction, cannot respond to anomalously wet conditions and therefore tends to smooth or underestimate peak-flow events."

Changes introduced in the Conclusions:

"This study conducted a comparative assessment of four monthly streamflow forecasting models for a data-scarce basin, all based on the SARIMA framework and progressively enhanced through the inclusion of exogenous predictors. While all configurations showed satisfactory performance, the LOG-NNSARIMAX-MVNL model provided the greatest added value under hydrologically critical conditions. Specifically, when performance was evaluated for high-flow events, this model achieved the lowest RMSE, demonstrating superior skill in reproducing peak flows. This advantage likely reflects its capacity to incorporate nonlinear precipitation patterns, macroclimatic variability, and temporal dependence, thereby capturing anomalous wet conditions that are not represented by the climatological benchmark. The robust performance of the hybrid model, including at lead times of up to 24 months, highlights its potential for anticipatory water-resources management and hydroclimatic risk mitigation in data-limited basins.
…"

Comment 3.

Editor: Include key datasets and outputs (e.g., observed discharge and main results) in the Supplementary Material.

Referee 1: Could you please share the raw data used and major outputs, including your observed discharge, in the supplementary document for transparency

The Supplementary Material has been expanded to include Table S2, which compiles the raw data used for streamflow, macroclimatic variables, and precipitation, together with the main outputs, including observed discharge

We thank the Reviewer for their careful evaluation of our manuscript and for the constructive comments provided throughout the review process. We have carefully revised the manuscript accordingly, clarified the remaining methodological aspects, strengthened the interpretation of the results, and expanded the Supplementary Material to improve transparency and reproducibility. We believe that these revisions have substantially improved the quality and clarity of the manuscript, and we respectfully submit the revised version for your consideration.