COMMENTS FOR THE AUTHOR:

Response to Reviewer 2

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: The description of the hybrid model is not sufficiently precise for reproducibility. It is unclear:

- How exactly the ANN interacts with SARIMA (residual modelling? parallel modelling? combined loss?).
- What is the ANN architecture (layers, activation functions, training epochs, optimizer).
- How exogenous variables are lagged to avoid information leakage.
- How the model moves from open-loop to closed-loop.

A fully explicit mathematical formulation and a schematic model diagram are required.

Answer: A detailed description of the neural network architecture and its associated parameters has been added to Section 2.4.1. This revision clarifies the structure of the neural network employed within the NLPCA framework, its role in processing CHIRPS precipitation data, and its integration into the SARIMAX modeling approach as an exogenous variable, thereby improving model transparency and reproducibility.

Comment 2: The autoencoder used for nonlinear PCA is described only conceptually. Please specify:

- Number of hidden layers and units
- Type of activation functions
- Loss function
- Training epochs
- Optimizer
- Reconstruction error achieved
- Rationale for selecting exactly two nonlinear components At the moment, NLPCA cannot be reproduced from the information provided

Answer: A detailed description of the autoencoder architecture used for nonlinear PCA and its associated training parameters has been added to Section 2.4.1. This revision provides the necessary information to fully specify the NLPCA implementation, thereby ensuring methodological transparency and reproducibility.

Comment 3: You report remarkably high R^2 values (0.75–0.78) for 24-month-ahead forecasts. This is unusual for tropical hydrology, where long-lead predictions are extremely difficult. Please:

- Include a baseline benchmark vs climatology and persistence
- Provide an alternative validation in which the training/validation split is shifted forward in time
- Discuss the potential risk of overfitting when many exogenous predictors are used for a single station
- Verify that no future information enters through the NLPCA step

Answer: We thank the reviewer for this important observation. Following the reviewer's suggestions, the analysis has been updated, resulting in more realistic R^2 values for long-lead forecasts and mitigating the risk of overfitting. Specifically:

- A baseline benchmark including climatology and persistence has been incorporated, enabling a clear comparison with the SARIMAX model.
- The potential risk of overfitting associated with the use of multiple exogenous predictors for a single station has been addressed and is discussed in the revised manuscript.
- The NLPCA transformation was recalculated using only the training period to ensure that no future information is inadvertently introduced.

Together, these revisions ensure that the reported long-lead forecast skill is both realistic and robust.

Comment 4: Table 2 shows extreme VIF values reaching 81,825, and two variables with infinite VIF. However, the process of variable selection is narrated but not clearly documented. I suggest:
• Provide a clear step-by-step table describing which variables were removed and why
• Avoid keeping multiple ENSO indices that are functionally redundant
• Confirm that the final chosen set has acceptable VIF values
This will make your variable selection reproducible.

Answer: The variable-selection process has been improved and clearly documented in Sections 2.4.3 and 3.2.3, where we now provide a step-by-step description of how the meteorological variables (MVs) were screened and selected. This revision clarifies how functionally redundant MVs were avoided and how the final set was retained based on the strongest correlation with the streamflow data. In addition, VIF values were recalculated for the retained MVs (see Table 2), confirming that all final predictors have acceptable multicollinearity levels (VIF < 10). As a result, the final set of exogenous MVs is shorter and more reproducible.

Comment 5: Pettitt p-value is reported as 1.99, which is impossible (p-values must be ≤ 1). Please revise all statistical outputs and their interpretations.

Answer: We thank the reviewer for this observation. The reported Pettitt p-value greater than 1 was due to a typographical error, which has been corrected in the revised manuscript. In addition, all statistical outputs and their corresponding interpretations have been carefully reviewed to ensure consistency and correctness throughout the manuscript.

Comment 6: The Breusch–Pagan test indicates heteroscedasticity, and the residuals vs. fitted plot confirms this. Please discuss or explore:
• Log transformation of discharge
• Box–Cox transformation
• Weighted regression or heteroscedasticity-aware models
• Impact on long-term forecast uncertainty

Answer: The log transformation of discharge was implemented and used as input data for all models. Model performance using log-transformed streamflow was superior compared to the non-transformed case. Accordingly, Section 2.6 has been added to the manuscript to document this procedure and its impact on model performance.

Comment 7: Forecasts are presented only as point estimates. Please include:
• Confidence intervals
• Prediction intervals
• Bootstrapped ensembles
• At minimum, a written discussion about uncertainty.
This is essential in hydrological forecasting.

Answer: Predictions and their associated confidence intervals were obtained, and Sections 2.7.1 and 3.3.1 were expanded to describe the corresponding methodology and results.

Comment 8: NLPC1 and NLPC2 remain abstract. Please include:
• Spatial loading maps • Seasonal cycle of each component
• Their relationship to ENSO/ITCZ migration
• Interpretation of their hydrological meaning This would strengthen the scientific contribution.

Answer: We thank the reviewer for this thoughtful and constructive comment. We agree that analyses of spatial loading patterns, seasonal cycles, and physical interpretations of the nonlinear principal components (NLPCs), including their relationships with ENSO or ITCZ migration, could provide valuable climatological and hydrological insights. However, the primary objective of this study is not the physical interpretation of the extracted components, but rather their use as predictive features within a streamflow forecasting framework.

In this context, NLPC1 and NLPC2 are employed as latent predictors designed to efficiently summarize large-scale climate variability and enhance forecasting skill, rather than to represent physically interpretable modes of variability. A detailed analysis of spatial loadings, seasonal behavior, and teleconnection mechanisms would require additional datasets, methodological developments, and extended discussion beyond the scope of the present work, and would substantially shift the focus of the manuscript from forecasting to process-based climate–hydrology analysis.

For clarity, the manuscript has been revised to explicitly state that the NLPCs are used as predictive variables without attempting a detailed physical interpretation. We consider a comprehensive diagnostic analysis of their climatological and hydrological meaning to be a valuable direction for future research.

Comment 9: Figure 1
- Caption should be more descriptive.
- Ensure all acronyms in the image are defined.

Answer: The caption of Figure 1 has been revised to be more descriptive, and all acronyms used in the figure have been defined.

Comment 10: Figure 2
- In IS the units for kilometers are km not Kms

Answer: The units in Figure 2 have been corrected to use km in accordance with the International System of Units (SI).

Comment 11: Table 1
- As said, Pettitt p-value impossible (1.99).
- Units missing for mean, min, max, and standard deviation (m^3/s).
- The title should specify “Streamflow characteristics at El Playón station”.

Answer: Pettitt p-value was reviewed. Regarding the descriptive statistics, the units (m^3/s) for mean, minimum, maximum, and standard deviation have now been explicitly indicated to avoid ambiguity. Finally, the title has been revised to explicitly specify “Streamflow characteristics at El Playón station”, as suggested by the reviewer.

Comment 12: Table 2 (VIF)
- Should include a final column indicating whether each variable was retained or dropped.

Answer: Table 2 has been modified to include an explicit indication of whether each variable was retained or removed.

Comment 13: Table 3
- Units missing for RMSE.
- AIC/BIC interpretations should be provided in a footnote.

Answer: The units for RMSE have been added, and interpretations of AIC and BIC have been included as a footnote in Table 3.

Comment 14: Table 4

- Clarify if these R^2 values are from training, validation, or test datasets.

Answer: The R^2 values reported in Table 4 correspond to the validation phase.

Comment 15: Stationarity section needs rewriting
Some expressions are grammatically incorrect:
"supporting weak stationarity assumptions component lacking any systematic trend"
Please rewrite this entire subsection more clearly.

Answer: Section 3.1.2 (Stationarity analysis) has been rewritten to improve clarity and grammatical accuracy, and the interpretation of the seasonal decomposition has been enhanced.

Comment 15: Improve logic flow between sections
Sometimes:

- The same idea is repeated (e.g., ENSO relevance).
- Paragraphs begin with generic phrases ("These results confirm…").

A more synthetic writing style would improve clarity.

Answer: The entire manuscript has been reviewed and rewritten to improve logical flow, reduce redundancy, and enhance overall clarity.

Comment 16: Please check all units
I found several missing or unclear units:

Answer: The entire manuscript has been reviewed to ensure that all units are clearly and consistently reported.

Comment 17: I strongly recommend adding:

- A short description of software used (Python, R, MATLAB).
- Version numbers for key libraries (TensorFlow, PyTorch, statsmodels, etc.).
- Pseudocode or model-training flowchart.

Answer: We thank the reviewer for this valuable suggestion. In response, a short description of the software environment has been added to the manuscript, indicating that the analyses were performed using Python and specifying the main libraries employed. The corresponding package information has been incorporated accordingly in the section 2.3.

Regarding the request for a model-training flowchart or pseudocode, we clarify that Figure 1 serves this purpose by visually summarizing the complete modeling workflow, from data preprocessing to model training, validation, and evaluation. To further address the reviewer's comment, Figure 1 has been revised to enhance clarity and better represent the sequence of steps involved in the modeling process. We believe these revisions improve the transparency and reproducibility of the study and adequately address the reviewer's recommendation.

Comment 18: My recommendation is Major Revision.
I hope my comments help you strengthen your manuscript. I would be very happy to re-evaluate a revised version.

We thank the reviewer for the thorough evaluation, constructive comments, and helpful recommendations. We have carefully addressed all observations and hope that the revisions adequately strengthen the manuscript. We would be pleased to have the opportunity for the revised version to be re-evaluated.