

Reviewer 2

This manuscript presents a dual-phase ensemble framework, which includes HEC-HMS, event selection from extreme scenarios, calibration, and a QRF post-processing stage, applied to reservoir inflow forecasting in the Nueces River Basin.

Overall comments:

The workflow is clear, but the methodology and scope are questionable. First, these steps, from the reviewer's perspective, are simply stacked together without a clear core logic. Hydrological forecasting is typically structured around uncertainty propagation or probabilistic prediction. In this paper, uncertainty is mentioned, but it is not formally structured.

(Take these papers as examples: <https://doi.org/10.5194/hess-26-2939-2022> , <https://doi.org/10.1007/s11269-025-04138-1>)

Response: We sincerely appreciate this important comment. Uncertainty in streamflow forecasting originates from multiple sources. While the meteorological input uncertainty is usually assumed to represent the largest sources of uncertainty in hydrologic forecasting, the uncertainties in initial conditions, structure, and parameters of the hydrologic models can also be significant (Cloke and Pappenberger, 2009; Zhang et al., 2025). An improved understanding of these uncertainties will contribute to the advances in multi-model and super-ensemble methods for reservoir inflow forecasts (Dion et al., 2021; Thebault et al., 2025).

The proposed method combines high-resolution parameter optimization, hydrological model setup to generate medium-range inflow forecasts for reservoir operations. The ensemble members are systematically generated based on the possible inflow patterns with consideration of historical inflow trend and analog years. Further, post-processing techniques are subsequently applied to further improve forecast accuracy and operational usefulness. The overall framework enhances forecast performance by reducing simulation and forecasting errors and improving agreement between forecasted and observed streamflow.

In the revision, we will enhance strengthen the development of ensemble members with the consideration of analog year forecasts and incorporate a weighted ensemble approach to better account for antecedent soil moisture conditions and forecasted precipitation patterns. Combined with analog-year simulations, this weighting scheme can improve the representation of dominant hydroclimatic conditions, reduce forecast uncertainty, and enhance the reliability of ensemble predictions. This enhancement is expected to reduce forecast bias and error dispersion while addressing the uncertainty and reliability of the ensemble forecasts.

References:

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of hydrology*, 375(3-4), 613-626.

Dion, P., Martel, J. L., & Arsenault, R. (2021). Hydrological ensemble forecasting using a multi-model framework. *Journal of Hydrology*, 600, 126537.

Thébault, C., Perrin, C., Legrand, S., Andréassian, V., Thirel, G., & Delaigue, O. (2025). What can be expected from a semi-distributed multi-model approach for streamflow forecasting?

Tailoring the structure and size of a super-ensemble on the Rhône basin. *Journal of Hydrology*, 661, 133589.

Zhang, J., Li, W., & Duan, Q. (2025). Quantifying the contributions of hydrological pre-processor, post-processor, and data assimilator to ensemble streamflow prediction skill. *Journal of Hydrology*, 651, 132611.

The study case is expected to be presented either at a disaggregated or aggregated level/region. However, the scope of the paper is not clear. From the title, the study focuses on reservoir inflow, but from Figures 1 and 2 and the results, multiple reservoir inflows are shown. Which one is the target?

Response: The targeted system is a cascade reservoir system that includes two reservoirs and four rivers as the sources of reservoir inflows. The flow observations at multiple gages are selected to fully cover all of the rivers for detailed setup and calibration of the hydrologic model at the basin scale.

The paper attempts to achieve multiple goals simultaneously: improving accuracy, representing uncertainty, and extending forecast lead time. This results in a lack of clear prioritization and no unified objective function. Even in the results (Figures 7a and 10), the estimated values are far from the observed inflows.

Response: The objective in the study is to maximize the operational reliability of the forecasting framework by addressing sources of uncertainty in the flow forecasts. By integrating different aspects of model calibration, ensemble modeling, and post-processing techniques, the study aims to establish a more robust and dependable forecasting framework for operational decision-making. The methodology and results will be enhanced to address the various sources of errors and provide high skill forecasts. The objective will also be focused on the reduction of bias in the forecast model and will be presented with more validation events in recent periods. The results will be compared with observation data to demonstrate the higher skill forecast results.

The regression-based forecast extension is overly simplistic. Ensemble forecasts are already affected by bias and dispersion errors, requiring more advanced correction methods. Simple statistical methods are often insufficient compared to modern approaches. The use of linear regression for extending forecasts is not consistent with state-of-the-art hydrological forecasting practice.

Response: Agreed. We will remove the part of linear regression in the revised manuscript.

The evaluation does not match the claimed goal. The paper claims probabilistic forecasting and uncertainty improvement, but the evaluation uses NSE, PBIAS, and CPI. Please provide references explaining why these metrics support the findings. Probabilistic forecasts should be evaluated using scoring rules such as CRPS if uncertainty representation is a key contribution.

Response: Accepted. We agree that probabilistic forecasts should be assessed using proper scoring rules that explicitly account for forecast uncertainty and ensemble spread. CRPS is particularly well suited for evaluating both the accuracy and calibration of ensemble predictions within a probabilistic forecasting framework. We will incorporate CRPS evaluation metrics in the revised manuscript to provide a more comprehensive assessment of ensemble forecast performance.

Line-by-line comments

1. Abstract: As mentioned above, probabilistic forecasts should be evaluated using proper scoring rules such as CRPS rather than deterministic metrics. (<https://doi.org/10.5194/hess-14-2545-2010>)

Response: We will update the abstract accordingly. Thank you.

2. Page 2, line 60: References are missing for “Dual-phase ensemble analysis...”.

Response: Noted, we will correct it.

3. Methodology: The combination of physical ensemble modeling and statistical post-processing is not novel (<https://doi.org/10.1080/27678490.2021.1936825>). The proposed framework should be positioned relative to SOTA (<https://doi.org/10.1080/27678490.2021.1936825>).

Response: We will enhance the ensemble framework through several important improvements, including the ensemble members selection methodology in consideration with analog year forecasts and provide more post-processing components by introducing a weighted ensemble approach that accounts for antecedent land-surface conditions and rainfall intensity. The revised integrated framework developed will combine multiple approaches to improve forecast reliability and operational applicability. The combination of these methodologies within a unified forecasting framework represents a novel contribution toward addressing forecast skill limitations commonly encountered in catchment and hydrological modeling studies. The study advances by integrating event selection methods systematically combined with analog year for reliable streamflow forecast modeling and post-processing. A performance-based weighting framework will be used for the evaluation of ensemble members using Continuous Ranked Probability Score (CRPS) to provide a more robust and operationally relevant forecasting approach for water resources management applications.

4. Pages 8–10: The ensemble consists of 12 members generated through parameter perturbation and initial condition adjustments. However, there are no detailed explanations or evaluations provided.

Response: We will provide a more detailed description of the ensemble members and explanation of the selection procedures in the revised manuscript.

5. Equation 7: Why are weights of 3, 2, and 1 chosen for A, B, and C? This lacks justification.

Response: We will provide a more detailed justification and discussion in the revised manuscript. Overall, different evaluation metrics place varying levels of emphasis on specific aspects of the forecast results. Assigning different weights to evaluation metrics is a justified practice when the importance of each metric varies and is widely supported in hydrological modeling (Bradley et al., 2015, Moriasi et al. 2015). In the revised manuscript we will propose a better way to determine different weights for the evaluation metrics.

References:

Bradley, A. A., Habib, M., & Schwartz, S. S. (2015). Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resources Research*, 51(9), 7382-7400.

Moriiasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.

6. Line 262: Why are the 5th, 50th, and 90th percentiles selected?

Response: The 5th and 90th percentile values are used to represent the lower and upper uncertainty bounds of the forecast, thereby illustrating the range of possible forecast outcomes. These percentile bands help capture the variability and uncertainty associated with the inflow forecasts. For example, Shrestha and Solomatine (2008) used these percentiles to define prediction intervals for seasonal river discharge. In contrast, the 50th percentile is intended to represent the central estimate of the forecasted flow, providing the most likely forecast condition. A central prediction interval, such as the interquartile range (25th–75th percentile) or a wider confidence band, provides a concise representation of the spread of likely outcomes and effectively communicates the uncertainty associated with simulated streamflow predictions (Seo et al., 2006; Verkade et al., 2017). Given the medium-range forecasting horizon of up to seven days, the study presents a broader ensemble spread around the mean forecast to better characterize forecast uncertainty and support flexible decision-making in reservoir operations.

References:

Shrestha, D. L., & Solomatine, D. P. (2008). Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. *International Journal of River Basin Management*, 6(2), 109-122.

Seo, D. J., Herr, H. D., & Schaake, J. C. (2006). A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions*, 3(4), 1987-2035.

Verkade, J. S., Brown, J. D., Davids, F., Reggiani, P., & Weerts, A. H. (2017). Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to Meuse and Rhine. *Journal of Hydrology*, 555, 257-277.

7. Pages 10–11: No data are presented. QRF is insufficiently justified, what's the input, how parameters are turned, and no comparison is provided. In addition, post-processing methods

differ significantly in their ability to improve reliability and sharpness and should be benchmarked (<https://doi.org/10.1002/hyp.9562>).

Response: The Quantile Regression Forest (QRF) is applied to provide predictive distribution and generate a range of forecast outcomes that explicitly accounts for predictive uncertainty. Unlike deterministic methods, QRF produces probabilistic forecasts by estimating conditional quantiles, thereby capturing the variability in forecasted streamflow. As a post processing tool, the QRF captures the growing uncertainty within the forecast. The QRF implementation follows the existing procedure in the referenced Python code, <https://doi.org/10.21105/joss.05976>.

In the revised manuscript, we will provide a more detailed description of the QRF methodology to improve clarity and reproducibility. We will also enhance the post-processing framework to better address uncertainties and reduce potential systematic errors in the ensemble member forecasts. This approach will be integrated with the weighted mean of the individual ensemble forecasts, following an evaluation of predictive performance using CRPS. The proposed methodology is expected to significantly enhance the post-processing framework, resulting in more accurate and reliable forecasts.

8. Page 11: The use of linear regression is overly simplistic, as mentioned above. Hydrological processes are nonlinear and forecast uncertainty increases with lead time. In your model, how do you ensure that errors do not propagate?

Response: Accepted. We will remove the linear regression section in the revised manuscript.

9. Pages 14–19: Calibration appears to be conducted using only one event. Some values in Table 1 show large errors for certain events, such as 0.54 for the CPI weight. How do you ensure this does not affect forecasting? Is validation conducted using a single event? Ensemble forecasting systems require multi-event validation; otherwise, the validation is insufficient. (<https://doi.org/10.1016/j.jhydrol.2021.126537>)

Response: We agree that ensemble forecasting typically requires multiple validation events to ensure robustness and reliability. In this study, we aim to evaluate the use of real forecast data from NOAA, and we will incorporate additional events since the submission of the manuscript. This will help strengthen the validation framework and improve the credibility of the ensemble forecasting results. Three events that have significant flows in the inflow stations from April to June 2026 will be included in addition to the existing events in the manuscript. Currently the model calibration are conducted with the events of 1997 (June 20 – July 6), 2002 (June 29 – July 29), 2015 (May 11 – June 7), and 2018 (October 7 – October 30). Similarly, for the Lake Corpus Christi inflow model, the selected events are 2002 (June 29 – July 29), 2015 (May 11 – June 17), 2016 (September 24 – October 30), and 2018 (October 7 – November 25). And the validation was conducted for the event between July 18-25, 2025. We will also introduce a weighted ensemble member evaluation within the post-processing stage to improve forecast performance. These additions will be included in the revised manuscript. Thank you for all valuable suggestions.