Dear Anonymous Referee,

We sincerely thank the reviewers for their constructive and insightful comments. Your feedback has substantially improved the quality, clarity, and organization of our work.

Comments from Reviewer #1:

This study proposes a hybrid model (LCEEMDAN-ASGGRU) integrating logarithmic transformation, CEEMDAN decomposition, and adaptive graph neural network for streamflow forecasting. The research topic holds practical application value, with systematic empirical studies conducted in the Poyang Lake Basin. However, the paper has several areas requiring improvement in terms of innovation, experimental design, and technical depth. Major revisions are recommended before considering publication.

Response:

We sincerely appreciate the reviewer's thoughtful and constructive comments. Below, we have provided responses to each comment and described how the suggested changes have been incorporated into the revised manuscript.

My comments are provided as follows:

Major Comments

1. The combination of CEEMDAN with deep learning has been explored in recent literature (e.g., Xu et al., 2024). Furthermore, logarithmic transformation represents a standard preprocessing technique that cannot be considered a primary contribution. The authors should clearly articulate what distinguishes their approach from existing methods and explicitly states the novel aspects of the proposed framework.

Response:

Thanks for your comment. We agree that both CEEMDAN and the logarithmic transform are established techniques, and that CEEMDAN-deep learning hybrids (Xu et al., 2024) have increasingly been applied in streamflow prediction. In the revised manuscript, we have tempered the novelty claims in the Introduction and contribution statement, and we now explicitly state the logarithmic transformation as a standard stabilizing preprocessing step rather than a primary contribution.

Our intention is not to present CEEMDAN or log-based preprocessing as a new technique, but to investigate how multiscale decomposition and graph neural networks interact in a multi-station setting, where both spatial connectivity and non-stationary dynamics are important for streamflow forecasting. To the best of our knowledge, most existing CEEMDAN-based hydrological studies couple CEEMDAN with LSTM/GRU/CNN at the level and do not explicitly encode spatial dependence among stations (e.g. Ghanbari-Adivi and Ehteram, 2025; Li et al., 2023), while GNN-based hydrological studies typically rely on graphs when representing spatial relationships across multiple stations.

In this respect, our contribution lies at the level of an integrated framework and its systematic evaluation. Specifically, the proposed model couples CEEMDAN's multi-scale temporal decomposition with an adaptive graph recurrent architecture, enabling the extraction of the intrinsic temporal characteristics of streamflow at different frequencies and their spatial propagation patterns across the watershed network.

In addition to proposing an integrated forecasting pipeline, the manuscript also aims to provide interpretive insights into spatial dependencies and predictive uncertainty. First, we analyze the adaptive adjacency matrix learned by ASGGRU on the original series (Section 4.4, Fig. 7), and show

that many strong directed edges coincide with known upstream-downstream relationships, while others reflect cross-basin connections plausibly driven by shared meteorological forcing. This indicates that learned graph captures hydrologically meaningful connectivity without explicitly encoding river topology. Second, each deterministic model is coupled with an HMM-GMR post-processor to quantify forecast uncertainty and evaluate coverage, sharpness and CRPS across models (Section 4.6). This provides a systematic assessment of uncertainty for the LCEEMDAN-ASGGRU hybrid framework and helps interpret performance gains.

Overall, we have revised the manuscript to temper the novelty claims around CEEMDAN and log-transform preprocessing and to more clearly articulate the contribution of combining multiscale decomposition, adaptive graph learning, and uncertainty quantification for multi-station streamflow forecasting.

[1] Ghanbari-Adivi, E., Ehteram, M., 2025. CEEMDAN-BILSTM-ANN and SVM models: two robust predictive models for predicting river flow. Water Resour. Manage. 39, 3235–3271. https://doi.org/10.1007/s11269-025-04105-w

[2] Li, H., Zhang, X., Sun, S., Wen, Y., Yin, Q., 2023. Daily flow prediction of the huayuankou hydrometeorological station based on the coupled CEEMDAN–SE–BiLSTM model. Sci. Rep. 13, 18915(2023). https://doi.org/10.1038/s41598-023-46264-z.

2. (Line 244) “Each variable (streamflow, precipitation, and temperature) at each station was decomposed into eight IMFs and one residual component.” The manuscript states that each variable is decomposed into eight IMFs plus one residual component without providing theoretical or empirical justification. This parameter selection requires thorough discussion including sensitivity analysis of IMF numbers, and comparison with alternative decomposition levels.

Response:

Thank you for the suggestion. In our implementation using the PyEMD library's CEEMDAN with its default parameters, the algorithm adaptively determined the number of IMFs, resulting in varying counts across different variables and stations. For our data, precipitation series decompose into 10 or 11 IMFs, streamflow into 9 or 10, and temperature into 8 or 9. Using these counts directly would make the input channel sizes inconsistent across variables/stations.

To align channels, we merge all components slower than IMF8 (i.e., IMFs with indices > 8) into a low-frequency aggregated residual. Thus, “8 IMFs + 1 residual” in the manuscript actually refers to “the first 8 CEEMDAN modes kept as separate channels, with all remaining modes aggregated into the residual”. This preserves exact reconstruction while ensuring a consistent number of channels across variables and stations. We have clarified this description in the revised Section 3.1.

We acknowledge your point that this design choice warrants empirical justification. To this end, we conducted a sensitivity analysis in which the proposed LCEEMDAN-ASGGRU model was re-trained using K = 5, 6, and 7 IMFs (plus the aggregated residual), while keeping all other settings unchanged. The results are summarized in Table A1. Compared with the K = 8 setting used in the main manuscript, using K = 6 or K = 7 leads to lower predictive skill. The K = 5 and K = 8 configurations achieve similarly high mean NSE (0.889 ± 0.040 and 0.888 ± 0.012), but K = 8 yields the smallest RMSE (264 ± 23 m^3/s) and the lowest run-to-run variability for both NSE and RMSE. This indicates that retaining eight IMFs offers a more accurate and robust representation, while still allowing higher-order modes to be safely aggregated into the residual. Consequently, we adopt K = 8 as the default decomposition level in the main experiments.

Table A1: Comparison of LCEEMDAN-ASGGRU performance under different numbers of IMFs.

The number of IMFs	5	6	7	8
NSE ± std	0.889 ± 0.040	0.817 ± 0.042	0.857 ± 0.063	0.888 ± 0.012
RMSE ± std (m^3/s)	275 ± 73	390 ± 77	299 ± 70	264 ± 23

*We have added a brief explanation of the decomposition-level choice in Section 3.1. In addition, we have created **Appendix A** in the revised manuscript, which provides the complete sensitivity analysis and the corresponding results in **Table A1**.*

3. In terms of the model design, LCEEMDAN-ASGGRU feeds nine decomposed components into ASGGRU individually to construct nine sub-models, with the final prediction result derived as the average of these sub-models. This design appears to overlook the interactivity among the decomposed components, rendering the approach more akin to an ensemble model rather than a truly integrated hybrid framework. It is therefore requested that the authors explain the rationale behind treating the nine sub-models as independent entities instead of exploring the interactive fusion of their training features. Why was the potential for synergistic information exchange between decomposed components not considered in the model architecture?

Response:

*Thank you for this thoughtful comment. We first clarify that, in our implementation, the final prediction is obtained by **summing**, not averaging, the outputs of the component-wise submodels. This summation is the inverse operation of the CEEMDAN decomposition and is the standard way to reconstruct the original signal from its components. We have revised Section 3.4 to make this reconstruction step explicit in both text and figure.*

*In LCEEMDAN-ASGGRU, nine scale-specific branches were constructed, each branch receives the k-th component (streamflow, precipitation and temperature at that scale, across all stations) as input, learns a scale-specific adaptive graph and temporal dynamics via ASGGRU, and outputs the predicted k-th component of streamflow. The final forecast is obtained by summing the predicted components. In this sense, the "nine sub-models" are **scale-conditioned modules** that mirror the CEEMDAN representation and allow the ASGGRU to specialize in different frequency bands.*

Regarding the reviewer's suggestion on "exploring the interactive fusion of their training features", we have implemented and tested a cross-scale fusion design. Specifically, all CEEMDAN components (IMFs and residual) were concatenated along the feature dimension and provided them as joint input to ASGGRU, so that each prediction step had access to the full multiscale representation. All other settings were kept unchanged.

The results of this experiment are summarized in Table R1 of this response. The cross-scale fusion variant leads to a substantial degradation in performance: the mean NSE over the 14 stations drops to about 0.63, which is markedly lower than the NSE of the original LCEEMDAN-ASGGRU, and the corresponding RMSE is consistently higher. We have also observed that this fused-input design requires a much larger input dimension and parameter count, increasing training time without delivering accuracy gains. These findings suggest that, for the present dataset and model capacity, explicitly mixing all scales at the input level tends to blur the scale-specific structure extracted by CEEMDAN and does not improve predictive skill.

Table R1: Comparison of predictive metrics between LCEEMDAN-ASGGRU and Cross-scale fusion

variant.

	LCEEMDAN-ASGGRU	Cross-scale fusion variant
NSE ± std	0.888 ± 0.012	0.686 ± 0.070
RMSE ± std (m^3/s)	264 ± 23	616 ± 95

Minor Issues

1. (Line 22) The phrase "mean root mean squared error of 264" lacks units, which is essential for RMSE (a dimensional metric). Please thoroughly review the entire manuscript and correct the notation for all dimensional units to ensure clarity and rigor.

Response:

Agree. In the revised manuscript, we have added the physical unit "m^3/s" wherever the root mean square error (RMSE) is reported. We have also carefully checked the entire manuscript, including the main text, tables, and figure captions, to ensure that dimensional quantities are consistently reported with their units.

2. (Figure 1) Elevation values in the figure require associated units. Additionally, the manuscript mentions five major rivers (Ganjiang, Fuhe, Raohe, Xinjiang, and Xiushui) as key tributaries of the Poyang Lake Basin, but these rivers are not explicitly displayed in Figure 1. Given the statement that "Figure 1 shows the spatial distribution of selected stations, covering the five major river systems that drain into the Poyang Lake," the five corresponding sub-basins should be clearly labeled in the figure to align with the textual description.

Response:

Thank you for your suggestions. We have revised Figure 1 as follows: (i) the elevation colour scale now includes units (m); (ii) the five major tributaries of the Poyang Lake basin (the Ganjiang, Fuhe, Raohe, Xinjiang, and Xiushui) have been clearly labelled.

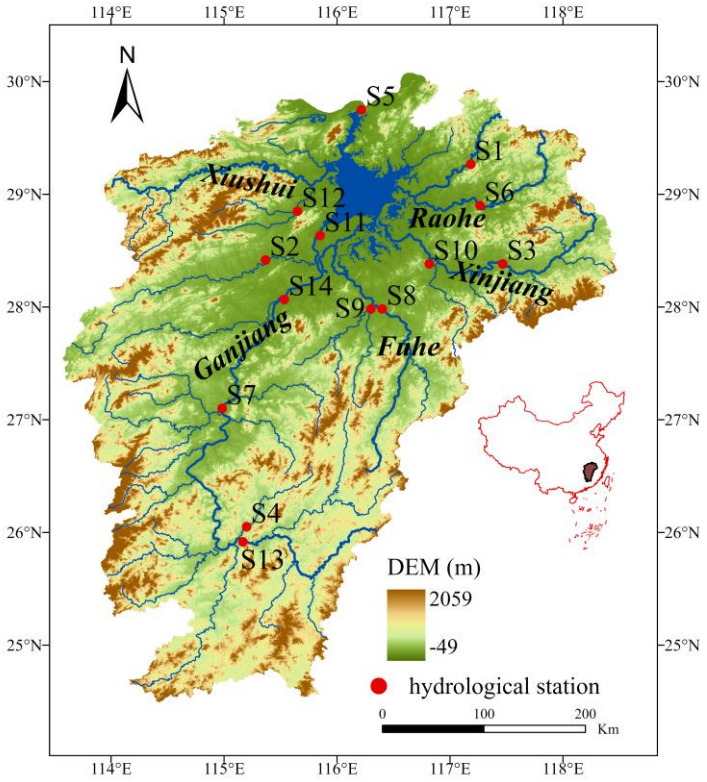

Figure 1: Location of hydrological stations.

3. (Figure 3) The model framework diagram necessitates optimization for clarity: (1) The meaning of "T1-T4383" following normalization is not clarified, please add explicit annotations to define these variables; (2) Relocating this framework diagram to Section 3.1 (Methods) would better facilitate readers' understanding of the model structure prior to discussing results; (3) The arrow placement between the 4th, 5th, and 6th sub-diagrams is ambiguous and fails to illustrate the data transfer and computational flow from the upper to lower layers of the model; (4) The boxes in the last two sub-diagrams are incomplete, and the label "LCEEMDAN-ASGGRU" is misleading, which should refer to the prediction output rather than the model name.

Response:

We thank the reviewer for these helpful suggestions regarding Figure 3. In the revised manuscript, Figure 3 has been redrawn to improve clarity and readability. The main changes are as follows:

(1) "T1–T4383" symbols represent the sequence of daily time steps in the full dataset. We have added an explicit explanation in the figure caption:

"The notation T1-T4383 represents the sequence of daily time steps in the full dataset (from day 1 to day 4383)."

(2) We updated the arrows to clearly illustrate the top-to-bottom flow of information. The revised figure now shows: (i) CEEMDAN outputs feeding into a "Normalization and sample construction" step; (ii) the construction of training and testing sequences (sliding windows T1-T12, …, Tm+13-Tm+24, etc.); and (iii) these sequence samples being used as inputs to the ASGGRU sub-models. This addresses the previous ambiguity in how information is transferred from the upper preprocessing steps to the lower model layers.

(3) The label in the final sub-diagram has been changed from "LCEEMDAN-ASGGRU" to "Prediction

output" to correctly refer to the model output rather than the model name.

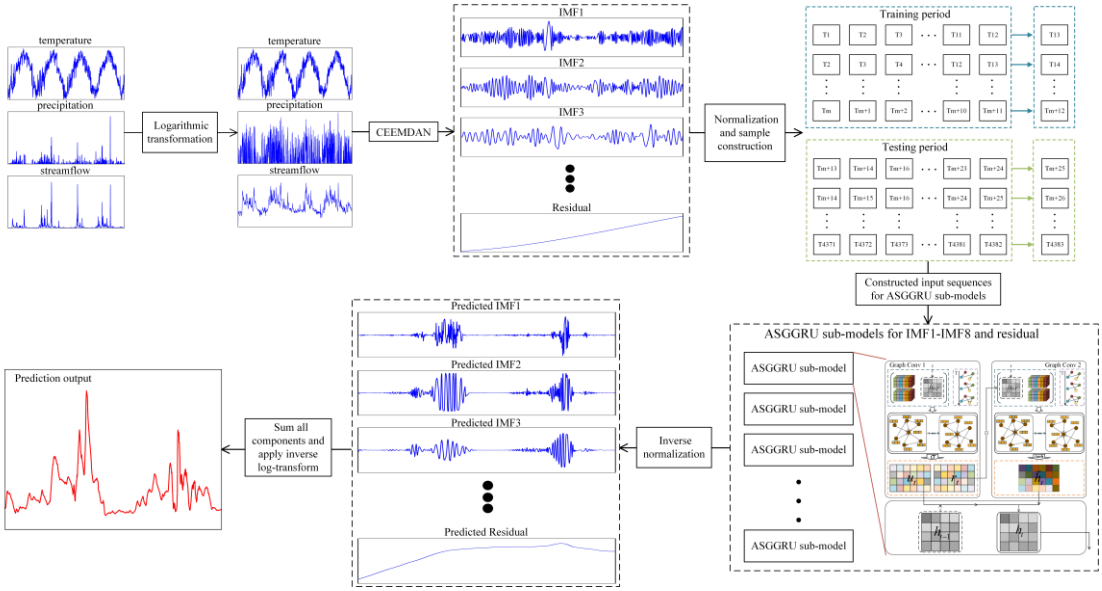

Figure 3: Framework of LCEEMDAN-ASGGRU. The notation T1-T4383 represents the sequence of daily time steps in the full dataset (from day 1 to day 4383).

We carefully considered the suggestion of moving this framework diagram to Section 3.1. Section 3.1 only introduces the CEEMDAN decomposition, whereas Section 3.4 describes the complete framework. For this reason, we keep Figure 3 in Section 3.4, where the full framework is first presented.

4. (Sections 4.1–4.2) The content in these sections focuses on experimental design and comparative model settings, which is misclassified under the Results section. It is recommended to create a dedicated chapter (e.g., "Experimental Setup" or "Comparative Models") following the Methods section to house this material. Furthermore, the abbreviations DTWSGGRU and FDSGGRU are used without prior definition.

Response:

Thank you for this helpful suggestion. We agree that the description of the experimental design and the comparative model settings fits more naturally within the methodological content rather than in the Results section.

In the revised manuscript, we have added a new subsection Section 3.7, titled "Experimental setup and comparative models". This subsection gathers the content from sections 4.1 and 4.2 of the original submission versions.

Within Section 3.7, all model abbreviations are now introduced and clearly defined. We have also checked the rest of the manuscript to ensure that all abbreviations are defined at their first occurrence.

5. (Section 4.5) The authors should explicitly justify the selection of stations S4 and S5 for detailed analysis, i.e., clarify their representativeness. Additionally, Figure 9 suffers from insufficient resolution, making its core message unintelligible.

Response:

Thank you for the comment. In the original manuscript, L409-L414, we noted that S4 is located in the southernmost part of the basin and exhibits the lowest mean streamflow, while S5 is located in the

northernmost part and records the highest mean streamflow. Since all 14 stations are modelled by GNN jointly, model performance at these low-flow and high-flow extremes provides a stringent test of model robustness across the remaining stations.

In the revised manuscript, we have made the rationale for selecting S4 and S5 more explicit:

L409-L412: "To further visualize these improvements, Figure 6 presents the observed and predicted streamflow of six models at two 410 representative stations: S4, located in the southernmost part of the basin and exhibiting the lowest mean streamflow, and S5, located in the northernmost end and recording the highest mean streamflow. Together, they provide a stringent complement to the overall performance metrics summarized in Figure 4."

This revision clarifies why S4 and S5 are used for detailed visualization and highlights their representativeness with respect to both flow magnitude and network position.

*Regarding **Figure 9**, we acknowledge that the original submission may have hindered interpretability. In the revised version, Figure 9 has been regenerated at substantially higher resolution, with enhanced line weights, larger font sizes, and clearer shading for the prediction intervals.*

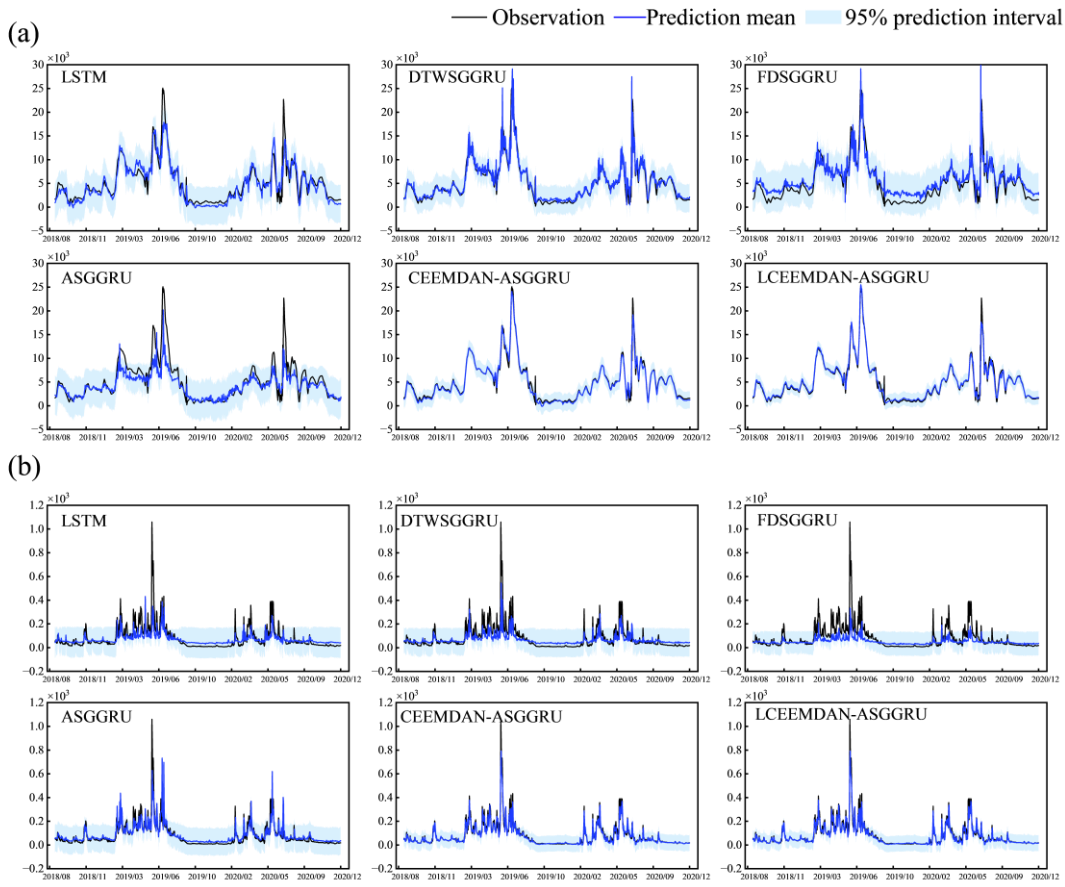

Figure 9: Time-series plots of predicted means and 95% prediction intervals for the six models at stations S5 (a) and S4 (b).

We have also revised the relevant descriptive text:

L481-L489:

"Figure 9 provides a time-series illustration of the probabilistic forecasts for stations S5 and S4, where the black line denotes the observed streamflow, the blue line represents the prediction mean (the

conditional expectation from the HMM-GMR predictive distribution), and the shaded area indicates the associated 95% prediction interval. This visualization allows a direct assessment of both point-forecast accuracy and the reliability of the uncertainty estimates.

At both stations, LCEEMDAN-ASGGRU exhibits the closest agreement between prediction means and observations while maintaining the narrowest and most responsive uncertainty bands among all six models. In the high-flow regime at S5 (Fig.9 (a)), the model tracks peak magnitudes without causing undue interval expansion, whereas in the low-flow regime at S4 (Fig.9 (b)), it captures subtle fluctuations with tight yet well-calibrated intervals, reflecting consistent coverage and stable predictive behaviour across contrasting hydrological conditions."

6. (Line 433) "In particular, connections such as S3→S10, S4→S14, S7→S11, and S13→S7 reflect physically consistent upstream–downstream dependencies that are embedded in the true hydrological topology." However, the figure also shows S10→S3. Therefore, it is necessary to provide a detailed explanation of the adjacency matrix, clarifying that it encodes not only hydrological upstream–downstream relationships but also spatial meteorological correlations (attributed to the inclusion of precipitation and temperature data) to resolve this apparent contradiction.

Response:

Thank you for the helpful comment. The adaptive adjacency matrix learned by ASGGRU is indeed constructed from ***multi-variable inputs****, including streamflow, precipitation and temperature at all stations. It is not restricted to a purely one-way river-routing structure, but can encode both hydrological routing and spatially coherent meteorological signals.*

In the revised manuscript, we clarify this interpretation in Section 4.4 and adopt more precise wording to describe how the adaptive adjacency matrix aligns with known hydrological connectivity.

We revised the description around L433:

"To investigate the spatial representations captured by the adaptive graph learning mechanism, we analyze the learned graph obtained from ASGGRU and evaluate its hydrological consistency and its relationship with static graph priors.

The learned adaptive adjacency matrix is visualized in Fig. 7. Several strong directional connections emerge from the learned structure, which are consistent with known flow-direction relationships within the river network. In particular, connections such as S3→S10, S4→S13, S7→S14 and S13→S7 consistent with physically plausible upstream-downstream dependencies inferred from the catchment topology. This indicates that the data-driven learning process is able to recover some key aspects of the underlying hydrological connectivity without any explicit spatial supervision. This result echoes findings by Bai and Tahmasebi (2023), who also observed that adaptive GNNs can learn spatial structures that align with physical understanding in groundwater forecasting and environmental modeling.

However, the adaptive adjacency matrix is learned jointly from streamflow, precipitation and temperature inputs. It also contains reciprocal and cross-basin links that go beyond a purely one-way river-routing graph, reflecting shared meteorological forcing or residual statistical dependence between stations. More generally, such indirect or cross-basin connections likely capture remote influences or common climatic drivers, complementing the river-network signal."

7. The terms "streamflow" and "runoff" are mixed throughout the manuscript.

Response:

Thank you, we standardized the terminology in the revision.

8. The standard deviations of the 10 repeated experiments should be reported.

Response:

Thank you for the suggestion. We have now reported the run-to-run standard deviations of the 10 repeated experiments for both NSE and RMSE. The results are summarized at the model level (mean ± standard deviation) and provided in ***Appendix A as Table A2****. This table quantifies variability across independent runs and demonstrates that the proposed LCEEMDAN-ASGGRU maintains low run-to-run variability while achieving the best overall accuracy among the six models.*

Table A2: Overall model performance (mean ± standard deviation) across 10 repeated runs.

Model	NSE ± std	RMSE ± std(m^3/s)
LSTM	0.710 ± 0.017	503 ± 14
DTWASGGRU	0.716 ± 0.017	496 ± 50
FDASGGRU	0.751 ± 0.061	479 ± 55
ASGGRU	0.779 ± 0.015	432 ± 28
CEEMDAN-ASGGRU	0.826 ± 0.036	327 ± 43
LCEEMDAN-ASGGRU	0.888 ± 0.012	264 ± 23