This manuscript proposes a hybrid VMD–Informer–LSTM framework for daily hydrological drought prediction in the Huaihe River Basin using the DEDI index derived from ERA5. The topic is relevant, and the combination of signal decomposition and deep learning architectures is timely. The authors present extensive experiments across different lead times and regions, and the results indicate consistent improvements over benchmark models. However, the manuscript needs further revision before it can be published.

The VMD step is critical to the model's performance, yet important details are missing: How was the number of modes (K) selected? Were K and the penalty parameter α fixed for all grid points, or tuned adaptively? Was sensitivity analysis performed to assess how VMD parameters affect prediction skill?

Reply: Thank you for your valuable comments. We agree that the number of modes (K) and the penalty parameter (α) have a significant impact on the quality of VMD decomposition and subsequent forecasting performance. The setting of these two parameters in this study was not arbitrary, but was based on existing research experience and combined with the characteristics of the data length.

In this study, we referred to previous VMD studies based on DEDI (e.g., Su et al., 2024). The penalty coefficient α (bandwidth constraint parameter) was empirically chosen based on the time series length, with a range set to 1.5–2.0 times the sample length. The final value of 1.75 times the sample length was selected as the consistent parameter for this study. This setting aims to balance frequency band separation ability and mode stability. The penalty coefficient α controls the bandwidth size of each mode: when α is small, the bandwidth of each IMF component is large, and an excessively large bandwidth may lead to spectral overlap between different modes, causing some IMF components to contain signals from other components, thus weakening the physical interpretability of the mode decomposition. On the other hand, when α is too large, the bandwidth is overly compressed, making the decomposition results more sensitive to noise. Additionally, the number of modes (K), which refers to the number of IMF components, was determined based on the frequency distribution characteristics of the decomposed signals. In this study, K was uniformly set to 7 to ensure that the main frequency information of the original DEDI sequence could be fully decomposed and retained, while avoiding redundancy from excessive decomposition. This parameter combination showed stable decomposition results and forecasting performance in preliminary experiments with multiple representative grid points.

To avoid introducing spatial overfitting risks from tuning parameters for individual grid points, while ensuring consistency and reproducibility of the method, the same set of VMD parameters (K and α) was uniformly applied across all grid points in the entire basin, rather than performing point-by-point adaptive tuning.

It should be noted that this study did not conduct a systematic sensitivity analysis of the VMD parameters. The main goal of this research is to validate the methodological feasibility of the VMD–Informer–LSTM framework for long-term drought index forecasting (methodological proof-of-concept), rather than to fine-tune the VMD parameters themselves or conduct a comparative study of parameter sensitivity. The parameter selection strategy has been further explained in the methods section of the revised manuscript to enhance transparency and reproducibility of the method. A systematic evaluation of the sensitivity of VMD parameters

and their impact on forecasting performance will be a valuable direction for future research.

Su, T., Liu, D., Cui, X., Dou, X., Lei, B., Cheng, X., Yuan, M., & Chen, R. (2024). Prediction of DEDI index for meteorological drought with the VMD-CBiLSTM hybrid model. JOURNAL OF HYDROLOGY, 641, 131805. https://doi.org/10.1016/j.jhydrol.2024.131805

The manuscript does not clearly describe: The train/validation/test split strategy (temporal split vs. random split). Whether hyperparameters were tuned using an independent validation set. And was the model tested on unseen data?

Reply: Thank you to the reviewer for pointing out the unclear explanation regarding the dataset division and model training process. We have supplemented and clarified the division strategy for the training, validation, and test sets, as well as the hyperparameter tuning process, in the revised methodology section to enhance the transparency and reproducibility of the study.

Specifically, for the DEDI time series of each grid point, we strictly performed forward splitting according to the time sequence, rather than random splitting, to avoid any future data leakage issues. We used 40 years of daily data, aiming to leverage the richness of long-term historical data to improve the model's stability and accuracy. Specifically, the training period is from 1984-1-1 to 2024-7-3, while the test period is from 2024-7-4 to 2024-12-31, mainly for evaluating the model's predictive ability on unseen data for the following 180 days. This division ensures that the model uses historical data during training and evaluates its ability to predict future data during testing.

During the model training process, we used Bayesian Optimization in the hyperparameter tuning to automatically search for key hyperparameters (such as network structure parameters, learning rate, batch size, etc.) within a predefined parameter space, with validation set error minimization as the optimization objective. The final optimal parameter combination was used to retrain the model on the training set and evaluate its performance on future data segments which are temporary independent of the training phase.

It should be noted that the training set, validation set, and test set in this study all come from the same ERA5-DEDI data system. Therefore, the evaluation in this study is focused on the out-of-time prediction capability within the consistent data system, rather than cross-data source validation.

Did the author use in-situ measurements? What does "Observed data" mean, is it in-situ measurements? If so, please clarify the data source.

Reply: Thank you for pointing out the ambiguity in this statement. No in-situ measurements were used in this study. The term "Observed data" in the manuscript actually refers to ERA5 reanalysis data (i.e., data used as a reference field/control sequence).

We agree that the term "Observed data" in the original manuscript may cause confusion. To avoid any misunderstandings, we have made unified modifications and clarifications in the revised manuscript, explicitly referring to it as "ERA5 reanalysis data." Additionally, in the Data and Methods section, we have clarified that the research is entirely based on the ERA5 reanalysis data system, with no independent site observational data introduced. The relevant

explanation has been added to the revised manuscript.

Discussion is a mixed of map results and its discussion. Please put the maps result to Section results, and expand the discussion content.

Reply: Thank you for the valuable comments from the reviewer. I acknowledge that in the original manuscript, the "Results" and "Discussion" sections were mixed together, which led to a lack of clarity in the structure. The reviewer's comment is completely correct, and we have made corresponding revisions to the paper. We have split the original "4 Results and Discussion" section into separate "4 Results" and "5 Discussion" sections. Specifically, we moved the map-based results, such as Figures 11 and 12, from the original manuscript to the "Results" section, while only Figure 13 remained in the "Discussion" section. Additionally, we have substantively expanded and rewritten the "Discussion" section. Instead of reiterating specific results, we now provide a comprehensive discussion of the research work, adding and systematically elaborating on the advantages, limitations, and shortcomings of the proposed method, as well as the model's applicability and potential value. These changes have been reflected in the corresponding chapters of the revised manuscript.

The specific additions are as follows:

The results show that the VMD–Informer–LSTM model exhibits high prediction accuracy within the 30–90 day forecast period at the time scale, while the prediction accuracy decreases during the longer forecast period of 120–180 days. Spatially, the prediction results for the upstream and downstream regions show the highest consistency with the observed values, followed by the Yishuisi River region, with the midstream region showing relatively weaker performance.

In summary, the VMD–Informer–LSTM framework proposed in this study demonstrates significant advantages in handling drought index series with prominent non-stationarity and multi-time scale features, using a multi-scale modeling strategy of "decomposition—parallel modeling—feature fusion." On the one hand, VMD effectively reduces the complexity of the original series, allowing for the separation of variation features at different time scales and modeling them individually. On the other hand, the parallel structure of Informer and LSTM focuses on capturing long-term background state changes and short-term fluctuations, enabling the model to represent both the persistence and phase-specific fluctuations of the drought process.

However, it should still be noted that there are certain limitations in this study: Moreover, the model evaluation in this study is based solely on "unseen data" in terms of time (i.e., future periods after the training samples), rather than independent generalization validation across data sources or independent observations. The current framework is entirely built on ERA5 reanalysis data and has not yet introduced independent ground-based observations or multi-source remote sensing data for external validation. Additionally, the model focuses on one-dimensional time series at each grid point and does not explicitly capture the spatial propagation of drought nor systematically assess prediction uncertainty. Therefore, the work in this study should be viewed as a methodological validation for long-term drought state forecasting rather than a tool that can directly replace operational forecasting systems. Nevertheless, the framework still holds potential application value in assessing drought background evolution

trends on seasonal to sub-seasonal scales and can provide references for mid- to long-term water resource regulation and risk assessment at the basin scale. Future research will further incorporate multi-source observational data and develop spatiotemporal coupling and uncertainty modeling methods to enhance the model's practical applicability and reliability.

Minor comments

Some figures (e.g. Figs. 7–9) are information-dense and difficult to read at journal scale. Consider simplifying or merging panels.

In line 170, please explain why 0.25°?

In line 317, please explain "an average" refer to average of what.

In Fig 7, please explain what are x-axis and y-axis values for.

Figure 11 should use same color with different gradients for three columns. It's more comparable.

Some figures (e.g. Figs. 7–9) are information-dense and difficult to read at journal scale. Consider simplifying or merging panels.

Reply: Thank you for your suggestions. We agree that some figures in the original manuscript (such as Figures 7–9) are quite information-dense and may affect readability at the journal's page size. Based on your feedback, we have made adjustments to the relevant figures. Specifically, the original Figure 7 has been moved to the supplementary materials (provided as an auxiliary figure) to reduce information overload in the main text and to make the figures in the main text more concise and focused on the core results. Through these adjustments, the overall readability of the figures in the main text has been improved. These changes have been reflected in the revised manuscript.

In line 170, please explain why 0.25°?

Reply: Thank you for your question. Regarding the choice of a 0.25° spatial resolution, this study primarily considers a balance between data availability, computational cost, and the needs of regional-scale modeling. ERA5 reanalysis data itself provide stable and consistent 0.25° resolution products, which have been widely adopted in numerous regional and global-scale hydrometeorological and drought studies as a mature and reliable working resolution.

For research at the scale of the Huaihe River Basin, a 0.25° resolution is well-suited to capture the spatial variation of regional droughts while ensuring that the sample size is sufficiently large, avoiding the significant computational overhead and instability issues that can arise with higher resolutions. Therefore, we chose the 0.25° resolution as a compromise between spatial detail representation and computational feasibility.

In line 317, please explain "an average" refer to average of what.?

Reply: Thank you for pointing out the lack of clarity in this statement. The term "an average" in the text refers to the spatial average of the model evaluation results for all grid points within the Huaihe River Basin. Specifically, in this study, the model was independently trained and run for each of the 108 grid points ($0.25° \times 0.25°$) within the basin. For each grid

point, prediction results and corresponding evaluation metrics were obtained. These evaluation results from the 108 grid points were then spatially averaged to derive the "average" performance metrics reported in the manuscript. This clarification has been added to the main text.

In Fig 7, please explain what are x-axis and y-axis values for.

Reply: Thank you for the reminder. The original Figure 7 (now moved to Supplementary Material Figure S1) shows a density scatter plot comparing the predicted values from different models with the ERA5-DEDI reference values for each sub-region of the Huaihe River Basin. The x-axis represents the DEDI reference values calculated from ERA5 reanalysis data, while the y-axis represents the corresponding model's predicted DEDI values. The gray diagonal line represents the ideal 1:1 line where the predicted values would exactly match the reference values. The color in the plot indicates the density distribution of the sample points.

Figure 11 should use same color with different gradients for three columns. It's more comparable.

Reply: Thank you for the reviewer's suggestion regarding the visualization of the figures. We understand that using a unified color scheme helps enhance intuitive comparisons between different columns. However, in Figure 11, the error metrics for different time phases (entire forecast period, first 90 days, last 90 days) have significant differences in their numerical ranges, especially in the last 90 days, where the error magnitude increases noticeably.

If the same color scale range and gradient were used for all three columns, the spatial differences in the first two phases would be compressed into a very narrow color range, thus weakening the discernibility of the spatial distribution details. Therefore, we adopted an adaptive color scale range for each time phase to ensure that the spatial difference structures within each column could be clearly presented, while also helping the reader better understand the relative changes in error levels across different time periods.

It should be noted that quantitative comparison between different phases is mainly done through statistical metrics (not just color visual intensity). We believe that the current visualization approach strikes a reasonable balance between "spatial distribution readability" and "inter-phase comparison explanation."