The authors present an innovative framework integrating Variational Mode Decomposition (VMD), Informer, and Long Short-Term Memory (LSTM) networks for long-term drought prediction in the Huaihe River Basin, China. The manuscript is well-structured, providing a clear explanation of the models and the indices used. The inclusion of SHAP (SHapley Additive exPlanations) to evaluate decision-making—identifying precipitation and temperature as dominant factors—adds significant value to the interpretability of the model.

However, several critical methodological concerns and limitations regarding data dependency and spatial representation must be addressed before the manuscript can be considered for publication.

Major Revisions

The model is trained and validated exclusively using ERA5 reanalysis data, including the DEDI index, which is itself formulated from ERA5 variables. This presents a significant limitation: the model may simply be learning the internal mathematical structure of the ERA5 atmospheric model rather than actual drought dynamics.

The authors must discuss how this dependency might amplify inherent biases within ERA5. For the results to be credible for public policy or operational use, the framework should be validated against observed in-situ station data to prove its real-world reliability.

Reply: Thank you for your valuable comments. We refer to several studies that use ERA5 data, which confirm the effectiveness and reliability of using ERA5 data in the Huaihe River Basin and similar regions, thus supporting our model application. For example, Gao et al. (2023) conducted a spatio-temporal analysis of cloud water resources in the Huaihe River Basin using ERA5 data, showing that ERA5 can effectively capture spatial variations in the region. Also, Zhang et al. (2021) constructed a Soil Moisture Deficit Index (SWDI) based on ERA5-Land data to assess agricultural drought in the Huaihe River Basin, validating the applicability of ERA5-Land data to drought analysis in this region. Furthermore, Li et al. (2025) studied hydrological drought modeling in the Huaihe River Basin using ERA5 data, demonstrating that ERA5 data performs excellently in drought forecasting for the basin. These studies collectively demonstrate that the use of ERA5 data in the Huaihe River Basin is feasible and effective.

Moreover, the application of ERA5 data to global-scale drought forecasting has also been validated. For instance, Xu et al. (2024) and Gupta et al. (2024) applied ERA5 data for drought prediction at the global and regional scales, respectively, proving that ERA5 data can effectively predict drought events even without ground-based observational data.

It should be noted that the statistical tests above do not eliminate the potential systemic biases in the ERA5 reanalysis data itself, but at least from the perspective of the time series structure, they demonstrate that the research object in this study is not a purely random process. In response to the reviewer’ s further concerns about data source dependence, we have added a discussion of the limitations of this issue to the revised manuscript and explicitly stated that future work will combine ground station observational data and multi-source remote sensing/reanalysis products for independent validation.

Xu, L., Zhang, X., Wu, T., Yu, H., Du, W., & Chen, N. (2024). *Global prediction of flash drought using machine learning*. *Geophysical Research Letters, 51*(21). https://doi.org/10.1029/2024GL111134.

Gupta, B. B., Gaurav, A., Attar, R. W., Arya, V., Bansal, S., Alhomoud, A., & Chui, K. T. (2024). Advance drought prediction through rainfall forecasting with hybrid deep learning model. SCIENTIFIC REPORTS, 14(1), 30459. https://doi.org/10.1038/s41598-024-80099-6

Zhang, R., Lu, L., Ye, Z., Huang, F., Li, J., Wei, L., Mao, T., Xiong, Z., & Wei, S. (2021). Assessment of Agricultural Drought Using Soil Water Deficit Index Based on ERA5-Land Soil Moisture Data in Four Southern Provinces of China. AGRICULTURE-BASEL, 11(5), 411. https://doi.org/10.3390/agriculture11050411

Li, M., Yao, Y., Feng, Z., and Ou, M.: Hydrological drought prediction and its influencing features analysis based on a machine learning model, Nat. Hazards Earth Syst. Sci., 25, 4299–4316, https://doi.org/10.5194/nhess-25-4299-2025, 2025.

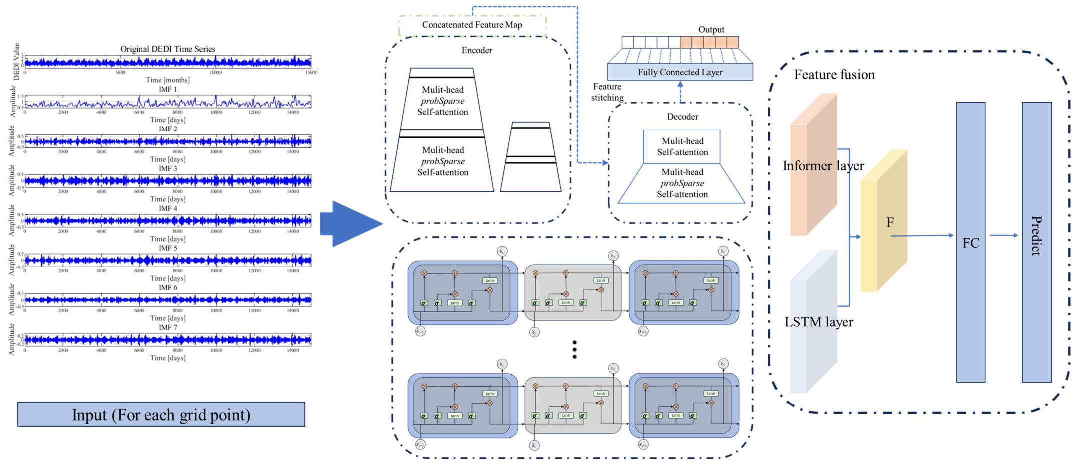
Gao, J., Feng, J., Cao, Y., & Zheng, X. (2023). Evaluation of Cloud Water Resources in the Huaihe River Basin Based on ERA5 Data. ATMOSPHERE, 14(8), 1253. https://doi.org/10.3390/atmos14081253

Methodological Clarity:

Point Selection: The process for the "systematic selection" of the 108 control points is unclear. The authors should explicitly define the criteria or algorithms used to ensure these points are representative of the basin’s hydro-climatic diversity.Spatial Connectivity: It is not specified whether these 108 grid points are treated as independent time series or if the model accounts for spatial interconnectivity. If the model treats them as isolated units, it ignores the spatial propagation of drought—a critical limitation that adds uncertainty to the findings.It is recommended that a workflow be incorporated in order to facilitate a better comprehension of the methodology.

Reply: Thank you for your valuable comments. The 108 control points used in this study were not manually or subjectively selected, but rather were the entire set of valid grid points that fall within the boundaries of the Huaihe River Basin after clipping the ERA5 DEDI data at the study resolution. Specifically, the procedure was as follows: first, the original grid data was clipped using the basin boundary mask, and then all grid points with centers located within the basin were retained. Therefore, these 108 points constitute a complete spatial sample of the study area at the current resolution, rather than sparse or selective sampling, and naturally cover the main climatic and hydrological gradients within the basin.

We thank the reviewer for pointing out this critical issue. In the current framework, each grid point is modeled as independent one-dimensional time series. The proposed VMD–Informer–LSTM focuses on capturing the long-term temporal dependencies of individual time series and does not explicitly incorporate spatial connectivity or spatial propagation processes. In other words, the modeling objective of this paper is a "point-based temporal prediction problem" rather than a complete spatio-temporal coupling model. To address this, we have added the following clarification in the methods section: "Each grid point is treated as an independent one-dimensional time series, with a focus on modeling its long-term temporal dependencies." We fully acknowledge that droughts exhibit significant spatial propagation and spatial correlation characteristics, and the current framework's failure to explicitly model this process is a limitation of this study. We have added the following to the discussion section of the revised manuscript: "This study has not explicitly modelled the spatial propagation process of droughts, which is one of the limitations of this research. Future work will introduce spatiotemporal coupling models for further improvement." We have also indicated that we will consider incorporating structures such as ConvLSTM, graph neural networks, or spatiotemporal transformers to model the spatiotemporal propagation process of droughts. The revised flowchart is shown below.



Model Robustness and Generalization: To demonstrate the model's true forecasting potential, the following points should be addressed:Evaluate the model's response to significant historical drought events that were excluded from the calibration period. This is essential to assess performance under extreme, "out-of-sample" conditions.Provide a technical justification for using the DEDI index derived from ERA5 data instead of more physically robust or internationally standardized methods, such as the FAO-56 Penman-Monteith equations for evapotranspiration.Justify why other high-resolution satellite or hybrid datasets (e.g., AgERA5, CHIRPS, IMERG, or MODIS) were not used to provide a more robust benchmarking of the results.

Reply: Thank you for your valuable comments. We sincerely appreciate this very helpful recommendation. We fully agree that independent testing of extreme drought events is crucial for assessing the model's generalization ability. However, it is important to note that the primary goal of this study is to validate the feasibility and effectiveness of the VMD–Informer–LSTM framework in long-term drought sequence forecasting (methodological proof-of-concept), rather than performing event-level reconstruction or scenario replay analysis for specific historical events. In the current data division framework, the model has been trained and tested using a strict time-series forward split method to avoid information leakage. Independent leave-out testing for specific extreme drought events (event-based cross-validation) would require a redesign of the sample construction strategy and experimental system, which would constitute an independent and systematic study. We clearly identified "independent generalization testing based on historical extreme drought events" as an important direction for future research in the discussion section of the revised manuscript.

The main consideration for choosing the DEDI index in this study is that it can be stable constructed based on ERA5 reanalysis data under long time series and complete spatial coverage conditions, and it has already been proven in previous studies (Zhang et al., 2022a; Zhang et al., 2022b) to effectively characterize regional drought evolution characteristics. In comparison, drought indices strictly based on the FAO-56 Penman–Monteith method for evapotranspiration or drought often have higher data dependency on ground meteorological variables and surface parameters (such as wind speed, radiation, crop parameters, etc.), leading to greater uncertainty in constructing long-term time series and ensuring consistency across the entire basin. The aim of this study is to validate the capacity of the proposed VMD–Informer–LSTM framework for medium- and long-term drought index forecasting. Therefore, DEDI is used as a representative variable for drought characterization in this study.

Thank you again for this very valuable suggestion. We fully agree that multi-data-source comparative validation is important for assessing the model's robustness. However, the focus of this study is mainly on proposing and validating an innovative long-term drought forecasting modeling framework (methodological framework), rather than conducting systematic comparative evaluation of multiple data products. It should be noted that systematically extending this framework to multi-source data such as CHIRPS, IMERG, MODIS, or AgERA5, and performing rigorous consistency comparison and cross-validation, would constitute a large-scale independent research effort. Due to the limitations of the study's scope and workload, this aspect has not been covered in this paper.

Zhang, X., Duan, Y., Duan, J., Chen, L., Jian, D., Lv, M., and Ma, Z. 2022a. A dail y drought index-based regional drought forecasting using the Global Forecast Syst em model outputs over China, Atmospheric Research, 273, 106166, https://doi.org/10.1016/j.atmosres.2022.106166

Zhang X, Duan Y, Duan J, Jian D, Ma Z. 2022b. A daily drought index based on evapotranspiration and its application in regional drought analyses. Science China Earth Sciences, 65(2): 317–336, https://doi.org/10.1007/s11430-021-9822-y

The current framework seems to overlook terrain characteristics (topography, land cover, soil type). Due to the resolution used, a single pixel may cover multiple climatic zones or land uses. The authors should discuss how this loss of sub-grid heterogeneity limits the model's accuracy in representing spatial drought reality.

Reply: Thank you for your valuable comments. We fully agree that, at the current spatial resolution, the individual grid cells of ERA5 reanalysis data essentially reflect the "area-averaged state" of the region, and the sub-grid scale heterogeneity, such as terrain variations, land use types, and soil differences, is inevitably smoothed. This issue is not only present in our study but is also a common challenge faced in all regional-scale studies based on reanalysis or medium-to-low resolution climate data, where scale mismatches occur. It should be noted that the main goal of this study is to validate the feasibility of the VMD–Informer–LSTM framework in addressing the problem of "long-term forecasting of regional drought index time series," rather than to model the fine-scale underlying processes within the basin. Therefore, the modeling object in this study is essentially the "composite drought state representation" at the grid scale, rather than the local differences at the sub-grid scale. We fully recognize that ignoring sub-grid scale heterogeneity may introduce uncertainties, particularly in areas with complex terrain or highly fragmented underlying surfaces. This is an important limitation of this study. We have clearly added this issue to the discussion section of the revised manuscript and pointed out that future work will consider integrating higher-resolution remote sensing data (such as land cover, soil moisture, DEM) or outputs from regional climate/hydrological models. These sub-surface characteristics will be incorporated as exogenous variables or introduced through spatiotemporal coupling models to further enhance the model's ability to represent real spatial drought processes.

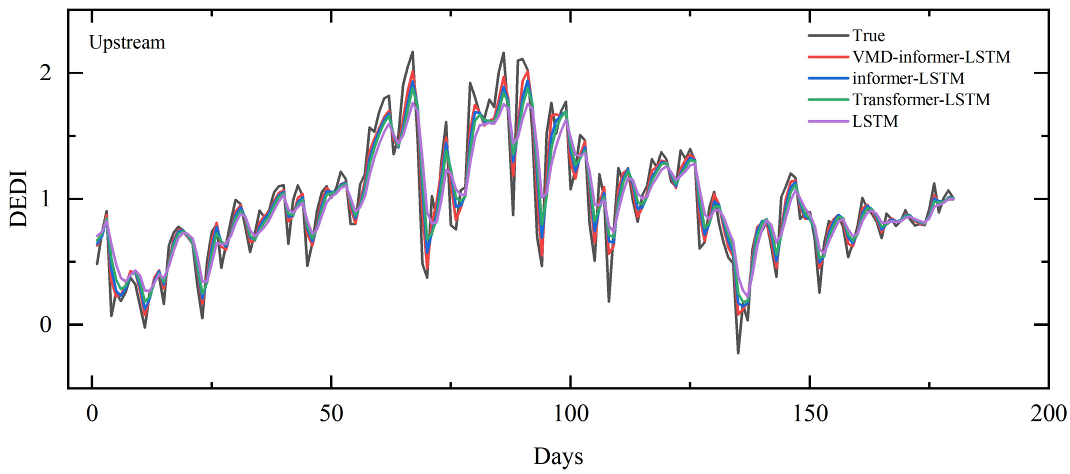
Minor Revisions

Figures 7 and 9 are currently non-intuitive. I recommend incorporating more descriptive labels or legends within the images and expanding the captions to ensure they are self-explanatory for the reader. Please provide a brief but detailed summary of the training/validation/testing split and the hyperparameter tuning process to ensure reproducibility.

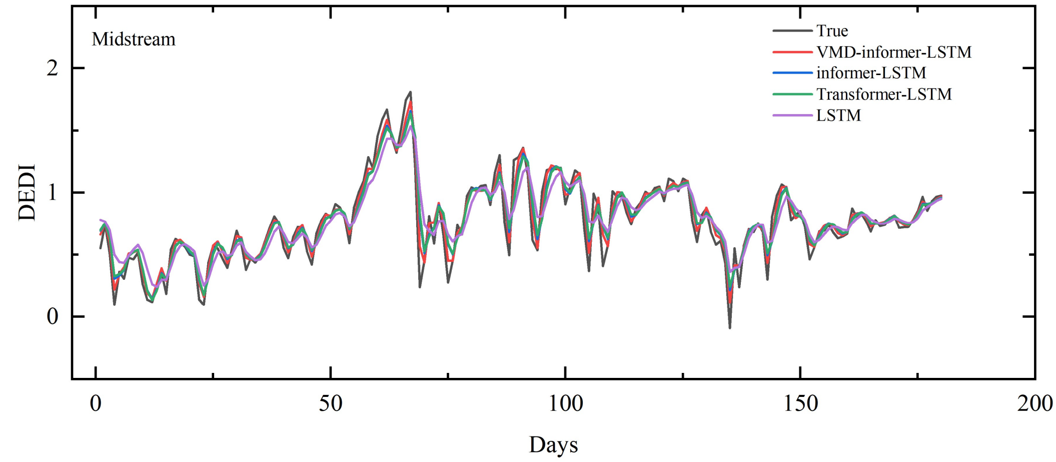
Reply: Thank you for your valuable comments. Regarding the readability of Figure 7 and Figure 9, we have made systematic adjustments based on your feedback. First, the original Figure 7 has been moved to the supplementary materials (now Figure S1) and no longer appears in the main text to avoid redundancy in the figures. Accordingly, the original Figure 9 has been renumbered as Figure 8 in the main text.

For the original Figure 9 (now Figure 8), we have enhanced its self-explanatory nature by modifying and specifying the figure title and caption. Specifically, Figure 8 has been revised to: "Figure 8 Taylor diagrams comparing the performance of different models for DEDI prediction in the Huaihe River Basin at different lead times: (a) 7 days; (b) 15 days; (c) 30 days; (d) 60 days; (e) 120 days; (f) 180 days," and the caption clearly indicates that each subplot corresponds to the forecast lead times of 7, 15, 30, 60, 120, and 180 days. Through these adjustments, the meaning of the relevant figures in the main text is now intuitive and clear. Readers can understand the comparison purpose without additional explanation. The corresponding changes have been reflected in the revised manuscript.

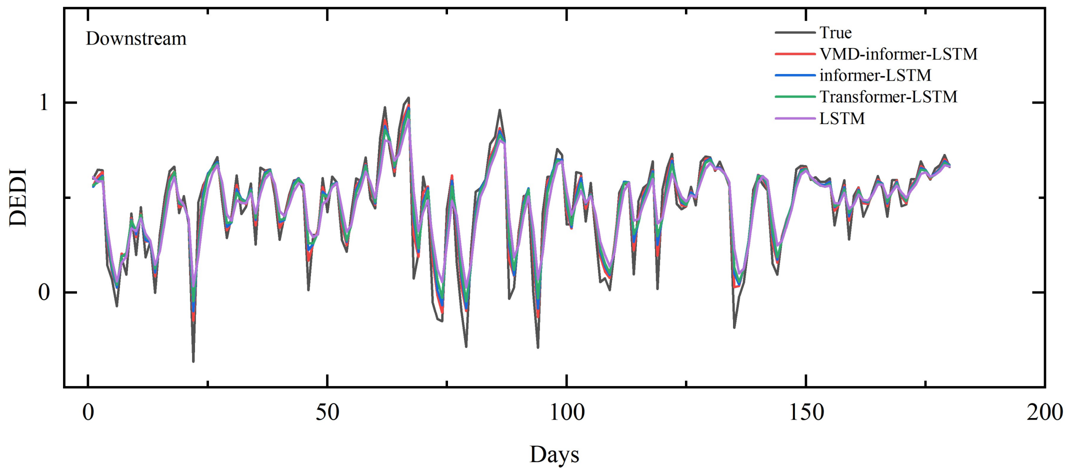
Specific image modifications are as follows:



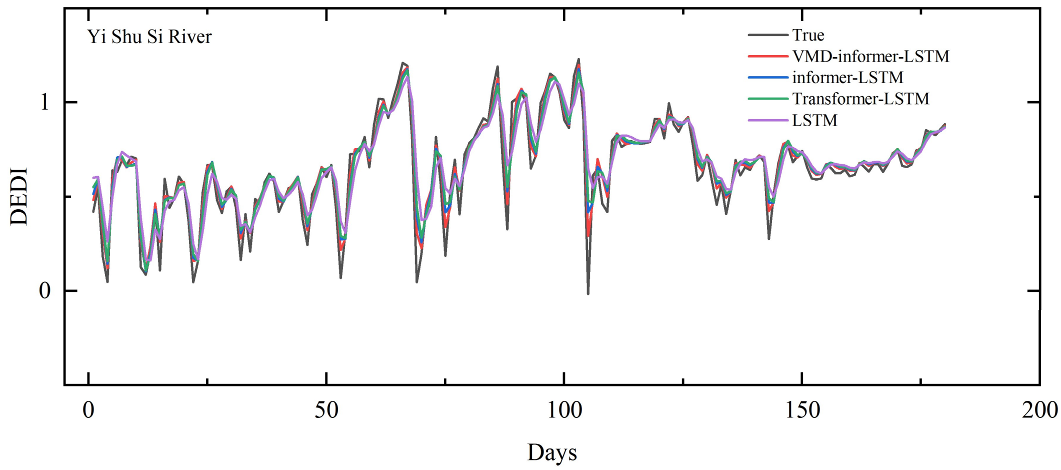
(a)



（b）



（c）



（d）

Figure 7 Line charts of different models' 180 - day predictions in four Huaihe River Basin Regions: (a) Upstream; (b) Midstream; (c) Downstream; (d) Yi Shu Si River

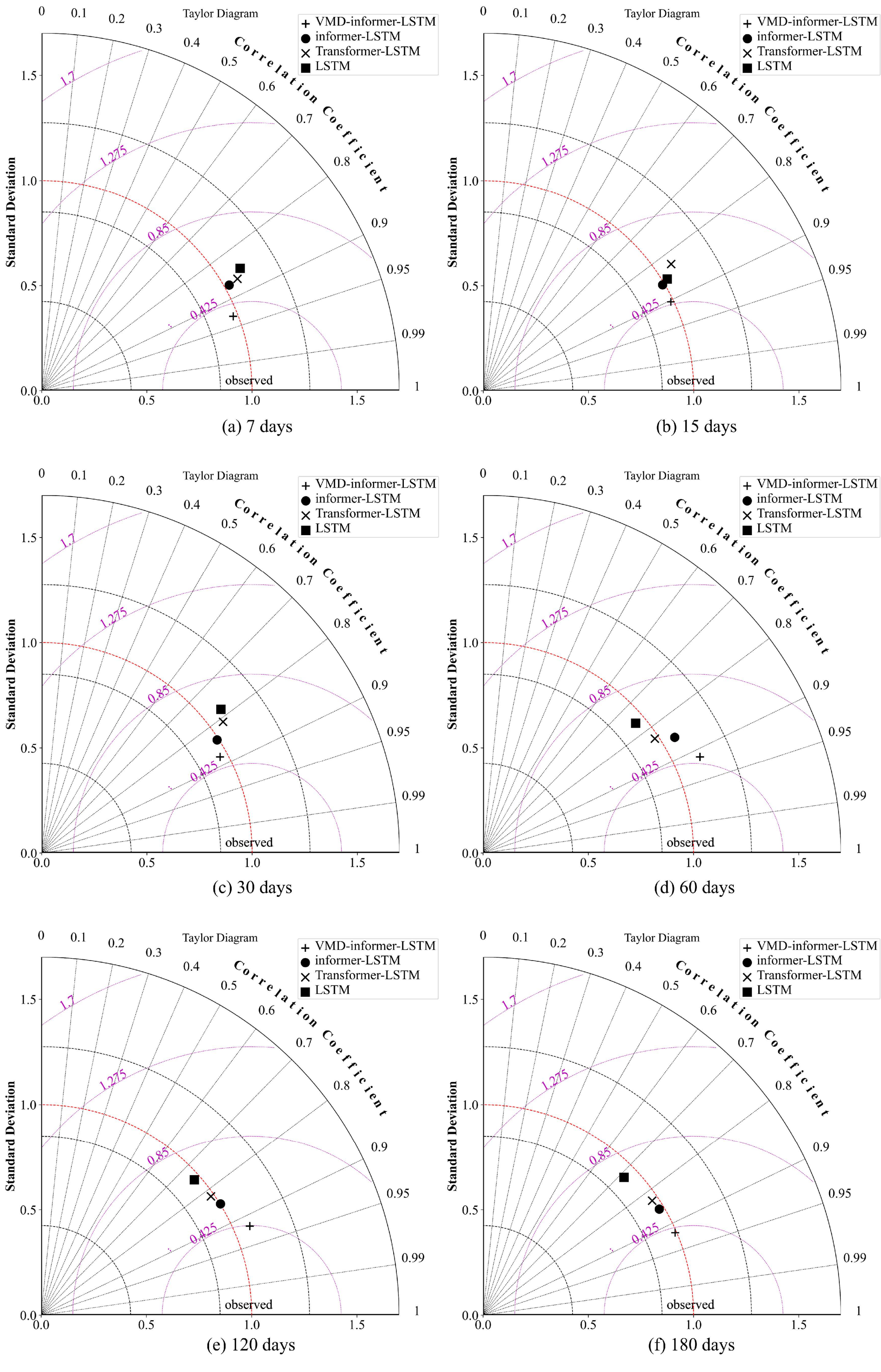


Figure 8 Taylor diagrams comparing the performance of different models for DEDI prediction in the Huaihe River Basin at different lead times: (a) 7 days; (b) 15 days; (c) 30 days; (d) 60

days; (e) 120 days; (f) 180 days.

We have added and clarified the model training, validation, and test set partitioning method, as well as the hyperparameter tuning process in subsection 3.5 "VMD-Informer-LSTM" of the revised manuscript to enhance the reproducibility of the study. Specifically, for each grid point's DEDI time series, a strict temporal split is applied: the training period covers data from 1984-1-1 to 2024-7-3, while the testing period spans from 2024-7-4 to 2024-12-31, to evaluate the model's performance on unseen data in the subsequent 180-day forecast period. This approach ensures that the model is trained on historical data and tested on future data, avoiding any potential information leakage.

Regarding the model's hyperparameter settings, we employ the Bayesian Optimization method to automatically search for key hyperparameters (such as hidden layer dimensions, learning rate, batch size, VMD decomposition parameters, etc.) within a predefined parameter space. Each candidate parameter combination is evaluated based on validation set performance, with the objective function being the minimization of validation error. The most appropriate parameter combination is then selected for model training and testing.