

Daily Drought Prediction in the Huaihe River Basin Using VMD-informer-LSTM

Min Li^{1,2}, Ming Ou¹, Yuhang Yao¹, Changman Yin¹

¹ College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou, China, 225000

² National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, China,
210098

Correspondence to: Min Li (limintju@126.com)

Abstract

Accurate drought prediction is a key challenge in water resource management and agricultural planning. This study proposes a novel drought prediction framework that integrates Variational Mode Decomposition (VMD), Informer, and Long Short-Term Memory (LSTM) networks to enhance hydrological drought forecasting in the Huaihe River Basin, China. The VMD-Informer-LSTM model decomposes complex non-stationary drought sequences into multi-scale components, effectively extracting long-term trends and short-term fluctuations. Results show that the model outperforms LSTM, Transformer-LSTM, and Informer-LSTM, improving R^2 , RMSE, MAE, and MAPE by 28.4%, 46.2%, 46.5%, and 50.8%, respectively, over the baseline LSTM. When the prediction period is 30 days, the VMD-Informer-LSTM achieves the highest prediction accuracy. During the 120–180 day prediction period, the prediction accuracy of all models declines, with drought intensity generally underestimated. Misclassifications are mainly concentrated in the transition zones between humid and semi-humid regions, with higher error frequency in semi-humid areas. Prediction accuracy is highest in the upstream and downstream regions, followed by the Yishuisi River Basin, while the midstream region performs poorly due to human interference. Shapley Additive Explanations (SHAP) further reveal that precipitation and temperature are the dominant meteorological drivers, jointly accounting for nearly half of the model's predictive power. These results confirm that the VMD-Informer-LSTM provides the most accurate predictions among the tested models, offering valuable support for drought risk assessment and water resource management in the Huaihe River Basin and other similar regions.

Keywords: Drought prediction, DEDI, Variational Mode Decomposition, informer, LSTM

1 Introduction

Drought represents one of the most spatially extensive, temporally persistent, and far-reaching natural disasters globally, with complex formation mechanisms involving intricate interactions between atmospheric circulation patterns, land surface processes, and human activities (Alsubih et al., 2021; Dai, 2013). Drought events not only directly threaten watershed water security and agricultural productivity but also profoundly affect regional ecosystem stability and socioeconomic development (Zhang et al., 2018). With intensifying climate change and increasing human activity intensity, drought events exhibit significant upward trends in frequency, intensity, and spatial extent (Cook et al., 2020; Trenberth et al., 2014). Therefore, accurate drought prediction is crucial for developing scientific disaster mitigation strategies, optimizing water resource allocation schemes,

39 and ensuring regional food and ecological security.

40 However, drought prediction faces major challenges due to the inherent complexity of drought
41 phenomena (Hao et al., 2017). Drought occurrence and evolution are controlled by multiple natural
42 and anthropogenic factors, including precipitation distribution, evapotranspiration processes,
43 topographic conditions, land use changes, and human interventions (AghaKouchak et al., 2015;
44 Vicente-Serrano et al., 2018). These factors generate highly complex, nonlinear, and non-stationary
45 spatiotemporal evolution patterns. Drought time series typically contain multi-scale periodic
46 oscillations, long-term trend changes, and stochastic fluctuation components that are mutually
47 coupled and interdependent, forming extremely complex dynamical systems (Belayneh et al., 2014;
48 Huang et al., 2015).

49 Traditional drought prediction methods rely on physics-based numerical models and st
50 atistical regression approaches (Dutra et al., 2014; Yuan & Quiring, 2017). Physics-based
51 methods include Global Climate Models (GCMs) such as the ECMWF and NCEP-CFSv2,
52 which can provide global-scale long-term climate predictions but have coarse spatial resolu
53 tions (typically 100-200 km) and cannot adequately capture regional drought characteristics
54 (Saha et al., 2014). Regional Climate Models (RCMs) employ dynamic downscaling techni
55 ques to achieve high resolutions (10-50 km) but inherit systematic biases from driving mo
56 dels and require substantial computational resources (Jacob et al., 2014; Rummukainen, 20
57 10). Land surface models such as Variable Infiltration Capacity (VIC), Community Land
58 Model (CLM), and Noah simulate coupled water cycle, energy balance, and vegetation dyn
59 amics processes but are highly sensitive to meteorological forcing data quality and paramet
60 erization schemes (Ek et al., 2003; Lawrence et al., 2011). Statistical methods include line
61 ar regression approaches, time series analysis and spectral/wavelet analysis techniques, etc
62 (Box et al., 2015; Modarres, 2007). Time series models, such as Autoregressive Moving A
63 verage (ARMA), Autoregressive Integrated Moving Average (ARIMA) models, Random Fo
64 rest (RF), demonstrate certain capabilities for stationary time series. However, their predicti
65 on performance significantly deteriorates on non-stationary, multi-periodic drought sequen
66 ces (Mishra & Desai, 2005; Mossad & Alazba, 2015). Despite contributions from these tradi
67 tional methods, fundamental limitations persist across both physics-based and statistical app
68 roaches (Hao et al., 2017; Morid et al., 2006). Rigid model structures in both GCMs and
69 RCMs cannot adaptively adjust to accommodate intrinsic data characteristics. Insufficient n
70 onlinear processing capabilities in land surface models and statistical methods (including A
71 RMA, ARIMA, and RF models) cannot capture complex feedback mechanisms and thresho
72 ld effects(AghaKouchak et al., 2015). Additionally, there are difficulties in multi-scale infor
73 mation integration and heterogeneous data fusion, particularly in land surface models and
74 GCMs (Wood et al., 2016). High parameter sensitivity affects the robustness and generaliz
75 ation capability of both physics-based and statistical models (including ARMA, ARIMA, a
76 nd RF models) (Svoboda et al., 2002). Finally, trade-offs between computational efficiency
77 and accuracy challenge the operational implementation requirements of RCMs and complex
78 statistical models (Mo, 2008; Yuan & Quiring, 2017).

79 Recent advances in artificial intelligence and big data technologies have fundamentally
80 transformed time series modeling and prediction across multiple fields (LeCun et al., 2015;
81 Shlezinger et al., 2023). Deep learning methods demonstrate significant advantages in automatically
82 capturing complex patterns and latent features without requiring pre-specified physical relationships,

83 possessing powerful nonlinear mapping and adaptive learning capabilities (Bengio et al., 2013;
84 Schmidhuber, 2015). Long-Short-Term Memory (LSTM) networks have been successfully applied
85 to various hydrological nonlinear sequence modeling tasks, including streamflow prediction, flood
86 forecasting, and water level estimation, demonstrating superior predictive performance (Kratzert et
87 al., 2018; Zhang et al., 2014). LSTM networks through their unique gate mechanisms and memory
88 cell design effectively address gradient vanishing problems in traditional recurrent neural networks
89 and exhibit excellent performance in capturing long-term dependency information (Greff et al.,
90 2017). However, single deep learning architectures still have limitations when processing complex
91 time series data, and multi-model ensemble and parallel architecture designs provide novel
92 approaches for further enhancing prediction performance (Mosavi et al., 2018; Sit et al., 2020).

93 To enhance model capability for processing non-stationary complex sequences, signal
94 decomposition algorithms for data preprocessing have become key strategies for improving time
95 series prediction performance. Variational Mode Decomposition (VMD), an advanced adaptive
96 signal decomposition technique proposed by Dragomiretskiy and Zosso in 2014, decomposes non-
97 linear, non-stationary complex sequences into multiple Intrinsic Mode Functions (IMFs) with
98 different center frequencies. Each IMF reflects the dynamic characteristics of the original sequence
99 at specific frequency levels, possessing relatively independent frequency bandwidth and amplitude
100 modulation properties (Dragomiretskiy & Zosso, 2014). Compared to traditional decomposition
101 methods such as Empirical Mode Decomposition (EMD) and Ensemble EMD (EEMD), VMD is
102 based on rigorous variational optimization theoretical frameworks, employs completely non-
103 recursive decomposition models, effectively avoids mode mixing and end-effect problems, and
104 possesses superior frequency separation effects and noise robustness.

105 In recent years, Transformer architectures have achieved major breakthroughs in time series
106 prediction, particularly the informer model specifically optimized for long time series prediction
107 tasks. Zhou proposed that informer reduces computational complexity from $O(L^2)$ to $O(L \log L)$
108 through Probabilistic Sparse Self-attention mechanisms, combined with self-attention distillation
109 operations that progressively compress sequence length layer by layer, significantly improving
110 efficiency and accuracy in processing lengthy sequences (Zhou et al., 2021a). This innovative
111 architecture provides novel technical pathways for capturing long-range temporal dependencies.

112 This study adopts the Daily Evapotranspiration Deficit Index (DEDI) as a drought monitoring
113 indicator, constructed based on daily actual and potential evapotranspiration from ERA5 reanalysis
114 data, which can effectively reflect the dynamic evolution processes of regional droughts (Hersbach
115 et al., 2020; Zhang et al., 2022; Zuo et al., 2020). Building upon this foundation, we propose a novel
116 drought prediction model integrating Variational Mode Decomposition, informer, and Long Short-
117 Term Memory networks (VMD-informer-LSTM), which is expected to provide a high-accuracy,
118 robust prediction framework for drought prediction applications by combining the technical
119 advantages of variational optimization decomposition, probabilistic sparse attention mechanisms,
120 and gated memory networks. Through a combination of multiscale feature decomposition and
121 hybrid deep learning architectures, this method effectively handles the non-stationary characteristics
122 of complex drought indices like DEDI, accurately capturing long-range climate trends and short-
123 term fluctuations. This study significantly improved prediction accuracy and reliability of complex
124 non-stationary drought time series and provided scientific foundations for regional water resource
125 management and drought risk assessment (Pozzi et al., 2013; Willmott & Matsuura, 2005).

126 The technical approach includes: (1) utilizing VMD for adaptive modal decomposition of

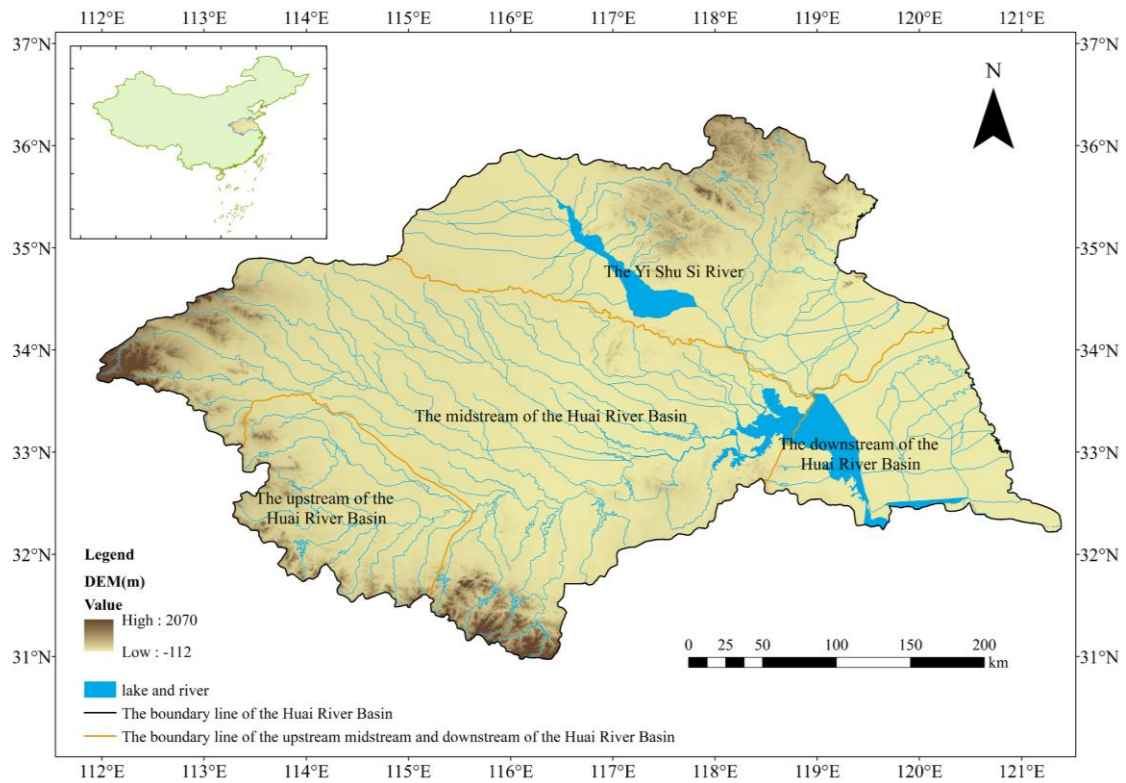
127 original DEDI sequences, deconstructing complex nonlinear time series into multi-frequency scale
128 IMF components to achieve structured extraction of multi-scale features (Dragomiretskiy & Zosso,
129 2014; Johny et al., 2022); (2) constructing dual-branch parallel architecture of informer and LSTM,
130 where informer efficiently captures global trends of long-range sequences through probabilistic
131 sparse attention mechanisms, while LSTM precisely models local temporal dynamics through
132 gating mechanisms (Zhou et al., 2021b); (3) fusing dual-source features through fully connected
133 layers to form hybrid feature representations possessing both long-range dependency analysis
134 capability and short-term fluctuation capture ability (Li et al., 2023; Zhang et al., 2019); (4) adopting
135 a three-stage design of decomposition-parallelization-fusion to obtain final drought prediction
136 results.

137 2 Study Area and data

138 2.1 Study Area

139 The Huaihe River Basin originates from the Tongbai Mountains in Nanyang City, Henan
140 Province, China. It is located between 111°55'-121°20'E longitude and 30°55'-36°20'N latitude,
141 covering approximately 270,000 km². The basin is situated in China's north-south climate transition
142 zone, with the area north of the Huaihe River belonging to the warm temperate zone and the area
143 south of the river belonging to the northern subtropical zone (Yao et al., 2024). The annual mean
144 temperature ranges from 11-16°C, with temperature variations increasing from north to south and
145 from coastal to inland areas. The Huaihe River Basin receives a multi-year average precipitation of
146 883 mm, with a spatial distribution characterized by higher precipitation in mountainous areas
147 compared to plains, increased precipitation along the coast than inland, and a decreasing gradient
148 from south to north. The multi-year average evaporation in the basin fluctuates between 650-1250
149 mm, primarily concentrated during May-August, with an overall decreasing trend from south to
150 north and from east to west. The studied region is shown in Fig. 1.

151 During the 62-year period from 1949 to 2010, the Huaihe River Basin experienced cumulative
152 drought-affected areas of 167 million hectares, with disaster-affected areas of 87.3 million hectares,
153 resulting in grain losses of 13.96 billion kg. On average, 2.698 million hectares of crops were
154 affected by drought annually, with 1.408 million hectares suffering disaster-level impacts (Gao et
155 al., 2015). Drought disasters have severely impacted industrial and agricultural production, urban
156 and rural water supply security, and ecological environments within the basin, becoming one of the
157 primary factors constraining rapid and sustainable socio-economic development in the region.
158 Therefore, providing reliable drought prediction methods is of great significance for accurate
159 drought forecasting and the scientific development of drought response strategies in the Huaihe
160 River Basin.



161
162

Figure 1 The study area of Huaihe river Basin

163

2.2 Data Sources

164 This study utilizes daily actual evapotranspiration, potential evapotranspiration, surface
 165 pressure, cloud cover, maximum temperature, mean temperature, wind speed, and precipitation data
 166 for the period 1980-2020, sourced from the fifth-generation high-resolution atmospheric reanalysis
 167 product (ECMWF Reanalysis v5, ERA5) developed by the European Centre for Medium-Range
 168 Weather Forecasts (ECMWF). ERA5 data are characterized by extensive coverage, long time series,
 169 and excellent spatio-temporal consistency. This makes them an optimal data choice for long-term
 170 and high-resolution analyses. To meet the specific spatial resolution requirements of this study, the
 171 downloaded ERA5 data were further subjected to interpolation processing. This converted them to
 172 a spatial resolution of $0.25^\circ \times 0.25^\circ$ to more precisely match the spatial characteristics of the study
 173 region (Muñoz-Sabater et al., 2021). Land use data for 2005, 2010, 2015, and 2020 with a spatial
 174 resolution of 1 km were obtained from the Data Center for Resources and Environmental Sciences,
 175 Chinese Academy of Sciences (<https://www.resdc.cn/>).

176

3. Methods

177

3.1 DEDI Index

178 The Daily Evapotranspiration Deficit Index (DEDI) is a daily drought index constructed based
 179 on daily evapotranspiration as well as potential evapotranspiration for monitoring and predicting
 180 regional drought events(Zhang et al., 2022). This index is calculated based on ERA5 data provided
 181 by ECMWF.

182 The DEDI is calculated as follows:

$$DEDI_i = \frac{D_i - D_{AVE}}{D_{STU}} \quad (1)$$

$$D_i = AET_i - PET_i \quad (2)$$

183 where i represents time, AET_i represents the actual evapotranspiration on day i (units:
 184 mm/day), PET_i represents the potential evapotranspiration on day i (units: mm/day), D_i
 185 represents the evapotranspiration deficit between AET and PET on day i , and D_{AVE} and D_{STU} are
 186 the multi-year climatological mean and standard deviation, respectively (Zuo et al., 2020).

187 3.2 VMD

188 Variational Mode Decomposition (VMD) was proposed by Konstantin Dragomiretskiy
 189 (Dragomiretskiy & Zosso, 2014) as a signal processing method designed to effectively overcome
 190 mode mixing and end effect problems existing in Empirical Mode Decomposition (EMD). Unlike
 191 the recursive decomposition principle of EMD, VMD determines the central frequency and
 192 bandwidth of each mode component. This is done by constructing and solving the optimal solution
 193 to a variational model. This represents a completely non-recursive decomposition model. This
 194 method searches for a set of mode components and their corresponding center frequencies through
 195 an iterative search. This ensures that each mode maintains smoothness after demodulation to
 196 baseband.

197 The adaptivity of VMD is reflected in its ability to automatically determine the number of
 198 modes decompositions according to signal characteristics and adaptably match the optimal center
 199 frequency and finite bandwidth for each mode, thereby achieving effective separation of Intrinsic
 200 Mode Functions (IMFs) and frequency domain partitioning of signals. Experimental results
 201 demonstrate that VMD exhibits strong robustness in sampling and noise aspects, is capable of
 202 reducing the non-stationarity of time series with high complexity and strong non-linearity and
 203 decomposing them into multiple sub-sequences with different frequency scales that are relatively
 204 stationary, making it particularly suitable for non-stationary signal processing.

205 VMD decomposes time series into simple high-frequency and low-frequency intrinsic mode
 206 functions through optimization processes, improving signal processing stability and accuracy. This
 207 method is not only theoretically innovative but also demonstrates superior performance in practical
 208 applications, providing an effective tool for non-stationary signal analysis. Research by Zhao et al.
 209 (2023) further proved VMD's excellent performance in handling boundary effects by adjusting
 210 parameters (such as decomposition levels, quadratic penalty terms, etc.) to effectively control
 211 deviations in decomposition results, thereby improving model adaptability (Zhang et al., 2023; Zhao
 212 et al., 2023).

213 In this study, two key parameters of the Variational Mode Decomposition (VMD) need to be
 214 predefined: the penalty factor α (bandwidth constraint parameter) and the number of modes K (i.e.,
 215 the number of intrinsic mode functions, IMFs). The penalty factor α is empirically determined based
 216 on the length of the time series, with its value ranging from 1.5 to 2.0 times the sample length,
 217 aiming to balance frequency band separation and decomposition stability. When α is relatively small,
 218 the bandwidth of each IMF becomes wider, which may lead to spectral overlap between different
 219 modes and thus reduce the physical interpretability of the decomposition results; whereas when α is
 220 excessively large, the bandwidth is overly constrained, making the decomposition more sensitive to
 221 noise. In this study, α is finally set to 1.75 times the sample length. In addition, the number of modes
 222 K is determined based on the frequency distribution characteristics of the decomposed signal and

223 preliminary experimental results. When K is less than 7, certain IMF components exhibit significant
 224 spectral mixing, making it difficult to effectively separate signals at different time scales; when K
 225 exceeds 7, the center frequencies of adjacent IMFs become too close and the energy distribution
 226 tends to be dispersed, resulting in redundant or noise-dominated modes and thus reducing the
 227 stability and physical interpretability of the decomposition. Therefore, K is uniformly set to 7 in this
 228 study to ensure effective separation of the dominant frequency components of the original DEDI
 229 series while avoiding redundancy caused by over-decomposition. Preliminary experiments
 230 conducted on several representative grid points indicate that this parameter combination yields
 231 stable decomposition results and satisfactory predictive performance. To maintain methodological
 232 consistency and avoid potential spatial overfitting, the same VMD parameter settings are applied
 233 uniformly across all grid points in the study area.

234 Assuming the original signal f is decomposed into k components, ensuring that the decomposed
 235 sequences are modal components with finite bandwidth and center frequencies, while minimizing
 236 the sum of estimated bandwidths of all modes, with the constraint that the sum of all modes equals
 237 the original signal, the VMD constrained variational model is as follows:

$$\min_{\{u_k, \omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (3)$$

$$\text{s. t. } \sum_k u_k = f(t) \quad (4)$$

$$u_k = \{u_1, u_2, \dots, u_k\} \quad (5)$$

$$\omega_k = \{\omega_1, \omega_2, \dots, \omega_k\} \quad (6)$$

238 where $f(t)$ represents the original data; K represents the number of modal components;
 239 $\delta(t)$ represents the Dirac function; $*$ represents convolution operation; ∂_t is the partial derivative
 240 operator; $\{u_k\}$ represents the k -th component function obtained through calculation; $\{\omega_k\}$
 241 represents the center frequency of the k -th component obtained through calculation.

242 The augmented LaGrange is introduced to solve this constrained optimization problem:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (7)$$

243 where α represents the penalty factor; $\lambda(t)$ represents the Lagrange multiplier.

244 The alternating direction method of multipliers (ADMM) is used to find the saddle point of the
 245 augmented LaGrange. In the frequency domain, the updates are:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k}^k \hat{u}_i^{n+1}(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (8)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k^{n+1}(\omega)| d\omega}{\int_0^\infty |u_k^{n+1}(\omega)| d\omega} \quad (9)$$

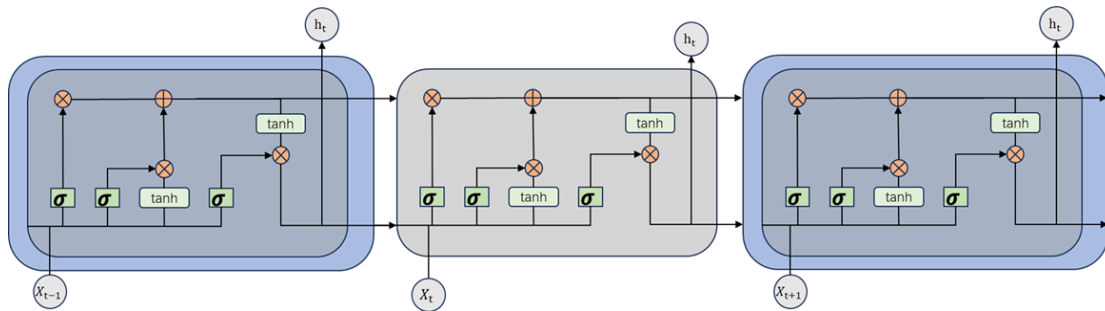
$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \pi \left[\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) \right] \quad (10)$$

246 where γ represents noise tolerance; the $\hat{u}_k^{n+1}(\omega)$ represent Wiener filtering residuals; $\hat{u}_i(\omega)$,
 247 and $\hat{\lambda}^n(\omega)$ represent the Fourier transforms of $u(t)$ and $\lambda(t)$, respectively.

248 3.3 LSTM

249 The Long-Short-Term Memory (LSTM) models are a special type of Recurrent Neural
 250 Network (RNN) variant that addresses the gradient vanishing problem in RNN processing of long
 251 sequence data by introducing memory cells (Cell States), specifically designed for processing time
 252 series data (Hochreiter & Schmidhuber, 1997). When recording long sequence data, RNN
 253 experiences gradient vanishing or exploding due to continuous information accumulation, making
 254 it difficult for the network to learn long-term dependencies, ultimately affecting the model's accurate
 255 capture of trends and periodicity in time series data (Jaseena & Kovoov, 2022).

256 LSTM aims to solve the gradient-vanishing or exploding problems encountered by traditional
 257 RNN when processing long sequence data. This is done mainly through gate mechanisms regulating
 258 information flow. The LSTM model consists of four interacting layers: input gate, forget gate, cell
 259 state gate, and output gate. It is shown in Fig. 2 below.



260
261 Figure 2 The construction of LSTM

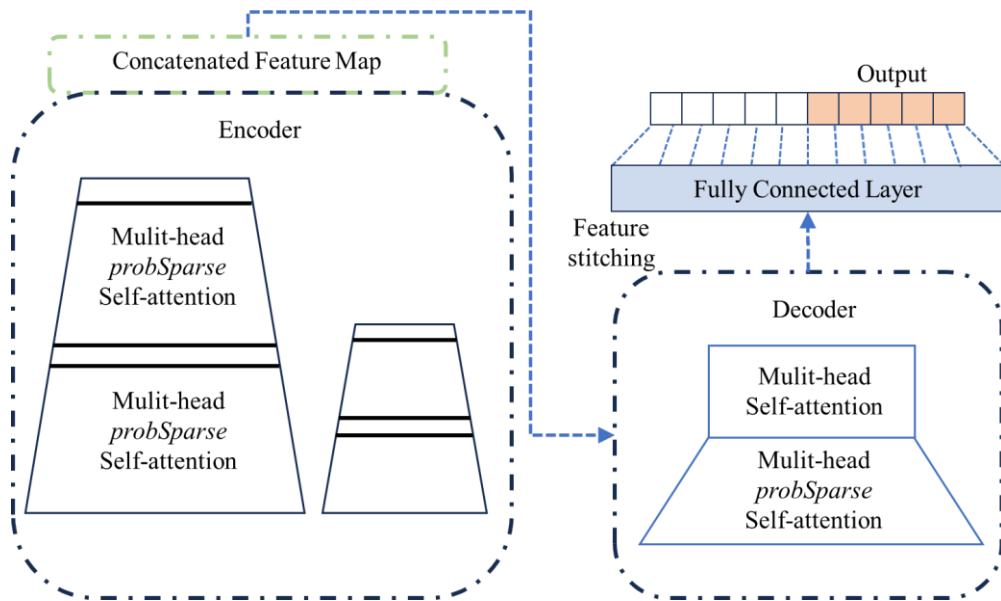
262 3.4 informer

263 informer is an improved and optimized version based on Transformer, specifically designed to
 264 enhance the speed and efficiency of processing long sequences and optimized for long-term time
 265 series prediction tasks (Zhou et al., 2021b). Transformer captures relationships between different
 266 positions in sequences through Self-Attention mechanisms. However, Transformer encounters
 267 difficulties when processing long time series data because its computational complexity grows
 268 quadratically with sequence length, becoming very slow or even unprocessable when dealing with
 269 very long time series.

270 informer proposes ProbSparse Self-attention to filter critical queries and reduce computational
 271 complexity, and introduces Self-attention Distilling to reduce dimensions and network parameters
 272 (Vaswani et al., 2017). As shown in Fig. 3, the informer architecture consists of an encoder-decoder
 273 structure. Input sequences first undergo convolutional encoding and positional embedding before
 274 being fed into the encoder; the encoder utilizes multi-layer probabilistic sparse self-attention and
 275 distillation operations to extract key information, outputting encoded features. The decoder then
 276 directly generates long sequence predictions under masked self-attention and cross-attention,
 277 followed by fully connected layers mapping to final values.

278 For long sequence prediction, the informer has three advantages: 1. Probabilistic sparse self-
 279 attention reduces time complexity from $O(L^2)$ to $O(L \log L)$, significantly reducing computational
 280 overhead; 2. Self-attention distillation progressively compresses the sequence length layer by layer,

281 simultaneously reducing computation and memory requirements; 3. Generative decoding outputs
 282 complete future sequences at once, avoiding error accumulation caused by step-by-step
 283 extrapolation.



284
 285

Figure 3 The construction of the informer

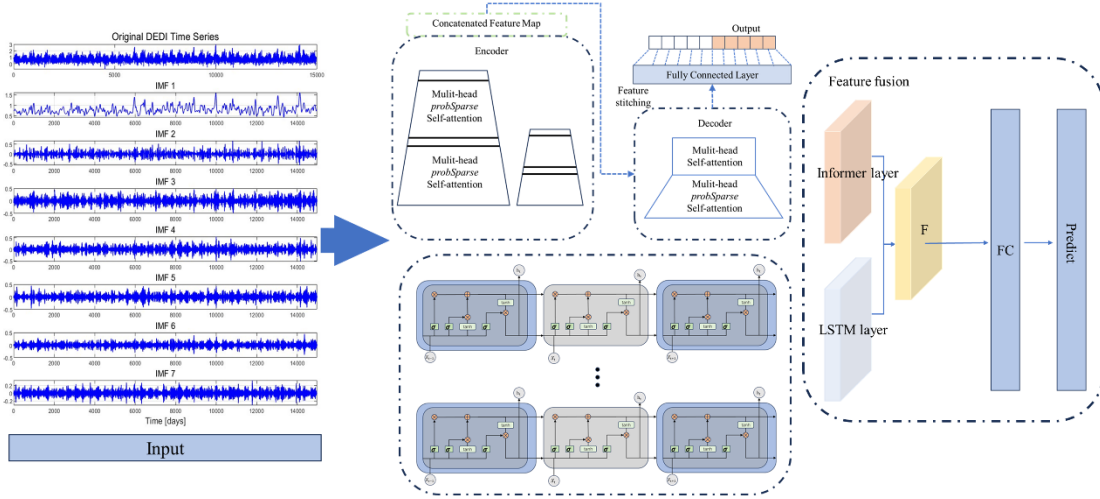
286 3.5 VMD-informer-LSTM

287 The VMD-informer-LSTM model employs Variational Mode Decomposition (VMD) to
 288 deconstruct DEDI time series into multi-frequency scale Intrinsic Mode Functions (IMFs),
 289 achieving the structured extraction of multi-scale features. Based on this foundation, it constructs a
 290 dual-branch parallel architecture of informer and LSTM, where informer efficiently captures global
 291 trends of long-range sequences through probabilistic sparse attention mechanisms, while LSTM
 292 precisely models local temporal dynamics through gate mechanisms. Finally, dual-source features
 293 are fused through fully connected layers to form hybrid feature representations possessing both
 294 long-range dependency analysis capability and short-term fluctuation capture ability. This model
 295 significantly improves prediction accuracy and reliability for complex time series data through a
 296 three-stage design of decomposition-parallelization-fusion, providing an innovative solution for
 297 time series prediction tasks. The construction of VMD-informer-LSTM is shown in Fig. 4 below.

298 Building upon this foundation, in order to ensure the standardization of the model training
 299 process and the reproducibility of the results, this study further establishes a unified design for
 300 dataset partitioning and hyperparameter optimization procedures. For the DEDI time-series data at
 301 each grid point, the dataset is strictly divided in chronological order using a forward-splitting
 302 strategy to prevent any potential leakage of future information. Specifically, the training period
 303 spans from January 1, 1984 to July 3, 2024, while the testing period covers July 4, 2024 to December
 304 31, 2024, which is used to evaluate the model's predictive performance on unseen data over the
 305 subsequent 180-day forecasting horizon. It should be noted that the 180-day prediction in this study
 306 does not correspond to traditional deterministic weather forecasting, but rather focuses on predicting
 307 the evolution of drought states.

308 Regarding hyperparameter configuration, this study adopts the Bayesian Optimization method
 309 to automatically search for key hyperparameters (including hidden layer dimension, learning rate,

310 and batch size) within a predefined parameter search space. Each candidate parameter combination
 311 is evaluated based on its predictive performance on the validation set, with the minimization of
 312 validation error serving as the optimization objective function. The optimal parameter combination
 313 is then selected for the final model training and testing evaluation. Through the design of the above
 314 training strategy and hyperparameter optimization procedure, the experimental process ensures
 315 methodological rigor, stability of results, and reproducibility of the research.



316

317

Figure 4 The construction of VMD-informer-LSTM

318 3.5 Shapley Additive Explanations

319 Since machine learning models are “black-box” models, although they can provide efficient
 320 predictions, their internal decision-making processes are complex and difficult to intuitively
 321 understand and explain. This may affect result analyses in the field of raster data, which require
 322 transparency and interpretability. To overcome such problems, SHAP values are introduced as an
 323 interpretive method. The SHAP value method was proposed by Lundberg et al (Lundberg & Lee,
 324 2017). In 2017, it is a method based on cooperative game theory that quantifies the contribution of
 325 driving factors to model prediction results. By calculating the marginal contributions of each factor
 326 to the model output under different combinations, it measures their importance in the overall
 327 prediction results. This helps us understand how the model makes decisions. The positive or
 328 negative values of SHAP indicate promotion or inhibition of prediction results. The absolute value
 329 reflects the degree of influence of the factor on the model prediction results. The larger the absolute
 330 value, the greater the influence of the factor on model prediction results (Wang et al., 2024). The
 331 formula is as follows:

$$\phi_i = \sum_{S \in N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (11)$$

332

333

334

In the formula, ϕ_i represents the SHAP value for feature i ; N is the set of all features; S is a
 subset of features, excluding feature i ; $f(S)$ is the model output using only the feature subset S for
 prediction; ($f(S \cup \{i\})$) is the predicted value after adding feature i to the featured subset S .

335 3.6 Evaluation Metrics

336

The coefficient of determination (R^2) measures the proportion of variance in the dependent

337 variable that is derived from the independent variable, ranging from 0 to 1, where values closer to
 338 1 indicate better model performance (Nash & Sutcliffe, 1970). The root mean square error (RMSE)
 339 quantifies the average magnitude of prediction errors, providing a measure of how well the model
 340 predicts actual values, with lower values indicating better accuracy (Willmott & Matsuura, 2005).
 341 Mean absolute error (MAE) represents the average absolute difference between predicted and
 342 observed values, offering a linear score that is less sensitive to outliers than RMSE. The range of
 343 mean absolute percentage error (MAPE) is $[0, +\infty]$. A MAPE of 0% indicates a perfect model, while
 344 a MAPE greater than 100% suggests a poor model (Myttenaere et al., 2016).

345 The mathematical expressions for these metrics appear as follows:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^N (y_{obs} - y_{pred})^2}{\sum_{i=1}^N \Sigma (y_{obs} - \bar{y}_{obs})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{obs} - y_{pred})^2} \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{obs} - y_{pred}| \quad (14)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N abs\left(\frac{y_{obs} - y_{pred}}{y_{obs}}\right) \quad (15)$$

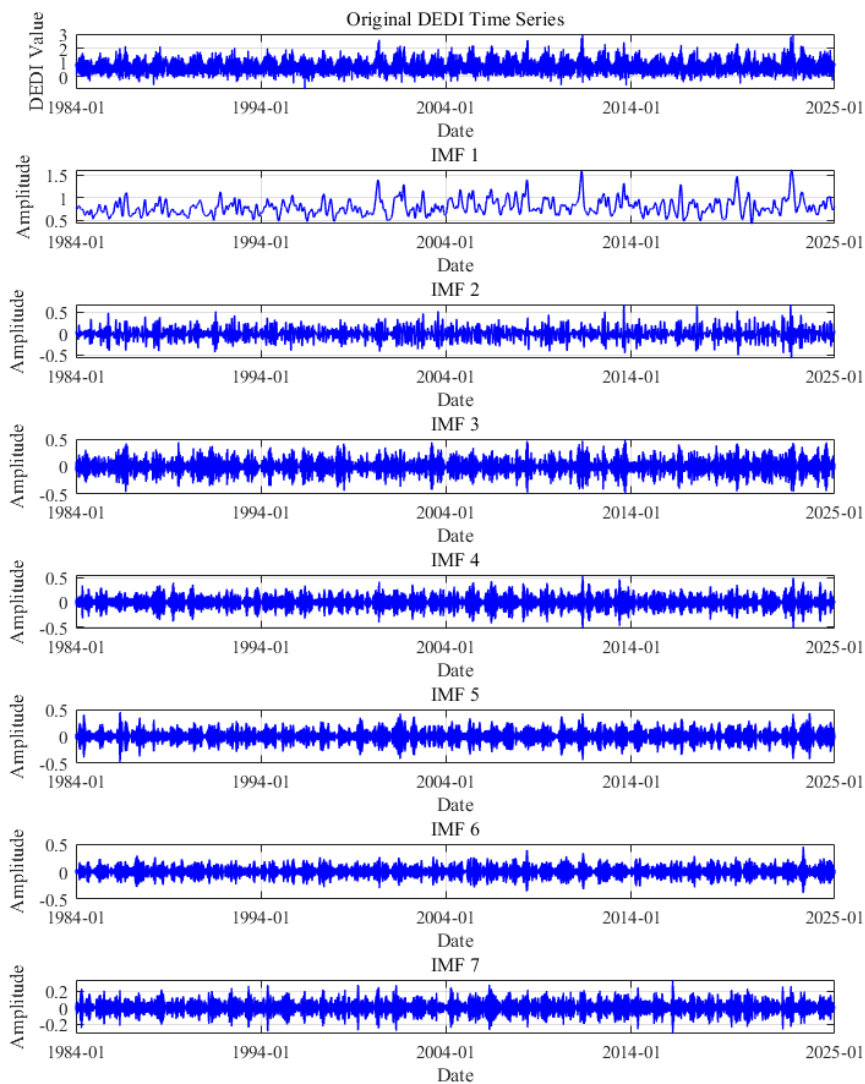
346 where y_{obs} represents observed values, y_{pred} represents predicted values, \bar{y}_{obs} is the mean of
 347 observed values, and n is the number of observations.

348

349 4. Results

350 4.1 VMD Decomposition Results

351 In this study, we systematically selected 108 grid points distributed across the Huaihe River
 352 Basin and performed VMD analysis of their corresponding daily DEDI time series data spanning
 353 1980-2020. Fig. 5 provides a comprehensive visual representation of the decomposed sub-sequences
 354 of DEDI data for a certain representative grid cell in the Huaihe River Basin. As illustrated in Fig.
 355 5, the daily DEDI values for a representative grid cell in the Huaihe River Basin exhibit substantial
 356 positive and negative fluctuations, with considerable variance between maximum and minimum
 357 values. These pronounced oscillations present significant challenges in capturing essential features
 358 during the prediction process, as the complex, non-stationary nature of the original signal obscures
 359 the underlying patterns and trends. The VMD algorithm successfully decomposes the complex, non-
 360 linear DEDI time series into multiple distinct IMFs, each characterized by specific frequency bands
 361 and temporal scales. These decomposed components reveal multi-scale variability patterns ranging
 362 from high-frequency short-term fluctuations to low-frequency long-term trends, facilitating more
 363 effective feature extraction and modeling processes. VMD technology demonstrates superior
 364 capability in accurately tracking changes in signal frequency components and effectively revealing
 365 the intrinsic and dynamic characteristics of time series, thereby substantially enhancing prediction



367
 368 Figure 5 The daily DEDI values of a certain grid in the Huaihe River Basin are decomposed into 7
 369 sub - sequence through variational mode decomposition (VMD).

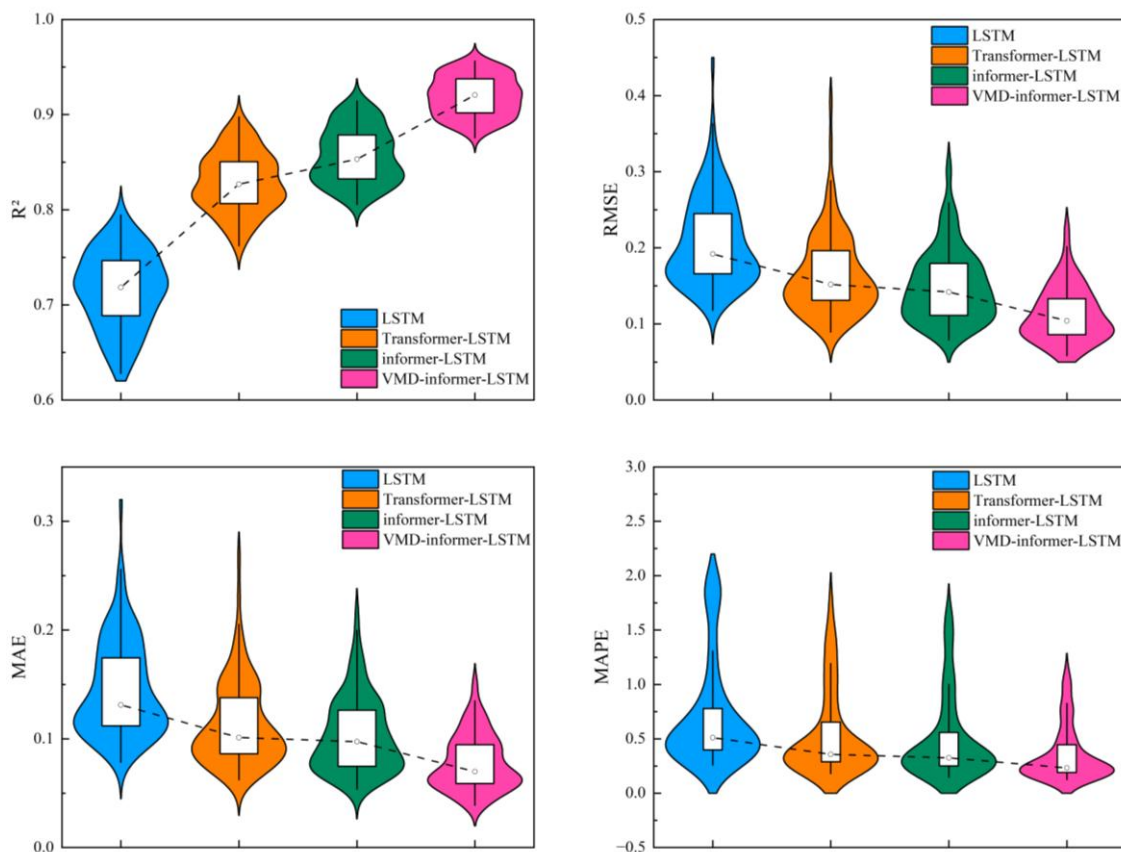
370 **4.2 Model Prediction Performance Evaluation**

371 To evaluate the performance of the proposed VMD-informer-LSTM model, we performed
 372 extensive comparative experiments involving four distinct modeling approaches: the proposed
 373 VMD-informer-LSTM model, the informer-LSTM model, the Transformer-LSTM model, and the
 374 standalone LSTM model. The evaluation of each model’s performance was based on four widely
 375 used statistical metrics for time series prediction evaluation. From Table 1, it can be seen that the
 376 VMD-informer-LSTM model performs the best in terms of predictive performance, with R^2 , RMSE,
 377 MAE, and MAPE reaching 0.9191, 0.1122, 0.0778, and 0.4056, respectively. In contrast, the
 378 informer-LSTM model without VMD decomposition has R^2 , RMSE, MAE, and MAPE of 0.8570,
 379 0.1498, 0.1031, and 0.5330, respectively. After calculation, VMD decomposition improves R^2 ,
 380 RMSE, MAE, and MAPE by 7.25%, 25.10%, 24.54%, and 23.90%, respectively. Furthermore, the
 381 traditional LSTM model shows relatively low predictive accuracy with four evaluation metrics of
 382 $R^2=0.7156$, $RMSE=0.2087$, $MAE=0.1454$, and $MAPE =0.8254$. The Transformer-LSTM model

383 achieves R^2 , RMSE, MAE, and MAPE of 0.8286, 0.1641, 0.1130, and 0.6266, respectively. This,
 384 although better than the basic LSTM model, is still not as good as the informer-LSTM model, let
 385 alone the VMD-informer-LSTM model. In summary, by comparing the predictive performance of
 386 the four models, it is evident that the VMD-informer-LSTM model has an advantage in time series
 387 prediction tasks. Especially after the introduction of VMD decomposition, its performance has
 388 significantly improved, further verifying the effectiveness of VMD decomposition in enhancing
 389 model predictive accuracy.

390 Table 1 Overall average evaluation indicators of various models within the Huaihe River Basin

Indicator	LSTM	Transformer-LSTM	informer-LSTM	VMD-informer-LSTM
R^2	0.7156	0.8286	0.8570	0.9191
RMSE	0.2087	0.1641	0.1498	0.1122
MAE	0.1454	0.1130	0.1031	0.0778
MAPE	0.8254	0.6266	0.5330	0.4056



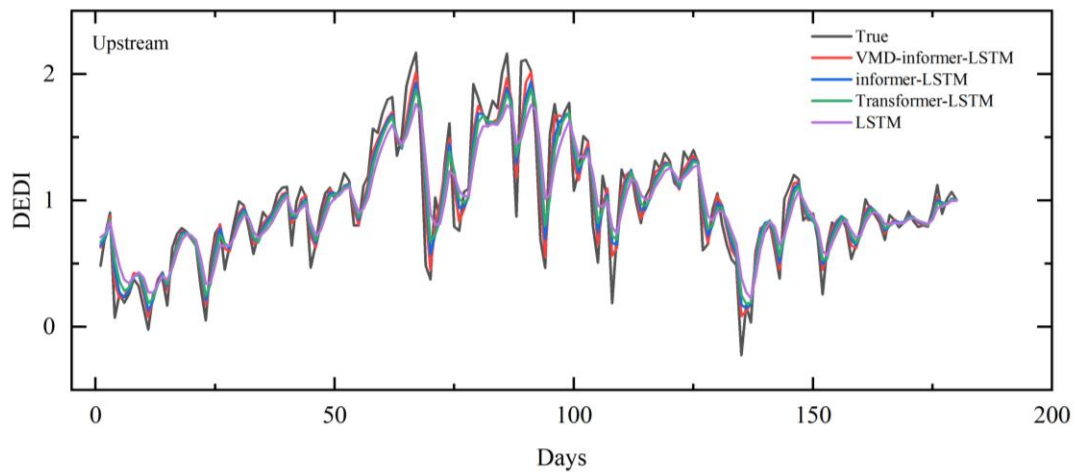
391
 392 Figure 6 Violin-box plots of evaluation indicators for different models in the Huaihe River Basin

393 Fig. 6 presents violin plots illustrating the distribution of evaluation metrics across all 108 grid
 394 points in the Huaihe River Basin. This provides insights into model performance variability and
 395 consistency. Violin plots reveal important patterns. The VMD-informer-LSTM model demonstrates
 396 the most concentrated distribution around optimal values, with R^2 distributions tightly clustered near
 397 0.92-0.95, indicating consistent and high performance across diverse geographical locations. Error
 398 metric distributions (RMSE, MAE and MAPE) show the VMD-informer-LSTM model has the
 399 narrowest spread and the lowest median values, suggesting robust prediction accuracy with minimal
 400 spatial variability.

401 The RMSE distributions demonstrate a clear monotonic improvement from LSTM (median:
402 0.2087) through Transformer-LSTM (0.1641), informer-LSTM (0.1498) to VMD-informer-LSTM
403 (0.1122), with progressively narrower interquartile ranges and fewer outliers. The VMD-informer-
404 LSTM model exhibits the most concentrated distribution with minimal dispersion (IQR: 0.05),
405 indicating superior prediction accuracy and enhanced stability across diverse geographical locations.
406 This concentrated performance distribution reflects the effectiveness of combining VMD
407 decomposition with hybrid deep learning architectures for robust drought prediction.

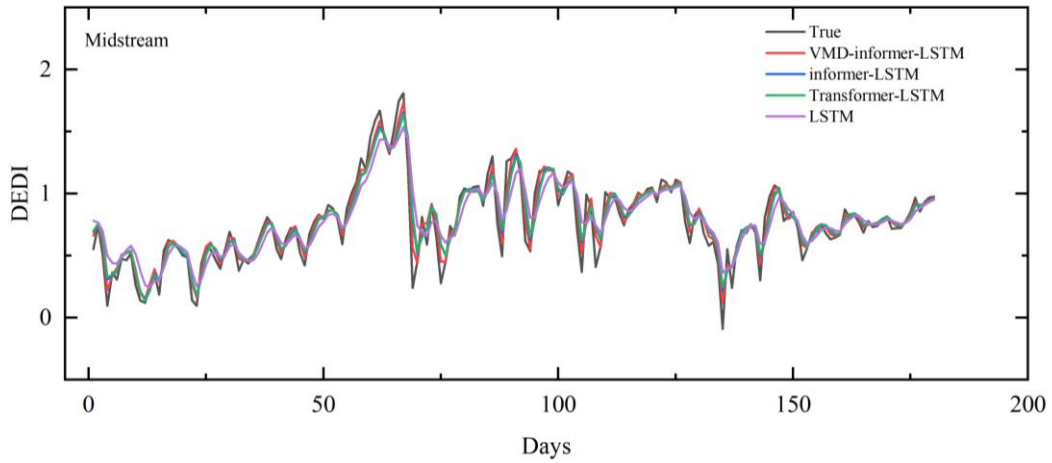
408 Comparing the four models, the VMD-informer-LSTM model consistently outperforms the
409 other three, with an average R^2 of 0.9191, RMSE of 0.1122, MAE of 0.0778, and MAPE of 0.4056.
410 The informer-LSTM model without VMD decomposition has an average R^2 of 0.8570, RMSE of
411 0.1498, MAE of 0.1031, and MAPE of 0.5330. The Transformer-LSTM model achieves an average
412 R^2 of 0.8286, RMSE of 0.1641, MAE of 0.1130, and MAPE of 0.6266. The basic LSTM model has
413 the lowest performance, with an average R^2 of 0.7156, RMSE of 0.2087, MAE of 0.1454, and
414 MAPE of 0.8254.

415 The performance improvement from LSTM to Transformer-LSTM to informer-LSTM is
416 evident, with each model showing better median values and reduced spread in the error metrics
417 compared to the previous one. However, the VMD-informer-LSTM model stands out with the most
418 significant enhancements in both median performance and reduced variance across all metrics. This
419 indicates that the integration of VMD decomposition with the informer-LSTM architecture provides
420 the most substantial benefits in terms of prediction accuracy and consistency.



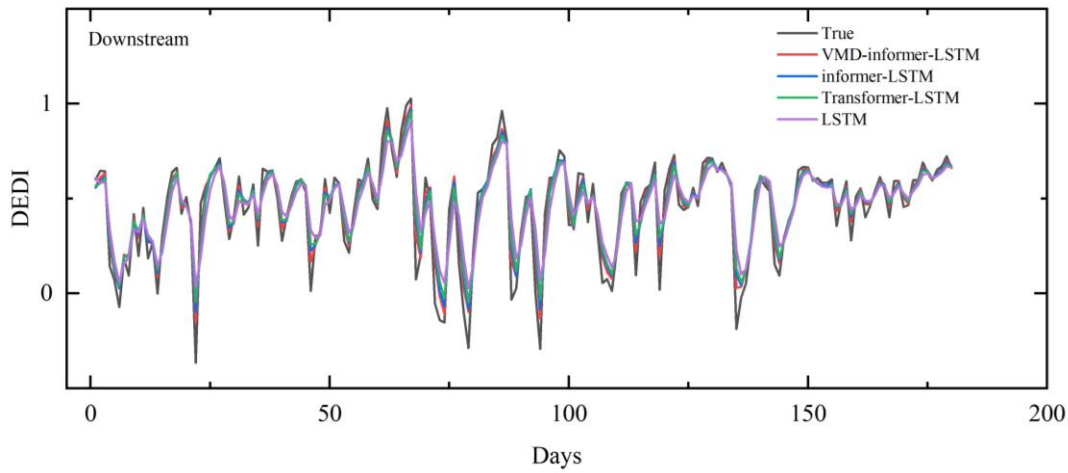
421
422

(a)



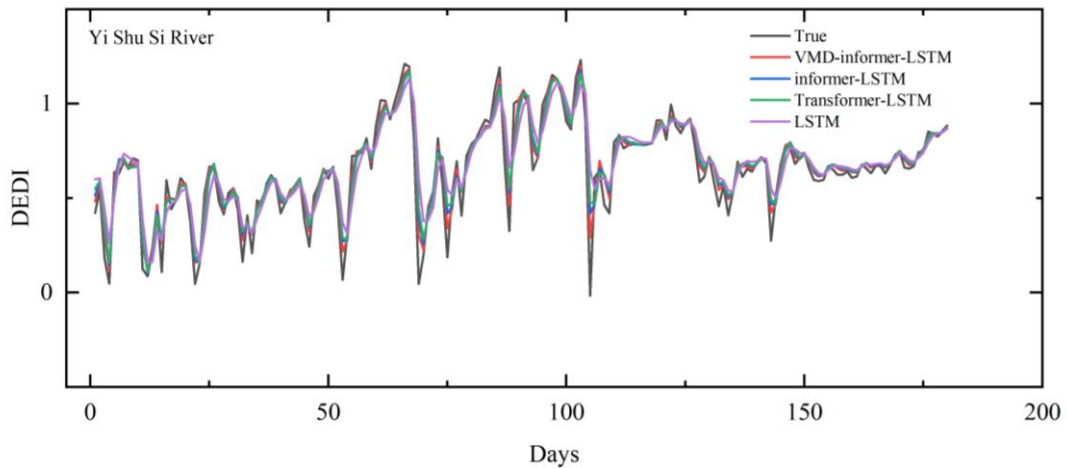
423
424

(b)



425
426

(c)



427
428

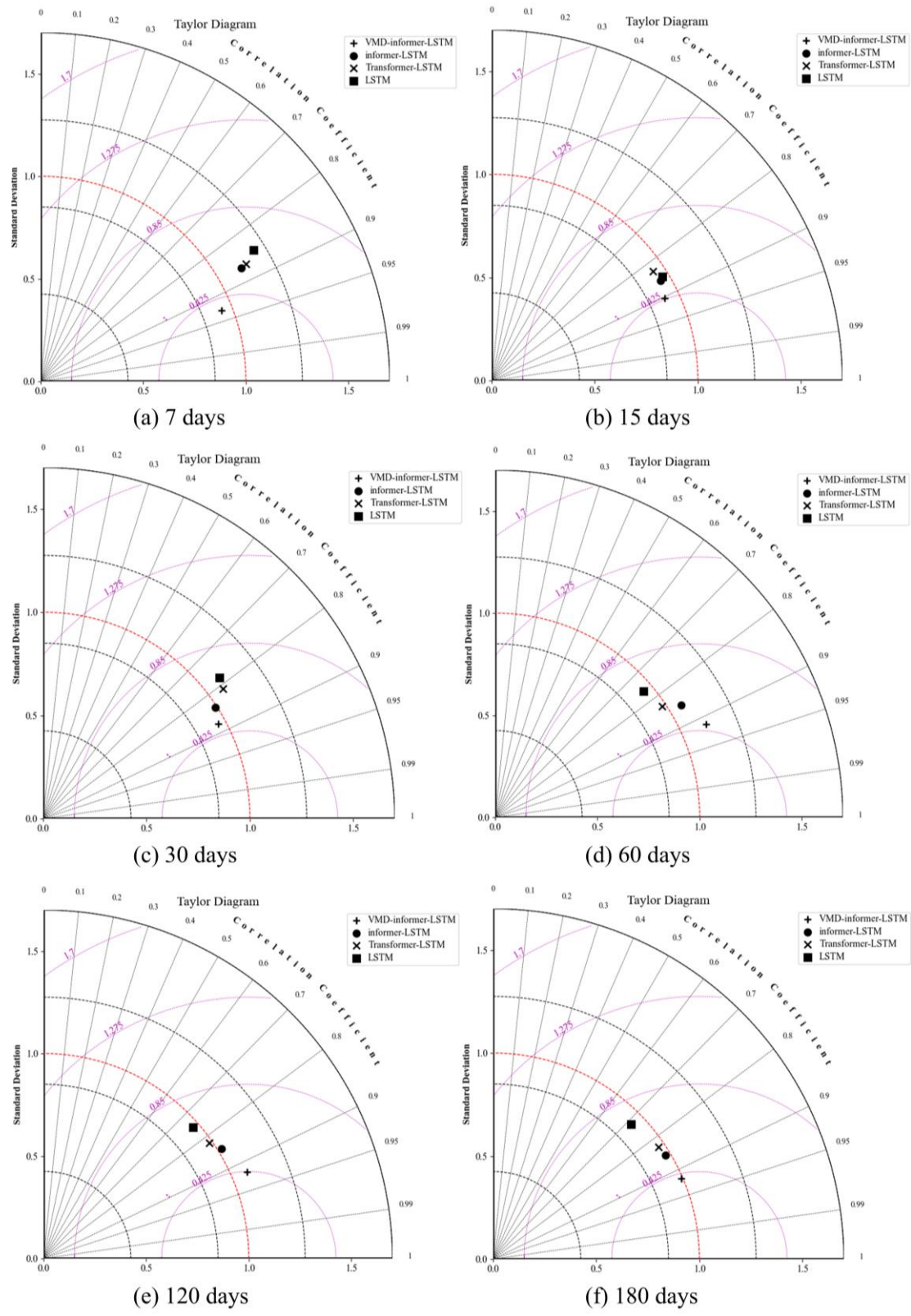
(d)

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

429
430
431
432 Fig. S1 and Fig. 7 show scatter plots and line charts of different models' 180-day predictions in different Huaihe River Basin regions. As illustrated in Fig. S1, LSTM yields the lowest R^2 values

433 (0.76–0.78) across all subplots and is therefore the poorest performer; its scatter points visibly
434 diverge from the 1:1 line, reflecting the largest prediction errors. Although informer-LSTM is less
435 accurate than VMD-informer-LSTM, it maintains a stable R^2 of about 0.89, and the scatter deviation
436 is markedly smaller than those of Transformer-LSTM and LSTM.

437 Fig. 7 further indicates that all four models successfully capture the overall drought trend, yet
438 differ in precision. When the line plots are examined in conjunction, it becomes clear that VMD-
439 informer-LSTM outperforms the other three models, delivering superior agreement between
440 simulated and ERA5 reanalysis values, higher prediction accuracy, and the best overall performance.
441 informer-LSTM and Transformer-LSTM rank next, whereas LSTM only roughly reproduces the
442 drought trend and performs poorly at predicting extremes, resulting in the lowest predictive
443 capability.



444

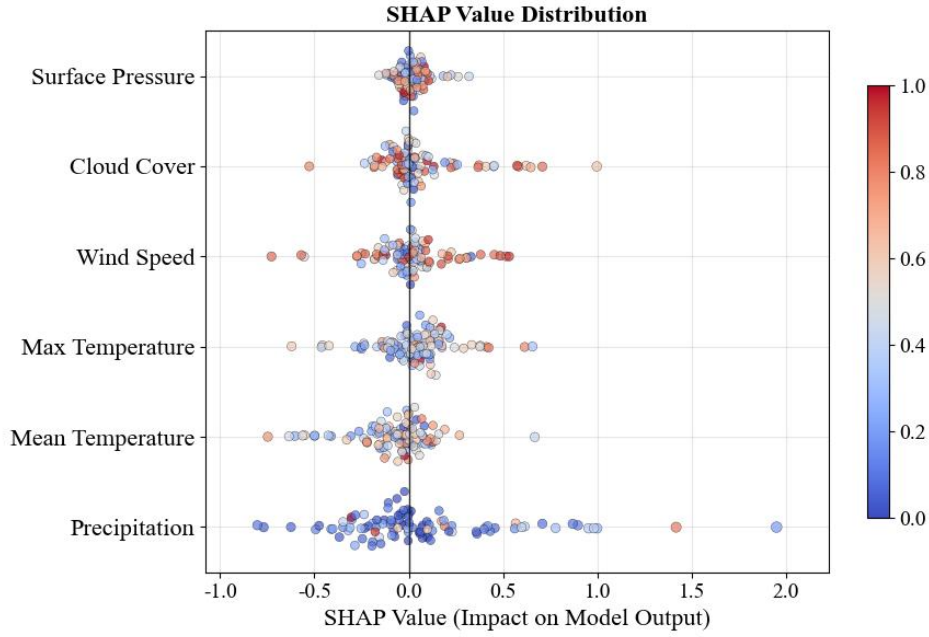
445 Figure 8 Taylor diagrams comparing the performance of different models for DEDI prediction in
 446 the Huaihe River Basin at different lead times: (a) 7 days; (b) 15 days; (c) 30 days; (d) 60 days; (e)
 447 120 days; (f) 180 days.

448 Fig. 8 presents Taylor diagrams that evaluate the predictive performance of various models
 449 across different lead times: (a) 7 days, (b) 15 days, (c) 30 days, (d) 60 days, (e) 120 days, and (f)

450 180 days. These diagrams provide a comprehensive assessment by integrating standard deviation
451 and correlation coefficients from the ERA5 reanalysis data. The VMD-informer-LSTM model
452 consistently shows superior predictive accuracy across all lead times, with its predictions closely
453 aligned with the ERA5 reanalysis data, particularly in the upper, middle, and lower regions of the
454 Huaihe River Basin and the Yi Shu Si River regions. At short lead times (7 and 15 days), all models
455 perform relatively well, but the VMD-informer-LSTM model slightly outperforms the others. As
456 the lead time extends to 30 and 60 days, the VMD-informer-LSTM model maintains high accuracy
457 while other models show performance degradation, indicating a reduced ability to capture the
458 underlying data patterns effectively. At long lead times (120 and 180 days), the VMD-informer-
459 LSTM model's advantage becomes more pronounced, with its predictions remaining significantly
460 closer to the ERA5 reanalysis data than other models, which exhibit more noticeable divergence.
461 This suggests that the VMD-informer-LSTM model is better equipped to handle the increased
462 uncertainty and complexity associated with long-term predictions, highlighting its robustness and
463 reliability in predictive tasks, especially crucial in drought forecasting where lead time is a critical
464 factor.

465 4.3 Model Influencing Factor Analysis

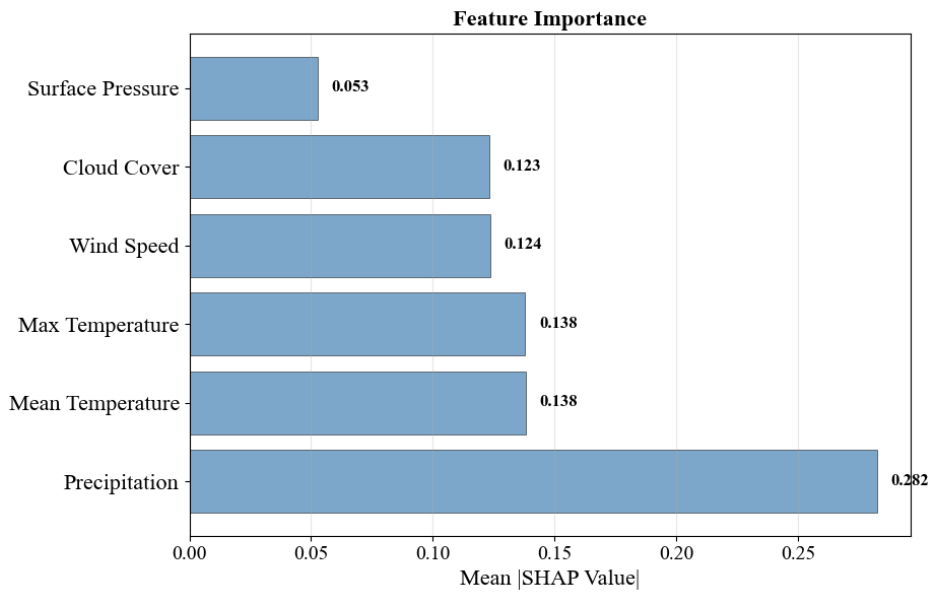
466 Fig. 9a presents a comprehensive SHAP value distribution analysis revealing the complex,
467 non-linear relationships between meteorological variables and drought prediction outcomes.
468 Precipitation emerges as the dominant predictor with predominantly positive SHAP contributions,
469 indicating its critical role in alleviating drought conditions through direct water supply augmentation.
470 However, the wide distribution of precipitation SHAP values (-1.00 to 2.00) suggests threshold-
471 dependent effects, where low precipitation events contribute negatively to drought mitigation while
472 high precipitation provides substantial positive contributions. Temperature exhibits a more complex
473 influence pattern, with SHAP values distributed across both positive and negative domains,
474 reflecting its dual role in drought dynamics: moderate temperatures may enhance vegetation water
475 use efficiency (positive contribution), while extreme temperatures intensify evapotranspiration
476 demands and soil moisture depletion (negative contribution). Fig. 9b quantifies that precipitation
477 and mean temperature collectively account for approximately 49.00% of the total model decision-
478 making process, with average SHAP magnitudes of 0.282 and 0.138 respectively. Surface pressure,
479 cloud cover, and maximum temperature demonstrate moderate but consistent influences (SHAP
480 magnitudes of 0.053 to 0.124), likely operating through indirect pathways affecting atmospheric
481 moisture transport, radiation balance, and boundary layer dynamics. The concentrated distribution
482 of these secondary variables' SHAP values suggests more linear, predictable relationships with
483 drought outcomes, contrasting with the high variability observed in precipitation and temperature
484 effects, which reflects the non-stationary nature of hydro-climatic processes and the model's
485 capacity to capture complex feature interactions across different drought severity conditions.



486

487

(a)



488

489

(b)

490 Figure 9 Relative contributions of meteorological variables to drought forecasting at a 180-day
491 lead time

492

493

494

495

496

497

498

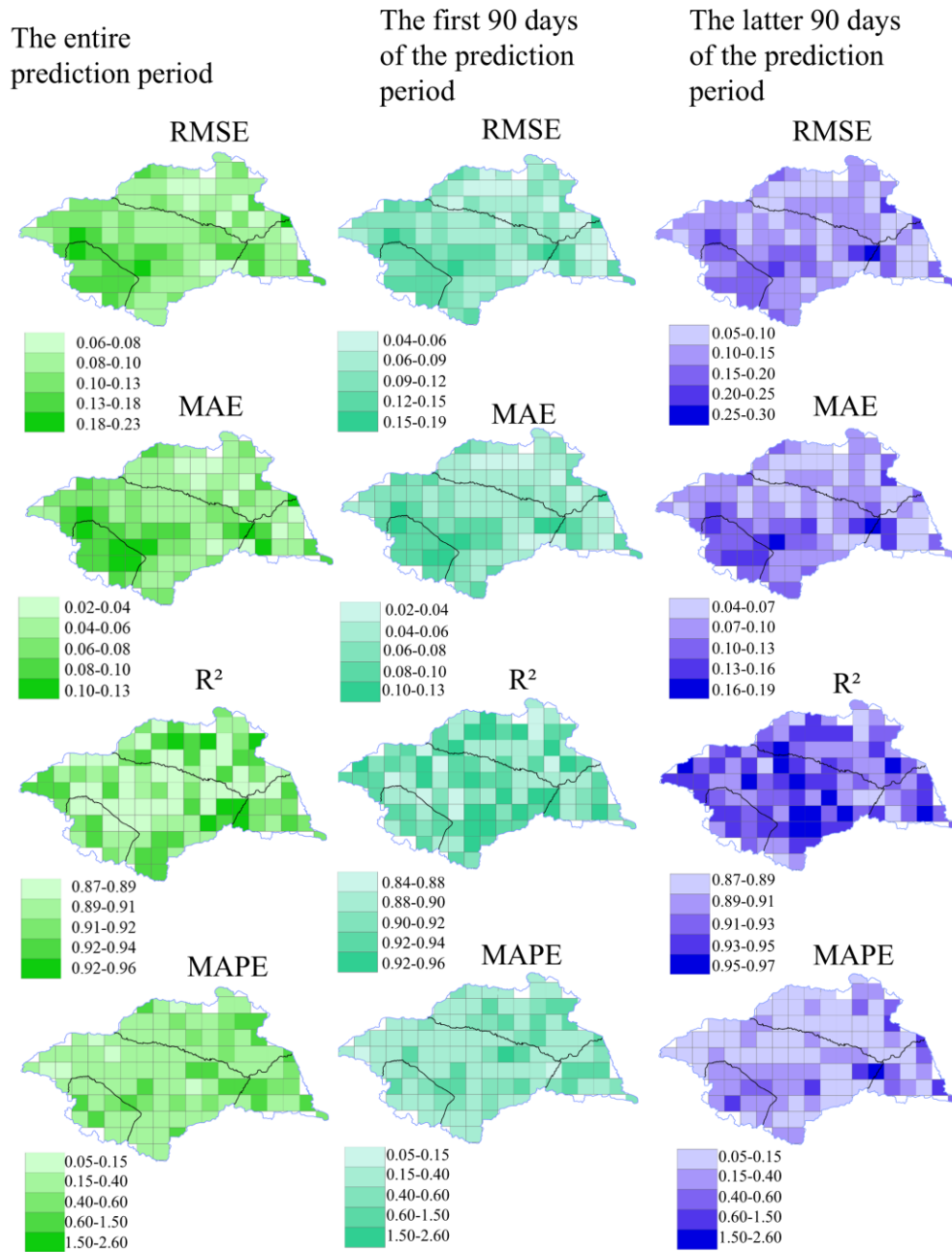
499

500

While this study demonstrates the superior performance of the VMD-informer-LSTM model through comprehensive comparative analysis, several aspects of model uncertainty warrant acknowledgement. The concentrated performance distributions observed in violin plots suggest relatively low model uncertainty across different geographical locations, with the VMD-informer-LSTM model showing the most stable performance (IQR: 0.05 for RMSE). However, systematic uncertainty quantification through techniques such as ensemble modeling, Monte Carlo dropout, or Bayesian approaches was not implemented in this study, representing a limitation that could be addressed in future research to provide confidence intervals for predictions and better understand prediction reliability under different hydroclimatic conditions.

501 The spatial analysis reveals that prediction accuracy varies across the basin, with slightly
502 higher uncertainties observed along catchment peripheries. This is potentially related to boundary
503 effects or data quality variations. Future work should incorporate explicit uncertainty quantification
504 methods to enhance the model's operational applicability and provide decision-makers with
505 confidence measures alongside drought predictions.

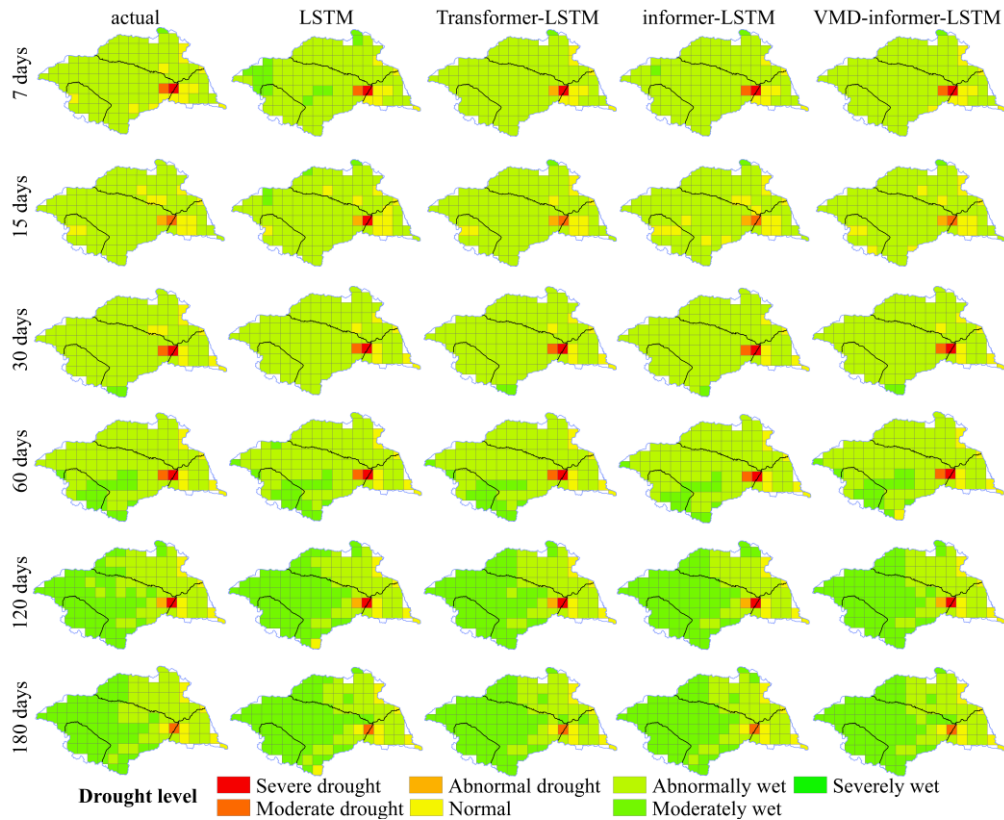
506 Fig. 10 illustrates the spatial prediction performance of the VMD-informer-LSTM model in
507 the Huaihe River Basin. Overall, the model achieves the highest accuracy in the first 90 days of
508 prediction, with an R^2 close to 0.92 and low error, effectively characterizing the evolution of drought.
509 As the prediction horizon extends beyond 90 days, the accuracy shows a noticeable decline,
510 reflecting the increasing difficulty of capturing long-term drought dynamics. In terms of spatial
511 distribution, the upstream area performs the best, accurately reproducing the observed drought
512 process. The downstream area follows, with the overall drought trend being tracked, though there
513 is a tendency of underestimation in drought intensity. The midstream area shows slightly larger
514 errors, but maintains high consistency with ERA5 reanalysis data on a long-term scale. The Yi Shu
515 Si River region shows a relatively balanced performance, with the model effectively reflecting
516 changes in drought levels.



517

518

Figure 10 The spatial distribution of VMD-informer-LSTM Model performance



519

520

Figure 11 Comparison of drought forecasting model performance at different time scales

521

522

Fig. 11 further reveals the characteristics of drought category distribution under different prediction horizons and reflects regional differences. In the short-term predictions (first 15 days), although some models exhibit prediction errors, the overall results in the upstream, midstream, downstream, and Yi Shu Si River regions remain broadly consistent with observations. At the 7-day horizon, the LSTM misclassifies 11 grid cells, identifying the abnormally wet category as moderately wet. At 15 days, the informer-LSTM shows errors in 4 grid cells, misclassifying abnormally wet conditions as normal.

527

528

529

In the medium-term predictions (30–60 days), the VMD-informer-LSTM demonstrates the best performance, with only 2 misclassified grid cells at both horizons, whereas the other three models show larger deviations, mainly concentrated in the upstream, midstream, and Yi Shu Si River regions. In the long-term predictions (120–180 days), prediction errors increase across all models; however, the VMD-informer-LSTM continues to maintain the highest overall accuracy. Across all prediction periods, models exhibit a common tendency to underestimate drought categories, with the VMD-informer-LSTM showing the smallest degree of underestimation.

532

533

534

535

Overall, it can be seen that the upstream and downstream regions achieved the highest observations consistency, followed by the Yi Shu Si River region, while the midstream region performed the weakest.

536

537

538

5. Discussion

539

540

541

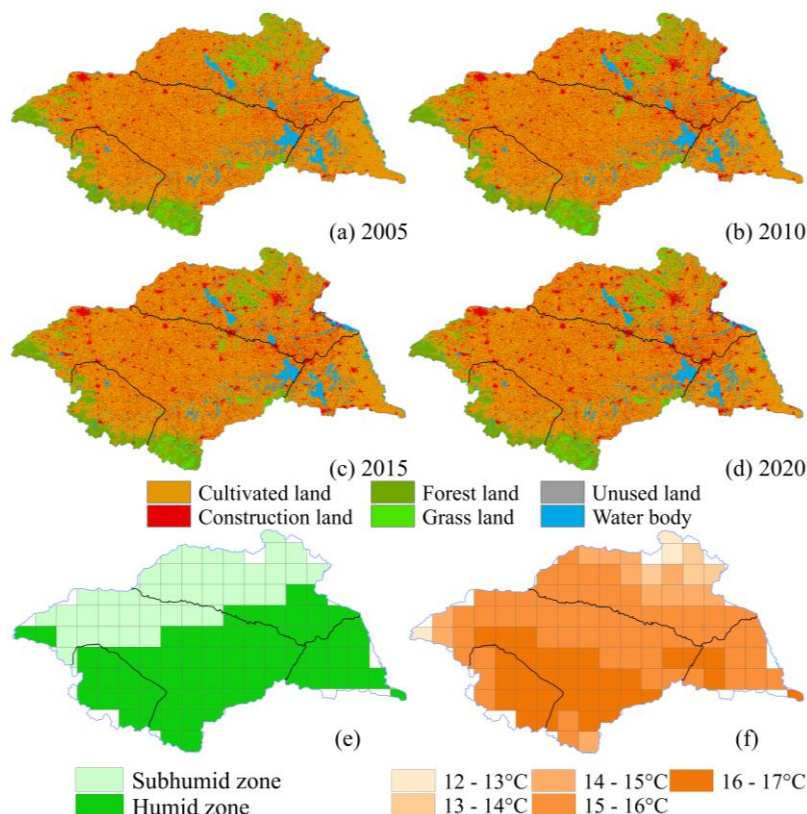
542

Further combining Table 2 and Fig. 12 provides a clearer understanding of the reasons for the differences in model prediction accuracy across different regions. From 2005 to 2020, the cultivated land in the basin continuously decreased (from 187,254 km² to 180,161 km²), the construction land significantly increased (from 34,739 km² to 42,801 km²), the grassland area markedly declined, and

543 the water area slightly increased. In the midstream region, the reduction of cultivated land, forest
 544 land, and grassland, coupled with the increase in construction land leading to the expansion of
 545 impervious surfaces, has altered the water cycle process. This has caused the model's prediction to
 546 underestimate drought severity, demonstrating that environmental changes have exacerbated
 547 hydrological droughts.

548 Table 2 Area of land use types in Huaihe River basin from 2005 to 2020(km²)

Year	Cultivated Land	Forest land	Grassland	Water Body	Building	Unused Land
2005	187254	16550	10149	13159	34739	213
2010	186078	16593	10117	13224	35841	212
2015	184597	16539	10049	13326	37326	231
2020	180161	16052	8083	14282	42801	168



549
 550 Figure 12 Distribution of Land Use, Climate Zones, and Multi-year Average Temperature in
 551 the Huaihe River Basin (a)–(d) Spatiotemporal distribution of land use in the Huaihe River Basin
 552 from 2005 to 2020; (e) Distribution map of semi-humid and humid zones in the Huaihe River
 553 Basin; (f) Distribution map of multi-year average temperature in the Huaihe River Basin

554 Meanwhile, Figure 12 shows distinct climatic zone characteristics: the upstream, southern
 555 midstream, downstream, and southern regions of the Yi Shu Si River are in the humid zone, where
 556 the annual mean temperature in the upstream and its adjacent southern midstream regions is close
 557 to 17°C, and the annual mean temperature in the downstream and its adjacent southern midstream
 558 regions as well as the southern regions of the Yi Shu Si River is between 13–16°C. In contrast, the
 559 northern midstream and northern regions of the Yi Shu Si River are in the semi-humid zone, with
 560 an annual mean temperature between 12°C and 15°C. Overall, the prediction results of several
 561 models underestimated hydrological drought, with misclassifications mainly concentrated in the
 562 transition zones between the humid and semi-humid regions. The semi-humid regions showed

563 relatively more frequent errors compared to the humid regions.

564 The results show that the VMD–Informer–LSTM model exhibits high prediction accuracy
565 within the 30–90 day forecast period at the time scale, while the prediction accuracy decreases
566 during the longer forecast period of 120–180 days. Spatially, the prediction results for the upstream
567 and downstream regions show the highest consistency with the ERA5 reanalysis values, followed
568 by the Yishuisi River region, with the midstream region showing relatively weaker performance. In
569 summary, the VMD–Informer–LSTM framework proposed in this study demonstrates significant
570 advantages in handling drought index series with prominent non-stationarity and multi-time scale
571 features, using a multi-scale modeling strategy of "decomposition—parallel modeling—feature
572 fusion." On the one hand, VMD effectively reduces the complexity of the original series, allowing
573 for the separation of variation features at different time scales and modeling them individually. On
574 the other hand, the parallel structure of Informer and LSTM focuses on capturing long-term
575 background state changes and short-term fluctuations, enabling the model to represent both the
576 persistence and phase-specific fluctuations of the drought process.

577 However, this study still has certain limitations. First, the model evaluation is primarily based
578 on ERA5 time-series data and lacks generalization validation using independent observations or
579 multi-source remote sensing data. Second, the model is constructed using one-dimensional time
580 series at individual grid points as input, without explicitly capturing the spatial propagation of
581 drought, thereby neglecting the spatial interconnections among different regions. Despite these
582 limitations, the proposed framework still demonstrates considerable potential in evaluating the
583 evolution of drought background conditions and can provide useful support for medium- to long-
584 term water resource regulation and risk assessment at the basin scale. Future work will incorporate
585 multi-source observational data and develop spatiotemporal coupling modeling approaches to
586 enhance the representation of drought evolution processes, thereby improving the reliability of
587 predictions.

588 6. Conclusion

589 This study takes the Huaihe River Basin in China as an example and constructs the VMD-
590 informer-LSTM model, and compares it with the LSTM, Transformer-LSTM, and informer-LSTM
591 models to verify its ability to predict hydrological drought in both temporal and spatial dimensions,
592 and influencing factors are quantitatively analyzed. The main conclusions are as follows:

593 (1) The VMD-informer-LSTM model shows clear advantages in drought forecasting across
594 different lead times and drought severity levels. Compared with the baseline LSTM, it improves R^2
595 by 28.4% ($0.7156 \rightarrow 0.9191$) and reduces RMSE by 46.2% ($0.2087 \rightarrow 0.1122$). Further
596 comparisons reveal that the single LSTM struggles to capture the complex drought process and
597 performs the poorest in all models' predictions.. Transformer-LSTM improves accuracy to some
598 extent, but there is still significant error accumulation in long-term predictions. The informer-LSTM,
599 leveraging the sparse self-attention mechanism and generative decoding approach of the informer,
600 can balance efficiency and accuracy in long time series prediction, showing greater stability than
601 Transformer-LSTM and being the best among the three benchmark models. However, the VMD-
602 informer-LSTM, by introducing VMD decomposition to further extract multi-scale features and
603 combining the long sequence modeling advantages of the informer with the local dynamic
604 characterization ability of LSTM, achieves multi-level information fusion. Therefore, the VMD-
605 informer-LSTM model achieves the highest precision in short-term forecasting (7 days), has the

606 smallest error growth rate in the medium term (30–60 days), and can still more effectively
607 characterize the evolution of drought in the long term (120–180 days). Its underestimation of
608 drought intensity is significantly lower than other models.

609 (2) The spatial prediction performance of the VMD informant LSTM model is influenced by
610 land use and climate regions. Within the entire watershed, all models showed excellent prediction
611 performance within a 30 day prediction period, with the VMD-informer-LSTM model being the
612 most accurate. However, within a prediction period of 120-180 days, the prediction accuracy of all
613 models significantly decreased throughout the watershed. Overall, the predicted drought intensity
614 was relatively mild, with misclassifications mainly concentrated in the transition zones between the
615 humid and semi-humid regions, and errors occurring more frequently in the semi-humid regions
616 compared to the humid regions. The model performs best in the upstream region, followed by the
617 downstream and Yishus River regions, while prediction accuracy in the midstream region is
618 relatively weak. With the extension of the forecast lead time, this downward trend is most evident
619 in the middle reaches, where the reduction of arable land, grassland, and forest land, as well as the
620 expansion of construction land, have changed the water cycle process, indicating that human
621 activities have exacerbated drought.

622 (3) The SHAP analysis enhances the interpretability of the VMD-informer-LSTM model by
623 revealing the relative importance of meteorological variables in drought prediction. Precipitation is
624 the dominant factor, contributing about 28.2% to model decisions, followed by mean temperature
625 (13.8%), while surface pressure, cloud cover, and maximum temperature together account for about
626 15–20%. These results confirm that the model effectively identifies key meteorological drivers and
627 enhances the interpretability of drought forecasting.

628 Author Contributions:

629 Conceptualization: Min Li, Ming Ou, Yuhang Yao, Changman Yin

630 Data curation: Ming Ou

631 Formal analysis: Min Li, Ming Ou

632 Funding acquisition: Min Li

633 Investigation: Min Li, Ming Ou

634 Methodology: Min Li, Ming Ou

635 Software: Min Li

636 Supervision: Ming Ou, Yuhang Yao, Changman Yin

637 Validation: Min Li

638 Visualization: Ming Ou, Yuhang Yao, Changman Yin

639 Writing – original draft: Min Li, Ming Ou

640 Writing – review & editing: Yuhang Yao, Changman Yin

641 Declaration of competing interest

642 The authors declare no conflicts of interest relevant to this study.

643 Data Availability Statement

644 The gridded daily precipitation, evaporation, potential evaporation, mean temperature,
645 maximum temperature, cloud cover, surface pressure, and wind speed data with a spatial r
646 esolution of 0.25° were obtained from ERA5 post-processed daily statistics on single level

647 s from 1940 to present (<https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics?tab=overview>), covering the period from January 1, 1984, to December 31, 20
648 24. Concurrently, the land use data were sourced from the Chinese Academy of Sciences
649 Resource and Environment Science Data Center (<https://www.resdc.cn/Default.aspx>). The core
650 implementation of the VMD–Informer–LSTM model used in this study is publicly available
651 on GitHub at: <https://github.com/OUman648/vmd-informer-LSTM>.
652

653 Acknowledgements

654 This work was financially Supported by Open Research Fund Program of National Key
655 Laboratory of Water Disaster Prevention, 2024490711, Natural Science Foundation of Jiangsu
656 Province, BK20250906.

657 References

- 658 AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., &
659 Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities.
660 *Reviews of Geophysics*, 53(2), 452–480. <https://doi.org/10.1002/2014RG000456>
- 661 Alsubih, M., Mallick, J., Talukdar, S., Salam, R., AlQadhi, S., Fattah, Md. A., & Thanh, N. V. (2021).
662 An investigation of the short-term meteorological drought variability over Asir Region of Saudi
663 Arabia. *Theoretical and Applied Climatology*, 145(1–2), 597–617.
664 <https://doi.org/10.1007/s00704-021-03647-4>
- 665 Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought
666 forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet
667 support vector regression models. *Journal of Hydrology*, 508, 418–429.
668 <https://doi.org/10.1016/j.jhydrol.2013.10.052>
- 669 Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New
670 Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–
671 1828. <https://doi.org/10.1109/TPAMI.2013.50>
- 672 Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting*
673 *and control*. John Wiley & Sons.
- 674 Cook, B. I., Mankin, J. S., Marvel, K., Williams, A. P., Smerdon, J. E., & Anchukaitis, K. J. (2020).
675 Twenty-First Century Drought Projections in the CMIP6 Forcing Scenarios. *Earth's Future*,
676 8(6), e2019EF001461. <https://doi.org/10.1029/2019EF001461>
- 677 Dai, A. (2013). Increasing drought under global warming in observations and models. *Nature*
678 *Climate Change*, 3(1), 52–58. <https://doi.org/10.1038/nclimate1633>
- 679 Dragomiretskiy, K., & Zosso, D. (2014). Variational Mode Decomposition. *IEEE Transactions on*
680 *Signal Processing*, 62(3), 531–544. <https://doi.org/10.1109/TSP.2013.2288675>
- 681 Dutra, E., Wetterhall, F., Di Giuseppe, F., Naumann, G., Barbosa, P., Vogt, J., Pozzi, W., &
682 Pappenberger, F. (2014). Global meteorological drought – Part 1: Probabilistic monitoring.
683 *Hydrology and Earth System Sciences*, 18(7), 2657–2667. <https://doi.org/10.5194/hess-18-2657-2014>
- 684
- 685 Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J.
686 D. (2003). Implementation of Noah land surface model advances in the National Centers for
687 Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research:*
688 *Atmospheres*, 108(D22), 2002JD003296. <https://doi.org/10.1029/2002JD003296>

689 Gao, C., Zhang, Z., Zhai, J., Qing, L., & Mengting, Y. (2015). Research on meteorological
690 thresholds of drought and flood disaster: A case study in the Huai River Basin, China.
691 *Stochastic Environmental Research and Risk Assessment*, 29(1), 157–167.
692 <https://doi.org/10.1007/s00477-014-0951-y>

693 Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A
694 Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10),
695 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>

696 Hao, Z., Hao, F., Singh, V. P., Ouyang, W., & Cheng, H. (2017). An integrated package for drought
697 monitoring, prediction and analysis to aid drought modeling and assessment. *Environmental*
698 *Modelling & Software*, 91, 199–209. <https://doi.org/10.1016/j.envsoft.2017.02.008>

699 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J.,
700 Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo,
701 G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J. (2020). The ERA5 global
702 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
703 <https://doi.org/10.1002/qj.3803>

704 Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8),
705 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

706 Huang, S., Huang, Q., Chang, J., Zhu, Y., Leng, G., & Xing, L. (2015). Drought structure based on
707 a nonparametric multivariate standardized drought index across the Yellow River basin, China.
708 *Journal of Hydrology*, 530, 127–136. <https://doi.org/10.1016/j.jhydrol.2015.09.042>

709 Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette,
710 A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler,
711 A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., ... Yiou, P. (2014). EURO-CORDEX:
712 New high-resolution climate change projections for European impact research. *Regional*
713 *Environmental Change*, 14(2), 563–578. <https://doi.org/10.1007/s10113-013-0499-2>

714 Jaseena, K. U., & Kovoov, B. C. (2022). Deterministic weather forecasting models based on
715 intelligent predictors: A survey. *Journal of King Saud University - Computer and Information*
716 *Sciences*, 34(6, Part B), 3393–3412. <https://doi.org/10.1016/j.jksuci.2020.09.009>

717 Johny, K., Pai, M. L., & S., A. (2022). A multivariate EMD-LSTM model aided with Time
718 Dependent Intrinsic Cross-Correlation for monthly rainfall prediction. *Applied Soft Computing*,
719 123, 108941. <https://doi.org/10.1016/j.asoc.2022.108941>

720 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling
721 using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*,
722 22(11), 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>

723 Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J.,
724 Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., & Slater, A. G. (2011).
725 Parameterization improvements and functional and structural advances in Version 4 of the
726 Community Land Model: PARAMETERIZATION IMPROVEMENTS AND FUNCTIONAL
727 AND STRUCTURAL ADVANCES. *Journal of Advances in Modeling Earth Systems*, 3(1),
728 n/a-n/a. <https://doi.org/10.1029/2011MS00045>

729 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
730 <https://doi.org/10.1038/nature14539>

731 Li, Z., Wu, H., Duan, S., Zhao, W., Ren, H., Liu, X., Leng, P., Tang, R., Ye, X., Zhu, J., Sun, Y., Si,
732 M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B., Yan, G., & Zhou, C. (2023). Satellite

733 Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and
734 Applications. *Reviews of Geophysics*, 61(1), e2022RG000777.
735 <https://doi.org/10.1029/2022RG000777>

736 Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*
737 (arXiv:1705.07874; Version 2). arXiv. <https://doi.org/10.48550/arXiv.1705.07874>

738 Mishra, A. K., & Desai, V. R. (2005). Drought forecasting using stochastic models. *Stochastic*
739 *Environmental Research and Risk Assessment*, 19(5), 326–339.
740 <https://doi.org/10.1007/s00477-005-0238-4>

741 Mo, K. C. (2008). Model-Based Drought Indices over the United States. *Journal of*
742 *Hydrometeorology*, 9(6), 1212–1230. <https://doi.org/10.1175/2008JHM1002.1>

743 Modarres, R. (2007). Streamflow drought time series forecasting. *Stochastic Environmental*
744 *Research and Risk Assessment*, 21(3), 223–233. <https://doi.org/10.1007/s00477-006-0058-1>

745 Morid, S., Smakhtin, V., & Moghaddasi, M. (2006). Comparison of seven meteorological indices
746 for drought monitoring in Iran. *International Journal of Climatology*, 26(7), 971–985.
747 <https://doi.org/10.1002/joc.1264>

748 Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models:
749 Literature Review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>

750 Mossad, A., & Alazba, A. (2015). Drought Forecasting Using Stochastic Models in a Hyper-Arid
751 Climate. *Atmosphere*, 6(4), 410–430. <https://doi.org/10.3390/atmos6040410>

752 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta,
753 S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-
754 Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-
755 the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–
756 4383. <https://doi.org/10.5194/essd-13-4349-2021>

757 Myttenaere, A. D., Golden, B., Grand, B. L., & Rossi, F. (2016). Mean Absolute Percentage Error
758 for regression models. *Neurocomputing*, 192, 38–48.
759 <https://doi.org/10.1016/j.neucom.2015.12.114>

760 Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A
761 discussion of principles. *Journal of Hydrology*, 10(3), 282–290.

762 Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., Heim, R. R., Brewer, M.
763 J., Svoboda, M., Westerhoff, R., Van Dijk, A. I. J. M., Lloyd-Hughes, B., Pappenberger, F.,
764 Werner, M., Dutra, E., Wetterhall, F., Wagner, W., Schubert, S., Mo, K., ... Lawford, R. (2013).
765 Toward Global Drought Early Warning Capability: Expanding International Cooperation for
766 the Development of a Framework for Monitoring and Forecasting. *Bulletin of the American*
767 *Meteorological Society*, 94(6), 776–785. <https://doi.org/10.1175/BAMS-D-11-00176.1>

768 Rummukainen, M. (2010). State-of-the-art with regional climate models. *WIREs Climate Change*,
769 1(1), 82–96. <https://doi.org/10.1002/wcc.8>

770 Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.,
771 Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang,
772 W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal of*
773 *Climate*, 27(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>

774 Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–
775 117. <https://doi.org/10.1016/j.neunet.2014.09.003>

776 Shlezinger, N., Whang, J., Eldar, Y. C., & Dimakis, A. G. (2023). Model-Based Deep Learning.

777 *Proceedings of the IEEE*, 111(5), 465–499. <https://doi.org/10.1109/JPROC.2023.3247480>

778 Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive
779 review of deep learning applications in hydrology and water resources. *Water Science and*
780 *Technology*, 82(12), 2635–2670. <https://doi.org/10.2166/wst.2020.369>

781 Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R.,
782 Palecki, M., Stooksbury, D., Miskus, D., & Stephens, S. (2002). THE DROUGHT MONITOR.
783 *Bulletin of the American Meteorological Society*, 83(8), 1181–1190.
784 <https://doi.org/10.1175/1520-0477-83.8.1181>

785 Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield,
786 J. (2014). Global warming and changes in drought. *Nature Climate Change*, 4(1), 17–22.
787 <https://doi.org/10.1038/nclimate2067>

788 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., &
789 Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing*
790 *Systems*, 30(1), 5998–6008.

791 Vicente-Serrano, S. M., Miralles, D. G., Domínguez-Castro, F., Azorin-Molina, C., El Kenawy, A.,
792 McVicar, T. R., Tomás-Burguera, M., Beguería, S., Maneta, M., & Peña-Gallardo, M. (2018).
793 Global Assessment of the Standardized Evapotranspiration Deficit Index (SEDI) for Drought
794 Analysis and Monitoring. *Journal of Climate*, 31(14), 5371–5393.
795 <https://doi.org/10.1175/JCLI-D-17-0775.1>

796 Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: A
797 comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*,
798 11(1), 44. <https://doi.org/10.1186/s40537-024-00905-w>

799 Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root
800 mean square error (RMSE) in assessing average model performance. *Climate Research*, 30,
801 79–82. <https://doi.org/10.3354/cr030079>

802 Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., & Clark, M. (2016). Quantifying
803 Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill. *Journal*
804 *of Hydrometeorology*, 17(2), 651–668. <https://doi.org/10.1175/JHM-D-14-0213.1>

805 Yao, T., Zhao, Q., Wu, C., Hu, X., Xia, C., Wang, X., Sang, G., Liu, J., & Wang, H. (2024). Spatio-
806 temporal Variation Characteristics of Extreme Climate Events and Their Teleconnections to
807 Large-scale Ocean-atmospheric Circulation Patterns in Huaihe River Basin, China During
808 1959–2019. *Chinese Geographical Science*, 34(1), 118–134. [https://doi.org/10.1007/s11769-](https://doi.org/10.1007/s11769-023-1398-1)
809 023-1398-1

810 Yuan, S., & Quiring, S. M. (2017). Evaluation of soil moisture in CMIP5 simulations over the
811 contiguous United States using in situ and satellite observations. *Hydrology and Earth System*
812 *Sciences*, 21(4), 2203–2218. <https://doi.org/10.5194/hess-21-2203-2017>

813 Zhang, J., Xin, X., Shang, Y., Wang, Y., & Zhang, L. (2023). Nonstationary significant wave height
814 forecasting with a hybrid VMD-CNN model. *Ocean Engineering*, 285, 115338.
815 <https://doi.org/10.1016/j.oceaneng.2023.115338>

816 Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., & Wei, M. (2019). A Review on Deep Learning
817 Applications in Prognostics and Health Management. *IEEE Access*, 7, 162415–162438.
818 <https://doi.org/10.1109/ACCESS.2019.2950985>

819 Zhang, Q., Kong, D., Shi, P., Singh, V. P., & Sun, P. (2018). Vegetation phenology on the Qinghai-
820 Tibetan Plateau and its response to climate change (1982–2013). *Agricultural and Forest*

821 *Meteorology*, 248, 408–417. <https://doi.org/10.1016/j.agrformet.2017.10.026>

822 Zhang, Q., Zhang, J., Yan, D., & Wang, Y. (2014). Extreme precipitation events identified using
823 detrended fluctuation analysis (DFA) in Anhui, China. *Theoretical and Applied Climatology*,
824 117(1–2), 169–174. <https://doi.org/10.1007/s00704-013-0986-x>

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

828 Zhao, L., Li, Z., Qu, L., Zhang, J., & Teng, B. (2023). A hybrid VMD-LSTM/GRU model to predict
829 non-stationary and irregular waves on the east coast of China. *Ocean Engineering*, 276, 114136.
830 <https://doi.org/10.1016/j.oceaneng.2023.114136>

831 Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021a). Informer: Beyond
832 Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI*
833 *Conference on Artificial Intelligence*, 35(12), 11106–11115.
834 <https://doi.org/10.1609/aaai.v35i12.17325>

835 Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021b). Informer: Beyond
836 Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI*
837 *Conference on Artificial Intelligence*, 35(12), 11106–11115.
838 <https://doi.org/10.1609/aaai.v35i12.17325>

839 Zuo, G., Luo, J., Wang, N., Lian, Y., & He, X. (2020). Decomposition ensemble model based on
840 variational mode decomposition and long short-term memory for streamflow forecasting.
841 *Journal of Hydrology*, 585, 124776. <https://doi.org/10.1016/j.jhydrol.2020.124776>

842