Revisiting the Efficacy of Signal Decomposition in AI-based Time Series Prediction

Dan Zhang [1, a], Kexin Jiang [2, b], Chuhan Wu[3,c], Yan Peng [1, d], Zhaolong Han[4,e], and Yaoran Chen[1, f,*1]

[1] Institute of Artificial Intelligence, Shanghai University, Shanghai, 200444, China;
[2] Department of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China;
[3] Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China;
[4] School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.

ABSTRACT

Time series prediction is a fundamental problem in scientific exploration and artificial intelligence (AI) technologies have substantially bolstered its efficiency and accuracy. A well-established paradigm in AI-driven time series prediction is injecting physical knowledge into neural networks through signal decomposition methods, and sustaining progress in numerous scenarios has been reported. However, we uncover non-negligible evidence that challenges the effectiveness of signal decomposition in AI-based time series prediction. We confirm that improper dataset processing with subtle future label leakage is unfortunately widely adopted, possibly yielding abnormally superior but misleading results. By processing data in a strictly causal way without any future information, the effectiveness of additional decomposed signals diminishes. Our work probably identifies an ingrained and universal error in time series modeling, and the de facto progress in relevant areas is expected to be revisited and calibrated to prevent future scientific detours and minimize practical losses.

Keywords: signal processing, time series prediction, label leakage, feature engineering, re-investigation.

1. INTRODUCTION

Time series, a fundamental data form composed of successive sequences of data points, is critical in many time-sensitive areas such as meteorology [1], economics [2], environics [3] and sociology [4]. Predicting the future trends of time series based on past observations enables researchers to understand the future patterns of natural and societal signals at different time scales, ranging from long-term sunspot activities and flares [5], hour-level tropical cyclone trajectories [6], and near-real-time fluctuations of volatile renewable energy like solar [7] and wind [8]. Thus, accurate time series prediction has tremendous potential for various research and industrial applications, such as decision-making [9], resource allocation [10], business planning [11], and risk management [12]

In the realm of physical-related areas such as oceanography and meteorology, canonical time series prediction models rely on numerical methods derived from physical modeling [13, 14]. Although these methods have been extensively used and improved over the decades [15, 16], the heavy computational costs hinder their applications in large-scale and real-time systems, and contradict the trend of green computing. To address this challenge, artificial intelligence (AI)-based methods have emerged as an alternative approach for predicting natural time series, offering the advantages of high efficiency, flexibility, and scalability [17]. With the rise of deep learning, natural time series prediction models based on deep neural networks have garnered widespread attention and achieved impressive results [18]. This technology trend has impacted a broader interdisciplinary community that involves data-driven time series analytics [19].

Deep learning time series models have demonstrated their effectiveness in discovering rich and complex patterns hidden in big data [20, 21]. However, these models are often criticized for their lack of prior knowledge and awareness of physical laws, resulting in non-robust predictions [22, 23]. To handle this issue, researchers have proposed to explicitly inject manually crafted information into AI models to empower them with accumulated domain knowledge [24, 25]. A representative example is time-frequency decomposition, which can help AI models deal with highly volatile and noisy

[1a] dan.zhang@shu.edu.cn, [b] jiangkexin98@163.com, [c] wuchuhan15@gmail.com, [d] pengyan@shu.edu.cn, [e] han.arkey@sjtu.edu.cn, [*f] ychen169@shu.edu.cn

The International Conference Optoelectronic Information and Optical Engineering (OIOE2024), edited by Yang Yue, Lu Leng, Proc. of SPIE Vol. 13513, 135130M · © 2025 SPIE · 0277-786X · © The Authors.

time series data [26, 27]. Through learning on stabilized decomposition signals, models can better capture inherent patterns that match physical phenomena [28, 29]. Consequently, a considerable amount of work along this research line has been extensively conducted in various areas [30], with encouraging and even striking results reported in many downstream applications [31, 32, 33].

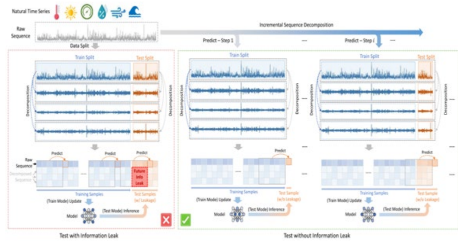

Figure. 1 The data leakage problem of leveraging signal decomposition in time series prediction. A widely employed method is performing sequence decomposition on the entire sequence, and then dividing the decomposed sequences into training and test splits. When training on the decomposed sequences, the information of the test set is leaked since most sequence decomposition methods are not strictly causal. Future information leakage usually causes considerable overestimation of model prediction accuracy. A correct method for utilizing sequence decomposition techniques without data leakage is using the sequences before the time of each test data point to compute decomposed subsequences for both training and test. In this way, the model is completely agnostic to future information.

While the direction being pursued is plausible, our systematic investigation suggests it may be misguided. Our findings uncover a frustrating issue that the effectiveness of signal decomposition methods is severely overestimated, probably due to the inconspicuous information leakage introduced by improper data processing practices that conduct signal decomposition on both past and future data (Fig. 1). Given the powerful function-fitting ability of deep learning methods, even subtle and implicit leakage of future information can lead to our hallucination of prediction accuracy enhancement. Unfortunately, the inability to use leaked information in practical contexts may yield severe model performance degradation and introduce unrecognized risks to scientific research and engineering endeavors. If not corrected promptly, such a course could lead to potential misleads and resource waste in both academia and industry, as the painful lesson learned from "cardiac stem cells" [34]. Our revisiting is expected to help calibrate researchers' estimation of the performance of signal decomposition-enhanced time series forecasting AIs, meanwhile providing valuable guidance and insight for engineering trouble-shooting, technological evolution, and scientific progress. Our work is also intended to inspire the research community to scrutinize certain well-established technological paradigms and proactively identify those neglected pitfalls.

2. RESULT

We conduct extensive experiments on six datasets in different domains, including significant wave height, wind speed, humidity, solar power, air pressure, and temperature (see details in Supplementary Methods). We select three representative signal decomposition methods, i.e., Empirical Mode Decomposition (EMD) [35], Discrete Wavelet Transform (DWT) [36], and Singular Spectrum Analysis (SSA) [37]. We use two different data processing procedures: (1) decomposition on all time steps, which introduces concealed future information leakage, and (2) decomposition on strictly restricted causal decomposition, which can only utilize past observations.

We test the results of the naive persistence baseline, the vanilla LSTM model, and LSTMs combined with leaked or non-leaked decomposed sequences (Fig. 2). On all six datasets, the vanilla LSTM consistently achieves lower errors than the persistence strategy, showing its expected predictability. The prediction errors decrease at an abnormal scale when incorporating decomposed signals with inherent future label leakage, while the errors are calibrated when using the non-leaked data. For all methods on all datasets, a slight performance degradation is observed compared with the vanilla LSTM when we incorporate decomposed sequences without a glance at future data points. SSA shows the most serious leakage problem among different decomposition methods, i.e., the prediction errors reduce up to more than 90% on all datasets.

These phenomena are unexpected but reasonable, since the sequences decomposed on the entire series encode rich clues about future trends, such as forthcoming peaks and valleys. Thus, the models learned and tested on sequences with potential data leakage usually show superior but inauthentic results. Moreover, the effectiveness of signal decomposition in natural time series prediction is challenged by the results, since it fails to boost the performance of the basic LSTM model. The above findings show some possible contradictions to some prior studies [38, 39, 40]. There is a small but non-negligible possibility that, this data leakage issue exists in a broad range of research in the natural sequence area, causing an over-estimation of the effectiveness of signal decomposition in predictive problems.

We also reproduce this worrying issue in different time series prediction models. By comparing the vanilla MLP, LSTM, and Transformer models and their variants combined with decomposed sequences (Fig. 3), we verify the universal impact of data leakage on all compared models. Although these models have different architectures, model parameter capacities, and computation processes, the implicit leakage of future data can be consistently captured by them. Even the time-agnostic MLP model can achieve rather low errors when the leak occurs. This result reveals that, a large family of machine learning models can be influenced by the improper data processing method.

To further analyze the impact on the prediction behaviors of time series prediction models, we visualize the real observations and the prediction curves on different datasets (Fig. 4). We use LSTM as the basic model and EMD as the decomposition method. The prediction curves with data leakage closely match the trends of real observations, which is consistent with the quantitative analysis above. However, for the model tested on non-leaked data, the predictions have lower sensitivities to frequent or drastic fluctuations, as shown in subplots A, B and D. This probably indicates a reason for the ineffectiveness of sequence decomposition in these cases, i.e., the frequency domain signals extracted from the past observations may enforce the model to follow past trends rather than discover upcoming patterns. The overfitting problem of these frequency signals makes models less sensitive to the quick changes in the time domain, since these fluctuations correspond to outliers in the frequency domain that are missing in past training data. Thus, the real effectiveness of signal decomposition in time series prediction is not as salient as we expect.

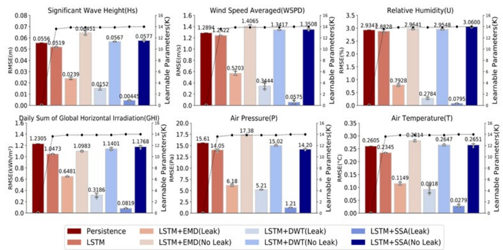

Figure. 2 Influence of potential data leakage in signal decomposition on prediction performance. Error bars are standard deviations of five repeated experiments. Comprehensive analysis is conducted on six datasets with different physical time series and different signal decomposition methods, including Empirical Mode Decomposition (EMD) [35], Discrete Wavelet Transform (DWT) [36], and Singular Spectrum Analysis (SSA) [37]. When introducing decomposed time series with future information encoded, the prediction errors consistently decrease at an unprecedented scale (e.g., 91.4% for SSA on the significant wave height dataset). The actual debiased performance with a strictly causal decomposition process substantially degrades, indicating rather limited and even negative impacts of decomposed signals on model predictability (the performance differences between leaked and non-leaked versions are significant in two-sided t-test, $p < 0.001$). Since the number of learnable parameters remains similar, it is highly suspected that data leakage may lead to a hallucination of the effectiveness of decomposed sequences.

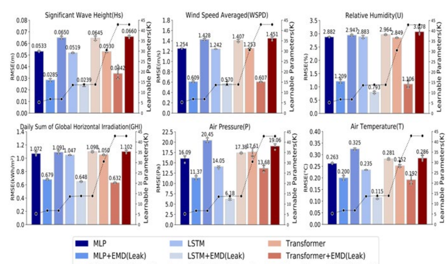

Figure. 3 The impact of data leakage on various time series prediction models. Error bars are standard deviations of five repeated experiments. We test three canonical neural network architectures, i.e., multi-layer perceptrons (MLP), long short-term memory (LSTM) [41] network, and Transformer [42], assisted by the represented signal decomposition method EMD. Although these models are diverse in their characteristics and model parameter sizes, consistent mismatches between models learned on leaked and non-leaked data are observed ($p < 0.001$ in two-sided t-test). It indicates a universal problem in the data processing procedure of signal decomposition-based time series prediction research.

Finally, we evaluate the impact of different components in the family of decomposed sequences. We take EMD as an exemplary method for analysis. We add one of the different components into the LSTM model and compare its performance change (Fig. 5). We uncover that the abnormal error decrease is majorly brought by the high-frequency components, and the first sequence has the most salient contribution. It implies that the model mainly overfits the fluctuation patterns encoded by high-frequency components, which have direct correlations with the actual values of future data observations. Based on this finding, we can explain the phenomenon that SSA leads to greater leakage impacts than EMD (Figure 2), since SSA draws the global picture of signals in the frequency domain while EMD only focuses on the patterns of local minima and maxima. This may further indicate that the improper data processing method may cast larger impacts on signal decomposition methods with heavier frequency information dependencies.

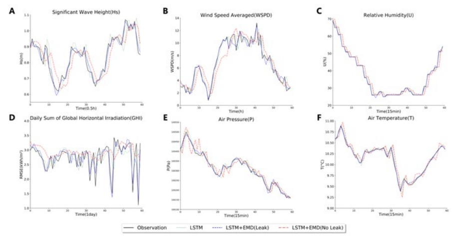

Figure. 4 The groundtruth curves and predicted curves of a randomly selected sample on different datasets. We compare the LSTM model with raw sequences only and EMD sequences with or without leakage. We observe different prediction patterns between leaked and non-leaked versions. For samples with frequent or drastic fluctuations (e.g., subplot A, B, and D), the non-leaked EMD version shows lower sensitivity and higher biases compared with the original LSTM model. It reveals a phenomenon that introducing extra frequency domain signals may weaken the model's capability of modeling strong fluctuations in the time domain, possibly due to the overfitting of past frequency domain data.

3. DISCUSSION

Time series prediction is a pivotal problem across various scientific research domains. With the accumulated studies and verifications from different scenarios, it seems that a consensus has been achieved that signal decomposition techniques are critical complementary to prediction models in terms of embedding physical-world knowledge. Unfortunately, after thorough analysis and scrutinization, we uncover an unexpected matter that the effectiveness of signal decomposition is exaggerated due to an inconspicuous type of data leakage. We confirm that, decomposing the time series on the entire data including test splits introduces considerable future information leakage, even if we incorporate the decomposed data points before the test time only. Such leakage causes a universal and dramatic performance overestimation of various models in various scenarios. We also uncover a phenomenon that high-frequency patterns encoded with future information majorly account for the label leakage, since they are probably more informative for forecasting short-term future trends. Our work whistleblows an alarming issue in the time series prediction area that a well-established practice seems to be questionable and needs comprehensive re-examination.

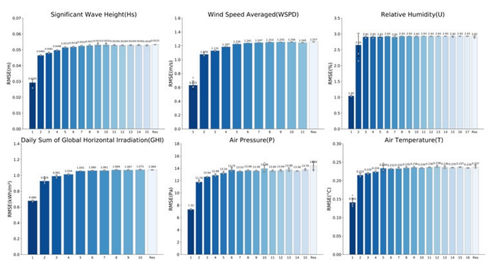

Figure. 5 Impact of residual and different Intrinsic Mode Functions (IMFs) on information leakage. Error bars are standard deviations of five repeated experiments. We add one of the IMFs in EMD or the residual part into the LSTM model in each experiment. Sequences with smaller indexes contain higher frequency components. The results show that the first component with the highest frequency has the most salient impact on the abnormal performance change. A few high-frequency components show major effects and the rest low-frequency ones do not significantly contribute to the label leakage. It shows the subtle local trends of time series play more important roles in time series prediction and the leakage of such clues causes more misleading conclusions.

The act of revisiting scientific findings holds immense value and significance as time series prediction is a fundamental aspect of many fields. Unfortunately, unintended mistakes have led to scientific errors, missed opportunities, and economic and social losses. This highlights the need to improve the systematic and reproducible nature of scientific research and encourage healthy and sustainable development of the scientific research ecology and the linkage between science and society. We aim to use our work as a flag in the process of accumulating scientific knowledge to champion these ideals and address problems hidden in other fields.

Our study has several limitations. Firstly, we cannot ensure that we have exhaustively reviewed all relevant literature and available approaches to investigate the existing issues. Secondly, there are diverse decomposition methods that can result in varying effects when applied in practice. We have only compared them based on typical and mainstream methods, and thus, other scenarios may exist. Our retrospective study aims to remedy past errors, and we cannot directly identify and correct existing flaws in actual systems. Therefore, it is necessary for researchers and technical experts to collaborate to minimize the impact of existing and potential issues.

4. METHODS

Here we introduce the detailed experimental protocol of our analysis (Figure 1). Given a time series prediction dataset D with M independent samples, we divide each sample into two parts chronologically for training and test. For each sample $x \in \mathrm{D}$, we denote the sub-sequence in the training and test sets as $x_{\mathrm{t}} = [\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_{\mathrm{P}}]$ and $x_{\mathrm{v}} = [\mathrm{x}_{\mathrm{P}+1}, \mathrm{x}_{\mathrm{P}+2}, \dots, \mathrm{x}_{\mathrm{Q}}]$, respectively, where P and $\mathrm{Q}-\mathrm{P}$ are their sequence lengths. We then use signal decomposition methods such as EMD, DWT, and SSA to obtain the decomposed sequences of the training sequence x_{t}, which are denoted as $\mathrm{X}_{\mathrm{d}} = [x_{\mathrm{t}}^1, x_{\mathrm{t}}^2, \dots, x_{\mathrm{t}}^{\mathrm{K}}]$, where K is the number of decomposed components. In a typical method [33] for incorporating decomposed signals, the raw sequence x_{t} and the series of decomposed sequences X_{d} are concatenated together to form the processed data X_{p}.

The time series prediction model receives the above sequences as the input. It can be implemented by various architectures, such as traditional machine learning methods (e.g., SVM [43]) and deep learning models like MLP [44], CNN [45], LSTM [46], and Transformer [47]. Denote the prediction window of this model as W, then the model conducts prediction on X_{p} in a sequential way as follows:

$$\begin{aligned} \hat{x}_{W+1} &= f\big(X_p[1{:}W];\theta\big) \\ &\dots \\ \hat{x}_{P-1} &= f\big(X_p[P-1-W{:}P-2];\theta\big), (1) \\ \hat{x}_{P} &= f\big(X_p[P-W{:}P-1];\theta\big), \end{aligned}$$

where $\mathrm{f}(\cdot;\theta)$ represents the non-linear mapping function learned by the model with parameters θ, $[\mathrm{i}{:}\mathrm{j}]$ stands for the slicing operation from the i-th to the j-th steps, and $\hat{\mathrm{x}}_{\mathrm{i}}$ means the predicted value at the i-th step. By comparing the predicted values with the real observations, we then compute the loss function $\mathcal{L}$. Taking the commonly used mean squared error as an example, the loss function on the sample x is computed as follows:

$$\mathcal{L} = \sum_{i=1}^{P-W} (x_{W+i} - \hat{x}_{W+i})^2 . (2)$$

The model parameters θ are therefore updated by minimizing $\mathcal{L}$ via gradient descent algorithms (e.g., Adam). For more efficient and robust training, we also use the batch model training strategy by simultaneously optimizing the model on multiple samples, until the model converges. For easy model hyperparameter tuning, we randomly select 10% of training samples as the validation set.

In the test phase, we use the model to predict the observations in the test split. We consider two types of data preparation methods. The first one is the leaked version, where the entire sequence x is used for signal decomposition. we denote its decomposed sequences (with Q observation values) as $\hat{\mathrm{X}}_{\mathrm{d}}$. For each prediction step i, we use the sliced subsequences of $\hat{\mathrm{X}}_{\mathrm{d}}$ as the model input for prediction as follows:

$$\hat{x}_{P+i} = f\big(\hat{X}_p[P+i-W{:}P+i-1];\theta\big), (3)$$

where $\hat{\mathrm{X}}_{\mathrm{p}}$ means the combination of the raw sequence x and $\hat{\mathrm{X}}_{\mathrm{d}}$. Since most signal decomposition methods like EMD and SSA are not strictly causal in theory, the decomposed sequences in fact contain the information on future observations, even though we only consider the data points before the prediction cutoff step.

By contrast, the second one is a non-leaked version, where only the data points before the prediction time are used for signal decomposition. For the i-th prediction step, we use the truncated sequence $x[1{:}\mathrm{P}+\mathrm{i}-1]$ to compute the decomposed sequences without future information leakage (denoted as $\hat{\mathrm{X}}_{\mathrm{d}}$). It is further combined with the original sequence x as the model input to infer the estimated value as follows:

$$\hat{x}_{P+i} = f\left(\tilde{X}_p[P+i-W{:}P+i-1];\theta\right), (4)$$

where $\widetilde{\mathrm{X}}_{\mathrm{p}}$ means the combination of the raw sequence x and $\widehat{\mathrm{X}}_{\mathrm{d}}$ without future information leakage. In this way, we ensure that the model can only be aware of the information before the time for prediction, thereby no label information is leaked. If there is a sufficiently significant difference between the prediction accuracy obtained by the two data processing methods, we can confirm that the improper data processing procedure indeed brings unwanted overestimation of the effectiveness of signal decomposition-based time series prediction methods.

5. APPENDIX

5.1 Dataset

We use 6 datasets for natural time series prediction in different scenarios. Their details are listed below.

Significant wave height (Hs) dataset, which comes from the Coastal Data Information Program (CDIP) (http://cdip.ucsd.edu/offline/wavecdf/wnc browse.php?ARCHIVE/150p1/150p1 historic). It was collected on-site at the designated location (34.142 N, 77.710 W) of UCSD (University of California, San Diego), and the time range was from January 1, 2017, to December 31, 2020, i.e., a total of four years. Significant wave height data has a time resolution of 30 minutes and a total of 70,128 points.

Wind Speed Averaged (WSPD) dataset, which comes from the measured data of National Data Buoy Center's East Coast site (Station 43.525 N, 70.141 W, https://www.ndbc.noaa.gov/). WSPD includes the average wind speed data of a total of 32,158 points from January 1, 2002 to December 31, 2005 with a time resolution of one hour.

Relative humidity (U) dataset, which comes from the Kaggle website (https: //www.kaggle.com/datasets/l3llff/electrical-grid-power-mw-20152021?resource=download), is the measured data at a power plant in Germany. The denotation "U" represents the relative humidity at a height of 2 meters above the surface. The dataset covers the time period from December 31, 2014 to July 8, 2021, with a time resolution of 15 minutes, resulting in a total of 228,526 data points.

The daily sum of global horizontal irradiation (GHI) data, which comes from Solargis website (https://solargis.com/products/evaluate/useful-resources). GHI is the site data generated at the location Plataforma Solar de Almeria, Spain (coordinates: 37.094 N, 2.360 W, elevation: 497.0m a.s.l.). The GHI contains data for a total of 9,952 points from January 1, 1994, to March 31, 2021, with a temporal resolution of one day.

Air pressure (P) dataset is a simulated one from National Renewable Energy Laboratory (NREL) (http://maps.nrel.gov/wind prospector), which measures the air pressure at an elevation of 100 meters at the west coast of the United States (41.812 N, 124.317 W). It includes four years of air pressure from January 1, 2014 to December 31, 2017, with a time resolution of 15 minutes and 140,160 data points.

Air temperature (T) dataset is also a simulated one from the same source as air pressure (P)(http://maps.nrel.gov/wind prospector). It measures the air temperature at the same elevation and location, with the same time resolution of 15 minutes and the number of data points of 140,160.

We present exemplary sequence diagrams (Fig. 6) and detailed statistics (Table 1) of the above 6 datasets.

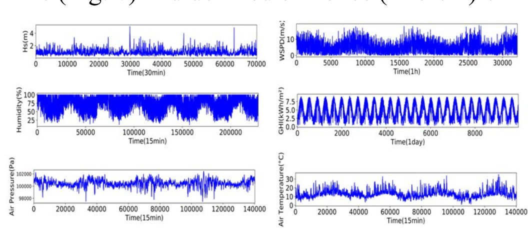

Figure. 6 Exemplary time series plots of the datasets used in our experiments. Different datasets are varied in their sequence lengths and resolutions.

Table 1. Statistics of the six datasets used in our experiments.

	Hs(m)	WSPD(m/s)	U(%)	GHI(kWh/m^2)	P(Pa)	T(°C)
Sequence Length	70,128	32,158	228,526	9,952	140,160	140,160

Number of Training Samples	52,596	24,118	171,394	7,464	105,120	105,120
Number of Test Samples	17,532	8,040	57,132	2,488	35,040	35,040
Mean	0.945	5.58	74.2	5.14	100.47k	13.1
Standard Deviation	0.414	3.18	19.5	2.25	555	4.12
Data Source Type	Observed	Observed	Observed	Observed	Simulated	Simulated

5.2 Experimental Setup

All data used in this study are processed using the information from the previous 12 steps to predict the subsequent step. The datasets are divided by time into a training set and a test set with a ratio of 0.75 to 0.25. Prior to training, the data are normalized using the MinMaxScaler technique. For both the MLP and LSTM models, the maximum training epoch is 1,000 and the batch size is 32. The optimizer we use is Adagrad for MLP and Adam for LSTM, and the learning rate is 0.0001. We set the epoch patience for early stopping as 30, i.e., the training will be terminated if the validation loss does not improve within the last 30 epochs. The MLP model contains three non-linear layers. The LSTM model consists of two LSTM layers followed by two fully-connected layers. The Transformer model we use has four Transformer encoder layers to extract features from the input sequence. The input feature dimension is set to 32. By using sinusoidal positional encoding, the model encodes the input sequence and adds it to the input so that positional information can be captured. In each Transformer encoder layer, the multi-head self-attention module has 8 attention heads and their output dimension is 32. The outputs from the Transformer encoders are aggregated using average pooling to obtain the hidden representation of the entire sequence. A fully connected layer with 128 neurons and ReLU activation function maps it to the predicted value. The optimizer employed is Adam, and a custom learning rate scheduler [42] is utilized to dynamically adjust the learning rate for better convergence. For the SSA method, we set the number of decomposed sequences to 3. For all models, the loss function and validation metrics are evaluated using the Mean Absolute Error (MAE). In the test phase, we use Mean Squared Error (MAE) as the main metric and include Mean Absolute Percentage Error (MAPE) as well as the coefficient of determination R2 as supplementary metrics.

5.3 Experimental Results in All Metrics

Here we present the complete results of different experiments in all three metrics (Table 2). The phenomena reflected by different metrics are consistent.

Table 2. The results of different models with different dataset processing methods. "Leak" indicates the data split method with future information leakage while "No Leak" indicates our causal data decomposition method. The results show that the leaked versions of all models show abnormal improvements over the original models without decomposed sequences. The performance decays when the information leakage is removed. The results in terms of different metrics exhibit consistent patterns.

Model	**Hs**			**WSPD**			**U**			**GHI**			**P**			**T**		
	MSE*	MAPE	R^2	MSE	MAPE	R^2	MSE	MAPE	R^2	MSE	MAPE	R^2	MSE	MAPE	R^2	MSE	MAPE	R^2
Persistence	3.092	4.175	0.9812	1.663	30.28	0.8494	8.613	1.935	0.9772	1.514	22.79	0.6843	243.7	0.0106	0.9994	0.0679	1.352	0.9971
MLP	2.838	4.005	0.9828	1.573	30.15	0.8575	8.304	1.987	0.9781	1.149	23.58	0.7603	259.1	0.0112	0.9993	0.0690	1.397	0.9970
MLP+EMD(Leak)	0.813	2.210	0.9951	0.371	13.42	0.9664	1.466	0.824	0.9960	0.462	12.62	0.9037	129.5	0.0077	0.9997	0.0403	1.162	0.9982
MLP+EMD(No Leak)	4.224	4.873	0.9743	2.038	36.03	0.8153	8.688	2.227	0.9770	1.191	24.39	0.7517	423.2	0.0144	0.9989	0.1056	1.801	0.9953
LSTM	35.76	3.775	0.9836	1.543	30.15	0.8602	8.313	1.902	0.9780	1.097	22.92	0.7712	188.9	0.0096	0.9995	0.0550	1.176	0.9976
LSTM+EMD(Leak)	0.573	1.808	0.9964	0.326	11.86	0.9704	0.630	0.396	0.9983	0.420	11.50	0.9124	38.19	0.0042	0.9999	0.0132	0.598	0.9994
LSTM+EMD(No Leak)	4.165	4.781	0.9747	1.979	34.49	0.8207	8.786	2.367	0.9768	1.206	24.09	0.7484	302.1	0.0122	0.9992	0.0792	1.542	0.9966
LSTM+DWT(Leak)	0.233	0.902	0.9986	0.131	6.148	0.9896	0.079	0.174	0.9998	0.103	5.485	0.9786	27.18	0.0036	0.9999	0.0086	0.437	0.9996
LSTM+DWT(No Leak)	3.209	4.229	0.9805	1.800	30.36	0.8369	8.731	2.396	0.9769	1.300	23.33	0.7289	225.7	0.0105	0.9994	0.0701	1.434	0.9970

LSTM+SSA(Leak)	0.020	0.264	0.9999	0.004	1.158	0.9996	0.006	0.087	0.9999	0.007	1.425	0.9986	1.477	0.0008	0.9999	0.0008	0.128	0.9999
LSTM+SSA(No Leak)	3.325	4.390	0.9798	1.825	30.32	0.8347	9.365	2.723	0.9753	1.385	25.94	0.7111	201.5	0.0099	0.9995	0.0703	1.315	0.9970
Transformer	2.810	3.924	0.9829	1.570	31.03	0.8577	8.116	1.957	0.9786	1.103	22.91	0.7698	310.7	0.0113	0.9992	0.0636	1.323	0.9973
Transformer+EMD(Leak)	1.184	2.276	0.9928	0.368	13.10	0.9666	1.233	0.644	0.9971	0.399	11.21	0.9168	187.6	0.0085	0.9995	0.0372	1.014	0.9984
Transformer+EMD(No Leak)	4.355	4.810	0.9735	2.106	36.86	0.8092	9.483	2.598	0.9749	1.216	24.53	0.7465	363.4	0.0130	0.9990	0.0817	1.584	0.9965

*The MSE data of the Hs dataset has been enlarged by 1000 times (the unit is m2*10−3).

6. ACKNOWLEDGMENTS

We acknowledge the financial support from the National Science Foundation of China (Nos. 61973208 and 62225308).

7. REFERENCES

[1] T. Geng, F. Jia, W. Cai, L. Wu, B. Gan, Z. Jing, S. Li, M. J. McPhaden, Increased occurrences of consecutive la nina events under global warming, ˜ Nature 619 (7971) (2023) 774–781.
[2] P. Ball, Economics prize for duo who resolved random results, Nature 425 (6959) (2003) 651–652.
[3] M. Cerf, S. C. Matz, M. A. MacIver, Participating in a climate prediction market increases concern about global warming, Nature Climate Change 13 (6) (2023) 523–531.
[4] Insights into the accuracy of social scientists' forecasts of societal change, Nature human behaviour 7 (4) (2023) 484–501.
[5] K. Ichimoto, J. Kubota, M. Suzuki, I. Tohmura, H. Kurokawa, Periodic behaviour of solar flare activity, Nature 316 (6027) (1985) 422–424.
[6] M. Ruttgers, S. Lee, S. Jeon, D. You, Prediction of a typhoon track using a ¨ generative adversarial network and satellite images, Scientific reports 9 (1) (2019) 6057.
[7] S. Poddar, J. P. Evans, M. Kay, A. Prasad, S. Bremner, Assessing australia's future solar power ramps with climate projections, Scientific Reports 13 (1) (2023) 11503.
[8] Y. Chen, Z. Dong, Y. Wang, J. Su, Z. Han, D. Zhou, K. Zhang, Y. Zhao, Y. Bao, Short-term wind speed predicting framework based on eemd-ga-lstm method under large scaled wind history, Energy Conversion and Management 227 (2021) 113559.
[9] Z. Lu, N. Ren, X. Xu, J. Li, C. Panwisawas, M. Xia, H. Dong, E. Tsang, J. Li, Real-time prediction and adaptive adjustment of continuous casting based on deep learning, Communications Engineering 2 (1) (2023) 34.
[10] J. Liu, D. Capurro, A. Nguyen, K. Verspoor, Early prediction of diagnosticrelated groups and estimation of hospital cost by processing clinical notes, NPJ digital medicine 4 (1) (2021) 103.
[11] X. Zhu, Z. Song, G. Sen, M. Tian, Y. Zheng, B. Zhu, Prediction study of electric energy production in important power production base, china, Scientific Reports 12 (1) (2022) 21472.
[12] K. Kuester, S. Mittnik, M. S. Paolella, Value-at-risk prediction: A comparison of alternative strategies, Journal of Financial Econometrics 4 (1) (2006) 53– 89.
[13] G. Evensen, The ensemble kalman filter: Theoretical formulation and practical implementation, Ocean dynamics 53 (2003) 343–367.
[14] B. Fox-Kemper, A. Adcroft, C. W. Boning, E. P. Chassignet, E. Curchitser, ¨ G. Danabasoglu, C. Eden, M. H. England, R. Gerdes, R. J. Greatbatch, et al., Challenges and prospects in ocean circulation models, Frontiers in Marine Science 6 (2019) 65.
[15] C. D. Argyropoulos, N. Markatos, Recent advances on the numerical modelling of turbulent flows, Applied Mathematical Modelling 39 (2) (2015) 693–732.
[16] G. Zhang, V. N. Mahale, B. J. Putnam, Y. Qi, Q. Cao, A. D. Byrd, P. Bukovcic, D. S. Zrnic, J. Gao, M. Xue, et al., Current status and future challenges of weather radar polarimetry: Bridging the gap between radar meteorology/hydrology/engineering and numerical weather prediction, Advances in Atmospheric Sciences 36 (2019) 571–588.
[17] R. P. Masini, M. C. Medeiros, E. F. Mendes, Machine learning advances for time series forecasting, Journal of economic surveys 37 (1) (2023) 76–111.

[18] Z. Han, J. Zhao, H. Leung, K. F. Ma, W. Wang, A review of deep learning models for time series prediction, IEEE Sensors Journal 21 (6) (2019) 7833– 7848.
[19] N. Davari, B. Veloso, G. d. A. Costa, P. M. Pereira, R. P. Ribeiro, J. Gama, A survey on data-driven predictive maintenance for the railway industry, Sensors 21 (17) (2021) 5739.
[20] M. Langkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature ¨ learning and deep learning for time-series modeling, Pattern recognition letters 42 (2014) 11–24.
[21] Y. Chen, Y. Wang, Z. Dong, J. Su, Z. Han, D. Zhou, Y. Zhao, Y. Bao, 2-d regional short-term wind speed forecast based on cnn-lstm deep learning model, Energy Conversion and Management 244 (2021) 114451.
[22] M. T. Ribeiro, S. Singh, C. Guestrin, ” why should i trust you?” explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
[23] F. Petropoulos, R. J. Hyndman, C. Bergmeir, Exploring the sources of uncertainty: Why does bagging for time series forecasting work?, European Journal of Operational Research 268 (2) (2018) 545–554.
[24] G. Ning, Z. Zhang, Z. He, Knowledge-guided deep fractal neural networks for human pose estimation, IEEE Transactions on Multimedia 20 (5) (2017) 1246–1259.
[25] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational physics 378 (2019) 686–707.
[26] J. Oh, K.-D. Suh, Real-time forecasting of wave heights using eof–wavelet– neural network hybrid model, Ocean Engineering 150 (2018) 48–59.
[27] Z. Qian, Y. Pei, H. Zareipour, N. Chen, A review and discussion of decomposition-based hybrid models for wind energy forecasting applications, Applied energy 235 (2019) 939–953.
[28] S. Soltani, On the use of the wavelet decomposition for time series prediction, Neurocomputing 48 (1-4) (2002) 267–277.
[29] X. Wang, H. Liu, J. Du, X. Dong, Z. Yang, A long-term multivariate time series forecasting network combining series decomposition and convolutional neural networks, Applied Soft Computing 139 (2023) 110214.
[30] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Conversion and Management 198 (2019) 111799.
[31] S. Shamshirband, K. Mohammadi, H. Khorasanizadeh, L. Yee, M. Lee, D. Petkovic, E. Zalnezhad, Estimating the diffuse solar radiation using a ´ coupled support vector machine–wavelet transform model, Renewable and Sustainable Energy Reviews 56 (2016) 428–435.
[32] J. Yan, L. Mu, L. Wang, R. Ranjan, A. Y. Zomaya, Temporal convolutional networks for the advance prediction of enso, Scientific reports 10 (1) (2020) 8055.
[33] S. Yang, Z. Deng, X. Li, C. Zheng, L. Xi, J. Zhuang, Z. Zhang, Z. Zhang, A novel hybrid model based on stl decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast, Renewable Energy 173 (2021) 531–543.
[34] R. Bolli, X.-L. Tang, Y. Guo, Q. Li, After the storm: an objective appraisal of the efficacy of c-kit+ cardiac progenitor cells in preclinical models of heart disease, Canadian journal of physiology and pharmacology 99 (2) (2021) 129–139.
[35] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.- C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences 454 (1971) (1998) 903–995.
[36] I. Daubechies, Ten lectures on wavelets, SIAM, 1992.
[37] R. Vautard, P. Yiou, M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D: Nonlinear Phenomena 58 (1-4) (1992) 95–126.
[38] M. Jamei, M. Ali, A. Malik, M. Karbasi, P. Rai, Z. M. Yaseen, Development of a tvf-emd-based multi-decomposition technique integrated with encoder-decoder-bidirectional-lstm for monthly rainfall forecasting, Journal of Hydrology 617 (2023) 129105.
[39] P. C. Deka, R. Prahlada, Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time, Ocean Engineering 43 (2012) 32–42.
[40] S. R. Moreno, L. dos Santos Coelho, Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system, Renewable energy 126 (2018) 736–754.
[41] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[43] J. Mahjoobi, E. A. Mosabbeb, Prediction of significant wave height using regressive support vector machines, Ocean Engineering 36 (5) (2009) 339– 347.
[44] S. C. James, Y. Zhang, F. O'Donncha, A machine learning framework to forecast wave conditions, Coastal Engineering 137 (2018) 1–10.
[45] Y. Jing, L. Zhang, W. Hao, L. Huang, Numerical study of a cnn-based model for regional wave prediction, Ocean Engineering 255 (2022) 111400.
[46] S. Fan, N. Xiao, S. Dong, A novel model to predict significant wave height based on long short-term memory network, Ocean Engineering 205 (2020) 107298.
[47] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, Advances in Neural Information Processing Systems 34 (2021) 22419–22430.