Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.35325

TFF-Former: A Time-Frequency Fusion Transformer for Multivariate Weather Time Series Forecasting

Zhixiang Zhang

Faculty of Arts And Sciences, Beijing Normal University, Beijing, China
3022804109@qq.com

Abstract. Multivariate weather time series forecasting is highly challenging due to the complex coexistence of short-term dynamic fluctuations and long-term periodic patterns. Existing models predominantly focus on the time domain, lacking explicit mechanisms to capture frequency priors, or rely on complex spatial reconstructions that struggle with long-range temporal dependencies. To address these limitations, we propose the Time-Frequency Fusion Transformer (TFF-Former), a novel dual-branch architecture. Specifically, the time-domain branch utilizes a Transformer encoder to extract local dynamics and variable dependencies, while the frequency-domain branch employs the real Fast Fourier Transform (rFFT) combined with a parallel Transformer encoder to explicitly model multi-scale periodic structures from amplitude spectra. The representations from both domains are subsequently concatenated to achieve feature complementarity. Experiments on the Jena Climate dataset demonstrate that TFF-Former achieves state-of-the-art performance, outperforming strong baselines including TimesNet and PatchTST across MAE, MSE, RMSE, and SMAPE metrics. Ablation studies further validate that the time-frequency fusion mechanism significantly surpasses single-domain modeling.

Keywords: Multivariate Time Series Forecasting, Weather Prediction, Transformer, Time-Frequency Fusion, Fast Fourier Transform

1. Introduction

Multivariate Time Series Forecasting (MTSF) is widely applied in fields such as energy and transportation [1], and weather time series forecasting is a highly typical and important component of this field [2, 3]. Compared with other time series, some variables in weather sequences show obvious periodicity. Typical examples include temperature and humidity, which follow both a daily cycle formed by day-night alternation and an annual cycle driven by seasonal changes. Furthermore, as a highly coupled multivariate system, weather sequences such as temperature and humidity follow periodic patterns while also experiencing sharp short-term fluctuations caused by local weather disturbances, presenting a combined feature of short-term volatility and long-term regularity.

Among existing mature models, ResNet relies on convolution to extract local patterns and performs moderately on short-term changes, but has limited capacity for long-range dependency modeling and periodic structure capture. Transformer and its variants have achieved great success in

DOI: 10.54254/2755-2721/2026.35325

long sequence modeling via the attention mechanism, but they lack dedicated mechanisms to explicitly extract and leverage the hidden frequency and periodic information in sequences. In addition, although models such as TimesNet attempt to introduce frequency-domain analysis, their core still leans toward structural reconstruction and 2D convolution modeling, and remains inadequate in flexibly integrating periodic features with temporal dynamic dependencies.

To fill this gap, this paper proposes a Time-Frequency Fusion Transformer (TFF-Former). The model adopts a parallel dual-branch structure to capture data features in the time domain and frequency space respectively:

1. Time-domain branch: It still uses a TransformerEncoder to track the trend of historical time series and extract dynamic changes and coupling dependencies among variables.

2. We first apply real Fast Fourier Transform (rFFT) to the sequence to extract the amplitude of the frequency spectrum, and then use a TransformerEncoder for modeling. This allows the model to capture periodic patterns directly in the frequency space.

3. Finally, we concatenate and fuse the outputs of the two branches to achieve time-frequency complementarity.

Based on the above motivations and innovations, the main contributions of this paper are summarized as follows:

1. A novel multivariate forecasting framework TFF-Former is proposed, which effectively combines time-domain sequence dependencies and frequency-domain periodic patterns through an explicit dual-branch design.

2. Systematic experiments on the Jena Climate dataset show that TFF-Former outperforms baseline models including Transformer, TimesNet, PatchTST and ResNet, and achieves state-of-the-art (SOTA) results on four metrics: MAE, MSE, RMSE and SMAPE.

3. Rigorous ablation studies are conducted to verify the necessity of collaborative fusion between time-domain and frequency-domain modeling, proving that the time-frequency fusion strategy brings significant performance gains.

2. Related work

Time series forecasting based on conventional deep learning. With the development of deep learning technology, neural network-based methods have gradually replaced traditional statistical models and become the mainstream for time series forecasting [4]. Among them, Convolutional Neural Network (CNN)-based models such as ResNet are widely used for local time series forecasting [5, 6]. Such models perform acceptably on short-term changes, but their deficiencies in long-range dependency modeling and periodic structure capture make them hard to fully meet the demands of complex weather time series forecasting tasks.

Transformer-based time series forecasting. To overcome the receptive field limitation of convolutional networks, the Transformer architecture has been introduced to capture global information in sequences. It can well model long-range temporal dependencies and exhibits strong performance in multivariate weather series forecasting [7-9]. Moreover, recent improved models represented by PatchTST further split long sequences into patches and model them with Transformer, achieving a good balance between local patterns and global dependencies [10]. However, these models operate entirely in the time domain, underutilize the periodic information in weather series such as temperature, and lack explicit modeling capability for frequency and periodic features.

Time series forecasting with frequency and periodic feature fusion. In view of the prominent periodicity in weather time series, recent frontier work has begun to explore explicit utilization of

DOI: 10.54254/2755-2721/2026.35325

periodic information [11, 12]. For example, TimesNet extracts dominant periods via frequency analysis and maps 1D time series into multi-period 2D structures for convolution modeling, which is suitable for capturing periodic patterns in time series [13]. However, despite its strong periodic modeling ability, TimesNet still relies on convolution after structural reconstruction, and is less flexible than pure attention-based models in handling global temporal dynamic dependencies. In contrast, the proposed TFF-Former directly applies Transformer encoding to the spectrum amplitude via an independent frequency-domain branch and fuses it with time-domain representations, which can integrate periodic information and long-range temporal dependencies more flexibly and directly.

3. Methodology

3.1. Problem definition

We first define the multivariate weather time series forecasting task. From observed data, we obtain a historical observation sequence $\mathscr{X} = X_1, X_2, \ldots, X_T \in \mathbb{R}^{T\times C}$, where T denotes the historical time step length of the input sequence, and C denotes the number of weather variable features (e.g., temperature, humidity, air pressure, wind speed, etc.).

The task aims to predict the target sequence $\mathscr{Y} = Y_{T+1}, Y_{T+2}, \ldots, Y_{T+P} \in \mathbb{R}^{P\times D}$ for the next P time steps based on the historical sequence $\mathscr{X}$, where P is the prediction length and D is the dimension of target variables to be predicted.

This forecasting process can be abstracted as a mapping function f_θ :

$$\mathscr{Y} = f_\theta(\mathscr{X}) \tag{1}$$

where θ represents all learnable parameters of the proposed TFF-Former model.

3.2. Overall architecture

To address the coexistence of short-term fluctuations and long-range periodic patterns in weather data, the proposed TFF-Former (Time-Frequency Fusion Transformer) innovatively adopts a parallel dual-branch architecture followed by feature fusion, solving the problem that traditional models (such as pure time-domain Transformer or spatial-reconstruction based frequency-domain networks) can hardly balance temporal dynamic dependencies and explicit periodic frequency features. As shown in Figure 1, TFF-Former mainly consists of three core modules:

1. Time-Domain Branch: It takes the raw weather sequence X as input and uses a standard Transformer encoder to mine the contextual dependencies and short-term evolution trends of multivariate sequences directly along the time dimension.

2. Frequency-Domain Branch: It first maps the raw time series into the frequency space via Fast Fourier Transform (FFT) and extracts the amplitude spectrum, then uses a parallel Transformer encoder to capture the latent coupling relations among different frequency components.

3. Feature Fusion and Prediction Head: It concatenates the high-dimensional representations from both time and frequency domains, and performs direct multi-step mapping from the mixed feature space to the future target sequence via a linear network.

DOI: 10.54254/2755-2721/2026.35325

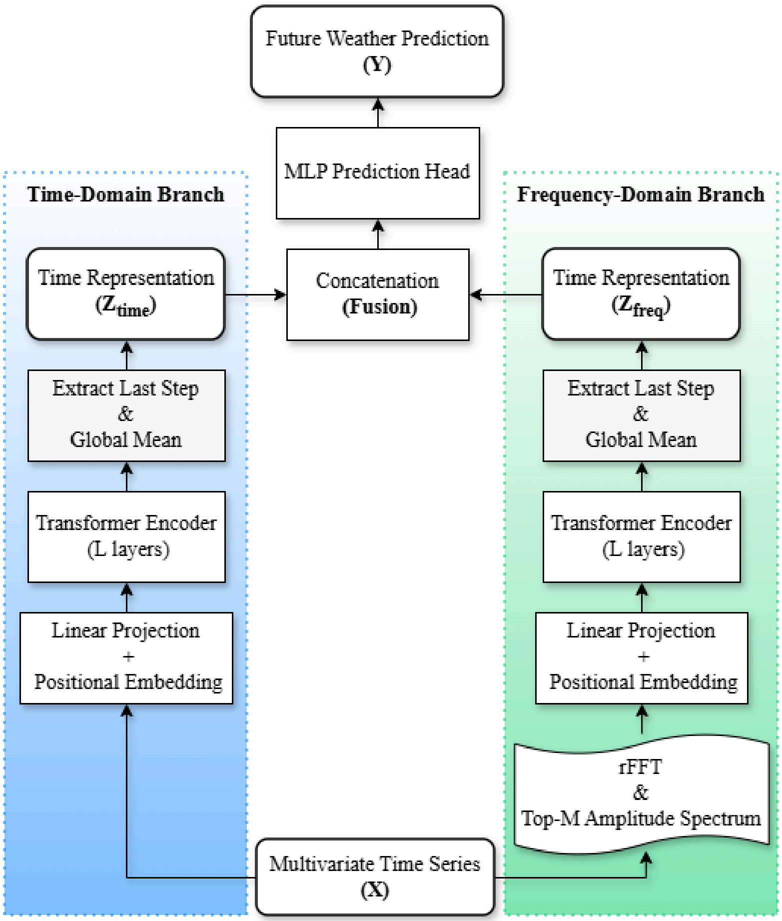

Figure 1. The overall architecture of the proposed Time-Frequency Fusion Transformer (TFF-Former)

3.3. Time-frequency branches and fusion

The core of TFF-Former consists of two parallel branches: the time-domain branch and the frequency-domain branch. This section elaborates the feature extraction principle of each branch and the final fusion and prediction mechanism.

3.3.1. Time-domain branch

The time-domain branch is designed to capture local dynamic fluctuations and long-range temporal dependencies in the input sequence X, and we use the standard Transformer architecture here. First, the multivariate input is projected to a higher feature dimension d_{model} via a linear projection layer, and positional embedding is added to preserve the order information of time steps:

$$H_{time}^{(0)} = \mathrm{Linear}_{time}(\mathscr{X}) + \mathrm{PosEmb}_{time}(\mathscr{X}) \tag{2}$$

Then the feature matrix $H_{time}^{(0)}$ is fed into a standard Transformer encoder with L layers. The core of the encoder is the self-attention mechanism, whose mathematical expression is:

DOI: 10.54254/2755-2721/2026.35325

$$\text{Attention}\left(Q,K,V\right)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q,K,V are query, key and value matrices obtained by projecting the feature matrix with different weights, and d_k is the scaling factor. Through this mechanism, the model can adaptively focus on the time steps in the historical sequence that are most informative for the current prediction.

After L encoding layers, the high-level time-domain representation $H_{time}^{(L)}$ is obtained. To integrate the latest local state and the global macro context, we extract the feature of the last time step and the global average feature of the whole sequence, then concatenate them as the final output of the time-domain branch Z_{time} :

$$Z_{time}=\text{Concat}\left(H_{time}^{(L)}\left[-1,:\right],\text{Mean}\left(H_{time}^{(L)}\right)\right) \tag{4}$$

3.3.2. Frequency-domain branch

Weather time series are often intertwined with periodic components of different frequencies (e.g., daily cycle, annual cycle), and the frequency-domain branch is the key module for TFF-Former to capture multi-scale periodic patterns. First, we apply real Fast Fourier Transform (rFFT) to the input sequence $\mathscr{X}$ along the time dimension:

$$F=\text{rFTT}(\mathscr{X}) \tag{5}$$

The transformed F lies in the complex domain $\mathbb{C}$. We extract its magnitude spectrum $A=|F|$ to measure the intensity of different frequency components. To control computational complexity and focus on dominant frequency features, we only keep the first M low-frequency components to filter out high-frequency noise, obtaining the truncated magnitude spectrum A_M.

Different from traditional 2D convolution reconstruction methods, this model treats A_M as a frequency sequence, projects it to the d_{model} dimension and adds frequency-domain positional embedding:

$$H_{freq}^{(0)}=\text{Linear}_{freq}\left(A_M\right)+\text{PosEmb}_{freq}\left(A_M\right) \tag{6}$$

Then a Transformer encoder is applied again for deep feature mining, yielding the high-level frequency-domain representation $H_{freq}^{(L)}$. Here the self-attention mechanism computes the latent coupling relations among different frequency components instead of dependencies between time steps. Similar to the time-domain branch, the final output of the frequency-domain branch Z_{freq} is also formed by concatenating the last frequency feature and the average frequency feature:

$$Z_{freq}=\text{Concat}\left(H_{freq}^{(L)}\left[-1,:\right],\text{Mean}\left(H_{freq}^{(L)}\right)\right) \tag{7}$$

3.3.3. Feature fusion and prediction head

After obtaining the time-domain representation Z_{time} and frequency-domain representation Z_{freq}, the model enters the feature fusion stage. Since the two branches perform symmetric computations

DOI: 10.54254/2755-2721/2026.35325

under the unified deep dimension d_{model}, we adopt the straightforward and effective concatenation operation to achieve complementary cross-domain feature fusion:

$$Z_{fuse} = \mathrm{Concat}(Z_{time}, Z_{freq}) \tag{8}$$

Finally, the fused multi-domain features Z_{fuse} are fed into a prediction head composed of Multi-Layer Perceptron (MLP), which directly maps and outputs the prediction sequence $\mathcal{Y}$ for the next P steps.

4. Results and discussion

4.1. Datasets and preprocessing

4.1.1. Datasets and task setup

This experiment uses the Jena Climate dataset provided by the Jena weather station in Germany for evaluation [14]. The dataset has a sampling frequency of 10 minutes, covers 8 years from 2009 to 2016, and includes multiple weather variables such as temperature, humidity, air pressure, wind speed and vapor pressure.

Based on this dataset, we formulate a short-term forecasting task with multivariate input and univariate output: temperature T (degC) is set as the target prediction variable, while the remaining variables and temporal features serve as input features.

The dataset is split chronologically into training set (70%), validation set (10%) and test set (20%) with a ratio of 7:1:2. Samples are constructed via the sliding window method, with the historical input length set to 96, prediction length set to 24, and sliding step set to 6 (i.e., using observations from the past 16 hours to predict temperature changes in the next 4 hours, and shifting the window by 1 hour after each prediction).

4.1.2. Data preprocessing

To make the dataset suitable for the experiment, we perform the following preprocessing steps:

1. Outlier repair: For abnormal placeholder values of -9999.0 in fields such as wind speed, we treat them as missing data and fill them via a combination of linear interpolation, forward filling and backward filling.

2. Temporal feature construction: Since the sliding window method truncates long sequences and may lose macro global temporal periodic information, we manually construct sine and cosine encodings for daily and annual cycles (sin_day, cos_day, sin_year, cos_year) by parsing timestamps, and append them to the original data as additional input features. This step explicitly encodes the critical day-night alternation and seasonal variation information of weather series, making temporal information available to the model.

3. Standardization: We use StandardScaler to standardize the data. To strictly avoid data leakage, the mean and standard deviation for standardization are computed only on the training set, and then applied to the entire dataset.

DOI: 10.54254/2755-2721/2026.35325

4.2. Metrics and experimental settings

4.2.1. Baselines and metrics

To evaluate the performance of TFF-Former comprehensively and objectively, four representative deep learning baseline models in time series forecasting are selected for comparison:

• Transformer: Uses a lightweight Transformer encoder as the pure time-domain baseline, to verify the ability to capture long-range dependencies in multivariate sequences.

• TimesNet: A state-of-the-art model combining frequency-domain analysis and convolution, specialized in extracting dominant periods and mapping 1D time series into multi-period 2D structures for modeling.

• PatchTST: Combines patch splitting strategy with attention mechanism, achieving a good balance between local patch patterns and global long-range dependency modeling.

• ResNet: Uses a 1D residual convolutional network to extract local temporal patterns, as the representative of pure convolution-based forecasting frameworks.

For model evaluation, we select four widely used metrics in time series forecasting to quantify prediction accuracy: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Symmetric Mean Absolute Percentage Error (SMAPE). Lower values of these four metrics indicate smaller deviation between predictions and ground truth, and better overall forecasting performance.

4.2.2. Experimental setup and environment

For fair comparison, all models follow a unified training strategy:

• Batch size: Uniformly set to 8.

• Training epochs: Maximum epoch number set to 20.

• Optimizer and loss: AdamW optimizer is used, and Mean Squared Error Loss (MSELoss) is adopted as the loss function for all models.

• Early stopping: To prevent overfitting, training is terminated early if the validation loss does not decrease for 7 consecutive epochs (Patience=7).

• Learning rate: According to the convergence characteristics of different models, initial learning rates are set as follows: 0.0001 for TimesNet, and 0.0003 for Transformer, PatchTST, ResNet and the proposed TFF-Former.

All experiments are implemented on the deep learning framework with Python 3.10 and PyTorch 2.2. Data processing relies on scientific computing libraries including NumPy 1.26, Pandas 2.1 and Scikit-learn 1.4. Experiments run on Ubuntu 20.04, equipped with a server-grade high-performance CPU and a single NVIDIA A100 (80GB) GPU.

4.3. Main results

The quantitative evaluation results are shown in Table 1. It can be clearly observed that the proposed TFF-Former achieves state-of-the-art (SOTA) performance on all four metrics: MAE, MSE, RMSE and SMAPE.

DOI: 10.54254/2755-2721/2026.35325

Table 1. Comparison of multivariate prediction performance of different models on the Jena Climate dataset

Models	MAE(degC)	MSE(degC^2)	RMSE(degC)	SMAPE(%)
Transformer	0.721	1.041	1.02	13.47
TimesNet	0.781	1.208	1.099	14.74
PatchTST	0.744	1.106	1.052	13.93
ResNet	0.861	1.392	1.18	16.21
TFF-Former	0.684	0.963	0.981	12.86

Combined with the data in Table 1, we draw the following in-depth experimental analysis and conclusions:

1. Time-frequency dual-domain fusion significantly breaks through the bottleneck of pure time-domain modeling. Pure time-domain models represented by vanilla Transformer and PatchTST show strong ability in modeling long-range dependencies of multivariate sequences via self-attention or patch splitting. Among them, PatchTST achieves a good balance between local patterns and global dependencies, and performs best among the baselines. However, they are confined to time-domain sequence evolution and lack explicit extraction of periodic features (day-night alternation and seasonal variation) hidden in weather data. TFF-Former introduces an independent frequency-domain branch on this basis, enabling the model to perceive frequency distribution directly in the spectrum space, thus reducing MAE from 0.744 of PatchTST to 0.684, achieving notable error reduction.

2. Compared with traditional spatial reconstruction, computing attention directly on spectrum amplitude is more efficient. TimesNet also addresses the periodicity of time series and performs significantly better than ResNet. But its core logic relies on reconstructing 1D sequences into 2D structures followed by convolution extraction. In contrast, TFF-Former obtains the spectrum amplitude via rFFT and uses a Transformer encoder to model coupling relations among different frequency components. This mechanism is more flexible and direct in integrating periodic information and temporal dependencies, so its MSE (0.963) is notably better than that of TimesNet (1.208).

3. Pure local convolution structures struggle with long-term evolution of complex weather sequences. Among all evaluated models, ResNet performs the worst (MAE up to 0.861, SMAPE over 16%). This indicates that relying solely on 1D convolution to extract local temporal patterns can capture local weather disturbances, but fails to establish long-range temporal dependencies, let alone effectively capture macro periodic structures. This also reversely confirms the necessity of introducing global attention mechanism and frequency periodic features in TFF-Former.

4.4. Ablation study

To verify the specific contribution and necessity of the time-frequency dual-branch architecture in TFF-Former, we design and conduct ablation studies under the same experimental environment and parameter settings. We remove specific modules from the full TFF-Former to construct two variant models for independent evaluation:

• TFF-Former-TimeOnly: Only retains the time-domain branch, with frequency-domain feature extraction and fusion modules removed, serving as the pure temporal baseline.

• TFF-Former-FreqOnly: Only retains the frequency-domain branch, with direct input of raw time series removed, to verify the independent forecasting ability of frequency-domain features.

DOI: 10.54254/2755-2721/2026.35325

The quantitative comparison of these ablation variants with vanilla Transformer and the full TFF-Former is shown in Table 2.

Table 2. Ablation study results of TFF-Former on the Jena Climate dataset

Models	MAE(degC)	MSE(degC^2)	RMSE(degC)	SMAPE(%)
Transformer	0.721	1.041	1.02	13.47
TFF-Former-FreqOnly	0.759	1.155	1.075	14.09
TFF-Former-TimeOnly	0.706	1.012	1.006	13.18
TFF-Former (Full)	0.684	0.963	0.981	12.86

Through in-depth analysis of Table 2, we draw the following key conclusions on model architecture design:

1. The task-customized time-domain architecture outperforms vanilla Transformer. The experimental results show that although TFF-Former-TimeOnly essentially degenerates into a pure time-domain attention model, its error metrics (e.g., MAE drops to 0.706, MSE drops to 1.012) are all better than the basic Transformer. This indicates that the network structure design and prediction head form optimized for the weather forecasting task are inherently reasonable, laying a solid performance foundation for the subsequent dual-branch fusion.

2. A single frequency-domain branch cannot independently support complex time series forecasting. Among all ablation variants, TFF-Former-FreqOnly performs the worst (MSE rises to 1.155, SMAPE exceeds 14%). This reveals the essential characteristic of weather data: although frequency-domain information can effectively describe macro seasonal and diurnal periodic structures, it lacks dynamic evolution information from the raw time series. Without perception of local weather disturbances and adjacent time-step dependencies, the model can hardly make accurate short-term predictions. Therefore, frequency-domain features can only serve as a supplement, not a replacement for time-domain features.

3. Collaborative fusion of time-frequency dual branches achieves complementary advantages. The full TFF-Former, after fusing time-frequency dual-domain features, outperforms all single-branch models across all metrics. This strongly proves the high complementarity of time domain and frequency domain in weather forecasting: the time-domain branch dominates tracking temporal dependencies and dynamic changes, while the frequency-domain branch injects clear periodic patterns and frequency structure priors into the model. This two-pronged strategy effectively breaks the information bottleneck of single-domain modeling, verifying the success of the core design idea of this work.

5. Conclusion and future work

To address the challenge of intertwined short-term dynamic fluctuations and long-range periodic patterns in multivariate weather time series forecasting, this paper proposes a novel time-frequency fusion model — TFF-Former. Through a parallel dual-branch architecture of time domain and frequency domain, the model effectively realizes collaborative extraction of temporal dynamic dependencies and frequency periodic features.

Comparative experiments on the Jena Climate dataset show that TFF-Former outperforms state-of-the-art baseline models such as TimesNet and PatchTST on all four metrics: MAE, MSE, RMSE and SMAPE. Rigorous ablation studies further confirm the indispensability of the time-frequency dual-branch structure: the frequency-domain branch effectively enhances the model's periodic

DOI: 10.54254/2755-2721/2026.35325

modeling ability, while the time-domain branch captures dependencies of raw time series. Their fusion can more fully exploit multi-scale features in weather sequences. This fully proves the efficiency and validity of the time-frequency fusion mechanism for complex weather time series forecasting.

Although TFF-Former achieves significant performance improvement with a relatively simple structure, there is still room for further exploration and optimization. Future work can focus on the following two directions:

1. Explore more advanced feature fusion methods. The current model uses straightforward concatenation to fuse dual-domain features. Future work can introduce advanced mechanisms such as cross-attention to enable bidirectional interaction and deep coupling of time-domain and frequency-domain features at earlier stages of the network.

2. Realize end-to-end adaptive extraction of ultra-long periods. Limited by the current sliding window length (96 steps), this study uses manually constructed temporal encodings to compensate for the lack of macro periodic information. Future work can explore how to expand the model's receptive field without significantly increasing computational cost, so that it can automatically learn ultra-long periodic patterns from raw data and build a fully end-to-end forecasting paradigm.

References

[1] Wen, Q., et al. (2022). Transformers in time series: A survey. arXiv preprint arXiv: 2202.07125.

[2] Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2026). Pangu-Weather: A 3D high-resolution model for fast and accurate global weather forecast. arXiv.org. https: //arxiv.org/abs/2211.02556v1

[3] Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Advances in Neural Information Processing Systems. Curran Associates, Inc. https: //proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html

[4] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https: //doi.org/10.1162/neco.1997.9.8.1735

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). https: //openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[6] Bai, S., Kolter, J. Z., & Koltun, V. (2018, April 19). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv. arXiv: 1803.01271. doi: 10.48550/arXiv.1803.01271.

[7] Vaswani, A. et al. (2017). Attention is all you need. In Advances in Neural Information Processing Systems. Curran Associates, Inc. Retrieved May 29, 2026, from https: //proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[8] Zhou, H. et al. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting.Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106–11115. doi: 10.1609/aaai.v35i12.17325.

[9] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In Advances in Neural Information Processing Systems (pp. 22419–22430). Curran Associates, Inc. Retrieved May 29, 2026, from https: //proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html

[10] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (n.d.). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv.org*. Retrieved May 29, 2026, from https: //arxiv.org/abs/2211.14730v2

[11] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022, June). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Proceedings of the 39th International Conference on Machine Learning (pp. 27268–27286). PMLR. https: //proceedings.mlr.press/v162/zhou22g.html

[12] Lee-Thorp, J., Ainslie, J., Eckstein, I., & Ontañón, S. (2022, July). FNet: Mixing tokens with Fourier transforms. In M. Carpuat, M.-C. de Marneffe, & I. V. Meza Ruiz (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4296–4313). Association for Computational Linguistics. https: //doi.org/10.18653/v1/2022.naacl-main.319

DOI: 10.54254/2755-2721/2026.35325

[13] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv: 2210.02186.
[14] Max Planck Institute for Biogeochemistry, "Jena Climate Dataset." 2009. [Online]. Available: https: //www.bgc-jena.mpg.de/wetter/