Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada
DOI: 10.23977/acss.2026.100214
ISSN 2371-8838 Vol. 10 Num. 2

Air Quality Index Forecasting Based on a CEEMDAN-KPCA-TCN-BiGRU-AMRC Hybrid Model

Honglin Ke[1,a,*], Chaohui Hao[1,b]

[1]*School of Labor Economics, China University of Labor Relations, Beijing, China*
[a]*happil1n@163.com*, [b]*13722358337@163.com*
**Corresponding author*

Keywords: Air Quality Index; CEEMDAN; Kernel Principal Component Analysis; TCN-BiGRU; Adaptive Markov Residual Correction

Abstract: To address the nonlinear, non-stationary, and abrupt fluctuations in the Air Quality Index (AQI) series, influenced by pollutant emissions, meteorological conditions, and random disturbances, this paper proposes a hybrid prediction model integrating Adaptive Noisy Complete Ensemble Empirical Mode Decomposition (CEEMDAN), Kernel Principal Component Analysis (KPCA), Temporal Convolutional Network (TCN), Bidirectional Gated Recurrent Unit (BiGRU), and Adaptive Markov Residual Correction (AMRC). First, CEEMDAN is used to decompose the original AQI series at multiple scales to separate high-frequency disturbances, periodic fluctuations, and long-term trend information. Second, KPCA is employed to extract nonlinear features from pollutants and meteorological variables, reducing input redundancy and enhancing the representation of key features. Subsequently, a TCN-BiGRU prediction network is constructed, extracting local temporal patterns through a temporal convolutional structure and learning sequence dependencies using a bidirectional recurrent structure. Finally, based on the state transition characteristics of the prediction residuals, an AMRC mechanism is introduced to dynamically correct the initial prediction results. Experimental results show that the proposed model outperforms the comparative model in terms of RMSE, MAE, MAPE, and R^2 and exhibits better tracking capabilities during local peaks, drastic fluctuations, and medium-to-high pollution stages. Residual distribution, autocorrelation analysis, and state transition heatmaps further validate the corrective effect of AMRC on systematic errors.

1. Introduction

With the continuous advancement of urbanization and industrialization, air pollution has become a significant issue affecting urban ecological environment, residents' health, and public governance [1-2]. The Air Quality Index (AQI), as a core indicator for measuring the degree of air pollution, plays a crucial role in pollution early warning and environmental management. Wuhan, with its dense population and frequent traffic activity, experiences interactions between industrial emissions, meteorological conditions, and regional transport, resulting in AQI sequences exhibiting significant nonlinearity [3-4], non-stationarity, and sudden fluctuations. Therefore, constructing a high-precision and stable AQI prediction model is of great importance for improving urban pollution early warning capabilities and environmental governance.

In recent years, air quality prediction methods have gradually evolved from traditional statistical models to machine learning, deep learning, and hybrid prediction models. Machine learning methods have improved nonlinear modeling capabilities but are insufficient in characterizing time dependencies. Meanwhile, although deep learning methods can learn temporal features, they are still prone to problems such as smoothing prediction curves, insufficient peak tracking, and large errors during high-pollution phases when facing strong fluctuations, strong noise, and sudden pollution events. Therefore, introducing sequence decomposition, nonlinear feature extraction, and residual correction mechanisms into air quality prediction has become an important direction for improving model performance.

Existing research still has certain shortcomings. First, AQI sequences typically contain multiple components, including trends, periods, short-term disturbances, and random noise, making direct modeling prone to interference between different frequency information. Second, there is a complex nonlinear coupling relationship between pollutants and meteorological variables, making it difficult for traditional linear dimensionality reduction methods such as Principal Component Analysis (PCA) to fully preserve key features [5-6]. Third, single deep learning models are prone to systematic biases in medium-to-high pollution ranges and periods of severe fluctuation. Furthermore, existing research has largely focused on overall error indices, with less analysis of the distribution characteristics, autocorrelation structure, and state transition patterns of prediction residuals, resulting in insufficient explanation of the intrinsic reasons for model performance improvements.

To address the aforementioned issues, this paper proposes a hybrid prediction model that integrates Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Kernel Principal Component Analysis (KPCA), Temporal Convolutional Network (TCN), Bidirectional Gated Recurrent Unit (BiGRU), and Adaptive Markov Residual Correction (AMRC), namely the CEEMDAN-KPCA-TCN-BiGRU-AMRC model. This model first utilizes CEEMDAN to perform multi-scale decomposition of the air quality index series, reducing the impact of non-stationary features on prediction. Then, KPCA is used to extract nonlinear principal features from pollutants and meteorological variables. Next, the TCN-BiGRU model learns local fluctuations and long-term dependencies to obtain initial prediction results. Finally, the AMRC mechanism is introduced to dynamically correct prediction biases under different conditions. The main contributions of this paper are: constructing a hybrid prediction framework that integrates sequence decomposition, nonlinear dimensionality reduction, deep temporal modeling, and residual state correction; enhancing the ability to express complex features using CEEMDAN and KPCA; improving prediction stability during medium-to-high pollution and volatile phases through AMRC; and conducting multi-dimensional validation of the model's effectiveness by combining CEEMDAN decomposition diagrams, variable distribution combination diagrams, KPCA clustering structure diagrams, locally magnified prediction diagrams, residual kernel density estimation (KDE) diagrams, residual autocorrelation function (ACF) diagrams, and state transition heatmaps.

2. Methodology

2.1. CEEMDAN Decomposition

The AQI sequence usually contains multiple components, including long-term trends, periodic fluctuations, short-term disturbances, and random noise. Therefore, it exhibits obvious nonlinear and non-stationary characteristics. If the original AQI sequence is directly fed into a deep learning model, information at different frequency scales may be mixed together, which increases the learning difficulty of the model and weakens its ability to capture local mutations and high-frequency fluctuations. Therefore, this study first introduces CEEMDAN to conduct multi-scale decomposition of the original AQI sequence, so as to reduce the complexity of the sequence and separate variation information at different temporal scales [7].

CEEMDAN is an adaptive signal decomposition method developed on the basis of Empirical Mode Decomposition (EMD) and Ensemble Empirical Mode Decomposition (EEMD) [8]. Compared with traditional EMD, CEEMDAN introduces adaptive white noise and performs ensemble averaging several times, which can effectively alleviate the mode mixing problem. Compared with EEMD, CEEMDAN can obtain more complete Intrinsic Mode Functions (IMFs) at each decomposition stage, thereby improving the stability and reconstruction accuracy of the decomposition results.

Let the original AQI sequence be denoted as x(t). After CEEMDAN decomposition, it can be expressed as:

$$x(t) = \sum_{i=1}^{K} IMF_i(t) + R(t) \tag{1}$$

where $IMF_i(t)$ denotes the i-th Intrinsic Mode Function, K represents the number of decomposed IMF components, and $R(t)$ denotes the residual trend term [9-10]. In general, the first few IMF components mainly reflect high-frequency disturbances and short-term random fluctuations in the AQI sequence. The middle IMF components mainly represent periodic variations and stage-wise fluctuations, while the later IMF components and the residual term mainly reflect long-term trend information. Through this decomposition process, the original complex sequence is transformed into multiple relatively stable sub-sequences with clearer frequency characteristics, which helps the subsequent model learn variation patterns at different scales.

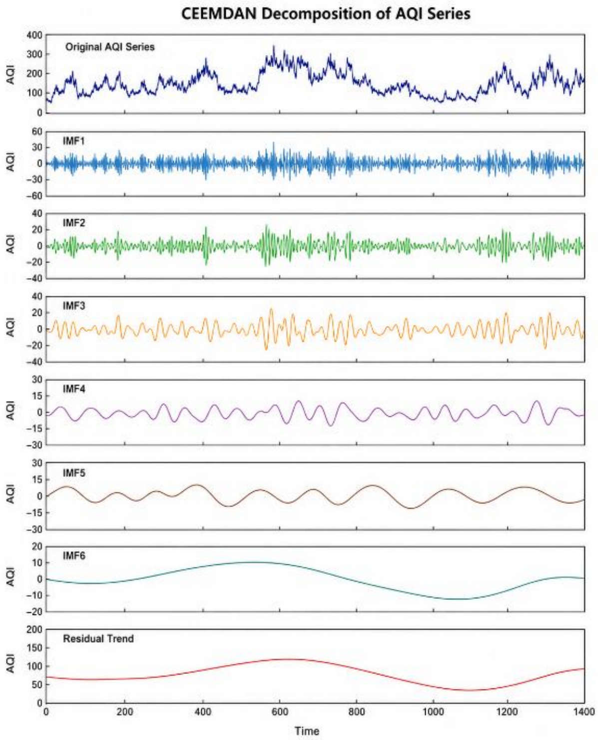


Figure 1: Visualization of decomposition results.

In this paper, the CEEMDAN decomposition results will be presented visually, as shown in Figure 1. The top of the figure shows the original AQI sequence, the middle shows several IMF components, and the bottom shows the residual trend term. This figure allows for a direct observation of the AQI sequence's variation characteristics at different frequency scales. High-frequency IMF mainly corresponds to short-term fluctuations, low-frequency IMF mainly corresponds to periodic changes, and the residual trend term reflects the long-term evolution trend of the air quality index.

2.2. Nonlinear Feature Extraction Based on Kernel Principal Component Analysis

Air quality forecasting involves various pollutant indicators and meteorological variables, such as PM2.5, PM10, sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone (O_3), temperature, humidity, and sunshine duration [11-12]. These variables are not simply linearly related but exhibit significant correlations, redundancy, and nonlinear coupling. Directly using all original variables as input to a deep learning model increases the number of model parameters and training complexity. Furthermore, the introduction of redundant information and noise also affects prediction performance.

Traditional PCA can extract key variable information through linear transformations. However, PCA can only characterize linear correlation structures and cannot fully represent the complex nonlinear features in air quality data. Therefore, this study uses KPCA to perform nonlinear dimensionality reduction on multi-source influencing variables. The basic idea of KPCA is to map the original low-dimensional input to a high-dimensional feature space using a kernel function, and then perform principal component analysis in the high-dimensional space to extract nonlinear principal component features.

Let the original input feature matrix be:

$$X = \{x_1, x_2, \ldots, x_n\} \tag{2}$$

where x_i represents the multi-source feature vector of the n-th sample. KPCA first maps the original data into a high-dimensional feature space through a nonlinear mapping function $\phi(\cdot)$:

$$x_i \rightarrow \phi(x_i) \tag{3}$$

In the high-dimensional feature space, the inner product between samples can be expressed by a kernel function:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{4}$$

In this study, the Radial Basis Function (RBF) kernel can be adopted as the kernel function of KPCA, which is expressed as [13-14]:

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \tag{5}$$

where γ is the kernel parameter used to control the decay rate of sample similarity. By constructing the kernel matrix and performing eigenvalue decomposition, several nonlinear principal components can be obtained. According to the cumulative contribution rate criterion, the first several principal components that retain the main information are selected as the input features of the subsequent TCN-BiGRU model.

2.3. TCN-BiGRU Forecasting Model

TCN is a convolutional neural network structure suitable for sequence modeling tasks. Unlike traditional recurrent neural networks, TCN captures temporal dependencies mainly through causal convolution and dilated convolution. Causal convolution ensures that the prediction at the current time step only depends on current and historical information, thereby avoiding future information leakage. Dilated convolution can expand the receptive field without significantly increasing the number of parameters, enabling the model to capture long-range historical dependencies.

Let the input sequence be denoted as $X = \{x_1, x_2, \ldots, x_T\}$. The one-dimensional dilated convolution in TCN can be expressed as:

$$F(t)=\sum_{k=0}^{K-1} f(k)\cdot x_{t-d\cdot k} \tag{6}$$

where $F(t)$ denotes the convolution output at time t, $f(k)$ represents the convolution kernel parameter, K is the kernel size, and d is the dilation factor. As the dilation factor increases, TCN can gradually expand its receptive field layer by layer, thus learning short-term disturbances, stage-wise changes, and local trend features in the AQI sequence. In addition, TCN is usually combined with residual connections to alleviate the gradient vanishing problem during deep network training and improve model stability.

While TCN effectively extracts local temporal patterns, its reliance on convolutional structures for feature modeling limits its ability to capture dynamic relationships between sequences. To further enhance the model's ability to learn temporal dependencies, this study introduces a BiGRU after TCN. A GRU is an improved recurrent neural network architecture that controls the retention and forgetting of historical information through update and reset gates.

The basic update process of a GRU can be represented as:

$$z_t=\sigma(W_z x_t+U_z h_{t-1}+b_z) \tag{7}$$

$$r_t=\sigma(W_r x_t+U_r h_{t-1}+b_r) \tag{8}$$

$$\tilde{h}_t=\tanh(W_h x_t+U_h(r_t\odot h_{t-1})+b_h) \tag{9}$$

$$h_t=(1-z_t)\odot h_{t-1}+z_t\odot\tilde{h}_t \tag{10}$$

where z_t is the update gate, r_t is the reset gate, $\tilde{h}_t$ is the candidate hidden state, h_t is the current hidden state, $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\odot$ represents element-wise multiplication. BiGRU extracts sequence information through forward and backward GRU branches and concatenates the hidden states from both directions:

$$h_t=[\overrightarrow{h_t};\overleftarrow{h_t}] \tag{11}$$

where $\overrightarrow{h_t}$ denotes the forward hidden state and $\overleftarrow{h_t}$ denotes the backward hidden state. Through this structure, the model can more fully learn the variation patterns of the AQI sequence from different temporal directions.

The proposed TCN-BiGRU forecasting model first uses TCN to extract local fluctuation features and multi-scale temporal patterns from the input sequence. Then, the high-dimensional temporal features output by TCN are fed into BiGRU to further learn long-term dependencies and dynamic evolution characteristics. Finally, the initial AQI prediction is generated through a fully connected layer. This structure combines the local pattern extraction capability of TCN and the temporal dependency modeling capability of BiGRU, which can effectively improve the model’s ability to fit AQI fluctuation trends and peak variations.

2.4. AMRC-Based Adaptive Markov Residual Correction

The core idea of Adaptive Markov Residual Correction (AMRC) is to regard the prediction error that cannot be fully explained by the deep learning model as a residual signal with state transition characteristics [15], and to model its evolution using a Markov chain. For air quality forecasting, the error distribution of the model is often inconsistent across different AQI levels. For example, in low-pollution intervals, the prediction error may be small and stable, whereas in medium-to-high pollution intervals or rapidly changing periods, the model may produce obvious underestimation or lagging effects. This indicates that prediction residuals have certain interval differences and state dependencies [16-17].

Let the real AQI value be y_t, and the initial prediction value output by the TCN-BiGRU model be $\hat{y}_t$. The prediction residual is defined as:

$$e_t = y_t - \hat{y}_t \tag{12}$$

where $e_t$represents the prediction residual at time t. If the residual sequence exhibits a certain transition pattern between adjacent time steps, it can be divided into several state intervals, and a Markov chain can be used to describe the transition probabilities among residual states.

Let the residual state set be defined as:

$$S = \{S_1, S_2, \dots, S_m\} \tag{13}$$

where $m$denotes the number of residual states. According to the distribution characteristics of residuals in the training set, the residuals can be divided into different states, such as large negative error, small negative error, near-zero error, small positive error, and large positive error. If the residual state at time $t$is S_i, and the residual state at time $t+1$is S_j, the state transition probability can be expressed as:

$$p_{ij} = P(e_{t+1} \in S_j \mid e_t \in S_i) \tag{14}$$

All state transition probabilities form the Markov state transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \tag{15}$$

where the sum of elements in each row of the matrix is equal to 1. This matrix describes the probability distribution of residual transitions from the current state to the next state[18-19].

After obtaining the state transition matrix, this study further calculates the expected residual correction value for the next time step. Let $c_j$be the representative value of the j-th residual state. When the current residual state is S_i, the residual correction term for the next time step can be expressed as:

$$\hat{e}_{t+1} = \sum_{j=1}^{m} p_{ij} c_j \tag{16}$$

The final corrected prediction value is then given by:

$$\hat{y}_{t+1}^{AMRC} = \hat{y}_{t+1} + \hat{e}_{t+1} \tag{17}$$

where $\hat{y}_{t+1}^{AMRC}$ denotes the final prediction after AMRC correction, $\hat{y}_{t+1}$ denotes the initial prediction output by TCN-BiGRU, and $\hat{e}_{t+1}$denotes the residual correction term estimated based on Markov state transition. The overall technical route is shown in Figure 2.

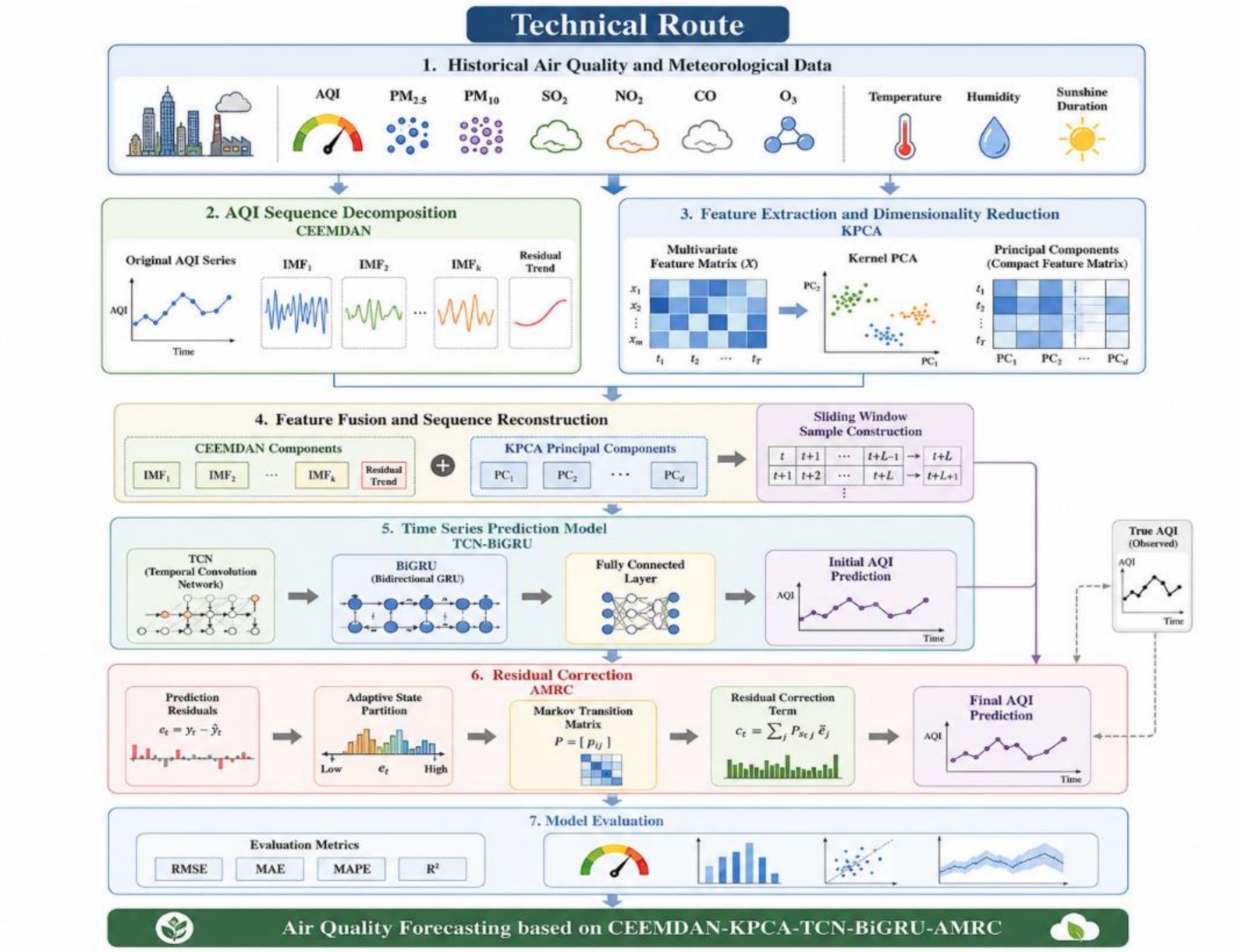


Figure 2: Overall technology roadmap.

3. Experimental Setup

3.1. Experimental Environment

The experiments in this paper were conducted on a local computing platform with an Intel Core i9-14900HX processor, 16 GB of RAM, and Windows 11 operating system. Model training and result analysis were performed in the same experimental environment to ensure the comparability of results from different models.

This paper utilizes the Python platform for model construction and experimental analysis. Data preprocessing, feature extraction, and evaluation metric calculation primarily employed NumPy, Pandas, and Scikit-learn packages; the CEEMDAN decomposition process was implemented using PyEMD methods; and the deep learning model was implemented using the PyTorch framework.

3.2. Data Division and Evaluation Indicators

This paper divides the air quality data according to time sequence, with the first 80% of the data used as the training set and the last 20% used as the test set.

The evaluation indicators adopted are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{19}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{21}$$

where $y_i$denotes the actual value, $\hat{y}_i$denotes the predicted value, $\bar{y}$denotes the mean of the actual values, and $n$denotes the number of test samples.

3.3. Model Comparison

To verify the predictive performance of the proposed model, this paper selects several typical time series prediction models as comparison models, including single recurrent neural network models, convolutional time series models, and models combining different modules. The specific comparison models are as follows:

LSTM; GRU; BiGRU; TCN; TCN-BiGRU; CEEMDAN-TCN-BiGRU; KPCA-TCN-BiGRU; CEEMDAN-KPCA-TCN-BiGRU; CEEMDAN-KPCA-TCN-BiGRU-AMRC. These models represent the complete model proposed in this paper and are used to verify the improvement effect of the adaptive Markov residual correction mechanism on the final prediction accuracy.

Through the above model comparisons, the impact of each module on the air quality index prediction performance can be systematically analyzed from three levels: single model, combined model, and complete model.

4. Results and Discussion

4.1. Predictive Performance Comparison Analysis

Table 1: Comparison of prediction performance of different models.

Model	RMSE	MAE	MAPE / %	R^2
LSTM	27.526	21.368	35.897	0.229
GRU	26.842	20.754	34.621	0.267
BiGRU	24.916	19.203	31.845	0.351
TCN	23.684	18.426	29.736	0.412
TCN-BiGRU	21.357	16.842	27.184	0.506
CEEMDAN-TCN-BiGRU	18.936	14.725	23.518	0.621
KPCA-TCN-BiGRU	19.482	15.316	24.203	0.598
CEEMDAN-KPCA-TCN-BiGRU	16.275	12.608	20.146	0.711
CEEMDAN-KPCA-TCN-BiGRU-AMRC	13.824	10.157	16.932	0.812

To verify the effectiveness of the proposed method in air quality index prediction, the model described above was selected as a comparison model. All models used the same data partitioning method and evaluation index, and the experimental results are shown in Table 1.

After introducing CEEMDAN decomposition, the RMSE, MAE, and MAPE of the CEEMDAN-TCN-BiGRU model decreased to 18.936, 14.725, and 23.518%, respectively, indicating that CEEMDAN can decompose the original air quality index series into subsequences at different frequency scales, thereby reducing the interference of nonstationarity on the prediction process. The prediction performance of the KPCA-TCN-BiGRU model is also better than that of TCN-BiGRU, indicating that KPCA can extract more compact nonlinear principal features from pollutants and meteorological variables, reduce input redundancy, and improve the model's feature representation ability.

When CEEMDAN and KPCA are introduced simultaneously, the prediction performance of the CEEMDAN-KPCA-TCN-BiGRU model is further improved, with the RMSE decreasing to 16.275 and the R^2 increasing to 0.711. This demonstrates that sequence decomposition and nonlinear feature extraction are complementary; the former enhances the model's ability to characterize multi-scale fluctuations in the target sequence, while the latter improves the nonlinear representation of multi-source influencing factors. Ultimately, the proposed CEEMDAN-KPCA-TCN-BiGRU-AMRC complete model achieves optimal results across all evaluation metrics, with RMSE, MAE, and MAPE decreasing to 13.824, 10.157, and 16.932%, respectively, and R^2 increasing to 0.812. This indicates that the adaptive Markov residual correction mechanism can further reduce prediction bias and improve the model's explanatory power for air quality index variations.

Figure 3 shows the comparison between the actual values of the test set and the predicted values of each model. It should include a main plot of the complete test interval, a magnified plot of the local fluctuation interval, and an error shaded area. The main plot is used to show the ability of different models to track the overall trend of air quality index changes; the magnified plot is used to highlight the local response capability of the model when the AQI rises rapidly, falls rapidly, or reaches a peak; the error shaded area is used to visually reflect the range of deviation between the predicted value and the actual value.

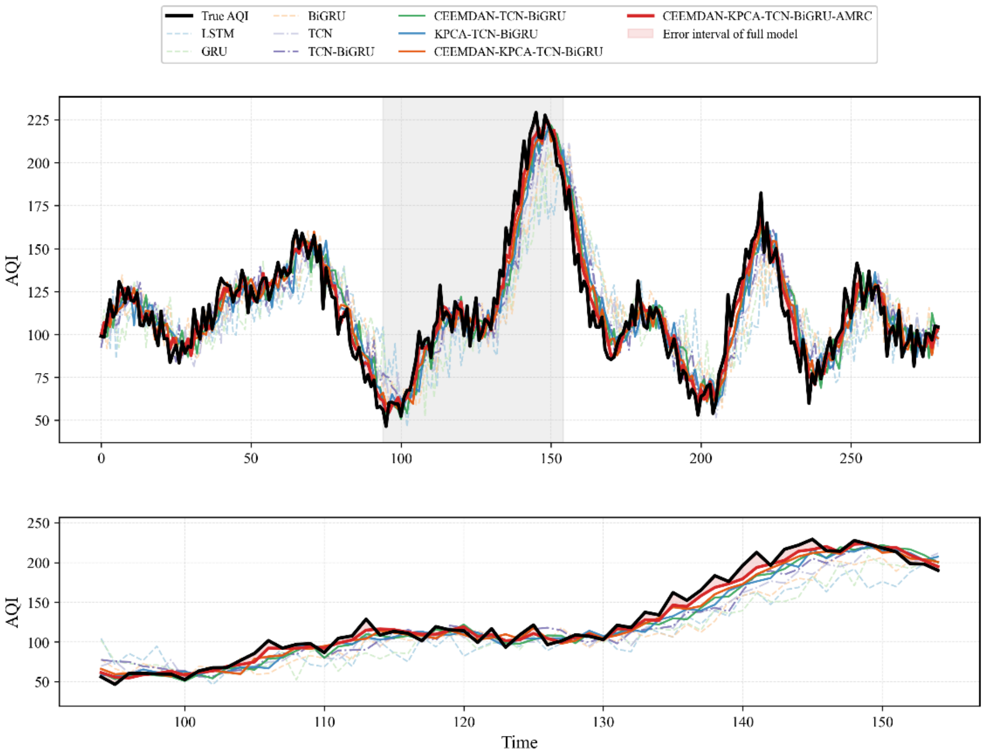


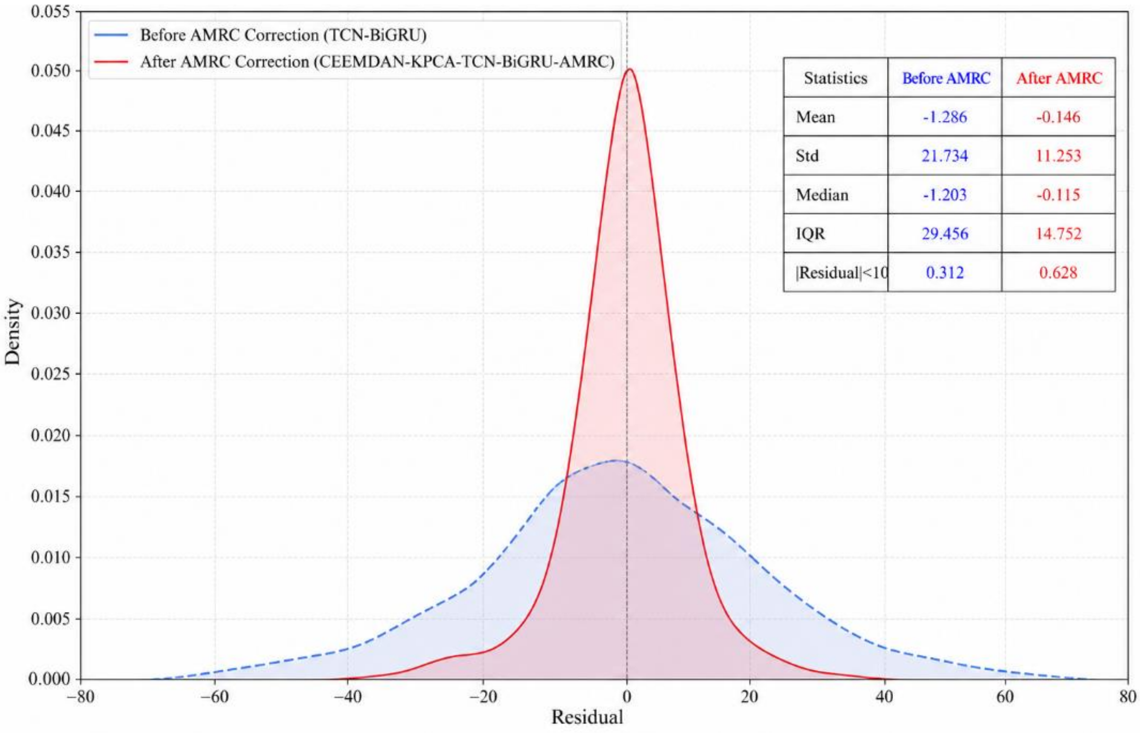
Figure 3: Comparison chart of actual test set values and prediction results from different models.

As shown in Figure 3, the CEEMDAN-TCN-BiGRU and KPCA-TCN-BiGRU models further improved the prediction performance compared to TCN-BiGRU. Specifically, CEEMDAN-TCN-BiGRU performed better in areas with frequent fluctuations, indicating that sequence decomposition effectively reduces the mutual interference between different frequency components in the original sequence. KPCA-TCN-BiGRU showed greater stability in fitting the overall trend, suggesting that nonlinear dimensionality reduction helps extract key features related to air quality changes. The CEEMDAN-KPCA-TCN-BiGRU model combines the advantages of both, with its predicted curve closer to the actual value, but some residuals still exist near local peaks.

After AMRC correction, the predicted curve of the CEEMDAN-KPCA-TCN-BiGRU-AMRC model is closest to the actual AQI curve, especially in the medium-to-high pollution range and during local peak stages, where the model can follow changes in the actual value more promptly, and the prediction lag is significantly reduced.

4.2. Residual Distribution Analysis

To further analyze the effect of the adaptive Markov residual correction mechanism on the improvement of prediction error, this paper performs kernel density estimation analysis on the prediction residuals before and after correction, as shown in Figure 4. The residual distribution can intuitively reflect the concentration and dispersion of prediction error. The closer the residuals are to zero and the more concentrated the distribution, the more stable the model prediction results and the smaller the error fluctuations.



Statistics	Before AMRC	After AMRC
Mean	-1.286	-0.146
Std	21.734	11.253
Median	-1.203	-0.115
IQR	29.456	14.752
|Residual|<10	0.312	0.628

Figure 4: Residual distribution.

As shown in Figure 4, the residual distribution before AMRC correction is relatively dispersed, with a low peak value and long tails, indicating that there are still a certain number of large error samples in the initial prediction results. This shows that although TCN-BiGRU can learn the main variation patterns of the AQI sequence, it still produces significant deviations in some drastic fluctuation stages. After AMRC correction, the residual distribution curve clearly concentrates near zero error, the peak value increases, and the tail shortens, indicating that the number of large error samples decreases and the overall prediction error is more stable.

This result shows that the AMRC module can effectively utilize residual state transition information to compensate for systematic deviations in the initial prediction results. Especially during periods of rapid AQI changes or high pollution levels, the residual correction mechanism can reduce the dispersion of model prediction errors, making the final prediction results closer to the true values.

4.3. Residual Autocorrelation Analysis

While residual distribution alone can illustrate changes in error magnitude and concentration, it doesn't fully reflect whether a time-dependent structure still exists in the residual sequence. Therefore, this paper further analyzes the residual sequences before and after correction using the autocorrelation function, as shown in Figure 5. The autocorrelation function reflects the correlation of residuals at different lags. If the residuals still exhibit strong correlation across multiple lags, it indicates that the model has not fully extracted the predictable information from the sequence.

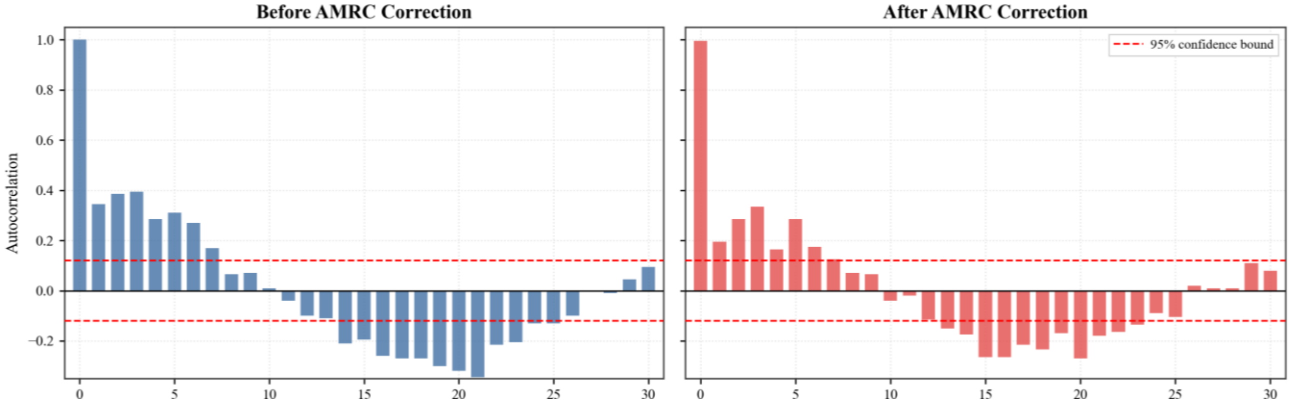


Figure 5: Autocorrelation function of residuals before and after AMRC correction.

Figure 5 shows that before AMRC correction, the residual sequence still exhibits significant autocorrelation at some lags, indicating that the error output by the initial prediction model is not completely random but contains a certain time-dependent structure. This also proves the necessity of further modeling and correcting the residuals. After AMRC correction, the overall residual autocorrelation coefficient decreases, and the correlation at most lags weakens significantly, indicating that the predictable structure in the residuals is further eliminated.

These results demonstrate that the AMRC module not only reduces the residual amplitude but also weakens the time correlation in the residual sequence, making the corrected error closer to random error. In other words, after residual correction, the model has further utilized error information that the original deep learning model failed to capture, thereby improving the reliability of the overall prediction results.

4.4. Residual State Transition Analysis

To further explain the mechanism of the AMRC module, this paper constructs a residual state transition probability matrix and displays it using a heatmap, as shown in Figure 6. The horizontal axis represents the residual state at the next time step, the vertical axis represents the residual state at the current time step, and the color intensity represents the magnitude of the state transition probability. This graph allows observation of the transition patterns of the prediction residual between different states.

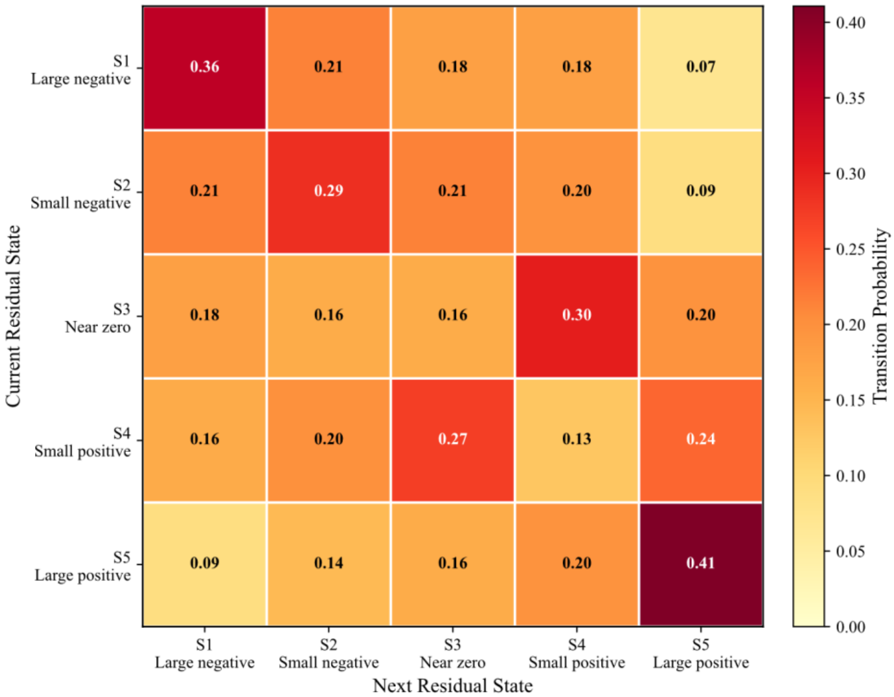


Figure 6: Markov transition probability heatmap of residual states.

As shown in Figure 6, the transitions between residual states are not completely random but exhibit a certain degree of state dependence. Some states have a high probability of maintaining their original state at the next time step, indicating a certain degree of state persistence in the residuals. The transition probabilities between adjacent states are also relatively large, indicating that the residuals usually exhibit gradual changes rather than completely random jumps. This state transition characteristic suggests that the prediction residuals do indeed contain usable dynamic information.

The AMRC module, based on this state transition law, estimates the expected value of the residuals at the next time step and incorporates it as a correction term into the initial prediction. In this way, the model can further compensate for systematic biases based on the original TCN-BiGRU prediction results, especially improving prediction performance during medium-to-high pollution and drastic fluctuation phases. Therefore, the residual state transition heatmap not only verifies the rationality of Markov residual correction but also explains the reason for the improved prediction performance of the complete model at a mechanistic level.

5. Conclusion

This paper addresses the problems in AQI forecasting, such as strong nonlinearity, significant nonstationarity, difficulty in capturing sudden fluctuations, and insufficient utilization of residual structure. A hybrid forecasting model, CEEMDAN-KPCA-TCN-BiGRU-AMRC, is proposed. First, the model utilizes adaptive noise-complete ensemble empirical mode decomposition to perform multi-scale decomposition of the original AQI sequence, splitting the complex sequence into intrinsic mode components and residual trend terms at different frequency scales to reduce the interference of sequence nonstationarity on the forecasting process. Then, kernel principal component analysis is used to extract nonlinear features from pollutants and meteorological variables, reducing input variable redundancy and enhancing the expressive power of key features. Based on this, a deep time-series forecasting model is constructed using a temporal convolutional network and a bidirectional gated recurrent unit, simultaneously learning local fluctuation characteristics and long-term dependencies to obtain initial forecast results. Finally, an adaptive Markov residual correction mechanism is introduced to model the state transition law of the forecast residuals and dynamically compensate for the initial forecast results, thus forming a complete AQI forecasting framework.

Experimental results show that the proposed CEEMDAN-KPCA-TCN-BiGRU-AMRC model outperforms other single models and various combined improved models in prediction performance. Compared with single deep learning models, the proposed model can better characterize the overall trend of AQI series changes and exhibits stronger tracking ability in medium-to-high pollution ranges, local peak stages, and periods of severe fluctuations. The prediction result comparison chart shows that the prediction curve of the complete model is closer to the true value, and the error shadowing interval is significantly reduced. The residual kernel density estimation chart shows that after adaptive Markov residual correction, the residual distribution is more concentrated near zero error, and the tail is significantly shortened. Residual autocorrelation function analysis further illustrates that the temporal correlation of the corrected residuals is weakened. The residual state transition heatmap verifies the existence of exploitable state transition patterns in the prediction residuals. These results collectively demonstrate that the proposed method not only reduces the overall prediction error but also improves the stability and reliability of the model prediction from the residual structure level.

Overall, the hybrid prediction model constructed in this paper effectively improves the accuracy of AQI prediction and provides a feasible method for time series prediction in complex environments. However, this study still has some limitations. For example, the data sources are mainly concentrated in a single region, and more external factors such as regional pollution transport, emission inventories, and spatially adjacent sites have not been fully considered. Furthermore, the adaptive Markov residual correction is still a post-processing correction method and has not yet been integrated with the deep prediction model for end-to-end joint training. Future research could further integrate heterogeneous data from multiple cities, sites, and sources, and attempt to introduce Transformer structures or end-to-end residual correction mechanisms to further improve the model's generalization ability and practical application value in complex pollution scenarios.

Acknowledgements

This research was funded by China University of Labor Relations, grant number “25xjx031" and "The APC was funded by China University of Labor Relations".

References

[1] Saxena V. Water quality, air pollution, and climate change: investigating the environmental impacts of industrialization and urbanization[J]. Water, Air, & Soil Pollution, 2025, 236(2): 73.

[2] Das S. Unraveling the urban climate crisis: Exploring the nexus of urbanization, climate change, and their impacts on the environment and human well-being–A global perspective[J]. AIMS Public Health, 2024, 11(3): 963.

[3] Shreyas D. Deciphering seasonal dynamics and time series forecasting of urban air quality: a case study in Urban Cities of Andhra Pradesh[J]. Scientific Reports, 2026.

[4] Nguyen A T. Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization[J]. Journal of big data, 2024, 11(1): 71.

[5] Serani A, Diez M. A survey on design-space dimensionality reduction methods for shape optimization[J]. Archives of Computational Methods in Engineering, 2026, 33(2): 1671-1698.

[6] Zheng Z. Response to letter to the editor from Y. Takefuji on “Beyond principal component analysis: Enhancing feature reduction in electronic noses through robust statistical methods”[J]. Trends in Food Science & Technology, 2025, 157: 104918.

[7] He S. A CEEMDAN-GNN-transformer hybrid model for air quality index forecasting: a case study of Chang’an town, Dongguan, China[J]. Frontiers in Environmental Science, 2025, 13: 1653446.

[8] Deng Y, Xu T, Sun Z. A hybrid multi-scale fusion paradigm for AQI prediction based on the secondary decomposition[J]. Environmental Science and Pollution Research, 2024, 31(22): 32694-32713.

[9] Ghazali M F, Yusop M H, Suyono H. A hybrid hilbert–huang and integrated kurtosis framework for automated leak

detection and characterisation in pipeline systems[J]. Measurement, 2026: 122324.
[10] Kameshwar Reddy R. Shivaprakash (2025) Disentangling Climatic Volatility from Long-Term Productivity Trends in Semi-Arid Agriculture: An Empirical Mode Decomposition Approach[J]. Indian Journal of Science and Technology, 2025, 18(43): 3445-3455.
[11] Rawat A. Advancing air pollution forecasting: a review of physical, statistical, and machine learning methods[J]. Environmental Science and Pollution Research, 2026, 33(16): 7373-7398.
[12] Al-Rousan N, Al-Najjar H, Elhaty I A. Machine learning framework for forecasting air pollution: Evaluating seasonal and climatic influences in Istanbul, Turkey[J]. PLoS One, 2025, 20(10): e0330716.
[13] Louifi A. Enhancing fault detection in stochastic environments using interval-valued KPCA: a cement rotary kiln case study[J]. IEEE Access, 2025, 13: 102951-102963.
[14] Madukpe V N, Ugoala B C, Zulkepli N F S. Topological Approach and Kernel Principal Component Analysis for Air Pollution Source Apportionment[J]. International Journal of Environmental Research, 2025, 19(6): 260.
[15] Varshney R P, Sharma D K. Optimizing Time-Series forecasting using stacked deep learning framework with enhanced adaptive moment estimation and error correction[J]. Expert Systems with Applications, 2024, 249: 123487.
[16] LI Y, WANG Z, YANG A, et al. Integrating evolutionary algorithms and enhanced-YOLOv8+ for comprehensive apple ripeness prediction[J]. Scientific Reports, 2025, 15: 7307. DOI: 10.1038/s41598-025-91939-4.
[17] LI Y, SONG Z, XIA L, et al. Application of the adaptive sparrow search algorithm in medical supply engineering[J]. Scientific Reports, 2025, 15: 35775. DOI: 10.1038/s41598-025-11793-2.
[18] FU Z, GAO F, LI Y, et al. How compound climate shocks reshape urban resilience and scenario-adaptive pathways: New insights from the Yangtze River Economic Belt[J]. Environmental Impact Assessment Review, 2026, 121: 108563. DOI: 10.1016/j.eiar.2026.108563.
[19] CHEUNG Y, GUO Z, LI Y. Foliated generative adversarial networks for rolling bearings fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2026, 75: 1-24. DOI: 10.1109/TIM.2026.3684644