Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Optimal Feature Extraction and Prediction for Air Quality Using a Swarm-Refined Hybrid System

**Jiahao Guo**
Central South University of Forestry and Technology
**Yingfang Zhu**
zhuyingfang@csuft.edu.cn
Central South University of Forestry and Technology
**Huihua Huang**
Central South University of Forestry and Technology
**Juyang Liao**
Hunan Botanical Garden
**Hao Huang**
Central South University of Forestry and Technology

**Article**

**Keywords:** Air quality prediction, Time series analysis, SARIMA, Temporal Convolutional Networks (TCN), Sand Cat Swarm Optimization (SCSO), AQI

**Posted Date:** September 7th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10787872/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Optimal Feature Extraction and Prediction for Air Quality Using a Swarm-Refined Hybrid System

**Jiahao Guo[a], Yingfang Zhu[a,*], Huihua Huang[a], Juyang Liao[b], Hao Huang[a]**

[a] College of Computer Science and Mathematics, Central South University of Forestry and Technology, Changsha, 410004, China

[b] Hunan Changsha-Zhuzhou-Xiangtan City Cluster National Research Station of Ecosystem, Hunan Botanical Garden, Changsha, 410116, China

E-mail: coamp0604@gmail.com, zhuyingfang@csuft.edu.cn, huanghuihua@csuft.edu.cn, liaojuyang@163.com, 13973113175@163.com

**Keywords:** Air quality prediction, Time series analysis, SARIMA, Temporal Convolutional Networks (TCN), Sand Cat Swarm Optimization (SCSO), AQI

**Abstract.** Air quality prediction plays a crucial role in environmental governance, public health protection, and urban planning optimization. However, due to the increasing complexity of pollution sources and the dynamic evolution of climate conditions, traditional single models are often inadequate in capturing both linear trends and nonlinear variations in atmospheric data. This study proposes a hybrid model integrating the Seasonal Autoregressive Integrated Moving Average (SARIMA), Temporal Convolutional Networks (TCN), and Sand Cat Swarm Optimization (SCSO) algorithms. The model firstly extracts the seasonal and long-term trends of air quality data by SARIMA with optimal parameter combinations, then inputs the residual information into TCN to capture the complex nonlinear dynamics, and finally globally optimizes the TCN hyper-parameters by the SCSO algorithm to enhance the model robustness. Empirical results based on multi-site monitoring data indicate that the hybrid model significantly outperforms traditional SARIMA, with RMSE reduced by 52.8% and $R^2$ improved by 51.9%. This study offers effective methodological support for regional air quality monitoring and pollution control, and provides a solid theoretical foundation for further research.

## Introduction

In the global environmental protection strategy, accurately predicting the changing trend of air quality has become an important goal to achieve effective pollution prevention and control. With the rapid development of urbanization and industrialization, the problem of air pollution is becoming more and more serious. The Air Quality Index (AQI) is a comprehensive indicator of air quality, which quantitatively reflects the severity of air pollution by combining the concentration of various pollutants in the atmosphere, such as $PM2.5$, $PM10$, $NO_2$, $SO_2$, $O_3$ and $CO$. AQI not only directly affects public health

and ecological security, but its accurate prediction is of great significance for formulating pollution prevention and control measures in advance and optimizing environmental governance policies.

Currently, AQI forecasting is mainly centered on statistical methods, machine learning methods and deep learning methods. Traditional statistical methods are widely used in this field, among which the Autoregressive Integrated Moving Average model (ARIMA) and its improved models dominate in time series forecasting. For example, Muzakki *et al.* [1] used the ARIMAX model combined with meteorological factors to forecast the AQI of Jakarta, verifying the important role of meteorological data in improving the prediction accuracy; Du *et al.* [2] applied the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to forecast urban air quality in China, demonstrating its applicability to AQI series with clear seasonal variations. Similarly, Ao *et al.* [3] analyzed the seasonal characteristics of PM2.5 concentrations in Hefei and applied the SARIMA model to forecasting, demonstrating the applicability of SARIMA to PM2.5 forecasting with pronounced seasonal patterns. In particular, Singh *et al.* [4] used the Holt-Winters Exponential Smoothing (Holt-Winters) and Long Short-Term Memory (LSTM) methods to forecast the AQI of the Indian region. The results show that in the short-term forecasting scenario, the traditional Holt-Winters model performs better, while LSTM fails to significantly outperform this traditional method. Although statistical methods have advantages in characterizing linear trends and seasonal changes, they have shortcomings in capturing complex nonlinear relationships and long-term dependencies.

To overcome this deficiency, researchers have introduced machine learning methods to model nonlinear relationships more efficiently. For example, Pradhan *et al.* [5] proposed a hybrid model employing Autoregressive Fractionally Integrated Moving Average (ARFIMA) and Support Vector Machine (SVM), which effectively combines the stability of statistical models and the nonlinear modeling ability of machine learning; Liu *et al.* [6] optimized the initial parameters of Kernel Extreme Learning Machine (KELM) through Genetic Algorithm (GA), solved the performance fluctuation problem caused by traditional KELM random weights, and realized real-time prediction while reducing computing costs; Umoh *et al.* [7] applied K-Nearest Neighbors (KNN) model to predict AQI in several cities in Nigeria, and the results show that the method has strong regional adaptability. In addition, Liu *et al.* [8] combined meteorological and pollutant data and used the Light Gradient Boosting Machine (LightGBM) model to achieve high-precision classification prediction of AQI. However, the performance of machine learning methods is still limited by feature engineering and parameter optimization, which limits the generalization ability of the model.

In recent years, deep learning methods have received widespread attention for their advantages in capturing complex nonlinear relationships. Unnikrishnan & Rajeswari [9] used a combination of fuzzy time series and deep learning to improve the stability and accuracy of AQI prediction; Wu *et al.* [10] proposed a method based on Periodic Attention Network (PAN) that can efficiently capture the cyclical changes of AQI, thus

improving the accuracy of long-term prediction; Sunku *et al.* [11] used a bidirectional LSTM (Bi-LSTM) for AQI prediction, and the experimental results show that it has certain advantages in dealing with time series data. At the same time, Transformer architectures [12] have been widely applied to long-horizon time-series forecasting. Zhou *et al.* [13] proposed Informer, which substantially reduces the computational and memory overhead for long-sequence modeling via mechanisms such as ProbSparse self-attention and self-attention distillation. In the context of AQI forecasting, Feng *et al.* [14] further proposed Spatiotemporal Informer by incorporating spatiotemporal embeddings and spatiotemporal attention, thereby improving multi-site AQI forecasting accuracy and enhancing the model's ability to capture spatiotemporal dependencies. For the challenge of multi-dimensional feature fusion, Zhang *et al.* [15] innovatively introduced Wavelet Transform and Detrended Cross-Correlation Analysis (DCCA) to decompose the multiscale features of AQI data and extract the nonlinear correlations among pollutants, and finally realized high-precision prediction by LSTM; while Xia *et al.* [16] further integrated multi-site time series and satellite remote sensing images, and used Res-GCN hybrid architecture to capture spatio-temporal heterogeneity of air quality, providing a new idea for refined prediction of urban agglomeration scale. Deepan & Saravanan [17] proposed a method combining SARIMA and Transfer Learning Long Short-Term Memory (TLSTM), which significantly improved the robustness of prediction in small-sample scenarios by capturing the seasonal characteristics of time series and transferring the knowledge of cross-region data. However, deep learning methods are sensitive to hyperparameters and training configurations, and their prediction stability and generalization performance are often limited when only AQI sequences are used for prediction.

Taken together, each type of method has its advantages, but there are also shortcomings: statistical methods can effectively portray linear trends, but it is difficult to capture nonlinear and long-term dependencies; machine learning methods have improved the prediction accuracy to a certain extent, but they rely on feature engineering and parameter tuning; deep learning methods can model complex nonlinear dependencies, but their performance is often sensitive to hyperparameter choices and training configurations, and they may fail to simultaneously capture the multi-scale temporal structure inherent in time-series data. Therefore, how to combine the advantages of traditional statistical methods and deep learning models to realize synchronous capture of linear trends and nonlinear temporal relationships has become the key to improving prediction accuracy and stability.

To this end, this study proposes an innovative hybrid model, which integrates SARIMA and Temporal Convolutional Networks (TCN), and utilizes Sand Cat Swarm Optimization (SCSO) algorithm to optimize the hyperparameters of TCN, thus effectively overcoming the limitations of existing methods and significantly improving the accuracy and stability of AQI prediction.

## Method

In order to elaborate the implementation details of this hybrid model, this paper constructs the framework shown in Figure 1, which fully integrates the advantages of statistical modeling and deep learning methods, and divides the time series data prediction process into three key modules.

**SARIMA module:** Firstly, the SARIMA model with the best combination of parameters is used to model the raw air quality data, effectively capture the linear trend and seasonal fluctuations in the data, and generate the preliminary prediction values. This part mainly deals with the problem of the description of data stationarity and periodicity.

**TCN Module:** Aiming at the limitations of the SARIMA model in capturing nonlinear dynamic changes and long-time dependencies, the residuals of the SARIMA model are used as inputs, and the TCN model further extracts potential nonlinear information from the data. It employs causal convolution and dilated convolution structures, which help rapidly expand the receptive field and capture long-range dependencies, thereby enhancing the model's ability to recognize complex temporal patterns [18, 19].

**SCSO Module:** Considering that the TCN model is highly sensitive to hyper-parameter settings, to search for the optimal parameter configurations in the global range, this study introduces the SCSO algorithm. The SCSO algorithm globally optimizes the key hyperparameters of the TCN through an adaptive strategy to effectively reduce the prediction error, and to improve the model's responsiveness and the overall stability of the model to extreme events [20].

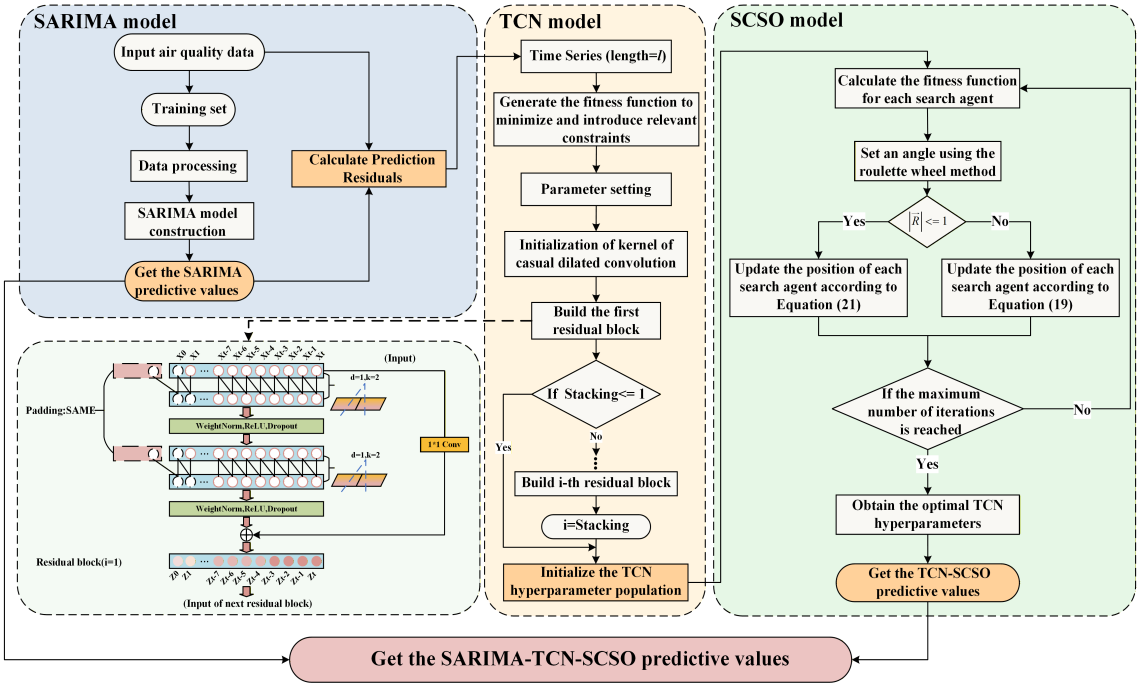



Figure 1: SARIMA-TCN-SCSO model frame diagram

By integrating the optimization modules of SARIMA, TCN, and SCSO, the framework innovatively constructs a set of complete prediction processes covering data preprocessing, nonlinear feature extraction, and global hyper-parameter tuning. The multi-model cooperative work not only effectively compensates for the limitation of a single model in the prediction of complex meteorological conditions, but also provides more accurate decision support for regional environmental management and pollution prevention and control strategies.

*SARIMA model*

The SARIMA model extends the traditional ARIMA framework to account for seasonal variations in time series data. The Box–Jenkins framework provides the classical foundation for SARIMA modeling [21]. This approach employs both non-seasonal and seasonal differencing techniques to remove underlying trends and seasonal patterns, thereby ensuring stationarity. With the transformed series meeting the stationarity prerequisites, an appropriate ARMA model can then be developed for robust forecasting [22].

When dealing with time series with cyclical fluctuations (e.g., monthly, quarterly, or weekly data), the SARIMA model introduces a seasonal difference operator based on the ARIMA model, defined as follows:

$$\Delta_s = 1 - L^s \tag{1}$$

where $s$ denotes the seasonal period and $L$ is the lag operator. The first seasonal

difference of a non-stationary time series $y_t$ can be expressed as:

$$\Delta_s y_t = (1 - L^s)\, y_t = y_t - y_{t-s} \tag{2}$$

If a seasonal difference still cannot make the data stable, it is usually necessary to carry out the $D$-order seasonal difference, that is:

$$(1 - L^s)^D y_t \tag{3}$$

For the series that reach smoothness after seasonal differencing, a seasonal ARMA model can be constructed. Its expression is as follows:

$$A_P\left(L^s\right) \Delta^D y_t = B_Q\left(L^s\right) \mu_t \tag{4}$$

where $P$ and $Q$ denote the maximum lag order of the seasonal Autoregressive (AR) term and Moving Average (MA) term, $\Delta^D$ is the seasonal difference operator, and $\mu_t$ is the white noise error term. $A_p\left(L^s\right)$ and $B_Q\left(L^s\right)$ denote the seasonal AR and MA polynomials respectively [23], with the following expressions:

$$A_P\left(L^s\right) = 1 - A_1 L^s - A_2 L^{2s} - \cdots - A_P L^{Ps} \tag{5}$$

$$B_Q\left(L^s\right) = 1 + B_1 L^s + B_2 L^{2s} + \cdots + B_Q L^{Qs} \tag{6}$$

In addition to seasonal components, time series usually contain trends or fluctuations with non-seasonal components. For a series that is smooth after non-seasonal $d$-order differencing, it can be modeled using a non-seasonal ARMA model with the following expression:

$$\phi_p\left(L\right) \Delta^d u_t = \theta_q\left(L\right) v_t \tag{7}$$

where $u_t$ and $v_t$ are the non-seasonally differenced time series and their error terms respectively. $p$ and $q$ denote the maximum lag order of the non-seasonal AR and MA terms. $\phi_p\left(L\right)$ and $\theta_q\left(L\right)$ denote the non-seasonal AR and MA polynomials respectively, with the following expressions:

$$\phi_p\left(L\right) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \tag{8}$$

$$\theta_q\left(L\right) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \tag{9}$$

Combining seasonal and non-seasonal components, the generalized expression of the SARIMA model [24] is as follows:

$$\phi_p\left(L\right) A_P\left(L^s\right) \Delta^d \Delta_s^D y_t = \theta_q\left(L\right) B_Q\left(L^s\right) v_t \tag{10}$$

The model is usually denoted as:

$$SARIMA\left(p, d, q\right) \times \left(P, D, Q\right)_s \tag{11}$$

This model can effectively deal with time series with complex trends and seasonal fluctuations, and provide a strong guarantee for air quality prediction.The detailed construction process of the model is shown in the following steps:

**Step 1:** Before constructing the SARIMA model, this study conducted a smoothness test on the AQI time series of Changsha City, Hunan Province, for the

period from 2021 to 2024 [25]. Considering the requirement of the SARIMA model for the smoothness of the series or the smoothness that can be achieved by differencing, the original AQI series was analyzed by the unit root test in this study. This method constructs the autoregressive characteristic equation, tests whether the characteristic root is located in the unit circle, and takes "the existence of a unit root" as the original hypothesis to calculate the Dickey-Fuller (DF) statistic. If the test result rejects the original hypothesis (i.e., the $p$-value is smaller than the set significance level), it means that the AQI series has smooth characteristics. Table 1 presents a detailed breakdown of the test outcomes.

Table 1: Unit root test results

| **Statistic** | ***t*-value** | ***p*-value** |
|---|---|---|
| DF Test Statistic | -4.7261 | 0.0001 |
| 1% Significance Level | -3.4349 | — |
| 5% Significance Level | -2.8635 | — |
| 10% Significance Level | -2.5678 | — |

The unit root test results reveal that the DF test statistic is considerably lower than the critical values at each significance levels, and the associated $p$-value is well below 0.05, so the original hypothesis is rejected, implying that the AQI time series does not exhibit a unit root and the series has smoothness. Therefore, when constructing the SARIMA model, there is no need for differential processing, and the differential order $d$=0 can be determined so that the original series can be modeled and forecasted directly.

**Step 2:** In the SARIMA model, the seasonal difference order $D$ is mainly used to capture and eliminate the non-stationarity in seasonal fluctuations. If the data tend to be stable after the seasonal difference treatment, the appropriate seasonal difference order $D$ can be determined. In this study, an Additive Decomposition Model is used to seasonally decompose the time series, and the data are decomposed into trend, seasonality and residual components to identify seasonal characteristics and provide a basis for the determination of the model parameters [26].

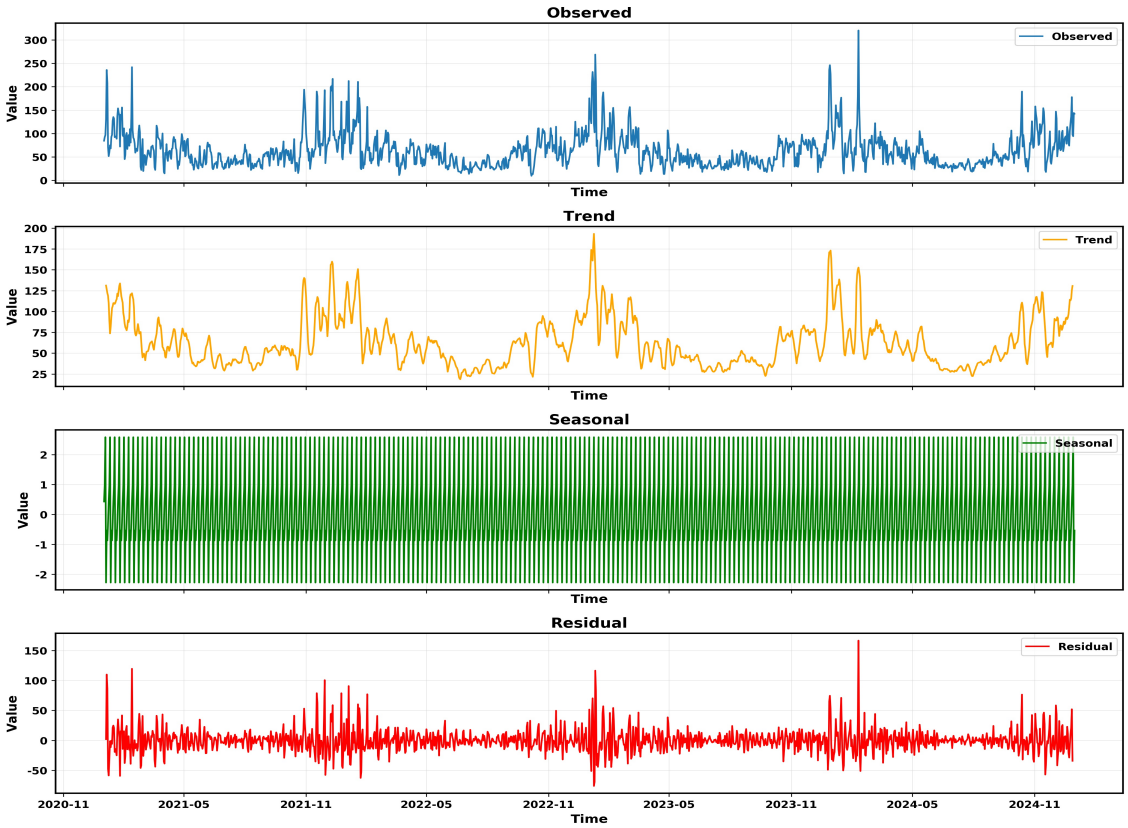



Figure 2: Seasonal survey diagram

Analyzed in conjunction with Figure 2, the observed data show a stable long-term trend, and the seasonal fluctuations are highly stable but lack significant cyclical repetitions, indicating that the seasonal component of the data has been adequately captured in the preliminary decomposition. Meanwhile, the residual components are similar to the white noise distribution, which further verifies that there is no obvious non-stationary component or uncaptured seasonal structure in the data. Based on these characteristics, it can be inferred that the data have satisfied the statistical requirement of seasonal smoothness, and there is no need to eliminate the non-smoothness through additional seasonal difference processing. Therefore, the seasonal differencing order $D$=0 in the SARIMA model can be determined.

**Step 3:** In this study, the auto-ARIMA method was used for the parameter selection of the SARIMA model, and the optimal parameter combinations of the model were determined through the optimization process [27, 28]. Since the data are hourly granular observations, the unit of the seasonal period $s$ corresponds to hours. To systematically assess the impact of seasonal cycles, we select three candidate cycles $s = \{7, 24, 48\}$ for comparison, where $s = 24$ corresponds to a daily cycle, $s = 48$ corresponds to a two-day cycle, and $s = 7$ represents a 7-hour short-term cycle. During the model construction process, both the non-seasonal difference order $d$ and the seasonal difference order $D$ are set to 0, suggesting that the time series is already stationary and does not require any additional differencing.

By auto-ARIMA method, the $p$, $q$, $P$ and $Q$ parameters of the model are searched in the following ranges: $p \in [1, 5]$, $q \in [1, 5]$, $P \in [0, 2]$, $Q \in [0, 2]$. The Akike Information Criterion (AIC) is used as the model evaluation criterion in the parameter selection process. In order to avoid overfitting of the model, we use rolling-origin cross-validation on different candidate periods $s$ to evaluate the prediction performance of out-of-sample. The experimental results are detailed in Figures 3 and 4.

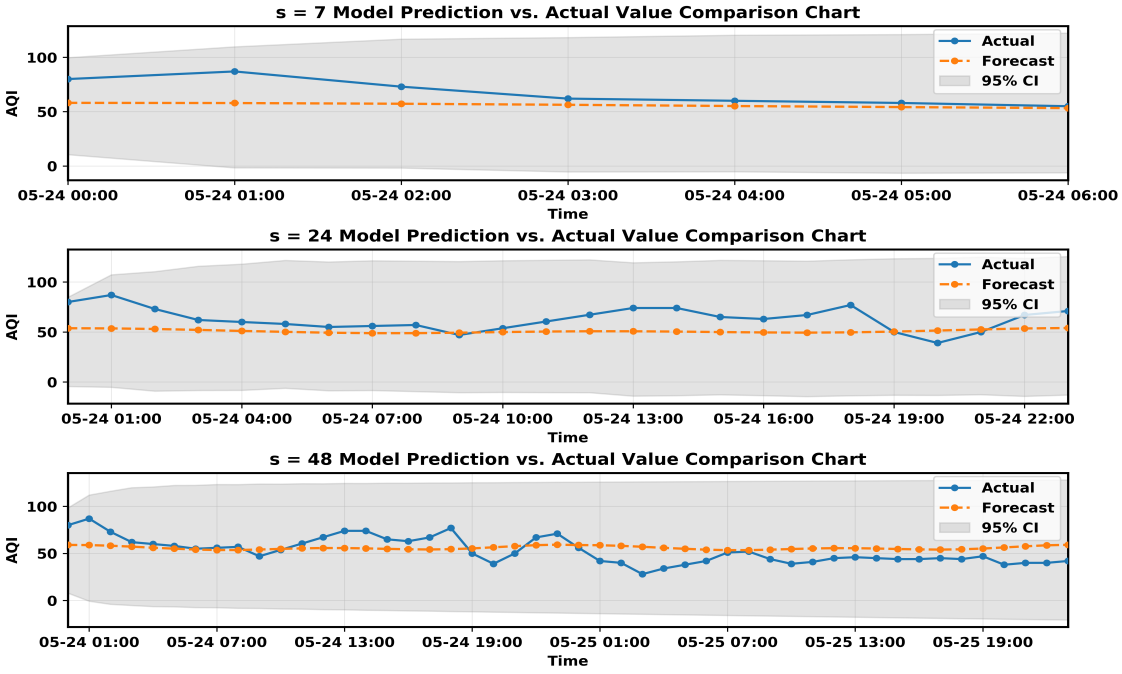



Figure 3: Rolling-origin cross-validation — Comparison of SARIMA Model Forecast with Actual Values under candidate seasonal period

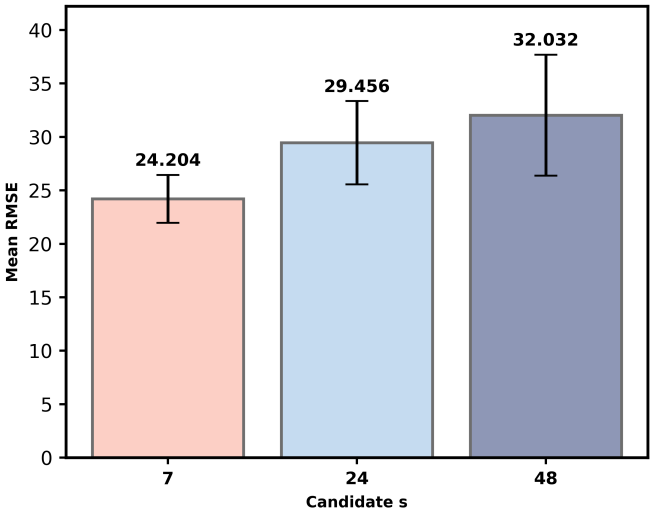



Figure 4: Rolling-origin cross-validation — Boxplot of out-of-sample $RMSE$ by candidate seasonal period

The results show that in the candidate set, based on the comprehensive comparison between the AIC value of the whole series and the $RMSE$ of the outer sample of rolling-origin cross-validation, the model corresponding to $s = 7$ is better than $s = 24$ and $s = 48$ in terms of fitting and prediction, so the seasonal period $s$ is selected as 7. The step-by-step search of auto-ARIMA screened out the optimal parameter combination of the model as SARIMA(3,0,2)$(0,0,1)_7$, with an AIC value of 13283.239. In the case of $s = 7$, all candidate parameter combinations are ranked in ascending order by AIC, and the top 10 are reported in Table 2.

Table 2: auto-ARIMA model selection

| **Model** | **Description** | **AIC** |
|---|---|---|
| SARIMA(3,0,2)$(0,0,1)_7$ | $pdq$(3,0,2), $PDQ$(0,0,1) ($s$): 7 | 13283.239 |
| SARIMA(4,0,2)$(0,0,1)_7$ | $pdq$(4,0,2), $PDQ$(0,0,1) ($s$): 7 | 13283.737 |
| SARIMA(3,0,2)$(0,0,0)_7$ | $pdq$(3,0,2), $PDQ$(0,0,0) ($s$): 7 | 13283.956 |
| SARIMA(3,0,2)$(0,0,2)_7$ | $pdq$(3,0,2), $PDQ$(0,0,2) ($s$): 7 | 13285.067 |
| SARIMA(3,0,2)$(1,0,1)_7$ | $pdq$(3,0,2), $PDQ$(1,0,1) ($s$): 7 | 13285.306 |
| SARIMA(4,0,3)$(0,0,1)_7$ | $pdq$(4,0,3), $PDQ$(0,0,1) ($s$): 7 | 13285.508 |
| SARIMA(2,0,2)$(0,0,0)_7$ | $pdq$(2,0,2), $PDQ$(0,0,0) ($s$): 7 | 13291.475 |
| SARIMA(3,0,2)$(1,0,0)_7$ | $pdq$(3,0,2), $PDQ$(1,0,0) ($s$): 7 | 13291.629 |
| SARIMA(4,0,3)$(0,0,0)_7$ | $pdq$(4,0,3), $PDQ$(0,0,0) ($s$): 7 | 13293.717 |
| SARIMA(4,0,1)$(0,0,1)_7$ | $pdq$(4,0,1), $PDQ$(0,0,1) ($s$): 7 | 13296.251 |

**Step 4:** To comprehensively evaluate the suitability and effectiveness of the selected SARIMA(3,0,2)$(0,0,1)_7$ model, we will perform a white noise test on its residuals and use the Ljung-Box Q statistic to assess the autocorrelation of the residuals [29].

The results presented in Table 3 indicate that the $Q$ statistic is 0.23, with a corresponding $p$-value of 0.63, which is significantly greater than the 0.05 threshold. This suggests that the model residuals exhibit no significant autocorrelation and can therefore be regarded as white noise, confirming the model sufficiently captures the inherent patterns in the data.

Table 3: SARIMA(3,0,2)$(0,0,1)_7$ White noise test

| **Ljung-Box ($Q$)** | **$p$-value** |
|---|---|
| 0.23 | 0.63 |

As shown in Table 4, the $p$-values of the parameters of the SARIMA(3,0,2)$(0,0,1)_7$ model are all less than 0.05, indicating that the estimated model parameters are statistically significant and contribute effectively to capturing the time series structure of the data.

Table 4: SARIMA$(3,0,2)(0,0,1)_7$ Parameter test

| **Parameter** | **coef** | **std err** | $z$ | $p > \|z\|$ | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| ar.L1 | 0.7959 | 0.033 | 23.864 | 0.000 | 0.731 | 0.861 |
| ar.L2 | 0.7837 | 0.057 | 12.252 | 0.000 | 0.591 | 0.816 |
| ar.L3 | -0.5000 | 0.031 | -15.995 | 0.000 | -0.561 | -0.439 |
| ma.L2 | -0.854 | 0.028 | -29.965 | 0.000 | -0.910 | -0.798 |
| ma.S.L7 | -0.0501 | 0.024 | -2.109 | 0.035 | -0.097 | -0.004 |
| sigma2 | 513.7929 | 9.484 | 54.175 | 0.000 | 495.205 | 532.381 |

To verify the effectiveness of the SARIMA$(3,0,2)(0,0,1)_7$ model for air quality forecasting in Changsha, Hunan Province, this study conducted both in-sample and out-of-sample predictive analyses based on the complete observed series from 2021 to 2024. For the in-sample evaluation, the model was fitted using the training set and its predictive performance was assessed. For the out-of-sample evaluation, after model fitting, the model's forecasting capability was validated using an independent test set. The prediction results are shown in Figure 5, demonstrating the model's forecasting performance on the dataset.

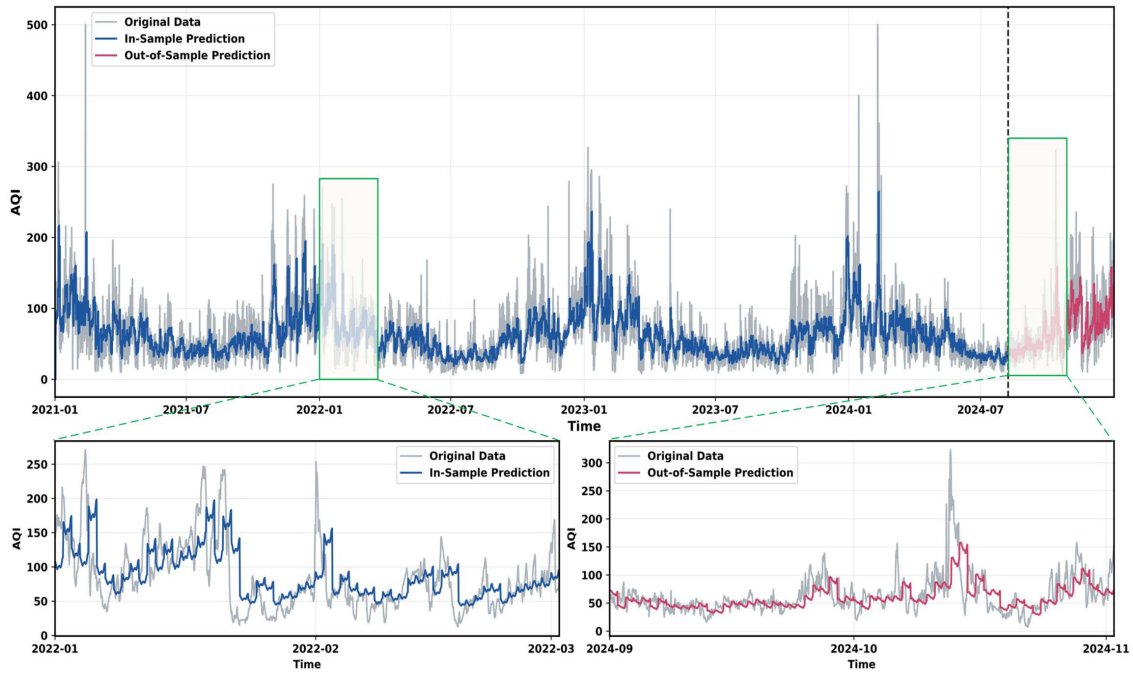



Figure 5: SARIMA$(3,0,2)(0,0,1)_7$ model Prediction Performance Diagram

Combined with the analysis results in Figure 5, it can be seen that the current model can better capture the linear dependence relationship in the data, and shows high accuracy in the overall trend prediction, especially in the interval of steady change,

the fitting effect is superior. It can be seen from the two subplots that the model can effectively fit the data in the in-sample prediction and accurately capture the trend of smooth change. However, in out-of-sample predictions, the model performs poorly in extreme value predictions, especially in highly fluctuating intervals. This exposes the limitations of the model in identifying complex nonlinear patterns and long-term dependencies. The results validate the effectiveness of the traditional SARIMA model in modeling stationary time series trends, while also revealing its inherent bottlenecks in dealing with extreme events driven by complex meteorological factors.

*TCN model*

In order to make up for the shortcomings of SARIMA in capturing nonlinear patterns and long-term dependencies, the prediction residuals of the SARIMA model are inputted into TCN for modeling in this study. In this way, the TCN can effectively learn the complex patterns that SARIMA model fails to capture, and further improve the prediction ability of the model [30]. The TCN efficiently captures the long-term dependencies by stacking multiple residual blocks, and realizes efficient parallel computation by utilizing causal dilated convolutions. The core idea of the model is to ensure that the output at moment t only depends on the data in the past moments, rapidly expanding the receptive field through the dilation convolution, and at the same time combining with the residual connection to efficiently transfer the gradient, so as to enhance model's ability to capture long-term dependence, and to ensure a high training stability [31]. By combining TCN and SARIMA, the respective advantages can be utilized to enhance the prediction accuracy, especially when dealing with extreme events driven by complex weather conditions, showing more flexible and accurate prediction capability.

Specifically, the residual data $X = [x_o, x_1, \cdots, x_{L-1}]$ predicted by the SARIMA model is mapped to a higher-dimensional feature space through the initial layer of the model. To ensure temporal causality, TCN employs causal convolution operations, whose basic formula is:

$$y\left[t\right] = \sum_{i=0}^{k-1} w\left[i\right] \cdot x\left[t - d \cdot i\right] \tag{12}$$

where $k$ is the convolution kernel size, $d$ is the dilation factor, and $w\left[i\right]$ are the weights of the convolution kernel. The causal convolution ensures that each output $y\left[t\right]$ depends only on moments earlier in the input sequence $t$ values, and the dilation convolution significantly expands the receptive field by increasing the dilation rate ($d = 1, 2, 4, \cdots$), so that even the shallower network can capture the dependencies in the long sequence.

Residual Block is the core component of the TCN model. Each residual block contains two layers of causal dilation convolution, and each layer of convolution is equipped with ReLU activation function, Dropout regularization, and Weight Normalization processing to enhance the nonlinear representation of the model and

prevent overfitting [32]. Let the input of the $l$-th residual block be $X^{(l)}$, then the output of the first layer of causal dilation convolution is:

$$H^{(l,1)} = \sigma\left(WeightNorm\left(Conv_d^{(1)}\left(X^{(l)}\right)\right)\right) \tag{13}$$

where $Conv_d^{(1)}(\cdot)$ represents a causal convolution operation with a dilation rate of $d_1$, and $\sigma(\cdot)$ is the ReLU activation function. After Dropout regularization, the output is obtained through the second layer of dilated convolution:

$$H^{(l,2)} = \sigma\left(WeightNorm\left(Conv_d^{(2)}\left(Dropout\left(H^{(l,1)}\right)\right)\right)\right) \tag{14}$$

Then, through a 1×1 convolution (denoted as $W_{(l)}^{1\times1}$), $H^{(l,2)}$ is mapped to the same dimension as $X^{(l)}$ , and a residual connection is performed with $X^{(l)}$ to obtain the output of the $l$-th layer residual block:

$$X^{(l+1)} = X^{(l)} + W_{(l)}^{1\times1}\left(Dropout\left(H^{(l,2)}\right)\right) \tag{15}$$

This residual connection not only effectively mitigates the gradient vanishing problem in the deep network, but also promotes the direct accumulation of features in each layer, thus enhancing the fusion of multi-scale information [33].In Figure 1, the output of each residual block will be used as the input of the next block until, after passing a preset number of stacking (i.e., Stacking times), the features are finally mapped to the output dimension by a 1×1 convolutional layer for performing prediction or classification tasks. The parameter settings of the TCN model are detailed in Table 5.

Table 5: TCN architecture parameters

| **Parameter** | **Description** | **Value** |
|---|---|---|
| $L$ | Number of convolutional layers in the TCN | 4 |
| num_channels | Output channel count for each layer | Search space: each layer $\in [16, 150]$; optimized by SCSO. |
| $k$ | Convolution kernel width | Search space: $\{2, 3, 4\}$; optimized by SCSO. |
| $d$ | Dilation factor for each layer | $[1, 2, 4, 8]$ |
| Padding | Padding used to keep time alignment | Computed per layer as $padding = (k-1) \times d$. |
| Dropout | Dropout rate applied in TCN layers | Search space: 0.10–0.40; optimized by SCSO. |
| Optimizer / LR | Optimizer and learning rate used for TCN training | Adam, lr = 0.001 |

The advantages of the TCN model are mainly reflected in the following aspects: firstly, its parallel computing characteristics make the training efficiency significantly

higher than that of the traditional recurrent neural network (RNN); secondly, the dilation convolution mechanism allows the model to cover the entire time range of the sequence in a small number of layers, so as to effectively capture the long-distance dependencies; finally, the residual structure ensures the stable information transfer and gradient flow of the deep network during the training process, thus further improving the performance and stability of the model.

However, TCN is highly sensitive to hyperparameters (e.g., convolutional kernel size, expansion rate, network depth, etc.), which directly affects its ability to model complex nonlinear dependencies [34]. In order to further improve the prediction performance of TCN, this study introduces the SCSO algorithm to globally optimize and regulate the key hyper-parameters, so as to obtain a better model performance.

*SCSO model*

The design of the SCSO algorithm is inspired by the predatory behavior of the sand cat, an animal that can perceive low-frequency signals below 2 kHz and exhibits excellent prey-digging skills, which enables it to survive in harsh environments. Based on this biological revelation, the SCSO algorithm, achieves a balance between the exploration phase and the exploitation phase by modulating the values of the control vector $\vec{R}$ and at the same time the algorithm has higher flexibility in practical applications through fewer parameter settings and simpler implementation [35].

The SCSO algorithm proposed in this study for optimizing TCN hyperparameters can be divided into four main phases: initialization population, search phase, attack phase and iterative termination phase. The parameter settings of the SCSO model are detailed in Table 6.

Table 6: SCSO architecture parameters

| **Parameter** | **Description** | **Value** |
|---|---|---|
| Population | Number of candidate individuals per generation | 50 |
| Iterations | Number of optimization generations | 100 |
| Variables optimized | Hyperparameters searched by SCSO | num_channels, $k$, dropout |
| Complexity penalty | Penalizes overly large models | $complexity_penalty = sum(channels)/100$ |

**Step 1:** In the optimization process, the first step is to determine the value ranges for the key hyperparameters of the TCN model, such as the number of layers, kernel size, learning rate, etc. Each candidate solution (i.e., 'Sand Cat') is represented by a

$1 \times d$ vector, where $d$ is the number of hyperparameters to be optimized. Each variable $x_i$ must satisfy the constraint $x_i \in [lower, upper]$, thereby forming a candidate solution matrix of size $N_{pop} \times N_d$, where $N_{pop}$ represents the population size and $N_d$ is the number of variables to be optimized. This process not only sets a clear hyperparameter search space for the TCN model, but also lays a solid foundation for the subsequent search, iteration and final determination of the optimal hyperparameter configuration.

**Step 2:** In the search phase, the SCSO algorithm introduces the sensitivity parameter $\vec{r}_G$ to model the sensitivity of sand cats to low-frequency signals, which in turn affects the accuracy of the search [36]. Equation (16) shows the linear decline of $\vec{r}_G$ from 2 to 0.

$$\vec{r}_G = S_M - \left( \frac{2 \times S_M \times iter_c}{iter_{\max} + iter_{\max}} \right) \tag{16}$$

where $S_M$ is the initial sensitivity, $iter_{\max}$ is the maximum number of iterations, and $iter_c$ is the current iteration number. As the number of iterations increases, the sensitivity gradually decreases, indicating that the algorithm transitions from broad exploration to local optimization. Next, by calculating the vector $\vec{R}$ using Equation (17), $\vec{R}$ can control the way the sand cat's position is updated.

$$\vec{R} = 2 \times \vec{r}_G \times rand\,(0,1) - \vec{r}_G \tag{17}$$

When $\left|\vec{R}\right| > 1$, the algorithm is in the search phase. In this phase, the sand cat updates its position by combining its current position with the global best candidate position $\overrightarrow{Pos_{bc}}(t)$ and the agility factor $\vec{r}$. Specifically, the agility factor $\vec{r}$ is calculated using Equation (18).

$$\vec{r} = \vec{r}_G \times rand\,(0,1) \tag{18}$$

Then update the current position according to Equation (19):

$$\overrightarrow{Pos}(t+1) = \vec{r} \cdot \left( \overrightarrow{Pos_{bc}}(t) - rand\,(0,1) \cdot \overrightarrow{Pos_c}(t) \right) \tag{19}$$

This process ensures that the TCN hyperparameter space is fully explored on a global scale, effectively avoiding prematurely falling into local optimal solutions.

**Step 3:** When the algorithm detects $\left|\vec{R}\right| \leq 1$, it enters the attack phase. At this phase, the sand cat adjusts its position based on the distance between its current position in the TCN hyperparameter space and the personal best position $\overrightarrow{Pos_b}(t)$. Specifically, the distance between the current sand cat and the personal best position is first calculated using Equation (20).

$$\overrightarrow{Pos_{rnd}}(t) = |rand(0,1) \cdot (\overrightarrow{Pos_b}(t) - \overrightarrow{Pos_c}(t))| \tag{20}$$

Subsequently, a random angle $\theta$ was selected for each sand cat through a roulette wheel selection [37], and update the position by Equation (21).

$$\overrightarrow{Pos}(t+1) = \overrightarrow{Pos_b}(t) - \overrightarrow{r} \cdot \overrightarrow{Pos_{rnd}}(t) \cdot cos(\theta) \tag{21}$$

This process simulates the flexibility of the sand cat in catching prey, which allows the algorithm to fine-tune the hyperparameters of the TCN model based on local optimization, thus further improving the overall performance of the model.

**Step 4:** Upon reaching the preset maximum iterations, the SCSO algorithm halts and outputs the globally optimal candidate solution. The optimal solution is the best hyperparameter setting of the TCN model, thus realizing the global optimization of TCN hyperparameters.

## Results and discussion

This section focuses on the prediction performance of the SARIMA-TCN-SCSO model on real data. The experiment is divided into two parts: firstly, it demonstrates the prediction effect of the model on the AQI data of Changsha City, Hunan Province for the period of 2021-2024, and secondly, it further evaluates its generalization ability and robustness by validating the model using the data of Beijing, Shanghai, Chongqing and Xi'an. Finally, through the comparison experiment with different models, the significant advantages of the model in key indicators are verified, and its excellent performance is demonstrated.

*Global trends and local flexibility*

As can be seen from Figure 6, the SARIMA-TCN-SCSO model continues the accurate fitting ability of the traditional SARIMA model in the stationary interval in terms of overall trend prediction. At the same time, in some extreme values and complex fluctuation intervals, the model shows higher flexibility and can better capture rapid changes caused by meteorological factors. Especially in the out-of-sample prediction part, the model effectively solves the shortcomings of the traditional SARIMA model in predicting extreme values, and shows the advantages of nonlinear modeling combined with TCN and SCSO models and the global optimization ability of hyperparameters. This result shows that the SARIMA-TCN-SCSO model has strong performance in both global trend prediction and local extreme value prediction, overcoming the limitations of a single SARIMA model in extreme change scenarios.

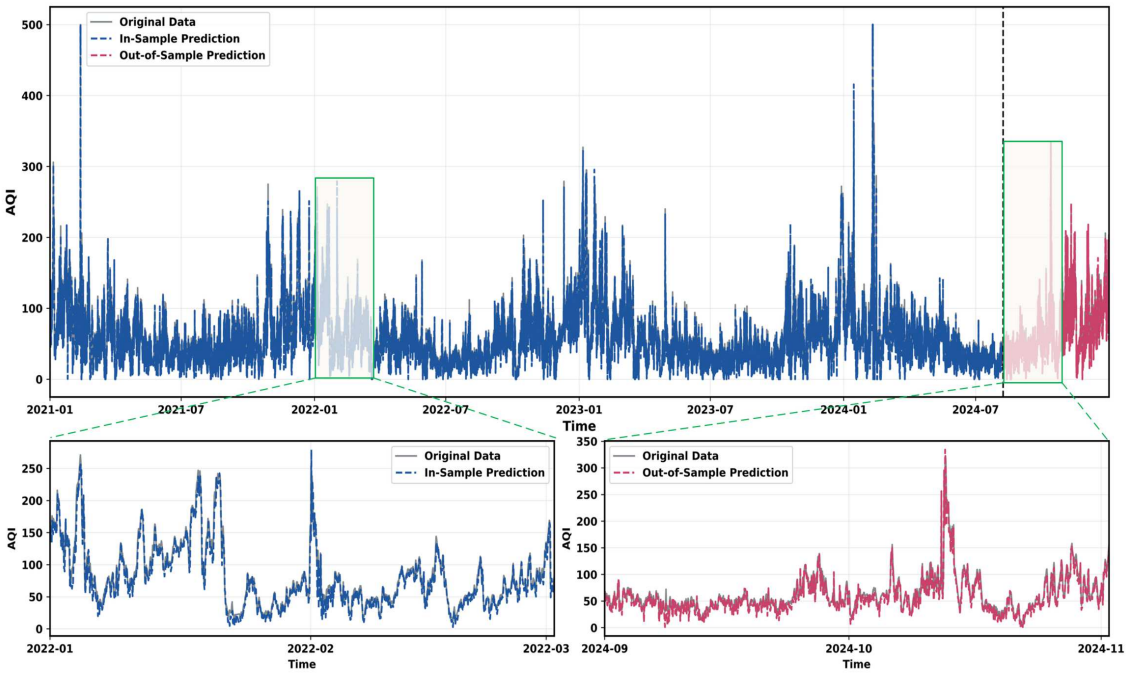



Figure 6: SARIMA-TCN-SCSO model Prediction Performance Diagram

This study employs a range of performance metrics to thoroughly assess the model's predictive accuracy [38]. These metrics include the Mean Squared Error ($MSE$), Root Mean Squared Error ($RMSE$), Mean Absolute Error ($MAE$), Mean Absolute Percentage Error ($MAPE$), and the Coefficient of Determination ($R^2$). The formulas for each of these metrics are provided below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{22}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{23}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{24}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_i)^2} \tag{26}$$

Among them, $y_i$ is the true value of the $i$-th sample, $\hat{y}_i$ is the predicted value of the $i$-th sample, and $n$ is the number of samples.

In terms of performance indicators, the model excels in overall prediction accuracy and stability. Specifically, the $MAE$ is 8.158, the $MSE$ is 115.767, and the $RMSE$ is

10.760, which indicates that the deviation between the predicted values and the actual observed values is very small. Meanwhile, the $R^2$ is as high as 0.931, which shows that the model is able to accurately capture the main trends, and the $MAPE$ is about 17.73%, which fully proves the model's excellent performance in capturing the extreme fluctuations of pollution. Combined with the above indicators, the overall prediction effect of the model is significantly improved compared with the traditional method.

*Regional generalization capability validation*

Based on the good results obtained in the experiment in Changsha, Hunan Province, this study extends the same data preprocessing procedure to the AQI data of Beijing, Shanghai, Chongqing and Xi'an, and utilizes the SARIMA-TCN-SCSO model for prediction. The selected cities represent different climatic and geographic conditions in China, and the adaptability and generalization ability of the model can be comprehensively evaluated. Among them, Beijing, as the capital city of China, is heavily polluted by industry and transportation, and its AQI data can adequately test the model's prediction performance in a highly polluted environment [39]; Shanghai, as a typical coastal city with humid climate and diverse pollution sources, is suitable for investigating the model's adaptability to coastal economic regions [40]; The complex topography and humid climate of Chongqing help to evaluate the model's prediction ability in mountainous areas, while Xi'an, as an inland arid city, provides a positive basis for verifying the model's adaptability in arid regions because of its industrial-based pollution characteristics [41].

After the experimental analysis, the corresponding SARIMA-TCN-SCSO model prediction performance indicators for each region and the best SARIMA model parameter combinations in this model are shown in Table 7.

Table 7: Prediction results of SARIMA-TCN-SCSO model in different regions

| **City** | **SARIMA parameter** | ***MSE*** | ***RMSE*** | ***MAE*** | ***MAPE*** | $\boldsymbol{R^2}$ |
|---|---|---|---|---|---|---|
| Beijing | $(1,0,1)(1,0,0)_7$ | 155.508 | 12.470 | 5.735 | 12.228 | 0.935 |
| Shanghai | $(1,0,4)(0,0,2)_7$ | 110.933 | 10.532 | 7.915 | 20.374 | 0.873 |
| Chongqing | $(2,0,2)(0,0,1)_7$ | 54.659 | 7.393 | 4.933 | 13.901 | 0.893 |
| Xi'an | $(1,0,1)(2,0,1)_7$ | 514.502 | 22.683 | 20.356 | 36.429 | 0.849 |

The experimental results show that the model achieves low error and high goodness-of-fit in AQI prediction in all cities. Among them, the $R^2$ of Beijing reaches 0.935, the $MSE$ of Chongqing is only 54.659, and the prediction error of Xi'an, which is more polluted, is still within the acceptable range. This shows that the model has strong adaptability and generalization ability in different regions and climatic conditions, which

further proves its robustness and superior performance in the task of predicting diverse time series data.

*Ablation and comparative studies*

Based on the AQI forecasting results for Changsha, Hunan Province, to further assess the effectiveness of SCSO for hyperparameter search, we compared several commonly used bio-inspired metaheuristic optimizers in addition to the ablation study. The results are summarized in Table 8. The compared optimizers include Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Cuckoo Search (CS), together with the SCSO algorithm adopted in this study. To ensure a fair comparison, all optimizers were evaluated under the same search space and an equivalent search budget using the same evaluation protocol.

Table 8: Ablation study results

| **Model** | ***MSE*** | ***RMSE*** | ***MAE*** | ***MAPE*** | $\boldsymbol{R^2}$ |
|---|---|---|---|---|---|
| SARIMA | 519.956 | 22.803 | 14.724 | 27.656 | 0.613 |
| SARIMA-TCN | 473.658 | 21.764 | 14.219 | 27.801 | 0.631 |
| SARIMA-TCN-PSO | 197.174 | 14.042 | 9.932 | 19.550 | 0.883 |
| SARIMA-TCN-GWO | 158.055 | 12.572 | 8.429 | 16.600 | 0.906 |
| SARIMA-TCN-CS | 271.896 | 16.489 | 10.628 | 20.920 | 0.839 |
| **SARIMA-TCN-SCSO** | **115.767** | **10.760** | **8.158** | 17.730 | **0.931** |

The experimental results show that although the prediction performance of the traditional SARIMA model can basically meet the requirements, there is still room for improvement in the overall accuracy; After TCN was used to modify the residual error of SARIMA model, the model was significantly enhanced in fitting complex nonlinear models, and all evaluation indicators were improved. Furthermore, after global optimization of TCN hyperparameters with different bio-inspired metaheuristic optimizers, the performance of each hybrid model is significantly better than that of the baseline model without tuning parameters. In particular, the hyperparameter optimization effect based on the SCSO algorithm is the best: compared with the SARIMA model baseline, the resulting hybrid model reduces $MAE$, $MSE$, $RMSE$, and $MAPE$ by approximately 44.6%, 77.7%, 52.8%, and 35.9%, respectively, while increasing $R^2$ by about 51.9%.

To evaluate the relative advantages and robustness of the proposed method across different modeling paradigms, we selected traditional statistical model, typical machine-learning model, and representative deep time-series models as baselines for comparison. All models were trained and tested on the same dataset under identical evaluation

493 settings. Multiple metrics were used to comprehensively assess forecasting performance
494 in terms of both error magnitude and goodness of fit. The comparative results are
495 reported in Table 9.

Table 9: Comparative experimental results

| Model | $MSE$ | $RMSE$ | $MAE$ | $MAPE$ | $R^2$ |
|---|---|---|---|---|---|
| ARIMAX | 264.478 | 16.263 | 11.125 | 15.450 | 0.835 |
| ARFIMA | 518.612 | 22.773 | 14.592 | 27.610 | 0.615 |
| Holt-Winters | 546.004 | 23.367 | 16.126 | 30.930 | 0.595 |
| LightGBM | 680.869 | 26.094 | 18.477 | 42.620 | 0.601 |
| LSTM | 365.544 | 19.119 | 13.175 | 25.420 | 0.716 |
| Informer | 433.010 | 20.809 | 14.463 | 32.120 | 0.737 |
| **SARIMA-TCN-SCSO** | **115.767** | **10.760** | **8.158** | 17.730 | **0.931** |

496 The experimental results indicate that, while a single statistical model can
497 effectively capture linear and seasonal components, its error reduction is limited when
498 confronted with nonlinear disturbances and abrupt fluctuations that are common in AQI
499 series. End-to-end deep time-series models are more sensitive to sample size and training
500 configurations, leading to less consistent performance gains. In contrast, the SARIMA-
501 TCN-SCSO model exhibits more stable overall advantages across multiple error metrics
502 and goodness-of-fit measures, highlighting its stronger capability to capture complex
503 temporal dynamics and its robustness to hyperparameter uncertainty.

## 504 Conclusions

505 In this study, an innovative hybrid model, SARIMA-TCN-SCSO, is proposed, which
506 integrates traditional statistical analysis, in-depth time-series modeling and intelligent
507 optimization strategies, aiming to improve the accuracy and stability of AQI prediction
508 and make up for the inadequacy of a single method in capturing the complex time-series
509 dynamics, thus providing a new solution for environmental time series prediction.

510 The model gives full play to the advantages of SARIMA in portraying linear trends
511 and seasonal fluctuations, combines the ability of TCN in extracting complex nonlinear
512 dependencies, and realizes the global optimization of hyper-parameters using the SCSO
513 algorithm. In this study, the validity of the model is verified by the AQI data of
514 Changsha City, Hunan Province, and the experimental results show that the SARIMA-
515 TCN-SCSO model performs excellently in several assessment metrics, especially in
516 capturing the extreme pollution fluctuations and recognizing the complex time-series
517 patterns. Subsequent tests on the AQI datasets of Beijing, Shanghai, Chongqing

and Xi'an further demonstrate the superb generalization ability and adaptability of the model under different geographic and climatic conditions, showing good regional applicability and wide application potential. Comprehensive ablation study results and cross-model comparison results show that the SARIMA-TCN-SCSO model has significant advantages in various predictive performance indicators, which further verifies the effectiveness and superiority of the proposed method.

In summary, the SARIMA-TCN-SCSO model can not only significantly reduce the prediction error, but also improve the response ability to extreme pollution events, which provides powerful technical support for environmental monitoring and pollution prevention under complex meteorological conditions. Future research will further apply the model to different fields, such as urban traffic flow forecasting and energy demand forecasting, to expand the scope of application of the model and improve its forecasting ability in various scenarios; At the same time, optimize the SCSO algorithm to improve its scalability and computing efficiency to meet the real-time prediction needs of large-scale data. In addition, the application potential of different optimization strategies and model architectures in other environmental and economic fields will be explored to promote the wide application of intelligent prediction technology in environmental management and policy decision-making.

## Statements and Declarations

*Ethics approval and consent to participate*

Not applicable. This study does not involve human participants or animals.

*Consent for publication*

Not applicable. This study does not contain any details, images, or videos relating to individual participants.

*Availability of data and material*

The AQI data used in this study were obtained from the National Urban Air Quality Real-Time Release Platform (https://air.cnemc.cn:18007/). This publicly accessible dataset provides real-time and historical AQI records across multiple urban areas in China, ensuring high accuracy and reliability for research purposes.

*Competing interests*

The authors declare that they have no competing interests.

*Funding*

This research was supported by the Education Department of Hunan Province of China (No.22A0197 & No.24C0131), the Hunan Provincial Natural Science Foundation

(No.2024JJ5235) and the Hunan students innovation and entrepreneurship project (No.S202410538091).

*Author contributions*

Jiahao Guo completed the writing of the thesis and the preparation of the experiments; Yingfang Zhu provides technical guidance and thesis supervision; Huihua Huang and Hao Huang provided the dataset and relevant experimental guidance; Juyang Liao provided guidance on the research paper and financial support.

*ORCID iDs*

Jiahao Guo https://orcid.org/0009-0009-2201-1150
Yingfang Zhu https://orcid.org/0000-0003-1098-5133

## References

1. Muzakki, N. F., Putri, A. Z., Maruli, S. & Kartiasih, F. Forecasting the air quality index by utilizing several meteorological factors using the ARIMAX method (case study: Central Jakarta City). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)* **8,** 569–586 (2024).
2. Du, Y., You, S., Liu, W., Basang, T.-x. & Zhang, M. Spatiotemporal evolution characteristics and prediction analysis of urban air quality in China. *Sci. Rep.* **13,** 8907 (2023).
3. Ao, X., Yuan, H. & Zhang, D. PM2.5 analysis and prediction based on seasonal time series model. *IOP Conf. Ser. Earth Environ. Sci.* **371,** 052006 (2019).
4. Singh, T., Sharma, N., Satakshi & Kumar, M. Analysis and forecasting of air quality index based on satellite data. *Inhal. Toxicol.* **35,** 24–39 (2023).
5. Pradhan, S., Panigrahi, S., Purohit, S. & Dash, J. Study and development of hybrid and ensemble forecasting models for air quality index forecasting. *Expert Syst.* **40,** e13449 (2023).
6. Liu, C., Pan, G., Song, D. & Wei, H. Air quality index forecasting via genetic algorithm-based improved extreme learning machine. *IEEE Access* **11,** 67086–67097 (2023).
7. Umoh, M. D., Evans, U. F. & Utting, C. Air quality index prediction using machine learning algorithms for certain locations in Nigeria. *J. Pollut. Eff. Community Health* **3** (2024).
8. Liu, Q., Cui, B. & Liu, Z. Air quality class prediction using machine learning methods based on monitoring data and secondary modeling. *Atmosphere* **15,** 553 (2024).

9. Unnikrishnan, A. & Rajeswari, S. Optimal prediction of air quality index in metropolitan cities using fuzzy time series with deep learning approach. *Ecol. Eng. Environ. Technol.* **25,** 341–359 (2024).
10. Wu, J., Wu, Y. & Chen, C. Periodic attention networks for air quality index prediction. In *Proceedings of the 2023 4th International Conference on Machine Learning and Computer Application* 316–320 (Association for Computing Machinery, Hangzhou, China, 2024).
11. Sunku, V. S. R. P., Mukkamala, R. & Namboodiri, V. Air quality index prediction using multivariate deep neural networks: A case study of a proposed state capital in India. *J. Air Pollut. Health* **8,** 257–268 (2023).
12. Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems* **30,** 5998–6008 (Curran Associates, Inc., 2017).
13. Zhou, H. *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **35,** 11106–11115 (2021).
14. Feng, Y. *et al.* Spatiotemporal informer: A new approach based on spatiotemporal embedding and attention for air quality forecasting. *Environ. Pollut.* **336,** 122402 (2023).
15. Zhang, Z., Chen, H. & Huang, X. Prediction of air quality combining wavelet transform, DCCA correlation analysis and LSTM model. *Appl. Sci.* **13,** 2796 (2023).
16. Xia, H., Chen, X., Wang, Z., Chen, X. & Dong, F. A multi-modal deep-learning air quality prediction method based on multi-station time-series data and remote-sensing images: Case study of Beijing and Tianjin. *Entropy* **26,** 91 (2024).
17. Deepan, S. & Saravanan, M. Air quality index prediction using seasonal autoregressive integrated moving average transductive long short-term memory. *ETRI J.* **46,** 915–927 (2024).
18. Ji, Q. *et al.* Short-term prediction of the significant wave height and average wave period based on the variational mode decomposition–temporal convolutional network–long short-term memory (VMD–TCN–LSTM) algorithm. *Ocean Sci.* **19,** 1561–1578 (2023).
19. Wang, M. & Qin, F. A TCN-linear hybrid model for chaotic time series forecasting. *Entropy* **26,** 467 (2024).
20. Li, X., Qi, Y., Xing, Q. & Hu, Y. IMSCSO: An intensified sand cat swarm optimization with multi-strategy for solving global and engineering optimization problems. *IEEE Access* **11,** 122315–122344 (2023).
21. Box, G. E. P. & Jenkins, G. M. *Time Series Analysis: Forecasting and Control* (Holden-Day, San Francisco, CA, 1970).
22. Zhang, W. & Dong, W. *SPSS Statistical Analysis Advanced Course* 3rd ed. (Higher Education Press, Beijing, 2018).

23. Kenyi, M. & Yamamoto, K. A hybrid SARIMA-Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan. *Discov. Appl. Sci.* **6,** 457 (2024).
24. Wu, W., An, S., Guan, P., Huang, D. & Zhou, B. Time series analysis of human brucellosis in mainland China by using Elman and Jordan recurrent neural networks. *BMC Infect. Dis.* **19,** 414 (2019).
25. Zhang, J. *Based on DTW-SARIMA-CNN-LSTM combination model air quality prediction research* Master's thesis (Shanxi University, 2023).
26. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* 3rd ed. (OTexts, Melbourne, Australia, 2021).
27. Szostek, K., Mazur, D., Drałus, G. & Kusznier, J. Analysis of the effectiveness of ARIMA, SARIMA, and SVR models in time series forecasting: A case study of wind farm energy production. *Energies* **17,** 4803 (2024).
28. Hyndman, R. J. & Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **27,** 1–22 (2008).
29. Tadesse, K. B. & Dinka, M. O. The SARIMA model-based monthly rainfall forecasting for the Turksvygbult Station at the Magoebaskloof Dam in South Africa. *J. Water Land Dev.* **53,** 100–107 (2022).
30. Chen, J., Cui, Y., Zhang, X., Yang, J. & Zhou, M. Temporal convolutional network for carbon tax projection: A data-driven approach. *Appl. Sci.* **14,** 9213 (2024).
31. Shi, T., Yang, W., Qi, A., Li, P. & Qiao, J. LASSO and attention-TCN: a concurrent method for indoor particulate matter prediction. *Appl. Intell.* **53,** 20076–20090 (2023).
32. Xu, Z. *et al.* PhaCIA-TCNs: Short-term load forecasting using temporal convolutional networks with parallel hybrid activated convolution and input attention. *IEEE Trans. Netw. Sci. Eng.* **11,** 427–438 (2024).
33. Wang, Y. *et al.* Short-term load forecasting for industrial customers based on TCN-LightGBM. *IEEE Trans. Power Syst.* **36,** 1984–1997 (2021).
34. Geng, Y., Zhang, F. & Liu, H. Multi-scale temporal convolutional networks for effluent COD prediction in industrial wastewater. *Appl. Sci.* **14,** 5824 (2024).
35. Seyyedabbasi, A. & Kiani, F. Sand cat swarm optimization: A nature-inspired algorithm to solve global optimization problems. *Eng. Comput.* **39,** 2627–2651 (2023).
36. Yao, L. *et al.* Multi-strategy improved sand cat swarm optimization: Global optimization and feature selection. *Biomimetics* **8,** 492 (2023).
37. Zhang, K., He, Y., Wang, Y. & Sun, C. Improved multi-strategy sand cat swarm optimization for solving global optimization. *Biomimetics* **9,** 280 (2024).
38. Zhao, Z. *et al.* Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China. *BMC Infect. Dis.* **23,** 71 (2023).

665 39. Zhou, Y. *et al.* Pollution characteristics and sources of ambient air dustfall in urban
666 area of Beijing. *Atmosphere* **15,** 544 (2024).
667 40. Yang, J. *et al.* Characteristics of atmospheric pollution in a Chinese megacity:
668 Insights from three different functional areas. *Sustainability* **15,** 2429 (2023).
669 41. Zhao, B. *et al.* Potential source area and transport route of atmospheric particulates
670 in Xi'an, China. *Atmosphere* **14,** 811 (2023).