



## Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models

Yong Cheng<sup>a</sup>, Qiao Zhu<sup>a</sup>, Yan Peng<sup>a</sup>, Xiao-Feng Huang<sup>a,b</sup>, Ling-Yan He<sup>a,\*</sup>

<sup>a</sup> Key Laboratory for Urban Habitat Environmental Science and Technology, School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, 518055, China

<sup>b</sup> Environmental Laboratory, PKU-HKUST Shenzhen-Hong Kong Institution, Shenzhen, 518057, China

### ARTICLE INFO

Handling Editor: Bin Chen

**Keywords:**

Multiple strategies  
Ozone  
Hybrid forecasting algorithm  
Data decomposition

### ABSTRACT

Ground-level ozone is an air pollutant that has adverse impacts on human health and vegetation growth. The accurate prediction of ozone concentrations is essential for developing strategies for ozone mitigation. To obtain a better forecasting model to predict ozone, this study provides a detailed discussion of the application of three model optimization strategies (i.e., adding decomposition algorithms, adding data and adding factors) to benchmark models, including long short-term memory (LSTM) and support vector regression (SVR), to predict ozone concentrations in Shenzhen. The results showed that adding a decomposition strategy, particularly the wavelet decomposition (WD) algorithm, provided the greatest improvement to the prediction accuracy. Based on this, a novel hybrid forecasting model (WD-LSTMSVR) was further developed that first used the WD algorithm to convert the original data from one dimension to multiple dimensions. Subsequently, each layer of the data set was trained and forecast by the LSTM and SVR models, which involved parameters that were optimized by the autoregressive integrated moving average (ARIMA) partial algorithm and particle swarm optimization (PSO) algorithm. The hybrid forecasting model had the best prediction accuracy performance compared with the benchmark models and optimization models in this study. Our results indicate that the developed hybrid forecasting model is a good technique to provide accurate ozone concentration prediction results.

### 1. Introduction

The near-surface formation process of ozone, a secondary pollutant, is highly complex, as it is produced by a series of intricate photochemical reactions of NOx and VOCs (Haagen-Smit, 1952). High concentrations of ozone not only damage the ecological environment but also seriously affect human health (U.S.EPA, 2013). In recent years, the ambient ozone concentrations have continued to increase in China. From 2013 to 2017, the 8-h ambient ozone concentrations (8-h average) increased from 139.2 µg/m<sup>3</sup> (69.6 ppb) to 163.0 µg/m<sup>3</sup> (81.5 ppb) in 74 key cities in China (Huang et al., 2018). Therefore, the accurate and reliable prediction of ozone is very important for preventing air pollution and protecting citizens' health.

In general, ozone forecasting models can be divided into two categories: numerical models and data-driven models. Numerical models (Afzali et al., 2017; Konovalov et al., 2009) are based on classic physical and chemical theories. Due to the difficulty of obtaining source

inventory data and some high-resolution data and the complexity of atmospheric conditions, numerical models are complex and difficult to establish. In our study, we mainly consider data-driven models due to their advantages in terms of software and data accessibility; these types of models are based on historical data characteristics (Knüsel et al., 2020). Data-driven models also achieve a higher prediction accuracy in many cases and can be used as supplements and extensions to numerical models.

Data-driven models include statistical models and hybrid models. In recent years, statistical models have developed rapidly in the field of air pollution prediction. Lu and Wang (2014) used two machine learning models, the multilayer perceptron (MLP) and support vector machine (SVM), to predict ozone and compared the risk applicability of the two models. It was found that the SVM model has good predictive performance, but its generalization ability still needs to be improved. Other studies have used ANNs (AlOmar et al., 2020; Prasad et al., 2016), convolutional neural networks (CNNs) (Sayeed et al., 2020) and other predictive models to forecast ozone. These models improve the capture

\* Corresponding author.

E-mail addresses: [yongcheng@pku.edu.cn](mailto:yongcheng@pku.edu.cn) (Y. Cheng), [zhuqiao2013@pku.edu.cn](mailto:zhuqiao2013@pku.edu.cn) (Q. Zhu), [pengyan@pku.edu.cn](mailto:pengyan@pku.edu.cn) (Y. Peng), [huangxf@pku.edu.cn](mailto:huangxf@pku.edu.cn) (X.-F. Huang), [hely@pku.edu.cn](mailto:hely@pku.edu.cn) (L.-Y. He).

Abbreviations	
NOx	Nitrogen oxide
VOCs	Volatile organic compounds
ANN	Artificial neural network
RNN	Recurrent neural network
LSTM	Long short-term memory
SVR	Support vector regression
RBF	Radial basis function
ARIMA	Autoregressive integrated moving average
PSO	Particle swarm optimization
WD	Wavelet decomposition
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise
IMFs	Intrinsic mode functions
ReLU	Rectified linear unit
MSE	Mean square error
MAE	Mean absolute error
RMSE	Root mean squared error
MAPE	Mean absolute percentage error
ACF	Autocorrelation function
PACF	Partial autocorrelation function
AIC	Akaike information criterion
MLR	Multiple linear regression
RF	Random forest
SC	Schwarz criterion
AQI	Air quality index
MF	Meteorological factor
MS	Monitoring station
AP	Atmospheric pollutant
AvT	Average temperature
MaxT	Maximum temperature
MinT	Minimum temperature
AvRH	Average relative humidity
MinRH	Minimum relative humidity
AvWS	Average wind speed
MaxWS	Maximum wind speed
MaxWD	Maximum wind direction
EWS	Extreme wind speed
EWD	Maximum wind direction
SH	Sunshine hours
P	Precipitation
LCE	Large-scale evaporation
AvP	Average pressure
MaxP	Maximum pressure
MinP	Minimum pressure
AvST	Average surface temperature
MaxST	Maximum surface temperature
MinST	Minimum surface temperature
XGBoost	Extreme gradient boosting

of daily ozone changes, but most of these studies used relatively few models for comparison, and most of the compared models were of the same type. In addition, Wu and Lin (2019) predicted the air quality between two cities in China and found that compared with other algorithms, the LSTM deep learning model achieves good performance and has great potential for time series prediction. Due to its excellent performance, the LSTM model has recently become a highly popular research model in deep learning and has been the subject of extensive research in other areas (Chen et al., 2019; Niu et al., 2020). However, there have been few studies related to the use of LSTM models for ozone prediction.

Despite progress in statistical models, better accuracy and applicability are still needed for pollutant data with different time series characteristics. Because the meteorological parameters and photochemical reactions involved in the formation of ozone are very complicated, it is difficult to develop an accurate statistical model. A hybrid model combines the advantages of multiple algorithms to attain better prediction results (Liu et al., 2020). The main principle of a hybrid model is to apply a single benchmark model to the hierarchical sequence generated by the decomposition algorithm. This produces more reasonable results, thereby improving the feature representation and information utilization capabilities. The nonlinear original time series is further decomposed into more stable and regular subsequences, and the final prediction result is obtained by reconstructing the predicted values of each subsequence. Decomposition algorithms commonly used in recent research include WD (AlOmar et al., 2020; Cheng et al., 2019), EMD (Zhu et al., 2017), EEMD (Song and Fu, 2020; Wang et al., 2019) and CEEMDAN (Mo et al., 2020), which have been confirmed to effectively improve the prediction accuracy of the model (Liu et al., 2020). However, few studies have considered the influences of relevant factors that affect the target predictor on the forecasting result, weakening the theoretical and practical support for the forecasting model. In addition, due to the complexity of using decomposition algorithms, most studies usually use a single decomposition algorithm, and there have been few systematic experiments on the differences in performance and applicability of classic decomposition algorithms.

Thus, in this study, we proposed three model optimization strategies based on the results of previous studies: adding decomposition algorithms, adding data, and adding factors. The main contributions and innovations of this study are as follows:

- (a) Three model optimization strategies were proposed for ozone concentration prediction, analyzed experimentally, and discussed in detail.
- (b) Four classic decomposition algorithms (i.e., WD, EMD, EEMD and CEEMDAN) were applied to assess the strategy of adding decomposition algorithms to predict ozone concentrations. The optimal algorithm was determined by comparing the results, which can also provide valuable experimental support for other studies.
- (c) Models for the strategy of adding factors were constructed by adding multiple factors and a single factor. The factors included meteorological variables, other pollutants, and ozone near stations. The best prediction results were determined based on a comparative analysis.
- (d) In the model experiments, the forward step size judgment method in the ARIMA algorithm was used to determine the step size of the prediction model, and the PSO algorithm was used to optimize the parameter selection of the benchmark model.
- (e) By combining WD, LSTM and SVR, a novel ozone concentration prediction model was developed. This novel model can better predict ozone concentrations by efficiently mining high-level nonlinear features from ozone sequences.

The remainder of this paper is organized as follows. Section 2 briefly introduces the data sources and study methods and describes the three model optimization strategies. In Section 3, a multistrategy case study is carried out, and on this basis, a novel hybrid ozone concentration forecasting model is developed, introduced and compared in detail. Finally, the conclusion is provided in Section 4.

## 2. Data and methodology

In this section, the details of the data are first described. Then, the parameter optimization algorithms, the methodology for the benchmark models and the three model optimization strategies are introduced; finally, the model evaluation method is presented.

The overall introduction process follows the order of the flowchart of this research, as shown in Fig. 1. The whole flowchart consists of three parts. The first part introduces the model parameter optimization algorithms and benchmark models. The second part describes the three model optimization strategies and methods to improve the model prediction accuracy. The third part presents a novel hybrid forecasting algorithm developed based on the first two research steps. More details are described in the following sections.

### 2.1. Study area and data description

Shenzhen is located in Southeast China and links Hong Kong to the Chinese mainland. As of the end of 2018, the resident population of Shenzhen was 13,026,600, of which the registered population was 4,547,000, but the local police and authorities estimated that the real population was over 20 million (<http://www.sz.gov.cn>). Although Shenzhen is situated approximately one degree south of the Tropic of Cancer, due to the Siberian anticyclone, it has a warm, monsoon-

influenced, humid subtropical climate; as a result, this area experiences long sunshine hours throughout the year, with an average annual sunshine duration of approximately 2120.5 h. Shenzhen's unique geographical and climatic conditions also pose certain challenges to air pollution control in this region. Based on the 2019 report on the state of the environment in Guangdong Province, the pollution problem of PM<sub>2.5</sub> in Shenzhen has improved significantly. However, ozone has become the primary pollutant affecting the ambient air quality in Shenzhen and is the key factor restricting the improvement of air quality (<http://gdei.gd.gov.cn>).

In this study, the main data used are air pollutant data and meteorological data. Air pollution data were collected from the China Air Quality Online Monitoring and Analysis Platform (<https://www.aqistudy.cn>) for 11 automatic air monitoring stations in Shenzhen, namely, the LY, HH, GL, NY, XX, YT, MS, LS, KC, NA and HQC stations. The geographical location of each station is shown in Fig. 2. The automatic monitoring indicators at each station included SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, CO and O<sub>3</sub>. The meteorological data were collected from the China Meteorological Data Network (<http://data.cma.cn>). Nineteen indicators were collected: AvT, MaxT, MinT, AvRH, MinRH, AvWS, MaxWS, MaxWD, EWS, EWD, SH, P, LCE, AvP, MaxP, MinP, AvST, MaxST, and MinST. The specific indicators and statistics are shown in Table 1.

Data sets were collected over a five-year daily scale from August 31,

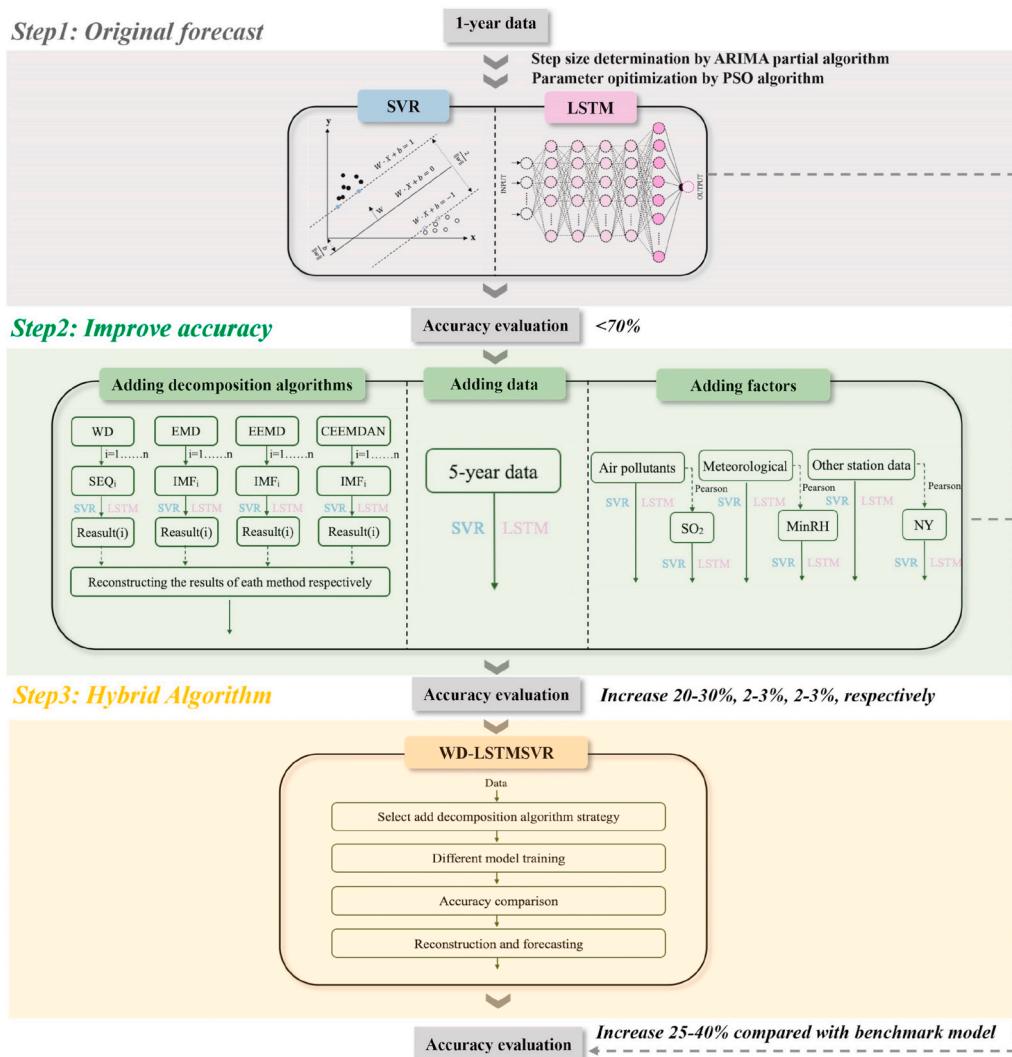
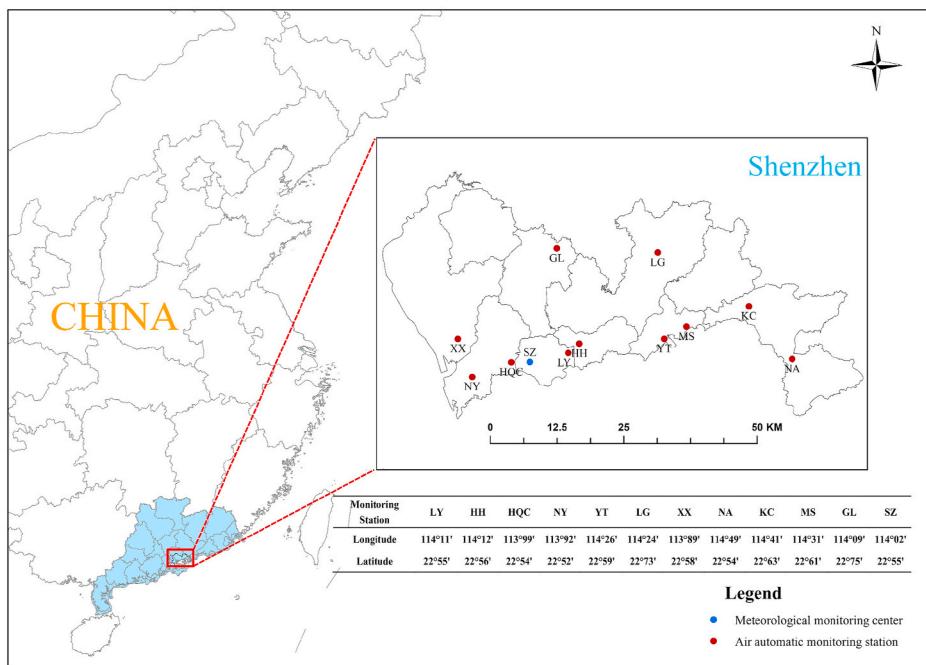


Fig. 1. Flowchart of the research.



**Fig. 2.** Study area of Shenzhen, China.

**Table 1**

The basic statistics of the indicators (HQC station).

Indicator	Mean	Minimum	Maximum	Standard Deviation
O <sub>3</sub> ( $\mu\text{g}/\text{m}^3$ )	56.77	2.50	166.91	30.65
CO ( $\text{mg}/\text{m}^3$ )	0.57	0.22	1.17	0.14
NO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	30.69	6.14	124.79	16.78
PM <sub>10</sub> ( $\mu\text{g}/\text{m}^3$ )	37.70	3.83	159.67	22.70
PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )	23.25	2.33	109.92	15.09
SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	6.07	1.83	14.83	2.50
AvT (°C)	24.06	11.40	30.70	4.80
MaxT (°C)	27.79	13.70	35.10	4.73
MinT (°C)	21.52	8.00	28.70	4.95
AvRH (%)	75.07	29.00	98.00	12.62
MinRH (%)	57.58	18.00	93.00	13.95
AvWS (m/s)	1.85	0.30	5.50	0.76
MaxWS (m/s)	4.39	2.00	8.50	1.36
MaxWD	6.82	1.00	16.00	4.10
EWS (m/s)	8.25	4.10	17.90	2.34
EWD	6.92	1.00	16.00	3.87
SH (h)	5.84	0.00	11.70	3.67
P (mm)	4.04	0.00	86.10	12.38
LCE (mm)	3.73	0.70	8.30	1.56
AvP (kPa)	100.62	99.31	101.91	0.58
MaxP (kPa)	100.83	99.51	102.17	0.61
MinP (kPa)	100.40	98.51	101.75	0.59
AvST (°C)	27.47	12.60	39.60	5.63
MaxST (°C)	43.36	18.30	62.90	8.76
MinST (°C)	20.59	5.30	28.70	5.37

2016, to August 30, 2020. Based on the location of the station and the needs of follow-up research, we choose HQC as the main research station because the meteorological center station is closest to the HQC station. In addition, the NA station, which is farthest from the meteorological center station, was selected as the model verification station. The NA station is located in an area with a high incidence of atmospheric ozone in Shenzhen and thus has special research significance.

The models used in this paper were completed in MATLAB R2019a and Python 3.6. The LSTM model was based on the Keras deep learning framework.

## 2.2. Parameter optimization algorithms

When building a forecasting model, the model parameters need to be configured and optimized. In this study, two optimization algorithms were used. The ARIMA partial algorithm was used to determine the data forward step size, and the PSO algorithm was used to optimize the hyperparameters of the two benchmark models (LSTM and SVR).

### 2.2.1. Autoregressive integrated moving average (ARIMA) partial algorithm

The ARIMA model treats the time series as a random process and considers the statistical characteristics of the time series. Although the sequence value of the time series is random, the seemingly random sequence changes exhibit periodicity. In other words, it is a specific description of time series autocorrelation with its own dynamic memory, and modeling is the process of quantifying dynamic memory (Cheng et al., 2019).

After preprocessing, the data exhibited smooth changes, and the next step was to identify and determine the forward step size of the model. The main method used was to draw the ACF diagram and the PACF diagram and initially determine the autocorrelation coefficient and partial autocorrelation coefficient. Then, the AIC and SC values were combined to determine the specific values of the parameters  $p$  and  $q$ , respectively. The forward step size of all models established in this paper was determined by this method.

### 2.2.2. Particle swarm optimization (PSO)

PSO was used to optimize the parameters in LSTM and SVR. PSO is a nonlinear algorithm derived from simulating the foraging behavior of a group of birds (Kennedy and Eberhart, 2002). It has a simple design, strong global search capability and fast convergence speed. In addition, PSO can solve various nonlinear compound optimization problems and is very suitable for rapid optimization under dynamic and multiobjective optimization conditions. This algorithm is not only suitable for scientific research but also suitable for engineering applications. It is a fast and efficient parallel search optimization algorithm (Ghimire et al., 2018).

The SVR algorithm involves many adjustable parameters. In our experiment, the penalty factor  $C$  and parameter  $g$  in the RBF kernel function were optimized. In the LSTM algorithm, PSO was used to determine the batch number and the number of neurons in the hidden

layer. The setting of hyperparameters in the model can directly affect the training and prediction accuracy of the network.

### 2.3. Benchmark model forecasting

In this study, two benchmark models were used: LSTM and SVR. These two benchmark models have a wide range of applications in the field of machine learning, and both have good prediction performance.

#### 2.3.1. Long short-term memory (LSTM)

LSTM is a special RNN (Hochreiter and Schmidhuber, 1997) that has a chain repeating network structure similar to a standard RNN, but its network structure is more complex than an RNN. The LSTM model is briefly introduced below; more details can be found in Liu and Long (2020).

The gating mechanism of LSTM has three gating units (Ma et al., 2020): the input gate, forget gate and output gate. The gating mechanism mutually determines what information needs to be kept in the network. As a special RNN, the LSTM model has a judgment logic unit that can judge whether the input information is useful, and it can also avoid the problems of gradient explosion and gradient disappearance during data training. The structure of the LSTM model is shown in Fig. 3, which shows the memory block of one LSTM unit. In this figure, the blue solid circle represents multiplication, and the red circle represents the memory unit. Unlike a standard RNN, LSTM uses memory blocks to store previous state information.

#### 2.3.2. Support vector regression (SVR)

The SVM model was proposed by Vapnik (1999). It is widely used in various fields, such as pattern recognition, regression estimation and probability density function estimation. The SVR algorithm is an extension of the SVM model and has applications in the field of fitting regression functions. In this section, only a brief introduction is given; specific details can be found in Ko and Lee (2013).

The SVR machine learning method was developed based on statistical learning theory. Compared with traditional machine learning methods, SVR solves problems characterized by small sample sizes, high dimensionality, and local minima and has a strong generalization ability. In particular, this model is suitable for classification and time series forecasting (Ortiz-García et al., 2010). In general, the key aspect of SVR is the kernel function, of which there are many types. Therefore, choosing the proper kernel function is particularly important. Commonly used kernels include linear kernels, polynomial kernels, and RBF kernels. This study used the RBF kernel function, which has high computational efficiency and maps each sample point to an infinite dimension feature space that linearly inseparable data linearly separable (Abo-Khalil and Dong-Choon, 2008; Nie et al., 2020).

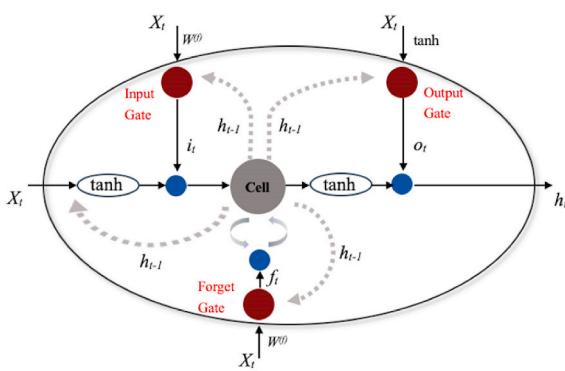


Fig. 3. LSTM memory unit.

### 2.4. Three model optimization strategies

To improve the prediction accuracy and generalization ability of the benchmark models (LSTM and SVR), three model optimization strategies (adding decomposition algorithms, adding data, and adding factors) were proposed based on in-depth research and discussion. The specific indicators are shown in Fig. 1.

#### 2.4.1. Adding decomposition algorithms strategy

Four decomposition algorithms (WD, EMD, EEMD and CEEMDAN) were used in combination with the two benchmark models, which have good performance when used with time series data for prediction. The decomposition of different algorithms is shown in Fig. 4. In the following sections, each algorithm is briefly introduced, and more details can be found in AlOmar et al. (2020), Zhu et al. (2017), Wang et al. (2019) and Mo et al. (2020).

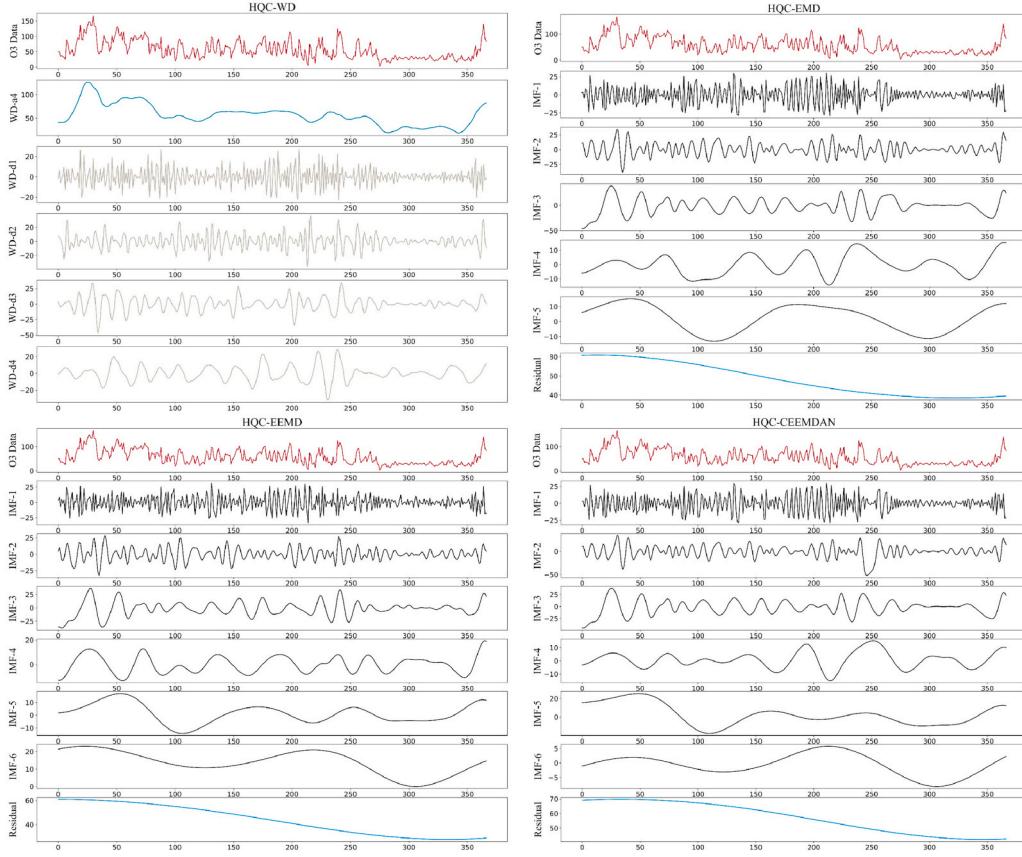
**2.4.1.1. Wavelet decomposition (WD).** WD is based on the multi-resolution principle proposed by Mallat (1989) and has been widely used in signal and image processing and other fields. The wavelet basis function is a method of applying signal analysis in the time and frequency domains. Compared with the Fourier transform, WD can more accurately decompose the required information from the original signal. Due to the multiscale characteristics of WD, the local characteristics of the signal can be better represented in the time-frequency domain, and the window function can be self-adjusted according to the frequency. The use of a narrower window function at high frequencies and a wider window function at low frequencies also solves the problem of the inability to conduct detailed signal analysis in the Fourier transform.

**2.4.1.2. Empirical mode decomposition (EMD).** Multiscale analysis is becoming increasingly popular in the field of complex systems analysis. The basic idea is to decompose the original system into different internal factors of different time scales.

The EMD algorithm proposed by Huang et al. (1998) uses experience and an adaptive basis to extract internal factors. The assumption is that any signal is composed of different simple internal oscillation modes, and the most prominent feature is that it works adaptively and thus can be used for the analysis of nonlinear and dynamic signals. EMD can decompose a signal into a set of finite oscillation components called IMFs. Each IMF must meet the following two conditions: (1) For the entire signal, the numbers of extreme points and zero crossing points of the IMF must be equal or have at most one difference. (2) At any moment, the average value of the upper envelope defined by the maximum point and the next envelope defined by the minimum point is zero.

**2.4.1.3. Ensemble empirical mode decomposition (EEMD).** The essence of EEMD is to repeatedly apply EMD to the signal using a given number of trials and then average all the decomposition results. In each EEMD test, white noise uniformly distributed in the frequency range is added. The noise auxiliary signal not only has a uniform decomposition scale but also smooths abnormal conditions such as impulse interference. As a result, the problem of EMD mode mixing can be effectively alleviated.

**2.4.1.4. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN).** EMD has great advantages in processing nonstationary and nonlinear signals but is still subject to the problem of mode mixing, which means that there are very similar oscillations in different modes or very different amplitudes in the modes. The EEMD algorithm greatly reduces the mode mixing problem in the EMD algorithm by adding white noise to the signal. However, the EEMD algorithm is prone to causing reconstruction errors because white noise cannot be completely eliminated after the signal is reconstructed. To solve these problems, some scholars have proposed using an improved version of



**Fig. 4.** Results for the four decomposition algorithms (WD, EMD, EEMD and CEEMDAN) (HQC station).

CEEMDAN for decomposition, which can more effectively solve the problem of mode mixing and reduce the reconstruction error.

#### 2.4.2. Adding data strategy

In other studies, the impact of the amount of data on the prediction accuracy has rarely been studied and discussed. It is true that the more data there are, the higher the accuracy of the model. However, for models with different predictive performances, the amount of data used is different, and the results and applicability are also different. Therefore, relevant experiments were conducted in our study.

To evaluate the adding data strategy, a 1-year data set was mainly used to build the model, and then a 5-year data set was added to build the corresponding model. The aim was essentially to increase the amount of data in the training data set while leaving the data size of the test set unchanged.

#### 2.4.3. Adding factors strategy

The formation process of ozone is very complicated, as various factors in the environment can affect its formation and distribution. In general, ozone is generated by local sources or is the result of regional transmission. From this perspective, three different factors that may affect ozone were separately studied to assess the strategy of adding factors. These three factors included all 6 air pollutant indicators, 19 meteorological indicators and ozone data for all stations in Shenzhen from the previous day. Thus, six models were established: LSTM-AP-All, SVR-AP-All, LSTM-MF-All, SVR-MF-All, LSTM-MS-All and SVR-MS-All. The specific indicators are described in Section 2.1.

Not all of the collected factors are helpful to the prediction result, and some factors may even become noise in the model, which increases the difficulty of training and reduces the prediction accuracy. To verify that the factors with high correlation can actually improve the accuracy of the model, Pearson's correlation coefficient was applied to analyze

the correlations in ozone levels predicted by the three factors, as shown in Fig. 5. Then, the three most correlated indicators were extracted: SO<sub>2</sub>, MinRH, and NY station ozone data. Next, six models were established: LSTM-AP-SO<sub>2</sub>, SVR-AP-SO<sub>2</sub>, LSTM-MF-MinRH, SVR-MF-MinRH, LSTM-MS-NY and SVR-MS-NY.

#### 2.5. Model evaluation method

In this study, the MAE, RMSE, MAPE and R<sup>2</sup> were used to evaluate the prediction accuracy of each model. Their definitions are shown in equations (1)–(4):

$$MAE = \frac{1}{n} \sum_{t=1}^n |AC(t) - FC(t)| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n |AC(t) - FCST(t)|^2}{n}} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{AC(t) - FCST(t)}{AC(t)} \right| \times 100\% \quad (3)$$

$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum_{t=1}^n |FCST(t) - AVG(t)|^2}{\sum_{t=1}^n |AC(t) - AVG(t)|^2} \quad (4)$$

where AVG is the average of sample *t*, AC is the actual value of sample *t*, FCST is the predicted value of sample *t*, and *n* is the number of samples. Lower MAE, RMSE, and MAPE indicate a smaller prediction error of the model. R<sup>2</sup> is the ratio of the explained sum of squares (ESS) to the total sum of squares (TSS), where a higher R<sup>2</sup> indicates a better fit.

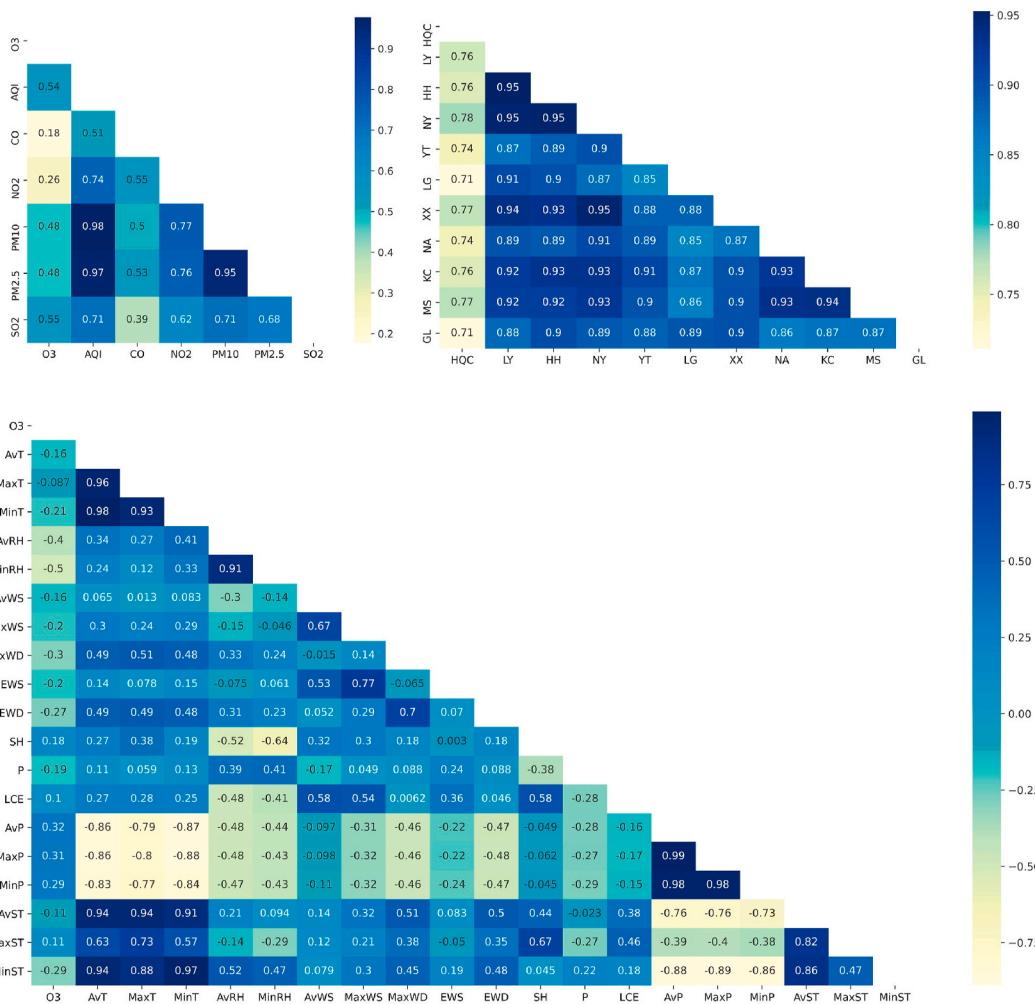


Fig. 5. Factor correlation coefficients.

### 3. Results and discussion

In this section, a variety of benchmark models are first compared. The compared models include MLR, RF, XGBoost, RNN, LSTM and SVR models. Then, based on the three different strategies, the detailed comparison and discussion of various models are carried out. Finally, the proposed hybrid forecasting model developed based on the three strategies is introduced in detail. The experiment involved ozone concentration data collected at the Shenzhen HQC station and NA station.

#### 3.1. Case study

The data used at the HQC and NA stations were divided into two parts. (1) The first part of the data set time scale was from August 31, 2019, to August 30, 2020; 306 days from the period of August 31, 2019, to July 1, 2020, were used as the training data set, and 60 days from the period of July 2, 2020, to August 30, 2020, were used as the test data set to test the prediction performance of the model. (2) The second part of the data set time scale was from August 31, 2016, to August 30, 2020; 1389 days from the period of August 31, 2016, to July 1, 2020, were used as the training data set, and 60 days from the period of July 2, 2020, to August 30, 2020, were used as the test data set. The important parameters and value ranges of the model are shown in Table 2.

In the comparison experiments of the benchmark model, we used the grid search (GS) method to optimize the parameters of each model. The results are shown in Table 3. The prediction results of the LSTM and SVR models at the two stations are better than those of the MLR model, tree

Table 2

The important parameters in the models (the format of the value range is as follows: [start, end, interval]).

Model	Parameter	Value range
RF	n_estimators	[500, 1000, 50]
	min_samples_split	[1, 10, 1]
	learning_rate	[0.01, 0.1, 0.01]
	n_estimators	[500, 1000, 50]
	max_depth	[1, 10, 1]
	min_child_weight	[1, 10, 1]
	subsample	0.5
XGBoost	colsample_bytree	0.8
	gamma	0
	reg_alpha	0
	reg_lambda	1
	Hidden_size	1, 5, 10, 15, 20
	Epoch	[100, 500, 100]
	Dropout	[0, 0.4, 0.1]
RNN & LSTM	Batch_size	1, 4, 8, 12, 16
	C	1, 5, 10, 50, 100
	gamma	0.01, 0.1, 1, 10
SVR		

models (RF and XGBoost) and the RNN model. The basic principles and characteristics of the LSTM and SVR models differ. As a typical deep learning model, the LSTM model has excellent predictive performance in dealing with time series forecasting problems, while the SVR model, as a typical traditional machine learning model, is very suitable for dealing with high-dimensional, small-sample prediction problems. Both have

**Table 3**

Results of the comparison of multiple models at the two stations (1-year data set).

Stations	Model	RMSE	MAE	MAPE	R <sup>2</sup>
HQC	MLR	16.8589	12.0243	36.0685	0.4347
	RF	14.8032	10.7221	32.9955	0.5642
	XGBoost	12.3641	10.2876	23.6543	0.5882
	RNN	14.3457	10.3383	25.5857	0.5907
	LSTM	14.0588	9.8924	28.7607	0.6068
	SVR	12.8742	8.7796	24.0158	0.6703
NA	MLR	12.9509	10.0415	24.8766	0.4915
	RF	12.5241	9.6376	22.6543	0.5182
	XGBoost	11.2060	9.4338	23.2398	0.5196
	RNN	12.5840	9.5198	23.6762	0.5097
	LSTM	12.2361	9.3810	23.2028	0.5226
	SVR	11.6485	7.9279	18.0897	0.5673

good robustness and generalization performance. Therefore, the strategy experiments conducted in this study are mainly based on the LSTM and SVR models.

### 3.1.1. Experiment I: HQC station

From Table 4, the following conclusions can be drawn:

- (1) The comparison of the LSTM and SVR models shows that the RMSE, MAE, and MAPE values of the SVR model are all lower than those of the LSTM model, indicating that the robustness and accuracy of the SVR model are higher than those of the LSTM model.
- (2) The analysis of the three model optimization strategies shows that the proper application of the three strategies can improve the prediction accuracy of the benchmark model, and the accuracy of the model established under the decomposition algorithm strategy improved by approximately 27%. The accuracy of the model established by combining the other two strategies is relatively low, only approximately 2.5%. Analyzing the effects of four

different decomposition algorithms, we can see that the WD algorithm produces the best result, followed from high to low by EEMD, CEEMDAN, and EMD.

Compared with the accuracy of the benchmark model, the model established under the strategy of adding factors is uncertain, indicating that the potential correlation between the added factor and the target value affects the actual prediction result. To confirm this conjecture, the indicators with the highest Pearson correlation coefficients among the three factors were selected (other air pollutants, meteorological factors and ozone factors at other surrounding stations), followed by SO<sub>2</sub>, MinRH, and ozone data from the NY station. The two benchmark models were each combined with three indicators to establish six models: LSTM-AP-SO<sub>2</sub>, SVR-AP-SO<sub>2</sub>, LSTM-MF-MinRH, SVR-MF-MinRH, LSTM-MS-NY and SVR-MS-NY. Compared with the benchmark model, the accuracy steadily improved. SVR-MS-NY and SVR-MF-MinRH have relatively high accuracy, which confirms that selecting appropriate indicators can improve the accuracy of the prediction model. The greater the correlation between the parameter factor and the target value is, the better the model training and prediction effects.

The models established under the strategy of adding data include LSTM-5Y and SVR-5Y. Compared with the benchmark model using the 1-year data set, the prediction accuracy of these two models increased steadily, which indicates that increasing the amount of training data in the training set can enable the benchmark model to obtain more time-series feature learning.

(3) Table 4 shows that the accuracy of the model established by combining the decomposition algorithm strategy and the adding data strategy is higher than that of the model incorporating only a single strategy.

The accuracy of the WD-LSTM-5Y model is higher than that of the WD-SVR-5Y model, which is the only case in Experiment I where the accuracy of the LSTM exceeds the SVR, with R<sup>2</sup> values of 0.9485 and 0.9395, respectively. Compared with the benchmark model, the accuracy is improved by approximately 34% and 27%. This shows that the combination of the decomposition algorithm strategy and the adding data strategy can significantly improve the accuracy of the benchmark model, and the improvement of the LSTM model is greater than that of the SVR model.

Among all of the established models, the accuracy of the models based on adding the decomposition algorithm strategy shows the greatest improvement. The next section provides key research on this point and a related discussion. Figs. 6 and 7 are both composed of two parts. The upper part (a-k) of the figure is a comparison chart of model predictions, and the lower part (l) is a box distribution chart of the model-predicted values and actual values. The prediction result of the benchmark model is not good; compared with the actual value, the prediction result has a relatively obvious lag. The results may often not meet the forecasting accuracy requirements. Obviously, the prediction accuracy of the model based on different decomposition algorithms is significantly better than that of the benchmark model. The model incorporating the decomposition algorithm can make a good prediction of the actual ozone value change, and the problem of the prediction lag is greatly improved. Among the four decomposition algorithms, the deviation between the predicted and actual values is smallest for the model combined with the WD algorithm.

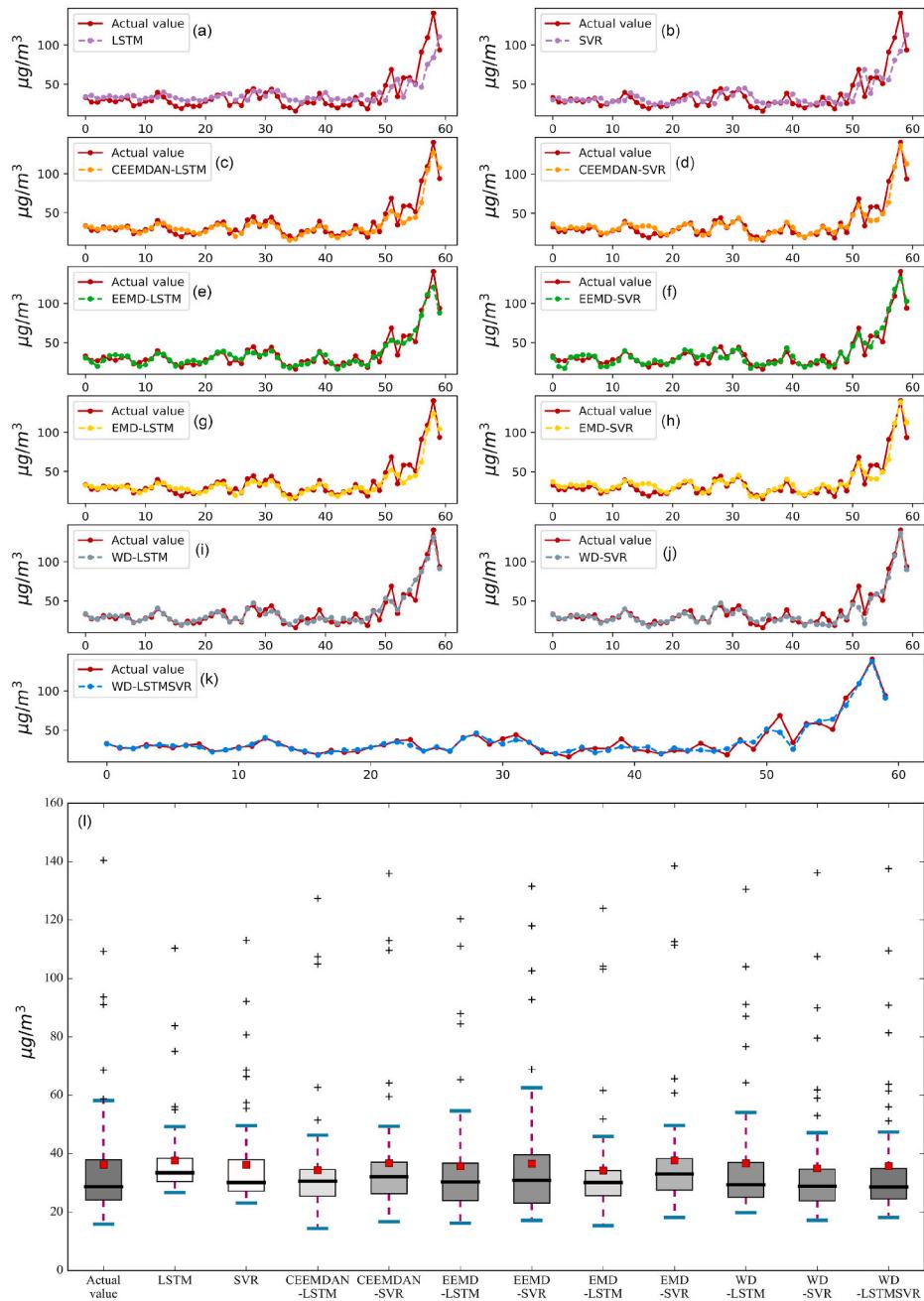
### 3.1.2. Experiment II: NA station

A good model should have a high generalization ability. Therefore, the purpose of Experiment II was to verify the prediction performance of the different models established under the optimal strategy. In addition, the applicability of the model was further discussed and verified. A representative station in Shenzhen, the NA station, was selected as the verification station.

In Experiment II, further research on the strategy of adding factors was not performed for two main reasons: (1) According to the results of

**Table 4**  
Model accuracy evaluations (HQC station).

Model		RMSE	MAE	MAPE	R <sup>2</sup>
Benchmark Model	LSTM	14.0588	9.8924	28.7607	0.6068
	SVR	12.8742	8.7796	24.0158	0.6703
Improved Model	WD-LSTM	5.9526	3.8921	11.4672	0.9295
	WD-SVR	5.8230	3.7826	11.2113	0.9325
CEEMDAN-SVR	EMD-LSTM	7.6063	5.2295	13.8080	0.8849
	EMD-SVR	7.0750	4.7501	14.5391	0.9004
CEEMDAN-SVR	EEMD-LSTM	6.1369	4.5792	13.6930	0.9250
	EEMD-SVR	5.9783	4.5272	13.8285	0.9289
CEEMDAN-SVR	CEEMDAN	7.5303	5.2685	14.0482	0.8872
	LSTM	6.9572	4.4075	12.9928	0.9037
CEEMDAN-SVR	LSTM-5Y	13.6025	9.2588	26.5690	0.6320
	SVR-5Y	12.3838	8.5556	23.8071	0.6949
CEEMDAN-SVR	LSTM-AP-All	14.1535	10.3353	32.1639	0.6289
	SVR-AP-All	14.2782	10.5346	32.3546	0.6224
CEEMDAN-SVR	LSTM-MF-All	16.3360	11.9123	34.7986	0.5057
	SVR-MF-All	15.8146	11.6586	36.3810	0.5367
CEEMDAN-SVR	LSTM-MS-All	14.7699	9.1198	23.2407	0.5959
	SVR-MS-All	14.7420	9.4271	23.7806	0.5974
CEEMDAN-SVR	LSTM-AP-SO <sub>2</sub>	14.2449	10.7515	33.1946	0.6241
	SVR-AP-SO <sub>2</sub>	12.9601	8.9529	25.1326	0.6889
CEEMDAN-SVR	LSTM-MF-	14.1378	9.8784	27.8293	0.6298
	MinRH	12.7148	9.0862	26.3011	0.7005
CEEMDAN-SVR	SVR-MF-MinRH	13.9959	9.4249	25.9204	0.6371
	LSTM-MS-NY	12.5415	8.4170	22.8302	0.7086
CEEMDAN-SVR	WD-LSTM-5Y	5.0845	3.6842	10.6997	0.9485
	WD-SVR-5Y	5.5153	3.7478	10.4921	0.9395
CEEMDAN-SVR	WD-LSTMSVR	5.0293	3.3461	10.0479	0.9496
	WD-LSTMSVR-5Y	4.3623	3.0842	8.8788	0.9621

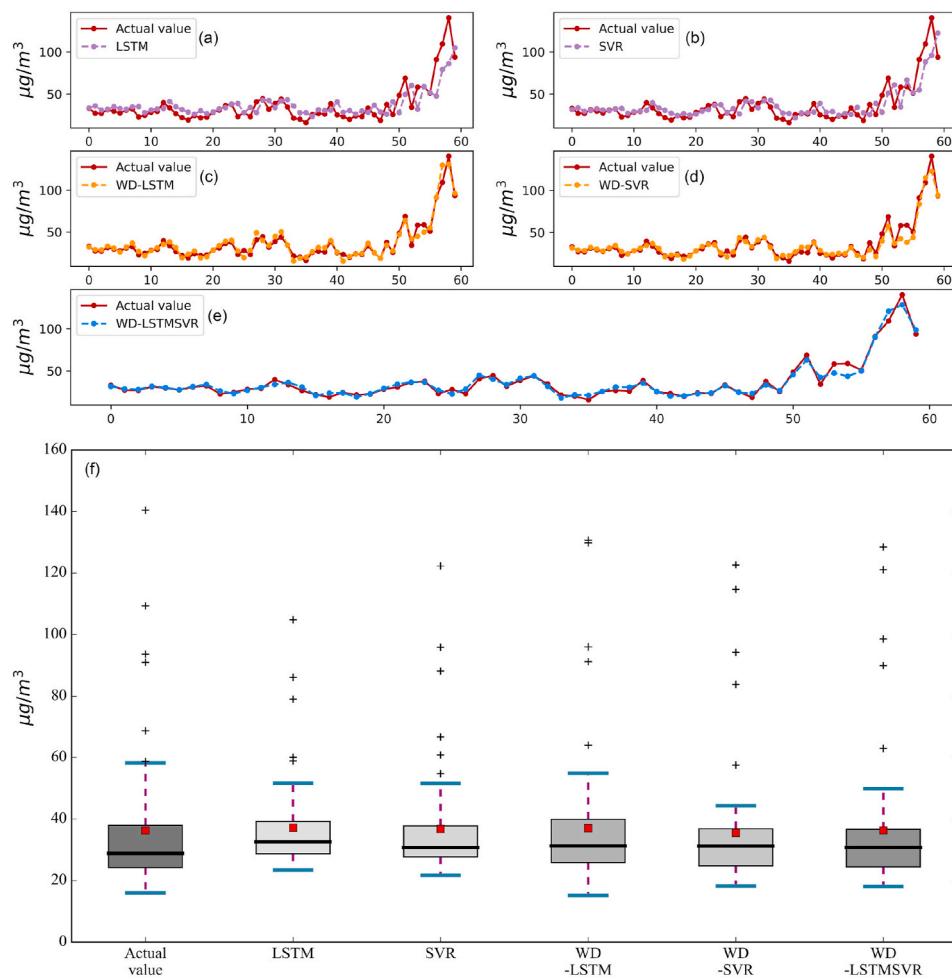


**Fig. 6.** Comparison between the model prediction and actual value (HQC station 1-year data set of 60-day forecast results from July 2, 2020, to August 30, 2020; the closer the color of the box is to the actual value box, the higher the accuracy of the model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Experiment I, the accuracy of the model established under the strategy of adding factors was less improved than that of the other models or even lower than that of the benchmark models. In addition, available data are insufficient for further research (e.g., there are no air pollutant data and meteorological data at the same point). (2) The NA station is the farthest from the meteorological center station. However, meteorological changes are often expressed in local areas and microscale features, and meteorological centers cannot provide reasonable and effective indicators for ozone prediction at NA station.

Table 5 provides some conclusions that differ from those of Experiment I. First, the accuracy of the benchmark model established in Experiment II is lower than that in Experiment I; a similar pattern is evident when the other established models are compared. Second, in Experiment II, comparing the combined LSTM and combined SVR

models under the same conditions revealed that all combined SVR models have higher accuracy. In Experiment I, the prediction accuracy of the WD-LSTM-5Y model was higher than that of the WD-SVR-5Y model, whereas the opposite result was obtained in Experiment II. Compared with the benchmark model, the prediction accuracy of the WD-LSTM-5Y model increased by approximately 39% and that of the WD-SVR-5Y model increased by approximately 37%. Therefore, we can still conclude that combining the adding decomposition algorithm and adding data strategies can significantly improve the accuracy of the LSTM model and that the performance is better than that of the SVR model. Finally, the accuracy of the model established under the decomposition algorithm strategy is improved the most. Although the combination of the WD algorithm is still the best among the four decomposition algorithms, the performance ranking of the other



**Fig. 7.** Comparison of model predictions with actual value (HQC station 5-year data set of 60-day forecast results from July 2, 2020, to August 30, 2020; the closer the color of the box is to the actual value box, the higher the accuracy of the model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 5**  
Model accuracy evaluations (NA station).

Model		RMSE	MAE	MAPE	$R^2$
Benchmark Model	LSTM	12.2361	9.381	23.2028	0.5226
	SVR	11.6485	7.9279	18.0897	0.5673
Improved Model	WD-LSTM	5.5846	3.7374	8.896	0.9005
	WD-SVR	4.9812	3.5407	8.4923	0.9208
	EMD-LSTM	7.7044	5.3012	11.8072	0.8107
	EMD-SVR	5.7153	4.5545	11.2521	0.8958
	EEMD-LSTM	7.5266	5.7755	13.9186	0.8193
	EEMD-SVR	7.4315	5.949	14.9225	0.8239
	CEEMDAN-LSTM	7.1681	5.3151	12.4301	0.8361
	CEEMDAN-SVR	5.9113	4.635	11.3538	0.8886
	LSTM-5Y	11.8877	8.7993	21.1559	0.5494
	SVR-5Y	11.1964	7.5702	17.1598	0.6003
Hybrid Model	WD-LSTM-5Y	5.3023	3.6045	8.4109	0.9104
	WD-SVR-5Y	4.3898	2.9895	6.6183	0.9385
	WD-LSTMSVR	4.7383	3.3206	7.9348	0.9284
	WD-LSTMSVR-5Y	3.7402	2.7042	6.3132	0.9553

decomposition algorithms in this experiment is different from that in Experiment I. Among the combined SVR models, WD > EMD > CEEMDAN > EEMD; among the combined LSTM models, WD > CEEMDAN > EEMD > EMD. These three decomposition algorithms have no absolute advantages or disadvantages in time series forecasting, and the

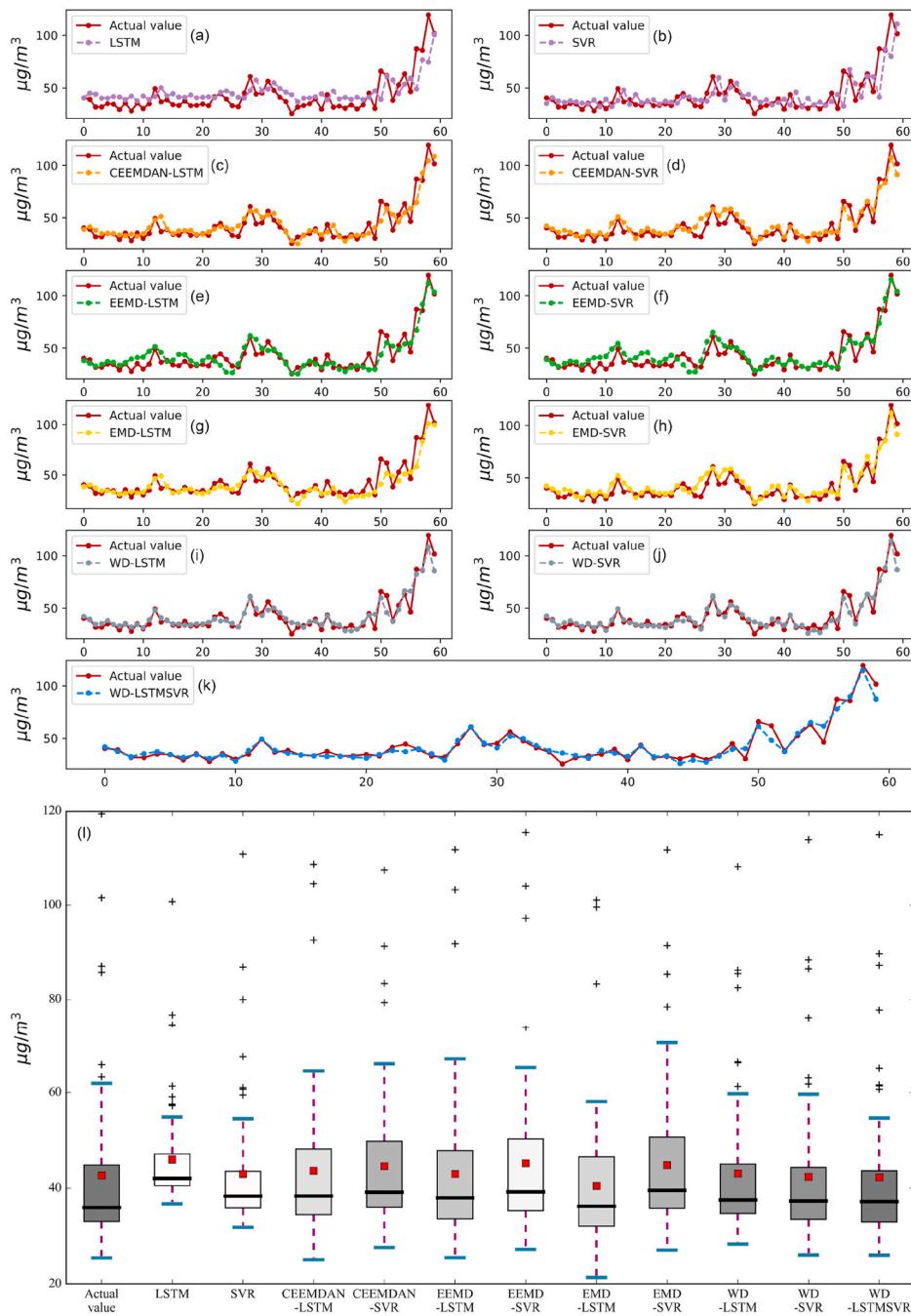
performance of the three algorithms will be different for data with different characteristics.

Figs. 8 and 9 show that the prediction results of the model based on different decomposition algorithms are significantly better than those of the benchmark models and can predict the actual ozone value changes well; a forecast lag is rarely seen.

### 3.2. Discussion of the three model optimization strategies

The analysis in Sections 3.1.1 and 3.1.2 showed that the adding data strategy is effective and stable. Increasing training data within a certain scale allows the model to mine more features from the data and thus improve the accuracy of modeling. For the adding factors strategy, adding all other pollutants provided the best performance, followed by adding ozone data from nearby stations and, finally, adding meteorological parameters. These results indicate that ozone from nearby regions plays the most important role in ozone prediction.

Considering the possibility of ozone transmission from other regions, greater distances between the meteorological station and the ozone monitoring station may lead to an underestimation of the impact of meteorological factors on the model prediction. Therefore, we believe that ozone data from nearby areas are more important for ozone prediction, followed by the influence of meteorological factors and the influence of other air pollutants. Of course, this does not mean that other air pollutants are not important for ozone prediction. In contrast, we think this may be the breakthrough point in ozone prediction. Because

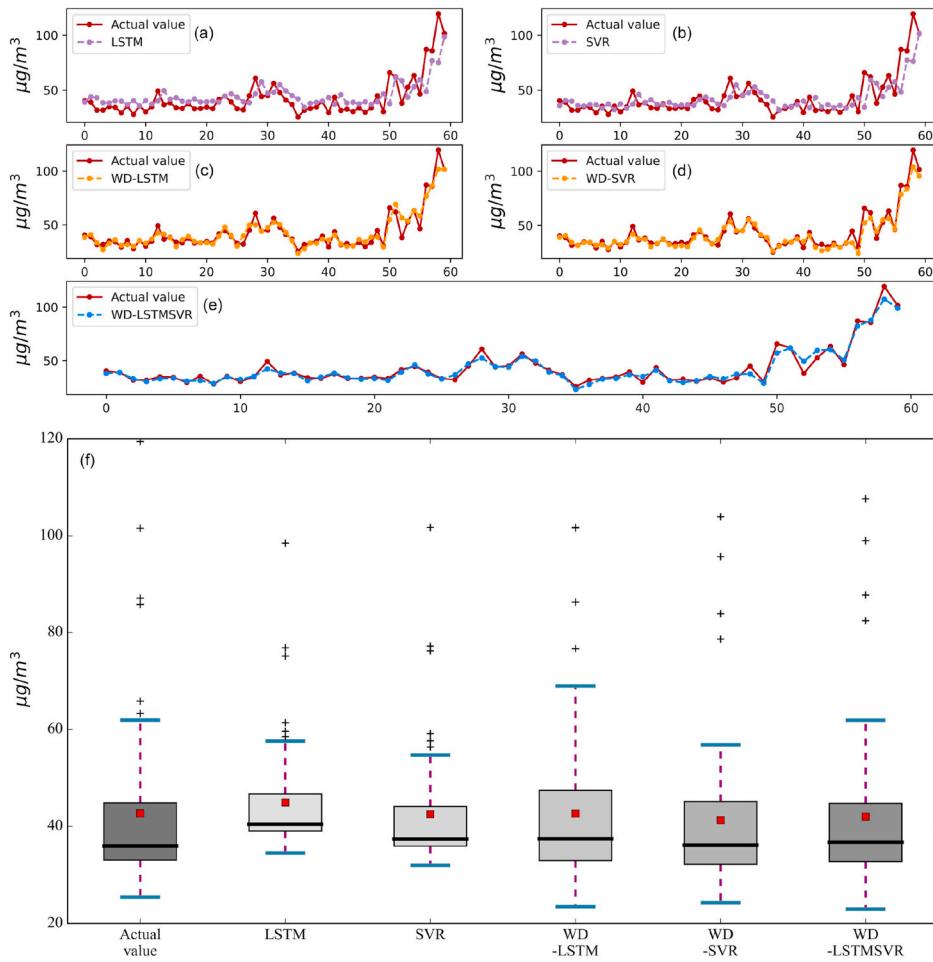


**Fig. 8.** Comparison between model prediction and actual value (NA station 1-year data set of 60-day forecast results from July 2, 2020, to August 30, 2020; the closer the color of the box is to the actual value box, the higher the accuracy of the model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the chemical mechanism of ozone formation is very complicated, it is difficult to find suitable characteristic pollutants and time periods for prediction, which also requires further research. We recommend considering the addition of NO<sub>x</sub> and VOCs data to the model.

The HQC station is the closest station to the meteorological center, but there is still a certain geographic distance. To aid further discussion and analysis of the relationship between the meteorological factors and prediction models, additional data collection at the same location is warranted. In addition, we suggest considering meteorological trajectory models such as HYSPLIT for regional airflow trajectory simulation and then incorporating the station data on the trajectory into the target station prediction model. Doing so will improve the scientific underpinnings of the model.

In this study, four different decomposition methods were used and compared in two experiments. The results show that the accuracy of the model combined with the WD algorithm is higher than that of the EMD algorithms (EMD, EEMD and CEEMDAN). This result may reflect differences in the working principles of the WD and EMD algorithms (Monjoly et al., 2017). The WD algorithm primarily uses a filter based on the mother wavelet, while the EMD algorithm mainly calculates the decomposed data in an adaptive and data-driven manner. When the decomposed data meet certain conditions, the decomposition will stop. The screening process and boundary effect issues are considered shortcomings of the EMD algorithm. The WD algorithm is not subject to this restriction because of its strict mathematical algorithm. The results show that the computational efficiency and results of the WD algorithm are



**Fig. 9.** Comparison between model prediction and actual value (NA station 5-year data set of 60-day forecast results from July 2, 2020, to August 30, 2020; the closer the color of the box is to the actual value box, the higher the accuracy of the model). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

better than those of the EMD algorithms, which also means that the WD algorithm is more suitable for large-scale signal analysis and prediction.

### 3.3. The novel hybrid forecasting model

#### 3.3.1. The process of developing the model

Based on the analysis and discussion of the three model optimization strategies, we found that using the decomposition algorithm strategy can significantly improve the accuracy of the benchmark model; among the four decomposition algorithms, the WD algorithm is best. In addition, we found that the different models differ in their performance in predicting different feature data. As a typical deep learning model, the LSTM model has excellent predictive performance in dealing with time series trend forecasting problems. This feature is often very suitable for processing approximate sequences and high-level detailed sequences in decomposition algorithms, while the SVR model is a typical traditional machine learning model, which is very suitable for dealing with high-dimensional, small-sample prediction problems. In general, the SVR model has high predictive performance on low-level detailed sequences.

Therefore, the key to the hybrid algorithm is to apply the WD algorithm to the original data. First, the WD algorithm converted the original one-dimensional data into multidimensional information and mined the hidden information of the original data. Then, the LSTM and SVR models were used to train each layer of the sequence training set data. Because the two benchmark models have different prediction effects on the decomposition sequence of different features, we compared the

accuracy of the training data of each layer of the different models and selected the benchmark model with higher training accuracy to predict and reconstruct the corresponding layer sequence. The details are shown in Fig. 10. Finally, the final forecasting results were combined, and the accuracy was evaluated. The key procedures of the algorithm are as follows.

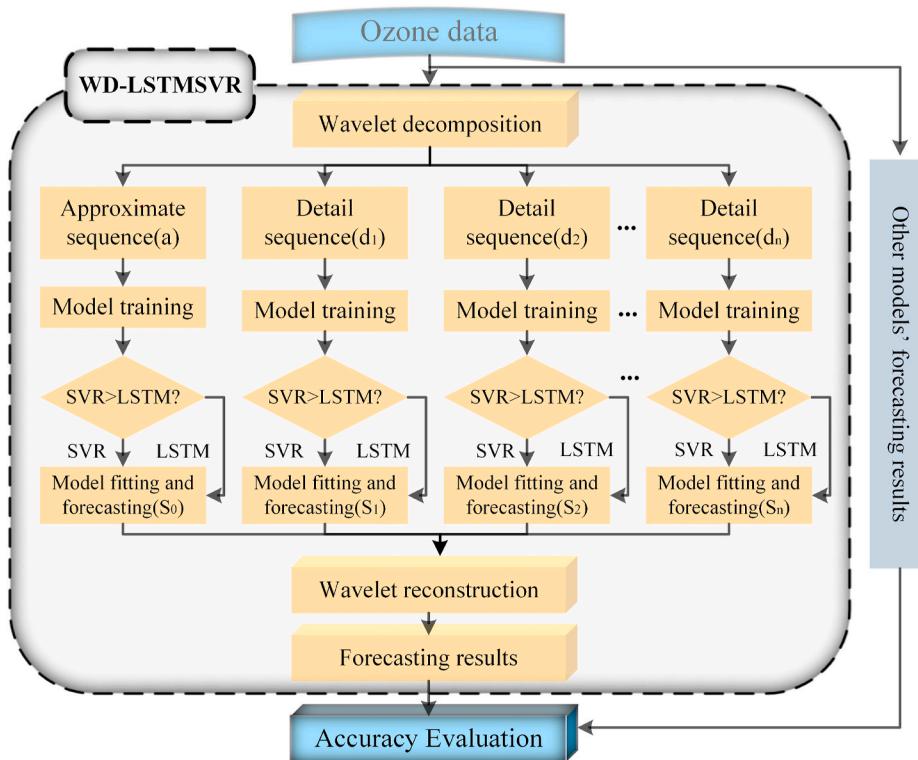
**Input:** Time series of ozone concentrations.

**Output:** Forecasted value of ozone concentrations and the accuracy of the model.

**Step 1.** The ozone data were used as the model input data after discrete standardization.

**Step 2.** A suitable wavelet basis function was selected, WD was performed on the input data, and the input one-dimensional data were decomposed into approximate sequences ( $a_i$ ) and detailed sequences ( $d_1, d_2, \dots, d_n$ ). In this study, the  $db4$  wavelet basis function was chosen to decompose the ozone concentration sequence into an approximate sequence and four detailed sequences.

**Step 3.** The ARIMA partial algorithm and the PSO algorithm were used to optimize the parameters of the WD sequence obtained in **Step 2**. The forward step size was mainly determined by the *ACF* and *PACF* diagrams and the *AIC* and *SC* values and was finally set to 3. The activation function used in this study was the ReLU, which uses MSE as a loss function and a "descent" method to avoid overfitting. In addition, the adaptive Adam algorithm was used to optimize the weights and biases of the LSTM network, and the number of iterations was 500. The important



**Fig. 10.** Framework of the novel hybrid model.

parameters and related value ranges used in this step are shown in [Table 6](#). Furthermore, the LSTM and SVR models were established to train the training set data of each layer of the sequence.

**Step 4.** The model with high training data accuracy was selected for each layer in [Step 3](#) to perform sequence prediction of the corresponding layer, and the sequence prediction result of each layer was obtained.

**Step 5.** Wavelet reconstruction was performed on the prediction results of each layer obtained in [Step 4](#) to form the final ozone prediction value, which is the prediction result of the hybrid forecasting model.

**Step 6.** We compared and evaluated the prediction accuracies of the novel hybrid forecasting model (WD-LSTMSVR) and other models established in this study.

### 3.3.2. Model performance

Based on the above steps, a novel hybrid forecasting model (WD-LSTMSVR) was developed. To verify the performance of the novel hybrid model, 28 models established in this study were compared. The specific model accuracy comparison is shown in [Tables 4 and 5](#) and [Figs. 6–9](#), and the following conclusions can be drawn:

**Table 6**  
The important parameters in the algorithms.

Algorithm	Parameter	Value range
PSO	number of iterations of the particles	50
	population size	100
	inertia factor of velocity	0.5
	acceleration factor $C_1$ & $C_2$	1.5
WD-LSTMSVR	Hidden_size	[1, 20]
	Dropout	[0, 0.1]
	Batch_size	[1, 12]
	C	[1, 100]
	gamma	[0.01, 10]

- (1) The analysis of the novel hybrid model in this study shows that the hybrid model has significantly higher accuracy than all other models.
- (2) The novel hybrid model has the best performance, and the overall trend is basically the same as the actual value. As shown in [Figs. 6–9](#), the outlier distribution of the hybrid model is consistent with the actual ozone outlier distribution, which shows that the hybrid model can predict the extreme values in the data well and that the stability and applicability have been further improved compared with the other models.
- (3) The developed hybrid model combines the advantages of the different algorithms well and provides good performance in terms of both the model accuracy and the generalization ability. This hybrid algorithm idea may provide support for other models and pollutant predictions.

### 3.3.3. Sensitivity analysis of the model parameters

To test the stability of the model, sensitivity analysis of the parameters of the main benchmark models in the novel hybrid model, i.e., LSTM and SVR, was conducted. More specifically, the number of hidden neurons of LSTM, the number of epochs, the dropout and batch size, the penalty parameter C and the kernel coefficient g in the SVR model were further studied. The changes in the performance of the models in response to changes in these parameters were observed to measure the stability of the models.

In the sensitivity analysis experiment, we used the GS method to perform modeling experiments with different parameter combinations. Due to the large number of potential parameter combinations, we were unable to evaluate every parameter combination. Instead, we present selected parameter combinations in [Tables 7 and 8](#). The following conclusions can be drawn:

- (1) The best parameters for the prediction model for ozone concentration data differ slightly at different points, but changes in the basic parameters have similar effects on the model.

**Table 7**  
Model parameters sensitivity analysis (HQC station).

Parameter		RMSE	MAE	MAPE	R <sup>2</sup>
Hidden_size	1	31.3659	29.8994	105.6787	-0.9567
	5	6.5560	4.2413	11.5642	0.9145
	10	6.9122	4.5330	12.5700	0.9050
	15	6.2064	4.0143	11.2533	0.9234
	20	6.4617	4.2606	11.6769	0.9170
Epoch	100	7.4272	5.6346	16.8803	0.8903
	200	6.4817	4.1922	11.3974	0.9164
	300	6.1513	4.0162	11.5459	0.9247
	400	6.0138	3.8266	11.2281	0.9251
	500	6.0954	3.8187	11.1997	0.9263
Dropout	0	6.5874	4.3439	11.9386	0.9137
	0.1	17.3342	14.8249	41.7263	0.4024
	0.2	33.7382	32.3419	99.4217	-1.2639
	0.3	48.0589	46.7778	146.0828	-3.5936
	0.4	62.5193	61.1772	192.3142	-6.7738
Batch_size	1	6.2382	4.2158	11.8382	0.9226
	4	6.0802	4.2122	12.2179	0.9265
	8	6.7610	4.4345	11.9810	0.9091
	12	14.7905	13.7767	47.4206	0.5649
	16	14.9220	13.8564	48.2250	0.5571
C = 1	g = 0.01	6.9950	4.5434	12.8412	0.9027
C = 1	g = 0.1	6.2413	4.0578	11.4177	0.9225
C = 1	g = 1	5.8859	4.0876	11.7881	0.9256
C = 1	g = 10	9.1702	6.4056	19.8886	0.8328
C = 5	g = 0.01	6.2726	3.9922	11.2289	0.9217
C = 5	g = 0.1	5.9525	3.8249	10.9656	0.9242
C = 5	g = 1	5.9391	3.8952	10.9976	0.9299
C = 5	g = 10	8.3034	5.9862	19.2740	0.8629
C = 10	g = 0.01	6.1107	3.9161	11.3082	0.9223
C = 10	g = 0.1	6.0993	3.8147	11.8103	0.9242
C = 10	g = 1	5.9945	3.8231	11.5116	0.9251
C = 10	g = 10	8.3226	5.9810	19.1619	0.8622
C = 50	g = 0.01	6.1120	3.9399	11.2041	0.9257
C = 50	g = 0.1	6.0181	4.1013	11.5627	0.9261
C = 50	g = 1	5.9823	3.9807	11.1194	0.9298
C = 50	g = 10	8.5815	6.1400	19.5514	0.8535
C = 100	g = 0.01	6.8163	4.5319	12.1097	0.9156
C = 100	g = 0.1	6.2563	3.9856	11.6782	0.9288
C = 100	g = 1	5.9817	3.8869	11.3939	0.9293
C = 100	g = 10	8.6454	6.1141	19.4852	0.8513

- (2) In the HQC and NA data sets, increasing Hidden\_size, Dropout and Batch\_size did not further increase the model performance. This is mainly because the larger the values of these parameters are, the greater the risk of model overfitting. As the number of epochs increases, the performance of the model continues to improve, thereby indicating that the parameters in this range of the model have not reached the point of performance degradation. However, increasing the number of epochs greatly limits the calculation speed of the model, easily causes overfitting, and significantly reduces the generalization ability.
- (3) The performance of the models in terms of the parameter sensitivity differs slightly between the two station data sets. Among the evaluated parameters, Dropout and Batch\_size have the greatest impact on the model. In other words, the model is more sensitive to changes in these parameters. In addition, we found that the sensitivity of the model to changes in C and g differs between the two data sets. The sensitivity of the model established at the NA station to the g value is significantly greater than that of model established at the HQC station. For the parameter C, there is little difference between the two stations.
- (4) As Hidden\_size and Batch\_size change, the performance of the model changes accordingly, and the model becomes prone to overfitting or easily converges. When Hidden\_size increases, the complexity of the model may increase, and overfitting may occur; however, the modeling effect may be improved to a certain extent. As Batch\_size increases, the number of epochs required to achieve the same accuracy increases. In fact, the model often does

**Table 8**  
Model parameters sensitivity analysis (NA station).

Parameter		RMSE	MAE	MAPE	R <sup>2</sup>
Hidden_size	1	33.5797	32.1619	88.3373	-2.5952
	5	5.6518	3.9406	9.1164	0.8982
	10	6.0114	4.2183	9.8540	0.8848
	15	5.6686	3.8454	8.8784	0.8975
	20	5.6668	3.8930	9.0536	0.8976
Epoch	100	6.2922	4.2721	9.5775	0.8738
	200	6.0034	4.4917	10.9005	0.8751
	300	5.9835	4.1819	9.4814	0.8843
	400	5.9708	3.9704	9.2413	0.8849
	500	5.6392	3.9075	9.1080	0.8925
Dropout	0	5.5294	3.7513	9.0603	0.8938
	0.1	14.4658	13.0918	31.1419	0.3328
	0.2	27.5964	26.6190	65.2878	-1.4281
	0.3	42.2098	41.1847	101.5266	-4.6806
	0.4	52.6740	51.6788	128.0853	-7.8463
Batch_size	1	5.9270	3.9074	9.0107	0.8978
	4	13.9244	12.9340	34.7653	0.3818
	8	5.6568	3.8990	8.9934	0.8980
	12	15.7053	14.7210	39.9619	0.2136
	16	17.0295	15.9543	43.4646	0.0754
C = 1	g = 0.01	5.9338	4.0461	9.1705	0.8877
C = 1	g = 0.1	5.7090	3.9914	9.3099	0.8961
C = 1	g = 1	6.3421	4.8896	11.8235	0.8718
C = 1	g = 10	12.0931	8.9559	24.1771	0.5337
C = 5	g = 0.01	5.3695	3.6669	8.4628	0.9081
C = 5	g = 0.1	5.3413	3.7457	8.9036	0.9090
C = 5	g = 1	5.7486	4.5529	11.2444	0.8946
C = 5	g = 10	11.5125	8.5166	22.7517	0.5774
C = 10	g = 0.01	5.2080	3.6194	8.7390	0.9135
C = 10	g = 0.1	5.3662	3.7126	8.8820	0.9082
C = 10	g = 1	5.6185	4.4613	11.0975	0.8994
C = 10	g = 10	11.4263	8.4535	22.5023	0.5837
C = 50	g = 0.01	5.1800	3.5952	8.7282	0.9144
C = 50	g = 0.1	5.1758	3.5824	8.7073	0.9146
C = 50	g = 1	5.8506	4.5376	11.2627	0.8909
C = 50	g = 10	11.6039	8.3917	22.4227	0.5707
C = 100	g = 0.01	5.1836	3.5626	8.5206	0.9143
C = 100	g = 0.1	5.1090	3.5359	8.6280	0.9168
C = 100	g = 1	5.9681	4.5837	11.4021	0.8864
C = 100	g = 10	11.7775	8.5068	22.8250	0.5577

not configure a very high number of epochs, so the accuracy of the model decreases at a large Batch\_size.

#### 4. Conclusion

In this study, we proposed three model optimization strategies, i.e., adding decomposition algorithms, adding data and adding factors, to model and predict surface ozone concentrations. After comparing the prediction results of models with different optimization strategies, we found that the strategy of adding a decomposition algorithm provides a more stable and improved performance, particularly when adding the WD algorithm, which yields the most accurate prediction results. Based on this, we developed the hybrid WD-LSTMSVR model, in which the WD algorithm converts one-dimensional input data into multidimensional data and then trains each layer of the data using LSTM and SVR models. Furthermore, the best model to forecast the corresponding layer sequence was selected. Two observational data sets at the HQC and NA stations were used to test the performance of this novel hybrid forecasting model. Compared with the results for other benchmark and optimization model results, the hybrid model gave the smallest deviation and the most accurate predictions, suggesting that the WD-LSTMSVR model developed in this study can improve the accuracy of ozone prediction in Shenzhen. In addition, the WD-LSTMSVR model can provide support and ideas for the prediction of air pollutants in other regions or cities after appropriate parameter changes and optimization.

## CRediT authorship contribution statement

**Yong Cheng:** Conceptualization, Methodology, Software, Writing – original draft. **Qiao Zhu:** Methodology, Writing – review & editing. **Yan Peng:** Data curation, Writing – review & editing. **Xiao-Feng Huang:** Supervision, Writing – review & editing. **Ling-Yan He:** Supervision, Methodology, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the IER foundation (NO. HT-JD-CXY-201903) and Science and Technology Plan of Shenzhen Municipality (JCYJ20180713112202572).

## References

- Abo-Khalil, A.G., Dong-Choon, L., 2008. MPPT control of wind generation systems based on estimated wind speed using SVR. *IEEE Trans. Ind. Electron.* 55 (3), 1489–1490. <https://doi.org/10.1109/tie.2007.907672>.
- Afzali, A., Rashid, M., Afzali, M., Younesi, V., 2017. Prediction of air pollutants concentrations from multiple sources using AERMOD coupled with WRF prognostic model. *J. Clean. Prod.* 166, 1216–1225. <https://doi.org/10.1016/j.jclepro.2017.07.196>.
- AlOmar, M.K., Hameed, M.M., AlSaadi, M.A., 2020. Multi hours ahead prediction of surface ozone gas concentration: robust artificial intelligence approach. *Atmospheric Pollution Research* 11 (9), 1572–1587. <https://doi.org/10.1016/j.apr.2020.06.024>.
- Chen, J., Yu, J., Song, M., Valdmanis, V., 2019. Factor decomposition and prediction of solar energy consumption in the United States. *J. Clean. Prod.* 234, 1210–1220. <https://doi.org/10.1016/j.jclepro.2019.06.173>.
- Cheng, Y., Zhang, H., Liu, Z., Chen, L., Wang, P., 2019. Hybrid Algorithm for Short-Term Forecasting of PM2.5 in China. *Atmospheric Environment*, pp. 264–279. <https://doi.org/10.1016/j.atmosenv.2018.12.025>.
- Ghimire, S., Deo, R.C., Downs, N.J., Raj, N., 2018. Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Rem. Sens. Environ.* 212, 176–198. <https://doi.org/10.1016/j.rse.2018.05.003>.
- Haagen-Smit, A.J., 1952. Chemistry and physiology of Los Angeles smog. *Ind. Eng. Chem.* 44 (6), 1342–1346. <https://doi.org/10.1021/ie50510a045>.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Huang, J., Pan, X., Guo, X., Li, G., 2018. Health impact of China's Air Pollution Prevention and Control Action Plan: an analysis of national air quality monitoring and mortality data. *The Lancet Planetary Health* 2 (7), e313–e323. [https://doi.org/10.1016/s2542-5196\(18\)30141-4](https://doi.org/10.1016/s2542-5196(18)30141-4).
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* 454, 903–995. <https://doi.org/10.1098/rspa.1998.0193>, 1971.
- Kennedy, J., Eberhart, R., 2002. Particle Swarm Optimization, Icn95-International Conference on Neural Networks.
- Knüsel, B., Baumberger, C., Zumwald, M., Bresch, D.N., Knutti, R., 2020. Argument-based assessment of predictive uncertainty of data-driven environmental models. *Environ. Model. Software* 134. <https://doi.org/10.1016/j.envsoft.2020.104754>.
- Ko, C.-N., Lee, C.-M., 2013. Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter. *Energy* 49, 413–422. <https://doi.org/10.1016/j.energy.2012.11.015>.
- Konovalov, I.B., Beekmann, M., Meleux, F., Dutot, A., Foret, G., 2009. Combining deterministic and statistical approaches for PM10 forecasting in Europe. *Atmos. Environ.* 43 (40), 6425–6434. <https://doi.org/10.1016/j.atmosenv.2009.06.039>.
- Liu, H., Long, Z., 2020. An improved deep learning model for predicting stock market price time series. *Digit. Signal Process.* 102. <https://doi.org/10.1016/j.dsp.2020.102741>.
- Liu, H., Yin, S., Chen, C., Duan, Z., 2020. Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. *J. Clean. Prod.* 277 <https://doi.org/10.1016/j.jclepro.2020.124023>.
- Lu, W.-Z., Wang, D., 2014. Learning machines: rationale and application in ground-level ozone prediction. *Appl. Soft Comput.* 24, 135–141. <https://doi.org/10.1016/j.asoc.2014.07.008>.
- Ma, J., Li, Z., Cheng, J.C.P., Ding, Y., Lin, C., Xu, Z., 2020. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total Environ.* 705, 135771. <https://doi.org/10.1016/j.scitotenv.2019.135771>.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7), 674–693. <https://doi.org/10.1109/34.192463>.
- Mo, Y., Li, Q., Karimian, H., Fang, S., Tang, B., Chen, G., Sachdeva, S., 2020. A Novel Framework for Daily Forecasting of Ozone Mass Concentrations Based on Cycle Reservoir with Regular Jumps Neural Networks, 220. *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2019.117072>.
- Monjoly, S., André, M., Calif, R., Soubdhan, T., 2017. Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach. *Energy* 119, 288–298. <https://doi.org/10.1016/j.energy.2016.11.061>.
- Nie, Z., Shen, F., Xu, D., Li, Q., 2020. An EMD-SVR model for short-term prediction of ship motion using mirror symmetry and SVR algorithms to eliminate EMD boundary effect. *Ocean. Eng.* 217 <https://doi.org/10.1016/j.oceaneng.2020.107927>.
- Niu, D., Wu, F., Dai, S., He, S., Wu, B., 2020. Detection of long-term effect in forecasting municipal solid waste using a long short-term memory neural network. *J. Clean. Prod.* 290 <https://doi.org/10.1016/j.jclepro.2020.125187>.
- Ortiz-García, E.G., Salcedo-Sanz, S., Pérez-Bellido, A.M., Portilla-Figueras, J.A., Prieto, L., 2010. Prediction of hourly O<sub>3</sub> concentrations using support vector regression algorithms. *Atmos. Environ.* 44 (35), 4481–4488. <https://doi.org/10.1016/j.atmosenv.2010.07.024>.
- Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmos. Environ.* 128, 246–262. <https://doi.org/10.1016/j.atmosenv.2016.01.007>.
- Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., Jung, J., 2020. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Network* 121, 396–408. <https://doi.org/10.1016/j.neunet.2019.09.033>.
- Song, C., Fu, X., 2020. Research on different weight combination in air quality forecasting models. *J. Clean. Prod.* 261 <https://doi.org/10.1016/j.jclepro.2020.121169>.
- U.S.EPA, 2013. Integrated Science Assessment for Ozone and Related Photochemical Oxidants EPA/600/R-10/076F Office of Research and Development. National Centre for Environmental Assessment-RTP.
- Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE Trans. Neural Network* 10 (5), 988–999. <https://doi.org/10.1109/72.788640>.
- Wang, J., Bai, L., Wang, S., Wang, C., 2019. Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system. *J. Clean. Prod.* 234, 54–70. <https://doi.org/10.1016/j.jclepro.2019.06.201>.
- Wu, Q., Lin, H., 2019. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Sci. Total Environ.* 683, 808–821. <https://doi.org/10.1016/j.scitotenv.2019.05.288>.
- Zhu, S., Lian, X., Liu, H., Hu, J., Wang, Y., Che, J., 2017. Daily air quality index forecasting with hybrid models: a case in China. *Environ. Pollut.* 231 (Pt 2), 1232–1244. <https://doi.org/10.1016/j.envpol.2017.08.069>.