Development of a high-performance machine learning model to predict ground ozone pollution in typical cities of China

Yong Cheng, Ling-Yan He, Xiao-Feng Huang [*]

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

ARTICLE INFO

ABSTRACT

High ozone concentrations have adverse effects on human health and ecosystems. In recent years, the ambient ozone concentration in China has shown an upward trend, and high-quality prediction of ozone concentrations has become critical to support effective policymaking. In this study, a novel hybrid model combining wavelet decomposition (WD), a gated recurrent unit (GRU) neural network and a support vector regression (SVR) model was developed to predict the daily maximum 8 h ozone. We used the ground ozone observation data in six representative megacities across China from Jan. 1, 2015 to Jun. 15, 2020 for model training, and we used data from Jun. 15 to Dec. 31, 2020 for model testing. The results show that the developed model performs very well for megacities; against observations, the model obtains an average cross-validated R^2 (coefficient of determination) ranging from 0.90 for Shanghai to 0.97 for Chengdu in the one-step predictions, thereby indicating that the model outperformed any single algorithm or other hybrid algorithms reported. The developed model can also capture high ozone pollution episodes with an average accuracy of 92% for the next five days in inland cities. This study will be useful for the environmental health community to prevent high ozone exposure more efficiently in megacities in China and shows great potential for accurate ozone prediction using machine learning approaches.

1. Introduction

High concentrations of ozone (O_3) can not only cause respiratory and cardiovascular diseases but also restrict plant growth, reduce agricultural production, and adversely affect ecosystem health (U.S.EPA, 2013; Crouse et al., 2015). Since the Action Plan on the Prevention and Control of Air Pollution was implemented by the Chinese government in 2013, the government has taken a variety of powerful measures to reduce pollution, resulting in a significant drop in the emissions and environmental concentrations of particulate matter (PM), nitrogen oxides (NOx) and sulfur dioxide (SO_2) (Zheng et al., 2018). However, the ambient ozone concentration in China is still increasing. From 2013 to 2017, the outdoor ozone concentration (8 h average) increased from 139.2 μg/m³ (69.6 ppb) to 163.0 μg/m³ (81.5 ppb) in 74 key cities in China, exceeding the Grade II criteria (160 μg/m³) set by the Chinese Ambient Air Quality Standards in 2017 (Huang et al., 2018). Due to the differences in the geographical environment and emission conditions between regions, the reasons for and changing patterns of ozone growth in various regions of China also differ. For example, in Beijing, the main

reason for the increase in O_3 is the reduction of NOx and PM, while in Shanghai, the reduction of NOx and the increase in volatile organic compounds (VOCs) emissions are the main reasons (Liu and Wang, 2020). Therefore, the characteristics and changing trends of ozone pollution are vary by area, and developing an ozone prediction model that can be well applied to areas with different ozone pollution characteristics and can achieve high-quality prediction is essential for preventing air pollution and protecting public and ecological health.

Ozone prediction models can generally be divided into two categories: numerical-driven models and data-driven models. The main numerical-driven models include the community multiscale air quality (CMAQ) model (Foley et al., 2010) and WRF-Chem model (Chuang et al., 2011; Hong et al., 2020). The characteristic of numerical-driven models is that prediction accuracy depends mainly on the degree of agreement between the actual atmospheric conditions and the model. Because there is a clear causal relationship between pollution sources and air pollution, such models are very interpretable. However, the main difficulties in establishing numerical-driven models for forecasting are as follows: (1) It is difficult to collect air pollution emission source

---

data, and the data on actual pollution conditions have a relatively obvious information lag. (2) The collected pollution emission source data are highly uncertain. (3) Due to the complexity of data calculation, the requirements for equipment, software and technical personnel are relatively high, and it is usually difficult for the corresponding resources to meet the requirements.

The main data-driven models include linear, machine learning (ML) and hybrid models. Because data-driven models have the characteristics of fast calculation, high accuracy and easy implementation, these models have irreplaceable advantages in the realm of pollutant prediction. Early related forecasting research used mainly linear models, such as the autoregressive (AR) model and autoregressive integrated moving average (ARIMA) model (Slini et al., 2002; Duenas et al., 2005). With the advancement of science and technology and the rapid development of hardware, ML-related theories can gradually be realized and can effectively solve the problems encountered in reality (Xiao et al., 2018); therefore, ML models (which include mainly artificial neural networks (ANNs) (Kolehmainen et al., 2001; Tsai et al., 2009; Antanasijevic et al., 2019; Lagesse et al., 2020), random forest (RF) model (Zhan et al., 2018; Chen et al., 2021) and the support vector regression (SVR) (Lu and Wang, 2005; Su et al., 2020)) are becoming more and more popular. Currently, many scholars use machine learning models to perform spatial interpolation prediction of historical remote sensing air pollution data (Liu et al., 2020b; Ma et al., 2021; Zhang et al., 2021), the main target of these scholars is to improve or reconstruct the resolution of historical remote sensing air pollution datasets and provide higher-quality pollution data products. These studies are of great significance to environmental and epidemiological research. Additionally, many scholars use machine learning models to predict the changes in air pollutants in advance (Su et al., 2020; Huang et al., 2021); such use is of great significance for protecting human health and pollution control before the arrival of pollution episodes, which is also the key direction of this study. In recent years, there have been many studies on forecasting pollutant change trends. Sayeed et al. (2020) used a deep convolutional neural network (CNN) to predict ozone, thereby greatly improving the trend of daily ozone changes and prediction accuracy. Feng et al. (2019) used various models, such as extreme learning machine (ELM), multi-layer perceptron (MLP), random forest (RF) and recurrent neural network (RNN) to analyze and predict ground ozone in Hangzhou, China. The results show that the RNN model performs the best. Wu and Lin (2019) used a long short-term memory (LSTM) neural network to predict the air quality of two cities in China and found that compared with other algorithms, the LSTM deep learning model performs well in time series prediction. Huang et al. (2021) used deep learning models, including an RNN, an LSTM neural network and a gated recurrent unit (GRU) neural network, to predict the $PM_{2.5}$ concentration in Beijing, and the results showed that the GRU neural network had the highest prediction accuracy. In addition, as a variant of the LSTM model, the GRU neural network can process time series faster through its gating device and can overcome the vanishing and exploding gradient problems. The GRU neural network has more potential than the LSTM model. Currently, the GRU neural network has good predictive performance in many fields (Gao et al., 2019; Zhang and Liu, 2020), but there are few studies on the GRU neural network in ozone prediction research.

Although much progress has been made in research on ML models, for pollutant data with different time series characteristics, models with better accuracy and applicability are needed. The hybrid model combines the advantages of multiple algorithms, such as data decomposition and benchmark prediction algorithms, to achieve better prediction and generalization performance (Liu et al., 2020a). As a key part of the hybrid model, the data decomposition algorithm can effectively improve the prediction performance of the model by decomposing the original pollutant time series into multiple more stable and more regular sublayers. The decomposition algorithms commonly use include wavelet decomposition (WD) (Cheng et al., 2019; AlOmar et al., 2020), empirical mode decomposition (EMD) (Zhu et al., 2017; Huang et al., 2021),

ensemble empirical mode decomposition (EEMD) (Bai et al., 2019; Wang et al., 2019), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) (Mo et al., 2020).

Currently, many scholars are committed to predicting long-term pollutant trends. However, due to the large differences in the discharge characteristics and formation mechanisms of different target pollutants, it is often difficult to obtain accurate data on precursor contaminants. In general, the long-term prediction results of a model are quite different from the actual results. Therefore, we believe that before achieving effective long-term ozone pollution prediction, short-term ozone prediction studies should be performed to adequately prepare and explore long-term ozone prediction. Based on this point of view, our current main research focuses on short-term ozone prediction and proposes a novel hybrid ozone prediction model based on the WD algorithm, GRU neural network and SVR model; the proposed model can achieve high-precision short-term prediction of urban ground ozone and has excellent generalization performance. In the proposed model, the WD algorithm was used to decompose the original ozone data into multiple sublayers. Then, the GRU neural network and SVR model were combined to predict the data of each layer, and finally, the prediction results of each sublayer were summarized to obtain the final prediction results.

2. Materials and methods

2.1. Study area

This paper selected six representative cities in China as our research area (Fig. 1). Beijing is the capital of China and has a temperate monsoon climate. Xi'an is an important city in western China with a warm temperate semihumid continental monsoon climate. Shanghai is one of the four municipalities under the direct administration of the central government of China and has a subtropical monsoon climate. Wuhan is the central city in central China and has a subtropical monsoon climate. Chengdu is an important central city in western China and has a monsoon-influenced humid subtropical climate. Shenzhen is one of China's megacities and has a subtropical maritime climate. Among the six cities, Beijing, Xi'an, Wuhan and Chengdu are inland cities, and Shanghai and Shenzhen are coastal cities.

The ground ozone data (national control stations) in the six urban areas were collected from the website (http://106.37.208.233:20035), which publishes hourly data of different airborne pollutants, including ground ozone data, from the China National Environmental Monitoring Center, CNEMC. The collected station datasets belong to a time series, and the time scale ranges from the hourly average concentration of ozone from Jan. 1, 2015 to Dec. 31, 2020. After preprocessing, the daily moving average maximum 8 h value of each city was generated. More details can be found in Section S1.1.

2.2. The WD-GRU-SVR model

In the WD-GRU-SVR model, the GRU neural network and the SVR model were mainly used to train the sequence obtained by using the WD algorithm as a decomposition algorithm. Then, based on the training results, the optimal model was used to make the next prediction. The essence of ozone concentration prediction is to use actual past data values to estimate the value at the next moment. Since the GRU neural network and the SVR model have rapid and accurate characteristics for small sample data, these two models were selected as the reference models of the hybrid model in ozone concentration prediction. The algorithms involved in this study were all implemented on the Anaconda 5.2 platform, the Python version was 3.6, and the deep learning model was based on the Keras deep learning framework. The experimental environment is a 64-bit Windows 10 operating system, the CPU is an Intel Core i7-6700HQ, and the GPU is an NVIDIA GTX 960 M.

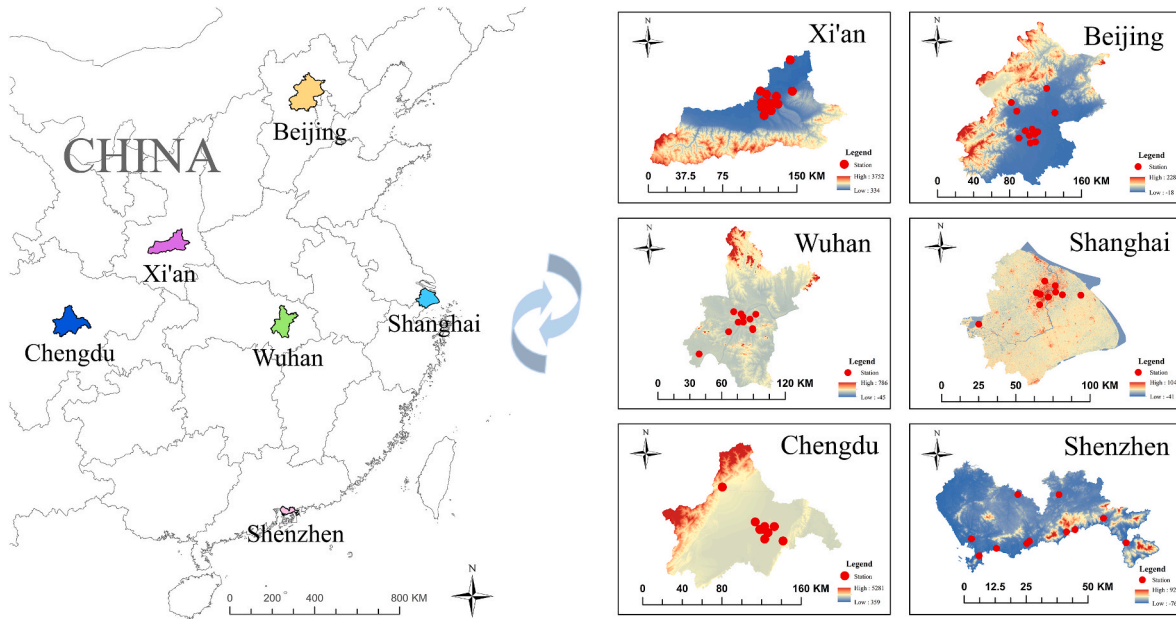The entire research process can be divided into three parts. The

Fig. 1. Research area in China.

preprocessing of relevant data samples was conducted in the first part. The key part of the novel hybrid model, mainly for data decomposition and model prediction, was implemented in the second part. In the third part, multiple models were built to compare the prediction performance of the proposed model. The entire main flowchart of the research is shown in Fig. 2. The specific process is as follows:

1. In the data preprocessing stage, linear interpolation was used to fill in the missing values in the data sample, and an algorithm was written to calculate the maximum 8-h value of the daily moving average of each target city as the original data input of the model. More details can be found in Section S1.1.
2. Then, the WD algorithm was used to decompose the 8 h O_3 concentration data into multiple sublayers. More details of the WD algorithm can be found in Section S1.2.
3. Because the decomposition sequence of each layer has different characteristics, the training data of each layer were trained with the GRU neural network and SVR model, and the parameters of the model were selected by the particle swarm optimization (PSO) algorithm and the grid search (GS) method, respectively. Then, according to the corresponding evaluation index, the optimal model was selected as the final model of the decomposition data of the current layer to predict the test set. Finally, the prediction result data of each layer are merged to obtain the final prediction result. The details of the GRU neural network and SVR model are presented in section 2.3.
4. In this study, 20 models were used to compare the prediction performance of the proposed model. The comparison models include the SVR model, RNN model, LSTM neural network, GRU neural network, WD-SVR model, WD-RNN model, WD-LSTM model, WD-GRU model, EMD-SVR model, EMD-RNN model, EMD-LSTM model, EMD-GRU model, EEMD-SVR model, EEMD-RNN model, EEMD-LSTM model, EEMD-GRU model, CEEMDAN-SVR model, CEEMDAN-RNN model, CEEMDAN-LSTM model and CEEMDAN-GRU model. Detailed information on the model evaluation indexes is provided in Section S1.3.

2.3. Machine learning methodology

2.3.1. GRU

Before introducing the GRU neural network, first introduce two models that are very closely related to the GRU neural network: the RNN and the LSTM neural network. The RNN is an advanced version of the ANN and consists of a series of modules with the same basic structure. Fig. 3 shows the basic structure of a traditional RNN, where x_t is the input vector at time step t, and $\widehat{y}_t$ is the final output of the RNN cell at time step t. In the RNN structure, h_t refers to the hidden state, which is calculated based on the hidden state of the previous time step t-1 and the input vector of the current time step t. In theory, no matter how long the sequence is, the RNN can use all the input information. However, due to the vanishing and exploding gradient problems, when the length of the processed time series becomes longer, the accuracy of the RNN model will be significantly reduced.

Hochreiter and Schmidhuber (1997) proposed the LSTM neural network in 1997, which can overcome the vanishing and exploding gradient problems that often occur in the RNN model and is known for its excellent memory function. The chain shape of the LSTM neural network is similar to that of the RNN, but the internal structure of the LSTM neural network is more complicated, which is also the key to enabling the LSTM neural network to learn more long-term effective information. The LSTM neural network is an effective model that automatically learns features from sequence data. The output of the LSTM neural network can be a variable length sequence that can be used for ozone prediction. For example, for time series $x_t(t = 1, 2, 3, 4, 5 \ldots)$, the output of the LSTM unit before each step x_t is regarded as the input of the current LSTM unit. In addition, the current LSTM unit will output a result. In the LSTM neural network, the memory unit generates a state vector for the current time. The structure of the LSTM neural network is shown in Fig. 3. The following equations introduce the specific process of the LSTM workflow:

$$i_i = \sigma\left(W^{(i)}H + b_i\right) \tag{1}$$

$$f_i = \sigma\left(W^{(f)}H_{(f)} + b_f\right) \tag{2}$$

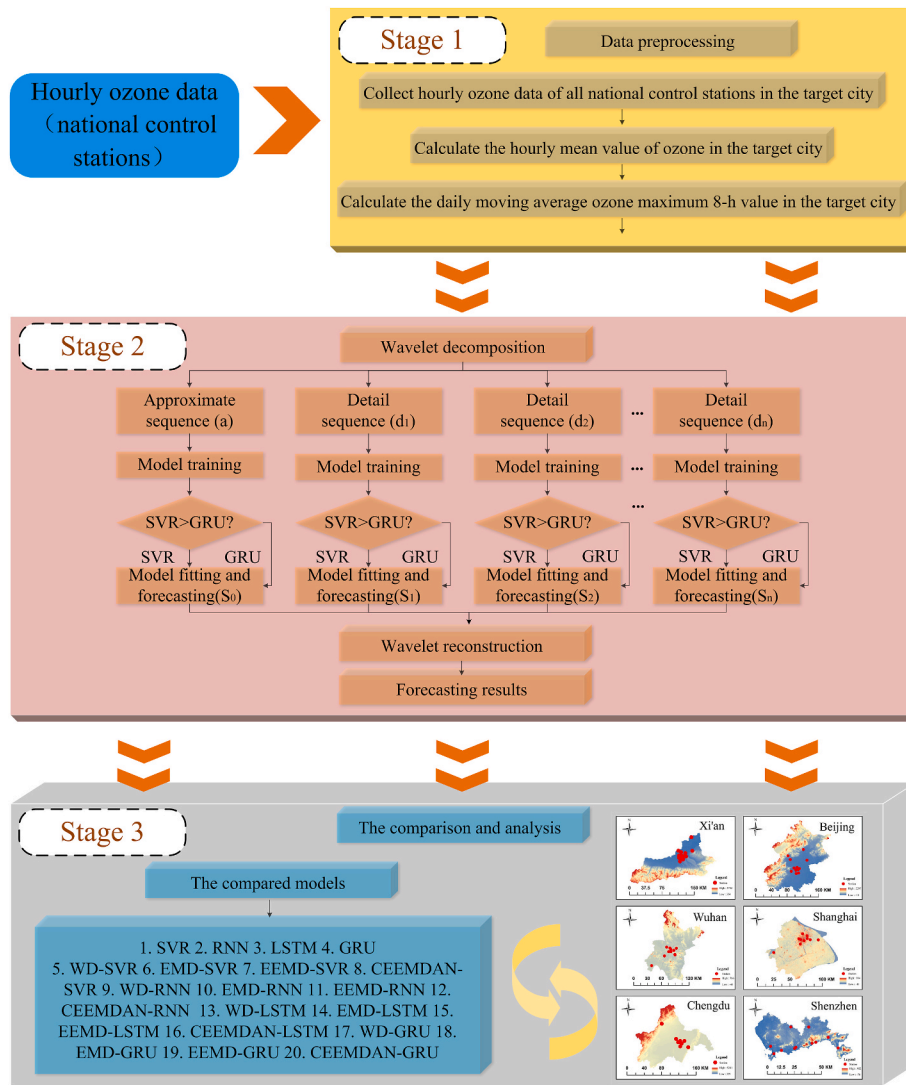$$o_i = \sigma\left(W^{(o)}H_{(o)} + b_o\right) \tag{3}$$

Fig. 2. Main flowchart of the hybrid WD-GRU-SVR model.

$$\widetilde{C}_i = \tanh\left(W^{(C)}H_C + b_C\right) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_i \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

The LSTM neural network has a relatively good time series prediction performance, but due to its complex structure, the training process of developing the LSTM neural network usually takes a long time. In order to speed up the training process, the GRU neural network was proposed. At present, this model has become one of the most promising algorithms in the RNN family. The GRU can be considered a variant of LSTM, and its main task is to handling the vanishing gradient problem that often occurs in an RNN. In most cases, both the GRU neural network and LSTM neural network can provide excellent results. The GRU structure, which is composed of a control gate and temporary output, is shown in Fig. 3. The hidden state h_t and the cell state are merged into one in the GRU neural network.

In the GRU unit, the control gate includes an update gate and a reset gate. The function of the update gate is mainly to control the status information of the previous time step *t-1* to be brought into the current time step *t*. The larger the update gate value is, the more state information there is in the previous time step. The function of the reset gate is to control the amount of information passed into the current candidate

set from the previous state. The smaller the reset gate is, the less information including the previous state there is. Because of the existence of these two gating devices in the GRU neural network, the exploding and vanishing gradient problems are effectively solved. The following equation introduces the specific process of the GRU workflow:

$$z_t = \sigma(wz*[h(t-1)], x_t) \tag{7}$$

where x_t and $h(t-1)$ are multiplied by their weights and added together. Then, sigmoid activation is used to convert the result to a value between 0 and 1. The reset gate r_t is calculated using the following equation at time step *t*:

$$r_t = \sigma(wr*[h(t-1)], x_t) \tag{8}$$

The reset gate in the unit helps the model determine how much past information needs to be forgotten. The current memory content involves resetting the gate. In addition, starting with the introduction of new memory content, the content will use the reset gate and store past related information. The mathematical equation is as follows:

$$h_t = tanh(w*[r_t*h(t-1)], x_t) \tag{9}$$

The reset gate r_t is multiplied by the previous output $h(t-1)$. This only allows information about the past to be passed. Then, the results of the two calculations are added together and the *tanh* function is applied.
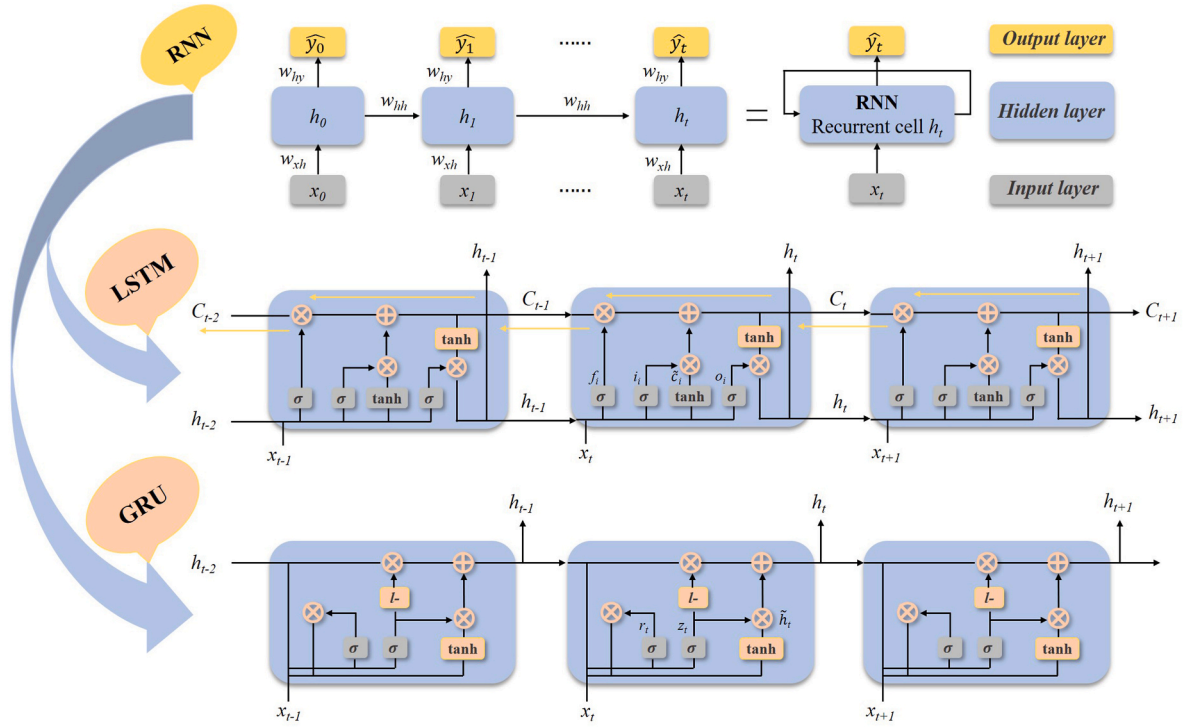
Fig. 3. The architecture of the RNN, LSTM and GRU neural networks.

Finally, the unit must calculate vector h_t of the information stored in the current unit and pass the information further to the network. The update gate z_t plays a key role in it. The mathematical equation is as follows:

$$h_t = (1 - z_t)*h(t - 1) + (z_t*h_t) \tag{10}$$

The overall analysis shows that if the vector z_t is close to 0, a large part of the current content will be ignored because it has nothing to do with prediction. Furthermore, since z_t will be close to 0 at this time, $1 - z_t$ will be close to 1 so that most of the past information can be retained.

In all experiments in this study, the PSO algorithm was used to determine three hyperparameters for the GRU neural network, including the time step, batch number, and number of neurons in the hidden layer. The setting of the hyperparameters in the model will directly affect the training and prediction accuracy of the network. When the input data are positive, the rectified linear unit (ReLU) function does not have the vanishing gradient problem; therefore, this study used the activation function to select the ReLU, and the loss function is defined as the mean absolute error (MAE) to avoid overfitting. In addition, the Adam optimizer optimization function was used, the learning rate was set to 0.001, and the number of iterations was 280. The hyperparameter value range to be optimized in our experiments was set as follows: the time step size was set to [1, 10], the batch size was set to [1, 20], and the number of hidden layer neurons was set to [2, 50]. The number of particles in the group was configured to be 40, the inertia factor of the velocity was 0.6, and the acceleration factor was $C_1 = C_2 = 1$. Next, through adaptive PSO iteration and optimization of the hyperparameters of the time step, the batch size and the number of hidden layer neurons in the GRU neural network are determined. Finally, an ozone prediction model was established based on the best hyperparameters obtained through optimization.

2.3.2. SVR

The SVR model is a general machine learning method developed based on the framework of statistical learning theory (Vapnik, 1999). Compared with the neural network method, the SVR model solves the problems of small samples, high dimensionality, local minima, etc., and exhibits a strong generalization performance; therefore, the SVR model is widely used in regression estimation, density estimation, system identification, process modeling and other fields. Compared with traditional machine learning methods, this model has stronger robustness and is particularly suitable for time series forecasting. In the SVR, the learning machine will use nonlinear mapping to map the training data to a high-dimensional space. Then, in the high-dimensional feature space, the hyperplane can form a nonlinear relationship between the training data and the output data, which is called the SVR function and can also be expressed as a convex optimization problem. The specific process is as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \left(\alpha_i^* - \alpha_i \right) \left(\alpha_j^* - \alpha_j \right) K(w_i, v_i) + \varepsilon \sum_{i=1}^{l} \left(\alpha_i^* + \alpha_i \right) - \sum_{i=1}^{l} y_i \left(\alpha_i^* - \alpha_i \right) \tag{11}$$

$$\text{subject to} \quad \sum_{i=1}^{l} \left(\alpha_i^* - \alpha_i \right) = 0, \ 0 \le \alpha_i, \ \alpha_i^* \le \frac{C}{l}, \ i = 1, 2, ..., l, \tag{12}$$

where $K(w_i, v_i)$ is the kernel function and α_i and α_i^* are Lagrange multipliers. The resulting regression function $f(x)$ is as follows:

$$f(x) = \sum_{i=1}^{l} \left(\alpha_i^* - \alpha_i \right) K(x, w_i) + b \tag{13}$$

The above formula shows that the key to SVR mainly lies in the choice of the kernel function and the hyperparameter setting. In this study, because the model sample is not large, to improve the prediction accuracy of the model as much as possible, the GS was used to configure the model parameters, which include the time advance step size, three kernel functions (linear kernel, polynomial kernel, and radial basis function kernel), penalty factor C and parameter g. The GS is a method to perform a detailed search on the specified parameter range value. From the selection of all candidate parameters, the GS uses a loop traversal method to test various possibilities and obtain the optimal parameters as the final result. For example, if there are 4 possibilities each for parameters a and b, 16 results will be produced. The advantages

of this method are that it is simple and easy to understand and implement, it achieves high efficiency and stable results, and it has few parameter adjustment requirements. Finally, an ozone prediction model was established based on the optimal kernel function and hyperparameters obtained by the GS method.

2.4. Prediction and validation strategies

After data preprocessing, the 8 h O_3 series of each city composed 2192 samples, as shown in Fig. S1; the samples ranged from Jan. 1, 2015, to Dec. 31, 2020. The 1–1992 samples (Jan. 1, 2015, to Jun. 14, 2020) of each city's 8 h O_3 series were used as the training dataset, and the 1993–2192 samples (Jun. 15, 2020, to Dec. 31, 2020) were used as the test dataset. In this study, we chose to use the direct prediction strategy for model experiments and conducted experiments and discussions on the multistep predictions of the model. The specific prediction strategy structure is shown in Fig. 4, where the size of forward time step in this study was set from 1 to 10, and the specific value will depend on the model training result.

In the one-step predictions strategy, assuming that x_0 to x_t are the model's training dataset (t days in total), the model's prediction result for the t+1st day is P_{t+1}. By adding the actual data x_{t+1} into the model to continue prediction, the predicted result is P_{t+2} (which is the model for the t+2nd day, and so on) by constantly adding actual data into the model. Thus, the prediction results for the next day are generated. Finally, we integrate all prediction results and compare them with the actual values. In the two-step prediction strategy, assuming that x_0 to x_t are the training dataset of the model (t days in total), the prediction result P_{t+2} on the t+2nd day is directly predicted. In contrast with the one-step predictions strategy, the change is that the direct forecast for the first day is changed to the direct forecast for the second day after each time. Using the same forecasting strategy, this study carried out at most five-step forecasting.

3. Result and discussion

3.1. The model performance

In the experiments involving the six urban areas, first, we used a one-step predictions strategy and 20 ozone prediction models as comparison models, which were produced mainly by a mixture of four benchmark models (namely, the SVR, RNN, LSTM and GRU neural network) and four typical decomposition algorithms (namely, the WD, EMD, EEMD, and CEEMDAN algorithms). Then, on the basis of the result of one-step predictions, we used a multistep (2- to 5-step) predictions strategy respectively and selected four benchmark models and four hybrid models with better prediction performance as comparison models. Among these models, the neural network architecture involving the deep learning model is consistent with the WD-GRU-SVR model. In order to ensure the accuracy of the experiments, all experiments used the same training dataset and test dataset, and they were all completed in the same experimental environment. In addition, in order to reduce the influence of the randomness of some models on the prediction results, the models involved in this study, including the WD-GRU-SVR model, were randomly repeated 6 times in each study area. Therefore, the model accuracy values shown in Tables S2–S7 and Tables S14–S19 are the averages of the results obtained from 6 experiments.

In the case of one-step predictions strategy, as shown in Tables S2–S7, compared with other 20 prediction models, the WD-GRU-SVR model achieved the most satisfactory performance. We mapped the prediction results of the six urban areas to fully demonstrate the verification results (Fig. 5). The ozone prediction results in different cities all show that the WD-GRU-SVR model has accurate and stable prediction performance, can well adapt to the ozone characteristics of the environment in different regions, and can make efficient and accurate predictions. In the six cities of Beijing, Shanghai, Shenzhen, Wuhan, Chengdu and Xi'an, the MAEs predicted by the WD-GRU-SVR model are approximately 7.86, 7.97, 5.50, 7.54, 6.55 and 8.70, respectively; the RMSEs are approximately 10.72, 12.09, 7.77, 11.01, 9.04 and 12.56, respectively; the MAPEs are approximately 11.58%, 9.60%, 7.61%, 11.46%, 10.56% and 14.40%, respectively; and the R^2s are approximately 0.96, 0.90, 0.96, 0.94, 0.97 and 0.94, respectively.

Fig. 5 shows that the prediction performance of the WD-GRU-SVR model will vary from region to region. The Chengdu area had the highest R^2 of 0.97; the Shanghai area had the lowest R^2 of 0.90 among the 6 regions studied, but the performance was still very good. The average R^2 of the six areas is 0.945, which is an important result indicating that the model we proposed is very suitable for ozone prediction in China. More importantly, the WD-GRU-SVR model also has outstanding predictive performance for sudden changes in the ozone
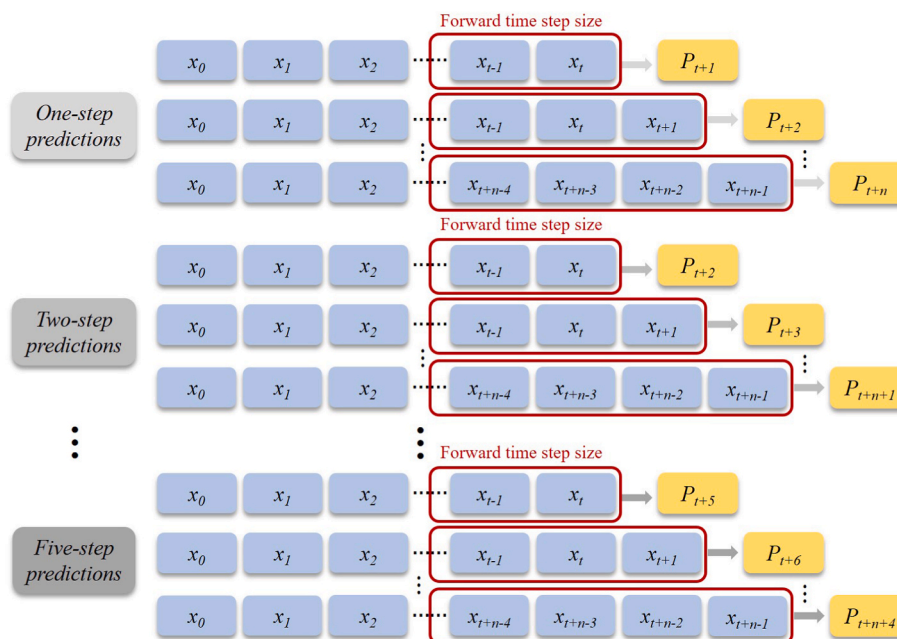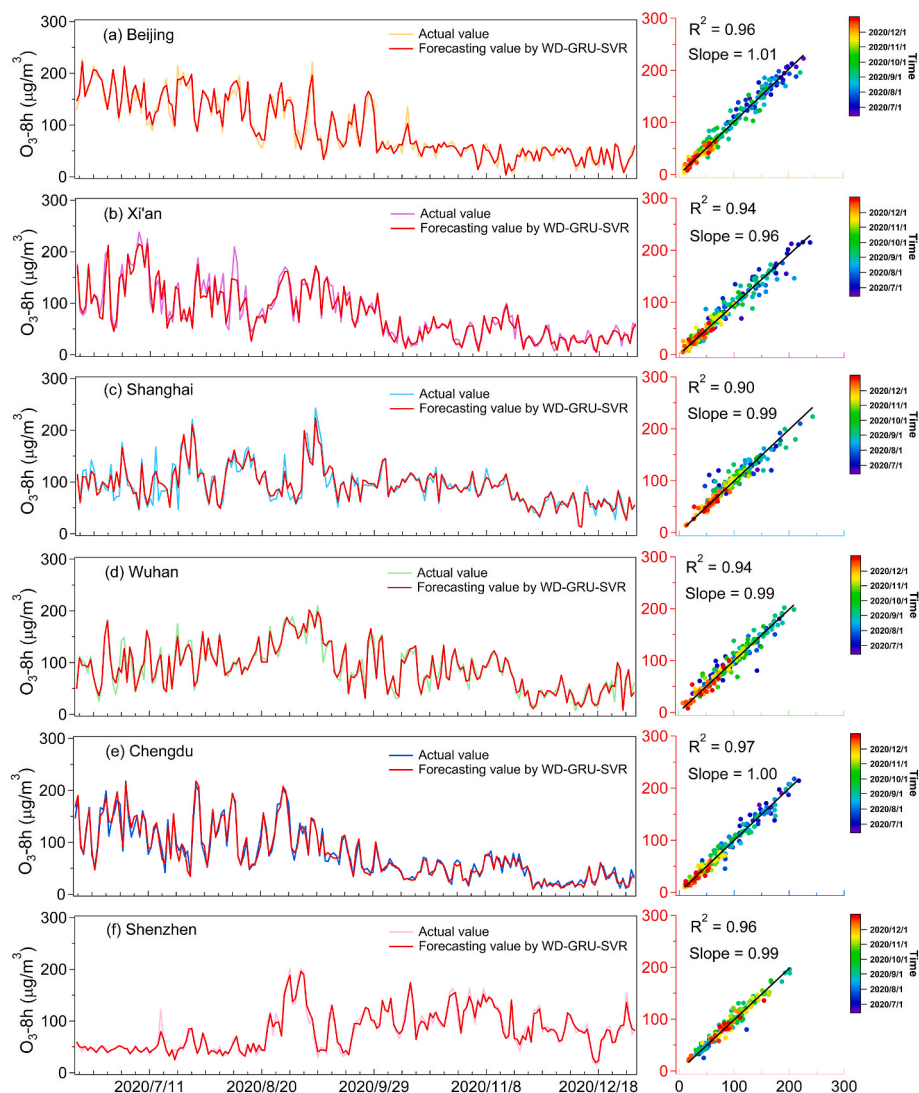

Fig. 4. The structure of multistep predictions strategy.

Fig. 5. Time series plots of the forecasted and actual values for the WD-GRU-SVR model by one-step predictions in Beijing, Xi'an, Shanghai, Wuhan, Chengdu and Shenzhen.

concentration. For example, in Beijing, the actual ozone concentration on Aug. 24, 2020 was 111.49 $\mu g/m^3$, and the concentration on Aug. 25 rose to 161.83 $\mu g/m^3$, where the proposed model predicted 163.12 $\mu g/m^3$ on August 25. In Shanghai, the actual ozone concentration on Jun. 29, 2020 was 72.65 $\mu g/m^3$, and the concentration on Jun. 30 rose to 176.28 $\mu g/m^3$, where the proposed model predicted 166.64 $\mu g/m^3$ on Jun. 30. In Shenzhen, the actual ozone concentration on Nov. 16, 2020 was 133.24 $\mu g/m^3$; the concentration on Nov. 17 fell to 66.22 $\mu g/m^3$, where the proposed model predicted 66.42 $\mu g/m^3$ on Nov. 17.

In the case of the multistep predictions strategy, as shown in Tables S14–S19, we can find that as the number of prediction steps increases, the performance of the ozone prediction model established in each city decreases. For example, for the one-, two-, three-, four- and five-step predictions in Beijing, the MAEs predicted by the WD-GRU-SVR model are approximately 7.86, 13.15, 14.62, 17.30 and 18.45, respectively; the RMSEs are approximately 10.72, 17.80, 19.65, 22.40 and 23.70, respectively; the MAPEs are approximately 11.58%, 18.25%, 22.03%, 26.76% and 27.92%, respectively; and the R^2s are approximately 0.96, 0.90, 0.87, 0.84 and 0.81, respectively. Due to the different characteristics of ozone pollution in each city, the prediction performance decline of the model also varies. However, in general, the WD-GRU-SVR model still always achieved the most satisfactory performance in each step predictions strategy.

Overall, the proposed hybrid model has satisfactory predictive performance. Tables S2-S7 and S14-S19 also show that the model we proposed performs best in terms of various evaluation indexes.

3.2. The prediction for high pollution episodes

In this study, we also counted the number of days exceeding the Grade II criteria (160 $\mu g/m^3$) set by the Chinese Ambient Air Quality Standards of 8-h maximum ozone concentration in each study area during the model prediction period (Jun. 15, 2020 to Dec. 31, 2020). In addition, we calculated the mean absolute percentage error for each city. If the Grade II criteria were reached within the uncertainty range of the predicted value, it was counted as an ozone exceeding day. Table 1 shows that the region where the model achieves the best prediction results in the one-step predictions of the number of days that ozone exceeds the standard is Beijing, where the model achieves a prediction accuracy rate of 100%; in the Xi'an and Shanghai regions, the model achieves the worst prediction accuracy rate at 85%. Taking all the regions together, the overall forecast accuracy rate of the one-step predictions is 92%. Concerning the two-step and three-step predictions, the model achieves the best prediction accuracy rate at 100% in the Wuhan and Chengdu regions and the worst prediction accuracy rate at 62% in the Shanghai region. The overall forecast accuracy rates of the two-step

Table 1
The number of days with exceeding 160 μg/m^3 of the 8 h maximum ozone concentration from Jun. 15, 2020 to Dec. 31, 2020 in six cities in China.

Step	Statistics	Beijing	Xi'an	Shanghai	Wuhan	Chengdu	Shenzhen	Total
	Actual the number of ozone exceedances (day)	30	20	13	15	19	9	106
One	Predict the number of ozone exceedances (day)	30	17	11	14	17	8	97
	The accuracy of prediction (%)	100%	85%	85%	93%	89%	89%	92%
Two	Predict the number of ozone exceedances (day)	28	19	8	15	19	6	95
	The accuracy of prediction (%)	93%	95%	62%	100%	100%	67%	90%
Three	Predict the number of ozone exceedances (day)	27	19	8	15	19	6	94
	The accuracy of prediction (%)	90%	95%	62%	100%	100%	67%	89%
Four	Predict the number of ozone exceedances (day)	27	19	8	15	18	4	91
	The accuracy of prediction (%)	90%	95%	62%	100%	95%	44%	86%
Five	Predict the number of ozone exceedances (day)	27	18	8	15	17	4	89
	The accuracy of prediction (%)	90%	90%	62%	100%	89%	44%	84%

and three-step predictions are 90% and 89%, respectively. Concerning the four-step and five-step predictions, the model achieves the best prediction accuracy rate at 100% in the Wuhan region and the worst prediction accuracy rate at 44% in the Shenzhen region. The overall forecast accuracy rates of the four-step and five-step predictions are 86% and 84%, respectively.

A detailed analysis of the results shows that the application of the established model in each city varies. Among these results, the prediction results in Beijing, Xi'an, Wuhan, and Chengdu are better, and the prediction accuracy of the model can reach more than 90% in the five-step predictions. The model performed the best in Wuhan by achieving 100% accuracy in the two-to five-step predictions; the model performed better in the one-step predictions in Shanghai and Shenzhen, but in the two-to five-step predictions, the prediction results were unsatisfactory. We believe that the direct cause of the difference in the prediction of ozone pollution in various regions by the model is that the ozone data in the two regions have more irregular fluctuations in a short period, thus increasing the difficulty of model training. A common feature of the four cities of Beijing, Xi'an, Wuhan and Chengdu is that they are all inland cities, while Shanghai and Shenzhen are both coastal cities. In China, the surface ozone concentration depends largely on emissions and meteorology (Lu et al., 2019). Anthropogenic and natural emissions from local and external transmissions can provide precursors for the high level of ozone pollution (Ni et al., 2018). Meteorology can adjust the chemical and physical processes of ozone formation, thereby affecting the changes in ozone concentration, and the ozone formation mechanism in coastal areas is often more complicated than that in inland regions (Wang et al., 2017). For example, in coastal regions, particulate chloride (Cl$^-$) in sea-salt can contribute to the formation of ozone (Dai et al., 2020). Additionally, the daily changes in ozone are very sensitive to the weather and climate system, and downdrafts in the peripheral circulation of a typhoon system can strongly enhance surface ozone before the typhoon lands in eastern or southern China (Shu et al., 2016; Han et al., 2020). Therefore, considering the differences in the geographical locations of various urban areas, we believe that this model is suitable for all cities for predicting ozone pollution days one day in advance, and can achieve good prediction results. However, when it is necessary to predict ozone pollution days 2–5 days in advance, we believe that this model is more suitable for inland cities and can achieve high prediction accuracy. For coastal cities, we do not recommend using this model for multistep predictions in advance because it is difficult for us to guarantee the accuracy of the model.

3.3. The comparisons and analysis

Table S8 and Fig. S6 show that the prediction accuracy of the WD-GRU-SVR model is significantly higher than those of the SVR, RNN, LSTM and GRU neural networks. For example, in (a) Beijing, compared with the SVR, RNN, LSTM and GRU neural networks, the WD-GRU-SVR model promoted the MAE by 70.75%, 70.73%, 70.76% and 70.60%, respectively; the WD-GRU-SVR model promoted the RMSE by 66.51%,

66.30%, 66.33% and 66.29%, respectively; the WD-GRU-SVR model promoted the MAPE by 71.49%, 74.38%, 74.24% and 73.72%, respectively; and the WD-GRU-SVR model promoted the R^2 by 107.43%, 104.35%, 104.71% and 104.18%, respectively. Similar results can be observed for the different evaluation indexes and multistep predictions (see Section S1.4, Tables S9–S13, Figs. S7–S9 and Tables S14–S19 in the Supplementary).

The results of the model comparison experiments in the one-step predictions show that the improvement of the benchmark model using different decomposition algorithms will vary due to the different research areas. We conducted statistical analysis on the overall results and used the R^2 as the evaluation index of the model accuracy. The average R^2 of single benchmark models is 0.523, that of hybrid models combined with the WD algorithm is 0.889, that of hybrid models combined with the EMD algorithm is 0.658, that of hybrid models combined with the EEMD algorithm is 0.662, and that of hybrid models combined with the CEEMDAN algorithm is 0.630. The above results, in general, show that the decomposition algorithms can improve the prediction performance of the benchmark model. Among the models, the performance of the benchmark model combined with the WD algorithm is the most improved, and the CEEMDAN algorithm is the least improved.

Among all the ozone prediction models involved in this study, the WD-GRU-SVR model has the best prediction performance. The reasons for such a significant promotion in forecasting can be explained as follows: First, compared with the EMD, EEMD and CEEMDAN decomposition algorithms, the WD is the decomposition of the specified number of layers. To a certain extent, the decomposition adaptability can be maintained at a higher level. However, the EMD-type decomposition algorithm is an adaptive decomposition algorithm, which easily causes excessive decomposition. Second, in the proposed prediction model, the GRU neural network and SVR model were used to predict the decomposition sequence. The GRU neural network often tends to have better prediction performance for low-frequency sequences, while SVR models tend to have better prediction performance for high-frequency sequences. Moreover, in this study, the proposed algorithm trains and compares two models (GRU and SVR) for each decomposition sequence, which effectively combines the strengths of the two models. The GRU neural network is an improved type of RNN that has all the advantages of RNN neural networks and overcomes the vanishing and exploding gradient problems. The GRU has strong nonlinear processing capabilities and is suitable for ozone prediction calculations. Finally, in the proposed hybrid forecasting model, the SVR model is the most typical of machine learning. The SVR model is very suitable for processing small sample data and has the advantages of fast training speed and high accuracy. Therefore, it has good performance in memory optimization and nonlinear data processing. Based on the successful combination of the above algorithms, the proposed WD-GRU-SVR model has satisfactory short-term prediction results.

Additionally, we found that the model has different applicability in each city. We believe that in addition to the stability difference of the algorithm of the benchmark model, the WD algorithm plays a key role.

Due to the different fluctuation characteristics of 8 h O_3 data in various cities, the decomposition results also vary. Comparing the characteristics of the cities in the WD results shows that the approximate sequence curves of Beijing and Xi'an cities are relatively similar in amplitude and smoothness, thus indicating that the data decomposition results are relatively good. Concerning Chengdu and Wuhan, the approximate sequence shows 6 major cycles overall, but the degree of smoothness is slightly worse than that of Beijing and Xi'an. In Shanghai, the approximate sequence presents 6 periods of gradually decreasing fluctuation amplitude, the smoothness of the curve gradually worsens, and the sudden change value increases. In Shenzhen, the approximate sequence has no obvious change period, the noise characteristics are obvious, the overall change amplitude tends to increase, and the smoothness of the curve is relatively poor. Since Shanghai and Shenzhen are both coastal cities, their boundary conditions of ozone transport may vary greatly (such as alternation of sea and land breeze), thus making the prediction by the model inferior to predictions for inland cities. The above analysis shows that the curve and fluctuation characteristics of the WD result can illuminate the model prediction result. When the approximate sequence shows regular fluctuations, the one-step and multistep prediction performances of the model are often good, but if the approximate sequence changes irregularly and the noise characteristics are more obvious, the multistep predictions results will be greatly discounted, and it is difficult to guarantee the model prediction accuracy.

4. Conclusions

In this study, we proposed a novel hybrid short-term ozone prediction model based on the WD algorithm, GRU neural network and SVR model. In the proposed hybrid WD-GRU-SVR model, the key step was to use the WD algorithm to decompose the original ozone data into multiple layers and then combine the GRU neural network and the SVR model to predict the decomposed data of each layer. With an average R^2 of 0.945 in the one-step predictions for six megacities, the hybrid model outperformed other prediction models overall. The results show that the WD algorithm can stably and effectively improve the prediction accuracy of the machine learning algorithm. In summary, this study can be useful for the environmental health community in preventing and controlling ozone pollution more effectively and accurately, especially in urban-scale research. Furthermore, this study predicts the daily moving average maximum 8 h value of ozone, and most of the 8 h data will be collected before evening. Therefore, high ozone pollution episodes of the next day can be predicted at night to remind sensitive people to use good protection. For the effective prediction of ozone pollution episodes many days in advance, we recommend using this model in inland cities, where, with an average accuracy rate of 92%, the model can still maintain high accuracy even five days in advance. However, for coastal cities with complex underlying surfaces and meteorology, we do not recommend using this model to make predictions many days in advance. Generally, this study is a valuable case study in applying machine learning in the environmental field, and the results show that deep learning has great potential in ozone pollution prediction.

Credit author statement

Yong Cheng: Conceptualization, Methodology, Software, and Writing. **Ling-Yan He:** Supervision and Reviewing. **Xiao-Feng Huang:** Supervision, Methodology, and Writing.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2021.113670.

References

AlOmar, M.K., Hameed, M.M., AlSaadi, M.A., 2020. Multi hours ahead prediction of surface ozone gas concentration: robust artificial intelligence approach. Atmos. Pollut. Res. 11 (9), 1572–1587. https://doi.org/10.1016/j.apr.2020.06.024.

Antanasijevic, D., Pocajt, V., Peric-Grujic, A., Ristic, M., 2019. Urban population exposure to tropospheric ozone: a multi-country forecasting of SOMO35 using artificial neural networks. Environ. Pollut. 244, 288–294. https://doi.org/10.1016/j.envpol.2018.10.051.

Bai, Y., Zeng, B., Li, C., Zhang, J., 2019. An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting. Chemosphere 222, 286–294. https://doi.org/10.1016/j.chemosphere.2019.01.121.

Chen, G., Chen, J., Dong, G.-h., Yang, B.-y., Liu, Y., Lu, T., Yu, P., Guo, Y., Li, S., 2021. Improving satellite-based estimation of surface ozone across China during 2008–2019 using iterative random forest model and high-resolution grid meteorological data. Sust. Cities Soc. 69 https://doi.org/10.1016/j.scs.2021.102807.

Cheng, Y., Zhang, H., Liu, Z., Chen, L., Wang, P., 2019. Hybrid algorithm for short-term forecasting of PM2.5 in China. Atmos. Environ. 200, 264–279. https://doi.org/10.1016/j.atmosenv.2018.12.025.

Chuang, M.-T., Zhang, Y., Kang, D., 2011. Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States. Atmos. Environ. 45 (34), 6241–6250. https://doi.org/10.1016/j.atmosenv.2011.06.071.

Crouse, D.L., Peters, P.A., Hystad, P., Brook, J.R., van Donkelaar, A., Martin, R.V., Villeneuve, P.J., Jerrett, M., Goldberg, M.S., Pope, r.C.A., Brauer, M., Brook, R.D., Robichaud, A., Menard, R., Burnett, R.T., 2015. Ambient PM2.5, O₃, and NO₂ exposures and associations with mortality over 16 Years of follow-up in the Canadian census health and environment cohort (CanCHEC). Environ. Health Perspect. 123 (11), 1180–1186. https://doi.org/10.1289/ehp.1409276.

Dai, J., Liu, Y., Wang, P., Fu, X., Xia, M., Wang, T., 2020. The impact of sea-salt chloride on ozone through heterogeneous reaction with N2O5 in a coastal region of south China. Atmos. Environ. 236 https://doi.org/10.1016/j.atmosenv.2020.117604.

Duenas, C., Fernandez, M.C., Canete, S., Carretero, J., Liger, E., 2005. Stochastic model to forecast ground-level ozone concentration at urban and rural areas. Chemosphere 61 (10), 1379–1389. https://doi.org/10.1016/j.chemosphere.2005.04.079.

Feng, R., Zheng, H.J., Zhang, A.R., Huang, C., Gao, H., Ma, Y.C., 2019. Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison:A case study in hangzhou, China. Environ. Pollut. 252 (Pt A), 366–378. https://doi.org/10.1016/j.envpol.2019.05.101.

Foley, K.M., Roselle, S.J., Appel, K.W., Bhave, P.V., Pleim, J.E., Otte, T.L., Mathur, R., Sarwar, G., Young, J.O., Gilliam, R.C., Nolte, C.G., Kelly, J.T., Gilliland, A.B., Bash, J. O., 2010. Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7. Geosci. Model Dev. 3 (1), 205–226. https://doi.org/10.5194/gmd-3-205-2010.

Gao, X., Li, X., Zhao, B., Ji, W., Jing, X., He, Y., 2019. Short-term electricity load forecasting model based on EMD-GRU with feature selection. Energies 12 (6). https://doi.org/10.3390/en12061140.

Han, H., Liu, J., Shu, L., Wang, T., Yuan, H., 2020. Local and synoptic meteorological influences on daily variability in summertime surface ozone in eastern China. Atmos. Chem. Phys. 20 (1), 203–222. https://doi.org/10.5194/acp-20-203-2020.

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.

Hong, J., Mao, F., Min, Q., Pan, Z., Wang, W., Zhang, T., Gong, W., 2020. Improved PM2.5 predictions of WRF-Chem via the integration of Himawari-8 satellite data and ground observations. Environ. Pollut. 263 (Pt A), 114451. https://doi.org/10.1016/j.envpol.2020.114451.

Huang, G., Li, X., Zhang, B., Ren, J., 2021. PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768, 144516. https://doi.org/10.1016/j.scitotenv.2020.144516.

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.

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting 35 (5), 815–825. https://doi.org/10.1016/S1352-2310(00)00385-X.

Lagesse, B., Wang, S., Larson, T.V., Kim, A.A., 2020. Predicting PM2.5 in well-mixed indoor air for a large office building using regression and artificial neural network models. Environ. Sci. Technol. 54 (23), 15320–15328. https://doi.org/10.1021/acs.est.0c02549.

Liu, H., Yin, S., Chen, C., Duan, Z., 2020a. 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.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., Bi, J., 2020b. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach. Environ. Int. 142, 105823. https://doi.org/10.1016/j.envint.2020.105823.

Liu, Y., Wang, T., 2020. Worsening urban ozone pollution in China from 2013 to 2017-Part 2: the effects of emission changes and implications for multi-pollutant control. Atmos. Chem. Phys. 20 (11), 6323–6337. https://doi.org/10.5194/acp-20-6323-2020.

Lu, W.Z., Wang, W.J., 2005. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. Chemosphere 59 (5), 693–701. https://doi.org/10.1016/j.chemosphere.2004.10.032.

Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T.-M., Zhang, Q., 2019. Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences. Atmos. Chem. Phys. 19 (12), 8339–8361. https://doi.org/10.5194/acp-19-8339-2019.

Ma, R., Ban, J., Wang, Q., Zhang, Y., Yang, Y., He, M.Z., Li, S., Shi, W., Li, T., 2021. Random forest model based fine scale spatiotemporal O3 trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017. Environ. Pollut. 276, 116635. https://doi.org/10.1016/j.envpol.2021.116635.

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. Atmos. Environ. 220 https://doi.org/10.1016/j.atmosenv.2019.117072.

Ni, R., Lin, J., Yan, Y., Lin, W., 2018. Foreign and domestic contributions to springtime ozone over China. Atmos. Chem. Phys. 18 (15), 11447–11469. https://doi.org/10.5194/acp-18-11447-2018.

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.

Shu, L., Xie, M., Wang, T., Gao, D., Chen, P., Han, Y., Li, S., Zhuang, B., Li, M., 2016. Integrated studies of a regional ozone pollution synthetically affected by subtropical high and typhoon system in the Yangtze River Delta region, China. Atmos. Chem. Phys. 16 (24), 15801–15819. https://doi.org/10.5194/acp-16-15801-2016.

Slini, Th, Karatzas, K., Moussiopoulos, N., 2002. Statistical analysis of environmental data as the basis of forecasting: an air quality application. Sci. Total Environ. 288 (3), 227–237. https://doi.org/10.1016/S0048-9697(01)00991-3.

Su, X., An, J., Zhang, Y., Zhu, P., Zhu, B., 2020. Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine

using wavelet transformation and partial least squares methods. Atmos. Pollut. Res. 11 (6), 51–60. https://doi.org/10.1016/j.apr.2020.02.024.

Tsai, C.H., Chang, L.C., Chiang, H.C., 2009. Forecasting of ozone episode days by cost-sensitive neural network methods. Sci. Total Environ. 407 (6), 2124–2135. https://doi.org/10.1016/j.scitotenv.2008.12.007.

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. https://www.epa.gov/isa/integrated-science-assessment-isa-ozone-and-related-photochemical-oxidants.

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.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. Sci. Total Environ. 575, 1582–1596. https://doi.org/10.1016/j.scitotenv.2016.10.081.

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.

Xiao, Q., Chang, H.H., Geng, G., Liu, Y., 2018. An ensemble machine-learning model to predict historical PM2.5 concentrations in China from satellite data. Environ. Sci. Technol. 52 (22), 13260–13269. https://doi.org/10.1021/acs.est.8b02917.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M.L., Zhang, M., Di, B., 2018. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. Environ. Pollut. 233, 464–473. https://doi.org/10.1016/j.envpol.2017.10.029.

Zhang, G., Liu, D., 2020. Causal convolutional gated recurrent unit network with multiple decomposition methods for short-term wind speed forecasting. Energy Convers. Manag. 226 https://doi.org/10.1016/j.enconman.2020.113500.

Zhang, X., Zhang, Y., Lu, X., Bai, L., Chen, L., Tao, J., Wang, Z., Zhu, L., 2021. Estimation of lower-stratosphere-to-troposphere ozone profile using long short-term memory (LSTM). Rem. Sens. 13 (7) https://doi.org/10.3390/rs13071374.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., Zhang, Q., 2018. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. Atmos. Chem. Phys. 18 (19), 14095–14111. https://doi.org/10.5194/acp-18-14095-2018.

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.