Front. Environ. Sci. Eng. 2023, 17(2): 21
https://doi.org/10.1007/s11783-023-1621-4

RESEARCH ARTICLE

Prediction and cause investigation of ozone based on a double-stage attention mechanism recurrent neural network

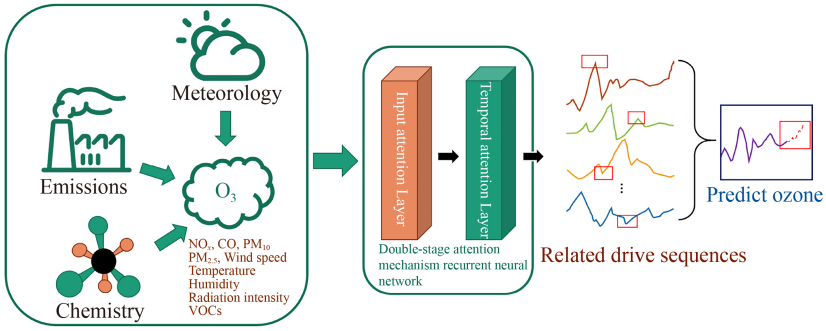
Yuanxin Zhang[1], Fei Li (✉)[1], Chaoqiong Ni[2], Song Gao (✉)[1], Shuwei Zhang[1], Jin Xue[1], Zhukai Ning[1], Chuanming Wei[1], Fang Fang[2], Yongyou Nie[3], Zheng Jiao (✉)[1]

1 School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China
2 Shanghai Jinshan Environmental Monitoring Station, Shanghai 201500, China
3 School of Economics, Shanghai University, Shanghai 200237, China

HIGHLIGHTS

- Used a double-stage attention mechanism model to predict ozone.
- The model can autonomously select the appropriate time series for forecasting.
- The model outperforms other machine learning models and WRF-CMAQ.
- We used the model to analyze the driving factors of VOCs that cause ozone pollution.

GRAPHIC ABSTRACT



ARTICLE INFO

Article history:
Received 3 May 2022
Revised 5 July 2022
Accepted 27 July 2022
Available online 10 September 2022

Keywords:
Ozone prediction
Deep learning
Time series
Attention
Volatile organic compounds

ABSTRACT

Ozone is becoming a significant air pollutant in some regions, and VOCs are essential for ozone prediction as necessary ozone precursors. In this study, we proposed a recurrent neural network based on a double-stage attention mechanism model to predict ozone, selected an appropriate time series for prediction through the input attention and temporal attention mechanisms, and analyzed the cause of ozone generation according to the contribution of feature parameters. The experimental data show that our model had an RMSE of 7.71 μg/m^3 and a mean absolute error of 5.97 μg/m^3 for 1-h predictions. The DA-RNN model predicted ozone closer to observations than the other models. Based on the importance of the characteristics, we found that the ozone pollution in the Jinshan Industrial Zone mainly comes from the emissions of petrochemical enterprises, and the good generalization performance of the model is proved through testing multiple stations. Our experimental results demonstrate the validity and promising application of the DA-RNN model in predicting atmospheric pollutants and investigating their causes.

1 Introduction

The emission of atmospheric pollutants is gradually increasing with the acceleration of industrialization and urbanization, and society is becoming increasingly concerned with environmental issues (Hui et al., 2018). Air pollution has evolved into a more complex type of pollution with a high ozone content. Ozone, as an air quality index (AQI) contaminant, has become one of the most serious environmental pollution issues in many cities and regions (Ou et al., 2015).

As it is more common in East China, especially during

✉ Corresponding authors
E-mails: lifeisparkle@shu.edu.cn (F. Li); njulegao@163.com (S. Gao); zjiao@shu.edu.cn (Z. Jiao)

Special Issue—Artificial Intelligence/Machine Learning on Environmental Science & Engineering (Responsible Editors: Yongsheng Chen, Xiaonan Wang, Joe F. Bozeman III & Shouliang Yi)

summer, ozone has become the primary air pollutant in urban areas (Wang et al., 2018). High levels of ozone have negative effects on human health, climate change, and the ecological balance (Kampa and Castanas, 2008). Volatile organic compounds (VOCs) and NO_x are major precursors. They catalyze the conversion of oxygen to ozone, which is affected by several atmospheric variables such as temperature, humidity, radiation intensity, particulate matter concentration, and other unclear factors (Avery, 2006). Each VOC can play a different role during ozone formation. The complexity of VOCs inevitably makes this situation extremely complicated and difficult to predict. The establishment of a reliable ozone forecasting model, based on the non-linear relationship between the affecting components and conditions, will be critically important to predict its formation and set up strategies for its elimination (Shao et al., 2009).

The establishment of ozone prediction systems is mainly based on two types of conventional methods: chemistry-transport models (CTMs) and statistical methods (Donnelly et al., 2015). CTMs commonly use the method of solving mathematical equations to numerically calculate the chemical and physical processes of pollutant emission, diffusion, and transformation, and obtain the fitting results of the relationship between the atmosphere and emissions (Vautard et al., 2007). Generally, there are three CTMs: the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006), the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) (Zhang et al., 2010), and the Nested Air Quality Prediction Modeling System (NAQPMS) (Ge et al., 2014). CTMs have the merits of accuracy of the emission inventory and meteorological data and completeness of the simulated chemical mechanism (Zhang et al., 2011). It often takes a long time to perform preliminary preparations and calculation simulations to obtain accurate results (Liu et al., 2021). Frequently used statistical methods include the autoregressive moving average (ARMA) model (Erdem and Shi, 2011), autoregressive integrated moving average (ARIMA) model (Kavasseri and Seetharaman, 2009), multiple linear regression (MLR) (del Carmen Bas et al., 2017), and geographically weighted regression (GWR) (Wang et al., 2014). Statistical methods are limited to capturing the complex nonlinear relationship between air quality data in a time series regardless of high-dimensional data (Delle Monache et al., 2006). Meanwhile, both types of methods struggle to solve the problem of dimension disaster owing to the lack of a flexible multi-scale framework, resulting in poor prediction performance (Garaud and Mallet, 2011). Recently, machine learning techniques have become the standard for researchers to solve problems. For example, Chen et al. used a combined representation learning algorithm for the super-resolution reconstruction of face images (Chen et al., 2021). Xia et al. (2022) used an improved algorithm to reduce occlusion interference during target tracking and improve the object recognition accuracy. Zhang et al. (2022a) used a multi-feature fusion algorithm to achieve a better success rate and accuracy than other trackers. Therefore, some researchers have begun using machine learning models to solve the problem of forecasting air pollutants. For example, Hu et al. (2017) estimated ground-level $PM_{2.5}$ concentrations in the United States using a random forest model. Ding et al. (2021) used the CatBoost model to reconstruct satellite aerosol optical depth (AOD) data. An object tracking framework with recapture based on correlation filters and Siamese networks developed by Zhang et al. (2022b) has application prospects in the visual detection of environmental pollutants. Wong et al. (2021) used an XGBoost-based boosting algorithm to predict NO_2 concentrations.

Deep learning technology has been developed owing to its ability to extract features from large amounts of data using neural networks; this process has the potential to address the limitations of conventional air quality prediction methods (Molina-Gomez et al., 2021). Ozone concentration prediction can be genuinely recognized as a multidimensional time-series prediction problem from the perspective of deep learning (Jia et al., 2021). According to the complexity of the multi-dimensional data series and the spatiotemporal characteristics of deep learning (Xu et al., 2021), the main structures suitable for time series forecasting include the recurrent neural network (RNN) (Qin et al., 2017), long short-term memory (LSTM) network (Wang et al., 2021), and gated recurrent unit (GRU) network (Tao et al., 2019). Based on these structures, corresponding models can extract spatiotemporal characteristics and apply them to air quality prediction (Reichstein et al., 2019). However, atmospheric pollution is a complex and long-time scale problem. These structures hardly capture long-time scale correlations and identify potential nonlinear relationships to improve prediction accuracy (Bengio et al., 1994).

In this study, we used a neural network model based on a double-stage attention mechanism recurrent neural network (DA-RNN) to predict ozone concentration. This model appropriately captures the long-term temporal dependence based on the previous values of the multi-dimensional time series and multiple driving sequences, and selects the relevant driving sequence for prediction. The importance and relevance of all characteristics to ozone were analyzed, and multiple evaluation standards were selected to assess the accuracy of the model prediction. The performance of the optimized DA-RNN model was compared with that of the CMAQ for the final test and verification. The main contributions of this study are as follows: 1) The DA-RNN model was used to accurately predict the ozone concentration in the Jinshan Industrial Zone. 2) Based on the importance of these features, we found that petrochemical enterprises were

the main causes of ozone pollution in the Jinshan Industrial Zone. 3) The model has a good generalization performance and good application prospects in the field of ozone prediction.

2 Study station and dataset

2.1 Research station and monitoring data

The selected station in this study was Jinshan station, which is a representative atmospheric monitoring site in the Shanghai Chemical Industry District. Shanghai, a chemical industry district, is located near the southern border of Shanghai, north of Hangzhou Bay, and close to the East China Sea. With a planned area of 29.4 km^2 and a management area of 36.1 km^2, it ranges within the Jinshan and Feng Xian districts. Jinshan is located near a chemical engineering park and has high research value. Therefore, except for ozone, significant precursors, including hourly monitoring data for NO_x, CO, PM_{10}, $PM_{2.5}$, and VOCs, were identified. Weather conditions, including wind speed, temperature, humidity, and radiation intensity, which could promote ozone formation, were also recorded. To study the model's generalization performance, we also studied the Jinshan industrial district station and Jinshan boundary station, using the same data categories as for the Jinshan station.

All data were obtained from the Shanghai Jinshan Environmental Monitoring Center. Fig. 1 shows scatter plots of hourly ozone concentration values versus NO_x, CO, PM_{10}, $PM_{2.5}$, wind speed, temperature, humidity, and radiation intensity; there was a significant non-linear relationship between the variables and ozone concentration. The data were used as training and test sets for the DA-RNN model.

2.2 Data preprocessing

The dataset used in this study contained outliers removed from the original records. However, the maintenance of equipment at a station can result in missing values. Therefore, in order to make the dataset as realistic as possible, the missing data must be filled in. Furthermore, it is well known that the variation in atmospheric pollutant concentrations can be considered a continuous, non-discrete process. Therefore, we can choose linear interpolation to complete the missing data. The linear interpolation equation is as follows:

$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0), \tag{1}$$

where (x_0, y_0) and (x_1, y_1) are the coordinates of the two points before and after the value to be filled, and (x, y) are the coordinates of the points that need to be supplemented.

In total, 8760 records were shared in the final dataset. Seventy percent of the data were used as the training set and the remaining 30 % as the testing set to evaluate the prediction performance of the model.

3 Method

In this study, the model was divided into two parts based on the input attention and temporal attention mechanisms. In the first stage, a new input attention mechanism that can automatically select a suitable time series was introduced. In the second stage, we used a temporal attention mechanism to select the relevant encoder hidden states across all time steps. Through these two attention mechanisms, the DA-RNN can adaptively select the most relevant input features and capture the long-term temporal dependence of the time series. A flowchart of the DA-RNN model is shown in Fig. 2. The details of the double-stage attention mechanism recurrent neural network are described in Algorithm 1.

Algorithm 1: DA-RNN Algorithm

Input: A set of multivariate time series: $\{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^n\}$
time step: t ; the parameter of model: θ
Output: The predicted result: y_t

for *each traing iteration* **do**
  for *t steps* **do**
    Compute $\widetilde{x}_t$ by the previous hidden state h_{t-1} and
    Input attention Layer:
    $\widetilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \cdots, \alpha_t^n x_t^n)^\top$
    Then into the encoder LSTM unit.
    Compute c_t by the previous hidden state h_{t-1} and
    Temporal attention Layer:
    $c_t = \sum_{i=1}^{T} \beta_t^i h_i$
    The generated context vector c_t is then used as an
    input to the decoder LSTM unit.
    Compute y_t to the given target sequence
    $\{y_1, y_2, \ldots, y_T\}$ and c_t :
    $y_t = w\,[y_{t-1}; c_{t-1}] + b$
    The output y_t of the last decoder LSTM unit is the
    predicted result.
  end
end

3.1 Input attention mechanism

The input attention mechanism captures dependencies on a multidimensional time series. For the time series, $X = (x_1, x_2, \ldots, x_T)$ with $x_t \in X$. There is mapping from x_t to h_t (at time step t) with

$$h_t = f(h_{t-1}, x_t), \tag{2}$$

where h_t is the hidden state of the input attention layer at time t and f is the non-linear activation function of the LSTM. Each LSTM is controlled by the control unit C_t at time t. Access to the control unit is controlled by three S-shaped doors: the forget gate f_t, input gate i_t, and output

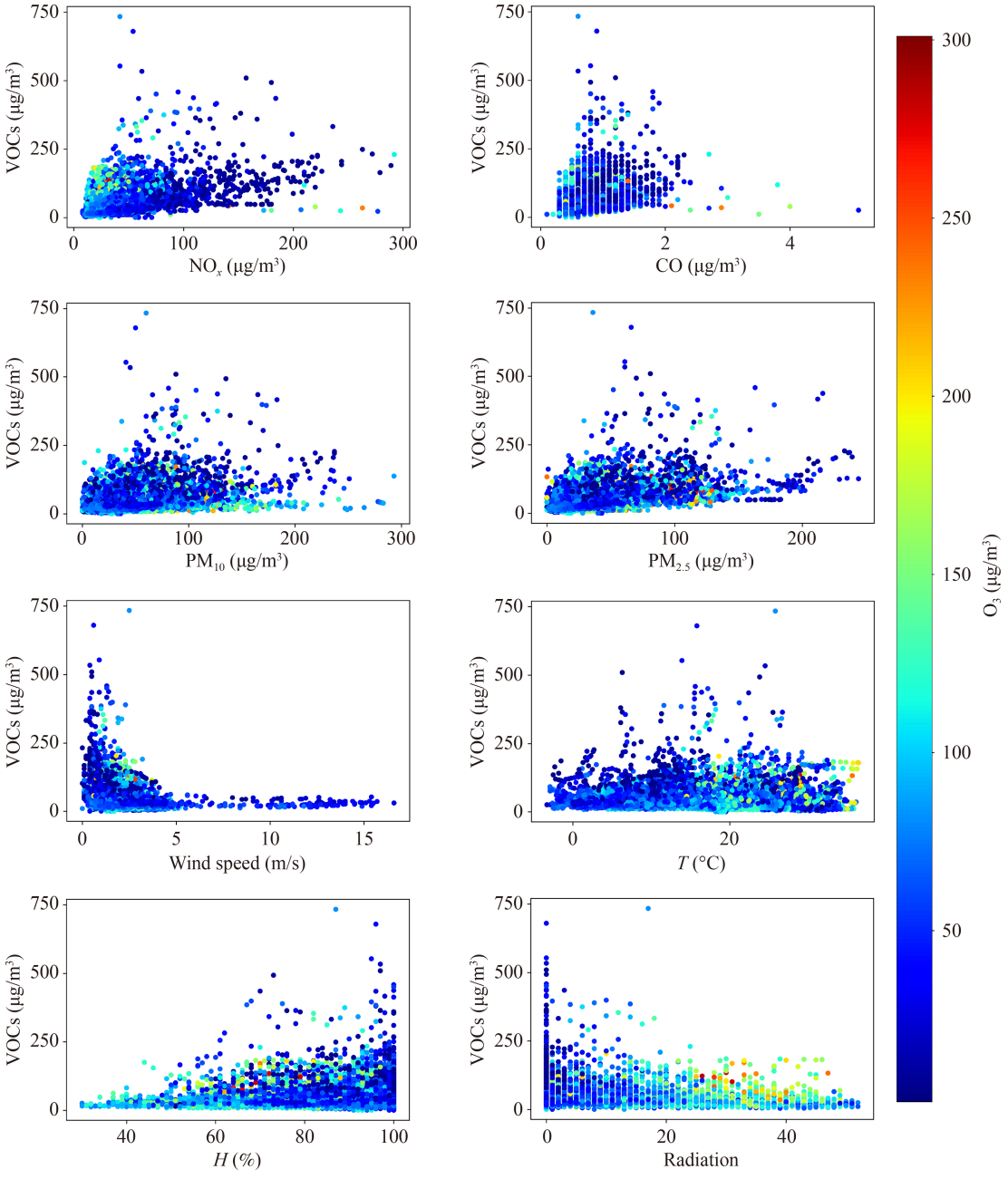


Fig. 1 Scatter plot of ozone concentration versus other variables.

gate o_t. The status update of LSTM can be represented by the following equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}; x_t] + b_f), \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}; x_t] + b_i), \quad (4)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}; x_t] + b_C), \quad (5)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}; x_t] + b_o), \quad (7)$$

$$h_t = o_t * \tanh(C_t), \quad (8)$$

where $[h_{t-1}; x_t]$ is the gate that reads the last hidden state h_{t-1} and the current input x_t. $W_f, W_i, W_C,$ and W_o, and $b_f, b_i, b_C,$ and b_o are the parameters that the model must learn. σ is the sigmoid function that is used as the activation function of the neural network to map variables between 0 and 1. There is one memory cell per LSTM cell with state C_t at time t. $\tilde{C}_t$ represents the cell state update value obtained from input data x_t and hidden state h_{t-1}. The reason for choosing LSTM is that it can overcome the problem of gradient disappearance caused by a long time, to better capture the correlation of time series over a long time span.

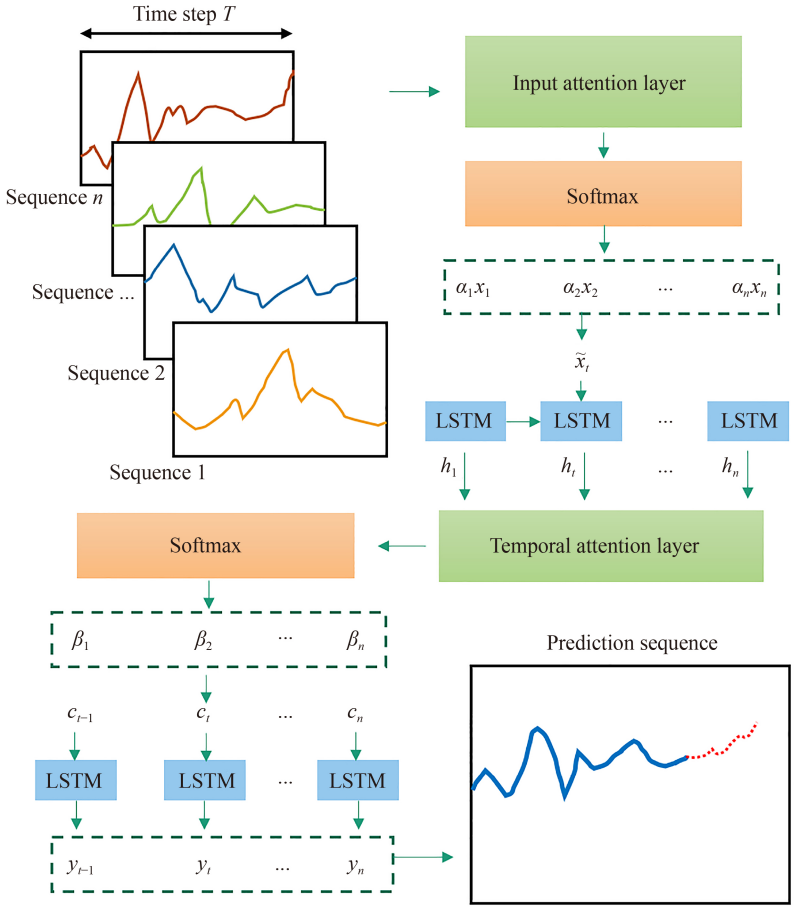


Fig. 2 Flow chart of the DA-RNN model.

In this structure, we use an algorithm based on input attention, which can adaptively select the relevant time series, given the k-th input time series $x^k = (x_1^k, x_2^k, \ldots, x_T^k)$, by reading the previous hidden state h_{t-1} and unit state C_{t-1} in the LSTM cell; thus, we can construct the input attention mechanism through the model:

$$e_t^k = v_e^T \tanh\left(W_e \cdot [h_{t-1}; C_{t-1}] + U_e x^k\right), \tag{9}$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{t=1}^{n} \exp(e_t^i)}, \tag{10}$$

where v_e, U_e, and W_e are the parameters that the model must learn. e_t^k is the attention weight of the k-th input sequence at time t. α_t^k is the attention weight, which measures the importance of the k-th input time series at time t. In Eq. (10), we use the softmax function to ensure that the sum of all the attention weights is 1. Using these attention weights, the driving sequence can be extracted adaptively.

According to this structure, we obtain a time series with attention by combining the attention weights:

$$\tilde{x}_t = (\alpha^1 x_t^1, \alpha^2 x_t^2, \ldots, \alpha^n x_t^n). \tag{11}$$

According to the obtained $\tilde{x}_t$, h_t is updated as follows:

$$h_t = f(h_{t-1}, \tilde{x}_t). \tag{12}$$

Therefore, in the input attention mechanism, the relevant time series can be selected according to the attention weight for the subsequent training.

3.2 Temporal attention mechanism

To make accurate predictions, we introduced an LSTM structure based on an RNN in the temporal attention mechanism. As the length of the time series increases, the performance of the model decreases rapidly. Therefore, we used a temporal attention mechanism to adaptively select the appropriate hidden state at the time step. By reading the previous hidden state h_{t-1} and unit state C_{t-1} in the LSTM cell, we constructed a temporal attention mechanism using the following model:

$$l_t^i = v_d^T \tanh(W_d[h_{t-1}; C_{t-1}] + U_d h_i), \tag{13}$$

$$\beta_t^i = \frac{\exp(l_t^i)}{\sum_{j=1}^{T} \exp\left(l_t^j\right)}, \tag{14}$$

where v_d, W_d, and U_d are the parameters that the model must learn. l_t^i is the temporal attention weight of the i-th input sequence at time t. β_t^i represents the importance of

the i-th encoder hidden state for prediction. In eq. (14), we use the softmax function to ensure that the sum of all the temporal attention weights is 1.

Then, we obtain the context vector c_t according to the time attention weight β_t^i and all hidden state sequences $(h_1, h_2, \ldots, h_T)$:

$$c_t = \sum_{i=1}^{T} \beta_t^i h_i. \tag{15}$$

Then, according to the given target sequence $(y_1, y_2, \ldots, y_T)$, combined with the context vector, the predicted sequence can be obtained as follows:

$$y_t = w[y_{(t-1)}; c_{(t-1)}] + b \tag{16}$$

where $[y_{t-1}; c_{t-1}]$ is a concatenation of the target sequence y_{t-1} and the context vector c_{t-1}. The parameters w and b were obtained by training the model. The calculated y_t can be used to update the hidden state h_t at time t as follows:

$$h_t = f(h_{t-1}, y_t). \tag{17}$$

The newly obtained hidden state h_t is used for the next prediction, and the prediction sequence is finally obtained. f is an LSTM cell that can be computed using Eqs. (3)–(8).

The input sequence of the model was the concentration of various pollutants and meteorological parameters per hour; the time step was set to 24, and the output sequence was the ozone concentration within the predicted time period.

3.3 Comparison with other machine learning models and WRF-CMAQ models

Machine learning models have been widely used to predict air pollutant concentrations. To prove the prediction performance of the model used in this study, we selected CatBoost, LightGBM, LightGBMLarge, XGBoost, Random Forest, ExtraTrees, NeuralNet, KNeighbors, and several common machine learning models for comparison.

Traditional chemistry-transport models are still widely used, so we also compared the WRF-CMAQ model, including the Weather Research and Forecasting Model (WRF; version 3.7), Sparse Matrix Operator Kernel Emission System (SMOKE; version 3.7), Model of Emissions of Gases and Aerosols from Nature (MEGAN; version 3), and Community Multi-scale Air Quality Modeling System (CMAQ; version 5.3.2).

Three levels of nesting were adopted for the simulation area of the model. The center latitude and longitude was 120.3 °E, 33.7 °N; we used the Lambert projection; the horizontal resolution of the three levels of nesting was 36 km, 12 km, and 4 km, respectively. The grid numbers were 185 × 148, 147 × 240, and 204 × 228. The model adopted the Yonsei University (YSU) scheme to parameterize the boundary-layer process (Hong et al., 2006). The parameterizing interaction of the land-atmosphere was generated using the Noah land surface scheme (Ek et al., 2003). The Purdue-Lin microphysical scheme was used to reproduce cloud and precipitation processes (Efstathiou et al., 2013). Meteorological input data were obtained from the meteorological reanalysis data (FNL) of the National Center for Environmental Prediction (NCEP) with a grid resolution of 1° × 1° and a time resolution of 6 h. The emission inventory was obtained from the 2017 National Air Pollution Source Emission Inventory of Tsinghua University (MEIC). On this basis, RRTM long-wave (Chin et al., 2002) and Goddard shortwave (Mlawer et al., 1997) radiation schemes were used to reflect the radiation situation.

3.4 Model evaluation method

To measure the effectiveness of the time-series forecasting method, three distinct evaluation indicators were used to assess the differences between the predicted and observed values. The three evaluation indicators were the mean absolute error (MAE), root mean square error (RMSE), and Pearson correlation coefficient (PMC).

The MAE is the sum of the absolute values of the difference between the target and predicted values. The calculation formula is as follows:

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |(y_i - \hat{y}_i)|, \tag{18}$$

where y_i is the value of the i-th forecast, $\hat{y}_i$ is the value of the i-th observation, and m is the total amount of data.

The RMSE was used to measure the error between the target and predicted values. The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2}, \tag{19}$$

where y_i is the value of the i-th forecast, $\hat{y}_i$ is the value of the i-th observation, and m is the total amount of data. The lower the MAE and RMSE, the higher the prediction accuracy of the model.

The Pearson correlation coefficient is generally used in the feature selection stage before training to determine the linear relationship between the features. This is calculated as follows:

$$PMC = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}, \tag{20}$$

where X, Y are two variables that require correlation coefficients, $\mathrm{Cov}(X, Y)$ is the covariance of X and Y; $D(X)$ is the variance of X, and $D(Y)$ is the variance of Y.

4 Results and discussion

To cover most factors affecting ozone generation, we se-

lected hourly monitoring data, including reacting substances (NO_x, CO, PM_{10}, $PM_{2.5}$, and typical VOCs), resulting ozone concentrations, and weather conditions (wind speed, temperature, humidity, and radiation intensity) from the Jinshan station in 2020. The Photochemical Assessment Monitoring Station (PAMS) program introduced 57 species of VOCs, which are prone to photochemical reactions for easy detection. Accordingly, 33 VOC species were selected: ethane, propane, n-butane, isobutane, n-pentane, isopentane, n-hexane, n-heptane, n-octane, 2,2,4-trimethyl pentane, 2,3,4-trimethyl pentane, ethylene, acetylene, propylene, cis-2-butene, trans-2-butene, 1,3-butadiene, isoprene, benzene, toluene, ethyl benzene, ortho-xylene, styrene, 1,2,3-trimethylbenzene, 1,2,4-trimethylbenzene, 1,3,5-trimethylbenzene, cumene, n-propylbenzene, 1,1-dichloroethane, 1,2-dichloroethane, trichloroethylene, perchloroethylene, chlorobenzene, and 1,2,4-trichlorobenzene.

4.1 Analysis of feature correlation

In essence, the machine learning model completes the prediction by mining the correlation between the features and the predicted values. If a certain set of features and predicted values are completely independent, then it is useless for modeling. Therefore, it was necessary to verify the correlation prior to training. In this study, characteristic correlation analysis enabled us to understand the correlation between different pollutants in Jinshan. The Pearson correlation heat map shown in Fig. 3 proves a good non-linear relationship between the selected species, which provides theoretical support for follow-up work.

4.2 Analysis of feature importance

In machine learning, it is essential to analyze the

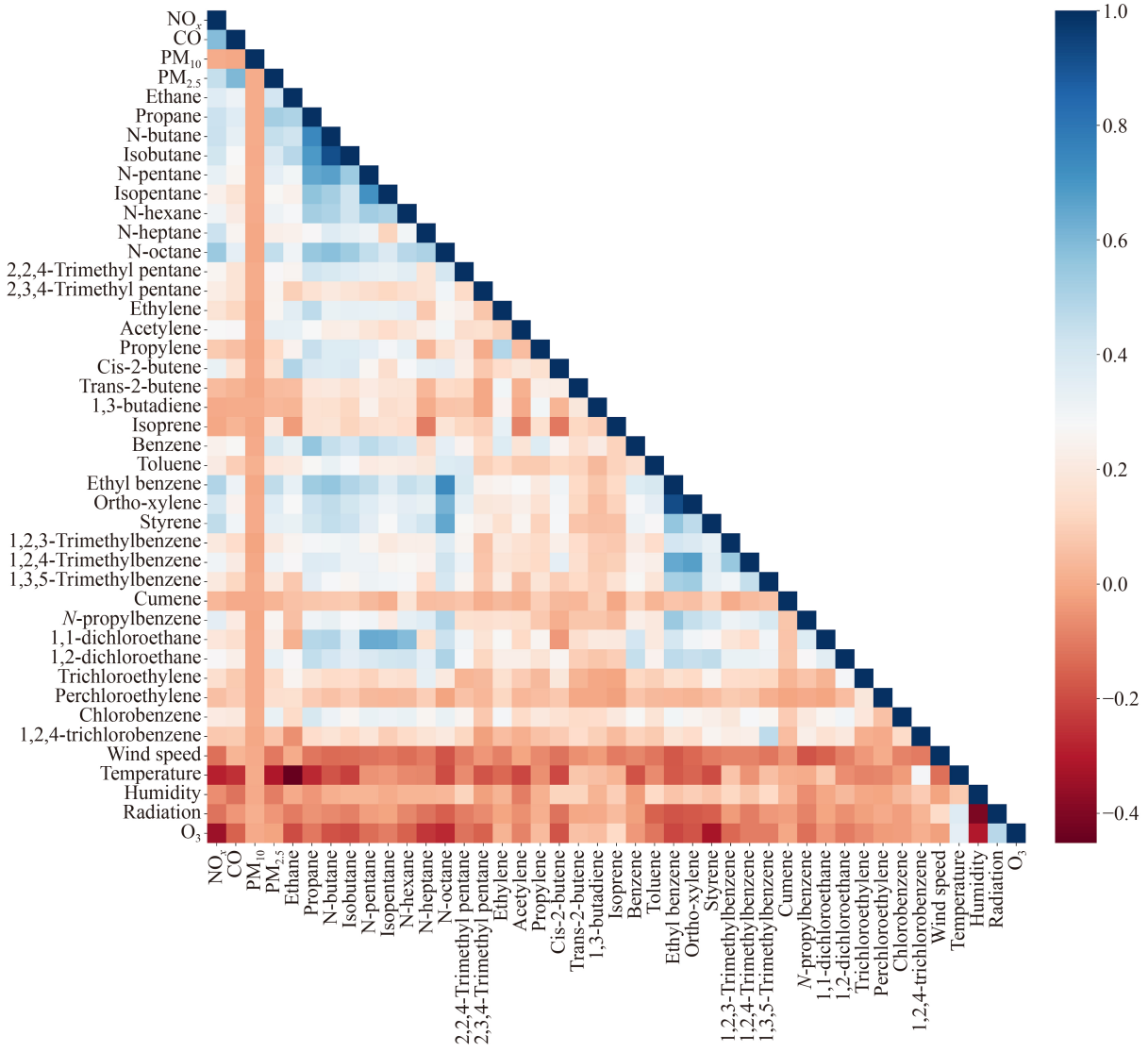


Fig. 3 Pearson correlation heat map.

importance of the features in the training model to improve the generalization ability of the model and the understanding of model features. Based on this model, the importance of features can help us understand the relative contributions of different species and conditions to ozone formation in Jinshan.

Fig. 4 shows a graph of the importance of weights of different features for the ozone concentration. For convenience, we divided VOCs into alkanes, olefins, aromatic compounds, and chlorinated compounds. As shown in Fig. 4(a), VOCs, as a large category of features, had the highest impact on ozone, indicating that changes in the concentration of VOCs had the greatest impact on ozone. Therefore, we can infer that the Jinshan station is in a VOC-controlled area. The Shanghai chemical industry zone is dominated by petrochemical and other chemical plants, and the main pollutants emitted by the petrochemical industry are alkanes and aromatic compounds, which accounted for more than 50 % of the importance of VOCs. Developed industries are often accompanied by massive energy consumption and transportation needs; therefore, olefins and chlorides also account for an essential portion of the critical weight of VOCs. At the same time, the impact of radiation intensity on ozone also has a higher weight because radiation can accelerate photochemical reactions and the production of ozone (Atkinson, 2000). As NO_x plays a similar catalytic role in the formation of ozone, the amount of NO_x directly affects the ozone yield. Fig. 4(b) shows that alkanes and aromatic hydrocarbons, as the two species with the highest percentages, have a combined importance of over 50 %, which indicates that ozone pollution in the Jinshan Industrial District is mainly influenced by VOCs from petrochemical enterprises and vehicle exhaust emissions. Fig. S1 shows the importance of each of the 33 VOC species.

4.3 Predictive performance of the model

We used an RTX3060 graphics card for model training. We set the batch size to 128, epochs to 300, and learning rate to 0.001. The software environment of the model was Python 3.8, Pytorch 1.8.1, and CUDA 11.1. The total training time was 900 s.

We used the model to predict the ozone concentration at 1, 2, 3, 6, 12, and 24 h on the test set. Both MAE and RMSE increased with an increase in forecast duration (Table 1). The MAE was between 5.97 μg/m^3 and 14.71 μg/m^3; the RMSE was between 7.71 μg/m^3 and 20.57 μg/m^3. Accordingly, R^2 gradually deviated from 0.958 to 0.699. In conclusion, the model indicated a high prediction accuracy in 1 h. Although the prediction performance gradually weakened, it was still possible to accurately predict the ozone concentration within a period of time.

Fig. 5 shows the scatter plots for different forecast durations that illustrate the correlation between the forecast and observed values. The dashed line represents the perfect case, where the predicted value is equal to the observed value, the color represents the density of points, and the model's prediction performance gradually decreases as the prediction duration increases. However, at 24 h, there was still a strong correlation between the predicted and observed values.

4.4 Performance of the model at low and high ozone values

To verify the performance of the model in predicting low and high values, we selected July and August as the typical months with high ozone values and November and December as the typical months with low ozone values. In Shanghai, ozone pollution incidents occur frequently, especially during summer when the peak ozone concentration exceeds 250 μg/m^3. In winter, the ozone concentration is low because of the weak sunlight and low temperature. Fig. 6 shows a comparison between the model's predicted hourly ozone concentration and the observed values in July, August, November, and

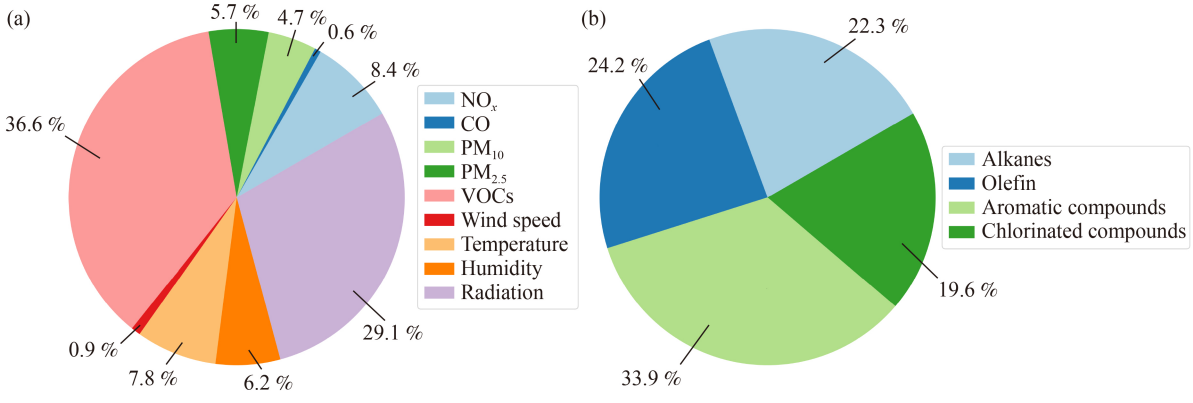


Fig. 4 Importance weight of different features on ozone concentration. (a) Importance weight of all features on ozone concentration, (b) importance weight of the four types of species in VOCs on ozone concentration.

Table 1 Forecast performance changes with forecast duration

Time (h)	MAE (μg/m^3)	RMSE (μg/m^3)	R^2
1	5.97	7.71	0.958
2	6.54	9.18	0.940
3	7.59	10.95	0.914
6	9.54	14.11	0.858
12	13.13	17.98	0.770
24	14.71	20.57	0.699

December 2020. The predicted values of the ozone concentration in most time periods were comparable to the observed values. The RMSE values for July, August, November, and December were 11.72 μg/m^3, 8.91 μg/m^3, 9.04 μg/m^3, and 7.85 μg/m^3, respectively. The results show that the DA-RNN model can be used to predict ozone concentration over a short period of time.

However, as shown in Fig. 6, Jinshan Station had ozone peaks exceeding 250 μg/m^3 in some time periods in July and August 2020. For example, at 16:00 on July 3, 2020, the observed value was 281.46 μg/m^3 and the predicted value was 196.37 μg/m^3. At 18:00 on August 24, 2020, the observed value was 262.43 μg/m^3 and the predicted value was 192.48 μg/m^3. As shown in Fig. 6, during the period when the ozone concentration was low, the prediction performance of the model was better. Therefore, the model can make accurate predictions for low ozone values but is not suitable for prediction at extremely high ozone values. Extremely high values appear less frequently in the training set, which could account for the model's impaired performance in identifying the characteristics at high ozone concentrations.

Figs. 7(a)–7(c) shows the prediction performance of the

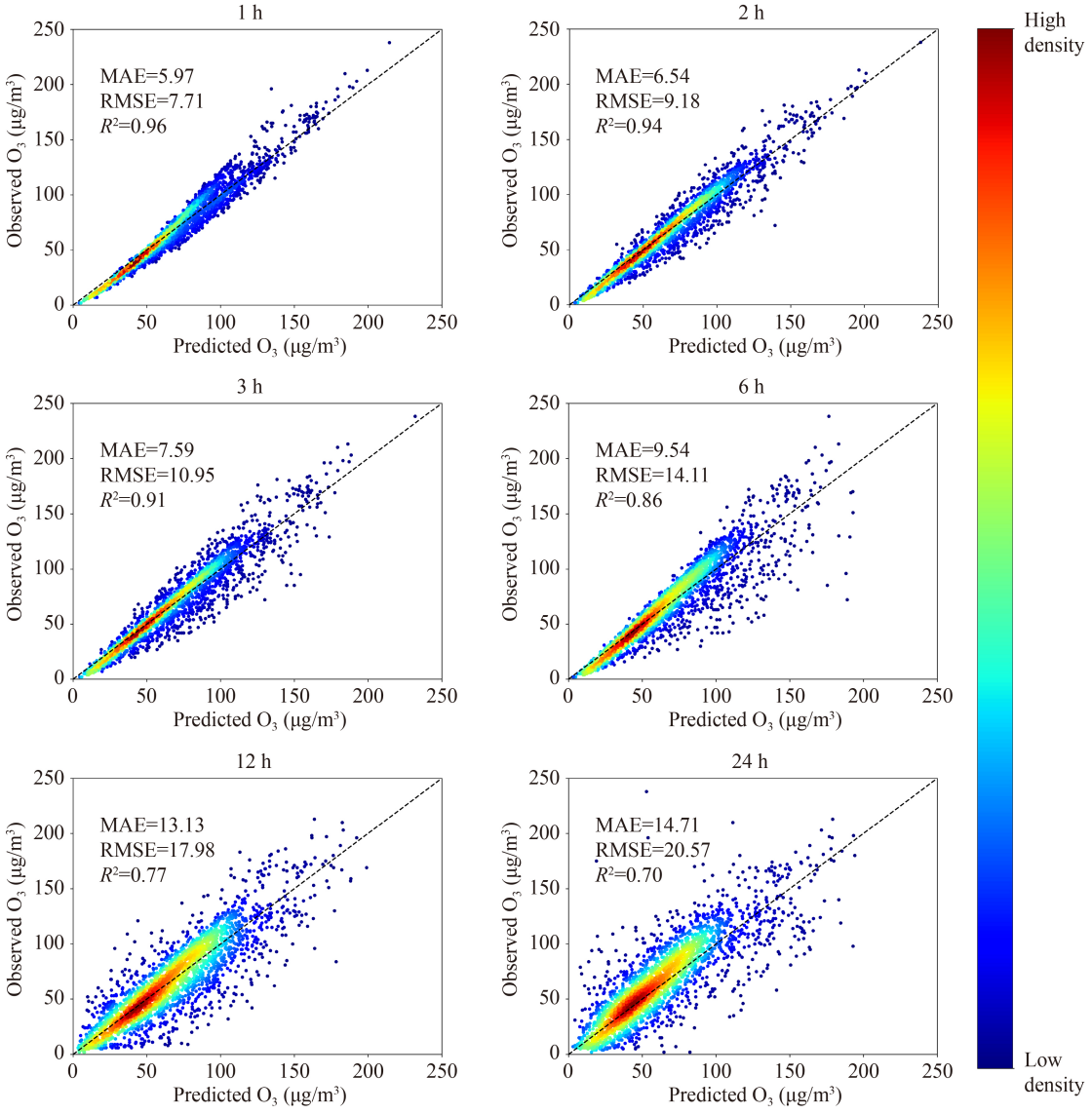


Fig. 5 Observed *vs.* predicted values of ozone concentration.

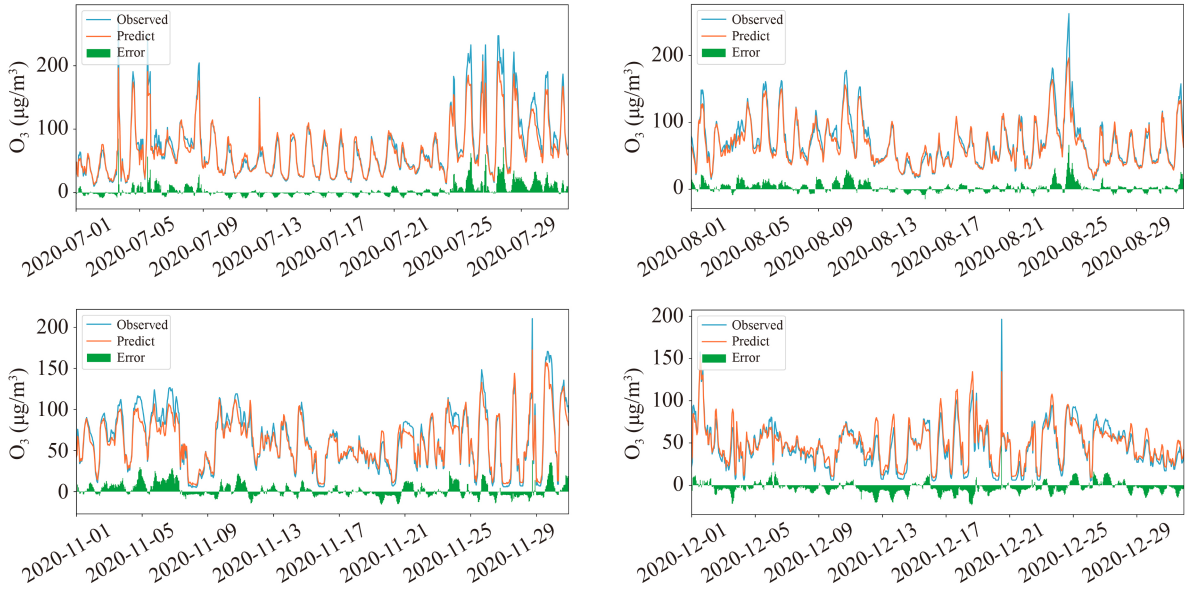


Fig. 6 Comparison of predicted and observed values of hourly ozone concentration in Jinshan Station in July, August, November, and December 2020.

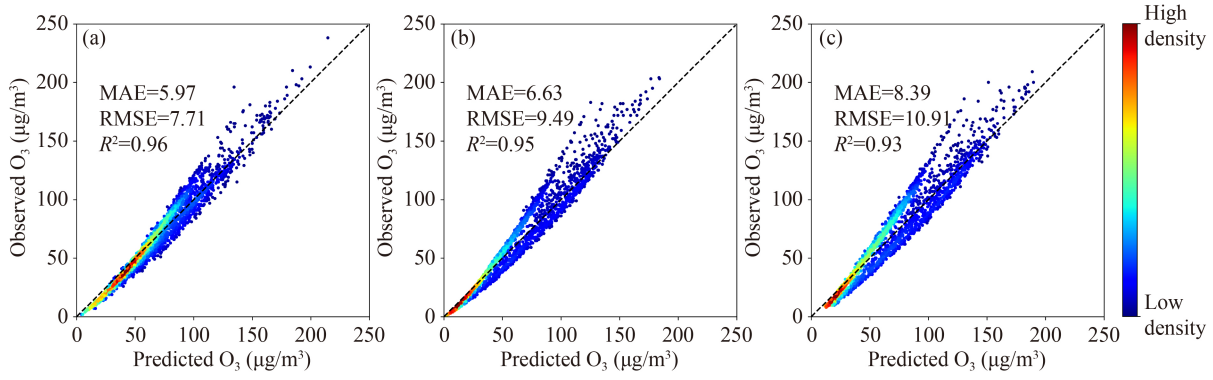


Fig. 7 Observed *vs.* predicted values of ozone concentrations at different stations. (a) Jinshan Station; (b) Jinshan Industrial District Station; (c) Jinshan Boundary Station.

DA-RNN model at different sites. For example, the RMSE (μg/m^3), MAE (μg/m^3), and R^2 values of the Jinshan Industrial District Station were 9.49, 6.63, and 0.95, respectively; the corresponding values of the Jinshan Boundary Station were 10.91, 8.39, and 0.93, respectively. This shows that the DA-RNN model has good generalization performance.

4.5 Comparing results with other models

Table 2 presents a comparison of the 1 h prediction performance of the DA-RAN model and other machine learning models. The DA-RNN model shows a significant improvement in the evaluation indicators of R^2, MAE, and RMSE compared with the other models. We speculate that this is due to the excellent ability of the DA-RNN model to capture multidimensional time-series features, and its performance on training sets with more dimensions is significantly better than that of other models.

We compared the prediction results of the DA-RNN model with those of the WRF-CMAQ model. As the simulation results of WRF-CMAQ are time-consuming, we chose only two representative months, July and November, 2020. We selected the grid points closest to Jinshan station and used the ozone concentration data at the grid points as the predicted values of the model. As shown in Fig. 8, the ozone concentration predicted by the DA-RNN model was closer to the observed value than the WRF-CMAQ model in most time periods, and the ozone concentration predicted by the WRF-CMAQ model was generally lower than the observed value. For extremely high values, neither the DA-RNN model nor the WRF-CMAQ model could predict very well, which was caused by too few extremely high-value samples.

Table 3 lists the prediction performance of the WRF-

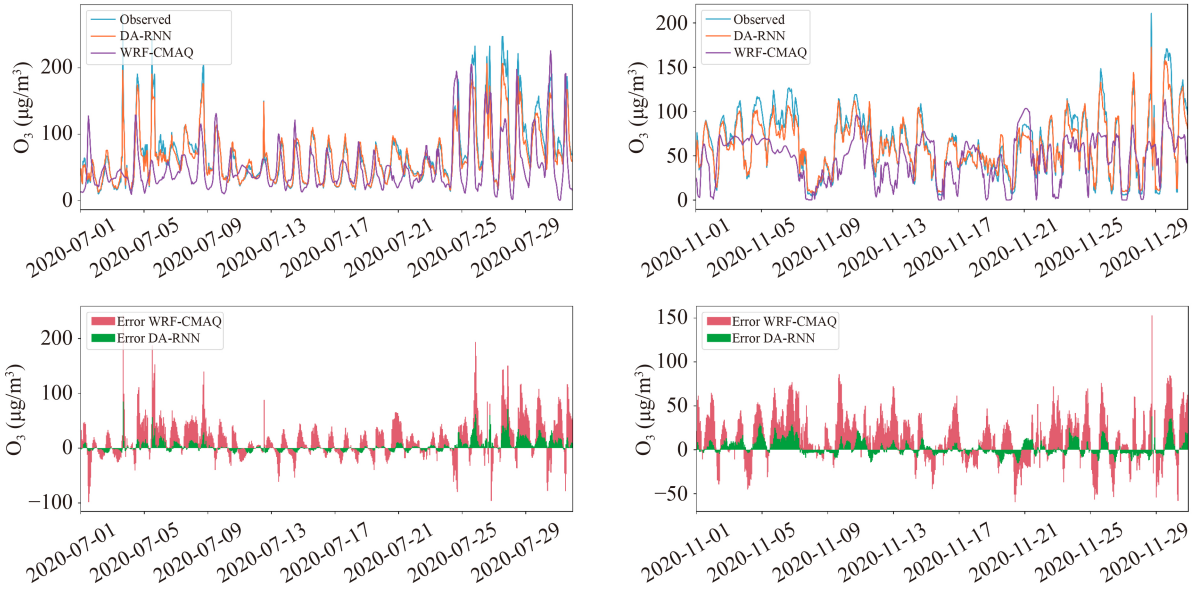


Fig. 8 Comparison of the WRF-CMAQ and DA-RNN models.

Table 2 Predictive performance of other machine learning models and the DA-RNN model

Model	R^2	MAE (μg/m³)	RMSE (μg/m³)
DA-RNN	0.958	5.97	7.71
CatBoost	0.901	8.46	13.25
LightGBM	0.896	8.72	13.61
LightGBMLarge	0.889	8.96	14.07
XGBoost	0.888	9.08	14.14
RandomForest	0.848	10.93	16.43
ExtraTrees	0.842	11.2	16.8
NeuralNet	0.832	11.39	17.31
KNeighbors	0.748	14.19	21.19

CMAQ and DA-RNN models. Both models have better predictions for low-value time periods, and the prediction performance of DA-RNN is better than that of WRF-CMAQ. Owing to the impact of the accuracy of the emission inventory, the accuracy of the meteorological data, and the completeness of the simulated chemical mechanism, WRF-CMAQ cannot predict the ozone concentration well, and it takes a long time to calculate. Compared with WRF-CMAQ, DA-RNN can predict future ozone concentration by learning the existing data samples to determine the dependencies and importance weights of features, which can efficiently and conveniently achieve good prediction results.

5 Conclusions

Currently, ozone has surpassed $PM_{2.5}$ and is the primary pollutant in some high-incidence seasons. Therefore, it is important to conduct real-time forecasts and ozone monitoring and study the direct synergistic relationship between ozone precursors and ozone reduction. In this study, we used a double-stage attention-based recurrent neural network (DA-RNN) to monitor and forecast ozone in real time and analyze the importance of these features. This model appropriately captures the long-term time dependence based on the previous value of the time series and multiple time series and selects the relatively important time series for prediction.

We calculated the results of different prediction periods, RMSE, MAE, and R^2 to assess the predictive performance of the model. The results showed that the short-term prediction results of the DA-RNN model were more accurate. Within the most precise prediction time (1 h), the values of MAE, RMSE, and R^2 were 5.97 μg/m³, 7.71 μg/m³, and 0.958, respectively. The model's accuracy decreased as the forecast time increased but still had good accuracy in the 24-h forecast, with MAE, RMSE, and R^2 values of 14.71 μg/m³, 20.57 μg/m³, and 0.70, respectively. By analyzing the importance of the characteristics of the Jinshan Industrial Zone, we found that ozone pollution in this site comes from other ozone precursors such as VOCs in the petrochemical industry. We explored the relationship between the concentrations of precursors and ozone based on the usual ozone prediction. We predicted the relationship between the concentration change of ozone and the evolution of the precursor concentration. Therefore, our model can effectively help environmental protection departments monitor and provide early warnings of ozone pollution and take targeted preventive measures.

To verify the model's prediction performance, we compared the DA-RNN model with other machine

Table 3 Predictive performance of the WRF-CMAQ and DA-RNN models

	1–31 Jul 2020		1–30 Nov 2020	
	WRF-CMAQ	DA-RNN	WRF-CMAQ	DA-RNN
MAE (μg/m^3)	31.22	7.03	27.15	6.74
RMSE (μg/m^3)	43.03	11.72	33.38	9.04

learning models. The DA-RNN model outperformed other machine learning models in 1-h predictions. We also compared the DA-RNN model with the WRF-CMAQ model. In the two representative months we selected, the average RMSE and MAE of the DA-RNN model were 8.63 μg/m^3, which was much lower than the 33.70 μg/m^3 of WRF-CMAQ. Compared with the WRF-CMAQ model, the DA-RNN model requires less data for prediction, is shorter and more time consuming, and has better versatility. To verify the generalization performance of the model, we selected three different types of stations for the experiments. In the 1-h prediction performance of the three sites, MAE, RMSE, and R^2 outperformed the other machine learning models. The excellent performance of the DA-RNN model in terms of generalization ability and generality was explained.

Owing to the structure of its deep neural network, the DA-RNN model can identify complex synergistic and nonlinear relationships from the concentration data of different pollutants through learning. Therefore, compared to traditional models, it can predict the concentration of pollutants more efficiently and accurately. However, owing to the limitation of the number of samples in the training set, the model cannot predict a high concentration value. Further research can improve the performance of the model in predicting extremely high values by collecting samples with extremely high values.

Acknowledgements The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study. We would like to acknowledge Ling Huang for his help with the simulation of the WRF-CMAQ model, Key Research and Development Projects of the Shanghai Science and Technology Commission (No. 20dz1204000).

Data Accessibility Statement Data not available due to the data sensitivity and software copyright restrictions.

Electronic Supplementary Material Supplementary material is available in the online version of this article at https://doi.org/10.1007/s11783-023-1621-4 and is accessible for authorized users.

References

Atkinson R (2000). Atmospheric chemistry of VOCs and NO_x. Atmospheric Environment, 34(12–14): 2063–2101

Avery R J (2006). Reactivity-based VOC control for solvent products: more efficient ozone reduction strategies. Environmental Science & Technology, 40(16): 4845–4850

Bas M D, Ortiz J, Ballesteros L, Martorell S (2017). Evaluation of a multiple linear regression model and SARIMA model in forecasting ^{7}Be air concentrations. Chemosphere, 177: 326–333

Bengio Y, Simard P, Frasconi P (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2): 157–166

Byun D, Schere K L (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 community multiscale air quality (CMAQ) modeling system. Applied Mechanics Reviews, 59(2): 51–77

Chen Y, Phonevilay V, Tao J, Chen X, Xia R, Zhang Q, Yang K, Xiong J, Xie J (2021). The face image super-resolution algorithm based on combined representation learning. Multimedia Tools and Applications, 80(20): 30839–30861

Chin M, Ginoux P, Kinne S, Torres O, Holben B N, Duncan B N, Martin R V, Logan J A, Higurashi A, Nakajima T (2002). Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. Journal of the Atmospheric Sciences, 59(3): 461–483

Delle Monache L, Nipen T, Deng X X, Zhou Y M, Stull R (2006). Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. Journal of Geophysical Research, 111(D5): D05308

Ding Y, Chen Z, Lu W, Wang X (2021). A CatBoost approach with wavelet decomposition to improve satellite-derived high-resolution $PM_{2.5}$ estimates in Beijing-Tianjin-Hebei. Atmospheric Environment, 249: 118212

Donnelly A, Misstear B, Broderick B (2015). Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmospheric Environment, 103: 53–65

Efstathiou G A, Zoumakis N M, Melas D, Lolis C J, Kassomenos P (2013). Sensitivity of WRF to boundary layer parameterizations in simulating a heavy rainfall event using different microphysical schemes: effect on large-scale processes. Atmospheric Research, 132–133: 125–143

Ek M B, Mitchell K E, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley J D (2003). Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale Eta model. Journal of Geophysical Research, 108(D22): 2002JD003296

Erdem E, Shi J (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. Applied Energy, 88(4): 1405–1414

Garaud D, Mallet V (2011). Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: application to air quality. Journal of Geophysical Research, 116(D19): D19304

Ge B Z, Wang Z F, Xu X B, Wu J B, Yu X L, Li J (2014). Wet deposition of acidifying substances in different regions of China and the rest of East Asia: modeling with updated NAQPMS. Environmental Pollution, 187: 10–21

Hong S Y, Noh Y, Dudhia J (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. Monthly Weather Review, 134(9): 2318–2341

Hu X, Belle J H, Meng X, Wildani A, Waller L A, Strickland M J, Liu Y (2017). Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach. Environmental Science & Technology, 51(12): 6936–6944

Hui L, Liu X, Tan Q, Feng M, An J, Qu Y, Zhang Y, Jiang M (2018). Characteristics, source apportionment and contribution of VOCs to ozone formation in Wuhan, Central China. Atmospheric Environment, 192: 55–71

Jia P, Cao N, Yang S (2021). Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence to sequence model. Atmospheric Environment, 244: 117917

Kampa M, Castanas E (2008). Human health effects of air pollution. Environmental Pollution, 151(2): 362–367

Kavasseri R G, Seetharaman K (2009). Day-ahead wind speed forecasting using f-ARIMA models. Renewable Energy, 34(5): 1388–1393

Liu C, Zhang H, Cheng Z, Shen J, Zhao J, Wang Y, Wang S, Cheng Y (2021). Emulation of an atmospheric gas-phase chemistry solver through deep learning: case study of Chinese mainland. Atmospheric Pollution Research, 12(6): 101079

Mlawer E J, Taubman S J, Brown P D, Iacono M J, Clough S A (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research, 102(D14): 16663–16682

Molina-Gómez N I, Diaz-Arevalo J L, Lopez-Jimenez P A (2021). Air quality and urban sustainable development: the application of machine learning tools. International Journal of Environmental Science and Technology, 18(4): 1029–1046

Ou J, Zheng J, Li R, Huang X, Zhong Z, Zhong L, Lin H (2015). Speciated OVOC and VOC emission inventories and their implications for reactivity-based ozone control strategy in the Pearl River Delta region, China. Science of theTotal Environment, 530–531: 393–402

Qin Y, Song D, Cheng H, Cheng W, Jiang G, Cottrell G W (2017). A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. Melbourne, Australia: 2627–2633

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743): 195–204

Shao M, Zhang Y, Zeng L, Tang X, Zhang J, Zhong L, Wang B (2009). Ground-level ozone in the Pearl River Delta and the roles of VOC and NO_x in its production. Journal of Environmental Management, 90(1): 512–518

Tao Q, Liu F, Li Y, Sidorov D (2019). Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. IEEE Access: Practical Innovations, Open Solutions, 7: 76690–76698

Vautard R, Builtjes P H J, Thunis P, Cuvelier C, Bedogni M, Bessagnet B, Honore C, Moussiopoulos N, Pirovano G, Schaap M, Stern R, Tarrason L, Wind P (2007). Evaluation and intercomparison of ozone and PM_{10} simulations by several chemistry transport models over four European cities within the CityDelta project. Atmospheric Environment, 41(1): 173–188

Wang B, Yuan Q, Yang Q, Zhu L, Li T, Zhang L (2021). Estimate hourly $PM_{2.5}$ concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network. Environmental Pollution, 271: 116327

Wang H, Xiang Z, Wang L, Jing S, Lou S, Tao S, Liu J, Yu M, Li L, Lin L, Chen Y, Wiedensohler A, Chen C (2018). Emissions of volatile organic compounds (VOCs) from cooking and their speciation: A case study for Shanghai with implications for China. Science of the Total Environment, 621: 1300–1309

Wang S, Fang C, Ma H, Wang Y, Qin J (2014). Spatial differences and multi-mechanism of carbon footprint based on GWR model in provincial China. Journal of Geographical Sciences, 24(4): 612–630

Wong P Y, Su H J, Lee H Y, Chen Y C, Hsiao Y P, Huang J W, Teo T A, Wu C D, Spengler J D (2021). Using land-use machine learning models to estimate daily NO_2 concentration variations in Taiwan, China. Journal of Cleaner Production, 317: 128411

Xia R, Chen Y, Ren B (2022). Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. Journal of King Saud University - Computer and Information Sciences

Xu X, Zhang C, Liang Y (2021). Review of satellite-driven statistical models $PM_{2.5}$ concentration estimation with comprehensive information. Atmospheric Environment, 256: 118302

Zhang H, Linford J C, Sandu A, Sander R (2011). Chemical mechanism solvers in air quality models. Atmosphere (Toronto), 2(3): 510–532

Zhang J M, Feng W J, Yuan T Y, Wang J, Sangaiah A K (2022a). SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking. Applied Soft Computing, 118: 108485

Zhang J M, Sun J, Wang J, Li Z P, Chen X (2022b). An object tracking framework with recapture based on correlation filters and Siamese networks. Computers & Electrical Engineering, 98: 107730

Zhang Y, Wen X Y, Jang C J (2010). Simulating chemistry-aerosol-cloud-radiation-climate feedbacks over the continental US using the online-coupled Weather Research Forecasting Model with chemistry (WRF/Chem). Atmospheric Environment, 44(29): 3568–3582