Journal of Cleaner Production 253 (2020) 119841

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach

Hong-Wei Wang [a], Xiao-Bing Li [b], Dongsheng Wang [a], Juanhao Zhao [c], Hong-di He [a, *], Zhong-Ren Peng [d, **]

[a] Center for Intelligent Transportation Systems and Unmanned Aerial Systems Applications, State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] Institute for Environmental and Climate Research, Jinan University, Guangzhou, 510632, China
[c] Department of Computer Science, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, 90089, USA
[d] International Center for Adaptation Planning and Design (iAdapt), School of Landscape Architecture and Planning, College of Design, Construction and Planning, University of Florida, P.O. Box 115706, Gainesville, FL, 32611-5706, USA

ARTICLE INFO

Article history:
Received 7 July 2019
Received in revised form
22 November 2019
Accepted 21 December 2019
Available online 27 December 2019

Handling editor: Kathleen Aviso

Keywords:
Ozone prediction
Spatiotemporal correlation
Air quality monitoring network
Sequence to sequence model
Deep learning

ABSTRACT

Ozone is one of the most important greenhouse gases and air pollutants in urban areas, and has significantly negative impacts both on the climate change and human health. In addition to alert the public for health concerns, accurate ozone prediction is also crucial to understand the formation mechanisms of surface ozone episodes, and has significant implications for making emission control strategies of ozone precursors such as methane, carbon monoxide, and volatile organic compounds. However, existing methods of ozone concentration prediction could not effectively capture temporal dependency, and most neglect spatial correlations. In this study, a hybrid sequence to sequence model embedded with the attention mechanism is proposed for predicting regional ground-level ozone concentration. In the proposed model, the inherent spatiotemporal correlations in air quality monitoring network are simultaneously extracted, learned and incorporated, and auxiliary air pollution and meteorological information are adaptively involved. The hourly data collected from 35 air quality monitoring stations and 651 meteorological girds in Beijing, China are used to validate the present model. The results demonstrate that the spatiotemporal correlations in the monitoring network present enormous advantages for the regional ozone prediction. Auxiliary data and time lags matching day-of-week or diurnal periods of ozone are confirmed to benefit the improvement of prediction accuracy. Monitoring stations in urban areas exhibit better prediction performances than stations in remote areas. The addressed model outperforms the baseline models, and is proven to have excellent performance in all monitoring station categories of Beijing and different months with significant disparity of ozone concentrations.
© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Tropospheric ozone is an important greenhouse gas and air pollutant, and has detrimental impacts both on climate change and human health. In recent years, great efforts have been made to reduce surface ozone levels by implementing stringent emission control measures of ozone precursors (NOx and VOCs)(Cheng et al., 2019), which is also closely associated with the reduction budget of carbon emissions around the world (Ye et al., 2019; Fasihi et al., 2019; Yusuf et al., 2019). However, severe ozone pollution episodes still frequently occurred over urban areas, especially in developing countries like China (Su et al., 2017). In comparison to aerosol pollution, people usually do not give many concerns on ozone pollution issues due to insignificant sensory stimulations and take necessary measures in advance to reduce their exposure levels to ozone pollution (He et al., 2017). Thus, it is extremely important to provide accurate forecasts in urban areas to alert the public when ambient ozone concentration will exceed preset thresholds (Gradišar et al., 2016). Compared with the prediction of ozone concentration only at monitoring stations (Awang et al., 2015; Gao et al., 2018), predicting the regional ozone concentration suggests higher spatial resolution and directly reflects the

* Corresponding author.
** Corresponding author.
E-mail addresses: hongdihe@sjtu.edu.cn (H.-d. He), zpeng@ufl.edu (Z.-R. Peng).

https://doi.org/10.1016/j.jclepro.2019.119841
0959-6526/© 2019 Elsevier Ltd. All rights reserved.

spatial distribution change of ozone, which assists to the pollution control and the emission reduction of ozone. Therefore, conducting accurate regional prediction for ground-level ozone concentration is of greatly significance for both public health protection and emission reduction of greenhouse gases.

Generally, ozone prediction models can be classified into two types, including deterministic models and empirical models (Li et al., 2017; Wen et al., 2019). Deterministic models refer to numerical simulation models that incorporate various physical and chemical mechanisms related to the emission, transformation, dispersion and transport of air pollutants (Sharma et al., 2016), such as the Weather Research Forecasting (WRF) model. In the WRF model, different combinations of macro-scale physical process parameterization schemes (e.g. planetary boundary layer schemes and land surface modelling schemes) have significant impacts on the accuracy of model results (Cai et al., 2016). Therefore, deterministic models are favorable for the analysis of air quality problems at large spatial scales, but have limited prediction accuracies at micro scales like urban areas (Gradišar et al., 2016). In empirical models, statistics and machine learning methods have been widely developed for general environmental predictions (e.g. ozone and aerosol pollution, renewable energy utilization and carbon emission reduction). Commonly used statistics models include Multiple Linear Regression (MLR)(Awang et al., 2015), Multivariate Adaptive Regression Splines (Srinivas et al., 2019) and Principal Component Analysis (PCA)(Huang et al., 2019). Recently, machine learning models are rapidly developed to improve the performances of general environmental predictions due to their nonlinear mapping ability, such as Random Forest (Feng et al., 2019; Srinivas et al., 2019), Bayesian Network (Mehdipour et al., 2018), Extreme Gradient Boosting (XGBoost)(Ma et al., 2019) and Support Vector Machine (SVM) (He et al., 2018; Mehdipour et al., 2018; Mehdipour and Memarianfard, 2019; Sumathi and Manivannan, 2019). These machine learning models take nonlinearity into account, and present strong abilities to capture temporal features of air pollutants and renewable resources.

As a form of machine learning, deep learning models are based on artificial neural networks with more complicated architectures, and have grown up to efficient approaches for handling multi-dimensional data due to flexible model structure, strong generalization, and powerful learning ability (Li et al., 2019; Hao et al., 2019). Recently, several deep learning methods have achieved outstanding performance in general environmental prediction issues (e.g., air pollution, carbon emission and renewable energy), such as Deep Neural Network (DNN)(Gao et al., 2018; Acheampong and Boateng, 2019), Stacked Autoencoder (SAE)(Bai et al., 2019), Recurrent Neural Network (RNN)(Feng et al., 2019), Long Short-Term Memory Recurrent Network (LSTM)(Li et al., 2017, Li et al., 2019; Zhou et al., 2019; Huang et al., 2019) and Gated Recurrent Network (GRU)(Athira et al., 2018). However, RNN-based models (i.e., LSTM and GRU) neglect neighboring factors in air quality monitoring network and are incapable of characterizing spatial correlations (Wen et al., 2019). Based on the idea of Convolutional Neural Network (CNN), CNN-LSTM (Wen et al., 2019) and Conv-LSTM (Shi et al., 2015) are employed to conduct temporal predictions considering spatial features in grid-structured data (e.g., images). These CNN-based models can effectively capture the spatiotemporal dependencies, but have limitations to handle long-term input data because of simply stacking multiple layers of LSTM units (Hao et al., 2019). In recent years, the emergence of Sequence to Sequence learning (Seq2Seq) (Cho et al., 2014) provides a more flexible and extendable framework on temporal modelling. However, the seq2seq model suffers from a potential issue that all historical data need to be compressed into an internal vector with fixed length (Bahdanau et al., 2015). This may limit the ability of the seq2seq model coping with long sequences (Cho et al., 2014; Bahdanau et al., 2015). Thus, attention mechanism methods (Bahdanau et al., 2015) are proposed to resolve the drawbacks of simple Seq2Seq and adept at modelling long-range dependencies (Bahdanau et al., 2015; Hao et al., 2019), such as long-term processes of chemical and physical mechanisms in air pollution episodes (Liu et al., 2019).

In this study, we aim to propose a hybrid sequence to sequence model embedded with the attention mechanism (HSA-Net) to predict the ozone concentration at a city scale. In order to fully consider the spatiotemporal correlations, the neighboring factors in the monitoring network are adaptively learned by the proposed model, and the spatial information of the entire monitoring network is well extracted and integrated. The dataset collected in Beijing, China is used for training, validating and testing the proposed model, and the auxiliary air pollution data (SO_2, NO_2, $PM_{2.5}$, PM_{10}, and CO), meteorological data and weather forecast data are also employed. The novelties of the presented study are summarized as follows: (1) The advantage of incorporating spatiotemporal correlations in the model is quantitatively explored and evaluated. (2) The spatiotemporal dependencies of air quality monitoring network are simultaneously captured. (3) The auxiliary air pollution and meteorological data are integrated into the model. (4) The ozone prediction performances of the proposed model are estimated monthly and geographically by different categories of monitoring stations.

2. Data and methods

The details of data are described in this section, then we introduce the methodology of RNN-series methods, Seq2Seq model and attention mechanism, and finally the experiment settings and network architecture are presented.

2.1. Data description

In this study, Beijing, the capital of China, is selected as the study area. The research area of interest approximately ranges from E115° to E118° and N39° to N41°. Hourly mean concentrations of six regulatory air pollutants (in China) including O_3 ($\mu g/m^3$), SO_2 ($\mu g/m^3$), NO_2 ($\mu g/m^3$), $PM_{2.5}$ ($\mu g/m^3$), PM_{10} ($\mu g/m^3$), and CO (mg/m^3) are obtained from 35 air quality monitoring stations from January 1, 2017 to May 31, 2018. Hourly gridding meteorological data (21×31 points) in the same period are processed by the Weather Research and Forecasting (WRF) model, and have a grid spacing of 5 km. Meteorological parameters including temperature, air pressure, relative humidity, wind speed, and wind direction are selected as the principal meteorological features due to their close relationships with the change of ground-level ozone concentration (Li et al., 2017; Wen et al., 2019; Bai et al., 2019). Fig. 1 shows the geographical locations of the air quality monitoring stations. The monitoring stations are officially divided into four categories according to their locations and surrounding environments. The stations labelled by 1–12, 13 to 17, 18 to 28 and 29 to 35 refer to urban, traffic, rural and cross reference monitoring stations, respectively. Urban and rural monitoring stations (Cheng et al., 2019) are used to assess the average condition and variation of air quality in urban and suburban environments, respectively; Traffic monitoring stations are used to monitor the impact of road traffic sources on ambient air quality; Cross reference monitoring stations are used to characterize the regional environmental background level and reflect the transmission of pollution within the area.

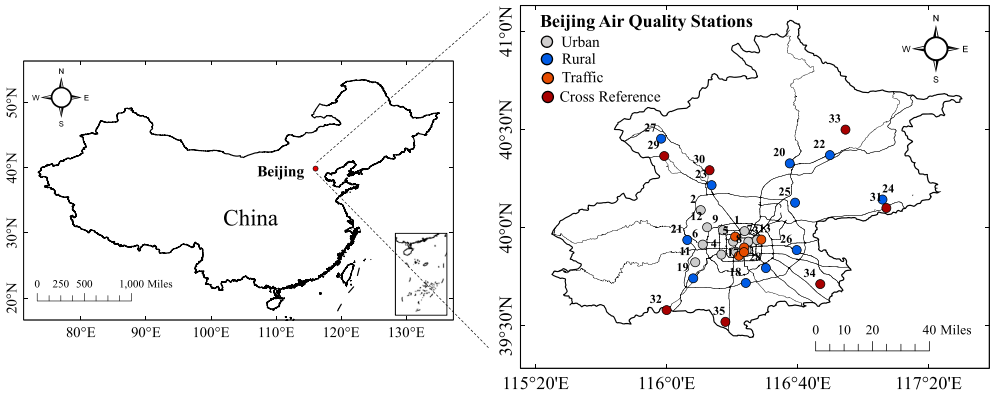


Fig. 1. Geographical locations of the air quality monitoring stations in Beijing.

2.2. Hybrid sequence to sequence model with attention (HSA-Net)

In this section, we first elaborate the methodologies involved in the HSA-Net model, from LSTM, GRU, Bidirectional LSTM (Bi-LSTM) to Seq2Seq architecture with attention mechanism, then introduce the structure details of the HSA-Net model, as shown in Fig. 2. In temporal modelling, the encoder is built by Bi-LSTM network that can read input data in both forward and backward directions, and the decoder is developed by GRU. The combination of the Bi-LSTM encoder and the GRU decoder is widely-used in Seq2Seq learning (Bahdanau et al., 2015; Liu et al., 2019). In spatial modelling, the Deep Neural Network (DNN) is implemented for extracting and

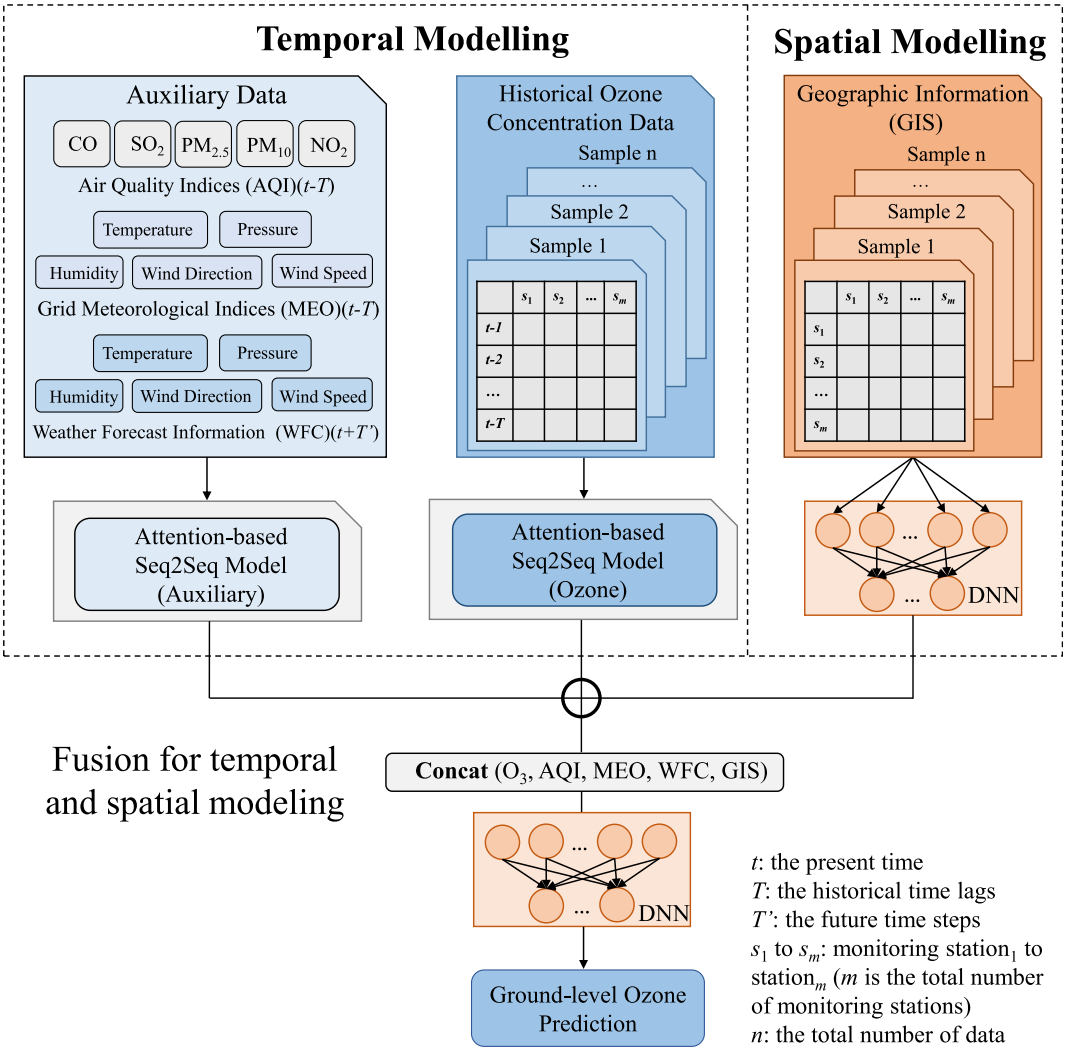


Fig. 2. The framework of the hybrid seq2seq model with attention (HSA-Net).

modelling the spatial information in the air quality monitoring network. To merge these features, reshaping, repeating and concatenating techniques are applied to transform the above variables into acquired sizes. In Fig. 2, t, T and T' respectively denote the present time, the historical time lags and the future time steps.

2.2.1. Recurrent neural network methods

Traditional recurrent neural network (RNN) can learn the past information to make predictions, but has limitation in modelling long input data due to the vanishing gradient problem (Cho et al., 2014). Therefore, a special Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) was proposed to fix the vanishing gradient issues, as shown in Fig. 3 (a). The LSTM is composed of one input layer, one output layer and one recurrent hidden layer within a basic unit called memory block. The memory block contains self-connected memory cells for recording the temporal state, and three multiplicative units (input, output and forget gates) to control write, read and reset operations in the block. According to the status of a memory cell, multiplicative gates can learn to open and close by minimizing the network error, and then allow the memory cell to store and access time series with long time spans. The equations of LSTM (Shi et al., 2015) are shown below:

$$i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \odot c_{t-1} + b_i) \tag{1}$$

$$f_t = \delta\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \odot c_{t-1} + b_f\right) \tag{2}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tan h(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{3}$$

$$o_t = \delta(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \odot c_t + b_o) \tag{4}$$

$$h_t = o_t \odot \tanh(c_t) \tag{5}$$

where i_t, f_t, o_t respectively represent the state of the input gate, forget gate and output gate at time t, c_t and h_t refer the memory cell state and the hidden state, W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf}, W_{xc}, W_{hc}, W_{xo}, W_{ho} and W_{co} are the trainable parameters, the variable b_i, b_f, b_c and b_o are the bias terms. The operator $\odot$ refers to the Hadamard product. δ and $tanh$ separately denote logistics sigmoid function and hyperbolic tangent function:

$$\delta(x) = \frac{1}{1+e^{-x}} \tag{6}$$

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{7}$$

Gated Recurrent Unit (GRU) is a powerful variant of Recurrent Neural Network (Cho et al., 2014) to adaptively capture dependencies of different time lengths, as shown in Fig. 3 (b). The initial GRU has two internal gated recurrent units (update gate and reset gate) to capture long-term dependencies. The equations of GRU (Cho et al., 2014) are shown as:

$$r_t = \delta(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{8}$$

$$u_t = \delta(W_{xu}x_t + W_{hu}h_{t-1} + b_u) \tag{9}$$

$$z_t = tanh(W_{xz}x_t + W_{hz}(r_t \odot h_{t-1}) + b_z) \tag{10}$$

$$h_t = u_t \odot h_{t-1} + (1-u_t) \odot z_t \tag{11}$$

where r_t, u_t respectively refer to the states of reset gate and update gate at time t, z_t and h_t respectively represent candidate states and hidden states, W_{xr}, W_{hr}, W_{xu}, W_{hu}, W_{xz}, W_{hz} are the trainable parameters, the variable b_r, b_u, b_z are the bias terms.

One shortcoming of LSTM and GRU is that only previous information can be utilized (Graves et al., 2013). To address this issue, Bidirectional Recurrent Neural Network (Bi-RNN) was proposed (Schuster and Paliwal, 1997) to increase the amount of input information available to the network, as shown in Fig. 3 (c). The basic idea of Bi-RNN is to connect two hidden layers of opposite directions to the same output, both considering backward and forward states of input information. The equations of Bi-RNN (Graves et al., 2013) are shown below:

$$\overrightarrow{h}_t = \delta(W_{x\overrightarrow{h}}x_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \tag{12}$$

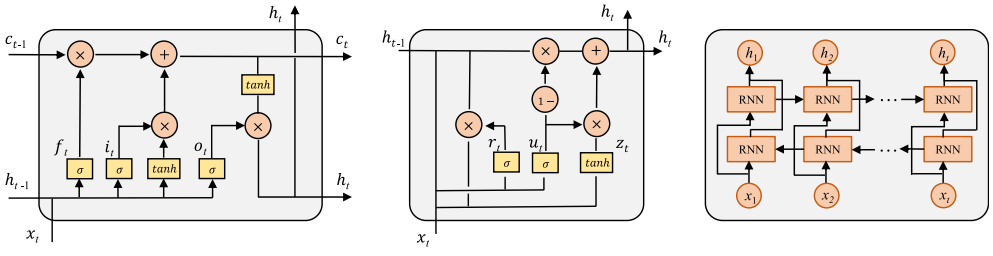
$$\overleftarrow{h}_t = \delta(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}) \tag{13}$$

$$y_t = W_{\overrightarrow{h}y}\overrightarrow{h}_t + W_{h\leftarrow y}h_t^{\leftarrow} + b_y \tag{14}$$

where, $\overrightarrow{h}_t$, $\overleftarrow{h}_t$ and y_t respectively represent forward hidden states, backward hidden states, and predictions at time t. $W_{x\overrightarrow{h}}$, $W_{\overrightarrow{h}\overrightarrow{h}}$, $W_{x\overleftarrow{h}}$, $W_{\overleftarrow{h}\overleftarrow{h}}$, $W_{\overrightarrow{h}y}$ and $W_{\overleftarrow{h}y}$ are the trainable parameters, the variable $b_{\overrightarrow{h}}$, $b_{\overleftarrow{h}}$ and b_y are the bias terms. Combining Bi-RNN with LSTM gives bidirectional LSTM (Bi-LSTM), which exhibits a superior ability to model long-range time series in both input directions (Bahdanau et al., 2015).

2.2.2. Attention-based Seq2Seq learning for temporal modeling

The Seq2Seq model (Cho et al., 2014) is an Encoder-Decoder architecture based on RNN units (e.g. LSTM and GRU). Historical data are read and encoded to a context vector by an encoder RNN, then the predictions are generated from that vector by a decoder RNN. However, the performance of a Seq2Seq network significantly



(a) LSTM (b) GRU (c) Bi-RNN

Fig. 3. The architectures of RNN-series models: (a) LSTM, (b) GRU, (c) Bi-RNN.

drops with the increased length of input series (Cho et al., 2014), and surfers in the fixed length of the internal vector (Bahdanau et al., 2015). Thus, the attention mechanism (Bahdanau et al., 2015) in the decoder is introduced to adaptively choose relevant hidden states in the encoder for each time step, which is able to model the dynamic temporal dependency between the intervals of source sequence and the target sequence, as shown in Fig. 4. In attention architecture, Bi-LSTM is usually used as the encoder to considering input information in both forward and backward directions (Bahdanau et al., 2015; Hao et al., 2019). The hidden states from a Bi-LSTM encoder are denoted by h_t, which is a concatenation of the forward hidden states $\overrightarrow{h}_t$ and the backward hidden states $\overleftarrow{h}_t$. Thus, the attention weights at future time $t^{'}$ in the decoder, denoted by $\alpha_{t^{'}t}$, is computed by:

$$\alpha_{t^{'}t} = \frac{\exp(e_{t^{'}t})}{\sum_{t=1}^{T_x}\exp(e_{t^{'}t})} \tag{15}$$

$$e_{t^{'}t} = v_e^{\mathrm{T}}\tanh(W_{se}s_{t^{'}-1} + W_{he}h_t) \tag{16}$$

where T_x denotes the length of input sequence, $e_{t^{'}t}$ is an alignment function to calculate the unnormalized match degree of the last hidden state $s_{t^{'}-1}$ in the decoder and the current hidden state h_t in the encoder, v_e, W_{se} and W_{he} are learnable parameters. Then, $e_{t^{'}t}$ is normalized by a softmax function to build the attention weights $\alpha_{t^{'}t}$. Thus, the context vector $c_{t^{'}}$ of the attention mechanism can be updated as:

$$c_{t^{'}} = \sum_{t=1}^{T_x}\alpha_{t^{'}t}h_t \tag{17}$$

Then the prediction output $y_{t^{'}}$ based on the attention mechanism can be also computed from the GRU decoder:

$$y_{t^{'}} = p(y_{t^{'}-1},\ s_{t^{'}}, c_{t^{'}}) \tag{18}$$

$$s_{t^{'}} = q(y_{t^{'}-1}, s_{t^{'}-1}, c_{t^{'}}) \tag{19}$$

where f, p and q refer to different nonlinear functions, respectively.

2.2.3. Deep neural networks for spatial modeling

The adjacency matrix of relative Euclidean distance $K \in \mathbb{R}^{m\times m}$ is the geographical features considered in spatial modelling, where m denotes the number of air quality monitoring stations. Given that K is a non-temporal variable and need to be injected in each time step when generating a temporal sequence inference. A pre-modelling step is supposed to be implemented at first: reshape K from the size of $(m\times m)$ to $(1,\ m\times m)$, then repeat it T times to obtain the size of $(T,\ m\times m)$ for fitting the input time lags T. Finally, a deep neural network (DNN) is employed to reduce the dimensions of K from $(T,\ m\times m)$ to the training size $(T,\ m)$. Thus, the final spatial modeling results, denoted by y^k, can be calculated as:

$$y^k = \mathscr{F}^N \cdots \mathscr{F}^n \cdots \mathscr{F}^1(W_k K + b_k) \tag{20}$$

where W_k is the trainable parameter, b_k is the bias term, and N denotes the layer numbers of the DNN model.

2.2.4. Fusion for temporal and spatial modeling

Another DNN model is utilized to combine temporal and spatial models. The prediction series of ozone concentration at future time $t^{'}$, denoted by $\mathscr{Y}_{t^{'}}$, is shown as:

$$\mathscr{Y}_{t^{'}} = \mathscr{C}^L \cdots \mathscr{C}^l \cdots \mathscr{C}^1\left(W_y\left[y_{t^{'}}^{ozone}; y_{t^{'}}^{aux}; y^k\right] + b_y\right) \tag{21}$$

where ';' refers the concatenation operation, L denotes the layer numbers of the DNN model, W_y and b_y are the trainable parameters and the bias terms, respectively. $y_{t^{'}}^{ozone}$ and $y_{t^{'}}^{aux}$ represent the temporal modelling results of ozone concentration and auxiliary features, including relevant air pollutant concentrations, meteorological features and weather prediction information.

2.3. Experiments details and network architecture

The experiment is deployed in a Ubuntu 16.04 Linux System with 128 GB RAM and Titan RTX Graphics Cards (24 GB GDDR5 VRAM). Python 3.6 with Pandas, Numpy, Tensorflow and Keras is applied to conduct data processing and modelling. We utilize Mean Squared Error (MSE) as the loss function and Root Mean Square Prop (RMSProp) as the optimizer for minimizing the errors between prediction and observation. In this study, 50% of the dataset are used for training, 25% are used for validation while the remaining 25% for testing. i.e., roughly speaking, the one-year data in 2017 are shuffled for training and validation, and the data from January to May, 2018 are used for testing. The hyper-parameters are determined by the performance of the model in the validation set. The EarlyStoping technique is implemented for effective learning, where a training process can be interrupted when the validation loss stops decreasing. In order to conduct rigorous statistical analysis, the nonparametric Kruskal-Wallis test (García et al., 2009a, 2009b; Acar and Sun., 2013) is employed to examine the statistical significance of model results and determine the differences among groups. For each numerical experiment, 30 independent runs are executed according to the requirements of the Kruskal-Wallis test.

Several statistical indices, including the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), the Pearson Correlation Coefficient (r) and the Coefficient of Determination (R^2), are employed to evaluate the performance of the proposed model, which are defined as:

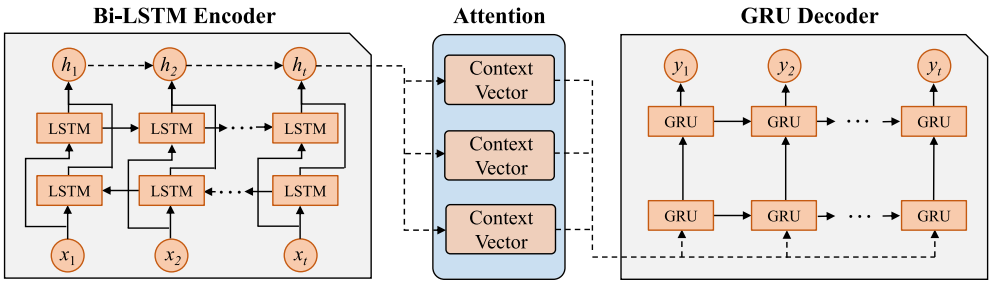


Fig. 4. The architecture of attention-based Seq2Seq model.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{22}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{23}$$

$$\frac{\mathrm{r} = \sum_{i=1}^{n}(O_i - \overline{O}_i)(P_i - \overline{P}_i)}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2}\sqrt{\sum_{i=1}^{n}(P_i - \overline{P}_i)^2}} \tag{24}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2} \tag{25}$$

where O_i refers to observed ozone concentrations, P_i denotes predicted ozone concentrations, n represents the number of samples, and $\overline{O}_i$, $\overline{P}_i$ are the mean of O_i, P_i respectively.

Table 1 summarizes the selection of units and layers in the model structures. The numbers of hidden units and layers respectively represent width and depth of the network, which are the direct representations of the learning capacity of a neural network. In the structure of temporal modeling, the structure of the encoder-decoder model is denoted as {number of units in each layer, number of encoder layers, number of decoder layers}. The prediction accuracy first increases and then drops, when the temporal structure changes from {256, 2, 2} to {512, 3, 3}. Meanwhile, the time cost rises continuously with the increase of hidden units and layers. Thus, we choose {256, 3, 3} as the temporal structure. In the structure of spatial modeling, the structure of the DNN model is denoted as {number of units in each layer, number of layers}. When the spatial structure changes from {256, 1} to {512, 2}, the running time continuously increases. However, the prediction accuracy first grows up and then drops after exceeding {256, 2}. In the structure of hybrid modelling, similar behaviors are observed for varying {256, 1} to {512, 2}. Thus, finally {256, 2} and {256, 2} are chosen as the spatial structure and the hybrid structure both considering prediction accuracy and time cost.

3. Results and discussion

In this section, the superiority of incorporating the spatiotemporal correlation is first quantitatively explored. Then, we evaluate the prediction performance of ground-level ozone concentration. Finally, the spatiotemporal characteristics of both observed and predicted ozone concentrations are estimated and discussed for further insights.

3.1. Quantification of spatiotemporal correlation in the monitoring network

As indicated in previous studies, spatiotemporal correlations exist in the measurements of the stations in the monitoring network (Wen et al., 2019). The Pearson's correlation coefficient analysis demonstrates that there are linear relationships among the observed values in monitoring stations (Li et al., 2017; He et al., 2018). This result is one of the preconditions that the spatial information benefits to the prediction accuracy of models. However, the improvement of prediction performance driven by integrating spatiotemporal correlations is still not clearly reported.

In this section, we implement a comparison between the regional pattern and the local pattern to quantitatively evaluate the importance of spatiotemporal correlation. The regional pattern considers the spatial information and the relevant information of neighboring stations, achieving by training model with the entire monitoring network and generating predictions for all monitoring stations in a lump. The local pattern is lack of the spatiotemporal correlation of the monitoring stations, achieving by separately building a model for each station and conducting prediction one by one. To ensure fairness, we develop two HSA-Net models with roughly the same input features and training parameters, but one is modeling in the regional pattern and the other is in the local pattern. For the regional pattern, the data of whole monitoring stations are integrated for training, validation and testing, and only one model exists, which has a mapping from 35-station input to 35-station output. For the local pattern, the data of each monitoring station is utilized for training, validating and testing itself independently, and each monitoring station has its own model, which has a mapping from 1-station input to 1-station output.

Fig. 5 and Fig. 6 give quantitative comparisons of the regional pattern and the local pattern, from the temporal and the spatial perspectives. A 24-h prediction performance is evaluated in the test set, and 240-h (10 days) time periods in January, 2018 and May, 2018 are selected to clearly understand the difference of prediction performances between the regional pattern and the local pattern. Thus, we conduct 24-h predictions first, and then concatenate them to 240-h long time series for further discussion. As shown in Fig. 5 (a) and (b), the observed ozone concentrations present a significant

Table 1
The arithmetic mean (standard deviation) of structure selection in the HSA-Net model.

Case	Baseline	Units, Layers	MAE (μg/m^3)	RMSE (μg/m^3)	r	Running Time (s)
Temporal Structure (encoder-decoder) Evaluation	Spatial Structure {256, 2}, Hybrid Structure {256, 2}	256, 2, 2	17.06 (0.87)	23.50 (1.19)	0.80 (0.01)	1965.9 (361.02)
		512, 2, 2	16.92 (0.91)	23.30 (1.27)	0.80 (0.01)	2161.99 (357.07)
		256, 3, 3	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
		512, 3, 3	16.58 (0.83)	22.81 (1.20)	0.81 (0.01)	3105.21 (499.31)
p-value			0.00**	0.00**	0.00**	0.00**
Spatial Structure (DNN) Evaluation	Temporal Structure {256, 3, 3}, Hybrid Structure {256, 2}	256, 1	17.20 (1.55)	23.64 (2.03)	0.80 (0.01)	1956.3 (136.36)
		512, 1	16.63 (0.87)	22.85 (1.18)	0.81 (0.01)	1981.54 (115.14)
		256, 2	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
		512, 2	16.87 (1.13)	23.16 (1.54)	0.81 (0.01)	2281.54 (135.14)
p-value			0.00**	0.00**	0.00**	0.00**
Hybrid Structure (DNN) Evaluation	Temporal Structure {256, 3, 3}, Spatial Structure {256, 2}	256, 1	17.38 (1.62)	23.86 (1.93)	0.80 (0.01)	1826.17 (114.2)
		512, 1	16.66 (1.06)	22.97 (1.45)	0.81 (0.01)	1931.87 (117.27)
		256, 2	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
		512, 2	17.01 (1.45)	23.29 (1.69)	0.81 (0.01)	2346.73 (143.7)
p-value			0.00**	0.00**	0.00**	0.00**

Kruskal-Wallis test, *: $p < 0.05$, **: $p < 0.01$.

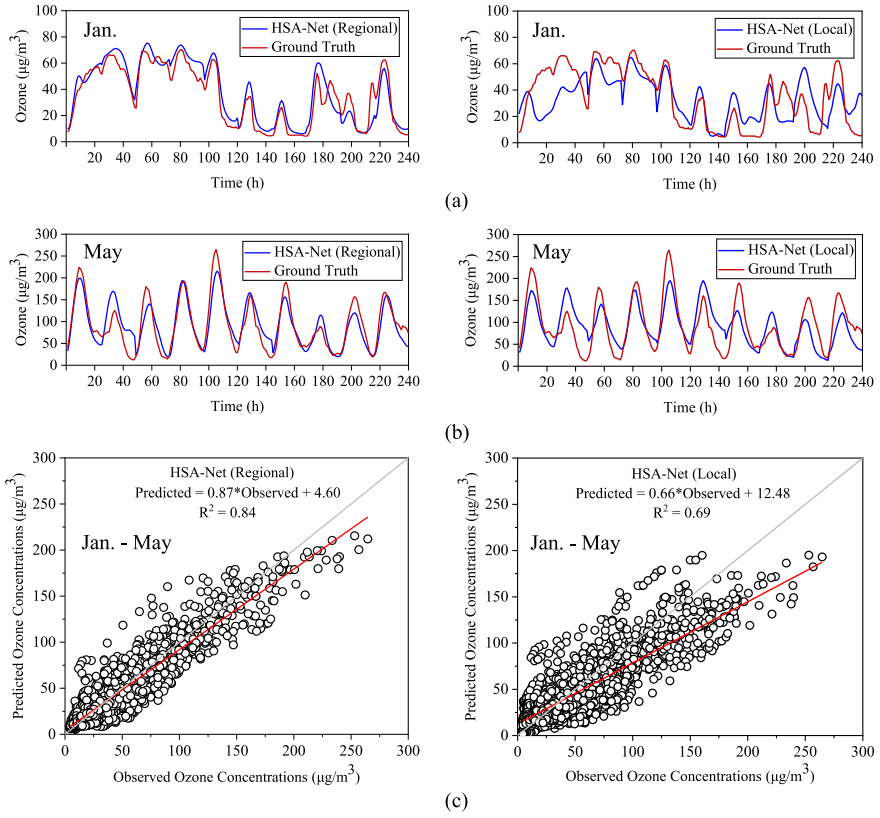


Fig. 5. Observed and predicted ozone concentrations of the regional pattern and the local pattern in the test dataset: (a) time series in January, (b) time series in May, (c) scatter plots from January to May.

increase from January to May in the test dataset. The forecasting results indicate that the regional model outperforms the local model both in the months with lower ozone concentration (January) and higher ozone concentration (May). In Fig. 5 (c), the R^2 values between predictions and observations demonstrate that 84% and 69% of the explained variance are captured by the regional pattern and the local pattern, respectively. The scatter plots also demonstrate that predictions are more consistent with the observations in the regional pattern. In Fig. 6 (a), the values of RMSE and r indicate that the regional model has dominated superior in prediction accuracy for most of the monitoring stations, except the station 12. In Fig. 6 (b), the regional model wins the comparisons in terms of the indices of RMSE and r for the four categories of monitoring stations. The results demonstrate that the adoption of spatiotemporal correlation can notably improve the prediction performance.

3.2. Performance evaluation of ground-level ozone predictions

In this section, the prediction performances of the HSA-Net model are carefully estimated. we first compare the proposed model with baseline methods, then evaluate the ability of multi-scale predictions, and finally examine the effects of auxiliary input variables.

3.2.1. Multi-scale predictions

This experiment evaluates the predictive ability of the HSA-Net model when the predictions of different lengths are required. In order to confirm the key historical period that significantly affects the prediction of ozone concentration, the influence of different input time lags is examined. We estimate the accuracy of predictions when the historical time lags are set as 24, 48, 120 and 168 h, respectively. In Table 2, when the input time lags increase from 24 h (1 day) to 168 h (7 days), prediction accuracy first drops and then grows up. However, the running time notably increases from 2026.57 s to 8518.81 s. This result indicates that the time lags corresponding to the diurnal variation period (24 h) of ozone concentrations (Su et al., 2017) and the weekend effects (168 h) of ozone concentrations caused by traffic activities (Gao, 2007) can achieve a better prediction performance, but longer input data result in longer computing time.

The multi-scale prediction performance is assessed under the scenarios that the lengths of prediction are set as 6, 12, 24, and 48 h, respectively. In Table 3, the result indicates that the prediction ability drops continuously from 17.75 to 23.23 in RMSE when the lengths of prediction increase from 6 h to 48 h. Meanwhile, the time cost rises notably from 692.45 s to 3498.88 s. The result demonstrates that longer-term predictions are more difficult and consume much more computing time.

3.2.2. Comparison of models

Two comparisons under the 24-h predictions are separately conducted for further performance estimation of the HSA-Net model in ozone predictions. The first comparison is implemented between the HSA-Net model and conventional models within the hybrid architecture. The second one is conducted between the HSA-Net model and the models without the hybrid architecture (e.g. Seq2Seq, Liu et al., 2019; LSTM, Li et al., 2017; GRU, Athira et al., 2018). All the results are presented in Table 4.

In the first comparison, the HSA-Net model has the optimal performance (22.08 in RMSE), and the hybrid Seq2Seq model and the hybrid Bi-LSTM model follow behind closely. There is an

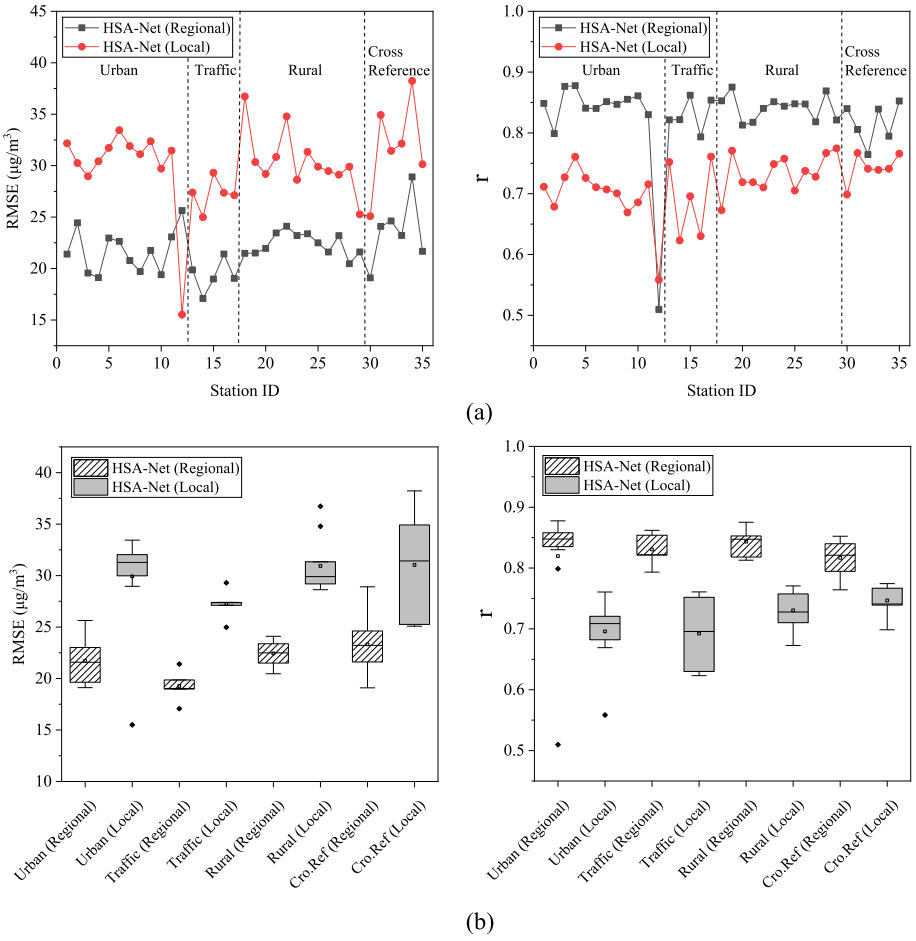


Fig. 6. Statistics indices comparisons between the regional pattern and the local pattern: (a) for all the 35 monitoring stations, (b) for urban, traffic, rural and cross reference monitoring stations.

Table 2
The arithmetic mean (standard deviation) of the effect of input time lags.

Time lags	MAE (μg/m^3)	RMSE (μg/m^3)	r	Running Time (s)
24 h (1 day)	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
48 h (2 days)	17.41 (1.66)	23.61 (1.93)	0.80 (0.01)	4006.14 (386.96)
120 h (5 days)	19.67 (2.27)	25.72 (3.03)	0.79 (0.01)	6879.66 (1010.11)
168 h (7 days)	17.53 (2.30)	22.79 (2.85)	0.81 (0.01)	8518.81 (2187.48)
p-value	0.00**	0.00**	0.00**	0.00**

Kruskal-Wallis test, *: $p < 0.05$, **: $p < 0.01$.

Table 3
The arithmetic mean (standard deviation) of multi-scale ground-level ozone predictions.

Predictions	MAE (μg/m^3)	RMSE (μg/m^3)	r	Running Time (s)
6-h prediction	13.21 (1.44)	17.75 (1.70)	0.88 (0.01)	692.45 (47.7)
12-h prediction	15.36 (1.53)	20.96 (1.74)	0.85 (0.01)	1128.6 (54.89)
24-h prediction	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
48-h prediction	16.99 (1.55)	23.23 (1.79)	0.79 (0.01)	3498.88 (625.08)
p-value	0.00**	0.00**	0.00**	0.00**

Kruskal-Wallis test, *: $p < 0.05$, **: $p < 0.01$.

obvious gap (about 4–5 in RMSE) between the RNN units (hybrid LSTM and hybrid GRU) and the Seq2Seq models (hybrid attention-based and hybrid Seq2Seq). However, the hybrid LSTM and hybrid GRU have the least running time (1643.22 s and 1517.23 s), compared with that of the HSA-Net model (2026.57 s). This result indicates that the attention mechanism can assist the Seq2Seq models to achieve better prediction performance, and also result in heavier computing workload. In the second comparison, the prediction abilities decrease significantly after removing the hybrid architecture from these models. The attention-based Seq2Seq model leads ahead (27.51 in RMSE), and the Seq2Seq model and the bidirectional LSTM exhibit slight inferior performance. Then a continuous drop (from 29.02 to 31.23 in RMSE) is observed when LSTM and GRU are employed. The result of above comparisons shows that the attention-based Seq2Seq models without the hybrid architecture can only achieve similar scores to the Hybrid LSTM, which suggests a significant improvement brought by the proposed hybrid architecture. However, the computing time in the second comparison (1371.43 s–1630.37 s) is significantly shorter than that in the first comparison (1517.23 s–2026.57 s). In sum, the proposed HSA-Net model outperforms these baseline models, but has longer running time.

3.2.3. Evaluation for the auxiliary data

This experiment aims to determine the optimal auxiliary data that contributes most to the improvement of the model predictions in accuracy. We remove temporal and spatial variables one by one and measure the change of prediction accuracy. Given that the prediction performance is strongly influenced by the diurnal variation period of ozone, this experiment is conducted under a 24-h prediction case. Thus, in terms of time steps, only the weather

Table 4
The arithmetic mean (standard deviation) of model comparison between the HSA-Net model and baseline models.

Models	MAE ($\mu g/m^3$)	RMSE ($\mu g/m^3$)	r	Running Time (s)
Hybrid Attention-based Seq2Seq	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
Hybrid Seq2Seq	17.10 (0.87)	23.44 (1.21)	0.81 (0.01)	1595.39 (390.43)
Hybrid Bidirectional LSTM	17.83 (1.71)	24.17 (2.25)	0.80 (0.01)	2168.42 (410.03)
Hybrid LSTM	20.43 (1.80)	27.56 (2.35)	0.79 (0.01)	1643.22 (327.99)
Hybrid GRU	20.83 (2.45)	27.61 (2.96)	0.79 (0.01)	1517.23 (330.28)
Attention-based Seq2Seq	20.41 (1.79)	27.51 (2.28)	0.78 (0.01)	1630.37 (470.59)
Seq2Seq	21.30 (1.84)	28.72 (2.34)	0.78 (0.01)	1469.53 (242.92)
Bidirectional LSTM	21.74 (2.18)	29.02 (2.51)	0.78 (0.01)	1503.89 (291.67)
LSTM	24.18 (2.97)	31.73 (3.52)	0.75 (0.05)	1375.62 (199.98)
GRU	23.97 (5.08)	31.23 (5.37)	0.77 (0.02)	1371.43 (200.5)
p-value	0.00**	0.00**	0.00**	0.00**

Kruskal-Wallis test, *: $p < 0.05$, **: $p < 0.01$.

forecast variable (WFC) group is set as t+24, and the other variable groups are set as t-24.

In Table 5, the predictive performance decreased significantly when we remove WFC and the geographic variable groups (GIS). Especially for WFC, there is an obvious gap (from 22.08 to 27.74 in RMSE) between before and after eliminating this variable group. This result demonstrates that the weather forecasting information and the spatial correlations have a strong influence on the ozone prediction. However, the predictive performances only exhibit light change (28.45 and 28.64 in RMSE respectively), when the meteorological variable group (MEO) or the air quality index group (AQI) are removed from the model input. But there is a drop (from 28.15 to 30.96 in RMSE) when we both remove MEO and AQI. It indicates that the effective information provided by these two relevant series sets is almost the same for the HSA-Net model, due to the slight multi-collinearity between meteorological features and air quality features (Li et al., 2017). But air quality and meteorological data are still important for the prediction of ground-level ozone concentration (Wen et al., 2019; Bai et al., 2019). The computing time roughly presents a continuous drop when variables groups are removed one by one. This result indicates that auxiliary features improve model prediction performance and also result in longer computing time.

3.3. Estimation of regional ozone predictions in the monitoring network

In order to fully discuss the characteristics of the proposed model, an overall estimation is implemented in this section. We conduct 24-h predictions by the HSA-Net model from January, 2018 to May, 2018 in the test dataset, and compare the predictions with ground truth in the same period. 240-hour (10 days) time periods in January, 2018 and May, 2018 are selected to clearly understand the detailed difference between observations and predictions. The 24-h prediction results are concatenated to 240-h long time series and divided into four categories for further discussions.

As shown in Fig. 7 and Fig. 8, the characteristics of ozone time series and prediction performance are firstly discussed. In Fig. 7 (a) and (c), the ozone concentrations in May are generally much higher than in January, and exhibit stronger diurnal variations. It indicates that the photochemical production of ozone in May is much stronger than that in Januray as the solar radiation increases. In both January and May, the ozone concentrations measured in traffic monitoring stations are lightly lower than that in other types of monitoring stations. This may be caused by the fact that ozone molecules could be rapidly titrated by traffic-emitted nitric oxide (NO) (Cheng et al., 2019). Additionally, the proposed model exhibits excellent predictions to match the curve of the observations in all four types of monitoring stations. In Fig. 7 (b), (d) and (e), the values of R^2 obtained by the present model in May are all slightly lower than in January. The result demonstrates that the proposed model exhibits better prediction performance in the months with lower ozone concentrations. Moreover, the values of R^2 in cross reference monitoring stations performance the worst (0.80 in January, 0.75 in May and 0.78 in the whole test dataset), and these values in urban, rural, traffic monitoring stations are higher and closer (0.81–0.85 in January, 0.78–0.80 in May and 0.80–0.84 in the whole test dataset). The above results demonstrate that the HSA-Net model can achieve excellent performance in the months with different ozone concentrations and in all four types of monitoring stations. In Fig. 8, we calculate the RMSE of four groups of multi-scale predictions (6, 12, 24 and 48 h) in the HSA-Net models, and then accumulated these RMSE in each station for further analysis. The results demonstrate that the traffic monitoring stations have the optimal prediction accuracy, while the cross reference monitoring stations accumulated the most errors in RMSE.

Combining the results in Figs. 7 and 8, the traffic monitoring stations have notably lower ozone concentrations, the best performance in RMSE and a good score of R^2. However, the ozone concentrations in cross reference monitoring stations are quite higher, and the values of RMSE and R^2 in these stations are both the worst. This is probably caused by the fact that the traffic pollution

Table 5
The arithmetic mean (standard deviation) of the effect of input variables in ozone concentration predictions.

Input variable groups	MAE ($\mu g/m^3$)	RMSE ($\mu g/m^3$)	r	Running Time (s)
O_3 + AQI + MEO + WFC + GIS	16.11 (1.05)	22.08 (1.29)	0.82 (0.01)	2026.57 (124.86)
O_3 + AQI + MEO + WFC	17.35 (1.19)	23.85 (1.70)	0.81 (0.01)	1919.32 (170.84)
O_3 + AQI + MEO + GIS	20.80 (1.35)	27.74 (1.83)	0.73 (0.02)	1732.25 (156.11)
O_3 + AQI + MEO	21.23 (1.57)	28.15 (1.94)	0.73 (0.01)	1682.61 (224.94)
O_3 + AQI	21.58 (1.01)	28.45 (1.29)	0.73 (0.01)	1269.44 (430.16)
O_3 + MEO	21.45 (1.19)	28.64 (1.58)	0.72 (0.01)	1460.08 (495.86)
O_3	22.83 (0.83)	30.96 (1.29)	0.72 (0.01)	1178.34 (496.38)
p-value	0.00**	0.00**	0.00**	0.00**

Kruskal-Wallis test, *: $p < 0.05$, **: $p < 0.01$.

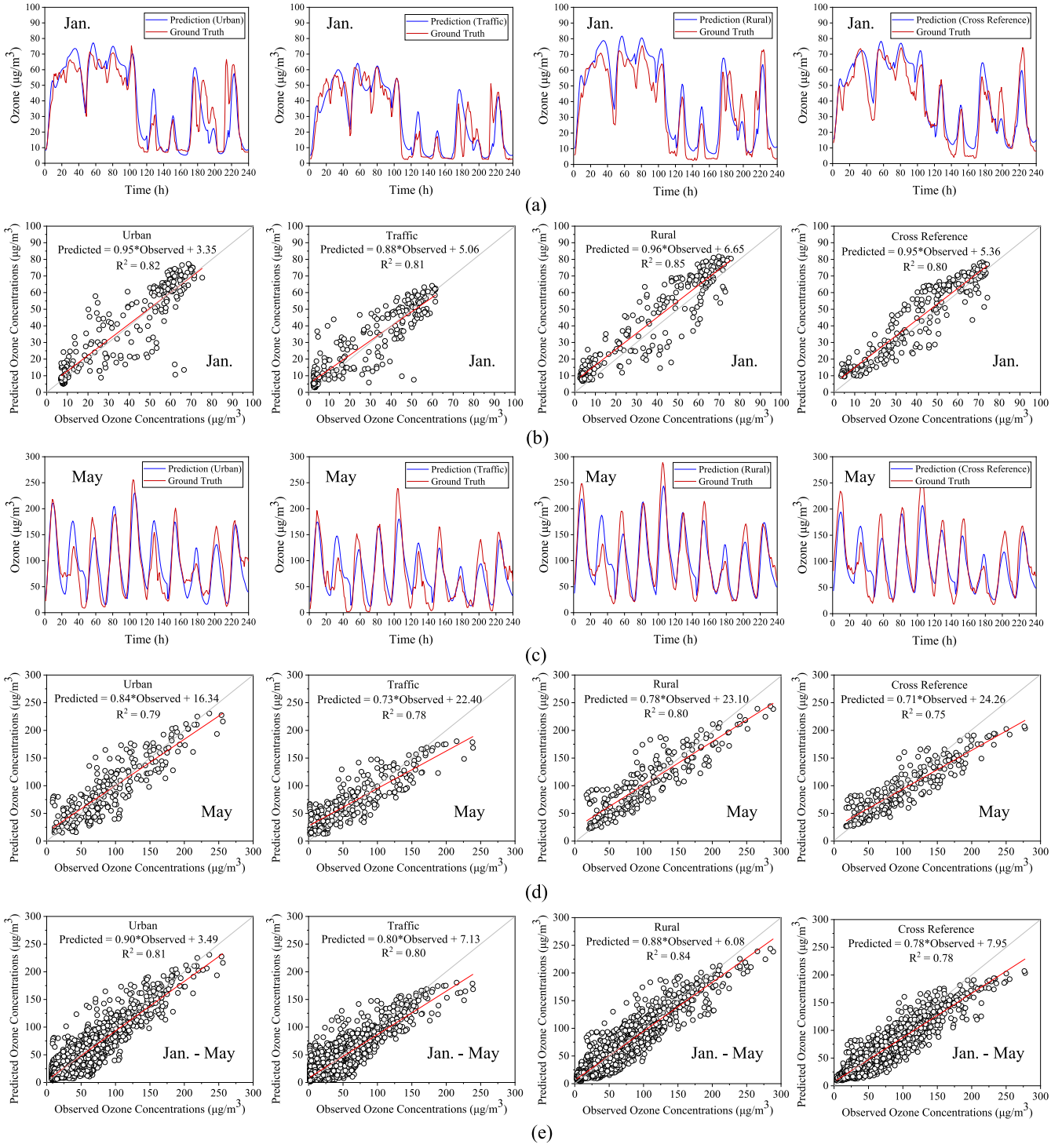


Fig. 7. Observed and predicted ozone concentrations of the HSA-Net model in the urban, traffic, rural and cross reference monitoring stations: (a) time series in January, (b) scatter plots in January, (c) time series in May, (d) scatter plots in May, (e) scatter plots from January to May.

monitoring stations are located in the downtown area of Beijing, where the ground-level ozone concentrations present much stronger diurnal variations due to the impacts of traffic emissions (Gao, 2007). This phenomenon benefits the deep learning models to learn the pattern of ozone generating and dispersing processes, and to achieve better performance. Cross reference monitoring stations are located in remote areas, and the diurnal variation patterns of ozone concentrations at these monitoring stations are affected by local photochemical production and transport of ozone from the downtown Beijing (Cheng et al., 2019). Thus, the special temporal patterns in such stations are quite challenging to predict.

Fig. 9 presents the spatial distribution of mean daily ozone concentrations and mean 24-h prediction performance in January, 2018, May, 2018 and the whole test dataset. The observed ozone

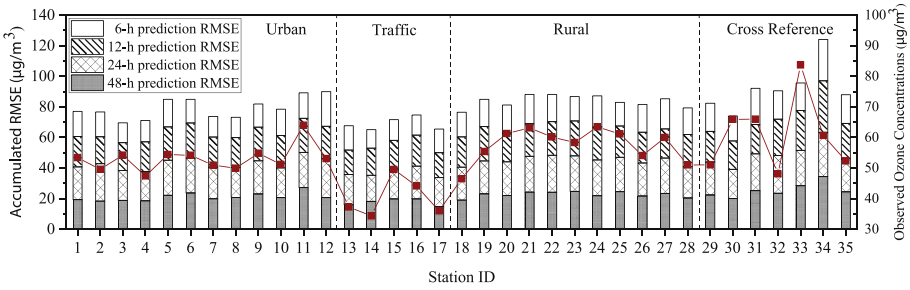


Fig. 8. Accumulative RMSE (μg/m^3) of the HSA-Net prediction (bars) and observed mean ozone concentrations (red line) for the 35 monitoring stations in the test dataset.

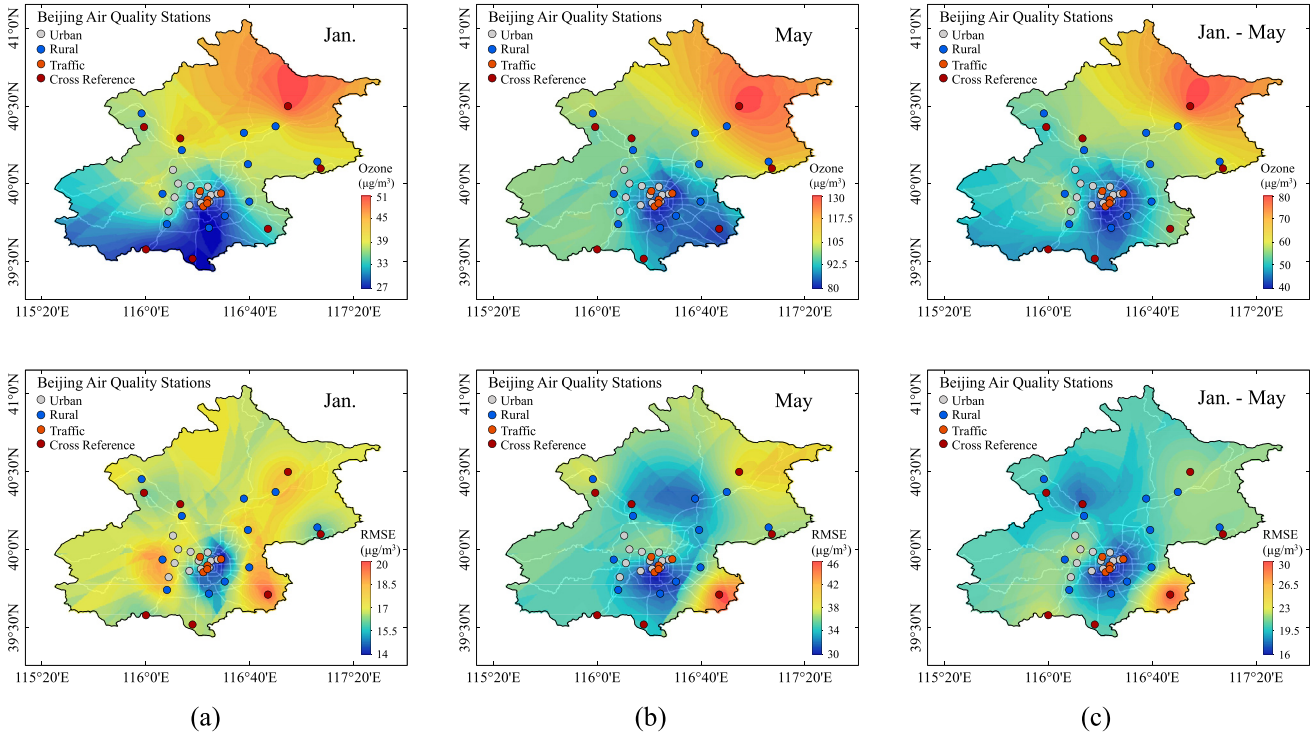


Fig. 9. Spatial distributions of observed mean daily ozone concentrations and mean RMSE (μg/m^3) of 24-h predicted ozone concentrations in the HSA-Net model for the 35 monitoring stations: (a) in January, (b) in May, (c) from January to May.

concentrations grow up notably from January to May, and are relatively lower over the urban areas. The spatial distribution patterns of ozone over Beijing are predominantly determined by meteorological and orographic factors (Cheng et al., 2019). Southeast, southwest and south winds prevailed over the period of January through May, 2018. The prevailing winds could blow ozone produced over urban areas to downwind areas in the north of Beijing. In addition, mountains are located to the north and northwest of Beijing, leading to evident accumulation of ozone to the north of Beijing. The RMSE of the predicted ozone concentrations also presents a significant increase from January to May. Additionally, the values of RMSE over the central urban area keep lower than the remote areas, and even show a more stable distribution in May. This result demonstrates that the prediction performance in strong spatial-correlated stations (such as traffic stations) is better than that in scattered stations (such as cross reference stations).

4. Conclusions

In this study, we propose a novel hybrid sequence to sequence model embedded with attention mechanism (HSA-Net) to predict the ground-level ozone concentration, fully considering the spatiotemporal correlations in the air quality monitoring network. The Attention-based Seq2Seq model and the deep neural network are both employed to simultaneously capture the spatiotemporal dependencies. In order to improve prediction performance, the neighboring factors and the spatial information in the air quality monitoring network are adaptively incorporated and learned in the model. The auxiliary air pollution and meteorological data are also integrated, which are related to the change of ground-level ozone concentration but were neglected in previous studies. The experiment results demonstrate that the proposed model can effectively predict the regional ground-level ozone concentration and outperform the competitors in terms of accuracy. Several meaningful findings are summarized as follows:

(1) Considering the spatiotemporal correlations in the air quality monitoring network presents enormous advantages for regional ozone prediction.
(2) The supplement of auxiliary data can notably improve the prediction ability.

(3) Prediction performance improves when time lags fit the day-of-week effects or diurnal variation period of ground-level ozone.
(4) Monitoring stations in urban areas present better prediction accuracy than stations in remote areas.
(5) The proposed model exhibits excellent performance in all monitoring station categories of Beijing and months with different ozone concentrations.

In terms of implications, the HSA-Net model is not limited to be applied in ozone prediction and is readily applicable to general cleaner production research, such as the forecast of carbon emissions and renewable energy (e.g. solar energy). In addition, this model could also assist environmental researchers to further improve the technologies of carbon capture and storage, and serve as tools for environmental policymakers to implement related policies. One limitation of this research is that the dataset only contains 17-month temporal series and 35-station spatial areas. Longer time spans and broader target regions can assist us to obtain more accurate prediction results and deeper insights. Future studies could consider the implications of the proposed model towards the prediction of carbon emissions and renewable energy, and the performance comparison among the HSA-Net model and widely-used machine learning methods in these topics.

Author contributions

Hong-Wei Wang designed and prepared the manuscript with contributions from all coauthors. **Xiao-Bing Li** designed experiments, analysed data and co-wrote the paper; **Dongsheng Wang** and **Junhao Zhao** performed data preprocessing; **Hong-di He** and **Zhong-Ren Peng** co-wrote the paper.

Declaration of competing Interest

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

Acknowledgements

This research is funded by the National Planning Office of Philosophy and Social Science (No. 16ZDA048), the National Natural Science Foundation of China (No. 11672176) and the China Scholarship Council (CSC) (No. 201506230117). The authors also thank Dr. Junchi Yan for his kind help on the model development.

References

Acar, E.F., Sun, L., 2013. A generalized kruskal-wallis test incorporating group uncertainty with application to genetic association studies. Biometrics 69 (2), 427–435.
Acheampong, A.O., Boateng, E.B., 2019. Modelling carbon emission intensity: application of artificial neural network. J. Clean. Prod. 225, 833–856.
Athira, V., Geetha, P., Vinayakumar, R., et al., 2018. DeepAirNet: applying recurrent networks for air quality prediction. Procedia Computer Science 132, 1394–1403.
Awang, N.R., Ramli, N.A., Yahaya, A.S., et al., 2015. Multivariate methods to predict ground level ozone during daytime, nighttime, and critical conversion time in urban areas. Atmos.Pollut. Res. 6 (5), 726–734.
Bahdanau, D., Cho, K., Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd ICLR arXiv preprint arXiv:1409.0473.
Bai, Y., Li, Y., Zeng, B., et al., 2019. Hourly PM2.5 concentration forecast using stacked autoencoder model with emphasis on seasonality. J. Clean. Prod. 224, 739–750.
Cai, C., Zhang, X., Wang, K., et al., 2016. Incorporation of new particle formation and early growth treatments into WRF/Chem: model improvement, evaluation, and impacts of anthropogenic aerosols over East Asia. Atmos. Environ. 124, 262–284.
Cheng, N., Li, R., Xu, C., et al., 2019. Ground ozone variations at an urban and a rural station in Beijing from 2006 to 2017: trend, meteorological influences and formation regimes. J. Clean. Prod. 235, 11–20.
Cho, K., Van Merriënboer, B., Gulcehre, C., et al., 2014. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation arXiv preprint arXiv:1406.1078.
Fasihi, M., Efimova, O., Breyer, C., 2019. Techno-economic assessment of CO2 direct air capture plants. J. Clean. Prod. 224, 957–980.
Feng, R., Zheng, H.J., Gao, H., et al., 2019. Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou, China. J. Clean. Prod. 231, 1005–1015.
Gao, H.O., 2007. Day of week effects on diurnal ozone/NOx cycles and transportation emissions in Southern California. Transp. Res. D Transp. Environ. 12 (4), 292–305.
Gao, M., Yin, L., Ning, J., 2018. Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. Atmos. Environ. 184, 129–139.
García, S., Fernández, A., Luengo, J., et al., 2009a. A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Computing 13 (10), 959.
García, S., Molina, D., Lozano, M., et al., 2009b. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. J. Heuristics 15 (6), 617.
Gradišar, D., Grašič, B., Božnar, M.Z., et al., 2016. Improving of local ozone forecasting by integrated models. Environ. Sci. Pollut. Control Ser. 23 (18), 18439–18450.
Graves, A., Jaitly, N., Mohamed, A.R., 2013. Hybrid speech recognition with deep bidirectional LSTM. In: Automatic Speech Recognition and Understanding (ASRU), 2013. IEEE Workshop, pp. 273–278.
Hao, S., Lee, D.H., Zhao, D., 2019. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. Transp. Res. C Emerg. Technol. 107, 287–300.
He, H.D., Li, M., Wang, W., et al., 2018. Prediction of PM2.5 concentration based on the similarity in air quality monitoring network. Build. Environ. 137, 11–17.
He, H.D., Qiao, Z., Pan, W., et al., 2017. Multiscale multifractal properties between ground-level ozone and its precursors in rural area in Hong Kong. J. Environ. Manag. 196, 270–277.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Huang, Y., Shen, L., Liu, H., 2019. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. J. Clean. Prod. 209, 415–423.
Li, X., Peng, L., Yao, X., et al., 2017. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. 231, 997–1004.
Li, L.L., Wen, S.Y., Tseng, M.L., et al., 2019. Renewable energy prediction: a novel short-term prediction model of photovoltaic output power. J. Clean. Prod. 228, 359–375.
Liu, B., Yan, S., Li, J., et al., 2019. A sequence-to-sequence air quality predictor based on the n-step recurrent prediction. IEEE Access 7, 43331–43345.
Ma, J., Ding, Y., Cheng, J.C.P., et al., 2019. Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. J. Clean. Prod. 118955 xxxx.
Mehdipour, V., Stevenson, D.S., Memarianfard, M., et al., 2018. Comparing different methods for statistical modeling of particulate matter in Tehran, Iran. Air Quality, Atmosphere & Health 11 (10), 1155–1165.
Mehdipour, V., Memarianfard, M., 2019. Ground-level O3 sensitivity analysis using support vector machine with radial basis function. Int. J. Environ. Sci. Technol. 16 (6), 2745–2754.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.
Sharma, S., Chatani, S., Mahtta, R., et al., 2016. Sensitivity analysis of ground level ozone in India using WRF-CMAQ models. Atmos. Environ. 131, 29–40.
Shi, X., Chen, Z., Wang, H., et al., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, pp. 802–810.
Srinivas, A.S., Somula, R., Govinda, K., et al., 2019. Predicting ozone layer concentration using machine learning techniques. In: Social Network Forensics. Cyber Security, and Machine Learning, pp. 83–92.
Su, W., Liu, C., Hu, Q., et al., 2017. Characterization of ozone in the lower troposphere during the 2016 G20 conference in Hangzhou. Sci. Rep. 7 (1), 17368.
Sumathi, D., Manivannan, S.S., 2019. Machine Learning-Based Algorithm for Channel Selection Utilizing Preemptive Resume Priority in Cognitive Radio Networks Validated by Ns-2. *Circuits, Systems, and Signal Processing*, pp. 1–21.
Wen, C., Liu, S., Yao, X., et al., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Sci. Total Environ. 654, 1091–1099.
Ye, B., Jiang, J., Zhou, Y., et al., 2019. Technical and economic analysis of amine-based carbon capture and sequestration at coal-fired power plants. J. Clean. Prod. 222, 476–487.
Yusuf, A., Giwa, A., Mohammed, E.O., et al., 2019. CO2 utilization from power plant: a comparative techno-economic assessment of soda ash production and scrubbing by monoethanolamine. J. Clean. Prod. 237, 117760.
Zhou, Y., Chang, F.J., Chang, L.C., et al., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J. Clean. Prod. 209, 134–145.