Estimation of Near-Ground Ozone With High Spatio-Temporal Resolution in the Yangtze River Delta Region of China Based on a Temporally Ensemble Model

Zhen Li, Heng Dong, Zili Zhang, Lan Luo, and Sicong He

***Abstract*—Recently, the near-ground ozone pollution has become an important factor restricting economic development and ecological environment protection. Due to the aging equipment of satellite sensors and the limitations of spatial resolution, the current approach utilizing satellite remote sensing observation faces challenges in effectively monitoring small-scale areas with sufficient data. Taking the near-ground ozone concentration as the research object, this article combined multiple classical machine learning (ML) methods based on tree models and developed a temporally ensemble model to achieve the estimation of near-surface ozone in the 1 km^2 area of the Yangtze River Delta region in China. In the ensemble model, the coefficient of determination (R^2) of the 10-fold cross-validation was 0.91, and the root-mean-square error was 9.21 μg/m^3. All evaluation indicators confirm that our approach was more accurate than some conventional ML models. The predicted spatial errors were evenly distributed, which indicated the superior spatial stationarity of the ensemble model. On the temporal scale, the ozone distribution predicted by the model agreed well with the results of ground-based meteorological station monitoring, both showing distinct seasonal trends. On the spatial scale, the model output reflected well the refined spatial variation of near-ground ozone at a small scale and captured the "medium-high-low" trend of near-ground ozone concentration in Shanghai and the trend of "low-medium" in Hangzhou, China. In contrast, the satellite observation data cannot well reflect the differences in details. In the future, this model will have good application potential in the refined monitoring of polluting gases across the country.**

Manuscript received 5 May 2023; revised 27 June 2023 and 13 July 2023; accepted 19 July 2023. Date of publication 1 August 2023; date of current version 3 August 2023. This work was supported in part by the Open Funding of Zhejiang Key Laboratory of Ecological and Environmental Big Data under Grant EED-2022-07, in part by the Fenghua Science and Technology Plan Project under Grant 202209204, and in part by the National Natural Science Foundation of China under Grant 52079101. *(Corresponding author: Heng Dong.)*

Zhen Li, Heng Dong, and Sicong He are with the School of Resources and Environment Engineering, Wuhan University of Technology, Wuhan 430070, China (e-mail: liyuchen980123@whut.edu.cn; simondong@whut.edu.cn; siconghe@whut.edu.cn).

Zili Zhang is with the Ecological Environment Monitoring Center of Zhejiang, Hangzhou 310012, China, and also with the Zhejiang Key Laboratory of Ecological Environment Monitoring, Early Warning and Quality Control Research, Hangzhou 310012, China (e-mail: xiahecia@gmail.com).

Lan Luo is with the Zhejiang Key Laboratory of Ecological and Environmental Big Data (2022P10005), Zhejiang Ecological and Environmental Monitoring Center, Hangzhou 310012, China (e-mail: luolan@zjemc.org.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2023.3298996, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2023.3298996

***Index Terms*—Air pollution, high spatio-temporal resolution, machine learning (ML), near-ground ozone, temporally ensemble model.**

I. Introduction

OZONE is a common atmospheric component [1], [2]. In the atmosphere, about 90% of ozone is distributed in the stratosphere, which can effectively weaken the direct solar ultraviolet radiation and is extremely important to the earth's ecological environment [3], [4], [5]. Ozone is also an oxidizing and reactive gaseous trace pollution gas, which can be used as a good tracer to cooperate with carbon emission monitoring to realize "carbon pollution from the same source" and complete the precise positioning of carbon sources [6]. As a photochemical product of volatile organic compounds (VOCs) and nitrogen oxides, ozone plays an important role in forming and driving photochemical smog and acid rain [7]. At the same time, ozone is also a harmful substance to the human body. Long-time exposure to high concentration of ozone can significantly increase the probability of several diseases such as asthma, respiratory infection, high blood pressure, and ischemic heart disease [8], [9], [10], [11]. According to the World Health Organization, the daily average ozone concentration that the human body can withstand should not exceed 100 mg/m^3. Therefore, refined monitoring of atmospheric ozone, especially the temporal and spatial distribution of near-ground ozone, can ensure the accurate implementation of "carbon pollution from the same source" and the effective implementation of air pollution evaluation, treatment, and traceability. It has important scientific research significance and practical value for protecting the physical and mental health of residents and winning the battle to defend the blue sky.

In order to monitor the instantaneous state of various components in the air, China had established air quality monitoring stations in every city. By the end of 2012, the initially formed ground monitoring network had been able to quickly and accurately obtain information on the state of the atmosphere near the ground [12], including pollutant gases such as nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3). However, China's ground monitoring network is not perfect, the spatial distribution of stations is uneven, and the air pollutant information it monitors can only represent a small range centered on the monitoring stations [13], [14], which means that it is difficult for ground stations to monitor air pollution conditions accurately and comprehensively. Remote sensing observations can obtain continuous atmospheric information at high frequency. In recent years, the quantitative inversion technology of air pollutants based on remote sensing has shown great application potential, and gradually becomes an important method of continuously monitoring the concentration of air pollutants in large areas [15], [16]. The huge amount of data obtained by remote sensing provides a new way for fine monitoring of regional ozone. At present, the mainstream atmospheric trace gas observation satellite sensors include the Ozone Monitoring Instrument (OMI) [17], the Global Ozone Monitoring Experiment-2 [18], and the Tropospheric Monitoring Instrument (TROPOMI) [19], and the observation system formed by these sensors can provide observation data from 1995 to the present. However, due to the hardware conditions of the remote sensing satellite itself and the limitations of external climate conditions, the existing ozone remote sensing retrieval products have a series of problems such as missing data, weak consistency between different sensors, and coarse spatial resolution. All these problems make it difficult to achieve fine monitoring of small areas such as urban areas by remote sensing observations alone [20], [21].

Although the current satellite sensors for monitoring atmospheric pollutants such as ozone have many shortcomings, satellite remote sensing technology has also provided new data for large-scale ozone monitoring research with its advantages of high efficiency, large-scale and dynamic acquisition of atmospheric information source. On this basis, many estimation methods for air pollutants have been developed [22], [23], [24]. Among them, machine learning (ML) models have been widely used in the estimation of the near-ground pollutants due to their ability to better deal with complex nonlinear relationships between variables, to better process and analyze large-scale data [25], [26]. As a classic algorithm in the ML model, the random forest (RF) model is gradually used in the estimation of air pollutants because of high-dimensional processing capability, big datasets, and strong antinoise ability [27], [28]. Compared with the traditional chemical transport model, the RF algorithm not only reduces the computational cost but also improves the computational accuracy [29]. Some scholars have explored the complex nonlinear relationship between near-ground ozone and explanatory variables using the extreme gradient boosting (XGB) model based on the boosting algorithm. The model outperforms other ML models in both site-based and sample-based cross-validation (CV) results [30], [31], [32]. In addition, some scholars have introduced the deep forest (DF) algorithm and the light gradient boosting (LGBM) algorithm into the estimation of ozone [33], [34], [35], [36]. As an upgraded version of the RF and XGB, DF and LGBM have avoided the computational pressure brought by a large amount of data and the overfit phenomenon caused by multifeature data [37], [38]. It can be seen that the ML model has a good prospect in estimating air pollutants. Some recent studies have built an ensemble model of "ML model + neural network" combined with satellite observations data, meteorological data, and human activity data [39], which has also been used in evaluating the near-ground air pollutants. The model avoids the problems of deviation and overfitting in different base models and obtains better estimation results. The ensemble multiple ML model method (FC-LsOA-KELM) successfully estimated the ozone exposure of 11 cities in the Fenwei Plain of China [40]. This model has built multiple machine learners, avoiding the deviation of prediction results caused by different algorithm principles of a single model, and achieved good results in monitoring fields such as nitrogen dioxide and ozone. At present, although ML models and ensemble learning models can effectively monitor the near-ground pollution gases, they can still be further explored and improved in terms of spatial resolution, time-series variation characteristic. Constructing a prediction model that takes into account the influence of time and obtains the results of near-ground ozone concentration distribution with high temporal and spatial resolution has become the focus of pollution gas estimation research at this stage.

According to the problems existing in the current research, this study combines concentration of ozone column data, reanalysis data, meteorological data, and other multisource data to propose a temporally ensemble model. Using this model, the near-ground ozone exposure information at 1 km^2 spatial resolution in the Yangtze River Delta region of China was successfully obtained, and the performance of four single-machine models and ensemble models were compared and analyzed by 10-fold CV. In addition, by comparing the model estimation results with interpolation results of satellite observation data on a monthly and seasonal scale, the advantages of the ensemble model in predicting the spatio-temporal distribution of the near-ground ozone were discussed. Finally, we selected two cities (Hangzhou and Shanghai in China) and analyzed the ability of the ensemble model to capture spatial heterogeneous characteristics at high spatial resolution. This work aims to develop a temporally ensemble model for near-ground ozone estimation with high spatio-temporal resolution, to explore the advantages of the model in predicting the spatio-temporal distribution of near-ground ozone, and to demonstrate its ability to capture fine-grained changes in ozone concentration in small regions. The high spatio-temporal resolution information obtained through this model will provide important data support for relevant Chinese departments in ozone monitoring and effective governance.

II. Materials and Methods

A. Data and Preprocessing

1) Distribution of Ground-Level Ozone Observation Stations: This study selected the Yangtze River Delta region of China as the research area (see Fig. 1), which included Shanghai, Jiangsu Province, Zhejiang Province, and Anhui Province (114.377°E-123.134°E, 26.167°N-35.527°N). The area covers an area of 40 000 km^2, with a total of 41 cities. The region is one of the most economically dynamic regions, with the most frequent industrial activities and the highest population density in China.

In this work, hourly ozone data from the China Environmental Monitoring Center in the Yangtze River Delta region from 2018

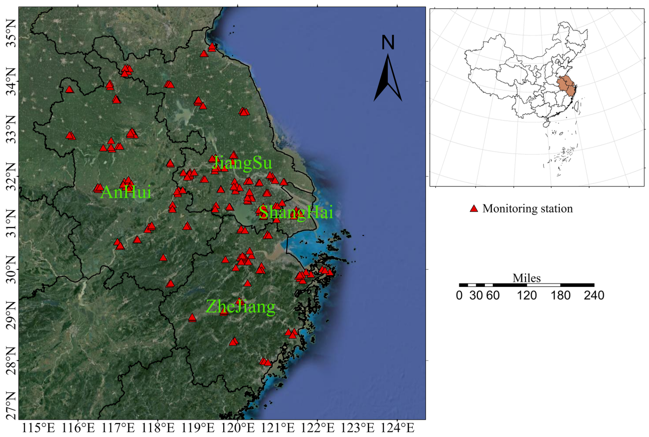

Fig. 1. Distribution map of ground air quality observation stations.

to 2020 (see Fig. 1) were collected. The Environmental Monitoring Center used ultraviolet photometers and equipment to monitor ozone generation to measure and calibrate near-ground ozone. The distribution of air quality monitoring stations in China does not cover every city (see Table SI). In fact, the layout of stations is affected by many factors, including city size, industrial development level, population density, and geographical location [41], [42]. As of 2021, there are 324 stations in this region, among which 138 in Jiangsu, 20 in Shanghai, 72 in Zhejiang, and 94 in Anhui. The layout of monitoring stations is strictly in accordance with the HJ664-2013 standard, which can represent the real situation of the near-ground ozone. In addition, in view of data omissions or outliers in the monitoring sites, this work eliminated the monitoring data of less than 15 h per day and obtained the daily average data of the sites after processing.

2) *TROPOMI O_3 Data:* The satellite ozone column concentration data used in this study were obtained from the Sentinel-5P TROPOMI O_3 Level 2 (S5P_L2_O3_TOT) released on NASA's official website with a spatial resolution of 7.0∗3.5 km^2 (5.5∗3.5 km^2 after August 6, 2019). S5P_L2_O3_TOT relies on a direct-fit algorithm (S5P_TO3_GODFIT) that uses nonlinear least squares inversion based on the difference between the model simulated radiation and the satellite-monitored radiation. This work obtained the tropospheric total vertical column ozone concentration from S5P_L2_O3_TOT, keeping data with quality control greater than 0.75 in the file to weaken the interference of the data caused by clouds and other problems.

3) Reanalysis Data: The atmospheric conditions, radiant heat, and ozone precursor information used in this work are derived from the ERA5 reanalysis data released by the European Center for Medium-Range Weather Forecasts (ECMWF). This dataset is based on the methods used by the Numerical Weather Prediction Center to optimally combine previously observed results with the most recent observations every 12 h to produce the best estimate of the state of the atmosphere and is the fifth generation of ECMWF reanalysis of global climate and weather over the past 40–70 years. This study selected 20 indicators from the ERA5 dataset (https://cds.climate.copernicus.eu), including atmospheric conditions, radiant heat, and ozone precursors, which may have an impact on near-ground ozone concentration changes, but not all indicators were involved in model construction. In addition, all the reanalysis data have undergone quality control, removal of missing values, and handling of outlier such as –9999 before being used as the modeling dataset.

4) Meteorological Data: Recently, studies on the estimation of near-ground ozone have found that meteorological factors are one of the critical factors affecting near-ground ozone concentration and distribution. The temperature and humidity in the meteorology can promote and delay the photochemical reaction of ozone, and the wind speed can affect the transmission and diffusion of ozone and precursors, which in turn affects the concentration and distribution of ozone.

The meteorological data were obtained from the National Meteorological Science Data Center of China. The land surface temperature (LST), wind speed (WS), daily average air pressure (AP), air temperature (AT), relative humidity (HD), and cumulative rainfall of the meteorological stations in the study area from 2018 to 2020 were selected. To ensure consistent spatial resolution of research variables, we performed kriging interpolation in ArcGIS software to obtain meteorological raster data with 1 km^2 spatial resolution after removing null values, –9999, and values that do not match the variable in the original data.

5) Other Geographic Data: The formation and distribution of ozone are the result of many factors: human migration and industrial production generate large amounts of VOCs, which are important precursors to ozone formation, including oxygen-containing organic compounds, nitrogen-containing organic compounds, and sulfur-containing organic compounds; the distribution of different types of surface objects also affects the variation of near-surface ozone. Therefore, this study incorporated road network density, population density, elevation, and land cover data to characterize these impacts.

The land cover data were obtained from the Climate Change Initiative (CCI) of the European Space Agency, with a spatial resolution of 300 m and a temporal resolution of one year. CCI divides all ground objects into 23 categories (see Table SII), which will reduce the stability of the model [43], [44]. Considering the differences in the driving mechanism of ozone by different land object types, this study reclassifies the land use types into six types: forest land, grassland, urban land, water body, agricultural land, and bare land (see Table SIII) [45], [46]. The population gridded data was obtained from Worldpop, and the road network density data were obtained from the OpenStreetMap website (https://www.openstreetmap.org) after projection transformation and statistical section length operations.

The near-ground ozone concentration distribution is closely related to altitude, so the digital elevation model (DEM) is also taken into consideration. The DEM data are derived from the Shuttle Radar Topography Mission (SRTM) of the USA (https://earthexplorer.usgs.gov/). For a specific description of the data, see supplementary material (see Table SIV).

6) Variable Selection and Matching: ML models are widely used for near-surface pollutant estimation due to the good handling of complex nonlinear relationships between variables [43], [44]. However, too many explanatory variables sometimes not only increase the complexity of the model and reduce the computational speed but also most of the explanatory variables have multicollinearity among them, which will lead to great noise and make the model unstable, resulting in overfitting, thus affecting the accuracy of the model prediction [47]. Therefore, it is necessary to properly screen the explanatory variables to control the variables within a reasonable range. In this study, the Pearson correlation coefficient [48], [49] (see Fig. S1) was used to screen the explanatory variables, and eliminate variables with low correlation coefficients and those with similar physical significance. Some variables such as human activities have a low correlation coefficient with ozone concentration, but always affect the concentration and distribution of ozone. Therefore, parts of them are also included.

Finally, 19 several predictors were introduced into the estimation model, namely, boundary layer height (blh), mean sea level pressure (msl), surface latent heat flux (slhf), top net solar radiation (tsr), 100 m wind speed (v100), relative humidity (HD), road network density (road), population (pop), land use data (lc), satellite-observed ozone column concentration (O3_5p), elevation (DEM), LST, time (time), and satellite-observed formaldehyde column concentration (HCHO_5P).

This study designed a grid with a spatial resolution of 1 km∗1 km to standardize the data from different resolutions for the construction of modeling datasets. First, all variables were reprojected into a unified projected coordinate system, and for data above 1 km^2 resolution were sampled into the standard grid using the nearest neighbor resampling method [50], and for data below 1 km^2 resolution were scaled to the standard grid using inverse distance weight interpolation [51]. Finally, the grid data that fell to the ground ozone monitoring station were collected, and 209 852 valid data were obtained after removing outliers and missing values for model construction and validation.

B. Methodology

Since the variables that affect ozone formation have time dependence and a single model is unable to fully adapt to multidimensional data, this study proposed a temporally ensemble model based on a variety of conventional ML models, taking into account the temporal correlation between variables.

The RF algorithm is based on the bagging ensemble constructed by the decision tree as the base learner, and further introduces the selection of random attributes in the training process of the decision tree. It can not only effectively deal with multisource data but also avoid overfitting phenomenon [52]. In contrast, the extreme random forest (EXT) algorithm uses all samples on the selected sampling set as the training set of decision trees, and randomly selects a feature value to divide the decision tree after selecting the division feature, which is to a certain extent improve the generalization ability of the model [53]. As an implementation of the gradient boosting decision tree (GBDT) framework in the boosting algorithm, the XGB algorithm adds a regular term to the loss function to control the overall complexity of the model, thereby greatly reducing the overfitting problem [54]. The LGBM algorithm is also the implementation of the GBDT algorithm; compared with the presorted algorithm used by XGB, LGBM uses a histogram-based decision tree algorithm. This method will greatly reduce the calculation amount of segmentation gain, thereby reducing memory usage and improving computing speed and ensuring higher accuracy different algorithms and structures determine that models have different advantages when dealing with data of different dimensions [55]. This study introduces the above models to compare the differences among models.

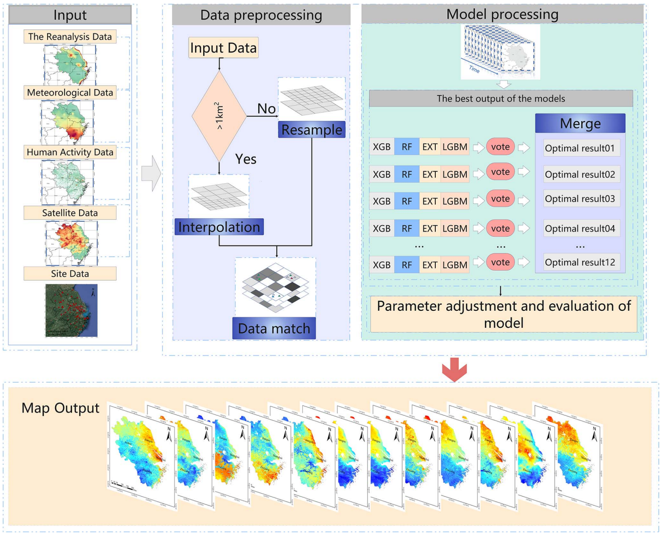
Input
The Reanalysis Data
Meteorological Data
Human Activity Data
Satellite Data
Site Data
Data preprocessing
Input Data
>1km^2
No
Resample
Yes
Interpolation
Data match
Model processing
Time
The best output of the models
Merge
XGB RF EXT LGBM vote Optimal result01
XGB RF EXT LGBM vote Optimal result02
XGB RF EXT LGBM vote Optimal result03
XGB RF EXT LGBM vote Optimal result04
...
XGB RF EXT LGBM vote Optimal result12
Parameter adjustment and evaluation of model
Map Output

Fig. 2. Framework of ensemble model.

A single model often considers the entire time series from an overall perspective to minimize the overall deviation. When facing long-term series and multisource data, these single models may not be able to capture the true characteristics of local time. Considering the temporal specificity of near-ground ozone, we constructed a temporally ensemble model framework, combined with a hard-voting ensemble approach [56], and developed an ensemble model based on classical ML models to estimate ground-level ozone concentration. First, we divided the modeling dataset into 12 monthly data according to the time, constructed XGB, RF, EXT, and LGBM models by month, and established the monthly mapping relationship between explanatory variables and near-ground ozone concentration (see (1) for details). In order to reduce the model deviation due to different algorithm principles, this study used Bayesian optimization to optimize the parameters of each model. The optimal hyperparameters for each model are shown in Table SV. Next, we employ a hard voting approach to comprehensively compare the estimation results of LGBM, XGB, RF, and EXT models for each monthly dataset using the coefficient of determination (R^2), the root-mean-square error (RMSE), and mean absolute error (MAE) metrics. Ultimately, the optimal model for each month is determined. Finally, the output results of the corresponding models for each month were ensemble to obtain the final ground-level ozone estimation result; the details of the model framework are shown in Fig. 2

$$Pozone_i = best\left(f_{ij}\left(blh, tsr, v100 \ldots O3\_5p\right)\right)\} \tag{1}$$

TABLE I
CV Results of the Five Regression Models on the Monthly Scale

		LGBM		XGB		RF		EXT		Ensemble	
Month	Counts	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
M01	N=13 599	0.8546	7.717	0.8299	8.266	0.7607	9.805	0.7585	9.848	0.8742	7.109
M02	N=12 363	0.8283	7.924	0.8036	8.475	0.7080	10.33	0.7295	9.945	0.8749	6.764
M03	N=13 497	0.7659	8.843	0.7317	9.466	0.6023	11.52	0.6335	11.063	0.8120	7.923
M04	N=12 939	8.8363	9.854	0.8092	10.638	0.7304	12.65	0.7418	12.376	0.8595	9.129
M05	N=17 671	0.8259	11.378	0.7969	12.289	0.7066	14.77	0.7107	14.667	0.8627	10.106
M06	N=19 709	0.8819	11.51	0.8616	12.46	0.7743	15.92	0.7730	15.961	0.9026	10.456
M07	N=20 801	0.8480	10.722	0.8172	11.758	0.7035	14.97	0.6876	15.37	0.8430	10.895
M08	N=20 364	0.8505	11.763	0.8137	13.132	0.7218	16.05	0.6885	16.981	0.8376	12.262
M09	N=19 574	0.8536	11.091	0.8283	12.011	0.7431	14.69	0.7449	14.641	0.8905	9.593
M10	N=20 278	0.8276	9.361	0.7841	10.478	0.6715	12.92	0.6752	12.851	0.8670	8.225
M11	N=19 833	0.8351	8.037	0.8074	8.686	0.7305	10.27	0.7198	10.475	0.8735	7.039
M12	N=18 998	0.7376	7.431	0.6983	7.966	0.6080	9.081	0.6036	9.132	0.7565	7.157

where $Pozone_i$ represents the optimal result of the model for each month, i represents the month,j represents the four ML models (XGB, RF, EXT, and LGBM), $best$ represents the optimal model corresponding to the month, f_{ij} represents the modeling dataset, and the $blh, tsr, v100 \ldots O3\_5p$, respectively, represent the boundary layer height, top net solar radiation, 100 m wind speed …satellite-observed ozone column concentration.

III. Results

A. Statistical Description

This study selected the top net solar radiation (tsr), surface latent heat flux (slhf), LST, satellite ozone column concentration (O3_5p), formaldehyde column concentration (HCHO_5p) mean sea level pressure (msl), boundary layer height (blh), and other 19 variables as the input of the model. The time resolution of these variables is daily (May 14, 2018–December 31, 2020), and the spatial resolution is 1 km^2. According to Fig. S2, different variables presented different spatial heterogeneity. The altitude in the study area behaved the highest with a value of 1869 m, and the lowest with 33 m, which was higher in the south and lower in the north. Meanwhile, the mean sea level and air pressure (msl) showed a trend of higher in the north and lower in the south, which was in line with the law of air pressure changing with altitude. In addition, O3_5p and HCHO_5P, v100, and msl showed similar distributions in space; the high value area of tsr was mainly located in aquatic regions, which reflected the real situation. In conclusion, the explanatory variables selected above could meet the modeling requirements and be used for estimating near-ground ozone.

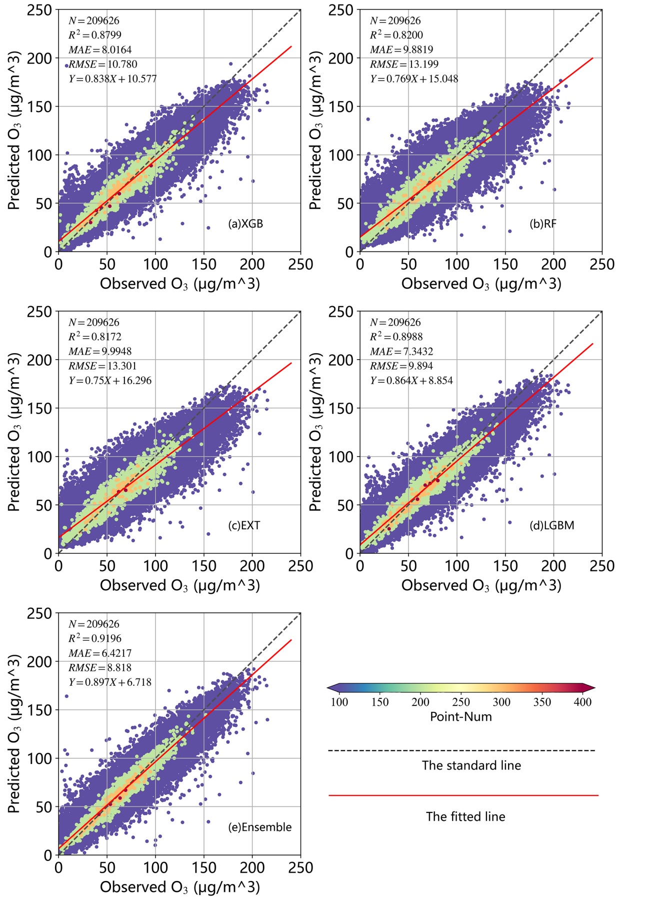

Fig. 3. 10-fold CV results of each model. (a) Validation results of XGB model. (b) Validation results of RF model. (c) Validation results of EXT model. (d) Validation results of LGBM model. (e) Validation results of ensemble model. The red solid line is the fitted line, and the black dotted line is the standard line.

B. Model Comparison

In this study, we selected four widely used single regression models, including XGB, RF, EXT, and LGBM for model comparison. The fitting results and CV accuracy of different models were shown in Fig. 3, and the sample size involved in the modeling was sufficient to support the credibility of the experiment. The R^2 values of XGB, RF, EXT, LGBM, and Ensemble reached 0.87, 0.82, 0.81, 0.89, and 0.91, respectively, the corresponding RMSE values were 10.78, 13.19, 13.30, 9.89, and 9.21 μg/m^3, respectively, and the corresponding MAE values were 8.01, 9.88, 9.99, 7.34, and 6.67, respectively. Overall, the five models underestimated the true concentration of near-ground ozone, but the slope of the predicted trend line of the ensemble model was closest to 1, which showed a value of 0.88. Compared with other models, the bias in the estimation had been corrected to a certain extent. In addition, we also compared the performance of five models in each month (see Table I). Among them, each model

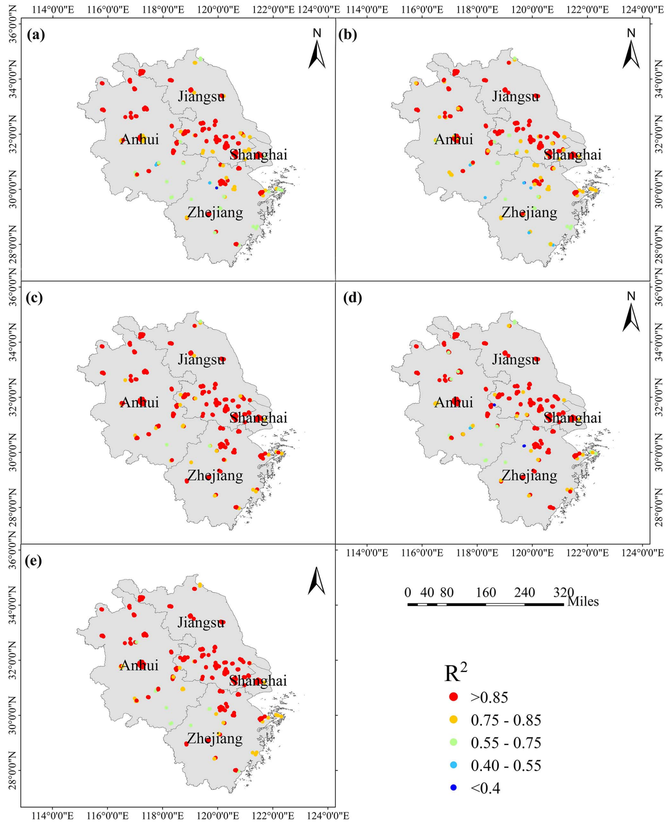

Fig. 4. Spatial distribution of R^2 of the ensemble model. (a) Distribution of the R^2 in spring. (b) Distribution of the R^2 in summer. (c) Distribution of the R^2 in autumn. (d) Distribution of the R^2 in winter. (e) Annual distribution of the R^2 for the ensemble model.

had the best fitting effect in June and the worst fitting effect in December. After the optimal results of each model, the ensemble model also behaved less fluctuation (absolute deviation: 0.028), second only to LGBM (absolute deviation: 0.026).

According to Fig. 4(e), 201 stations showed greater R^2 than 0.85, accounting for 83.05% of all observation stations. Most of these stations were located in urban areas, whose economy was developed and facilities were complete. Therefore, these areas could provide enough explanatory variable observation data, so that the model can fully capture the nonlinear relationship between explanatory variables and ozone concentration. Furthermore, the fitting result was the best in winter, and 76.98% of the stations had a fitting accuracy greater than 0.85 [see Fig. 4(a)–(d)]. The season with the worst fitting effect was summer, and only 149 stations showed a fitting coefficient greater than 0.85, accounting for 61.82% of all observed stations. The spatial distribution of RMSE (Fig. S3) and MAE (Fig. S4) was consistent with R^2. It could be seen that the ensemble model proposed had good adaptability in both spatial distribution and seasonal scale.

C. Sensitivity Analysis of Model Performance

The relative importance rankings of the independent variables of five regression models were shown in Fig. 5. Among the XGB, RF, and EXT models, tsr had the highest importance proportion,

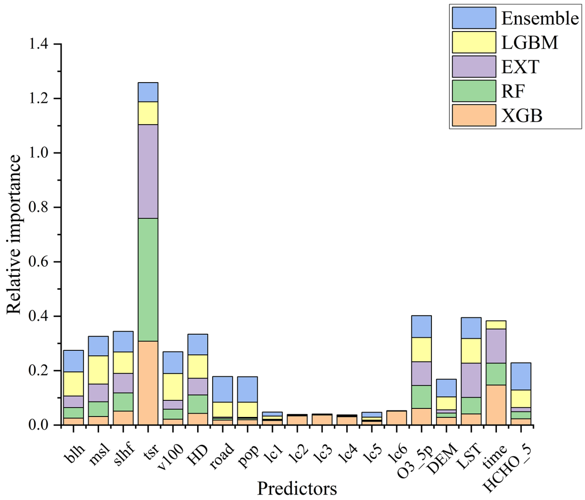

Fig. 5. Relative importance of explanatory variables in five regression models.

which were 30.8%, 45.1%, and 34.4%, respectively; among the LGBM models, msl had highest proportion, which was 10.3%; whereas in the Ensemble model, the importance of O3_5p was the highest at 10.3%, followed by tsr at 8.4%. In general, the data representing heat (slhf, tsr, and LST) were the most important factors affecting each model (XGB: 39.9%, RF: 57.8%, EXT: 54.1%, LGBM: 25.3%, Ensemble: 22.7%), the effects of human activity data (road, pop, and lc) were the least (XGB: 22.0%, RF: 1.6%, EXT: 1.7%, LGBM: 13.9%, Ensemble: 21.3%). In the ensemble model, except lc and time, the contribution of other explanatory variables was about 8%, which was relatively balanced.

The main reason why the relative importance of each explanatory factor in the ensemble model is different from other models is that RF, EXT, XGB, and LGBM only focus on the contribution of each factor in the overall data while ignoring the performance of explanatory variables in different time periods. The ensemble model took into account the temporal heterogeneity of the explanatory variables from the perspective of each month. For example, LST showed different seasonal distributions over time. This phenomenon of showing different numerical changes in different months was averaged in the ensemble model (see Fig. S5), which was why the relative importance of most explanatory variables showed a relatively balanced phenomenon.

D. Spatiotemporal Distribution Analysis

The distribution of ozone presented a regular distribution with each month (see Fig. 6). It mainly presented that the ozone concentration was high in the northeastern coastal area from January to March, the high-value area gradually moved from the northeast to the northwest during March to September, and ozone with high values was mainly distributed in the southern area from September to December. The monthly average concentration of ozone showed a “double peak” trend (see Fig. 6). From January to June, the ozone concentration gradually rose to a peak of 88.52 μg/m^3, and from June to August, the concentration gradually decreased to 65.42 μg/m^3. From August to September, the concentration reached the second peak of 76.81 μg/m^3,

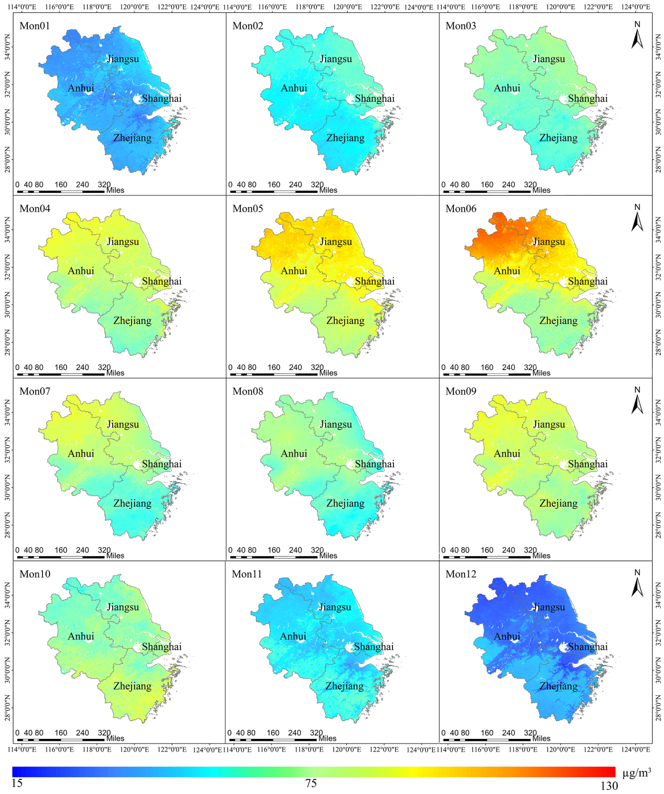
Mon01
Mon02
Mon03
Mon04
Mon05
Mon06
Mon07
Mon08
Mon09
Mon10
Mon11
Mon12
Jiangsu
Anhui
Shanghai
Zhejiang
0 40 80 160 240 320 Miles
114°0'0"E 116°0'0"E 118°0'0"E 120°0'0"E 122°0'0"E 124°0'0"E
34°0'0"N 32°0'0"N 30°0'0"N 28°0'0"N
N
15
75
130
μg/m³

Fig. 6. Monthly distribution of estimating results.

and gradually decreased to 34.86 μg/m^3 from September to December. The concentration of ozone in the spring and summer showed values ranged from 50 to 100 μg/m^3, which is mainly distributed in the northern Anhui and Jiangsu provinces (see Fig. S6). The concentration and distribution area of ozone were much higher than those in other seasons. The ozone concentration range in autumn was 48–85 μg/m^3, and the distribution of high-concentration ozone was relatively scattered. The ozone concentration in winter was lower than 60 μg/m^3, and the high-value areas were mostly distributed in the coastal areas of Jiangsu Province, Shanghai, and southern Zhejiang Province. The spatial changes of ozone concentration on the monthly and seasonal scales are consistent with the changes in temperature and solar radiation, which also confirms the research [57] that temperature is beneficial to production of ozone.

Fig. S7 shows the time series distribution map of the daily average ozone concentration predicted by the ensemble model from May 2018 to 2020. The highest value period was mainly concentrated in May to August, and the highest daily average concentration reached 137.76 μg/m^3. The lowest value period was mainly concentrated in November to February, and the lowest daily average concentration reached 7.81 μg/m^3. The daily average concentration of ozone showed a "W" shape change over time, with the concentration gradually increasing from February and reaching a peak in June, then slowly decreasing from June to October and reaching its lowest point in January. Overall, the results predicted by the model are in good agreement with the observations at ground stations, which indicates that the model has good applicability for long-term predictions.

IV. Discussion

A. Comparison of Existing Downscaling Models

Since our ensemble model takes into account the influence of temporal correlation among explanatory variables on ozone generation and distribution, it has better spatial and temporal resolution and better performance. Specifically, in terms of estimation accuracy, our model outperformed others. For example, the XGB model [58] (CV R^2: 0.89, RMSE: 4.75 μg/m^3), the RF model [59] (CV R^2: 0.69, RMSE: 26 μg/m^3), a daily-scale RF model [60] (CV R^2: 0.84, RMSE: 0.0059 ppm), the space-time extreme random tree model [61] (CV R^2: 0.87, RMSE: 21.10 μg/m^3), the deep neural network model [62], the feedforward back propagation neural network [63] (CV R^2: 0.88, RMSE: 10.74 μg/m^3), and the artificial neural network model [64] (CV R^2: 0.89, RMSE: 0.0066 ppm). This is because our model comprehensively considered that the input variables themselves are limited by time factors, which is also reflected in the comparison between the construction of the overall model and the ensemble model in this article (as shown in Fig. 3).

In addition, the model is also similar to the Geoi-LGB method [38] (CV R^2: 0.91, RMSE: 10.25 μg/m^3). The difference is that Chen constructed spatiotemporal autocorrelation factors and introduced them into the model as modeling data. In this work, all explanatory variables were divided into different datasets according to the time, and input into multiple base models, respectively, to get optimal solutions. But in general, the temporally ensemble approach of explanatory variables could improve the model's estimation accuracy. As for spatial resolution, the ensemble model provided ozone exposure information with a spatial resolution of 1 km^2, which could provide data support for preventing and controlling polluted gases. And few scholars have constructed a near-ground ozone concentration estimation model with a spatial resolution of 1 km^2. For instance, the RF generalized additive model [65] provided ozone exposure with a spatial resolution of 0.25°, and a statistical model obtained a product with a spatial resolution of 0.2° [66]. The RF model [67], the XGB model [68], and an inheritance algorithm combined with multisource geographic data [69], all of the above only provided ozone exposure information at a spatial scale of 0.1°. In contrast, our ensemble model performs better in both spatiotemporal resolution and performance.

B. Uncertainty Evaluation of the Model

In this study, we found that the four independent ML models (XGB, RF, EXT, and LGBM) involved in modeling showed different results in different months, which may be due to differences in model structure and explanatory variables driving ozone Mechanisms are different. In terms of model structure, RF and EXT are decision tree-based ensemble methods to reduce the variance of the model to build a model for the optimization goal [52], [53]. XGB and LGBM are methods based on gradient boosting trees, which mainly generate models by gradually reducing the residual error of the model [54], [55]. The different ways of building the model lead to differences in the strategy of generating the tree, which may cause the model to show different results in different months. In terms of explanatory variables, we observed that some explanatory variables involved in modeling have different effects on the generation and distribution of ozone in different time periods, which may lead to different performances of different models in different months. For example, in summer, the increase in air humidity in the Yangtze River Delta area makes the moisture content in the area skyrocket, which weakens the influence from solar radiation to a certain extent, and then weakens the photochemical reaction of ozone generation [70]. In addition, the increase in air humidity also leads to an increase in precipitation in the region. Unstable phenomena such as thunderstorms are not conducive to the deposition of ozone [71]. In addition, the increase in humidity will increase the concentration of OH radicals in the atmosphere, and a large amount of OH will combine with ozone in the atmosphere, which may lead to the degradation and reduction of ozone [72]. Temperature is an important factor affecting the change of ozone concentration. In summer, higher temperature may promote the photochemical reaction between ozone precursors. In winter, the temperature drops, and it is accompanied by rain, snow, and strong wind, which is not conducive to the formation and accumulation of ozone. In addition, low temperature will reduce the activity and metabolism rate of relevant biological enzymes in plant cells, thereby slowing down the respiration rate of plant cells, which hinders the production of VOCs by vegetation through respiration, and then affects the chemical reactions that produce ozone [73], [74]. Wind speed and direction also affect the generation and distribution of ozone. In summer, the Yangtze River Delta region of China is mostly affected by the East Asian monsoon. The warm and humid air flow from the southeast accelerates the photochemical reaction of ozone formation. In addition, this region is often affected by typhoons in summer. The greater wind speed makes the precursors of ozone transported from the high-concentration high altitude to the near ground, and the increase of reactants intensifies the chemical reaction to generate ozone, thus promoting the formation of ozone [75]. In winter, mainly affected by the dry and cold airflow from the northwest, the lower temperature may reverse the photochemical reaction that generates ozone, thereby reducing the ozone concentration in the atmosphere. In addition to the above factors, boundary layer height, air pressure, and terrain conditions [76], [77] are also important factors affecting the formation and distribution of ozone in the atmosphere.

C. Potential for Fine-Scale Ozone Monitoring

The refined ozone distribution was proved to be potential in capturing the spatial heterogeneity of a small area. In the study area, we found the fine distribution characteristics of the city and its surroundings, which could not be reflected by satellites.

As shown in Fig. 7, the distribution of ozone mainly concentrated on the central in the main urban area, dominated by Yangpu District, Putuo District, Xuhui District, and Pu dong New Area. High-value areas are in suburban, dominated by Chuansha Town, Zhoupu Town, Zhuqiao Town, Huinan Town, Anting Town, Nanxiang Town, Jiading District, and Baoshan District. Low-value areas are in farmland, dominated by the southern part of Qingpu District, the southeastern part of Fengxian District, and the central part of Songjiang District. The distribution of ozone concentration showed a trend of “medium-high-low” from the city center to the agricultural areas. The ozone concentration in suburban areas was significantly higher than that in other areas, where a large number of factories and chemical enterprises existed. The combustion of fossil fuels and emission of chemical gases generated a significant amount of ozone precursors such as NOx and VOCs [78]; as a result, the ozone concentration was higher than other areas. The ozone concentration in the main urban area was higher than farmland area, which was mainly because the farmland area had less population, transportation activities than the main urban area. Different from the distribution of ozone in the Shanghai area, the Hangzhou area presented a “low-medium” trend from the main urban area to away from the urban area [see Fig. 7(a)(1) and (b)(1)]. Among them, the median area was mainly distributed in the suburban area on the edge of the main urban area and the mountainous area far away from the urban area. The former had a large number of factories that produced massive ozone precursors, which made the ozone concentration in this area rise. The latter was covered with massive forest. On the one hand, plants’ respiratory processes generated a significant amount of VOCs [79], and the photosynthesis was strong in mountainous areas, which was conducive to the generation of ozone. On the other hand, the ozone imported from the outside could not dissipate with the wind and eventually accumulated in forested areas. Due to the high altitude, low forest canopy density, and sufficient sunlight, the reaction of photochemical to generate ozone was

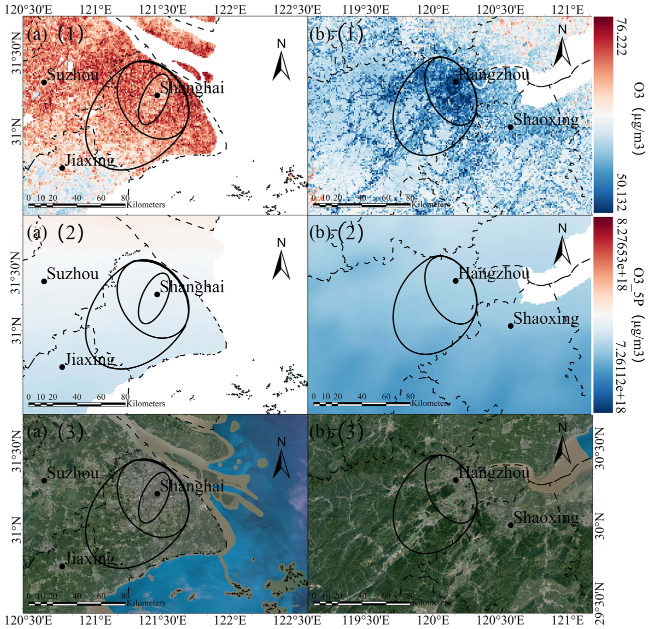

Fig. 7. Distribution of ozone in Shanghai and Hangzhou. (a)(1) Estimated results of model in Shanghai. (a)(2) Observation results of satellite in Shanghai. (a)(3) Real scene of remote sensing in Shanghai. (b)(1) Estimated results of model in Hangzhou. (b)(2) Observation results of satellite Hangzhou. (b)(3) Real scene of remote sensing in Hangzhou.

strong [80]. In addition, compared with Shanghai, the main reason why Hangzhou does not have high ozone concentration areas is that it has completed industrial adjustments in recent years (see Fig. S8), the proportion of the tertiary industry has reached half, and the GDP of the secondary industry is less than half of that of Shanghai. In general, compared with the interpolation results of satellite observations [see Fig. 7(a)(2) and (b)(2)], the ensemble model exhibited more details and provided richer information. However, the interpolation results of satellite observations in this area were smoothed and could not be seen that the Shanghai and Hangzhou regions presented a "medium-high-low" and "low-medium" trend. Therefore, the ensemble model has many advantages in the research of small areas, and the 1 km^2 spatial resolution products can offer data support for local ozone management.

D. Study Limitations

Our research also has some limitations. First, the time scale is not enough to analyze the evolution process of regional ozone under long-term series. Since TROPOMI only provides ozone column concentration data after 2018, long-term ozone monitoring cannot be performed. Although another satellite sensor, OMI, can provide ozone column concentration data from 2004 to the present, the spatial resolution of this product is relatively coarse and some data have been missing since 2013 due to equipment aging, which affects the accuracy of estimation. Second, meteorological conditions often exhibit a lag effect on air pollutants. When introducing meteorological explanatory variables, only the data of the current day is introduced without considering the influence of meteorological factors in the previous and subsequent days, which would lead to certain deviations in the model's prediction of ozone concentration and distribution. Third, there are large uncertainties in ensemble explanatory variables with different spatial resolutions into a 1 km^2 prediction grid using inverse distance weighted (IDW) interpolation and bilinear interpolation resampling methods. Due to the flaw in the IDW interpolation algorithm, which is very sensitive to noise, it may not be possible to accurately predict grid cell values. What is more, the bilinear interpolation method only considers the influence of the gray value of the four adjacent points around the predicted point, but does not take into account the influence of the change rate of the gray value between adjacent points, which leads to the loss of high-frequency components of the image.

V. Conclusion

Combining with multiple data sources such as reanalysis data, meteorological data, and human activity data, this study proposes a temporally ensemble model based on four ML algorithms: XGB, RF, EXT, and LGBM. Our aim is to address the temporal heterogeneity issues that affect near-ground ozone concentration and distribution. The model successfully predicted the near-ground ozone concentration in the Yangtze River Delta region of China and generated a spatial distribution map of the daily average near-ground O_3 concentration with a spatial resolution of 1 km^2. The main findings are as follows.

1) Due to its ability to capture subtle variations in explanatory variables, the temporal ensemble model exhibits higher estimation accuracy compared to most ML models.
2) Estimation models with high spatial resolution can capture changes in near-surface ozone concentrations over small regions.

The excellent performance of the ensemble model in terms of time and space would contribute to providing favorable data support for the prevention and comprehensive management of air pollutants in the future.

Acknowledgment

The authors would like to extend their special thanks to their colleagues for their technical support during the experimental process. They also express their gratitude to the free websites that provided data support for this study, including the Sentinel-5P TROPOMI O_3 Level 2 released on NASA's official website, hourly meteorological data provided by the China National Meteorological Science Data Center, population data provided by Worldpop, road network data provided by OpenStreetMap, DEM data provided by the SRTM3 file released by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, land cover data provided by the European Space Agency, and ERA5 reanalysis data released by the European Center for Medium-Range Weather Forecasts.

References

[1] E. Agathokleous et al., "Ozone affects plant, insect, and soil microbial communities: A threat to terrestrial ecosystems and biodiversity," *Sci. Adv.*, vol. 6, 2020, Art. no. eabc1176.

[2] T. Le et al., "Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China," *Science*, vol. 369, no. 6504, pp. 702–706, 2020.

[3] P. J. Nowack, N. L. Abraham, P. Braesicke, and J. A. Pyle, "Stratospheric ozone changes under solar geoengineering: Implications for UV exposure and air quality," *Atmos. Chem. Phys.*, vol. 16, no. 6, pp. 4191–4203, 2016.

[4] J. I. Steinfeld, "Atmospheric chemistry and physics: From air pollution to climate change," *Environ., Sci. Policy Sustain. Develop.*, vol. 40, no. 7, p. 26, Sep. 1998, doi: 10.1080/00139157.1999.10544295.

[5] J. E. Sickles, J. C. Suggs, and L. M. Vorburger, "Ozone indicators determined at rural sites in the eastern United States by two monitoring networks," *Environ. Monit. Assessment*, vol. 65, pp. 485–502, 2000.

[6] G. Lee, J. H. Park, M. S. Koh, M. Lee, J. S. Han, and J. C. Kim, "Current status and future directions of tropospheric photochemical ozone studies in Korea," *J. Korean Soc. Atmos. Environ.*, vol. 36, no. 4, pp. 419–441, 2020.

[7] B. Rani, U. Singh, A. K. Chuhan, D. Sharma, and R. Maheshwari, "Photochemical smog pollution and its mitigation measures," *J. Adv. Sci. Res.*, vol. 2, no. 4, pp. 28–33, 2011.

[8] J. Li et al., "The association between ozone and years of life lost from stroke, 2013–2017: A retrospective regression analysis in 48 major Chinese cities," *J. Hazardous Mater.*, vol. 405, 2021, Art. no. 124220.

[9] Y. Tian et al., "The impact of ambient ozone pollution on pneumonia: A nationwide time-series analysis," *Environ. Int.*, vol. 136, 2020, Art. no. 105498.

[10] R. Canella et al., "Tropospheric ozone effects on chlorine current in lung epithelial cells: An electrophysiological approach," *Free Radical Biol. Med.*, vol. 1, no. 96, pp. S58–S59, 2016.

[11] D. Nuvolone, D. Petri, and F. Voller, "The effects of ozone on human health," *Environ. Sci. Pollut. Res.*, vol. 25, pp. 8074–8088, 2018.

[12] W.-N. Wang et al., "Assessing spatial and temporal patterns of observed ground-level ozone in China," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 3651.

[13] X. Shi, C. Zhao, J. H. Jiang, C. Wang, X. Yang, and Y. L. Yung, "Spatial representativeness of PM2.5 concentrations obtained using observations from network stations," *J. Geophys. Res., Atmos.*, vol. 123, no. 6, pp. 3145–3158, 2018.

[14] R. Li, L. Cui, F. Hongbo, J. Li, Y. Zhao, and J. Chen, "Satellite-based estimation of full-coverage ozone (O_3) concentration and health effect assessment across Hainan Island," *J. Cleaner Prod.*, vol. 244, 2020, Art. no. 118773.

[15] Z. Wang et al., "Systematics of atmospheric environment monitoring in China via satellite remote sensing," *Air Qual., Atmos. Health*, vol. 14, pp. 157–169, 2021.

[16] Q. Sun et al., "Acute effect of multiple ozone metrics on mortality by season in 34 Chinese counties in 2013–2015," *J. Intern. Med.*, vol. 283, no. 5, pp. 481–488, 2018.

[17] X. Y. Zhang, L. M. Zhao, M. M. Cheng, and D. M. Chen, "Estimating ground-level ozone concentrations in eastern China using satellite-based precursors," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4754–4763, Jul. 2020.

[18] H. Wang et al., "Ground-based MAX-DOAS measurements of tropospheric aerosols, NO_2, and HCHO distributions in the urban environment of Shanghai, China," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1726.

[19] S. Wang et al., "A high-performance convolutional neural network for ground-level ozone estimation in Eastern China," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1640.

[20] L. Shen et al., "An evaluation of the ability of the ozone monitoring instrument (OMI) to observe boundary layer ozone pollution across China: Application to 2005–2017 ozone trends," *Atmos. Chem. Phys.*, vol. 19, no. 9, pp. 6551–6560, 2019.

[21] K. W. Bowman, "Toward the next generation of air quality monitoring: Ozone," *Atmos. Environ.*, vol. 80, pp. 571–583, 2013.

[22] C. I. Alvarez-Mendoza, A. Teodoro, and L. Ramirez-Cando, "Spatial estimation of surface ozone concentrations in Quito Ecuador with remote sensing data, air pollution measurements and meteorological variables," *Environ. Monit. Assessment*, vol. 191, pp. 1–15, 2019.

[23] A. Adam-Poupart, A. Brand, M. Fournier, M. Jerrett, and A. Smargiassi, "Spatiotemporal modeling of ozone levels in Quebec (Canada): A comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy–LUR approaches," *Environ. Health Perspectives*, vol. 122, no. 9, pp. 970–976, 2014.

[24] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating ground-level PM2.5 in China using satellite remote sensing," *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7436–7444, 2014.

[25] S. D. Latif, A. N. Ahmed, M. Sherif, A. Sefelnasr, and A. El-Shafie, "Reservoir water balance simulation model utilizing machine learning algorithm," *Alexandria Eng. J.*, vol. 60, no. 1, pp. 1365–1378, 2021.

[26] M. Sapitang, W. M. Ridwan, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Machine learning application in reservoir water level forecasting for sustainable hydropower generation strategy," *Sustainability*, vol. 12, no. 15, 2020, Art. no. 6121.

[27] Y. Zhan, Y. Luo, X. Deng, M. L. Grieneisen, M. Zhang, and B. Di, "Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment," *Environ. Pollut.*, vol. 233, pp. 464–473, 2018.

[28] J. W. You, B. Zou, X. G. Zhao, S. Xu, and R. He, "Estimating ground-level NO_2 concentrations across mainland China using random forests regression modeling," *China Environ. Sci.*, vol. 39, pp. 969–979, 2019.

[29] M. Ahmad, B. Rappenglück, O. O. Osibanjo, and A. Retama, "A machine learning approach to investigate the build-up of surface ozone in Mexico-City," *J. Cleaner Prod.*, vol. 379, 2022, Art. no. 134638.

[30] Y. Lyu, Q. Ju, F. Lv, J. Feng, X. Pang, and X. Li, "Spatiotemporal variations of air pollutants and ozone prediction using machine learning algorithms in the Beijing-Tianjin-Hebei region from 2014 to 2021," *Environ. Pollut.*, vol. 306, 2022, Art. no. 119420.

[31] J. Liu, "Mapping high resolution national daily NO_2 exposure across mainland China using an ensemble algorithm," *Environ. Pollut.*, vol. 279, 2021, Art. no. 116932.

[32] A. A. Heidari, M. Akhoondzadeh, and H. Chen, "A wavelet PM2.5 prediction system using optimized kernel extreme learning with Boruta-XGBoost feature selection," *Mathematics*, vol. 10, no. 19, 2022, Art. no. 3566.

[33] Y. Kang et al., "Estimation of surface-level NO_2 and O_3 concentrations using TROPOMI data and machine learning over East Asia," *Environ. Pollut.*, vol. 288, 2021, Art. no. 117711.

[34] A. Sarkar, S. S. Ray, A. Prasad, and C. Pradhan, "A novel detection approach of ground level ozone using machine learning classifiers," in *Proc. 5th Int. Conf. IoT Social, Mobile, Analytics Cloud*, 2021, pp. 428–432.

[35] X. Chen et al., "Estimating monthly surface ozone using multi-source satellite products in China based on deep forest model," *Atmos. Environ.*, vol. 307, 2023, Art. no. 119819.

[36] S. Zhu et al., "Learning surface ozone from satellite columns (LESO): A regional daily estimation framework for surface ozone monitoring in China," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 4108711.

[37] M. Li, Q. Yang, Q. Yuan, and L. Zhu, "Estimation of high spatial resolution ground-level ozone concentrations based on Landsat 8 TIR bands with deep forest model," *Chemosphere*, vol. 301, 2022, Art. no. 134817.

[38] J. Chen, H. Shen, X. Li, T. Li, and Y. Wei, "Ground-level ozone estimation based on geo-intelligent machine learning by fusing in-situ observations, remote sensing data, and model simulation data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, 2022, Art. no. 102955.

[39] S. He, H. Dong, Z. Zhang, and Y. Yuan, "An ensemble model-based estimation of nitrogen dioxide in a southeastern coastal region of China," *Remote Sens.*, vol. 14, no. 12, 2022, Art. no. 2807.

[40] D. Li and X. Ren, "Prediction of ozone hourly concentrations based on machine learning technology," *Sustainability*, vol. 14, no. 10, 2022, Art. no. 5964.

[41] S. Zhao et al., "Annual and diurnal variations of gaseous and particulate pollutants in 31 provincial capital cities based on in situ air quality monitoring data from China National Environmental Monitoring Center," *Environ. Int.*, vol. 86, pp. 92–106, 2016.

[42] L. Du, W. Lin, J. Du, M. Jin, and M. Fan, "Can vertical environmental regulation induce enterprise green innovation? A new perspective from automatic air quality monitoring station in China," *J. Environ. Manage.*, vol. 317, 2022, Art. no. 115349.

[43] Y. Feng, W. Zhang, D. Sun, and L. Zhang, "Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification," *Atmos. Environ.*, vol. 45, no. 11, pp. 1979–1985, 2011.

[44] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2021, Art. no. 5610819.

[45] M. Mohajane et al., "Land use/land cover (LULC) using Landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco," *Environments*, vol. 5, no. 12, 2018, Art. no. 131.

[46] Y. C. Li et al., "A study of the spatiotemporal dynamic of land cover types and the driving forces of grassland area change in Gannan Prefecture and Northwest Sichuan based on CCI-LC data," *Acta Prataculturae Sin.*, vol. 29, no. 3, pp. 1–15, 2020.

[47] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[48] R. Liu, Z. Ma, Y. Liu, Y. Shao, W. Zhao, and J. Bi, "Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach," *Environ. Int.*, vol. 142, 2020, Art. no. 105823.

[49] C. E. Reid et al., "Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning," *Environ. Sci. Technol.*, vol. 49, no. 6, pp. 3887–3896, 2015.

[50] P. Mahesh, P. Sujatha, C. B. S. Dutt, and S. Jose, "Comparative study of the tropospheric ozone derived from satellite data using different interpolation techniques," *Int. J. Remote Sens.*, vol. 36, no. 9, pp. 2409–2420, 2015.

[51] J. D. Berman, P. N. Breysse, R. H. White, D. W. Waugh, and F. C. Curriero, "Evaluating methods for spatial mapping: Applications for estimating ozone concentrations across the contiguous United States," *Environ. Technol. Innov.*, vol. 3, pp. 1–10, 2015.

[52] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[53] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, pp. 3–42, 2006.

[54] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.

[55] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 3149–3157.

[56] A.-Z. S. Bin Habib and T. Tasnim, "An ensemble hard voting model for cardiovascular disease prediction," in *Proc. 2nd Int. Conf. Sustain. Technol. Ind. 4.0*, 2020, pp. 1–6.

[57] X. Pu et al., "Enhanced surface ozone during the heat wave of 2013 in Yangtze River Delta region, China," *Sci. Total Environ.*, vol. 603, pp. 807–816, 2017.

[58] Y. Li, K. Qin, L. Ding, W. Fan, and H. Qin, "Estimation of ground-level ozone concentration based on GBRT," *China Environ. Sci.*, vol. 40, no. 3, pp. 997–1007, 2020.

[59] R. Pernak, M. Alvarado, C. Lonsdale, M. Mountain, J. Hegarty, and T. Nehrkorn, “Forecasting surface O_3 in Texas urban areas using random forest and generalized additive models,” *Aerosol Air Qual. Res.*, vol. 19, no. 12, pp. 2815–2826, 2019.

[60] W. Wang, X. Liu, J. Bi, and Y. Liu, “A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology,” *Environ. Int.*, vol. 158, 2022, Art. no. 106917.

[61] J. Wei et al., “Full-coverage mapping and spatiotemporal variations of ground-level ozone (O_3) pollution from 2013 to 2020 across China,” *Remote Sens. Environ.*, vol. 270, 2022, Art. no. 112775.

[62] Y.-W. Chen, S. Medya, and Y.-C. Chen, “Investigating variable importance in ground-level ozone formation with supervised learning,” *Atmos. Environ.*, vol. 282, 2022, Art. no. 119148.

[63] L. I. Ziwei, M. A. Qingxun, and L. Jie, “BP neural network for near-surface ozone estimation and spatial and temporal characteristics analysis,” *Bull. Surv. Mapping*, no. 6, pp. 28–32, 2021.

[64] A. Yafouz et al., “Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction,” *Alexandria Eng. J.*, vol. 61, no. 6, pp. 4607–4622, 2022.

[65] R. Li, Y. Zhao, W. Zhou, Y. Meng, Z. Zhang, and H. Fu, “Developing a novel hybrid model for the estimation of surface 8 H ozone (O_3) across the remote Tibetan Plateau during 2005–2018,” *Atmos. Chem. Phys.*, vol. 20, no. 10, pp. 6159–6175, 2020.

[66] P. Bogaert, G. Christakos, M. Jerrett, and H.-L. Yu, “Spatiotemporal modelling of ozone distribution in the State of California,” *Atmos. Environ.*, vol. 43, no. 15, pp. 2471–2480, 2009.

[67] T. Li, Y. Lu, X. Deng, and Y. Zhan, “Spatiotemporal variations in meteorological influences on ambient ozone in China: A machine learning approach,” *Atmos. Pollut. Res.*, vol. 14, no. 4, 2023, Art. no. 101720.

[68] X. Hu, J. Zhang, W. Xue, L. Zhou, Y. Che, and T. Han, “Estimation of the near-surface ozone concentration with full spatiotemporal coverage across the Beijing-Tianjin-Hebei region based on extreme gradient boosting combined with a WRF-Chem model,” *Atmosphere*, vol. 13, no. 4, 2022, Art. no. 632.

[69] T. Xue et al., “Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations,” *Environ. Int.*, vol. 123, pp. 345–357, 2019.

[70] M. Ma, G. Yao, J. Guo, and K. Bai, “Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors,” *J. Environ. Manage.*, vol. 288, 2021, Art. no. 112368.

[71] R. Szep, R. Keresztes, S. Tonk, A. Korodi, and M. E. Craciun, “The examination of the effects of relative humidity on the changes of tropospheric ozone concentrations in the Ciuc Basin, Romania,” *Rev. Chim*, vol. 68, pp. 642–645, 2017.

[72] T. Wang, L. Xue, P. Brimblecombe, Y. F. Lam, L. Li, and L. Zhang, “Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects,” *Sci. Total Environ.*, vol. 575, pp. 1582–1596, 2017.

[73] Y.-K. Cheng et al., “Spatial and temporal distribution characteristics of ozone concentration and population health benefit assessment in the Yangtze River Delta region from 2017 to 2020,” *Huan Jing Ke Xue = Huanjing Kexue*, vol. 44, no. 2, pp. 719–729, 2023.

[74] S. Rui, Z. Hong, W. Shuibing, and W. Youwen, “Research on the temporal and spatial distribution of ozone in typical cities in the Yangtze River Delta region and its correlation with meteorological factors,” *J. Atmos. Environ. Opt.*, vol. 16, no. 6, pp. 483–494, 2021.

[75] Y. Ke, Z. Zhu, C. Wenmi, W. Weiguo, and W. Haoyue, “Characteristic analysis of the relationship between the South Asian monsoon circulation and the temporal and spatial distribution of ozone,” *J. Yunnan Univ. (Natural Sci. Ed.)*, vol. 45, no. 1, pp. 111–123, 2023.

[76] R. J. Seguel, C. A. Mancilla, R. Rondanelli, M. A. Leiva, and R. G. Morales, “Ozone distribution in the lower troposphere over complex terrain in Central Chile,” *J. Geophys. Res., Atmos.*, vol. 118, no. 7, pp. 2966–2980, 2013.

[77] K. Biqin et al., “Spatial-temporal distribution characteristics and driving factors of surface ozone in North China,” *Chin. Environ. Sci.*, vol. 42, no. 4, pp. 1562–1574, 2022.

[78] S. Liang et al., “Estimation of health and economic benefits based on ozone exposure level with high spatial-temporal resolution by fusing satellite and station observations,” *Environ. Pollut.*, vol. 255, 2019, Art. no. 113267.

[79] J. Peñuelas and M. Staudt, “BVOCs and global change,” *Trends Plant Sci.*, vol. 15, no. 3, pp. 133–144, 2010.

[80] W. Duan et al., “Variation of ozone concentrations in three urban forests under different habitats of Shenzhen in summer,” *China Environ. Sci.*, vol. 37, no. 6, pp. 2064–2071, 2017.

Zhen Li received the B.S. degree in physical geography and resource environment from Henan Polytechnic University, Jiaozuo, China, in 2021. He is currently working toward the master’s degree in geography with Wuhan University of Technology, Wuhan, China.

His research interests include atmospheric environment remote sensing and geographic information modeling.

Heng Dong received the Ph.D. degree in cartography and geographic information system from Peking University, Beijing, China, in 2013.

He is currently an Associate Professor with the School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan, China. His research interests include environmental remote sensing and geographic information modeling.

Zili Zhang received the B.E. degree in cartography and geography information and M.S. degree in physical geography from Xinjiang University, Urumqi, China, in 2002 and 2005, respectively, and the Ph.D. degree in cartography and geography information system from Peking University, Beijing, China, in 2009.

He is currently the Senior Engineer with the Zhejiang Province Ecological Environment Monitoring Center, Hangzhou, China. His current research focuses on remote sensing of environment.

Lan Luo received the B.E. degree in computer science and technology from the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China, in 2015.

She has successively worked for several Chinese Internet technology companies, and is currently a staff member of the Big Data Center of Zhejiang Environmental Monitoring Center, Hangzhou, China.

Sicong He received the B.S. degree in geographic information science in 2019 and M.E. degree in geography in 2022, both from Wuhan University of Technology, Wuhan, China, where he is currently working toward the Ph.D. degree in environmental geographic information systems.

His research interests include atmospheric environment remote sensing and geographic information modeling.