Environmental Research Communications

LETTER

Estimations of ambient fine particle and ozone level at a suburban site of Beijing in winter

Weixuan Liu1,3,4, Zheng Yang1,4 and Qingyang Liu1,2

1 Beijing Milu Ecological Research Center, Beijing, 100076, People’s Republic of China
2 College of Biology and the Environment, Nanjing Forestry University, Nanjing, 210037, People’s Republic of China
3 The High School Affiliated to Renmin University of China, Beijing, 100080, People’s Republic of China
4 The authors contribute equally.

E-mail: qyliu@njfu.edu.cn

Keywords: machine learning, PM_{2.5}, O_3, suburban areas

Abstract

Estimates of PM_{2.5} and O_3 in suburban areas are of importance for assessing exposure risk and epidemiological studies of air pollution where large scale and long-term measurements network are absent. To fulfill this goal, our study develops a flexible approach to predict levels of PM_{2.5} and O_3 at a suburban site of Beijing using multilayer perceptron (MLP) neural network analysis with the inputs of gaseous air pollutants (CO, SO_2, NO, and NO_2) and meteorological parameters (wind direction, wind speed, temperature, pressure and humidity). Daily ambient data of PM_{2.5}, O_3, PM_{10}, CO, SO_2, NO, and NO_2 were estimated using hourly data collected from January 20 to March 10 in the years from 2016–2020 at a suburban site of Beijing, respectively. Ambient measured levels of PM_{2.5} and O_3 were compared with the output estimates of PM_{2.5} and O_3 through MLP neural network analysis with limited input variables. Overall, MLP neural network analysis could explain 97% of measured PM_{2.5} mass and 82% of measured O_3 level with R^2 values of 0.983 and 0.905, respectively. This approach could be helpful for reconstruct historical PM_{2.5} and O_3 levels in suburban areas.

1. Introduction

Air pollution is recognized as the largest environmental risk to human health by the World Health Organization (WHO) for population in areas where levels of air pollutants exceed the limit (D’Antoni et al 2019). Numerous studies have documented that air pollution exerted negative impacts on health through showing associations between exposures of air pollution and incidence rates of cardiovascular disease, autism spectrum disorder in children, leukemia as well as mortality of lung cancers (Lin et al 2017, Lin et al 2018, Lin et al 2020, Lin et al 2021). Air pollution is mixtures of various substances including ground-level ozone (O_3) and ambient fine particulate matter (de Vries et al 2014, LaCount et al 2021). Ground-level ozone is formed through the chemical reactions between nitrogen oxide (NO_x) and volatile organic compounds (VOCs) under sunlight (de Vries et al 2014). The meteorological conditions including amount of sunlight, wind and temperature are found to be controlling factors in ground-level ozone formation (de Vries et al 2014). Ambient fine particulate matter (<2.5 microns in aerodynamic diameter; PM_{2.5}) is mainly comprised of inorganic and organic components that emitted from emission sources (Liu et al 2014b, Liu et al 2016, Liu et al 2019). Meanwhile, gaseous pollutants including NO, NO_2, SO_2 and CO are co-emitted from the sources of PM_{2.5} and reacted with O_3 to form secondary chemicals (Luo et al 2015, Tang et al 2020). The meteorological parameters such as wind and temperature are found to be main factors in spreading PM_{2.5} pollution (Cheng et al 2016, Sun et al 2020).

Due to the negative impacts of ground-level O_3 and PM_{2.5}, large scale and long-term measurements of O_3 and PM_{2.5} were established in countries all around the world (de Vries et al 2014, LaCount et al 2021). The nationwide measurements of O_3 and PM_{2.5} were started in China when the National Ambient Air Quality...
Standards (NAQPS) for limits of O₃ and PM₂.₅ were issued in 2013 (Liu et al 2021). Recently, several advanced methods were developed to predict levels of PM₂.₅ and O₃ recently. For example, Eftiens et al (2012) used land use model with the inputs of several variables including traffic intensity, population, and land-use to estimate PM₂.₅ mass in 20 European study areas. The model could explain 35%–94% of PM₂.₅ mass collected at 20 sites in European areas. Wang et al (2016) incorporated land-use regression (LUR) with chemical transport modeling to calculate spatiotemporal variability in levels of O₃ and PM₂.₅ from 2000 to 2008 in the Los Angeles Basin. Huang et al (2021) established a hybrid spatiotemporal model that incorporates satellite, chemical transport model, geographic, and meteorological data to reconstruct daily PM₂.₅ concentrations in China from 2013 to 2019 at 1 km × 1 km grid cells. Xiao et al (2018) employed a set of machine learning approaches including random forest, generalized additive model, and extreme gradient boosting with the inputs of long satellite aerosol data to estimate daily PM₂.₅ in different areas of China, respectively.

Beijing, as the capital of China, has experienced serious air pollution with high levels of O₃ and PM₂.₅ in winter in recent years (Liu et al 2014a, Liu et al 2016, Miao et al 2018, Liu et al 2021, Ren et al 2021, Xu et al 2021, Yang et al 2021). Since NAQPS was launched in 2013, Beijing Municipal Environmental Monitoring Center has established 34 monitoring sites across Beijing to announce hourly and daily air quality data including O₃, PM₂.₅, PM₁₀, NO₂, SO₂, and CO (Hua et al 2021). These sites are mostly located in urban areas of Beijing (figure S1(A) (available online at stacks.iop.org/ERC/3/081008/mmedia)), while the sites are uneven distributed geographically across Beijing. Thus, it is necessary to develop subsidiary methods to present spatiotemporal levels of O₃ and PM₂.₅ across Beijing, especially in suburban areas (Huang et al 2021). As discussed above, machine learning approaches could generate high precisions of O₃ and PM₂.₅ estimates with appropriate inputs of variables.

In order to fill this gap, our study aims to estimate levels of O₃ and PM₂.₅ in suburban areas of Beijing using multilayer perceptron (MLP) neural network analysis with the inputs of ambient gaseous data. The study collected daily air quality and meteorological parameters including PM₂.₅, O₃, PM₁₀, CO, SO₂, NO₂, wind direction, wind speed, temperature, pressure and humidity from January 20 to March 10 in the years from 2016–2020 at a suburban area of Beijing, respectively. In total, ~3000 field data were included in MLP neural network analysis to generate high estimate precisions of O₃ and PM₂.₅ in suburban areas of Beijing. The estimated approaches from our study are insightful for local policy-makers to reconstruct unbiased historical O₃ and PM₂.₅ levels for exposure assessment and epidemiological studies of air pollution.

2. Materials and methods

2.1. Sampling site
The sampling site is situated at the Milu Park Ecological Research Center (figure S1(A), 39.78N, 116.47E), at a suburban area to the south-5th-ring road of Beijing (Liu et al 2021). Automated instruments were set up on the roof of a building in the Park at 15 m above ground. The sampling location is surrounded by artificial wetlands with plenty of vegetation and trees. There are no industrial facilities near the site with only a two-way road about 900 meters to the northwest. There are 34 monitoring sites for air quality affiliated with the Beijing Municipal Environmental Monitoring Center (http://bjmemc.com.cn/) in Beijing (figure S1(A)). Two monitoring sites including Jiugong, Daxing District (39.90N, 116.40E) and Yizhuang Development Zone (39.80N, 116.51E) are close to our sampling site (39.78N, 116.47E). The two sites are located at the downtown areas of Daxing District, which are ~6 km away from our sampling site (figure S1(B)). In this study, we choose the ambient air quality data collected at these two sampling sites (Jiugong, Daxing District and Yizhuang Development Zone) affiliated with the Beijing Municipal Environmental Monitoring Center to compare with those obtained from our sampling site (Milu Park Ecological Research Center).

2.2. Data collection
At the sampling site, hourly concentration data of gaseous pollutants including NO, NO₂, SO₂, CO and O₃ were collected with Thermo Scientific NO-NO₂-NOₓ Analyzer (Model 42i), SO₂ Analyzer (Model 43i), CO Analyzer (Model 48i) and O₃ Analyzer (Model 49i) respectively (Liang et al 2019, Bekbulat et al 2021). The flow rates for NO-NO₂-NOₓ Analyzer, SO₂ Analyzer, CO Analyzer and O₃ Analyzer were kept at 25 ml min⁻¹, with detection limits ranging from 10–50 ng m⁻³, respectively. Mass concentrations of PM₁₀ and PM₂.₅ were measured by Thermo Scientific Tapered Element Oscillating Microbalance (TEOM) 1405- Ambient Particulate Monitor with Filter Dynamics Measurement System (FDMS) (Liu et al 2021). Ambient particles were collected at a flow rate of 16.7 l min⁻¹ on a glass-fiber filter tape with PM₁₀ or PM₂.₅ cyclone inlets. These data were recorded automatically by the instrument and the calibration of the flow rate was done monthly under normal conditions by the technicians. Furthermore, hourly meteorological data including wind direction, wind speed,
temperature, pressure and humidity were collected using the WXT520 Meteorological measuring instrument (Vaisala).

Hourly data including NO, NO\textsubscript{2}, SO\textsubscript{2}, CO, O\textsubscript{3}, PM\textsubscript{10}, PM\textsubscript{2.5}, wind direction, wind speed, temperature, pressure and humidity during the periods from January 20 to March 10, 2020 (lockdown period, \(n = 1224\)) were selected and compared with those obtained from the same periods during the past four years (January 20 to March 10, 2016, \(n = 1211\); January 20 to March 10, 2017, \(n = 1200\); January 20 to March 10, 2018, \(n = 863\); January 20 to March 10, 2019, \(n = 1200\)) to assess the trends in pollutants over these years. Data in some sampling periods (i.e., \(\sim 7\%\) of the total samples) were absent because of rainfall.

The ambient hourly data (NO\textsubscript{2}, SO\textsubscript{2}, CO, O\textsubscript{3}, PM\textsubscript{10} and PM\textsubscript{2.5}) at the two sampling sites including Jiugong, Daxing District (39.90N, 116.40E) and Yizhuang Development Zone (39.80N, 116.51E) were collected from the website of the Beijing Municipal Environmental Monitoring Center (http://bjmemc.com.cn/) during January 20 to March 10 across the years of 2016–2020. Data in some sampling periods (i.e., \(\sim 5\%\) of the total samples) were absent because of rainfall. Meteorological data at the two sampling site including wind direction, wind speed, temperature, pressure and humidity were not available through the website of the Beijing Municipal Environmental Monitoring Center (http://bjmemc.com.cn/).

The daily concentration of gaseous pollutants (NO, NO\textsubscript{2}, SO\textsubscript{2}, CO and O\textsubscript{3}), ambient particulate matter (PM\textsubscript{10} and PM\textsubscript{2.5}) and meteorological data, including wind direction, wind speed, temperature, pressure and humidity were estimated with the use of hourly data. The wind directions influencing the sampling site across the five individual sampling periods were investigated. They were classified as follows: 337.5° < \(\leq\) 22.5°, 22.5° < \(\leq\) 67.5°, 67.5° < \(\leq\) 112.5°, 112.5° < \(\leq\) 157.5°, 157.5° < \(\leq\) 202.5°, 202.5° < \(\leq\) 247.5°, 247.5° < \(\leq\) W < 292.5°, 292.5° < \(\leq\) NW < 337.5° (Sulaymon et al. 2021).

### 2.3. Multilayer perceptron (MLP) neural network analysis

The MLP analysis is a feed-forward machine learning method, which designs for predicting one or more output variables using input variables (Desai and Shah 2021). The MLP model consists of three main components, i.e., (1) the input layer, (2) the hidden layer and (3) the output layer (figure S2). In general, the MLP model could use interconnected layer of artificial neurons with the input of data to produce a set of output data. Then, the MLP model could train the neural network through a back-propagation process (Simões Hoffmann et al. 2020). The hidden layer of the MLP model has demonstrated its predictive applicability in output of ambient air quality with the inputs having non-linear impacts (Desai and Shah 2021).

At the beginning step, the data for input and output variables are standardized in order to eliminate potential bias resulted from the range of variance and dimensional differences within the dataset (Arias del Campo et al. 2021). Each variable is standardized using z-score method that subtracts the mean observed value and then divided by the standard deviation (Li et al. 2020, Arias del Campo et al. 2021). Then, nine independent variables including NO, NO\textsubscript{2}, SO\textsubscript{2}, CO, wind direction, wind speed, temperature, pressure and humidity are fed in the input layer to the hidden layer for predicting dependent PM\textsubscript{2.5} as output variable. Six independent variables including NO, NO\textsubscript{2}, wind direction, wind speed, temperature, pressure and humidity are fed in the input layer to the hidden layer for predicting dependent O\textsubscript{3} as output variable. The equation for the hidden layer is described as follows:

\[
Z_j = H\left(\sum_{i=1}^{n} W_{ij}^G x_i + W_{0j}^G\right)
\]  

(1)

Where \(W_{0j}^G\) is the weight of the neuron between the input and hidden layer and \(W_{0j}\) an activation constant for neuron \(j\). The activation function \(H\) is often non-linear.

The hidden layer transfers a response onto the output layer through activation function \(H\). The functions of sigmoid and hyperbolic tangent (TanH) are tested in this study because these functions are appropriate for continuous dependant variables (Borlaza et al. 2021). The scaled conjugate and stochastic gradient descent optimization algorithms were applied in this study to obtain the optimal weights for both the input and output layers (Arias del Campo et al. 2021). The dataset were grouped into the training set (70% of the dataset) and testing set (30% of the dataset), respectively. For each run, the data in training set are responsible for train the MLP, while data in the testing set is adopted for monitoring errors during the training step independently. In the process of training step, the MLP is continually developed and verified until the weighting values between the nodes accurately predict the outcome (i.e. minimal possible errors). A step of stopping rules is utilized to terminate the training of the MLP if any of following scenarios occurs for preventing the model from overfitting. The following scenarios are described as follows: (1) there is no decrease in prediction error, (2) the maximum training time (30 min) is reached, (3) the minimum relative change (0.0001) in the training error is found and (4) the minimum relative change in the training error ratio (0.001) is observed. A maximum of 1000 data passes (epochs) are kept in memory before the training step is completed. Further, the results are verified in
the testing step to check the performance of the MLP model by comparing its forecasting capability on data points with those in the testing set. The MLP neural network analysis was carried out in this study by IBM SPSS Statistics for Windows, version 20 (IBM Corp., Armonk, NY, USA).

3. Results

3.1. Trends in ambient concentration

The daily concentrations of PM$_{2.5}$, O$_3$, PM$_{10}$, CO, SO$_2$, NO, and NO$_2$ collected in five individual sampling periods across five years were estimated using hourly data (figure S3 and table S1). At the sampling site (Milu Park), the lowest mean concentration of PM$_{2.5}$ was observed in 2018 (37.8 $\mu$g m$^{-3}$, n = 40), followed by 2019 (60.6 $\mu$g m$^{-3}$, n = 50), 2016 (67.3 $\mu$g m$^{-3}$, n = 51), 2020 (67.2 $\mu$g m$^{-3}$, n = 51) and 2017 (72.5 $\mu$g m$^{-3}$, n = 50) (table 1). Daily PM$_{2.5}$ concentrations in 72 of 242 samples were higher than the limit level of 24-hour PM$_{2.5}$ value standard (75 $\mu$g m$^{-3}$) (figure 1(A)). Similar mean levels of PM$_{2.5}$ were found in other two sampling sites (Jiugong and Yizhuang Development Zone) during the same periods across five years. At these two sampling sites, higher mean levels of PM$_{2.5}$ were observed in 2016 and 2017 ($\sim$80 $\mu$g m$^{-3}$) compared to those in 2018, 2019 and 2020 ($\sim$70 $\mu$g m$^{-3}$).

For the sampling site (Milu Park), the daily average concentration of PM$_{10}$ during the sampling period in 2016–2020 was 109.9 $\mu$g m$^{-3}$ (n = 51), followed by 103.6 $\mu$g m$^{-3}$ (n = 50), 76.3 $\mu$g m$^{-3}$ (n = 40), 87.6 $\mu$g m$^{-3}$ (n = 50) and 53.1 $\mu$g m$^{-3}$ (n = 51), respectively (table 1). The mean levels of PM$_{10}$ at Milu Park during the sampling periods across five years were comparable with those observed in other two sampling periods (Jiugong and Yizhuang Development Zone) (figure 1(B)).

The highest mean level of daily O$_3$ at the sampling site (Milu Park) during the sampling period in 2020 was found to be 46.6 $\mu$g m$^{-3}$ (n = 51), followed by 37.5 $\mu$g m$^{-3}$ (2018, n = 40), 18.7 $\mu$g m$^{-3}$ (2019, n = 50), 16.7 $\mu$g m$^{-3}$ (2016, n = 51) and 14.6 $\mu$g m$^{-3}$ (2017, n = 50) (table 1). The mean levels of daily O$_3$ at other two sampling sites were within the range of 30–40 $\mu$g m$^{-3}$ (table 1). Increases in daily O$_3$ level during lockdown period in 2020 ($\sim$46 $\mu$g m$^{-3}$) were observed at these three sampling sites relative to those observed in the years of 2016–2019 ($\sim$20–40 $\mu$g m$^{-3}$) (figure 1(C)).

Daily SO$_2$ levels in all samples at the three sampling sites were lower than the limit level of 24-hour value for SO$_2$ (150 $\mu$g m$^{-3}$) (figure 1(D)). At the Milu Park, the mean level of SO$_2$ during the sampling period in 2020 was recorded as 3.3 $\mu$g m$^{-3}$ (n = 51), which was relatively lower than that observed in 2016 (11.9 $\mu$g m$^{-3}$, n = 51), 2017 (26.0 $\mu$g m$^{-3}$, n = 50), 2018 (9.8 $\mu$g m$^{-3}$, n = 40) and 2019 (6.9 $\mu$g m$^{-3}$, n = 50). Mean levels of SO$_2$ at other two sampling periods (Jiugong and Yizhuang Development Zone) were in the range of 4–30 $\mu$g m$^{-3}$, which was comparable with those observed in Milu Park.

A higher mean level of NO$_x$ (i.e., NO + NO$_2$) was observed during the sampling period in 2020 (67.6 $\mu$g m$^{-3}$) relative to that in 2016 (41.7 $\mu$g m$^{-3}$), 2017 (47.0 $\mu$g m$^{-3}$), 2018 (48.7 $\mu$g m$^{-3}$) and 2019 (25.3 $\mu$g m$^{-3}$), respectively (table 1). Though the data of NO were absent at other two sampling periods (Jiugong and Yizhuang Development Zone), the mean level of NO$_2$ were similar as those observed in Milu Park, ranging from 30–60 $\mu$g m$^{-3}$.

At Milu Park, the mean daily levels of CO were recorded as 1.3 mg m$^{-3}$, 1.6 mg m$^{-3}$, 3.5 mg m$^{-3}$, 6.5 mg m$^{-3}$ and 0.9 mg m$^{-3}$ during the sampling period in 2016–2020, respectively (figure 1(F)). Extremely high daily concentrations of CO for all samples (n = 50) were observed during the sampling period in 2019, which all exceeded the limit levels of 24 h values for CO (4 mg m$^{-3}$). Yet, the mean level of CO decreased to 0.9 mg m$^{-3}$ during the sampling period in 2020. In contrast, the mean levels of CO were in the range of 0.8–1.3 mg m$^{-3}$ at other two sampling sites (table 1). No increases in mean daily levels of CO were found at other two sampling sites across five individual sampling periods.

3.2. Meteorological parameters

Daily meteorological parameters including temperature, pressure, humidity, wind speed and direction were summarized and presented in table S2. Average daily pressure during individual sampling periods across five years showed similar levels, ranging from 1015 hPa to 1026 hPa. Mean daily temperatures varied from $-0.8^\circ$C to 2.6 $^\circ$C. The mean levels of humidity including hourly and daily data during the sampling period in 2016–2019 were in the range of 28%–32%, which were lower than those in 2020 (59%). For wind speed, the mean level during the sampling period in 2020 was found to be $\sim2.5$ m s$^{-1}$, which is greater than those observed in the other four years (1.4–1.7 m s$^{-1}$) (figure S4). Higher wind speed seen during the sampling period in 2020 could be attributed to a dominant westerly wind, while lower wind speed related to wind directions during the sampling periods in the other four years (i.e., 2016–2019) were northerly or northwesterly (table S2).
Table 1. Concentration for the seven daily air pollutants (PM$_{2.5}$, O$_3$, PM$_{10}$, CO, SO$_2$, NO, and NO$_2$) across the five periods (January 20 to March 10, 2016, January 20 to March 10, 2017, January 20 to March 10, 2018, January 20 to March 10, 2019, and January 20 to March 10, 2020) at three sampling sites in Beijing, China.

| Site                                | Year | PM$_{2.5}$ (µg m$^{-3}$) | PM$_{10}$ (µg m$^{-3}$) | O$_3$ (µg m$^{-3}$) | SO$_2$ (µg m$^{-3}$) | NO$_x$ (µg m$^{-3}$) | NO (µg m$^{-3}$) | CO (mg m$^{-3}$) |
|-------------------------------------|------|--------------------------|-------------------------|---------------------|---------------------|-------------------|----------------|----------------|
| Milu Park (39.78N, 116.47E)        | 2016 | 67.3 ± 2.0               | 109.9 ± 3.2             | 16.7 ± 0.4          | 11.9 ± 1.3          | 24.8 ± 0.9       | 16.9 ± 1.3     | 1.3 ± 0.03     |
|                                     | 2017 | 72.5 ± 2.0               | 103.6 ± 2.4             | 14.6 ± 0.7          | 26.0 ± 2.9          | 71.0 ± 1.1       | 21.0 ± 1.3     | 1.6 ± 0.02     |
|                                     | 2018 | 37.8 ± 1.5               | 76.3 ± 1.7              | 37.5 ± 0.9          | 9.8 ± 2.2           | 38.4 ± 1.9       | 10.3 ± 2.0     | 3.5 ± 0.09     |
|                                     | 2019 | 60.6 ± 1.7               | 87.6 ± 1.3              | 18.7 ± 0.4          | 6.9 ± 0.6           | 16.3 ± 0.3       | 9.1 ± 0.6      | 6.5 ± 0.02     |
|                                     | 2020 | 67.2 ± 1.7               | 53.1 ± 1.3              | 46.6 ± 0.8          | 3.3 ± 0.1           | 33.8 ± 0.6       | 33.8 ± 0.8     | 0.9 ± 0.02     |
| Jiugong, Daxing Distric (39.90N, 116.40E) | 2016 | 111.1 ± 4.0              | 98.3 ± 2.2              | 41.6 ± 0.8          | 29.7 ± 1.5          | 50.0 ± 0.8       | N. A.          | 2.0 ± 0.06     |
|                                     | 2017 | 79.9 ± 1.3               | 111.9 ± 2.0             | 42.2 ± 0.7          | 23.1 ± 1.9          | 56.0 ± 1.1       | N. A.          | 1.2 ± 0.04     |
|                                     | 2018 | 50.9 ± 1.0               | 93.6 ± 1.7              | 40.2 ± 0.8          | 9.9 ± 1.2           | 42.2 ± 0.9       | N. A.          | 1.1 ± 0.09     |
|                                     | 2019 | 67.2 ± 1.1               | 107.3 ± 1.3             | 34.9 ± 0.8          | 5.9 ± 0.4           | 44.6 ± 1.1       | N. A.          | 1.0 ± 0.04     |
|                                     | 2020 | 67.8 ± 1.1               | 45.8 ± 0.8              | 49.5 ± 0.8          | 2.9 ± 0.5           | 30.6 ± 0.9       | N. A.          | 1.0 ± 0.03     |
| Yizhuang Development Zone (39.90N, 116.40E) | 2016 | 88.2 ± 3.0               | 98.0 ± 2.2              | 42.1 ± 0.7          | 24.5 ± 1.3          | 43.0 ± 1.1       | N. A.          | 1.3 ± 0.03     |
|                                     | 2017 | 87.7 ± 2.3               | 102.3 ± 2.0             | 46.3 ± 0.5          | 22.4 ± 2.9          | 57.6 ± 1.5       | N. A.          | 1.3 ± 0.02     |
|                                     | 2018 | 58.3 ± 1.8               | 69.9 ± 1.7              | 42.9 ± 0.6          | 10.8 ± 2.2          | 46.7 ± 1.3       | N. A.          | 0.9 ± 0.09     |
|                                     | 2019 | 57.9 ± 1.7               | 100.0 ± 2.3             | 40.6 ± 0.3          | 6.5 ± 0.5           | 44.3 ± 1.3       | N. A.          | 0.8 ± 0.02     |
|                                     | 2020 | 72.6 ± 2.1               | 40.7 ± 0.4              | 47.3 ± 0.5          | 3.9 ± 0.1           | 33.4 ± 1.1       | N. A.          | 0.9 ± 0.02     |

N.A.: not available.
3.3. Performance of MLP neural network analysis

With some interactions between PM$_{2.5}$ mass with sources of NO, NO$_2$, SO$_2$, and CO, as well as wind direction, wind speed, temperature, pressure and humidity (Tang et al 2020), MLP neural network analysis yielded a good performance for predicting PM$_{2.5}$ mass for the dataset obtained at Milu Park (figure 2). In modeling PM$_{2.5}$ mass, the activation function in hidden layer is TanH, while the area under receiver operating characteristic curve (ROC) ($0.9$) is higher than the threshold goodness of fit metrics of the model ($0.7$). Though MLP method did not fully capture some peaks of the measured PM$_{2.5}$ (figure 2), an optimization equation between modeled PM$_{2.5}$ using the inputs of NO, NO$_2$, SO$_2$, CO, wind direction, wind speed, temperature, pressure and humidity through MLP neural network analysis against measured PM$_{2.5}$ was estimated to be $y = 0.97x$ ($R^2 = 0.983$, $p < 0.001$) (figure 3(A)).
Since nonlinear interactions are observed between O₃ with NO and NO₂ (Ren et al 2021), as well as wind direction, wind speed, temperature, pressure and humidity, the equation between modeled O₃ equation with the inputs of NO, NO₂, wind direction, wind speed, temperature, pressure and humidity and measured O₃ equation was estimated to be $y = 0.82x$ ($R^2 = 0.905, p < 0.001$) (figure 3(B)). The activation function for modeling O₃ concentration in hidden layer is TanH and the area under receiver operating characteristic curve (ROC) (0.8) is greater than the threshold goodness of fit metrics of the model (0.7).

4. Discussion

In response to outbreak of the coronavirus 2019 (COVID-19) pandemic, Beijing government implemented restrictions on socio-economic activities including stay-at-home orders and limited non-essential travel starting in January 2020 (Hua et al 2021, Shi et al 2021). Consequently, emissions associated with business activity and
mobile source were markedly reduced and air pollution in urban area decreased correspondingly (Hua et al 2021). For the three sampling sites, the decline in mean concentrations of PM$_{10}$, SO$_2$, and CO in 2020 compared with those observed in the same period within the years of 2016–2019 was linked with emission reductions in primary sources including mobile source due to COVID-19 lockdown restrictions (Bekbulat et al 2021, Hua et al 2021). Yet, little variations in the mean PM$_{2.5}$ levels were observed between 2019 and 2020 at the three sampling sites though primary emissions associated with socioeconomic activities were reduced due to COVID-19 lockdown restrictions in 2020 at urban area of Beijing. This result could be attributed to the high contributions of secondary pollutants to PM$_{2.5}$ in winter during COVID-19 lockdown periods (Liu et al 2016, Xu et al 2018).

Figure 3. Relationship between measured concentrations of PM$_{2.5}$ (A) and O$_3$ (B) and modeled concentrations using multilayer perceptron neural network method of PM$_{2.5}$ and O$_3$ at a suburban site (Milu Park) of Beijing during the periods from January 20 to March 10, 2016, January 20 to March 10, 2017, January 20 to March 10, 2018, January 20 to March 10, 2019, and January 20 to March 10, 2020, respectively.
Increase in mean levels of $O_3$ at the three sampling sites were observed in 2020 relative to those observed in the same period within the years of 2016–2019, which is in consistent with several prior findings that enhanced level of $O_3$ occurred in Beijing and the surrounding areas during COVID-19 lockdown periods (Shi et al. 2021). Meanwhile, the mean levels of NO and NO$_2$ were observed to increase during COVID-19 lockdown periods. The enhancement of $O_3$ levels observed at the three sampling sites was not resulted from the reduced NO$_2$ levels, which differed from previous findings conducted by Li and his colleagues with the Goddard Earth Observing System Chemical Transport Model (Li et al. 2021). The enhanced levels of O$_3$ and NO$_2$ observed at our sampling sites may be ascribed to regional transport (Liu et al. 2021).

Multiple-layer neural network method is proven to be superiority in estimations of PM$_{2.5}$ and O$_3$, especially with limited input variables (Hung et al. 2020, Liao et al. 2021). However, the inaccurate selection of input variable could result in the prediction error of 30%–60% (Lin et al. 2017). For selections of input variables to model PM$_{2.5}$ mass, we input different combinations of variables including air pollutants (NO, NO$_2$, SO$_2$, and CO) with meteorological parameters and without meteorological parameters. The output of PM$_{2.5}$ estimates with the inputs of air pollutants (NO, NO$_2$, SO$_2$, and CO) and meteorological parameters can improve the prediction precision by 30% compared to those with the inputs without meteorological parameters. While in modeling O$_3$ level, the output of $O_3$ estimates with inputs of air pollutants (NO, NO$_2$, SO$_2$, and CO) and meteorological parameters could reduce the prediction precision by 15% relative to those with the inputs of air pollutants (NO and NO$_2$) and meteorological parameters. The optimization performance of MLP neural network analysis achieved better predictions of PM$_{2.5}$ and $O_3$ levels versus measured levels of PM$_{2.5}$ and $O_3$ with $R^2$ values of 0.983 and 0.905, which could explained 97% of measured PM$_{2.5}$ and 82% of measured $O_3$ on average respectively. We also performed the predictions of PM$_{2.5}$ and $O_3$ at other two sampling sites (Jiugong, Daxing District and Yizhuang Development Zone) with the input of air quality data obtained from Beijing Municipal Environmental Monitoring Center and meteorological parameters obtained from Milu sampling site. The prediction precisions (∼50% prediction error) of PM$_{2.5}$ and $O_3$ at these two sampling sites were not acceptable. A possible cause was the measurement loss of meteorological parameters for their own sampling sites. Our results suggest that MLP neural network analysis provides flexibility in predicting PM$_{2.5}$ and $O_3$ with the limited input variables. This method could be applied in reconstructing dataset of PM$_{2.5}$ and $O_3$ at suburban sites of Beijing in past periods when the network of PM$_{2.5}$ and $O_3$ measurement were absent and other air pollutants (NO, NO$_2$, SO$_2$, and CO) were available.

5. Conclusions

Seven hourly and daily air pollutants (PM$_{2.5}$, $O_3$, PM$_{10}$, CO, SO$_2$, NO, and NO$_2$) and meteorological parameters (wind direction, wind speed, temperature, pressure and humidity) were measured from January 20 to March 10 in the years of 2016–2020 at a suburban site in Beijing. The mean levels of three pollutants, PM$_{10}$, CO and SO$_2$ were observed to be lower from January 20 to March 10, 2020 (COVID-19 lockdown period), than those in the four past years (2016–2019) due to lockdown restrictions on socioeconomic activities. Although the contributions of primary source emission decreased in urban areas during the COVID-19 lockdown period, PM$_{2.5}$ mass exhibited little variations across the five sampling periods. This result suggests that secondary formation processes dominate the fine particle mass at the sampling sites. The increases in levels of O$_3$ and NOx were observed from January 20 to March 10, 2020, versus the four other sampling periods due to impacts of regional sources. (MLP) neural network analysis yielded relative high precisions of PM$_{2.5}$ and $O_3$ estimates with limited inputs variables, which explained 97% of measured PM$_{2.5}$ mass and 82% of measured $O_3$ level, respectively.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Author contributions

W L and Q L conceived and designed the experiments; W L and Y Z performed the experiments; W L, Y Z and Q L analyzed the data; W L and Q L wrote the paper. All authors have read and agreed to the published version of the manuscript.
Conflicts of interest

There are no conflicts of interest to declare.

ORCID iDs

Qingyang Liu © https://orcid.org/0000-0003-3193-5856

References

Arias del Campo F, Guevara Neri M C, Vergara Villegas O O, Cruz Sánchez V G, Ochoa Domínguez H D J and Garcia Jiménez V 2021 Auto-adaptive mul- ter layer perceptron for univariate time series classification. Expert Syst. Appl. 181 115147

Bekbudat B, Apte J S, Miller D B, Robinson A L, Wells K C, Presto A A and Marshall J D 2021 Changes in criteria air pollution levels in the US before, during, and after Covid-19 stay-at-home orders: evidence from regulatory monitors. Sci. Total Environ. 769 146693

Borlaza L S, Weber S, Jaffezo J L, Houdier S, Slama R, Rieux C, Albinet A, Micallef S, Trebluchon C and Uzu G 2021 Disparities in particulate matter (PM10) origins and oxidative potential at a city scale (Grenoble, France)—2: sources of PM10 oxidative potential using multiple linear regression analysis and the predictive applicability of multilayer perceptron neural network analysis Atmos. Chem. Phys. 21 9719–39

Cheng X, Zhao T, Gong S, Xu X, Han Y, Yin Y, Tang L, He H and He J 2016 Implications of East Asian summer and winter monsoons for interannual aerosol variations over central-eastern China Atmos. Environ. 129 216–28

D’Antoni D, Auyeung V, Walton H, Fuller G W, Grieve A and Weinman J 2019 The effect of evidence and theory-based health advice accompanying smartphone air quality alerts on adherence to preventative recommendations during poor air quality days: a randomised controlled trial Environ. Int. 124 216–35

Desai M and Shah M 2021 An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN) Clin. eHealth, 4 1–11

Eeftens M et al 2012 Development of land use regression models for PM2.5, PM10, absorbance, PM10, and PM2.5 in 20 European study areas; results of the ESCAPE project Environ. Sci. Technol. 46 11195–205

Han J, Zhang Y, Foy B, de, Shang J, Schauer J J, Mei X, Salaymon I D and Han T 2021 Quantitative estimation of meteorological impacts and the COVID-19 lockdown reductions on NO2 and PM2.5 over the Beijing area using generalized additive models (GAM) J. Environ. Manage. 291 112676

Huang C, Hu J, Xue T, Xu H and Wang M 2021 High-resolution spatiotemporal modeling for ambient PM2.5 exposure assessment in China from 2013 to 2019 Environ. Sci. Technol. 55 2152–62

Hung W T, Lu C H, Alessandrini S, Kumar R and Lin C A 2020 Estimation of PM2.5 concentrations in New York State: understanding the influence of vertical mixing on surface PM2.5 using machine learning Atmosphere 11 1303

LaCount M D, Haeuber R A, Macy T R and Murray B A 2021 Reducing power sector emissions under the 1990 clean air act amendments: a retrospective on 30 years of program development and implementation Atmos. Environ. 245 118012

Li K et al 2021 Ozone pollution in the North China Plain spreading into the late-winter haze season P. Natl. Acad. Sci. 118 e2015797118

Li X, Tripe D, Malone C and Smith D 2020 Measuring systemic risk contribution: the leave-one-out z-score method Financ. Res. Lett. 36 101316

Liang D, Wang Y Q, Wang Y J and Ma C 2019 National air pollution distribution in China and related geographic, gaseous pollutant, and socio-economic factors Environ. Pollut. 250 998–1009

Liao K, Huang X, Dang H, Ren Y, Zuo S and Duan C 2021 Statistical approaches for forecasting primary air pollutants: a review Atmosphere 12 686

Lin C K, Chang Y T, Lee F S, Chen S T and Christiani D 2021 Association between exposure to ambient particulate matters and risks of autism spectrum disorder in children: a systematic review and exposure-response meta-analysis Environ. Res. Lett. 16 063003

Lin C K, Hsu Y T, Brown K D, Polkarell B, Wei Y and Chen S T 2020 Residential exposure to petrochemical industrial complexes and the risk of leukemia: a systematic review and exposure-response meta-analysis Environ. Pollut. 258 113476

Lin C K, Hung H Y, Christiani D C, Forastiere F and Lin R T 2017 Lung cancer mortality of residents living near petrochemical industrial complexes: a meta-analysis Environ. Health 16 101

Lin C K, Lin R T, Chen P C, Wang P, De Marcellis-Warin N, Zigler C and Christiani D C 2018 A global perspective on sulfur oxide controls in coal-fired power plants and cardiovascular disease Sci. Rep. 8 2611

Liu Q, Baumgartner J, de Foy B and Schauer J J 2019 A global perspective on national climate mitigation priorities in the context of air pollution and sustainable development City Environ. Interact. 1 100003

Liu Q, Baumgartner J, Zhang Y, Liu Y, Sun Y and Zhang M 2014a Oxidative potential and inflammatory impacts of source apportioned ambient air pollution in Beijing Environ. Sci. Technol. 48 12920–9

Liu Q, Baumgartner J, Zhang Y and Schauer J J 2016 Source apportionment of Beijing air pollution during a severe winter haze event and associated pro-inflammatory responses in lung epithelial cells Atmos. Environ. 126 28–35

Liu Q, Liu Y, Yin J, Zhang M and Zhang T 2014b Chemical characterizations and source apportionment of PM10 during Asian dust storm and non-dust storm days in Beijing Atmos. Environ. 91 85–94

Liu Y, Yang Z, Liu Q, Qi X, Qi J, Zhang S, Wang X, Jia J and Zhu M 2021 Study on chemical components and sources of PM2.5 during heavy air pollution periods at a suburban site in Beijing of China Atmos. Pollut. Res. 12 188–99

Luo M, Shephard M W, Cady-Pereira K E, Henze D K, Zhu L, Bash J O, Pinder R W, Capps S L, Walker J T and Jones M R 2015 Satellite observations of tropospheric ammonia and carbon monoxide: global distributions, regional correlations and comparisons to model simulations Atmos. Environ. 106 262–77

Miao Y, Guo J, Liu S, Zhao C, Li X, Zhang G, Wei W and Ma Y 2018 Impacts of synoptic condition and planetary boundary layer structure on the trans-boundary aerosol transport from Beijing-Tianjin-Hebei region to northeast China Atmos. Environ. 181 1–11

Ren J, Hao Y, Simayi M, Shi Y and Xie S 2021 Spatiotemporal variation of surface ozone and its causes in Beijing, China since 2014 Atmos. Environ. 260 118556

Shi Z, Song C, Liu B, Lu G, Xu J, Vu T V, Elliott R J R, Li W, Blass W J and Harrison R M 2021 A abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns Sci. Adv. 7 eabd6696
Simões Hoffmann L F, Parquet Bizarria F C and Parquet Bizarria J W 2020 Detection of liner surface defects in solid rocket motors using multilayer perceptron neural networks Polym. Test. 88 106559
Sulaymon I D, Zhang Y, Hopke P K, Zhang Y, Hua J and Mei X 2021 COVID-19 pandemic in Wuhan: Ambient air quality and the relationships between criteria air pollutants and meteorological variables before, during, and after lockdown Atmos. Res. 250 105362
Sun P et al 2020 Impact of air transport and secondary formation on haze pollution in the Yangtze River Delta: in situ online observations in Shanghai and Nanjing Atmos. Environ. 225 117350
Tang L et al 2020 Iron and steel industry emissions and contribution to the air quality in China Atmos. Environ. 237 117668
de Vries W, Dobbertin M H, Solberg S, van Dobben H F and Schaub M 2014 Impacts of acid deposition, ozone exposure and weather conditions on forest ecosystems in Europe: an overview Plant Soil 380 1–45
Wang M, Sampson P D, Hu J, Kleeman M, Keller J P, Olives C, Srpiro A A, Vedal S and Kaufman J D 2016 Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM$_{2.5}$ Environ. Sci. Technol. 50 5111–8
Xiao Q, Chang H H, Geng G and Liu Y 2018 An ensemble machine-learning model to predict historical PM$_{2.5}$ concentrations in China from satellite data Environ. Sci. Technol. 52 13260–9
Xu X, Zhang H, Chen J M, Li Q, Wang X, Wang W, Zhang Q, Xue L, Ding A and Mellouki A 2018 Six sources mainly contributing to the haze episodes and health risk assessment of PM$_{2.5}$ at Beijing suburb in winter 2016 Ecotox Environ Safe 166 146–56
Xu X, Zhang W, Zhu C, Li J, Wang J, Li P and Zhao P 2021 Health risk and external costs assessment of PM$_{2.5}$ in Beijing during the ‘Five-year Clean Air Action Plan’ Atmos. Pollut. Res. 12 101089
Yang Z, Liu Q, Liu Y, Qi X and Wang X 2021 Cell cycle arrest of human bronchial epithelial cells modulated by differences in chemical components of particulate matter RSC Adv. 11 10582–91