A Machine-Learning-Based Classification Method for Meteorological Conditions of Ozone Pollution

Yang Cao1,2,3*, Xiaoli Zhao1, Debin Su3, Xiang Cheng1, Hong Ren4

1 Sichuan Meteorological Disaster Prevention Technology Center, Chengdu 610072, China
2 Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, Chengdu 610072, China
3 China Meteorological Administration Key Laboratory of Atmospheric Sounding, Chengdu University of Information Technology, Chengdu 610225, China
4 College of Resources and Environment, Chengdu University of Information Technology, Chengdu 610225, China

ABSTRACT

Ozone pollution is harmful to human health and ecosystem, which occurs in ecosystems and has occurred frequently in China in recent years, especially during the warm seasons. Meteorological conditions are among the important factors affecting the occurrence of ozone pollution. In this study, a classification method for meteorological conditions of ozone pollution levels based on a back propagation (BP) neural network was proposed to reflect the impact of meteorological conditions on the occurrence of ozone pollution. Ozone pollution was divided into three levels according to surface hourly ozone (O₃) concentrations and thus into three groups of meteorological conditions. The input physical parameters for the BP neural network were determined by evaluating the relationship between surface O₃ concentrations and meteorological parameters and precursors, including relative humidity, temperature, mixing layer height, precipitation, and nitrogen dioxide (NO₂) concentrations. The study area focused on 21 cities in Sichuan Province in southwestern China, which was divided into 12 BP classifiers according to the urban geographical location and sample number of each city, and a single BP classifier was trained for 21 cities. The classification results of the trained BP classifiers were verified by comparison to the observations. With 12 individual BP classifiers, the classification accuracy of all 21 cities was more than 60%, of which 18 cities were more than 70%, and 9 cities were more than 80%. With the single BP classifier, the classification accuracy of 20 cities was more than 60%, of which 18 cities were more than 70%, and 14 cities were more than 80%. Overall, the classification performance of the trained single model was better than trained 12 individual models. The classification method can comprehensively reflect the impact of meteorological conditions on the occurrence of ozone pollution.

Keywords: Ozone pollution, Meteorological conditions, Back propagation (BP) neural network, Classification method

1 INTRODUCTION

Ozone (O₃) in the troposphere is the product of complex oxidation reactions between volatile organic compounds (VOCs) and nitrogen oxides (NOx) in the presence of sunlight (Lu et al., 2018, 2020). Long-term exposure to high levels of surface ozone can adversely affect human health (Atkinson et al., 2016; Stowell et al., 2017; Nuvolone et al., 2018; Si and Tian, 2020). Surface ozone can also lead to corrosion and aging of building materials (Kucera and Fitz, 1995; Massey, 1999; Kuzmichev and Loboyko, 2016), as well as damage to vegetation, crops, and the ecological environment, such as reduction of crop productivity and inhibition of tree growth, which are issues that have attracted wider attention (Mills et al., 2011; Tai et al., 2014; Cailleret et al., 2018; Sharma et al., 2019; Juran et al., 2021). Due to accelerated urbanization and industrialization, as
well as the rapid increase in the volume of traffic, surface ozone pollution has occurred frequently in China in recent years, especially during warm seasons. O₃ has gradually become a primary pollutant in the atmosphere (Wang et al., 2019b; Shu et al., 2019; Xu et al., 2021; Li et al., 2021). As one of the most important detection methods to provide both surface ozone and vertical ozone structure, ozonesonde has been successfully developed by Chinese scholars and used to conduct long-term observation at local site (Zhang et al., 2014, 2021).

Timely and accurate predictions help decision-making agencies send out alarms and take effective measures and procedures to prevent or mitigate the occurrence of serious pollution events. The general public and all sectors of society may benefit from such useful advice to make reasonable and sensible arrangements for life, work and various social activities.

There are currently three types of prediction methods for ozone pollution from previous studies. (1) Statistical prediction models for O₃ concentration: These prediction models are established by comprehensively considering field measurements (e.g., meteorological and air quality data) that strongly affect the occurrence of ozone pollution (Pavón-Domínguez et al., 2014; Munir et al., 2015; Binaku and Schmeling, 2017; Nuñez-Alonso et al., 2019; Habeebullah, 2020; Cifuentes et al., 2021). This method is relatively simple, economical, and easy to implement, but it assumes that there is a linear relationship between meteorological elements and pollutant concentrations. (2) Numerical model systems for air quality are used to provide O₃ concentration predictions, which involve the use of chemical transport models and can simulate the dynamics of the atmosphere from the mathematical representation of different physical and chemical mechanisms (Hoshyaripour et al., 2016; Zeng et al., 2020; Sayeed et al., 2021). However, this method has a complex calculation process, high requirements for computer performance and high cost. (3) Artificial intelligence and machine learning methods have emerged in recent years as powerful tools for modeling and predicting in various fields, including air quality forecasting (Karimian et al., 2019; Wang et al., 2019a; Kumar et al., 2020; Davenport and Difenbaugh, 2021). The prediction method of O₃ concentrations based on machine learning mainly relies on the advantage that the machine learning method can effectively capture the hidden nonlinear characteristics in the change in atmospheric composition and can build a prediction model of atmospheric composition through characteristic variables (Mo et al., 2019; Aljanabi et al., 2020; Amato et al., 2020; Betancourt et al., 2021). Machine learning models trained with data from observations or physical models can produce reliable simulations without intensive high-end computing (Ojha et al., 2021). Previous studies focused mainly on the prediction of surface O₃ concentrations.

Surface O₃ concentrations are influenced by both meteorological conditions and emissions of precursors (Zhang et al., 2016; Yadav et al., 2016). When the emissions of precursors are relatively stable, meteorological conditions are the main factors affecting the occurrence of ozone pollution. Many studies focus on analyzing the relationship between weather conditions and O₃ concentrations, which can help better understand the formation and prediction of ozone pollution. It has been found that weather conditions with high temperatures, low humidity, high solar radiation, and weak diffusion environments are favorable for high O₃ concentrations (Jasaitis et al., 2016; Lu et al., 2019a, 2019b; Yang et al., 2020; Kim et al., 2021).

In this study, a classification method for meteorological conditions that are favorable for different ozone pollution levels based on a back propagation (BP) neural network was proposed to reflect the impact of meteorological conditions on the occurrence of ozone pollution. When the forecast meteorological parameters are used as the input parameters of the BP neural network, the occurrence of ozone pollution can be predicted. The data and associated preprocessing methods are presented in Section 2. Section 3 describes the methods and algorithms, including the analysis of the temporal and spatial distribution characteristics of surface O₃ concentrations in the study area, the determination of physical parameters that are closely related to the occurrence of ozone pollution, the establishment of a meteorological condition classification model for ozone pollution levels based on the BP neural network, and the verification of classification results. The results and analysis are illustrated in Section 4. Section 5 summarizes with conclusions and discussions.

2 STUDY AREA AND DATA

Due to intense emissions, complex terrain, and special meteorological conditions, Sichuan
Province in southwestern China suffers from severe ozone pollution in recent years (Yang et al., 2020; Chen et al., 2021). In this study, our study region covers the entire Sichuan Province (97°–109°E and 26°–35°N), including 21 prefecture-level cites. Fig. 1 shows the topographic distribution and location of cities in Sichuan Province, China, which is characterized by high values in the west and low values in the east, with elevations of 109 m to 7845 m.

Air quality monitoring data, automatic weather station (AWS) data, and China Meteorological Administration Land Surface Data Assimilation System (CLDAS) data were obtained for the study area. CLDAS is the only real-time operating system in the field of land data assimilation systems in China, which has integrated a large number of observations and can provide high-resolution, high-accuracy atmospheric forcing data and land surface model data (Liu et al., 2019; Sun et al., 2020; Liu et al., 2021).

Hourly ambient mass concentrations of surface pollutants, including O₃ and nitrogen dioxide (NO₂), in Sichuan Province for a 7-year (2015–2021) period were acquired from real-time data released by the air monitoring data center of the Ministry of Ecology and Environment of the People’s Republic of China. The number of observation sites in Sichuan Province increased from 38 in 2014 to 123 in 2021. The unit of O₃ and NO₂ concentrations is µg m⁻³. The number of observation sites in each city was small and mainly concentrated in the urban area; a prefecture-level city was taken as the research object in this study. The average hourly concentrations of surface pollutants from all effective observation sites in a city were used to represent the city’s pollution status.

The meteorological data for surface air temperature, relative humidity (RH), precipitation, wind speed (WS) and cloud fraction were acquired from CLDAS data for a 4-year (2018–2021) period, which had a temporal resolution of 1 hr and spatial resolution of 0.05° × 0.05° (approximately 5 km × 5 km). The atmospheric mixing layer height (MLH) was calculated from the temperature, RH, WS and cloud fraction data according to the Nozaki method (He et al., 2021). In addition, the gauge-observed hourly rainfall data for a 6-year (2015–2020) period were acquired from the Sichuan Meteorological Bureau, which was used to analyze the influence of precipitation on O₃ concentrations. Moreover, meteorological parameters and pollutant concentrations were matched at temporal and spatial scales to establish associations among variables.

![Fig. 1. Topographic distribution of Sichuan Province in southwestern China.](image-url)
3 METHODS

The key design features of the classification method for meteorological conditions corresponding to ozone pollution levels include four components: (1) analysis of temporal and spatial distribution characteristics of surface O₃ concentrations in the study area; (2) determination of physical parameters closely related to the occurrence of ozone pollution; (3) establishment of classification model for meteorological conditions of different ozone pollution levels based on BP neural network; (4) verification of the classification results.

3.1 Analysis of Ozone Distribution

The sample data for the BP training model were selected from the main ozone pollution periods, which were determined according to the characteristics of monthly and daily surface ozone distributions in the study area. The monthly and daily mean values, including O₃ concentrations, were estimated for a 3-year (2018–2020) period in 21 cities of Sichuan Province to analyze the spatial distribution characteristics.

3.2 Determination of Physical Parameters

Ozone pollution is influenced by both meteorological conditions and precursor contents. The NO₂ concentration is one of the important precursors affecting the formation of ozone. In this study, NO₂ concentration was considered a primary representative of precursors. Other studies have found that weather conditions with high temperature, low humidity, high solar radiation, and a weak diffusion environment are favorable for high O₃ concentrations (Jasaitis et al., 2016; Lu et al., 2019a, 2019b; Yang et al., 2020; Kim et al., 2021). The meteorological parameters considered in this study include RH, temperature, MLH, and precipitation. By evaluating the relationship between surface O₃ concentrations and meteorological parameters and precursors, the physical parameters that mainly affected ozone pollution were obtained as the input parameters for the BP training model. The correlation coefficient (CC) was calculated as follows (Kumar and Naseef, 2015):

\[
CC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}
\]

where \( n \) is the sample size, \( x \) is the O₃ concentration, \( y \) is the meteorological parameter or precursor, and \( \bar{x} \) and \( \bar{y} \) are the averages of \( x \) and \( y \).

In addition, the influence of precipitation on surface O₃ concentrations was analyzed by its scavenging effect, which was calculated based on the precipitation process. A precipitation process was discriminated according to the method adopted by Cao et al. (2020). The scavenging effect of surface O₃ concentrations by the precipitation process was calculated as follows:

\[
MSE = \frac{\sum_{i=1}^{n}(\rho_1 - \rho_2)}{\sum_{i=1}^{n}\rho_1} \times 100\%
\]

where \( n \) is the number of precipitation processes and \( \rho_1 \) and \( \rho_2 \) are the surface O₃ concentrations before and after the precipitation process, respectively. MSE tells the percentage of the pollutant being washed out of the atmosphere.

3.3 Classification Method for Meteorological Conditions of Ozone Pollution Levels

Fig. 2 presents an overview of the flow of the classification method for meteorological conditions of ozone pollution levels. The overall method used consisted of three main components: (1) the
3.3.1 Relationship between meteorological conditions and ozone pollution levels

With reference to the Technical Regulation on Ambient Air Quality Index in China, ozone pollution was divided into three levels according to hourly surface O₃ concentrations (≤ 160 µg m⁻³: excellent air quality; 160–200 µg m⁻³: good air quality; > 200 µg m⁻³: ozone pollution). Accordingly, the meteorological conditions are categorized into three corresponding levels. The relationship between meteorological conditions and ozone pollution levels is presented in Table 1. An hourly O₃ concentration of 0–160 µg m⁻³ is defined as level 1, and the corresponding meteorological conditions are considered very unfavorable to the occurrence of ozone pollution. An hourly O₃ concentration of 160–200 µg m⁻³ is defined as level 2, and the corresponding meteorological conditions are considered unfavorable to the occurrence of ozone pollution. An hourly O₃ concentration of > 200 µg m⁻³ is defined as level 3, and the corresponding meteorological conditions are considered conducive to the occurrence of ozone pollution.

3.3.2 Training

The sample data of input parameters for the BP training model, including meteorological parameters and precursor contents (e.g., NO₂ concentrations), were grouped together according to the relationship between meteorological conditions and ozone pollution levels, (2) training of the classifier based on the BP neural network, and (3) classification based on the trained BP classifier.

**Table 1.** The relationship between meteorological conditions and ozone (O₃) pollution levels.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Surface hourly O₃ concentrations (µg m⁻³)</th>
<th>Description of the corresponding meteorological conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>0–160</td>
<td>very unfavorable to the occurrence of ozone pollution</td>
</tr>
<tr>
<td>Level 2</td>
<td>160–200</td>
<td>unfavorable to the occurrence of ozone pollution</td>
</tr>
<tr>
<td>Level 3</td>
<td>&gt; 200</td>
<td>conducive to the occurrence of ozone pollution</td>
</tr>
</tbody>
</table>
to the corresponding ozone pollution levels (shown in Table 1). The study area was divided into multiple regions according to the spatial distribution characteristics of ozone pollution in 21 cities; each region has its own training model. In addition, a single BP training model for the 21 cities was trained. The classification performance of a single model will be compared with that of 12 individual models.

The BP neural network was used to train the classifiers, which was composed of input layer, hidden layer and output layer. The training of the BP neural network was through signal forward propagation and error backward propagation, and the adjustment of weight and threshold was repeated until the preset learning and training times, or the output error was reduced to the allowable value. The number of nodes in the input layer was determined by the number of input parameters (which is 5 in this study), including relative humidity, temperature, mixing layer height, precipitation, and NO2 concentrations in the previous night, and the number of nodes in the output layer was determined by the number of ozone pollution levels (which is 3 in this study). The number of hidden layers and nodes were comprehensively selected according to the training error and classification accuracy. The comparative tests were carried out to determine the number of hidden layers and nodes. Taking model of No. 1 as an example, comparison results showed that classification accuracies of one and two hidden layers were 67.4% and 86.5%. Therefore, the performance of the training model was better when the number of hidden layers was 2. In addition, as the number of hidden layer nodes increased, the training error decreased, but when the number of hidden layer nodes was too large, the overfitting problem of the models occurred. Therefore, to save computing resources and achieve ideal discrimination ability, it was appropriate to set the number of nodes in the first layer between 7 and 9 and the number of nodes in the second layer between 8 and 10. The mean squared error was used as the performance function of BP neural network. Each parameter was scaled to [0, 1] by minimum–maximum normalization in the training process. The activation function or transfer function of hidden layer and output layer were logarithmic sigmoid transfer function (logsig) and linear transfer function (purelin) respectively. The training function of BP neural network was gradient descent with momentum and adaptive learning rate backpropagation (traingdx).

In this study, the data of a major ozone pollution period from 2018 to 2020 were used for training and validation. After training, the obtained BP classifier acquired the “knowledge” to answer the question about the influence of meteorological conditions on ozone pollution.

3.3.3 Classification

With the trained BP classifier, when new meteorological parameters and precursor contents (e.g., NO2 concentrations) are provided, classification results can be obtained. The classification output would be either 1, 2, or 3, corresponding to very unfavorable, unfavorable, or conducive ozone pollution, respectively. For example, when new meteorological data arrive, which are used as input into the trained BP classifier, if the output is 1, it indicates that the meteorological conditions are very unfavorable to the occurrence of ozone pollution.

3.4 Verification of Classification Results

The performance of the trained BP classifier was qualitatively and quantitatively examined by comparison with observations. The classification accuracy was calculated to quantitatively verify the classification results, which was the percentage of correct samples in the total number of samples. A correct classification occurred when the classification result was consistent with the observation; otherwise, it was wrong. For example, when the classification result and observation are both 1, it is a correct classification.

4 RESULTS AND ANALYSIS

4.1 Analysis of Ozone Distribution

Figs. 3(a) and 3(b) present the monthly and daily distribution characteristics of surface O3 concentrations for a 3-year (2018–2020) period in Sichuan Province. The solid line of “Sichuan” is the average value of 21 cities in Sichuan Province. The dashed lines are three representative cities (Guangyuan, Chengdu, and Yibin) from the north to the south of Sichuan Province. The O3
concentrations from April to September were higher than those in other months, and two peaks appeared in May and August. The daily variation in O₃ concentrations was a “single peak single valley” type with a peak at 16:00 Beijing Time (BJT) and a valley at 08:00 BJT. The O₃ concentrations during the day (from 09:00 to 21:00 BJT) were higher than those at night (from 22:00 to 08:00 BJT the next day). The monthly and daily variation trends of O₃ concentrations were influenced by the changes in solar radiation (Liu et al., 2018; Zhao et al., 2022). Based on the analysis above, the main ozone pollution period of Sichuan Province was set from April to September every year and from 09:00 to 21:00 BJT every day.

Fig. 4 shows the spatial distribution of O₃ concentrations during the main ozone pollution period from 2018 to 2020 in Sichuan Province. There were 5 cities with O₃ concentrations larger

![Graph showing monthly and daily distribution characteristics of O₃ concentrations](image)

**Fig. 3.** (a) The monthly and (b) daily distribution characteristics of O₃ concentrations for a 3-year (2018–2020) period in Sichuan Province, China. The solid line of “Sichuan” shows the average value of 21 cities in Sichuan Province. The dashed lines show three representative cities (Guangyuan, Chengdu, and Yibin) from the north to the south of Sichuan Province.

![Spatial distribution of O₃ concentrations](image)

**Fig. 4.** Spatial distribution of O₃ concentrations in the main ozone pollution periods (set from April to September every year and from 9:00 to 21:00 BJT every day) from 2018 to 2020 in Sichuan Province, China.
than 90 µg m⁻³, including Meishan (97.34 µg m⁻³), Chengdu (94.82 µg m⁻³), Zigong (92.33 µg m⁻³), Ziyang (91.62 µg m⁻³) and Deyang (91.61 µg m⁻³). There were 4 cities with O₃ concentrations less than 70 µg m⁻³, including Guangyuan (69.26 µg m⁻³), Bazhong (67.03 µg m⁻³), A’ba (66.43 µg m⁻³) and Ganzi (69.22 µg m⁻³). The spatial distribution characteristics of O₃ concentrations might be affected by complex terrain, special climate conditions and the distribution of precursors in Sichuan Province (Yang et al., 2020; Chen et al., 2021). In addition, the average values of O₃ concentrations in Chengdu and Meishan were much higher than those in other cities in Sichuan Province, possibly due to the high traffic density and local emissions.

4.2 Determination of Physical Parameters

In the analysis of the relationship between hourly surface O₃ concentrations and meteorological parameters and precursor contents, Chengdu (the capital city) was used since more serious ozone pollution occurs there and it also represents more developed areas in Sichuan Province. Fig. 5 shows scatterplots of O₃ concentrations and meteorological parameters and precursor contents from April to September for a 3-year (2018–2020) period in Chengdu city. The colored bar indicates the probability density of the scattered points. The high density of scattered points indicates the aggregated distribution of meteorological parameters and precursor contents during the study.

![Fig. 5](https://example.com/fig5.png)

**Fig. 5.** Scatterplots of O₃ concentrations and meteorological parameters, including (a) temperature, (b) relative humidity, (c) mixing layer height and (d) NO₂ concentration in the previous night from April to September for a 3-year (2018–2020) period in Chengdu city, China. Night was defined as 22:00 to 08:00 BJT the next day. The colored bar indicates the probability density of the scattered points. The black dotted line indicates the fitted curve. The correlation coefficients between O₃ concentrations and meteorological parameters and precursor contents were calculated. Significance tests at the 0.01 level were carried out for the linear relationship between O₃ concentrations and four parameters. The p values of the significance test were less than 0.001.
Table 2. Mean scavenging effect (MSE) of O3 concentrations in 425 precipitation processes from January 2015 to December 2020 in Chengdu city, China.

<table>
<thead>
<tr>
<th>Maximum hourly rainfall (mm)</th>
<th>Cumulative rainfall (mm)</th>
<th>Rainfall duration (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Precipitation processes</td>
<td>MSE (%)</td>
</tr>
<tr>
<td>0.1–0.9</td>
<td>158</td>
<td>19.93</td>
</tr>
<tr>
<td>1.0–4.9</td>
<td>205</td>
<td>20.53</td>
</tr>
<tr>
<td>5.0–9.9</td>
<td>32</td>
<td>20.50</td>
</tr>
<tr>
<td>≥ 10.0</td>
<td>30</td>
<td>26.00</td>
</tr>
</tbody>
</table>

The temperature was mainly distributed at 18–28°C, RH was mainly distributed at 85–95%, MLH was mainly distributed at 0–3000 m, and the NO2 concentrations in the previous night were mainly distributed at 10–60 µg m⁻³. Night was defined as 22:00 to 08:00 BJT the next day according to the daily distribution of O3 concentrations (shown in Fig. 3(b)). The correlation coefficient between the hourly O3 concentrations and meteorological parameters was calculated. Moreover, the correlation coefficient between the daily maximum O3 concentrations and NO2 concentrations in the previous night was calculated. The O3 concentrations showed a negative correlation with RH and a positive correlation with temperature, MLH, and NO2 concentrations in the previous night during the study period, which were consistent with previous studies. Significance tests at the 0.01 level were carried out for the linear relationship between O3 concentrations and four parameters (temperature, RH, MLH, and NO2 concentration in the previous night). The p values of the significance test were less than 0.001.

In addition, precipitation had a significant influence on decreasing surface O3 concentrations. Table 2 shows the mean scavenging effect of surface O3 concentrations by 425 precipitation processes from January 2015 to December 2020 in Chengdu city according to maximum hourly rainfall, cumulative rainfall, and rainfall duration, which were divided into four levels for statistical analysis. Except for the level at which the rainfall duration was less than 1 hr, the mean scavenging effects at other levels of maximum rainfall, cumulative rainfall, and rainfall duration were greater than 19%. There was little difference in the mean scavenging effect in the four levels of maximum rainfall and cumulative rainfall. However, with the increase in rainfall duration, the mean scavenging effect gradually increased. The probably reason is cloud cover. The effect of clouds on solar radiation energy attenuation are not conducive to photochemical reaction. Under the background of precipitation, the longer the duration of rainfall is, the greater the negative influence on the generation of ozone. Moreover, the precursors are removed by precipitation, which will also affect the generation of ozone.

4.3 Classification Results for Meteorological Conditions of Ozone Pollution Levels

The spatial distribution characteristics of ozone pollution and the corresponding meteorological conditions are affected by topography, climate conditions, and precursor distribution. It is necessary to train the BP classifier for meteorological conditions of each ozone pollution level in each region based on urban geographical location and sample numbers. In addition, a single BP training model for the 21 cities was trained to compare with those individual models. Table 3 shows the sample numbers of three ozone pollution levels in the main ozone pollution periods (set from April to September every year and from 09:00 to 21:00 BJT every day) from 2018 to 2020 of 21 cities in Sichuan Province. According to the sample numbers of ozone pollution level 3, the cities were divided into "ozone pollution prone cities" (more than 100 samples of level 3 pollution, including Chengdu, Deyang, Zigong and Meishan) and "other cities". For the cities classified as "ozone pollution prone cities", an independent BP classifier was trained. The cities classified as "other cities" are further divided into groups according to their similarity in pollution characteristics and spatial distance, and a separate BP classifier is trained for each group. Furthermore, if a city in "other cities" has a large difference in pollution characteristics from adjacent cities, the pollution is more of a concern, and there is sufficient sample data, an independent BP classifier could be trained, such as Mianyang city. Overall, as shown in Table 4, the study area was...
Table 3. The sample numbers of three ozone pollution levels during the main ozone pollution periods from 2018 to 2020 in 21 cities in Sichuan Province, China.

<table>
<thead>
<tr>
<th>Cities</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chengdu</td>
<td>4501</td>
<td>415</td>
<td>266</td>
</tr>
<tr>
<td>Mianyang</td>
<td>4887</td>
<td>261</td>
<td>33</td>
</tr>
<tr>
<td>Deyang</td>
<td>4695</td>
<td>357</td>
<td>130</td>
</tr>
<tr>
<td>Zigong</td>
<td>4649</td>
<td>347</td>
<td>169</td>
</tr>
<tr>
<td>Meishan</td>
<td>4610</td>
<td>441</td>
<td>127</td>
</tr>
<tr>
<td>Nanchong</td>
<td>5056</td>
<td>123</td>
<td>10</td>
</tr>
<tr>
<td>Dazhou</td>
<td>4985</td>
<td>178</td>
<td>26</td>
</tr>
<tr>
<td>Guang’an</td>
<td>4922</td>
<td>244</td>
<td>20</td>
</tr>
<tr>
<td>Suining</td>
<td>4946</td>
<td>213</td>
<td>28</td>
</tr>
<tr>
<td>Ziyang</td>
<td>4837</td>
<td>309</td>
<td>36</td>
</tr>
<tr>
<td>Neijiang</td>
<td>4871</td>
<td>248</td>
<td>59</td>
</tr>
<tr>
<td>Yibin</td>
<td>4802</td>
<td>298</td>
<td>91</td>
</tr>
<tr>
<td>Luzhou</td>
<td>4793</td>
<td>308</td>
<td>87</td>
</tr>
<tr>
<td>Ya’an</td>
<td>5003</td>
<td>152</td>
<td>26</td>
</tr>
<tr>
<td>Leshan</td>
<td>4959</td>
<td>170</td>
<td>54</td>
</tr>
<tr>
<td>Liangshan</td>
<td>4885</td>
<td>110</td>
<td>14</td>
</tr>
<tr>
<td>Panzhihua</td>
<td>4898</td>
<td>88</td>
<td>16</td>
</tr>
<tr>
<td>Guangyuan</td>
<td>5132</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>Bazhong</td>
<td>4307</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>A’ba</td>
<td>5135</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ganzi</td>
<td>5126</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. The BP classifier training models used for 21 cities in Sichuan Province, China, and the classification accuracy of the trained BP classifiers for the test samples.

<table>
<thead>
<tr>
<th>Model number</th>
<th>Cities</th>
<th>Sample numbers of three levels for training and test</th>
<th>Classification accuracy of test samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1</td>
<td>Chengdu</td>
<td>Level 1: 266, Level 2: 266, Level 3: 266</td>
<td>Classification accuracy (%)</td>
</tr>
<tr>
<td>No. 2</td>
<td>Mianyang</td>
<td>33, 33, 33</td>
<td>30, 76.7</td>
</tr>
<tr>
<td>No. 3</td>
<td>Deyang</td>
<td>130, 130, 130</td>
<td>195, 72.3</td>
</tr>
<tr>
<td>No. 4</td>
<td>Zigong</td>
<td>169, 169, 169</td>
<td>255, 70.6</td>
</tr>
<tr>
<td>No. 5</td>
<td>Meishan</td>
<td>127, 127, 127</td>
<td>192, 71.9</td>
</tr>
<tr>
<td>No. 6</td>
<td>Nanchong, Dazhou, Guang’an</td>
<td>56, 56, 56</td>
<td>53, 66.0</td>
</tr>
<tr>
<td>No. 7</td>
<td>Suining, Ziyang, Neijiang</td>
<td>123, 123, 123</td>
<td>186, 72.6</td>
</tr>
<tr>
<td>No. 8</td>
<td>Yibin, Luzhou</td>
<td>178, 178, 178</td>
<td>268, 73.1</td>
</tr>
<tr>
<td>No. 9</td>
<td>Ya’an, Leshan</td>
<td>80, 80, 80</td>
<td>72, 63.9</td>
</tr>
<tr>
<td>No. 10</td>
<td>Liangshan, Panzhihua</td>
<td>30, 30, 30</td>
<td>18, 61.1</td>
</tr>
<tr>
<td>No. 11</td>
<td>Guangyuan, Bazhong</td>
<td>69, 69, 3</td>
<td>42, 90.5</td>
</tr>
<tr>
<td>No. 12</td>
<td>A’ba, Ganzi</td>
<td>160, 13, 0</td>
<td>60, 85.0</td>
</tr>
<tr>
<td>No. 13</td>
<td>21 cities</td>
<td>1421, 1274, 1195</td>
<td>1770, 69.6</td>
</tr>
</tbody>
</table>

In addition, the sample numbers of the three levels should be roughly the same when training the BP classifier. The levels with large sample numbers needed to be randomly screened and extracted. To ensure that the BP classifier training samples conformed to the actual distribution characteristics, random sampling was carried out in proportion according to the distribution frequency of O₃ concentrations in each interval. The specific extraction method was: first, determined the total number of samples (n) and the sample frequency at different O₃ concentration intervals (p, unit: %) at each level, and then calculated the product of p and n to obtain the number of...
samples to be randomly sampled at different O₃ concentration intervals at each level. The sample data were divided into training samples and test samples. The classification accuracy of the test samples for all 13 BP classifiers was greater than 60% (shown in Table 4). Among those classifiers, 8 BP classifiers had classification accuracies greater than 70%.

4.4 Verification of Classification Results

The regional ozone pollution event of 1 May 2021 in the study area was used to qualitatively examine the performances of the trained BP classifiers compared with observations. The hourly meteorological data obtained from CLDAS were used as the input parameters. As shown in Fig. 3(b), the peak O₃ concentrations occurred between 14:00 and 17:00 BJT every day. Fig. 6 shows the observations and classification results of 12 individual models and the single model for meteorological conditions of ozone pollution levels from 14:00 to 17:00 BJT. Some symbols were signed in the circle on the classification maps, which were helpful to understand the difference between the classification results and observations. A small plus sign and a bold plus sign indicate that the city is classified as one level higher and two levels higher, respectively. A small minus sign and a bold minus sign indicate that the city is classified as one level lower and two levels lower, respectively. Comparing Figs. 6(a), 6(b) and 6(c), it is seen that most cities with long-term observations of level 3 were successfully classified with 12 individual models and the single model, such as Chengdu, Deyang, Meishan, Leshan. Some cities in the Northeastern Sichuan with observations of level 1 were incorrectly classified as level 2 with 12 individual models and incorrectly classified as level 3 with the single model, such as Guangyuan, Bazhong, Nanchong, Dazhou. In addition, some cities with observations of level 2 were incorrectly classified as level 3 with 12 individual models and the single model, such as Ziyang, Suining, Yibin and Luzhou. In general, the consistency between the classification results of the single model and the observations was better than that of 12 individual models. The reason might be that training 21 cities into a single model could have more samples. Overall, the established classification method for meteorological conditions of ozone pollution levels based on the BP neural network can effectively indicate the occurrence of regional ozone pollution events.

The hourly data of the main ozone pollution periods from 1 April to 30 September 2021 were used to quantitatively verify the classification results of the trained BP classifiers by comparison with observations. According to observations, there were 9 cities with a large number of level 3 samples (indicating that meteorological conditions are conducive to the occurrence of ozone pollution), including Chengdu, Deyang, Mianyang, Meishan, Leshan, Neijiang, Zigong, Yibin and Luzhou, all of which were more than 20. Chengdu (79), Meishan (46) and Deyang (38) had the largest number of level 3 samples, which might be related to the high concentrations of NO₂ in these three cities. Fig. 7 shows the classification accuracy of trained 12 individual models (Fig. 7(a)) and the single model (Fig. 7(b)) in 21 cities in Sichuan Province during the research period. It was found that, with 12 individual models, the classification accuracy of all 21 cities was more than 60%, of which 18 cities were more than 70%, and 9 cities were more than 80%. And with the single model, the classification accuracy of 20 cities was more than 60%, of which 18 cities were more than 70%, and 14 cities were more than 80%. Furthermore, there were 17 cities with O₃ concentrations larger than 70 μg m⁻³ (shown in Fig. 4), which had greater chance of the occurrence of ozone pollution events. Among the 17 cities, the number of cities with classification accuracy over 80% were 5 for trained 12 individual models and 13 for the trained single model. Again, the classification performance of the trained single model was better than trained 12 individual models in Sichuan Province, China. As shown in Fig. 8, the classification accuracy of the trained 12 individual BP classifiers was higher for level 1, especially in cities with lower O₃ concentrations, such as A’ba, Ganzi, Guangyuan and Bazhong. The classification accuracy of level 3 was higher with the trained single model in cities with higher O₃ concentrations. The reason was that there was a large number of level 3 samples when training a single BP classifier for all 21 cities, which was helpful to improve the performance of the model.

5 SUMMARY

Regional ozone pollution events have occurred frequently in China in recent years, especially
Fig. 6. (a) Observations, classification results of (b) 12 individual models, and (c) the single model for meteorological conditions of ozone pollution levels at (1) 1400, (2) 1500, (3) 1600 and (4) 1700 BJT 1 May 2021 in Sichuan Province, China. The green, yellow and orange dots show level 1, level 2, and level 3, respectively. Some symbols are signed in the circle on the classification maps to understand the difference between the classification results and observations. A small plus sign and a bold plus sign indicate that the city is classified as one level higher and two levels higher, respectively. A small minus sign and a bold minus sign indicate that the city is classified as one level lower and two levels lower, respectively.
Fig. 7. Classification accuracy of trained (a) 12 individual models and (b) the single model in the main ozone pollution periods from 1 April to 30 September 2021 in 21 cities in Sichuan Province, China.

Fig. 8. Classification accuracy of trained (a) 12 individual models and (b) the single model for three levels in the main ozone pollution periods from 1 April to 30 September 2021 in 21 cities in Sichuan Province, China. The green, yellow and orange bars show level 1, level 2, and level 3, respectively.

in the warm seasons, which are harmful to human health, vegetation, ecosystems and building materials. Timely and accurate predictions help decision-making agencies raise alarms and take effective measures and procedures to prevent or mitigate the occurrence of serious pollution events. Meteorological conditions are one of the important factors affecting the occurrence of ozone pollution. The analysis of the relationship between weather conditions and ozone pollution
is helpful to forecast the occurrence of ozone pollution. In this study, a classification method for meteorological conditions of ozone pollution levels based on a BP neural network was proposed to indicate the occurrence of ozone pollution events in Sichuan Province, China, which suffered from serious ozone pollution due to complex terrain, special meteorological conditions, and intense emissions. Ozone pollution was divided into three levels according to surface hourly ozone concentrations and thus into three groups of meteorological conditions. The lower the level indicated that the meteorological conditions were more unfavorable to the occurrence of ozone pollution. Finally, the classification results were verified against observations.

To improve the performance of trained BP classifiers, the data of the main ozone pollution period should be selected as the sample data. The main ozone pollution period was determined according to the monthly and daily distribution characteristics of surface ozone in the research domain, which was set from April to September every year and from 09:00 to 21:00 BJT every day. The input physical parameters of the BP neural network were determined by evaluating the relationship between surface O3 concentrations and meteorological parameters and precursors, including relative humidity, temperature, mixing layer height, precipitation, and NOx concentrations in the previous night. The study area was divided into 12 BP classifiers according to the urban geographical location and sample number of 21 cities, and a single BP classifier was trained for 21 cities. The classification accuracy of the test data for 12 individual BP classifiers was greater than 60%, and 8 were greater than 70%. The classification accuracy of the test data for the single BP classifier was 69.6%. The comparison between classification results and observed validation data in the 21 cities showed that, with 12 individual BP classifiers, the classification accuracy of all 21 cities was more than 60%, of which 18 cities were more than 70%, and 9 cities were more than 80%. And with the single BP classifier, the classification accuracy of 20 cities was more than 60%, of which 18 cities were more than 70%, and 14 cities were more than 80%. Overall, the classification performance of the trained single model was better than trained 12 individual models in Sichuan Province.

The developed classification method can comprehensively reflect the impact of meteorological conditions on ozone pollution and has the potential to indicate the occurrence of ozone pollution, demonstrating a strong correlation with meteorological conditions. Additionally, a series of classification results in this study will provide a reference for further evaluating the impact of meteorological conditions on ozone pollution. When the forecast meteorological parameters are used as the input parameters of the model, the occurrence of ozone pollution can be predicted.

ACKNOWLEDGMENTS

This research was funded by Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, China (Grant No. SCQXJKQN202223), Key Laboratory of Atmospheric Chemistry, China Meteorological Administration (Grant No. 2020B05), and the National Natural Science Foundation of China (Grant No. 42007188). The authors greatly appreciate the immense assistance provided by Professor Xingang FAN at Western Kentucky University, USA. This manuscript was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at AJE.

REFERENCES


Si, Y., Tian, Y. (2020). The impact of ambient ozone pollution on pneumonia: A nationwide time-