Science of the Total Environment 864 (2023) 160928

Contents lists available at ScienceDirect

Science of the Total Environment

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

Estimation of near-surface ozone concentration and analysis of main weather situation in China based on machine learning model and Himawari-8 TOAR data

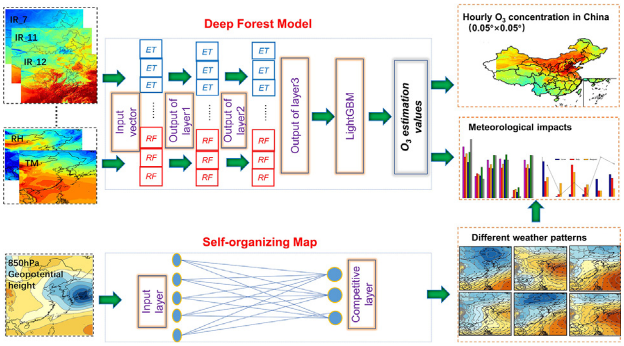
Bin Chen *, Yixuan Wang, Jianping Huang, Lin Zhao, Ruming Chen, Zhihao Song, Jiashun Hu

Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China
Collaborative Innovation Center for Western Ecological Safety, Lanzhou 730000, China

HIGHLIGHTS

- Linear and nonlinear relationships between Himawari-8 channels and O_3 were studied.
- Fine scale O_3 was obtained using the top of atmospheric radiation of Himawari-8.
- The hourly R^2 of interpretable deep learning model (TOAR-O_3) can reach 0.86–0.94.
- Relationship between weather patterns and O_3 was analyzed by self-organizing map.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Anastasia Paschalidou

Keywords:
Near-surface ozone concentration
Deep forest model
Pollution trend
Self-organizing map
Weather circulation patterns

ABSTRACT

Ozone (O_3) is an important greenhouse gas in the atmosphere. Stratospheric ozone protects human beings, but high near-surface ozone concentrations threaten environment and human health. Owing to the uneven distribution of ground-monitoring stations and the low time resolution of polar orbiting satellites, it is difficult to accurately evaluate the refinement and synergistic pollution of near-surface ozone in China. Besides, atmospheric circulation patterns also affect ozone concentrations greatly. In this study, a new generation of geostationary satellite is used to estimate the hourly near-surface ozone concentration with a spatial resolution of 0.05°. First, the Pearson correlation coefficient and maximum information coefficient were used to study the correlation between the top of atmospheric radiation (TOAR) of Himawari-8 satellite and O_3 concentration; seven TOAR channels were selected. Second, based on an interpretable deep learning model, the hourly ozone concentration in China from September 2015 to August 2021 was obtained using the TOAR-O_3 model. Finally, the self-organizing map method was used to determine six major summer weather circulation patterns in China. The results showed that (1) the near-surface O_3 concentration can be accurately estimated; the R^2 (RMSE: $\mu g/m^3$) values of the daily, monthly, and annual tenfold cross validation results were 0.91 (12.74), 0.97 (5.64), and 0.98 (1.75), respectively. The feature importance of the model showed that the temperature, TOAR, and boundary layer height contributed 38 %, 22 %, and 13 %, respectively. (2) The O_3 concentration showed obvious spatiotemporal difference and gradually increased from 10:00 to 15:00 (Beijing time) every day. In most areas of China, O_3 concentration had increased significantly. (3) The O_3 concentration in northern China was the highest under the circulation pattern of the Meiyu front over the Yangtze River Delta, while in southern China, it was the highest under the circulation pattern of the northeast cold vortex controlling most of China.

---

* Corresponding author at: Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China.
E-mail address: chenbin@lzu.edu.cn (B. Chen).

http://dx.doi.org/10.1016/j.scitotenv.2022.160928
Received 22 September 2022; Received in revised form 21 November 2022; Accepted 10 December 2022
Available online 17 December 2022
0048-9697/

1. Introduction

O_3 is an important trace gas in the atmosphere that prevents ultraviolet radiation and protects the Earth's organisms in the stratosphere, but high O_3 concentrations in the boundary layer can lead to air pollution (DeLang et al., 2021; Monks et al., 2015; Tang et al., 2021). Near-surface O_3 stimulates the respiratory mucosa and other lung tissues, causing pneumonia, chronic obstructive pulmonary disease, asthma, allergic rhinitis, and other respiratory diseases (Lei et al., 2019; Liu et al., 2018; Tian et al., 2020; Yang et al., 2019); O_3 pollution damages human health and plant growth (Chi et al., 2016; Feng et al., 2015; Lu et al., 2020). According to the IPCC 6th report, O_3 increased the global average surface temperature by 0.23 °C from 1750 to 2019 (IPCC, 2021). Air pollution caused by fine particulate matter has gradually decreased since the Chinese government took measures to prevent and control air pollution, however, O_3 pollution is on the rise (Dong et al., 2021; Gao et al., 2020; He et al., 2021b). An analysis of the characteristics of air pollution in China during the COVID-19 outbreak found that after the Chinese government imposed strict lockdown measures in Wuhan and restricted human activities, NO_2, an important precursor of O_3, decreased by 61.92 %, while the O_3 concentration increased (Dong et al., 2021), which was mainly due to seasonal variation in O_3 and the titration effect of NO (Wang et al., 2022a). The premise for near-surface O_3 pollution research is to obtain its concentration with high spatial and temporal resolution.

Ground-monitoring stations have been an effective way to monitor O_3 pollution for a long time. However, owing to the uneven distribution of ground stations and the low density of stations in sparsely populated areas, it is difficult to accurately evaluate the overall distribution characteristics of O_3. Satellite remote sensing has the advantage of continuity in space (Li et al., 2010a; Li et al., 2010b; Wei et al., 2021; Zhang and Li, 2015); the tropospheric ozone column density observed by the Ozone Monitoring Instrument (OMI) satellite during 2010–2012 was consistent with the surface ozone of six global atmospheric monitoring stations in China, with a mean correlation coefficient of 0.47 (Liu et al., 2019). However, it is worth noting that O_3 in the lowest layer of the OMI products did not directly correspond to O_3 in the boundary layer (Bai et al., 2016).

The machine learning methods show unique advantages for retrieving surface O_3 concentrations (Chen et al., 2022a; DeLang et al., 2021; Liu et al., 2020; Xue et al., 2020). Li et al. (2020) used a random forest generalized weighted model, developed from OMI data, to predict the maximum 8-h daily average (MDA8) concentration of O_3 over the Tibetan Plateau (TP), with a maximum R^2 of 0.76 and a minimum root mean square error (RMSE) of 14.41 $\mu g/m^3$. Zhang et al. (2020) estimated the monthly near-surface O_3 concentration in eastern China using OMI data based on the geo-weighted regression method and found that the R^2 was 0.81, with an absolute error (AE) of 7.38 $\mu g/m^3$. The surface O_3 concentration in East Asia was estimated using tropospheric ozone monitoring instrument (TROPOMI) data; the predicted R^2 was 0.65–0.78, and the RMSE was 19.6–24.7 % (Dong et al., 2021; Kang et al., 2021). Based on the two stage deep learning model, Moderate Resolution Imaging Spectroradiometer (MODIS) data were used to invert O_3 over the Beijing-Tianjin-Hebei region, the Yangtze River Delta, and Guangdong, Hong Kong and Macao, the maximum R^2 of the model was 0.78, and the minimum RMSE was 18.35$\mu g/m^3$ (Luo et al., 2022). Based on TROPOMI data, the O_3 concentration in North China in 2019 was estimated using the geo-intelligent light gradient lifting method with spatio-temporal correlation, and the R^2 can reach 0.912 (Chen et al., 2022c). The accuracy of O_3 retrieval based on machine learning methods to reverse O_3 concentrations is greatly improved by using polar-orbiting satellite data.

The top of atmospheric radiation (TOAR) data includes signals from the surface and atmosphere, from which atmospheric pollution information can be extracted (Wang et al., 2021). Zang et al. (2021) reversed the O_3 concentration in China using TOAR data from the MODIS, with R^2 reaching 0.71. Compared with polar-orbit satellites, geostationary satellites can remarkably improve the temporal and spatial resolution of the data (Zhang et al., 2021). Himawari-8 is a third-generation geostationary meteorological satellite (Bessho et al., 2016); the Himawari-8 satellite AOD data were used to analyze near-surface particle pollutants with a time resolution per hour and a spatial resolution of 0.05° (Chen et al., 2022b; Song et al., 2022). However, few studies have been conducted in China on near-surface O_3 concentration inversions using the TOAR of stationary satellites.

The change in O_3 concentration is affected by many factors, among which meteorological factors and weather situations are two of the key factors (Bei et al., 2022; Shu et al., 2020). Tropical cyclones and continental anticyclones are the major weather systems associated with O_3 events (Wang et al., 2017), while meteorological factors (mainly including surface temperature, humidity, and wind speed) also play an important role in O_3 pollution (Carro-Calvo et al., 2017; Chen et al., 2019; Pu et al., 2017; Yang et al., 2022). Using objective principal component analysis in the T-mode classification method to study the relationship between the weather patterns and O_3 pollution in the North China Plain in the summer of 2014–2018, Dong et al. (2020) found that the most serious O_3 pollution is related to the high-pressure anomalies over the Northwest Pacific Ocean and the obvious low-pressure centers in northeast China. The empirical orthogonal function (EOF) of daily O_3 changes in eastern China concluded that the stronger western Pacific subtropical high (WPSH) is associated with lower O_3 concentrations in southern China and higher concentrations in northern China (Zhao and Wang, 2017). A general drawback of these traditional weather classification is that, although they represent discrete phenomena of atmospheric systems, they generally cannot be organized into a continuum (Sheridan and Lee, 2011). Self-organizational mapping neuronal networks (SOM) have more advantages than the traditional EOF or principal component analysis (PCA) (Liu and Weisberg, 2011). Therefore, in this study, SOM was used to classify the seasons with the highest number of days of excessive O_3 pollution and to reveal the relationship among various weather types, meteorological elements, and O_3 concentrations.

In this study, first, the correlation between the 16 observation channels of the Himawari-8 and O_3 concentrations was determined. Then, spatiotemporal information, geographic information, meteorological factors, and TOAR data of seven highly correlated channels were input into an interpretable deep learning model as variables to build the TOAR-O_3 model, and its performance was tenfold cross-verified. Next, the model was used to estimate the ozone concentration, the hourly ozone concentration dataset over a 0.05° grid in China was developed, and the temporal and spatial distribution characteristics and trends of ozone concentration were analyzed. Finally, the main summer weather patterns in China were determined by SOM during the study period, and the effects of each weather pattern on ozone concentration were analyzed.

2. Data and method

O_3 site observation data, Himawari-8 TOAR data, ERA5 meteorological data and geographic data were used in the study. The period is from September 1, 2015 to August 31, 2021 (In this paper, winter refers to December of the previous year to February of the following year, spring refers to March to May, summer refers to June to August, and autumn refers to September to November. All hours are Beijing time).

2.1. Datasets

2.1.1. O_3 site observation

China National Environmental Monitoring Center (CNEMC) provides the average value of O_3 per hour, the O_3 monitoring data is calibrated and quality controlled according to the China National Standard (GB 3095–2012). During the study period, the total number of environmental monitoring stations was 1893, as shown by the black dots in Fig. 1(A), indicating that the stations were not evenly distributed. Studies have shown that O_3 has obvious regional differences, based on the factors such as large urban agglomerations and topography in China (Duan et al., 2022). Considering the regional division of China, we divided ten O_3

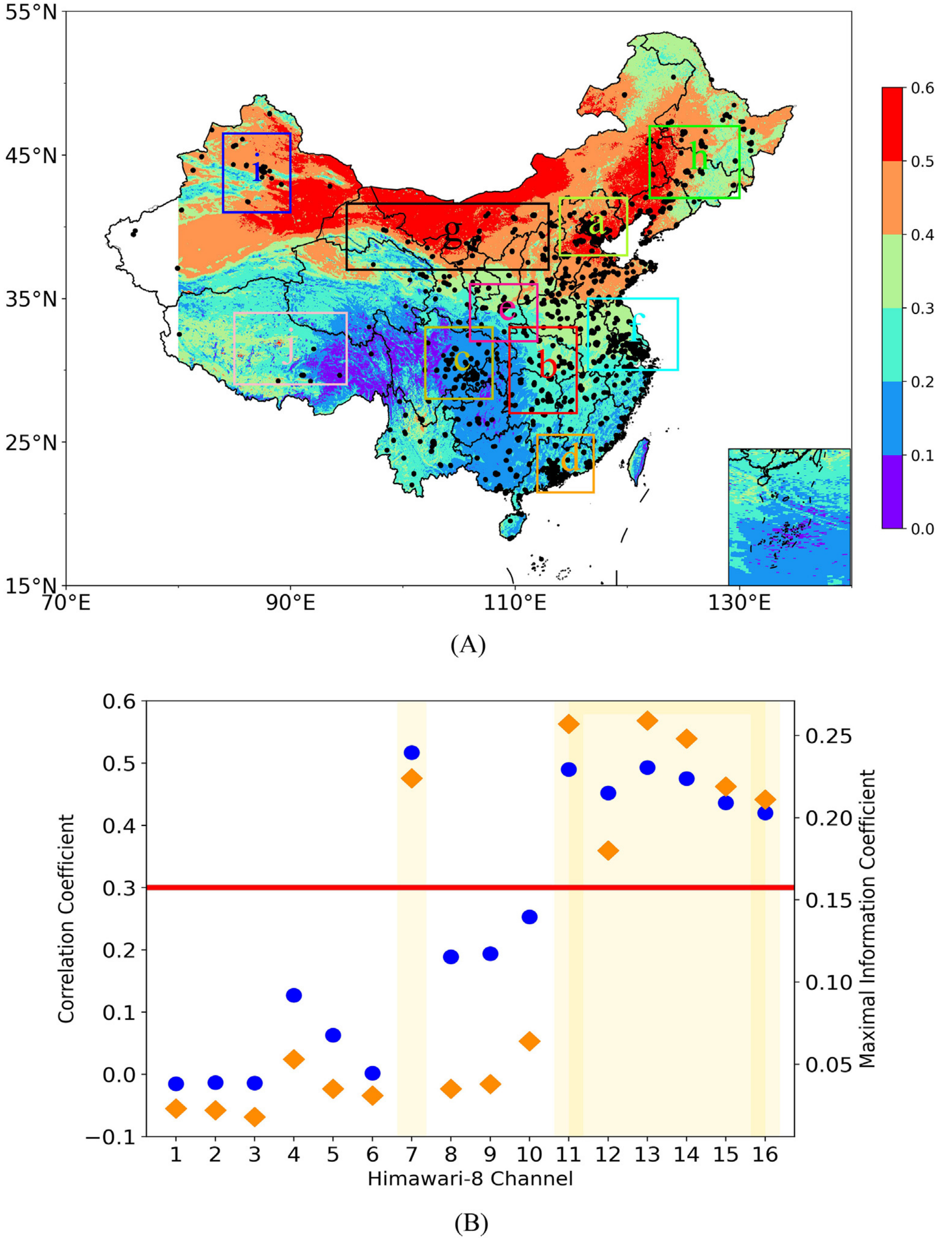


Fig. 1. (A) Distribution of China's environmental monitoring stations; the black dots represent environmental monitoring stations; the legend represents the coverage rate of TOAR after cloud removal. The boxes represent ten areas, a: Beijing-Tianjin-Hebei (BTH, greenyellow box), b: central China (CC, red box), c: Chengdu-Chongqing expressway (CY, yellow box), d: Yangtze River delta (YRD, cyan box), e: Pearl River delta (PRD, orange box), f: Guan-Zhong plain (GZP, dark pink box), g: Hexi corridor (HX, black box), h: Northeast of China (NEC, lime box), i: Northern of Xinjiang (NXJ, blue box) and the j: Tibetan plateau (TP, pink boxes); (B) Linearity and nonlinear relationship between Himawari-8 channels and O_3 concentration (Left axis: Pearson correlation coefficient, shown by blue dots, Correlation Coefficient; Right axis: MIC value, shown by orange diamond, Maximal Information Coefficient). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

research areas (as shown in the box in Fig. 1(A), and see supplementary materials Table S1 for regional longitude and latitude information).

2.1.2. Himawari-8 TOAR dataset

Himawari-8 was launched by the Japan Meteorological Agency in 2014 and officially began providing data in 2015. Compared to the Imager on Himawari-6/7, the Advanced Himawari Imager (AHI) on Himawari-8 can provide 16 optical channels of information (See Table S2 for data information). It includes 3 visible light channels, 3 near infrared channels and 10 infrared channels (Bessho et al., 2016). The O_3 channel with a central wavelength of 9.6 μm is included in infrared channels, it can be used for O_3 observation. The AHI has a time resolution of 10 min for the overall area (80°E − 160°W, 60°S − 60°N) and 2.5 min for the target area. Average coverage rate of Himawari-8 TOAR after cloud removal (ratio of effective data after cloud removal to all data) is shown in Fig. 1(A). With the gradual decrease of cloud cover from low to high latitudes in China, the coverage rate of TOAR increased from south to north. The average coverage rate of TOAR was 34.2 %, and the maximum coverage rate was over 60 %.

2.1.3. Meteorological factors and geographic information data

Meteorological and geographic factors influence O_3 concentration and transport (Ma et al., 2021). Miao et al. (2021) found that there is a nonlinear relationship between O_3 concentration and boundary layer height in North China. Zhan et al. (2018) found that temperature affects the photochemical reaction rate of O_3, and ultimately affects the concentration of O_3. Dong et al. (2020) believed that the most severe O_3 pollution in North China was related to the high-pressure center in the eastern part of north China. Therefore, the meteorological factors selected in this study including the surface 10 m wind U and V components (U10 and V10), boundary layer height (BLH), 2 m air temperature (TM), relative humidity (RH) and surface pressure (SP). They are derived from the ERA5 dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and have a temporal resolution of $0.25^\circ \times 0.25^\circ$ or $0.1^\circ \times 0.1^\circ$ in space per hour. In addition, high and low vegetation index (LH,LL), land cover type (LUCC), which were used to represent land cover type. HEIGHT is a srtm-3 elevation data jointly measured by NASA and the national mapping administration (NIMA) with a spatial resolution of 90 m. Population density data (PD) is based on the 2015 United Nations adjusted population data provided by the socio-economic data and applications center (SEDAC), spatial resolution is $0.04^\circ \times 0.04^\circ$. See Table S3 for details of the TOAR-O_3 model input data.

2.1.4. TAP O_3 dataset and China high O_3 dataset

We selected data from two commonly used MDA8 datasets namely the tracking air pollution (TAP) and China High Air Pollutants (CHAP), compared with our model products. TAP dataset is a multi-scale, near-real-time dataset of atmospheric aerosol and gaseous pollutant concentrations in China constructed by integrating ground observation, multi-source satellite remote sensing, emission inventory and model simulation data, it includes a MDA8 O_3 concentration dataset with a spatial resolution of ten km and a five-fold cross-validation R^2 of the data prediction model of 0.84 (Xiao et al., 2021; Xue et al., 2020). CHAP generated from big data (e.g., ground-based measurements, OMI multi-source satellite remote sensing products, atmospheric reanalysis and model simulations) with artificial intelligence by considering the spatiotemporal heterogeneity of air pollution),the cross-validation determination coefficient (CV-R^2) of the dataset is 0.87 (Wei et al., 2021).

2.2. Methods

2.2.1. Data matching

Bilinear interpolation was used to adjust the spatial resolution of the meteorological factors and geographical data to $0.05^\circ \times 0.05^\circ$ Himawari-8 TOAR data. Then, based on a grid of $0.05^\circ \times 0.05^\circ$, the O_3 hourly mean data recorded by environmental monitoring stations in China (80°E −136°E, 16°N − 54°N) were matched with the TOAR data. If there are multiple sites in a grid, the average of the O_3 hourly concentration of these sites was used (unless otherwise specified, the O_3 concentration presented herein refers to hourly O_3 concentration).

2.2.2. Linear and nonlinear relationship

The Himawari-8 channels may have a linear and nonlinear relationship with O_3 concentration. The linear relationship between channels and O_3 concentration was tested using Pearson's correlation, and the maximum information coefficient (MIC) was used to test the nonlinear relationship (Reshef et al., 2011). The MIC values range from zero to one; the higher the MIC value, the higher the nonlinear correlation between the two variables, that is, the stronger the dependence between the two variables.

2.2.3. Deep forest model

The deep forest (DF) model is based on an integrated model that can obtain a performance similar to that of a deep neural network (LeCun et al., 2015), but it is faster and more efficient than a deep neural network. The designed deep forest model had three hidden layers, each of which contained 12 ensemble learners (six extreme trees and six random forests). A detailed introduction to this model can be found in the literature (Chen et al., 2022b; Song et al., 2022).

The model performance was tested using the tenfold cross-validation method (Chen et al., 2022a; Di et al., 2017; Liu et al., 2020; Song et al., 2021), and the parameters used to describe the model performance included the determination coefficient (R^2), root mean square error (RMSE), mean absolute error (MAE), and bias. The formulae for these parameters can be found in literature (Chen et al., 2022a; Song et al., 2021).

2.2.4. Weather situation classification

Weather classification can be used to characterize the process of atmospheric movement at multiple scales and study the relationship between air pollution and weather circulation. SOM is an artificial neural network method for nonlinear unsupervised learning (Kohonen, 1990), which has been applied in the cluster analysis of atmospheric sciences (Crawford et al., 2016; Shu et al., 2020; Stauffer et al., 2018). SOM objectively performs nonlinear projections from the input data to two-dimensional node arrays while keeping the topological results and probability distributions of the output unchanged (Liao et al., 2018).

The geopotential height field at 850 hPa at 08 o'clock can effectively capture the weather circulation changes (Han et al., 2018). The summer geopotential height of 850 hPa at 08 o'clock from ERA5 was used as the SOM input in this study. Each SOM output node corresponds to the weather circulation type. Six major circulation types in China were identified. All 552 summer days of the selected study period were included in the clustering results.

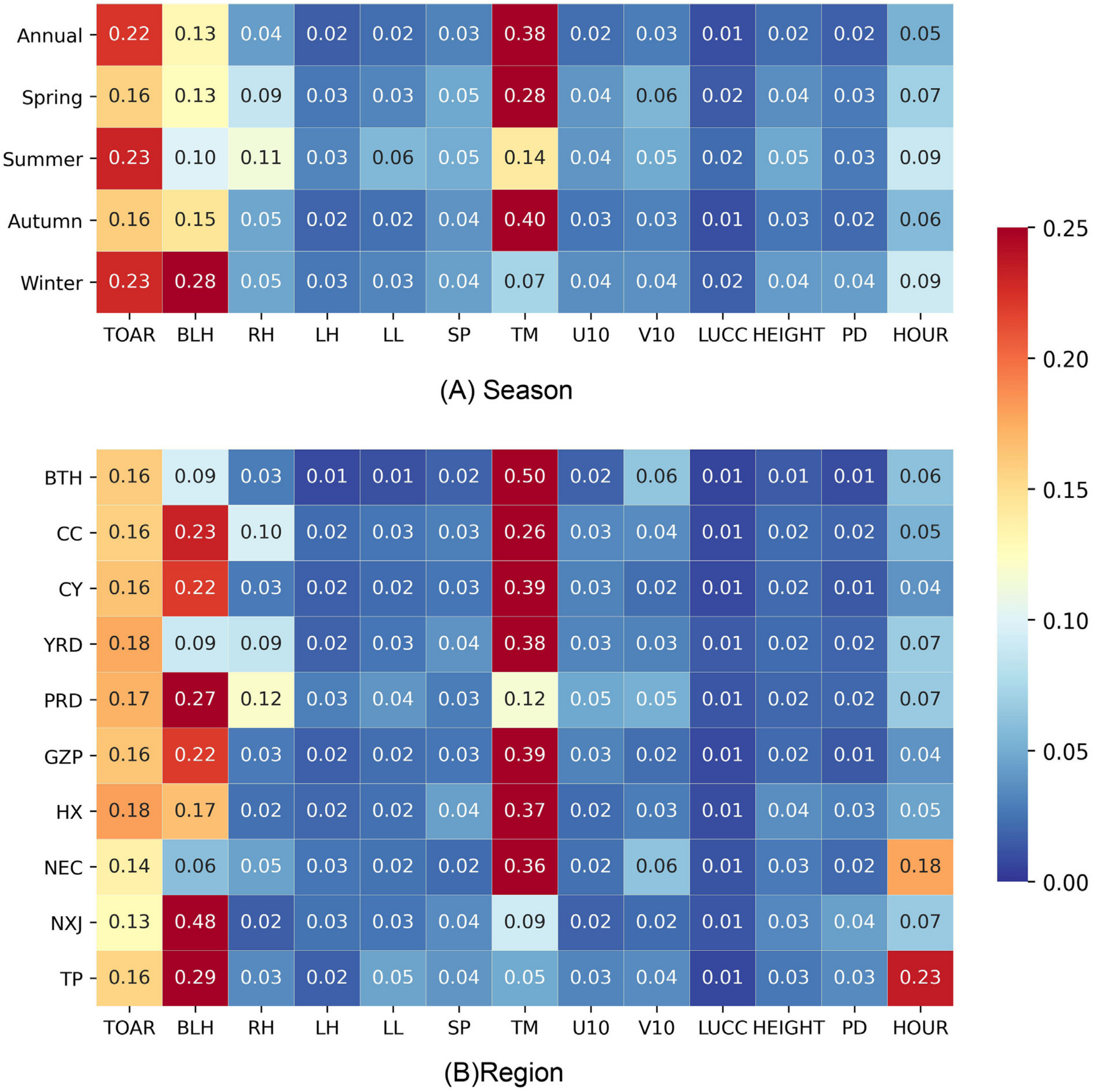
3. TOAR-O_3 model estimation

3.1. Correlation study between Himawari-8 channels and O_3 concentration

The Himawari-8 AHI contains 16 channels, of which the 12th channel (9.6 μm) is called the O_3 absorption channel. Fig. 1(B) shows the linear (Pearson correlation coefficient) and nonlinear (MIC value) relationships between the Himawari-8 channels and near-surface O_3 concentration. The dotted red line represents the correlation coefficient of 0.3 ($MIC = 0.159$). It can be seen from the Fig. 1(B) that seven channels show significant nonlinear correlation with O_3 concentration ($MIC > 0.159$), and the linear correlation between these seven channels and O_3 is also strong (Pearson correlation coefficient > 0.4). The considered seven channels are channels 7, and 11 to 16, notably channel 12, referred to as O_3 absorption channel, was also included. To maximize the utilization of channel information for the deep forest model to build its relationship with O_3, the TOAR data of these seven channels were selected as the TOAR-O_3 model input.

3.2. Feature importance of TOAR-O_3 model

The TOAR-O_3 model can directly give the importance of the features of the model, and the sum of the importance of each feature is one. Fig. 2 (A) and (B) show the characteristic importance of the TOAR-O_3 model in different seasons and regions. In summer, the TOAR had the greatest influence on the model, with a feature importance of 23 %. In winter, the contribution of TOAR (23 %) to the model was second only to that of BLH (28 %), mainly because BLH affects the diffusion and accumulation of O_3 (Miao et al., 2021). In spring and autumn, the temperature had the greatest influence on the model, accounting for 28 % and 40 %, respectively. Temperature affects the photochemical reaction rate of O_3 and ultimately affects the O_3 concentration (Zhan et al., 2018), He et al. (2021a) also highlight the important role of temperature in the photochemical reaction of O_3 precursors to generate O_3. The characteristics of the models are also different in different regions. Among the 10 regions divided in Fig. 1(A), the contribution of the TOAR to the model was relatively stable, averaging approximately 16 %. Temperature was the most important variable in the BTH, CC, CY, YRD, GZP, HX, and NEC regions, whereas BLH contributed the most in the PRD, NXJ, and TP regions. In addition, in the TP and NEC regions, HOUR is the second most important variable in the model, indicating that the O_3 concentration changes with time (Wang et al., 2017). Moreover, compared with other regions, temperature is the lowest importance in the TP, which may be related to the average altitude of the Tibetan Plateau. The above results indicate that there are remarkable differences in the influence of variables on the TOAR-O_3 model with seasons and regions.



	TOAR	BLH	RH	LH	LL	SP	TM	U10	V10	LUCC	HEIGHT	PD	HOUR
Annual	0.22	0.13	0.04	0.02	0.02	0.03	0.38	0.02	0.03	0.01	0.02	0.02	0.05
Spring	0.16	0.13	0.09	0.03	0.03	0.05	0.28	0.04	0.06	0.02	0.04	0.03	0.07
Summer	0.23	0.10	0.11	0.03	0.06	0.05	0.14	0.04	0.05	0.02	0.05	0.03	0.09
Autumn	0.16	0.15	0.05	0.02	0.02	0.04	0.40	0.03	0.03	0.01	0.03	0.02	0.06
Winter	0.23	0.28	0.05	0.03	0.03	0.04	0.07	0.04	0.04	0.02	0.04	0.04	0.09

	TOAR	BLH	RH	LH	LL	SP	TM	U10	V10	LUCC	HEIGHT	PD	HOUR
BTH	0.16	0.09	0.03	0.01	0.01	0.02	0.50	0.02	0.06	0.01	0.01	0.01	0.06
CC	0.16	0.23	0.10	0.02	0.03	0.03	0.26	0.03	0.04	0.01	0.02	0.02	0.05
CY	0.16	0.22	0.03	0.02	0.02	0.03	0.39	0.03	0.02	0.01	0.02	0.01	0.04
YRD	0.18	0.09	0.09	0.02	0.03	0.04	0.38	0.03	0.03	0.01	0.02	0.02	0.07
PRD	0.17	0.27	0.12	0.03	0.04	0.03	0.12	0.05	0.05	0.01	0.02	0.02	0.07
GZP	0.16	0.22	0.03	0.02	0.02	0.03	0.39	0.03	0.02	0.01	0.02	0.01	0.04
HX	0.18	0.17	0.02	0.02	0.02	0.04	0.37	0.02	0.03	0.01	0.04	0.03	0.05
NEC	0.14	0.06	0.05	0.03	0.02	0.02	0.36	0.02	0.06	0.01	0.03	0.02	0.18
NXJ	0.13	0.48	0.02	0.03	0.03	0.04	0.09	0.02	0.02	0.01	0.03	0.04	0.07
TP	0.16	0.29	0.03	0.02	0.05	0.04	0.05	0.03	0.04	0.01	0.03	0.03	0.23

Fig. 2. Feature importance of TOAR- O_3 model in (A) annual, seasonal and (B) ten regions. TOAR represents the radiation at the top of the atmosphere for seven selected channels, the legends and figures represent the proportion of feature importance in the model.

3.3. Model results

3.3.1. Time scale verification results

The performance of the TOAR-O_3 model was tested using a ten-fold cross-validation method. The cross-validation results of the TOAR-O_3 model from 10:00–15:00 during the study period are shown in the Fig. 3(A − F). From 10:00 to 14:00, the model fitting effect gradually increased, cross-validation R^2 was 0.86–0.92, fitting slope was 0.90–0.92, RMSE (μg/m^3) was 12.97–14.05, MAE (μg/m^3) was 8.60–9.87. At 15:00, the R^2 of the model was about 0.9, the RMSE was 15.27 μg/m^3, MAE was 10.08 μg/m^3.Which indicates that the TOAR-O_3 model can accurately estimate the hourly O_3 concentration and estimated results are in good agreement with the real data. As the gradually increase of solar radiation, the temperature was gradually rise, photochemical reaction was strengthened, and the formation of O_3 was increased (He et al., 2012; Zhan et al., 2018); therefore, from 10:00 to 14:00, the model fitting effect gradually enhanced. In contrast, R^2 of the TOAR-O_3 model also decreased (14:00–15:00) when the solar radiation reduced gradually. In general, the cross-validation R^2 of the TOAR-O_3 model was higher than 0.86, and even higher than 0.9 during 12:00–15:00, indicating that the TOAR-O_3 model has excellent performance in predicting O_3 concentration.

The TOAR-O_3 model performance also varied across time scales. From Fig. 3(G − J), the model performed well in autumn. The mean R^2, RMSE, and MAE of model were 0.92, 13.55 μg/m^3, and 9.24 μg/m^3, respectively. The model of other three seasonal performed relatively well, with cross-validation R^2 values of spring, summer, and winter being 0.89, 0.86, and 0.89, respectively. RMSE (μg/m^3) were 14.04, 18.54, and 10.1, MAE (μg/m^3) were 9.62, 13.09, and 6.92, respectively. In summer, RH is the third importance variable of the model after TOAR and temperature (Fig. 2(A)). Although temperature and solar radiation in summer are favorable for the photochemical reactions to generate O_3, higher humidity also affects O_3 generation (Zhan et al., 2018). In addition, clouds have a greater influence on O_3 generation in summer, and the number of samples matched by the TOAR-O_3 model is much smaller than that in other seasons; therefore, the model does not show better results in summer than in other seasons. In winter, the temperature is cooler, the solar radiation intensity is

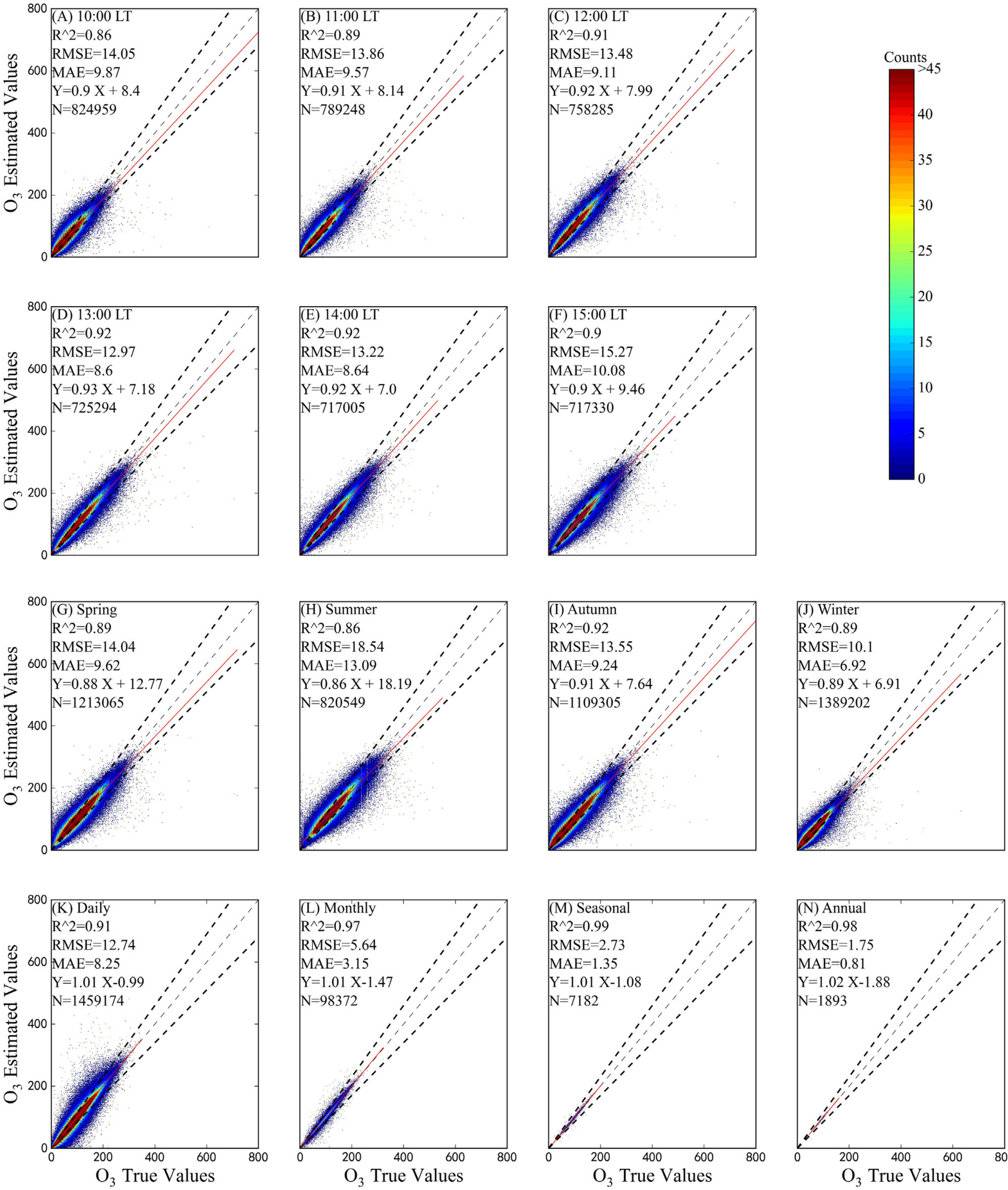


Fig. 3. Hourly (A − F), each season (G − J) and daily, monthly, seasonal, and annual (K − N) cross validation results of the TOAR-O_3 model based on grid points. (Black dotted lines represent expected error lines, light dotted lines represent 1:1 lines, and red solid lines represent linear regression fitting lines; N represents the sample size obtained each time). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

smaller, and the photochemical reaction rate of O_3 generation decreases; therefore, O_3 concentration is low (Wang et al., 2022b). In spring, the temperature gradually rises, and solar radiation strengthens, which is conducive to the generation of O_3. In winter, the TOAR and boundary layer height variables contribute >50 % to the model, temperature is only 7 %, while temperature in spring is 28 % to the model (Fig. 2(A)). In autumn, the contribution of temperature to the model can reach 40 %, which is the highest among the four seasons, and the sum contribution of the TOAR and boundary layer height to the model also exceeded 30 % (Fig. 2 (A)). In addition, the temperature in autumn was relatively high, and the number of samples was much higher than that in summer, resulting in better model performance. The estimated and measured values of O_3 were verified at different timescales, as shown in the Fig. 3(K − N). The R^2 values for daily, monthly, seasonal, and annual were 0.91, 0.97, 0.99, and 0.98,

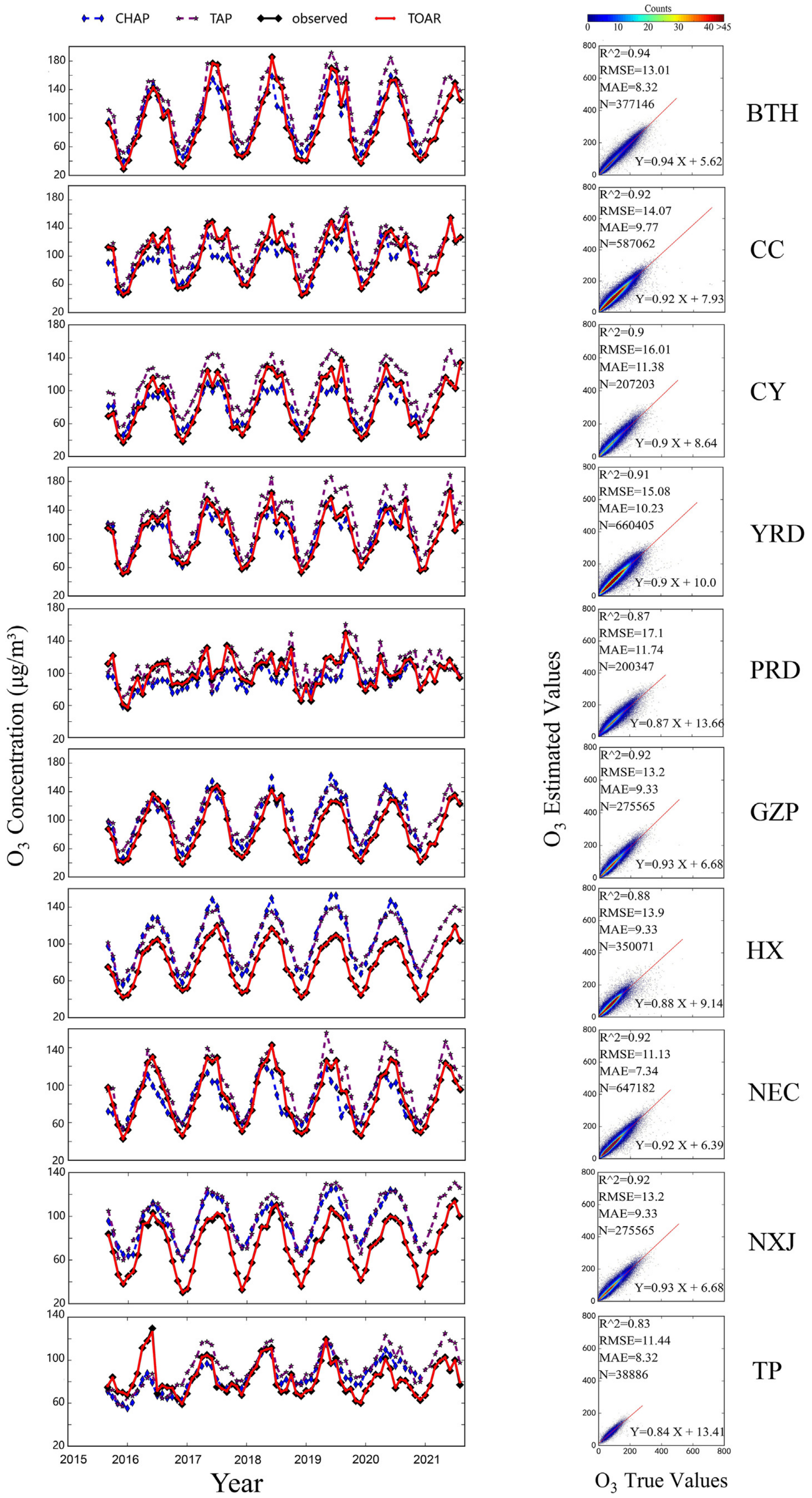


Fig. 4. Left column: Comparison of the monthly average O_3 concentration of TOAR-O_3 model prediction results with TAP and CHAP datasets and observation data; right column: hourly cross validation results of the TOAR-O_3 model based on grid points in typical regions (red solid lines represent linear regression fitting lines, N represents the sample size obtained each region). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively. RMSE (MAE) values were 12.74, 5.64, 2.73, and 1.75 μg/m^3 (8.25, 3.15, 1.35, and 0.81 μg/m^3), respectively. The fitting slopes were 1.01, 1.01, 1.01, and 1.02, respectively. The results show that the near-surface O_3 concentration estimated by the TOAR-O_3 model can provide reliable data for the change of O_3 pollution in China.

3.3.2. Spatial scale verification results

The TOAR-O_3 model shows good performance in all regions, as shown in the Fig. S1(A). The mean value of R^2 was above 0.9, but there were spatial differences in the models. Regions with a higher R^2 (A), lower RMSE (B) and MAE (C) were usually distributed in densely populated and economically developed areas, such as CY, BTH, and YRD. The R^2 of BTH and the surrounding areas was much higher than that of HX and other areas, mainly because of the high concentration of O_3 generated by photochemical reactions due to more anthropogenic O_3 precursors. Atmospheric transport has a significant impact on surface O_3 in central and eastern China (Li et al., 2021; Shen et al., 2022). In the TP region, due to sparse stations and the strong influence of meteorological factors such as wind speed, the performance of this model is relatively poor, with low R^2 (A), high RMSE (B) and MAE (C).

According to the 10 regions divided in Fig. 1(A), the cross-verification results of TOAR-O_3 model for each region were calculated. As shown in the Fig. 4 right column, the model performance was excellent in all regions, among which the model had the best performance at BTH, with R^2 reaching 0.94, RMSE of 14.09 μg/m^3 and MAE of 9.37 μg/m^3. The R^2 values of CC, GZP, NEC, and NXJ all reached 0.92, and HX's R^2 is slightly lower, with a mean of 0.88. The mean R^2 value of TP (0.83) was the lowest, which is mainly because temperature is an important factor affecting the formation of O_3 by photochemical reactions (Chang et al., 2021). The contribution of the temperature variable to the model was the weakest in the TP, at only 5 %. In addition, the sample size of TP was much smaller than that of other regions, adversely affecting the model performance. In general, the mean R^2, RMSE, and MAE of TOAR-O_3 model in each typical region can reach 0.90, 13.81 μg/m^3, and 9.51 μg/m^3, indicating that the model has universal applicability for O_3 estimation in each region of China. As shown in Fig. S2 deviation analysis results (the O_3 concentration range (A), typical area (B), different times (C), and different months (D)), the closer the median difference (orange dot) is to zero, the closer the model estimate is to its true value. Specifically, the deviation analysis of the O_3 concentration range showed that when the O_3 concentration was <300 μg/m^3, the median deviation was close to zero, and the model deviation was small. From Fig. S2(A), When O_3 concentration was >300 μg/m^3, the deviation was negative, indicating that it was underestimated. It was also overestimated in the CY (Fig. S2 (B)). and at 10:00(Fig. S2(C)), whereas it was underestimated at 15:00. From Fig. S2(D), O_3 concentrations from June to September were slightly overestimated.

3.3.3. Comparison of predicted data with other datasets

The average monthly O_3 concentration predicted by the TOAR-O_3 model was compared with the average monthly MDA8 O_3 concentration at the TAP and CHAP grid points and the hourly O_3 observation data at the site. As shown in the left column of Fig. 4, the red line represented the monthly average of the hourly O_3 concentration estimated by the TOAR-O_3 model, and the black line was the monthly average of the hourly O_3 concentration observed by the CNEMC. However, hourly O_3 concentration data are not available in the CHAP-O_3 dataset (blue line) and TAP-O_3 dataset (purple line), so the monthly mean value calculated by the MDA8 O_3 concentration was selected. The prediction result (red line) was closer to the observed data (black line), and the overall trend was consistent with the CHAP-O_3 and TAP-O_3 datasets.

3.3.4. Spatial distribution of O_3

In order to obtain the hourly spatial distribution of O_3 concentration, the O_3 estimation results were averaged by hour (10:00–15:00). As shown in the Fig. 5(A–F), the spatial distribution of near-surface O_3 concentration changes significantly between 10:00 and 15:00. Initially (10:00–11:00), it increased slowly, mainly because the solar altitude angle increased gradually with time, resulting in a moderate increase in solar radiation. From 12:00 to 15:00, near-surface O_3 showed an obvious upward trend, especially in densely populated areas such as CC and YRD, which is mainly because of favorable meteorological conditions (such as temperature and solar radiation intensity) (Chen et al., 2020), and anthropogenic nitrogen oxides (NOx) and volatile organic compounds (VOC_S) produce more O_3 through photochemical reactions (Luo et al., 2020).

The monthly variation characteristics of the ozone concentration also differed with month. As shown in the Fig. 5(H − S), except for PRD and CY, the monthly variation trend of O_3 concentration in other regions was roughly the same, increasing gradually from January to June and reached maximum, then decline from June to December, reaching the minimum value in December, which was mainly attributed to the seasonal variations of meteorological conditions and precursor emissions. The monthly variation trend of O_3 concentration in ten typical regions was analyzed (Fig. S3). The annual O_3 concentration in the CY, GZP, CC, YRD, and PRD increased significantly, and the increase in GZP was the most significant. The slope of the trend line reaches 0.22, indicating that the annual growth rate could reach 2.64 μg/m^3. In general, the slope of the O_3 concentration trend line in ten typical regions (the straight line in Fig. S3) was greater than zero, indicating that the O_3 concentration in all regions presented an upward trend, which is consistent with the results of Huang et al. (2021).

The seasonal changes of O_3 concentration are shown in the Fig. 6(A–D). The O_3 concentration was highest in the summer and lowest in the winter, which is mainly related to the temperature and solar radiation (Chen et al., 2020). In spring and summer, O_3 pollution was most serious in BTH, YRD, and CC (Han et al., 2013). In autumn, the central area of O_3 pollution shifted to the south of China. In winter, only a few cities in PRD experiencing high O_3 concentrations. To further clarify the location of high O_3 concentration area in each season, the proportion of days with daily O_3 concentrations >100 μg/m^3 at grid points in each season was calculated. As shown in the Fig. 6(E–H), there were few days with O_3 concentrations >100 μg/m^3 in winter. The number of days exceeding the standard of O_3 pollution in Southeast China was higher than that in Northwest China. The areas with more days exceeding the standard in spring were concentrated in and around the BTH, and the proportion exceeding the standard was about 20 %. Days exceeding the standard were lower in autumn and the most serious in summer. As shown in the Fig. 6(F), the O_3 concentrations in the YRD, BTH, NEC, GZP, and HX exceeded the standard (100 μg/m^3) most significantly, with the proportion up to approximately 40 %.

4. Influence of weather circulation patterns on O_3 concentration in summer

Traditional deep learning models are often regarded as “black boxes” and cannot effectively explain the results. In order to verify the accuracy of the feature importance of the DF model from the perspective of the influence of meteorological conditions on O_3 concentration, we have clustered the atmospheric circulation by SOM and the influence of meteorological factors on O_3 concentration under each circulation pattern was discussed. O_3 pollution in eastern China was more serious than that in western China (Tang et al., 2017) and in summer was the worst in all seasons (Fig. 6(B)), with the maximum number of exceeding days (Fig. 6(F)). Dong et al. (2020) found that the O_3 concentration is closely related to meteorological conditions generated by regional atmospheric circulation, and the most severe O_3 pollution in North China is related to the eastern high-pressure center. Shu et al. (2020) emphasized that summer near-surface O_3 concentrations are sensitive to major weather patterns. Considering seasonal O_3 pollution levels and atmospheric circulation influences on O_3 concentrations, the summer weather conditions during the research period were analyzed.

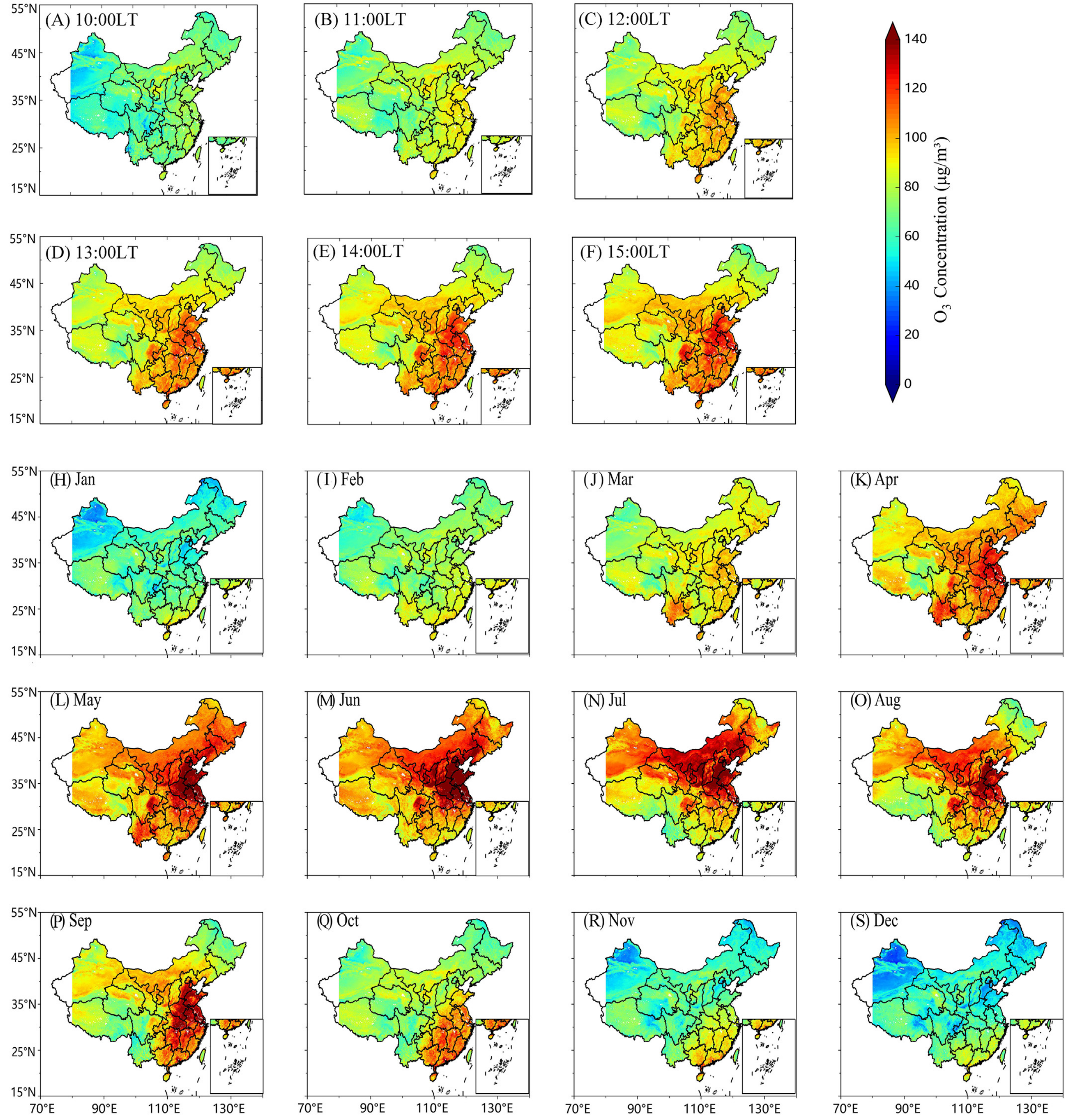


Fig. 5. Spatial distribution of hourly (A–F) and monthly (H–S) mean O_3 concentrations in China.

4.1. Classification of weather patterns

The influence of different weather systems on near-surface O_3 concentrations in summer was studied using SOM to classify the weather circulation. The main characteristics of summer weather circulation in China during the study period were represented by six major weather circulation patterns (P1, P2, P3, P4, P5 and P6, as shown in Fig. 7). Slightly different from Fig. 1 (A), the TP and NXJ regions are not shown in Fig. 7 due to the influence of altitude. P1 (Fig. 7(A)) is the most common circulation model, with a frequency of 28.08 %, and is mainly characterized by weak East Asian summer monsoons and abnormal anticyclones located in southeast China. P2 (Fig. 7(B)) is a typical typhoon system on the southeast coast of mainland China, with a frequency of 6.88 %. P3 (Fig. 7(C)) shows the weather circulation with the second highest frequency of 23.55 %, and is characterized by a zonal circulation dominant. P4 (Fig. 7(D)) is a weather system dominated by a subtropical high with a frequency of 9.78 %. P5 (Fig. 7(E)) is characterized by a frequency of 12.14 %. The main feature of this circulation pattern is the strong northeast cold vortex, Subtropical

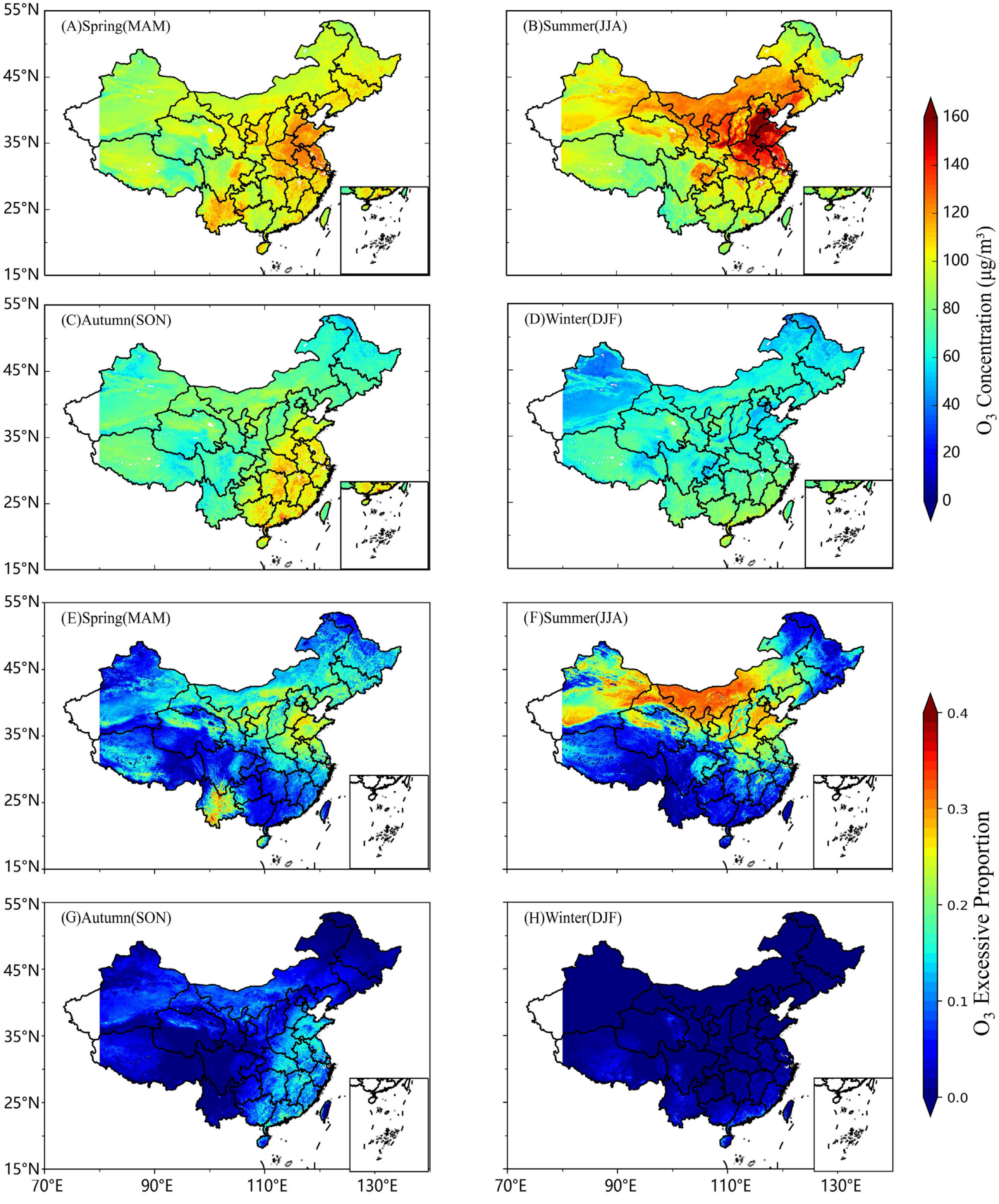


Fig. 6. (A-D) Spatial distribution of seasonal mean O_3 concentration in China; (*E*-H) proportion of days with O_3 concentration exceeding 100 μg/m^3 in each season.

highs recur under the impetus of cyclones (Zhao and Wang, 2017). P6 is a typical pattern of the Meiyu phenomenon in the middle and lower reaches of the Yangtze River (Fig. 7(F)) with a frequency of 19.57 %.

4.2. Relationship between circulation patterns and O_3 concentration

The P1 pattern occurred in June with the highest proportion (Fig. S4). The anticyclone brought clear marine air to southern China and reduced the O_3 concentration in the area. In the PRD, the O_3 concentration of P1 was 80 μg/m^3, which was lower than that of the other patterns. However, in northern China, due to the influence of weak cyclonic circulation, O_3 concentrations were relatively high, such as BTH, NEC, and HX (Fig. 8(A)).

In the P2 pattern, the subtropical high was pushed farther north by the typhoon system, and the occurrence rate in August was higher than that in the other two months (Fig. S4). The typhoon system resulted in the lowest concentration of surface O_3 in the YRD (Fig. 8(A)), but did not reduce

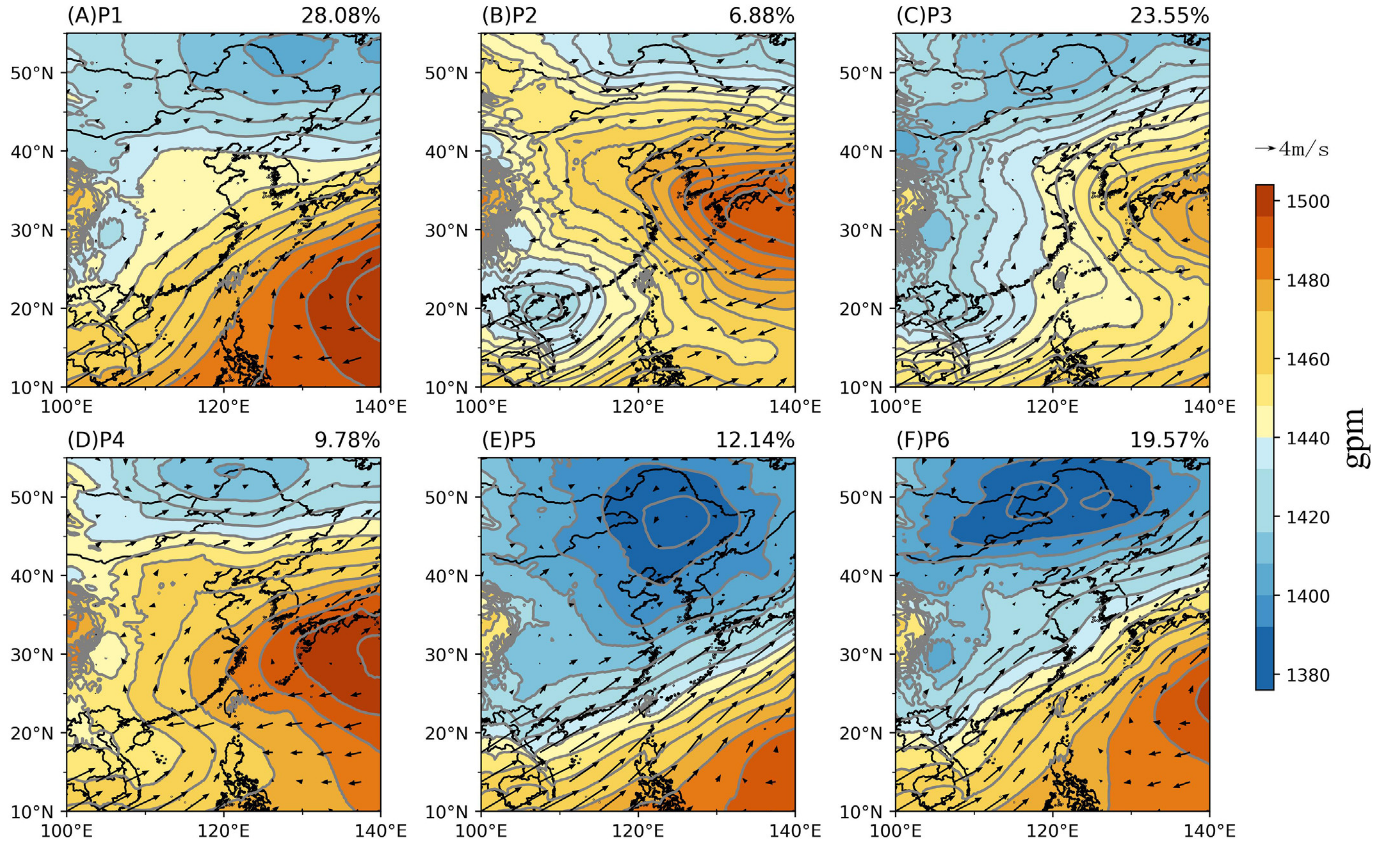


Fig. 7. 850 hPa geopotential height field and wind field for six weather circulation patterns. The number in the upper right corner indicates the occurrence frequency of each pattern.

the O_3 concentration in the PRD, which may be due to the enhanced transport of O_3 from Central China to Southeast China by cyclone circulation (Fig. 7(B)).

The P3 pattern mainly occurred in July and August (Fig. S4). In contrast to the P1 pattern, P3 was dominated by zonal circulation, which reduced O_3 transport, and the O_3 concentration in each region was higher than that in P2 (Fig. 8(A)).

The P4 pattern occurred mostly in July and August (Fig. S4). The steady weather conditions occur more frequently in the PRD; therefore, the O_3 concentration was higher (Fig. 8(A)). However, owing to the influence of the subtropical high rain belt in northern China, the O_3 concentrations of HX, GZP, and BTH was relatively low. (Fig. 8(A)).

The P5 pattern occurred mainly in June. In this weather pattern, the hourly mean O_3 concentration reached the maximum value of 130 μg/m^3 among the six patterns (Fig. S4). The O_3 concentrations were higher in all regions, reaching maximum values in the PRD and CC (Fig. 8(A)).

The Meiyu front (P6) is a typical circulation pattern of an important plum rain phenomenon in the middle and lower reaches of the Yangtze River from early June to mid-July (Fig. S4). In the northern region, owing to the control of the northeast cold vortex, a continuous sunny day occurred, and the O_3 concentrations in BTH, NEC, HX, and GZP reached the maximum. Especially in the BTH region, the O_3 concentration exceeded 150 μg/m^3, which was much higher than the other five weather circulation patterns (Dong et al., 2020).

4.3. Relationship between meteorological elements and O_3 concentration in various circulation patterns

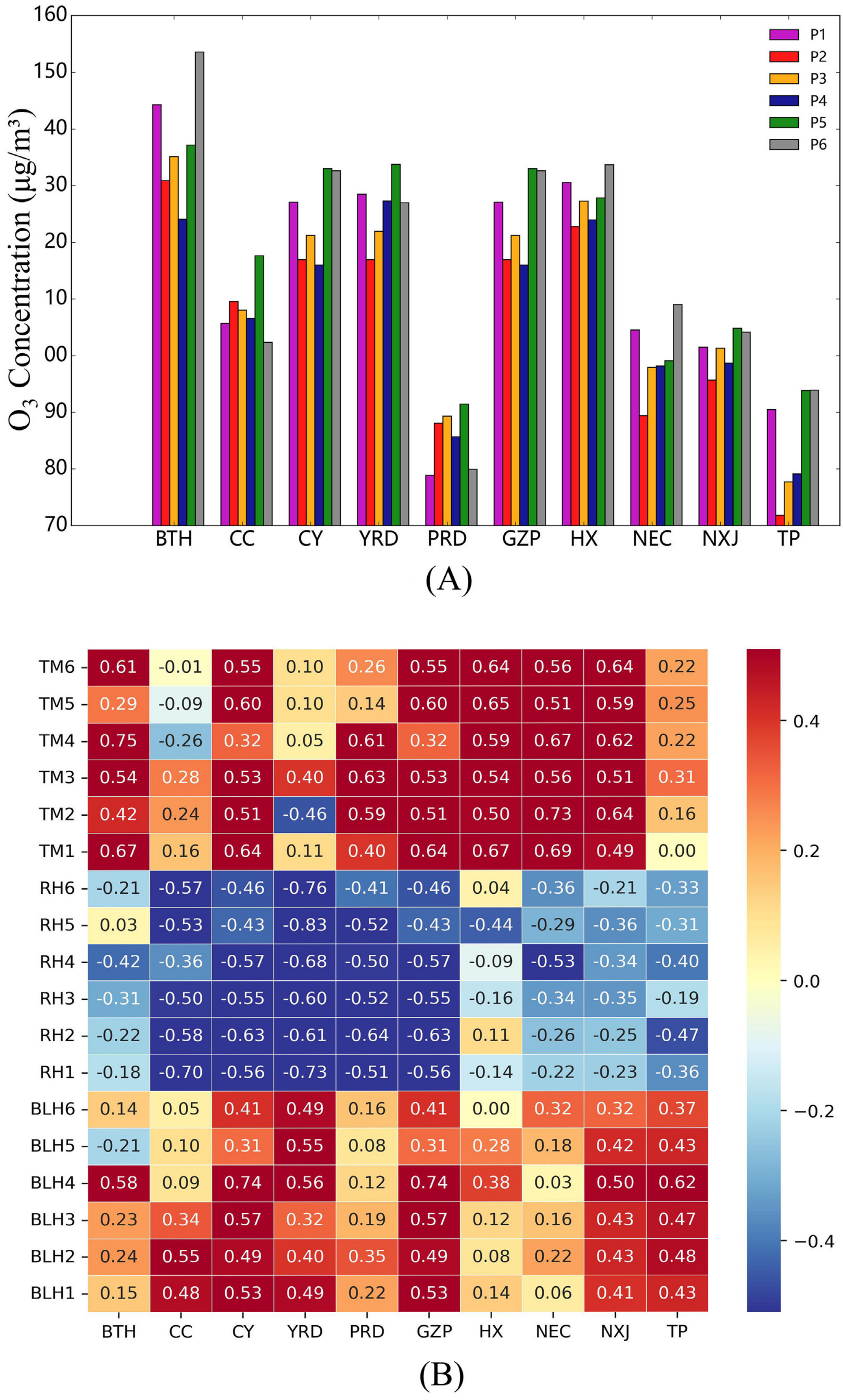
The change of O_3 concentration under different circulation patterns are mainly attributed to the change of local meteorological elements, which affects the photochemical reactions and transport process of O_3. The feature importance of the summer (Fig. 2(A)) showed that the meteorological elements have different contribution. Compared with other meteorological elements, the importance of TM, RH and BLH is >10 %. Therefore, in Fig. 8 (B), the correlation between these three meteorological elements and O_3 concentrations was discussed.

In general, the O_3 concentration was positively correlated with TM and BLH, and negatively correlated with RH. In BTH, the correlation was high for all three factors under P4 pattern, while the correlation between O_3 concentration and temperature was strong in all circulation patterns except P5. In CC, the correlation was higher for all three factors in P1 and P2, while the correlation between O_3 concentration and relative humidity was greater than the other two elements. In CY and GZP, these three meteorological elements were all highly correlated with O_3. In the YRD, the O_3 concentration was highly correlated with RH and the BLH, but less correlated with temperature except P3. In HX, NEC, and NXJ, O_3 concentration had a good correlation with temperature, while in the TP, it had a higher correlation with BLH than TM and RH, which was consistent with the area feature importance of TOAR-O_3 model (Fig. 2(B)).

5. Discussion and conclusions

For a large amount of satellite data, it is a challenge to build a fast computing model and ensure the optimal model performance. The Pearson correlation coefficient (PCC) and maximum information correlation (MIC) were used to study the correlation between the TOAR of the Himawari-8 satellite and the O_3 concentration; seven TOAR channels (channels 7 and 11–16) had significantly high linearity (PCC > 0.4) and nonlinear correlation (MIC>0.15), including O_3 at 9.6 μm absorption band. Compared with the model which all 16 bands inputs, the model with the selected 7 bands had the best performance (See Table S4 for details) and saved about 1/3 of calculation time.

Traditional deep learning models are often regarded as "black boxes" and cannot effectively explain the results. An interpretable deep learning



	BTH	CC	CY	YRD	PRD	GZP	HX	NEC	NXJ	TP
TM6	0.61	-0.01	0.55	0.10	0.26	0.55	0.64	0.56	0.64	0.22
TM5	0.29	-0.09	0.60	0.10	0.14	0.60	0.65	0.51	0.59	0.25
TM4	0.75	-0.26	0.32	0.05	0.61	0.32	0.59	0.67	0.62	0.22
TM3	0.54	0.28	0.53	0.40	0.63	0.53	0.54	0.56	0.51	0.31
TM2	0.42	0.24	0.51	-0.46	0.59	0.51	0.50	0.73	0.64	0.16
TM1	0.67	0.16	0.64	0.11	0.40	0.64	0.67	0.69	0.49	0.00
RH6	-0.21	-0.57	-0.46	-0.76	-0.41	-0.46	0.04	-0.36	-0.21	-0.33
RH5	0.03	-0.53	-0.43	-0.83	-0.52	-0.43	-0.44	-0.29	-0.36	-0.31
RH4	-0.42	-0.36	-0.57	-0.68	-0.50	-0.57	-0.09	-0.53	-0.34	-0.40
RH3	-0.31	-0.50	-0.55	-0.60	-0.52	-0.55	-0.16	-0.34	-0.35	-0.19
RH2	-0.22	-0.58	-0.63	-0.61	-0.64	-0.63	0.11	-0.26	-0.25	-0.47
RH1	-0.18	-0.70	-0.56	-0.73	-0.51	-0.56	-0.14	-0.22	-0.23	-0.36
BLH6	0.14	0.05	0.41	0.49	0.16	0.41	0.00	0.32	0.32	0.37
BLH5	-0.21	0.10	0.31	0.55	0.08	0.31	0.28	0.18	0.42	0.43
BLH4	0.58	0.09	0.74	0.56	0.12	0.74	0.38	0.03	0.50	0.62
BLH3	0.23	0.34	0.57	0.32	0.19	0.57	0.12	0.16	0.43	0.47
BLH2	0.24	0.55	0.49	0.40	0.35	0.49	0.08	0.22	0.43	0.48
BLH1	0.15	0.48	0.53	0.49	0.22	0.53	0.14	0.06	0.41	0.43

Fig. 8. (A) Hourly mean O_3 concentration in typical regions under different circulation patterns; (B) the correlation coefficient between temperature (TM), relative humidity (RH), boundary layer height (BLH) and O_3 concentrations under the six circulation patterns. The numbers behind TM, RH and BLH correspond to the six patterns in Fig. 7.

model (DF) was used to invert the near-surface O_3 concentration. The feature importance of the model showed that temperature, TOAR, and BLH contributed 38 %, 22 %, and 13 %, respectively. Since the time resolution of polar-orbiting satellites is difficult to meet the current research on refinement O_3 and synergistic pollution, this paper selects the TOAR data of geostationary satellites with relatively high time and spatial resolution for O_3 concentration inversion. The temporal resolution of the retrieved O_3 concentration is hours, and the spatial resolution is 0.05°. The R^2 (RMSE: μg/m^3) values of the daily, monthly, seasonal and annual 10-fold cross-validation results were 0.91 (12.74), 0.97 (5.64), 0.99 (2.73) and 0.98 (1.75), respectively. The ozone concentration in most areas of China has increased significantly, especially in GZP, with an annual growth rate of 2.64 μg/m^3.

Sheridan and Lee (2011) mentioned that traditional atmospheric circulation classification models, such as empirical orthogonal function and principal component analysis, represent discrete phenomena of atmospheric systems, but they usually cannot be organized into a continuum. The SOM classification results can represent a continuum of circulation patterns. Therefore, SOM was used in the paper to classify the summer circulation field, corresponding to the period of high ozone pollution. The O_3 concentration in northern China was the highest under the circulation pattern of the Meiyu front over the Yangtze River Delta, while in southern China, it was the highest under the circulation pattern of the northeast cold vortex controlling most of China. Nevertheless, the lowest O_3 concentration under the circulation pattern was characterized by the typhoon system in the Yangtze River Delta.

The feature importance of TOAR-O_3 model shows that the contribution of temperature is small while the BLH is the largest over the Qinghai-Tibet Plateau. Since temperature is closely related to the photochemical reaction that generates ozone, the photochemical reaction is not the main source of O_3. The BLH on the Qinghai-Tibet Plateau was closely related to the terrain height and the seasonal change (Xiang et al., 2021). Ozone at Mount Waliguan in the northeastern Qinghai-Tibetan Plateau mostly originated from the stratosphere, Americas, Europe and Africa (Zhu et al., 2016). However, Air masses from central and eastern China dominates the airflow at Mount Waliguan in summer, suggesting strong impact of anthropogenic forcing on the surface ozone on the Plateau (Xue et al., 2011). Therefore, the source of O_3 in TP needs to be further analyzed.

The O_3 concentration in Inner Mongolia is relatively high in summer (Fig. S5), but there is no major O_3 precursor source region in this region (Liang et al., 2020). There are several possible reasons. First, atmospheric transport largely regulates the regional interaction of O_3 pollution in China (Shen et al., 2022). The atmospheric circulation not only transports O_3 from BTH, GZP and other areas, but also transports O_3 precursors, and, 70 % of the circulation patterns is conducive to increase the ozone concentration in Inner Mongolia (Fig. S6). Second, according to the feature importance of the model in Inner Mongolia (Table S5), it can be seen that temperature has the greatest influence on the O_3 concentration, indicating that the ozone generated by photochemical reaction accounts for a large proportion of total O_3 in this region. Third, the feature importance of the BLH is next to TM (Table S5). When the BLH increases, the diffusion of ozone depleting pollutants (NO, etc.) accelerates, and the titration effect of NO on near-surface O_3 is weakened, and the ozone consumption decreases. In addition, the increase of BLH will also accelerate the diffusion of O_3 and reduce the near-surface O_3 concentration. When the increasing rate of O_3 is higher than its decreasing rate, the near-surface O_3 concentration will increase. With the increase of solar radiation, the BLH will also increase in normal, and the increase of solar radiation will lead to the increase of photochemical reaction of ozone generation. These comprehensive reasons lead to relatively high ozone concentration in summer and at 13–15 o'clock (Beijing Time) in Inner Mongolia, as shown in Fig. S5 and Fig. 5 (D-F).

In this study, Himawari-8 TOAR data were combined with meteorological factors and geographic information data to construct a TOAR-O_3 model based on DF. The O_3 concentration in China from September 2015 to August 2021 was estimated using the model, and the summer circulation patterns were classified using the SOM. The conception process is shown in the Fig. S7. In future work, further improvement is needed. In this paper, only the daytime O_3 concentration was retrieved, but the 24-h O_3 concentration was not obtained. As the DF model is a deep learning model, it will take a long time to calculate if there is a large amount of data. It remains difficult to accurately evaluate the performance of the TOAR-O_3 model in areas without O_3 monitoring sites. In addition, the inter-regional transport contribution of ozone in China has not been quantified.

CRediT authorship contribution statement

Bin Chen: Conceptualization, Methodology, Writing-Original draft preparation, Resources.
Yixuan Wang: Software, Methodology, Data curation, Writing-Original draft preparation, Visualization.
Jianping Huang: Supervision, Project administration.
Lin Zhao: Software, Data curation.
Ruming Chen: Software, Data curation.
Zhihao Song: Software, Methodology, Data curation.
Jiashun Hu: Software, Methodology, Data curation.

Data availability

I have shared the link to my data and code in the Manuscript

Declaration of competing interest

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

Acknowledgements

The work Supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP; Grant number 2019QZKK0602), the National Natural Science Foundation of China (Grant number 41775021), and the Fundamental Research Funds for the Central Universities (Grant number lzujbky-2022-ct06). The O_3 observation data were obtained from https://www.aqistudy.cn/historydata/. The Himawari-8 TOAR data provided by the Japan Meteorological Agency, download from: http://www.eorc.jaxa.jp/ptree/index.html. The ERA-5 data are available from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview. The TAP O_3 dataset is available from http://tapdata.org.cn. The CHAP O_3 dataset is obtained from https://zenodo.org/record/4400043. The authors would like to thank China National Environmental Monitoring Center, TAP team, CHAP team, and European Centre for Medium-Range Weather Forecasts for their datasets. The estimated data and data reading codes are available from doi:https://doi.org/10.5281/zenodo.7045738. All programs in this study are implemented based on Python3, https://www.python.org/ (Python3, 2022).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.160928.

References

Bai, K., Chang, N.-B., Yu, H., Gao, W., 2016. Statistical bias correction for creating coherent total ozone record from OMI and OMPS observations. Remote Sens. Environ. 182, 150–168. https://doi.org/10.1016/j.rse.2016.05.007.

Bei, N., Liang, J., Li, X., Wang, R., 2022. Worsening summertime ozone pollution in the Guanzhong Basin, China from 2014 to 2018: impacts of synoptic conditions and anthropogenic emissions. Atmos. Environ. 274. https://doi.org/10.1016/j.atmosenv.2022.118974.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., Yoshida, R., 2016. An introduction to Himawari-8/9— Japan’s new-generation geostationary meteorological satellites. J. Meteorol. Soc. Jpn. 94, 151–183. https://doi.org/10.2151/jmsj.2016-009.

Carro-Calvo, L., Ordóñez, C., García-Herrera, R., Schnell, J.L., 2017. Spatial clustering and meteorological drivers of summer ozone in Europe. Atmos. Environ. 167, 496–510. https://doi.org/10.1016/j.atmosenv.2017.08.050.

Chang, L., He, F., Tie, X., Xu, J., Gao, W., 2021. Meteorology driving the highest ozone level occurred during mid-spring to early summer in Shanghai,China. Sci. Total Environ. 785. https://doi.org/10.1016/j.scitotenv.2021.147253.

Chen, Z., Zhuang, Y., Xie, X., Chen, D., Cheng, N., Yang, L., Li, R., 2019. Understanding long-term variations of meteorological influences on ground ozone concentrations in Beijing during 2006–2016. Environ. Pollut. 245, 29–37. https://doi.org/10.1016/j.envpol.2018.10.117.

Chen, Z., Li, R., Chen, D., Zhuang, Y., Gao, B., Yang, L., Li, M., 2020. Understanding the causal influence of major meteorological factors on ground ozone concentrations across China. J. Clean. Prod. 242. https://doi.org/10.1016/j.jclepro.2019.118498.

Chen, B., Song, Z., Pan, F., Huang, Y., 2022a. Obtaining vertical distribution of PM2.5 from CALIOP data and machine learning algorithms. Sci. Total Environ. 805, 150338. https://doi.org/10.1016/j.scitotenv.2021.150338.

Chen, B., Song, Z., Shi, B., Li, M., 2022b. An interpretable deep forest model for estimating hourly PM10 concentration in China using Himawari-8 data. Atmos. Environ. 268. https://doi.org/10.1016/j.atmosenv.2021.118827.

Chen, J., Shen, H., Li, X., Li, T., Wei, Y., 2022c. 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 Obs. Geoinf. 112. https://doi.org/10.1016/j.jag.2022.102955.

Chi, G., Huang, B., Shi, Y., Chen, X., Li, Q., Zhu, J., 2016. Detecting ozone effects in four wheat cultivars using hyperspectral measurements under fully open-air field conditions. Remote Sens. Environ. 184, 329–336. https://doi.org/10.1016/j.rse.2016.07.020.

Crawford, J., Griffiths, A., Cohen, D.D., Jiang, N., Stelcer, E., 2016. Particulate pollution in the Sydney region: source diagnostics and synoptic controls. Aerosol Air Qual. Res. 16, 1055–1066. https://doi.org/10.4209/aaqr.2015.02.0081.

DeLang, M.N., Becker, J.S., Chang, K.L., Serre, M.L., Cooper, O.R., Schultz, M.G., Schroder, S., Lu, X., Zhang, L., Deushi, M., Josse, B., Keller, C.A., Lamarque, J.F., Lin, M., Liu, J., Marecal, V., Strode, S.A., Sudo, K., Tilmes, S., Zhang, L., Cleland, S.E., Collins, E.L., Brauer, M., West, J.J., 2021. Mapping yearly fine resolution global surface ozone through the Bayesian Maximum Entropy data fusion of observations and model output for 1990–2017. Environ. Sci. Technol. 55, 4389–4398. https://doi.org/10.1021/acs.est.0c07742.

Di, Q., Rowland, S., Koutrakis, P., Schwartz, J., 2017. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. J. Air Waste Manag. Assoc. 67, 39–52. https://doi.org/10.1080/10962247.2016.1200159.

Dong, Y., Li, J., Guo, J., Jiang, Z., Chu, Y., Chang, L., Yang, Y., Liao, H., 2020. The impact of synoptic patterns on summertime ozone pollution in the North China Plain. Sci. Total Environ. 735, 139559. https://doi.org/10.1016/j.scitotenv.2020.139559.

Dong, L., Chen, B., Huang, Y., Song, Z., Yang, T., 2021. Analysis on the characteristics of air pollution in china during the COVID-19 outbreak. Atmosphere 12. https://doi.org/10.3390/atmos12020205.

Duan, W., Wang, X., Cheng, S., Wang, R., 2022. Regional division and influencing mechanisms for the collaborative control of PM2.5 and O3 in China: a joint application of multiple mathematic models and data mining technologies. J. Clean. Prod. 337. https://doi.org/10.1016/j.jclepro.2022.130607.

Feng, Z., Hu, E., Wang, X., Jiang, L., Liu, X., 2015. Ground-level O3 pollution and its impacts on food crops in China: a review. Environ. Pollut. 199, 42–48. https://doi.org/10.1016/j.envpol.2015.01.016.

Gao, D., Xie, M., Chen, X., Wang, T., Liu, J., Xu, Q., Mu, X., Chen, F., Li, S., Zhuang, B., Li, M., Zhao, M., Ren, J., 2020. Systematic classification of circulation patterns and integrated analysis of their effects on different ozone pollution levels in the Yangtze River Delta Region,China. Atmos. Environ. 242. https://doi.org/10.1016/j.atmosenv.2020.117760.

Han, S.-Q., Zhang, M., Zhao, C.-S., Lu, X.-Q., Ran, L., Han, M., Li, P.-Y., Li, X.-J., 2013. Differences in ozone photochemical characteristics between the megacity Tianjin and its rural surroundings. Atmos. Environ. 79, 209–216. https://doi.org/10.1016/j.atmosenv.2013.06.045.

Han, H., Liu, J., Yuan, H., Jiang, F., Zhu, Y., Wu, Y., Wang, T., Zhuang, B., 2018. Impacts of synoptic weather patterns and their persistency on free tropospheric carbon monoxide concentrations and outflow in eastern China. J.Geophys.Res.Atmos. 123, 7024–7046. https://doi.org/10.1029/2017jd028172.

He, J., Wang, Y., Hao, J., Shen, L., Wang, L., 2012. Variations of surface O3 in August at a rural site near Shanghai: influences from the West Pacific subtropical high and anthropogenic emissions. Environ. Sci. Pollut. Res. Int. 19, 4016–4029. https://doi.org/10.1007/s11356-012-0970-5.

He, Y., Wang, H., Wang, H., Xu, X., Li, Y., Fan, S., 2021a. Meteorology and topographic influences on nocturnal ozone increase during the summertime over Shaoguan,China. Atmos. Environ. 256. https://doi.org/10.1016/j.atmosenv.2021.118459.

He, Z., Liu, P., Zhao, X., He, X., Liu, J., Mu, Y., 2021b. Responses of surface O3 and PM2.5 trends to changes of anthropogenic emissions in summer over Beijing during 2014–2019: a study based on multiple linear regression and WRF-Chem. Sci. Total Environ. 807, 150792. https://doi.org/10.1016/j.scitotenv.2021.150792.

Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, L., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., Cheng, W., Liu, W., Fu, Q., Liu, B., Chai, F., Davis, S.J., Zhang, Q., He, K., 2021. Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China. Natl. Sci. Rev. 8, nwaa137. https://doi.org/10.1093/nsr/nwaa137.

IPCC, 2021. Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.K., Kim, S., 2021. Estimation of surface-level NO2 and O3 concentrations using TROPOMI data and machine learning over East Asia. Environ. Pollut. 288, 117711. https://doi.org/10.1016/j.envpol.2021.117711.

Kohonen, T., 1990. The self-organizing map. Proceedings of the IEEE. 78, pp. 1464–1480. https://doi.org/10.1109/5.58325.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lei, R., Zhu, F., Cheng, H., Liu, J., Shen, C., Zhang, C., Xu, Y., Xiao, C., Li, X., Zhang, J., Ding, R., Cao, J., 2019. Short-term effect of PM2.5/O3 on non-accidental and respiratory deaths in highly polluted area of China. Atmos.Pollut.Res. 10, 1412–1419. https://doi.org/10.1016/j.apr.2019.03.013.

Li, C., Krotkov, N.A., Dickerson, R.R., Li, Z., Yang, K., Chin, M., 2010a. Transport and evolution of a pollution plume from northern China: a satellite-based case study. J. Geophys. Res. 115. https://doi.org/10.1029/2009jd012245.

Li, C., Zhang, Q., Krotkov, N.A., Streets, D.G., He, K., Tsay, S.-C., Gleason, J.F., 2010b. Recent large reduction in sulfur dioxide emissions from Chinese power plants observed by the Ozone Monitoring Instrument. Geophys. Res. Lett. 37. https://doi.org/10.1029/2010gl042594.

Li, R., Zhao, Y., Zhou, W., Meng, Y., Zhang, Z., Fu, H., 2020. Developing a novel hybrid model for the estimation of surface 8 h ozone (O3) across the remote Tibetan Plateau during 2005–2018. Atmos. Chem. Phys. 20, 6159–6175. https://doi.org/10.5194/acp-20-6159-2020.

Li, M., Wang, T., Shu, L., Qu, Y., Xie, M., Liu, J., Wu, H., Kalsoom, U., 2021. Rising surface ozone in China from 2013 to 2017: a response to the recent atmospheric warming or pollutant controls? Atmos. Environ. 246. https://doi.org/10.1016/j.atmosenv.2020.118130.

Liang, X., Sun, X., Xu, J., Ye, D., 2020. Improved emissions inventory and VOCs speciation for industrial OFP estimation in China. Sci. Total Environ. 745, 140838. https://doi.org/10.1016/j.scitotenv.2020.140838.

Liao, Z., Sun, J., Yao, J., Liu, L., Li, H., Liu, J., Xie, J., Wu, D., Fan, S., 2018. Self-organized classification of boundary layer meteorology and associated characteristics of air quality in Beijing. Atmos. Chem. Phys. 18, 6771–6783. https://doi.org/10.5194/acp-18-6771-2018.

Liu, Y., Weisberg, R.H., 2011. A review of self-organizing map applications in meteorology and oceanography. Self-organizing Maps: Applications And Novel Algorithm Design. 1, pp. 253–272.

Liu, H., Liu, S., Xue, B., Lv, Z., Meng, Z., Yang, X., Xue, T., Yu, Q., He, K., 2018. Ground-level ozone pollution and its health impacts in China. Atmos. Environ. 173, 223–230. https://doi.org/10.1016/j.atmosenv.2017.11.014.

Liu, N., Lin, W., Ma, J., Xu, W., Xu, X., 2019. Seasonal variation in surface ozone and its regional characteristics at global atmosphere watch stations in China. J. Environ. Sci. (China) 77, 291–302. https://doi.org/10.1016/j.jes.2018.08.009.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., Bi, J., 2020. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach. Environ. Int. 142, 105823. https://doi.org/10.1016/j.envint.2020.105823.

Lu, X., Zhang, L., Wang, X.L., Gao, M., Li, K., Zhang, Y.Z., Yue, X., Zhang, Y.H., 2020. Rapid increases in warm-season surface ozone and resulting health impact in China Since 2013. Environ.Sci.Technol.Lett. 7, 240–247. https://doi.org/10.1021/acs.estlett.0c00171.

Luo, H., Li, G., Chen, J., Lin, Q., Ma, S., Wang, Y., An, T., 2020. Spatial and temporal distribution characteristics and ozone formation potentials of volatile organic compounds from three typical functional areas in China. Environ. Res. 183, 109141. https://doi.org/10.1016/j.envres.2020.109141.

Luo, N., Zang, Z., Yin, C., Liu, M., Jiang, Y., Zuo, C., Zhao, W., Shi, W., Yan, X., 2022. Explainable and spatial dependence deep learning model for satellite-based O3 monitoring in China. Atmos. Environ. 290. https://doi.org/10.1016/j.atmosenv.2022.119370.

Ma, S., Shao, M., Zhang, Y., Dai, Q., Xie, M., 2021. Sensitivity of PM2.5 and O3 pollution episodes to meteorological factors over the North China Plain. Sci. Total Environ. 792, 148474. https://doi.org/10.1016/j.scitotenv.2021.148474.

Miao, Y., Che, H., Zhang, X., Liu, S., 2021. Relationship between summertime concurring PM2.5 and O3 pollution and boundary layer height differs between Beijing and Shanghai, China. Environ. Pollut. 268, 115775. https://doi.org/10.1016/j.envpol.2020.115775.

Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. 15, 8889–8973. https://doi.org/10.5194/acp-15-8889-2015.

Pu, X., Wang, T.J., Huang, X., Melas, D., Zanis, P., Papanastasiou, D.K., Poupkou, A., 2017. Enhanced surface ozone during the heat wave of 2013 in Yangtze River Delta region, China. Sci. Total Environ. 603–604, 807–816. https://doi.org/10.1016/j.scitotenv.2017.03.056.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C., 2011. Detecting novel associations in large data sets. 334, pp. 1518–1524. https://doi.org/10.1126/science.1205438.

Shen, L., Liu, J., Zhao, T., Xu, X., Han, H., Wang, H., Shu, Z., 2022. Atmospheric transport drives regional interactions of ozone pollution in China. Sci. Total Environ. 830, 154634. https://doi.org/10.1016/j.scitotenv.2022.154634.

Sheridan, S.C., Lee, C.C., 2011. The self-organizing map in synoptic climatological research. 35, pp. 109–119. https://doi.org/10.1177/0309133310397582.

Shu, L., Wang, T., Han, H., Xie, M., Chen, P., Li, M., Wu, H., 2020. Summertime ozone pollution in the Yangtze River Delta of eastern China during 2013–2017: synoptic impacts and source apportionment. Environ. Pollut. 257, 113631. https://doi.org/10.1016/j.envpol.2019.113631.

Song, Z., Chen, B., Huang, Y., Dong, L., Yang, T., 2021. Estimation of PM2.5 concentration in China using linear hybrid machine learning model. Atmos.Meas.Tech. 14, 5333–5347. https://doi.org/10.5194/amt-14-5333-2021.

Song, Z., Chen, B., Huang, J., 2022. Combining Himawari-8 AOD and deep forest model to obtain city-level distribution of PM2.5 in China. Environ. Pollut. 297, 118826. https://doi.org/10.1016/j.envpol.2022.118826.

Stauffer, R.M., Thompson, A.M., Witte, J.C., 2018. Characterizing global ozonesonde profile variability from surface to the UT/LS with a clustering technique and MERRA-2 reanalysis. J. Geophys. Res. Atmos. 123, 6213–6229. https://doi.org/10.1029/2018jd028465.

Tang, G., Zhu, X., Xin, J., Hu, B., Song, T., Sun, Y., Zhang, J., Wang, L., Cheng, M., Chao, N., Kong, L., Li, X., Wang, Y., 2017. Modelling study of boundary-layer ozone over northern China - part I: ozone budget in summer. Atmos. Res. 187, 128–137. https://doi.org/10.1016/j.atmosres.2016.10.017.

Tang, G., Liu, Y., Huang, X., Wang, Y., Hu, B., Zhang, Y., Song, T., Li, X., Wu, S., Li, Q., Kang, Y., Zhu, Z., Wang, M., Wang, Y., Li, T., Li, X., Wang, Y., 2021. Aggravated ozone pollution in the strong free convection boundary layer. Sci. Total Environ. 788, 147740. https://doi.org/10.1016/j.scitotenv.2021.147740.

Tian, Y., Wu, Y., Liu, H., Si, Y., Wu, Y., Wang, X., Wang, M., Wu, J., Chen, L., Wei, C., Wu, T., Gao, P., Hu, Y., 2020. The impact of ambient ozone pollution on pneumonia: a nationwide time-series analysis. Environ. Int. 136, 105498. https://doi.org/10.1016/j.envint.2020.105498.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. Sci. Total Environ. 575, 1582–1596. https://doi.org/10.1016/j.scitotenv.2016.10.081.

Wang, B., Yuan, Q., Yang, Q., Zhu, L., Li, T., Zhang, L., 2021. Estimate hourly PM2.5 concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network. Environ. Pollut. 271, 116327. https://doi.org/10.1016/j.envpol.2020.116327.

Wang, H., Huang, C., Tao, W., Gao, Y., Wang, S., Jing, S., Wang, W., Yan, R., Wang, Q., An, J., Tian, J., Hu, Q., Lou, S., Pöschl, U., Cheng, Y., Su, H., 2022a. Seasonality and reduced nitric oxide titration dominated ozone increase during COVID-19 lockdown in eastern China. <sb:contribution><sb:title>npj Clim. Atmos.</sb:title></sb:contribution><sb:host><sb:issue><sb:series><sb:title>Sci.</sb:title></sb:series></sb:issue></sb:host> 5. https://doi.org/10.1038/s41612-022-00249-3.

Wang, Y., Yuan, Q., Zhu, L., Zhang, L., 2022b. Spatiotemporal estimation of hourly 2-km ground-level ozone over China based on Himawari-8 using a self-adaptive geospatially local model. Geosci. Front. 13. https://doi.org/10.1016/j.gsf.2021.101286.

Wei, J., Li, Z., Li, K., Dickerson, R.R., Pinker, R.T., Wang, J., Liu, X., Sun, L., Xue, W., Cribb, M., 2021. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O3) pollution from 2013 to 2020 across China. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2021.112775.

Xiang, J., Zhou, J., Huang, S., 2021. The boundary layer height obtained by the spline numerical differentiation method using COSMIC GPS radio occultation data: a case study of the Qinghai-Tibet Plateau. J. Atmos. Sol. Terr. Phys. 215. https://doi.org/10.1016/j.jastp.2020.105535.

Xiao, Q., Geng, G., Xue, T., Liu, S., Cai, C., He, K., Zhang, Q., 2021. Tracking PM2.5 and O3 pollution and the related health burden in China 2013–2020. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c04548.

Xue, L.K., Wang, T., Zhang, J.M., Zhang, X.C., Deliger, Poon, C.N., Ding, A.J., Zhou, X.H., Wu, W.S., Tang, J., Zhang, Q.Z., Wang, W.X., 2011. Source of surface ozone and reactive nitrogen speciation at Mount Waliguan in western China: new insights from the 2006 summer study. J. Geophys. Res. 116. https://doi.org/10.1029/2010jd014735.

Xue, T., Zheng, Y., Geng, G., Xiao, Q., Meng, X., Wang, M., Li, X., Wu, N., Zhang, Q., Zhu, T., 2020. Estimating spatiotemporal variation in ambient ozone exposure during 2013–2017 using a data-fusion model. Environ. Sci. Technol. 54, 14877–14888. https://doi.org/10.1021/acs.est.0c03098.

Yang, P., Zhang, Y., Wang, K., Doraiswamy, P., Cho, S.H., 2019. Health impacts and cost-benefit analyses of surface O3 and PM2.5 over the U.S. under future climate and emission scenarios. Environ. Res. 178, 108687. https://doi.org/10.1016/j.envres.2019.108687.

Yang, Y., Li, M., Wang, H., Li, H., Wang, P., Li, K., Gao, M., Liao, H., 2022. ENSO modulation of summertime tropospheric ozone over China. Environ. Res. Lett. 17. https://doi.org/10.1088/1748-9326/ac54cd.

Zang, Z., Guo, Y., Jiang, Y., Zuo, C., Li, D., Shi, W., Yan, X., 2021. Tree-based ensemble deep learning model for spatiotemporal surface ozone (O3) prediction and interpretation. Int. J. Appl. Earth Obs. Geoinf. 103. https://doi.org/10.1016/j.jag.2021.102516.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M.L., Zhang, M., Di, B., 2018. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. Environ. Pollut. 233, 464–473. https://doi.org/10.1016/j.envpol.2017.10.029.

Zhang, Y., Li, Z., 2015. Remote sensing of atmospheric fine particulate matter (PM2.5) mass concentration near the ground from satellite observation. Remote Sens. Environ. 160, 252–262. https://doi.org/10.1016/j.rse.2015.02.005.

Zhang, X.Y., Zhao, L.M., Cheng, M.M., Chen, D.M., 2020. Estimating ground-level ozone concentrations in eastern China using satellite-based precursors. IEEE Trans. Geosci. Remote Sens. 58, 4754–4763. https://doi.org/10.1109/tgrs.2020.2966780.

Zhang, Y., Li, Z., Bai, K., Wei, Y., Xie, Y., Zhang, Y., Ou, Y., Cohen, J., Zhang, Y., Peng, Z., Zhang, X., Chen, C., Hong, J., Xu, H., Guang, J., Lv, Y., Li, K., Li, D., 2021. Satellite remote sensing of atmospheric particulate matter mass concentration: advances, challenges, and perspectives. Fundam.Res. 1, 240–258. https://doi.org/10.1016/j.fmre.2021.04.007.

Zhao, Z., Wang, Y., 2017. Influence of the West Pacific subtropical high on surface ozone daily variability in summertime over eastern China. Atmos. Environ. 170, 197–204. https://doi.org/10.1016/j.atmosenv.2017.09.024.

Zhu, B., Hou, X., Kang, H., 2016. Analysis of the seasonal ozone budget and the impact of the summer monsoon on the northeastern Qinghai-Tibetan Plateau. J.Geophys.Res.Atmos. 121, 2029–2042. https://doi.org/10.1002/2015jd023857.