

Article

A Framework to Predict High-Resolution Spatiotemporal PM_{2.5} Distributions Using a Deep-Learning Model: A Case Study of Shijiazhuang, China

Guangyuan Zhang ¹, Haiyue Lu ², Jin Dong ³, Stefan Poslad ¹, Runkui Li ^{3,4}, Xiaoshuai Zhang ¹ and Xiaoping Rui ^{2,*}

¹ IoT Laboratory, School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK; guangyuan.zhang@qmul.ac.uk (G.Z.); stefan.poslad@qmul.ac.uk (S.P.); xiaoshuai.zhang@qmul.ac.uk (X.Z.)

² School of Earth Sciences and Engineering, Hohai University, Nanjing 211000, China; 191309010002@hhu.edu.cn

³ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; dongjin19@mails.ucas.ac.cn (J.D.); lirk@ucas.edu.cn (R.L.)

⁴ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: ruixp@hhu.edu.cn

Received: 30 June 2020; Accepted: 29 August 2020; Published: 31 August 2020



Abstract: Air-borne particulate matter, PM_{2.5} (PM having a diameter of less than 2.5 micrometers), has aroused widespread concern and is a core indicator of severe air pollution in many cities globally. In our study, we present a validated framework to predict the daily PM_{2.5} distributions, exemplified by a use case of Shijiazhuang City, China, based on daily aerosol optical depth (AOD) datasets. The framework involves obtaining the high-resolution spatiotemporal AOD distributions, estimation of the spatial distributions of PM_{2.5} and the prediction of these based on a convolutional long short-term memory (ConvLSTM) model. In the estimation part, the eXtreme gradient boosting (XGBoost) model has been determined as the estimation model with the lowest root mean square error (RMSE) of 32.86 $\mu\text{g}/\text{m}^3$ and the highest coefficient of determination regression score function (R^2) of 0.71, compared to other common models used as a baseline for comparison (linear, ridge, least absolute shrinkage and selection operator (LASSO) and cubist). For the prediction part, after validation and comparison with a seasonal autoregressive integrated moving average (SARIMA), which is a traditional time-series prediction model, in both time and space, the ConvLSTM gives a more accurate performance for the prediction, with a total average prediction RMSE of 14.94 $\mu\text{g}/\text{m}^3$ compared to SARIMA's 17.41 $\mu\text{g}/\text{m}^3$. Furthermore, ConvLSTM is more stable and with less fluctuations for the prediction of PM_{2.5} in time, and it can also eliminate better the spatial predicted errors compared to SARIMA.

Keywords: PM_{2.5}; AOD; XGBoost; prediction; deep learning; ConvLSTM; SARIMA

1. Introduction

In recent years, many cities globally, e.g., in China, have suffered an increase in air pollution characterized by an increase in air-borne particulate matter [1], as this is easily perceivable by humans via sight and ease of breathing, provoking a widespread concern about the occurrence of more frequent severe air pollution incidents [2]. PM, an acronym for particulate matter, and PM_{2.5}, which means

particulate matter with a kinetic equivalent diameter of 2.5 microns or less in the air, have been a focus for researchers and governments for several years who are concerned about air pollution. The concentration of $PM_{2.5}$ (we use $PM_{2.5}$ only to represent $PM_{2.5}$ in the remaining paper) is mainly affected by various pollution sources and meteorological conditions such as automobile exhaust and wind speed [3–5]. There is a positive correlation between the concentration of fine particles and the incidence and mortality of humans due to cardiopulmonary and respiratory system diseases [6,7]. According to research, $PM_{2.5}$ can penetrate into the lungs and bronchus. Long-term exposure to $PM_{2.5}$ can increase the incidence and mortality of respiratory and cardiovascular diseases [2]. $PM_{2.5}$ has a small diameter and mass; it remains in the atmosphere for a long time and can be air-borne over longer distances as it is so light. Hence, it can seriously affect the visibility of the atmosphere [8] and people's health and daily activities [9]. Hence, there is an urgent need to predict $PM_{2.5}$ quickly and accurately [10,11].

However, predicting future levels of $PM_{2.5}$ is challenging, and there are several key limitations of the current work. First, many studies focus on a prediction based on data from air quality stations, which means they only predict the temporal distribution but not the spatial distribution, because such stations are very sparsely distributed. They hardly consider the spatial autocorrelation between the predictors, which can cause a much greater error in the prediction. Second, in terms of work to predict future spatial $PM_{2.5}$ distributions, they do not consider the $PM_{2.5}$ estimation to get the distributions in the whole continuous space. This entails using $PM_{2.5}$ values obtained from an air pollution monitoring station to estimate the $PM_{2.5}$ values when there is no station to obtain the data at specific spatial points. Note that this is a kind of prediction, but in this study, we call this spatial prediction an estimation. Combining the first and second issues, there is currently no complete framework that not only estimates the spatially continuous $PM_{2.5}$ distributions but, also, predicts the distributions in the future. Third, the spatial and temporal resolutions for the predicted $PM_{2.5}$ distributions do not have a high enough resolution—for example, daily for the temporal resolution combined with kilometers for the spatial resolution. Finally, the models that are used in the estimation and prediction processes often lack a base-line method for comparison and validation to better assess the accuracy of the new proposed models. Such challenges and limitations in the current research (discussed in more detail in Section 2) comprise the motivation for this current work.

China has realized the harmful influence from excessive $PM_{2.5}$ levels since 2013, and the government has made great efforts to reduce harmful emissions and, hence, haze. Despite the initial success in recent years, the exposure level of $PM_{2.5}$ in many areas of China, e.g., Hebei Province, is four times higher than the World Health Organization (WHO) standard in 2019, according to a report from the Centre for Research on Energy and Clean Air (CREA) (<https://energyandcleanair.org/wp/wp-content/uploads/2020/01/CREA-brief-China2019-Zh.pdf>). As the capital of the Hebei Province, Shijiazhuang was one of the areas most polluted by haze in China in 2016 [12]. The city still suffers from the most serious haze pollution amongst all the cities in China in recent years, especially in the first half of 2019, with a $39\text{-}\mu\text{g}/\text{m}^3$ monthly average concentration of $PM_{2.5}$ based on the National Urban Air Quality Report in June 2019 of China (<http://www.mee.gov.cn/hjzl/dqhj/ckqzlkzyb/201908/P020190821498490317309.pdf>). Thus, in this study, based on the raw datasets, including the daily aerosol optical depth (AOD), $PM_{2.5}$ from the monitoring sites, etc., we created a framework to predict the daily $PM_{2.5}$ distributions, exemplified for a use case for the city of Shijiazhuang in China. The time period of the study focuses on the first half of the year 2019. We aim to demonstrate (test) the prediction capability of our framework to predict about one month's of $PM_{2.5}$. The test data represents 20% of the whole dataset in many prediction models. Hence, we just need to use about five months' data in our study. So, we use only the first five months' data of 2019, which is from the 1 January 2019 to 31 May 2019 (151 days) in the case.

In this study, we aim to present a framework that not only estimates the spatially continuous $PM_{2.5}$ distributions but, also, predicts the distributions in the future at resolutions in time and space that are both high-resolution (this is our main novelty, but see also the discussion in Section 5). For both processes, we compare some popular regression models, machine-learning models and deep-learning

methods to select the most accurate ones for use in our framework. In the prediction process, we adopt a convolution long short-term memory (ConvLSTM) model, which is an improvement of the long short-term memory (LSTM) model by adding convolution operations to predict the temporal and spatial distributions of $PM_{2.5}$ for future day(s) based upon referring to historical data.

The remainder of this article is organized as follows: Section 2 presents a literature review. Section 3 presents the dataset we used and the method of our framework. The results are reported in Section 4, and these are discussed in Section 5. In Section 6, our conclusions are proposed, and our thoughts for future research are presented.

2. Literature Review

At present, $PM_{2.5}$ data is mainly obtained from ground-monitoring stations and satellite data [13]. $PM_{2.5}$ data from ground-monitoring stations is distributed as points. It is hard to gain a good spatial coverage using the limited number of fixed monitoring facilities to obtain global $PM_{2.5}$ data; therefore, spatial interpolation is used to deduce unknown data, based on known data in the same region, to make up for this deficiency. Commonly used methods to do this are inverse distance weighted (IDW) interpolation, trend surface (TS) interpolation, ordinary kriging (OK) interpolation, collaborative kriging (CK), radial basis function and so on. Obtaining PM_{10} and total suspended particles (TSP) data is considered easier than obtaining $PM_{2.5}$ and estimating missing $PM_{2.5}$. Hwa Lung et al., [14] proposed a Bayesian maximum entropy (BME) algorithm to calculate the ratio of $PM_{2.5}/PM_{10}$ and $PM_{2.5}/TSP$ in Taipei, integrated with PM_{10} and TSP to interpolate $PM_{2.5}$ data, to retrospectively estimate the spatial and temporal distributions of $PM_{2.5}$ in previous years. The stability of $PM_{2.5}/PM_{10}$ and $PM_{2.5}/TSP$ ratios is based on a yearly time period, instead of shorter time periods such as months, weeks and days. In areas where there is no $PM_{2.5}$ -monitoring facility, the most commonly used ancillary data is AOD (aerosol optical depth), which is the integration of the aerosol extinction coefficient in the vertical direction of the atmosphere that is relevant to the radiation wavelength, vertical profile, particle size distribution and aerosol size [13]. It has been found that the particle size, especially the $PM_{2.5}$ particle size range, is closely related to the inversion of AOD from 0.1–2 nm in the visible and near-infrared band ranges. The correlation between $PM_{2.5}$ and AOD is influenced by numerous elements such as meteorological factors (AOD vertical profile, temperature, humidity, wind speed, etc.) and geographical factors (regional categories, road distribution, forest cover, etc.), which are important to build the relationship between $PM_{2.5}$ and AOD. Therefore, establishing a relevant prediction model based on these auxiliary factors and AOD data can effectively obtain $PM_{2.5}$ data, so as to monitor and predict the $PM_{2.5}$. Considering that monitoring stations are limited and unevenly distributed, Rui et al., [15] used the AOD data, introduced a multiple linear regression model and determined the relationship between $PM_{2.5}$, AOD, meteorological factors and physical and chemical factors. Finally, these researchers established a quantitative model to interpolate $PM_{2.5}$ in Beijing. The results show that this model can more accurately analyze the spatial and temporal distributions of $PM_{2.5}$ in the study area. However, the model does not consider changes in time and the spatial distribution. The resolution of satellite data (from which the AOD data is derived) and the detailed composition of $PM_{2.5}$ both play a significant role in raising the precision of the model.

Much research has been undertaken to predict $PM_{2.5}$. Yuanhua et al., [16] adopted a back propagation (BP) artificial neural network to predict $PM_{2.5}$ in Beijing and found that it can reflect the variation of $PM_{2.5}$ well, and based upon this, the $PM_{2.5}$ is forecast. However, this model has high requirements for the rationality of the method, and the configuration of the model is complicated. Hongfu et al., [17] established a prediction model based on a GM (1,1) model where GM means grey model—a grey forecasting model. They used historical $PM_{2.5}$ data to predict $PM_{2.5}$ in Changchun, China for the next two days. The results showed that this model could be used to forecast haze events, but the temporal resolution was a little low over one day or two days, instead of one hour or two hours. Bingyue [18] used the eXtreme gradient boosting (XGBoost) model to monitor the air quality data in Tianjin and predict $PM_{2.5}$. This model has high precision, low overfitting probability

and better performance compared to other models in terms of the numerical calculations. However, only one monitoring station data is used in this research; the data is not multivariate, resulting in the limited prediction ability of the XGBoost model, in this case. Huang et al., [19] found that the forecast guidance provided by the NOAA (National Oceanic and Atmospheric Administration) NAQFC (National Air Quality Forecasting Capability) has significant seasonal biases: in winter, there is an overprediction phenomenon, while, in summer, there is an underprediction phenomenon. To reduce the bias, the researchers integrated an analog ensemble bias correlation approach with NAQFC to predict $PM_{2.5}$ in the Lower Middle, the Upper Middle, the Southeast, the Northeast, the Pacific Coast and the Rocky Mountains of the United States. The results show that this analog ensemble bias correlation approach can improve the prediction accuracy.

The consideration of spatial data is also important for the temporal prediction model. Lei et al., [20] proposed a spatial data-aided incremental support vector regression (SaIncSVR) model to predict $PM_{2.5}$ over 13 monitoring stations in Auckland, New Zealand. It is found that the model can satisfactorily deal with the short-term and missing data problems that exist in many prediction models compared with a pure temporal IncSVR prediction model. However, this model does not take into account the geographical features of the monitoring station, which are vital for the prediction ability of the model. Zong et al., [21] introduced a RNN (recurrent neural network) model, aiming to establish a universal prediction model by using the meteorological data and $PM_{2.5}$ data of Beijing, Chengdu and Shenyang. This found that the prediction model based on data from one of the cities can be generalized to the other two cities, indicating that there is a close intrinsic correlation between $PM_{2.5}$ source-sink dynamics (defined as a theoretical model to describe how $PM_{2.5}$ may affect the population growth or decline of organisms) and environmental drivers. In addition, this correlation is common among cities. Machine-learning or deep-learning models have a strong expressive ability when processing nonlinear data [22], but different models have a similar prediction accuracy when they use the same datasets. In order to improve the accuracy, Yegang [23] established a relationship between the AOD, hourly forecast value of meteorological factors and $PM_{2.5}$ based on an adaptive BP neural network model to predict $PM_{2.5}$ in Fuling, Chongqing. This model supports adaptive training and tuning; therefore, it can suppress the overfitting phenomenon well, but the prediction of the $PM_{2.5}$ time series is influenced by numerous factors such as temperature and people's activity, and the amount of historical data is insufficient, which reduces the accuracy of the model. Wei et al., [24] proposed using a PCA (principal component analysis) model and a least-squares support-vector machine (LSSVM) model as an improved support-vector machine (SVM) model. Here, PCA accurately extracts useful information and reduces the dimensionality of the input layer, and LSSVM reduces the computational complexity; this hybrid strategy not only improves the prediction accuracy but, also, greatly increases the prediction speed.

At present, there are many models that can be used to study, analyze and predict $PM_{2.5}$, but few of these consider spatial autocorrelation, and most of the models have a low temporal accuracy, which cannot be used to predict $PM_{2.5}$ in the next few hours or even days, which is of little use to the establishment of air pollution warning systems. R.A. Bahari et al., [25] proposed a multilayer perceptron (MLP) artificial neural network. Temperature inversion as a parameter was added to the model to predict $PM_{2.5}$ in the Tehran region for the next three days. It was found that temperature inversion can improve the model and that the prediction accuracy was improved. This model takes 12 h as a temporal unit, draws the map of the temperature, wind direction, wind speed and so on and gets the temperature inversion; finally, the $PM_{2.5}$ in the next hours was predicted. However, the prediction period of 12 h cannot adapt to a shorter period such as one hour. Ping et al., [26] proposed a hybrid strategy based on a high-dimension association rules (HDAR) model, a modified association framework (MAF) model, a learning vector quantization (LVQ) model and an adaptive fuzzy neural network (AFNN) model, termed HML-AFNN (HML is the combination of these three methods' initial), to analyze and predict $PM_{2.5}$ in the Beijing–Tianjin–Hebei region and the Pearl River Delta region. This works as follows: (1) The HDAR model selects the cities where $PM_{2.5}$ has strong

correlations with weather factors within the study area. (2) The MAF model selects the spatial-temporal factors and geographical factors that affect $PM_{2.5}$ in the center of the study area from the above cities. (3) The LVQ model divides all datasets into several datasets according to the $PM_{2.5}$. (4) The AFNN model analyzes and predicts the $PM_{2.5}$ based on the above datasets. The results show that this hybrid strategy has a better performance than any single model it uses, but it does not consider the spatial autocorrelation between variables, so its prediction accuracy is not very high.

Chaotic phenomenon, a dynamic, complex form of uncertainty in complex nonlinear systems, is a common phenomenon in nature. Lorenz et al., illustrated this with an example of a flying butterfly in South America that could cause a hurricane in Florida in North America. This is the so-called butterfly effect—that is, the chaotic system is sensitive to the initial value; a disturbance will completely deviate from the original evolution direction after a long time, no matter how small it is [27]. To improve the prediction accuracy, Yun et al., [28] introduced a multivariate chaotic time series model based on chaotic theory to predict $PM_{2.5}$ in Beijing. The phase space unit of the chaotic time series is first expanded into a multi-time series phase space unit. Then, based on this, a multi-time series phase space matrix is constructed. Finally, the radial basis function (RBF) neural network is introduced to predict $PM_{2.5}$. This is better than a traditional time-series prediction model; although it considers some indicators, such as air pressure, temperature, wind direction, wind speed, dew point and so on and has a high temporal accuracy, it does not consider the spatial autocorrelation between variables, which is also vital to the improvement of the prediction accuracy.

Considering that $PM_{2.5}$ data has strong nonlinear characteristics, and there are many difficulties to monitor and acquire it, Haiming et al., [29] introduced a RBF neural network model, which improved the classical BP neural network to imbue the model with a local learning ability. Normal air pollution monitoring data and meteorological factors were considered as variables to predict $PM_{2.5}$. The results demonstrate that the RBF model has a stronger prediction ability compared to the BP model, but because of the lack of samples, the prediction accuracy of some samples is reduced, which makes the model difficult to adapt to complex dynamic weather conditions. There is some redundant information, such as unused weather condition values in meteorological environmental data, so it is useful to filter the data and eliminate the redundant information before predicting $PM_{2.5}$ based on this. Chen et al., [30] proposed a combined multifractal dimension–artificial bee colony–support vector regression (MFD + ABC + SVR) hybrid strategy. MFD + ABC chooses the best feature dataset, in which MFD is used as the evaluation criterion for selecting the dataset. ABC provides the search strategy. Finally, an SVR model is used to predict $PM_{2.5}$ the next day in Guangzhou and Shanghai. This hybrid strategy optimizes the process of the input layer and improves the prediction accuracy, but the temporal accuracy is not very high, and it does not consider the spatial autocorrelation.

The ConvLSTM model is an improved LSTM model that has good spatial-temporal characteristics. It not only exhibits the temporal modeling ability of a LSTM, but it can also consider the influence from the spatial neighborhood of a target area. Liu et al., [31] used a ConvLSTM model to analyze historical traffic flow data and to make predictions, confirming that the ConvLSTM model has a better performance than other models, such as LSTM. Qiao et al., [32] considered spatial-temporal autocorrelation for the monitoring of health and proposed a time-distributed ConvLSTM model to extract the spatial-temporal characteristics of a multi-sensor time series. It is found that the model has a good applicability to health monitoring. Yuan et al., [33] proposed a heterogeneous ConvLSTM framework to integrate spatial image features and spatial model sets to solve the spatial heterogeneity problem in urban and rural areas. The results show that this framework can make reasonable predictions and effectively improve the prediction accuracy. Yuan et al., [34] adopted a ConvLSTM model to predict the spatial and temporal distributions of mobile phone users a day in advance based upon 15-day historical CDR (Call Detail Record) data. The results show that a ConvLSTM model that considers a spatial autocorrelation is more realistic and accurate compared with an autoregressive–moving-average model (ARMA) model and LSTM model.

Some models that predict $PM_{2.5}$ focus on decreasing the redundancy of input data and increasing the number of the related influence variables to improve the prediction accuracy, but models that improve the temporal accuracy and spatial autocorrelation are not used widely. Therefore, in this paper, we adopt a ConvLSTM model, which considers the spatial autocorrelation, to predict a $PM_{2.5}$. It has a high temporal precision hourly. The prediction time can be extended to 24 h or even several days. The LSTM model has the structure of a recursive neural network. It can process time series data, exhibiting time autocorrelation defined to represent the degree of similarity between a time series and a lagged version of itself over successive time intervals. A ConvLSTM model improves on a LSTM model through adding a convolution operation to the basic structure of the LSTM model, such that it not only establishes temporal relationships such as a LSTM but, also, extracts spatial temporal characteristics such as a CNN. This paper collects historical $PM_{2.5}$ data and inputs it into the ConvLSTM model to predict the temporal and spatial distributions of $PM_{2.5}$ in the next day or over several days.

3. Data and Method

In this section, we introduce the datasets and the method. We use 7 types of raw datasets: including the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm-based Level 2 gridded (L2G) aerosol optical thickness over land surfaces product (MCD19A2) data, the Goddard Earth Observing System forward-processing (GEOS FP) data, air quality observation data, surface meteorology observation data, SPOT/VEGETATION (Satellite Pour l'Observation de la Terre Vegetation), ERA-nterim (ECMWF Reanalysis Interim) and the Shuttle Radar Topography Mission (SRTM) datasets. These datasets are described in more detail in Section 3.1. Our method consists of 3 main parts, divided into 9 smaller steps (see Section 3.2).

3.1. Data

3.1.1. Study Area and Period

In this study area, we focused on the city of Shijiazhuang, China, but because of the shape of the city—it is an irregular polygon, while the predicted cell in this study consists of squares—the main study area was set as a big square area with 165 km in length, which covered all the Shijiazhuang area inside of it. The location of the city and the administrative units are shown in Figure 1. The study period was from 1 January 2019 to 31 May 2019, totaling 5 months and 151 days. The details of the data and the method of the framework are introduced in the next two subsections.

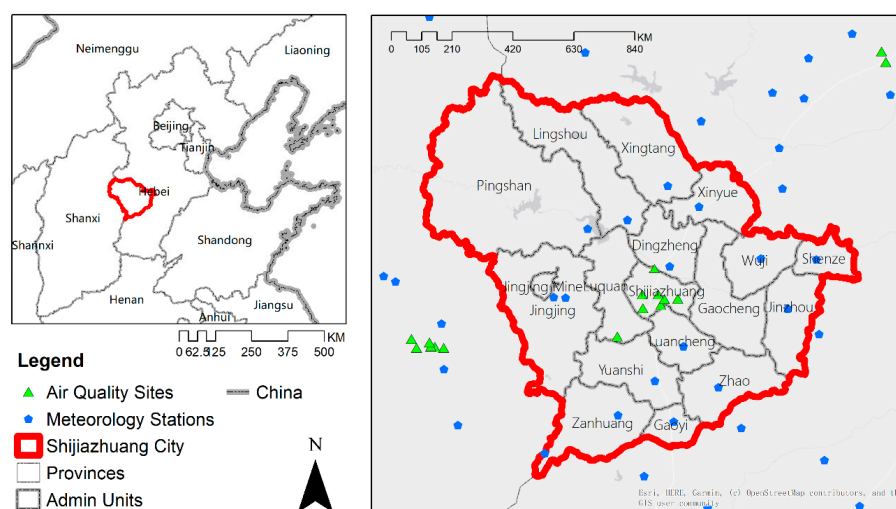


Figure 1. The study area (Shijiazhuang City) and the distributions of the air quality sites and meteorology stations in and around it.

3.1.2. Data Sources

MCD19A2 (<https://ladsweb.modaps.eosdis.nasa.gov>) is the abbreviation for the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm-based Level 2 gridded (L2G) aerosol optical thickness over land surfaces product, which is produced daily at a 1-km pixel resolution (<https://lpdaac.usgs.gov/products/mcd19a2v006/>). This product is a Moderate Resolution Imaging Spectroradiometer (MODIS) Terra and Aqua combined, which can make up for cloud coverage or missing data in other situations and has the highest temporal resolution and has the spatial resolution that is available to the public. The following Science Dataset (SDS) layers are involved in the MCD19A2 AOD data product: blue band AOD at 0.47 μm , green band AOD at 0.55 μm , fine mode fraction over water, AOD uncertainty, smoke injection height (m above ground), column water vapor over land and clouds (in cm), AOD QA (quality assurance), cosine of the solar zenith angle, AOD model at 900 m, cosine of the view zenith angle, scattering angle, relative azimuth angle and glint angle at 5 km. The product L2G also includes a low-resolution image to show the AOD of the blue band at 0.47 μm , which is created using a combination of all available satellite sensor orbits. From the product, we can obtain high-temporal and -spatial resolution spatial AOD datasets, i.e., daily for time and 900 m for space. In this study, we used the AOD at 0.55 μm , where AOD is a daily dataset that contains the data collected when a satellite passes over the image area, so the daily data contains different amounts of data (depending on the number of satellite transits in a day). In order to unify the temporal intervals between AODs over the study period, we computed the daily average AODs and converted them into an image format. There was a total of 151 images of AOD from 1 January to 31 May 2019.

GEOS FP files are produced with the Network Common Data Form (NetCDF-4) library; the underlying format of this library is the hierarchical data format version 5 (HDF-5). The standard of the GEOS FP file is `tavg1_2d_flux_Nx` (2D time-averaged surface flux diagnostics). This file contains: the planetary boundary layer height (PBLH), surface layer height (HLML), total precipitation (PRECTOT) and so on. Based on the GEO FP data in the Global Modeling and Assimilation Office (GMAO) (<https://gmao.gsfc.nasa.gov/>), we can obtain hourly planetary boundary layer height (PBLH) data from January 2019 to May 2019 in Shijiazhuang.

National air quality observation data in our study comes from the National Urban Air Quality Real-time Publishing Platform (<http://106.37.208.233:20035/>) of the China Environmental Monitoring Sites, including hourly $\text{PM}_{2.5}$, PM_{10} , SO_2 , NO, CO, etc. values. In our study area, there were 16 air quality sites, and we extracted $\text{PM}_{2.5}$ values on the sites as the dependent variables in the regression model. The temporal resolution of the datasets was one hour. The distribution of the air quality sites is shown in Figure 1.

The surface meteorology observation data was from the National Meteorological Science Data Center of China (<https://gmao.gsfc.nasa.gov/>), including hourly observations of air pressure, air temperature, relative humidity, wind speed, water vapor pressure precipitation and other factors. All of this data was obtained from surface meteorology stations. In our study area, there were 31 stations, and we extracted the wind speed (m/s) and precipitation (mm) from the stations as the key independent variables in the regression model according to previous studies [35], and the relative humidity (percentage) values were extracted for correction of the AOD. The temporal resolution of the datasets was set to be one hour. The distribution of the meteorology stations is shown in Figure 1.

The NDVI (normalized difference vegetation index) can precisely mirror the surface vegetation coverage. This dataset is derived from the VEGETATION sensor on-board the SPOT satellite platform [36]. The measurements of land surface reflectance in the visible and in the infrared domains can be obtained from the VEGETATION instruments on-board SPOT4 (launched on April 1998) and VEGETATION2 on-board SPOT5 (since February 2003) [37]. At present, the NDVI time series data has been extensively used for the study of land use or cover change detection, vegetation dynamic change monitoring, macro-vegetation cover classification and net primary productivity estimation. This dataset effectively reflects the distribution and diversification of vegetation coverage in different areas of the country at different spatial and temporal scales. In this investigation, we used the

annual NDVI data in 2018 for China with a 1-km² spatial resolution (<http://www.resdc.cn/DOI>, 2018, doi:10.12078/2018060601).

In this study, the daily albedo (a measure of the amount of light that hits a surface that is reflected without being absorbed) data that covers the study area and period is needed. This data can be obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF), which reanalyzes interim (ERA-interim) daily data with a horizontal resolution of 0.25° × 0.25°. This data with a global coverage is suitable for implementing climate studies in different areas of the world because of its long-term availability [38]. ERA-interim monthly averaged globally gridded meteorological data (fully described by Berrisford et al., [39]), from 1 January 2019 to 31 May 2019, is used to extract the corresponding daily albedo raster datasets.

The Shuttle Radar Topography Mission (SRTM) is a joint project between the National Geospatial-Intelligence Agency (NGA) and National Aeronautics and Space Administration (NASA), which offers an important step forward in the generation of digital elevation model (DEM) SRTM data [40,41]. In this study, to estimate PM_{2.5}, the elevation data is extracted as a significant feature from 1:1,000,000 geomorphological maps of China [42]. The spatial resolution of the data is 90 m.

3.2. Method

The framework consists of 3 parts and 9 steps based upon the methods and the process introduced in Figure 2. Part 1 illustrates the process to obtain high-resolution spatiotemporal AOD distributions. Based on this, the process to build a regression model to compute PM_{2.5} distributions is presented in part 2. Finally, based upon training high-resolution spatiotemporal PM_{2.5} distributions, we use ConvLSTM to predict the test distributions, compared with a seasonal autoregressive integrated moving average (SARIMA) model as a baseline to test the accuracy of the prediction.

3.2.1. Part 1: High-Resolution AOD Acquisition and Correction

The aim of the first part is to compute the spatiotemporal high-resolution AOD distributions with correction processes for the reasons as follows: the air humidity has an effect on the aerosol optical depth (AOD), with an increase in humidity, size of hygroscopicity and dissolved aerosol, the particles will increase accordingly. There is a positive correlation between air humidity and aerosol optical depth [43,44]. Besides, the PBLH has an impact on the relationship between surface PM_{2.5} and AOD. The greater the PBLH, the greater the AOD, but the surface PM_{2.5} could be low [45,46]. Considering such effects, we use relative humidity and PBLH to correct the AOD, which can make the relationship between AOD and PM_{2.5} clearer. There are three steps to compute the corrected AOD distributions.

In the first step, we process the raw datasets of MCD19A2, GEOS FP and surface meteorology observation data. We select and define AOD that is extracted from the MCD19A2 as the AOD-0 dataset. Then, based on the GEOS FP data, the natural neighbor interpolation [47,48] method is utilized to convert the PBLH into the same raster format files as the AOD dataset, which makes it easier to correct the AOD. In terms of the surface meteorology observation data, we only extract the hourly relative humidity data from the 31 stations, then compute the average values of the relative humidity for each day in every station. For the daily average relative humidity, we use the inverse distance weighting (IDW) [49,50] method to spatially interpolate the points data to the whole of the study area. The final relative humidity distributions are the image files where a cell represents a 900-m-square space area, which is the same as for the AOD-0 images.

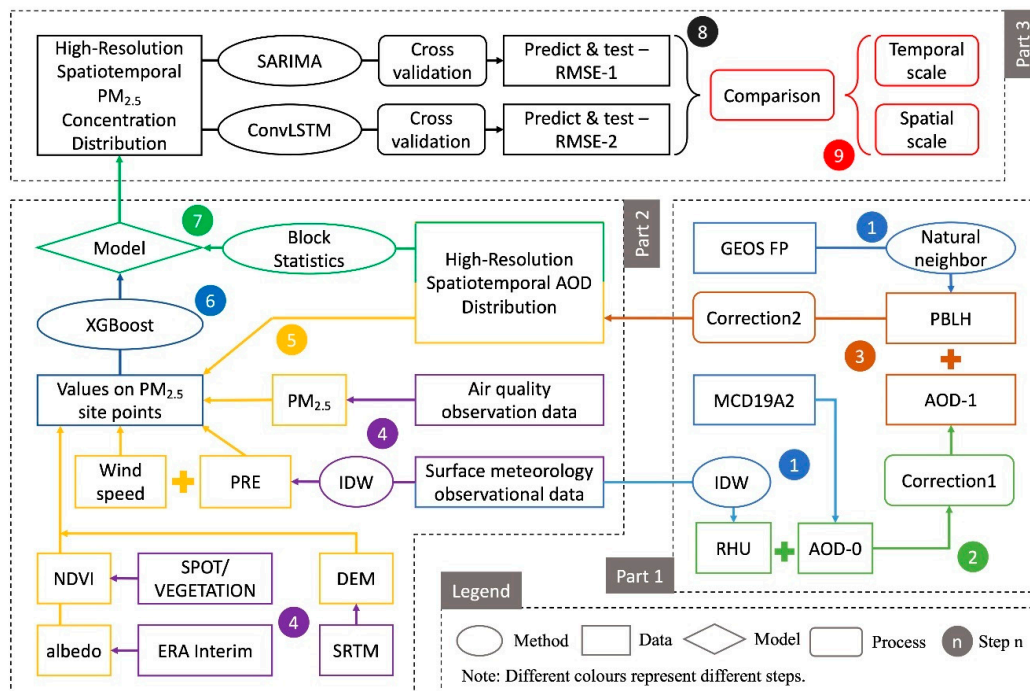


Figure 2. The workflow of the framework to predict PM_{2.5}. The definitions of abbreviations are as follows: GEOS FP (Goddard Earth Observing System forward-processing), PBLH (planetary boundary layer height), MCD19A2 (Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm-based Level 2 gridded (L2G)), AOD (aerosol optical depth), IDW (inverse distance weighted), RHU (relative humidity), PRE (precipitation), NDVI (normalized difference vegetation index), DEM (DEM), SPOT/VEGETATION (Satellite Pour l’Observation de la Terre Vegetation) ERA (ECMWF Reanalysis), SRTM (Shuttle Radar Topography Mission), XGBoost (eXtreme gradient boosting), ConvLSTM (convolutional long short-term memory), SARIMA (seasonal autoregressive integrated moving average). And the correspondence of the colors and steps are as follows: step 1 (blue), 2 (green), 3 (brown), 4 (purple), 5 (yellow), 6 (dark blue), 7 (dark green), 8 (black) and 9 (red).

In the second step, we correct the AOD-0 for the first time by using relative humidity distribution datasets. The correction equation is taken from [51,52]:

$$AOD_1 = AOD_0 * \left(1 - \frac{RH}{100}\right) \quad (1)$$

where the *RH* represents the relative humidity, *AOD*₀ is the AOD-0, and *AOD*₁ is the AOD-1. In terms of each corresponding cell, we use Equation (1) to correct AOD-0; then, we get the AOD-1 distributions.

The third step is based on the AOD-1 and PBLH datasets; we use the equation below to correct the AOD again [45,52]:

$$AOD_2 = \frac{AOD_1}{PBLH} \quad (2)$$

where *AOD*₁ is the AOD-1, and *AOD*₂ is the AOD-2. We get the AOD-2 distributions in the final step of part 1, which are the high-resolution spatiotemporal AOD corrected distributions.

3.2.2. Part 2: Regression Modeling to Compute High-Resolution PM_{2.5}

PM_{2.5} has a high correlation with AOD [43,51,53,54], but other meteorological factors also influence the concentration of PM_{2.5}. However, we can only obtain the accurate ground PM_{2.5} values from the sparsely populated ground air quality station sites. If we want to get the spatially continuous distributions of PM_{2.5}, we need to estimate the concentration based on the AOD and other key factors [13]. Thus, the purpose of this part is to build a regression model according to the ground truth

PM_{2.5} data based upon the air quality station site points, and then, the model is used to estimate the whole distribution in the whole city area. There are 4 steps to this part 2, beginning with step 4 and following from the 3 steps in part 1. For the fourth step, we preprocess the air quality observation data to extract the hourly PM_{2.5} values on the 16 air quality sites. The surface meteorology observation data are used again to extract the wind speed and precipitation values. We still use IDW to obtain the raster format image of the wind speed and participation values over the study period.

In the fifth step, we compute the daily average values of PM_{2.5}, wind speed and precipitation for the 16 sites in 151 days. The same preprocessing is used for the AOD-2 raster images. The PM_{2.5}-AOD relationship can establish a multivariate function, which is related to several influencing factors, according to previous studies [55–57]. Hence, the following parameters are selected to improve the PM_{2.5} estimation in our study, based upon albedo, precipitation, NDVI, wind speed and elevation, which are constructed for modeling, where PM_{2.5} is regarded as the dependent variable and the others are used as the independent ones in the next step.

The wind speed has a significant influence on personal exposure levels in PM_{2.5}, because the kind of coarse particles from resuspension shows a positive dependence on the wind speed [58,59]. Precipitation has a significant impact on the concentrations of PM_{2.5} [60,61], because precipitation has a wet scavenging effect on particles, which is the major process by which the atmospheric self-purification and a balance between the sources and sinks of atmospheric aerosols are maintained [62,63]. Elevation is known to impact PM_{2.5} distributions, because PM_{2.5} is difficult to reach in higher elevations [64,65], which results in a big difference between the lower region and higher one. Vegetation planted in urban areas, urban forests and parks can remove particle pollutants [66], which means PM_{2.5} is influenced by the vegetation in a city. Thus, the lack of NDVI, which is a satellite-based greenness index that measures and monitors plant growth and vegetation density, might limit the capability of using green spaces to explain PM_{2.5} changing [67]. Furthermore, albedo also plays a key role in the estimation process of PM_{2.5} [35,68]. There is a strong correlation between the NDVI and surface albedo, i.e., with the severity of desertification increased, vegetation coverage reduced and then the soil moisture decreased, resulting in an increased albedo [69]. However, considering albedo is a response to the heterogeneity of different land surface types, while NDVI focuses on the influence from the vegetation on PM_{2.5}, they have different priorities, and both of them contribute to the estimation of PM_{2.5} distribution, which is important for use in a machine-learning model and to promote a good performance; hence, we use these two key factors at the same time.

There are $151 \times 16 = 2416$ sets of data points from the 151 days and 16 stations as the output of this step.

For the sixth step, we need to build an estimation model to estimate PM_{2.5} in the whole study area. Although we use the 2416 sets of data points, including humidity, albedo, precipitation, NDVI, wind speed and elevation; yet, in this step, there are many missing values in the AOD distributions. Some of the missing values' regions cover some 16 stations, where the corresponding AOD cannot be obtained. Thus, less than 2416 sets of data points would be input, but they are all independent variables, while the corresponding PM_{2.5} values are the dependent variables. The output is the relationship model that is trained by the input data. The most accurate model would be used to estimate continuous spatial PM_{2.5} distributions based on the continuous spatial AOD distributions. According to the study of Yongming et al., a cubist regression model is the best choice to estimate the concentrations of ground-level PM_{2.5} [35]. At the same time, some other studies also test that the cubist model performs well in similar situations [70,71]. Thus, in this study, we also select some popular traditional regression models and machine-learning methods to build the estimation model between the PM_{2.5} and other independent variables and then compare their accuracy.

The regression models that are used and compared in this step include: (1) linear, (2) ridge, (3) Least Absolute Shrinkage and Selection Operator (LASSO), (4) cubist and (5) eXtreme gradient boosting (XGBoost). LASSO is a variable selection and regularization method that can force some secondary coefficients to be zero in order to shrink coefficients [72]. It can increase the explainability

of the model and decrease overfitting. Cubist is a rule-based tree model, which uses Quinlan's M5 theory [73] to generate multiple linear regression models in the end nodes of the tree. When forecasting the terminal nodes, the corresponding linear regression model can be used. The corresponding linear regression model makes a prediction at the terminal node, which is then smoothed by combining it with predictions from nearest-neighbor nodes within the tree to improve the prediction precision [74]. In addition, the cubist also builds several tree models (called committees), in which rule-based models are built into each individual tree model [75]. The final forecast can be obtained by averaging the forecasts of all committees. XGBoost is an integration tree method based on the principle of a gradient lifting framework [76]. It can control overfitting and the complexity of the model by using regularization techniques [77].

In the seventh step, we built a model between the $PM_{2.5}$, AOD and other necessary weather conditions, such as wind speed and humidity, and then, we needed to use this model to estimate the spatially continuous $PM_{2.5}$ distributions for the whole of the study area. Since there are many missing values of AOD in space, we needed to interpolate the missing data based on the existing data. The frequency of the 151 AOD maps that have missing values (null values) are shown in Figure 3. In the histogram, we can see more than a half of AOD maps have more than 50% cells missing values, and more than 27% (more than 41) AODs have 90%–100% missing-value cells. There are 14 AOD maps that are totally null (100% missing-values cells). In terms of this issue, we propose a strategy as follows: If there are even consecutive AODs, the first half would be replaced by the nearest AOD that has values. However, if there are odd consecutive AODs that are all totally null, apart from the middle one, the remaining AODs change to the even consecutive AODs; then, we use the same way as the last strategy to replace only one middle AOD, which would be represented by the average AODs of the two nearest ones. For example, here are 1 to 10 AODs; within these, the 3rd, 4th, 7th, 8th and 9th AODs are totally null; then, because the 3rd and 4th are even consecutive AODs, the 3rd one would be replaced by the 2nd one, and the 4th one would be replaced by the 5th one. On the other hand, if the 7th, 8th and 9th are odd consecutive AODs, the 7th one would be replaced by the 6th one, and the 9th one would be replaced by the 10th one, while the middle one, the 8th AOD, would be computed as the average maps of the 6th one and the 9th one.

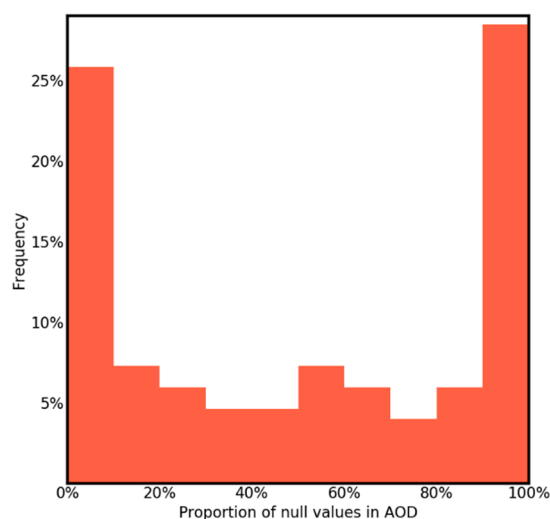


Figure 3. The frequency histogram of the AOD maps with missing values, where the orange bars illustrate the volumes of the frequency. For example, there are 40 AODs ($151 \times 26\% \approx 40$) that have 0%–10% missing value cells, which is shown as the first bar from the left.

In terms of the AODs that have at least one cell with a value (not a null value), here, we use a Block Statistics and Missing-value Padding (BSMP) method to turn the AODs into more spatially continuous distributions. The BSMP method applies to a raster grid or image and consists of two parts. The block

statistics tool conducts a neighborhood operation, which means inputting pixel calculation statistics information, where these pixels belong to a set of fixed nonoverlapping square areas, also called blocks or windows (in order to avoid ambiguity, we use “window” in the latter part of this paper). Such statistics (for example, maximum, average or sum) apply to all input pixels contained within each window. After getting the calculation result value of a single window, a window is specified, and the calculation result will be assigned to all pixel positions contained in the minimum boundary rectangle within the window. The missing-value padding (padded cell) is used to combine the original raster with the new raster that is generated after the block statistics. The cell that already has a value keeps the original value, while the missing-value cell obtains a new value from the block statistic. The workflow for the BSMP is shown in Figure 4.

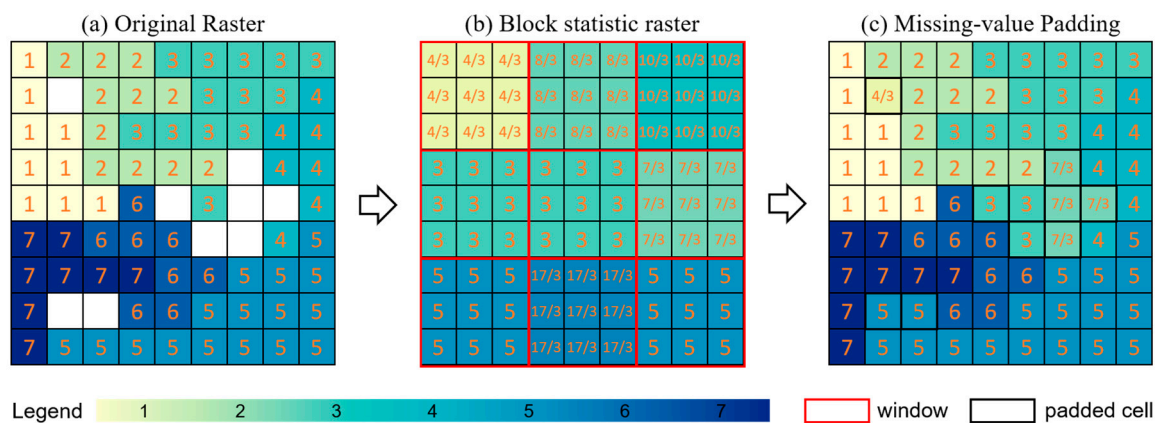


Figure 4. The workflow of the Block Statistics and Missing-value Padding (BSMP) method. Padded (white) cells in (a) have no values where (a) shows the original raster of a case of AOD distribution. (b) shows the process of the block statistic for the original raster, while the result of the BSMP after one round has been shown in (c).

The BSMP has two key characteristics. First, the size of the window can dictate the number of padded cells. The larger the window is, the larger the number of padded cells is. Second, according to Tobler’s First Law of Geography [78], “everything is related to everything else, but near things are more related to each other”.

There are a number of cells that have missing values in many cases of AOD distributions. The larger the window is, the lower the accuracy of the value in the padded cell that is computed. That is, we use the BSMP to let a cell that has no data get a value. If there are many cells that have data in the block, the non-cell can get a value with a higher accuracy, but if only a few cells exist, the non-cell can still get a value, but the accuracy is lower. In order to get higher, more accurate values for the padded cells, based on the characteristics of the BSMP method, a smaller-sized window is needed, which also means that only one round of BSMP would not be enough. Thus, one process might use BSMP for several rounds to make up all the missing values for one AOD distribution. However, with an increasing number of BSMP rounds, the computed value for the padded cell would get a lower accuracy, though the window size does not decrease. This is because the new round BSMP uses cells that just are filled. There are some new errors that are generated, but there were some old errors that were generated by the last round, so, the sum error increases with new rounds (defined as situation 1). However, if we enlarge the size of the window in the latter rounds of the BSMP (defined as situation 2), the cumulative error would be similar to strategy 1. At the same time, situation 2 can improve the efficiency of the BSMP, because it can decrease the number of BSMP rounds by increasing the size of the window.

Thus, combining the two key characteristics of the BSMP, we use multi-rounds of BSMP to interpolate the missing data based on the existing data, which, as the number of rounds increases, the size of the window increases as well, until all the missing values have been interpolated (see Section 4.1.2).

After applying the BSMP method, the missing values in the AOD are filled. In the final step, 7, of this part, we use the trained models that have been selected with the highest estimation accuracy to estimate the $PM_{2.5}$ based on the AOD datasets. In this process, we create a fishnet-like grid to divide the whole of the study area into square cells. Since the study region is a 165×165 km square, to make the data structure more suitable for the prediction models, the grid in this area consists of $50 \times 50 = 2500$ cells, and each cell is a 3300×3300 m square. Thus, all datasets would be reformatted from the original resolution (i.e., 900 m resolution of AOD) to 3300 m resolutions for every distribution that is in each time node (day), which is for the prediction part. Then, we can obtain the daily spatially continuous $PM_{2.5}$ distributions for the whole of the study area. These distribution maps are then used as the input into the next prediction processing part, 3.

3.2.3. Part 3: Prediction of High-Resolution $PM_{2.5}$ Distributions

In this last part, 3, that is the core component of the prediction framework, there are 2 steps to predict the high-resolution spatiotemporal $PM_{2.5}$. We continue to use the index in the previous part, 2, to present the eighth and ninth steps. Please note that in this 3rd part, we focus on the estimation of the high-resolution $PM_{2.5}$ in a spatial area by using the 7 variables to input to the estimation models. However, the datasets are extracted from the different dates, which means the time factor in this processing has not been considered. However, in this part, we focus on the spatiotemporal prediction for $PM_{2.5}$, so we need to define the spatial and temporal resolutions. The temporal resolution is a day, because daily AOD data is the highest resolution that we can obtain this data for (see Section 3.1.2), and the spatial resolution is set to be 3300 m in this prediction part for the reason given below.

In the eighth step, we applied two prediction models to the SARIMA dataset (used as a baseline) and ConvLSTM (used in the final prediction part of our framework).

The input of data is 151 $PM_{2.5}$ distribution images with a 3300 m spatial resolution. However, the approaches for inputting data into these two models are different. In terms of the SARIMA model, we regard each cell as a single input, which means every cell has a time series consisting of 151 concentration values. The model is built 2500 times, because there are 2500 cells.

The SARIMA model used in this study is originally proposed by Box and Jenkins, which is the most general form of a univariate class of models [79]. It has been widely studied and used in different fields, such as industry, economy and, more recently, in public health. In this study, this model is utilized for the prediction of a $PM_{2.5}$ time series.

Since the $PM_{2.5}$ distribution changes with time in a specific spatial area, such as within one city or province, we need to determine if they have a trend that increases or decreases values in the series in the dataset. The initial process is that a first-order difference of the series z_t is given by w_t , the difference between points in the series, calculated as:

$$w_t = z_t - z_{t-1}, \quad (3)$$

We may also write w_t in terms of the backward shift operator B as:

$$w_t = (1 - B)z_t, \quad (4)$$

and, hence, the d th order differencing is obtained as $(1 - B)^d z_t$.

Besides this trend, the seasonality needs to be considered as well. Thus, Box and Jenkins have extended the above idea of ordinary differencing by forming seasonal differences:

$$w_t = z_t - z_{t-1} = (1 - B^s)z_t \quad (5)$$

where s is the seasonal period of the data. Consequently, the seasonal autoregressive integrated moving average (SARIMA), which is the most general Box–Jenkins model, has the following form:

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d z_t = \theta(B)\Theta(B^s)a_t, \quad (6)$$

with:

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \\ \Phi(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{sP} \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ} \end{aligned} \quad (7)$$

where p represents the auto-regressive order; q represents the moving average order; d represents the number of differencing operations and P , D and Q represent the corresponding seasonal orders.

After removing the trend and seasonal components, the model-fitting process includes the identification, parameter estimation and diagnostic verification. Based on the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF), a tentative autoregressive moving average (ARMA) process is developed in the recognition phase. We compare the shape of the ACF and PACF of the PM_{2.5} time series with the shape of the theoretical model. In this comparison, we can define p and q .

It is worth mentioning that, in our study, for each validation, we need to build 2500 SARIMA models for 2500 PM_{2.5} time series in each cell; this is computationally intensive and is quite time-consuming. In order to simplify the process, for every distribution that consists of 2500 cells, we compute the average PM_{2.5}. One hundred and fifty-one-day average PM_{2.5} values are regarded as the input of the SARIMA, then, to calculate the best parameters. The best output parameters are defined as the global parameters that would be used for each of the 2500 cells in the next process. Then, we still build 2500 SARIMA models for the 2500 times series in each cell, but we use the same global parameters for all 2500 SARIMA models in each validation. We have 10 validation groups; thus, we used 10 sets of global parameters in total.

Long short-term memory (LSTM) is a kind of recurrent neural network (RNN) node structure, which can deal well with time series data that generally has temporal autocorrelations [33]. This model can successfully learn and generalize the attributes of a time series, such as a road traffic flow [80] and a financial stock options return [81]. The battery state affected by various interconnection gates is the core concept of LSTM. The unit state acts as the transmission highway, and the relevant information is transferred down the sequence chain as the “memory” of the network. A cell state can carry relevant information to the whole process of sequence processing. Therefore, even information from an earlier time step can enter a later time step, thus minimizing the impact of the short-term memory. With the development of the cell state, information is added or removed through the gate, just like a kind of neural network, which decides what information is allowed to exist in the cell state (during training) by learning the relevant information [82]. In a LSTM network, at each time step t , the hidden state h_t is updated by the current data, i.e., at the same time step X_t , the hidden states at the previous time step h_{t-1} , the input gate i_t , the forget gate f_t , the output gate o_t and a memory cell C_t are updated as well [83]. The basic principle of this model is the same as that of a ConvLSTM. Therefore, the equation is not repeated here, which will be described in the introduction of the ConvLSTM model later.

A ConvLSTM model is a variant of a LSTM, which is used to deal with spatial-temporal predictions. It was first proposed by Shi et al., [84]. At first, it was used for real-time precipitation predictions—amongst which, real-time prediction is used as a very short-range prediction technology using the estimated values of velocity and motion direction. In this paper, we follow the formulation of ConvLSTM according to [84], which includes inputs X_1, \dots, X_t ; cell outputs C_1, \dots, C_t ; hidden states h_1, \dots, h_t ; gates i_t, f_t and o_t and uses a three-dimensional (3D) tensor structure. In the three-dimensional

space-time tensor of the input elements of the ConvLSTM network, the first two dimensions are spatial dimensions, and the third dimension is the temporal dimension. Just like the original LSTM model, the transformation from input to state and from state to state involves the convolution of the three-dimensional output tensor. The following equation can be used to further build the model, where $*$ represents convolution operation, and \circ represents Hadamard product.

$$\begin{aligned}
 i_t &= \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + W_{ci} \circ h_{t-1} + b_i), \\
 f_t &= \sigma(W_{xf} * X_t + W_{hf} * h_{t-1} + W_{cf} \circ h_{t-1} + b_f), \\
 o_t &= \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + W_{co} \circ h_{t-1} + b_o), \\
 C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * h_{t-1} + b_c), \\
 h_t &= o_t \circ \tanh(C_t) \\
 \Theta(B^S) &= 1 - \Theta_1 B^S - \dots - \Theta_Q B^{SQ},
 \end{aligned}
 \tag{8}$$

In the above equations, i_t , f_t and o_t are the outputs of the input gate, forget gate and output gate for time step t . C_t is the cell output at time step t . h_t is the hidden state of a cell at time step t . Sigmoid (σ) is used as the gating function for the three gates, since it outputs a value between 0 and 1 corresponding to either how it does not allow an information flow through the gate or it allows a complete information flow through the gate. On the other hand, in order to overcome the problem of a vanishing gradient (which is the problem in using a gradient-based learning method and back propagation to train an artificial neural network), a function (\tanh) is needed. W and b are a weight matrix and deviation vector parameter, respectively, which need to be learned during training the ConvLSTM model. Then, we input the PM_{2.5} grids into the ConvLSTM model and use the validation strategies below.

As the dataset used in this study is a time series, we present a cross-validation strategy, as shown in Figure 5. We validate the prediction models 10 times, which utilizes 10 different data groups that are extracted from the original dataset. Each group consists of $151 - 9 = 142$ days of data, $142 \times 80\% \approx 114$ days that are used to train the models and the next 28 days' data is then used or prepared to be tested.

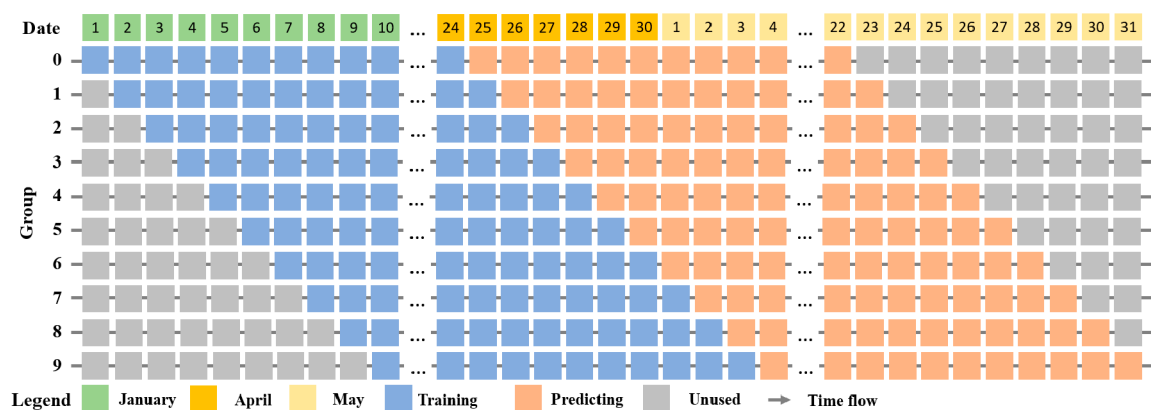


Figure 5. The validation strategy for time series PM_{2.5} distributions. Every square block represents one day's PM_{2.5} data. The training and prediction days' data have been differentiated using blue and orange colors. Some days, data has not been displayed, as it makes the figure too wide to be displayed and are represented by symbol ellipses ("..."), such as February and March.

Ninth step—the final step.

After training and predicting the two models 10 times, we assess the results by calculating the root mean square error (RMSE) and the R² (coefficient of determination) regression score function [85]

of the tested data and the predicted data. The RMSE represents the absolute fit between the model and the data and how close the observed data points are to the predicted values of the model, while R^2 reflects the relative fit result. We compute the results of these two indicators as the average results for each cell at one-time intervals (one hour). Nevertheless, for the purpose of assessments in different scales to compare the two models, we assess and present the two results in time and space, respectively, by only using RMSE to reflect the absolute value, which makes them easier to be distinguished.

In the temporal scale, for every fishnet (grid) that consists of 2500 cells, there are three features defined for each cell: the original $PM_{2.5}$ value (F1), predicted $PM_{2.5}$ value by SARIMA (F2) and the predicted $PM_{2.5}$ value using ConvLSTM (F3). We calculate the RMSE between the 2500 F1 and 2500 F2; the result is defined as R1. Then, we calculate the RMSE between the 2500 F1 and 2500 F3. The result is defined as R2. For every validation group, there are 28 days to be tested, and there are 28 R1 and 28 R2 values that can be arranged by time flow. While we build models in 10 validations, there are 10 group results. In every group, there are 28 R1s and 28 R2s. R1 and R2 reflect the accuracy of the two predicted results that can be assessed and compared between the two prediction models in time.

According to the discussion above, there are three features for each cell in every different fishnet. In terms of the same cell in different fishnets, we extracted all 28 F1, 28 F2 and 28 F3 values. Then, we calculated the RMSE between the 28 F1 and 28 F2 values, resulting in R3, and calculated the RMSE between the 28 F1 and the 28 F3 values as result R4. Finally, there were 2500 R3s and 2500 R4s. Since we validated the models 10 times, there were 10 groups of results. R3 and R4, which reflect the accuracy of the two predicted results, can be assessed and compared between the two prediction models in space.

4. Results

This section presents the results of applying our $PM_{2.5}$ prediction framework in Shijiazhuang City. First, it shows some examples of the preprocessing, the data used, including the distributions of the input variables for the estimation models, and the BSMP processing based on the remote-sensing dataset. Then, the results are presented and discussed. The estimation methods are compared. Then, the most accurate method is used to build a model to estimate the spatiotemporal high-resolution $PM_{2.5}$ over the whole study area and period. Finally, we show the prediction results using the SARIMA and ConvLSTM models with a cross-validation, and then, finally, we report the comparison of accuracy between these two models.

4.1. Data Reprocessing

4.1.1. Data Preparation for Estimation Models

In this part, we introduce an example of the preprocessing results for the used datasets in the estimation model. Six distributions, including albedo, precipitation, relative humidity, wind speed, elevation and NDVI, are illustrated in Figure 6, where albedo, precipitation, relative humidity and wind speed distributions are from 1 January 2019, and the NDVI distribution is from December 2018. For the six kinds of distributions, the precipitation, relative humidity and wind speed distributions are generated by an IDW method based on the data from meteorology stations (points). Albedo, precipitation, relative humidity and wind speed distributions have 151 maps over the study period, while the elevation and NDVI data are from one map.

These datasets have two functions: First, in step 5 of part 2, we extract the values from the air quality sites, then use them to train an estimation model. In this step, the value sets of albedo, precipitation, relative humidity and wind speed used are from the 151 maps, but the other two are from the individual maps. Second, after getting a complete estimation model, we can use the datasets by combining the AOD distributions to get the $PM_{2.5}$ distribution that covers the whole study area.

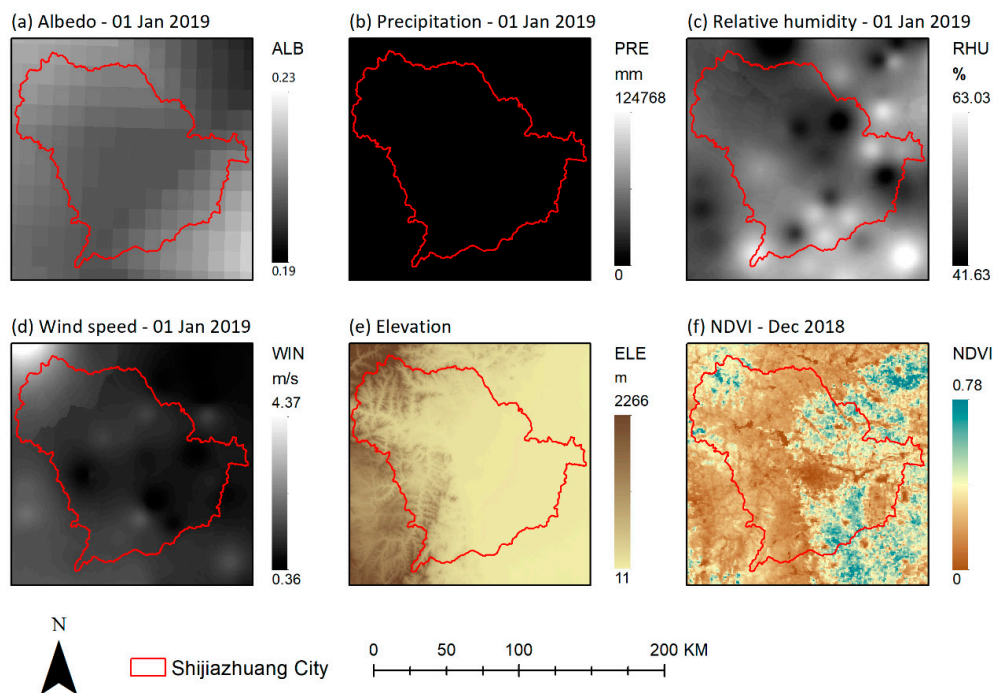


Figure 6. Examples of the distributions of the 6 factors that contribute to the PM_{2.5}-AOD relationship. (a–d) show cases of the albedo, precipitation, relative humidity and wind speed distributions in Jan 2019; (e) shows the elevation distribution of the study area, while (f) illustrates the NDVI that in Dec 2018. In (b), because there is no precipitation in that day, the black cells, which mean the 0 values, cover the whole study area.

4.1.2. BSMP Processing

We show the processing of the BSMP method that creates the spatial distributions for all 151 days’ AOD in Figure 7. In this process, different AOD distributions need different numbers of BSMP rounds depending on the number of cells that are missing values. We show an example of one of the total 151 AOD distributions in Figure 7 (1 January 2019), which needs five rounds of BSMP. The settings of the window sizes in the five rounds are 5 × 5, 10 × 10, 15 × 15, 20 × 20 and 25 × 25.

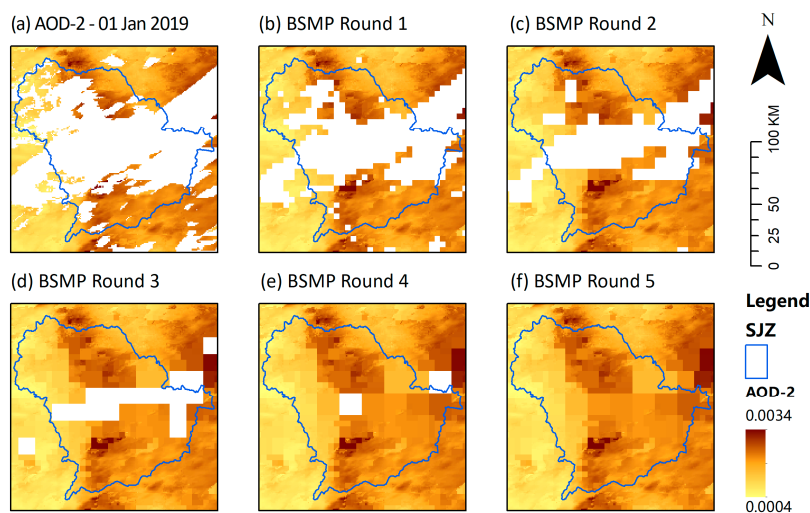


Figure 7. An example of the processing of the BSMP method using an AOD-2 distribution from the 1 January 2019. (a) shows the AOD-2, and (b–f) illustrate the 5 rounds of BSMP to get the final used AOD-2 distribution. “SJZ” is the abbreviation of Shijiazhuang City; “AOD-2” is not the original AOD but the corrected AOD using Equations (1) and (2).

Note that the dataset of the upgraded AOD distributions that have used BSMP 151 times has only one function: After getting a completed estimation model, it is used to get the $PM_{2.5}$ distribution that covers the whole study area by combining the other five key factors' distributions.

4.2. $PM_{2.5}$ Estimation Model

4.2.1. Model Performance and Comparison

Following the workflow of step 5 in part 2, we use five models/methods to estimate the $PM_{2.5}$ based upon six features. Since there are several null values in the AOD datasets for the results because of MODIS satellite orbit spacing issues, cloud coverage problems and the limitations of the inversion algorithms, we extracted 1159 arrays from 2416 data arrays as the input into the five models. Here, we use a five-fold validation strategy to test the models. The specific process used is as follows. We divide the datasets into five subsets. While the first four subsets have 232 arrays, the last subset has 231 arrays (total: 1159). Then, one of the subsets is used as the validation dataset, and the rest are used as the training dataset. The training is repeated 10 times, until all the subsets are used as the validation dataset at least once.

On the basis of comparison between the verification and training data, the root mean square error (CV-RMSE) and the coefficient of determination ($CV-R^2$) are used to validate the accuracy of each estimated model. While determining the best model for $PM_{2.5}$ estimation based on accuracy, a variable importance analysis was also performed to evaluate the contribution of each predictor in the $PM_{2.5}$ prediction. This method is based on the F Score (feature score) measure, which simply summarizes the number of times each feature is used in the decision trees.

Parameters of the machine-learning models were optimized with an auto-process, which means we set a range of every parameter for each model and then processed the cross-validation; the parameter set with the most accurate estimation results would be regarded as the determined parameters. Here, we report the determined parameters of the models below:

- Ridge: alpha = 0.001 (alpha is the regularization strength).
- LASSO: alpha = 0.001 (alpha is the constant that multiplies the L1 term).
- Cubist: committees = 1000.
- XGBoost: max_depth = 8, subsample = 0.8, colsample_bytree = 0.8, eta = 0.3 and num_boost_round = 1000. (The max_depth is the maximum depth of a tree, subsample is the subsample ratio of the training instances, colsample_bytree is the subsample ratio of columns when constructing each tree and num_boost_round is the number of boosting iterations.).

Based on the results of the optimized models, the CV-RMSE ranged from 32.86 $\mu\text{g}/\text{m}^3$ to 52.23 $\mu\text{g}/\text{m}^3$, and the $CV-R^2$ ranged from 0.17 to 0.71 (Table 1). Among all, XGBoost had the best performance, while cubist is the model with the worst performance determined by CV-RMSE. The results are different from a similar study [35], where the cubist model gave the best performance. The reason why this happened is that the performance of a regression or machine-learning method is based on different geographic and environmental situations. The results in [35] could not be used in our case, which demonstrated that this comparison is necessary. With the optimal parameters, the CV-RMSE and $CV-R^2$ of XGBoost were 32.86 $\mu\text{g}/\text{m}^3$ and 0.71. We chose XGBoost to build an estimation model in the last part of this paper.

Table 1. The cross-validation results of predictions from the 5 estimation methods. CV means the cross validation RMSE means the root mean square error and R^2 means the determination regression.

Model	CV-RMSE ($\mu\text{g}/\text{m}^3$)	CV- R^2
Linear	46.69	0.41
Ridge	49.67	0.33
LASSO	46.71	0.41
Cubist	52.23	0.17
XGBoost	32.86	0.71

4.2.2. XGBoost Estimation

A positive but medium correlation was observed based on the evaluation of an empirical relationship between the observed $PM_{2.5}$ and satellite-derived AOD (Figure 8a), with a correlation coefficient (R) of 0.58 ($R^2 = 0.34$), while the p -value < 0.01 , which provides the evidence to use the AOD to estimate the $PM_{2.5}$. For the best model, the predicted and observed values were well-aligned with the line of the best fit (Figure 8b), indicating the high accuracy of the $PM_{2.5}$ estimation using XGBoost. The RMSE of the final estimation is $13.31 \mu\text{g}/\text{m}^3$, while the R^2 is as high as 0.96, which means XGBoost is truly a good model with excellent performance for this use case.

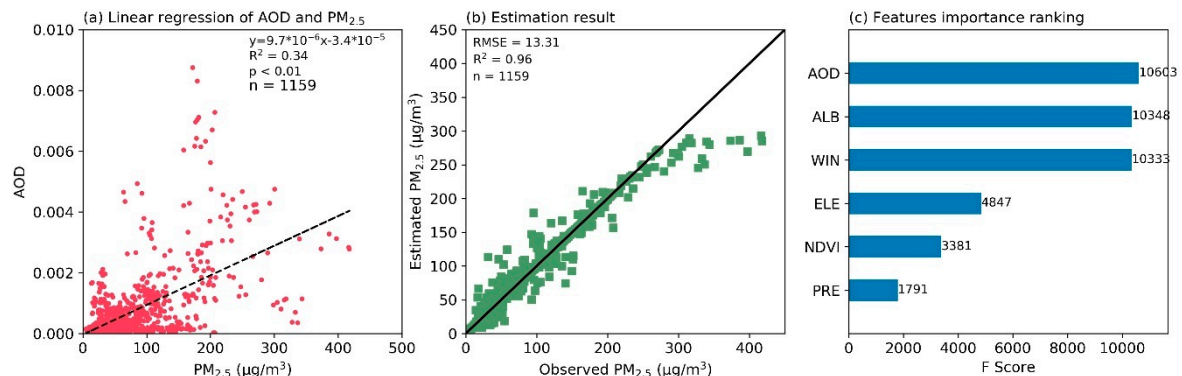


Figure 8. The relationship of the AOD and $PM_{2.5}$ and the regression results of XGBoost: (a) is the linear regression result of the AOD and $PM_{2.5}$ where red points are the fitting coordinate points, (b) shows the prediction and comparison results of the observed and estimated $PM_{2.5}$ illustrated by green points and (c) is the importance analysis result that shows the feature importance ranking for the trained XGBoost model with blue bars. ALB means albedo.

We notice that, in Figure 8b, there is a significant underestimation of values larger than 300. However, there are only 15 to 20 values with a significant underestimation out of a total of 1159 observations. Less than 2% of points could be regarded as outliers that might be caused by accidents, i.e., inputting corrupted data like all input independent variables are 0 or negative numbers. The settings of the XGBoost parameters may result in this situation as well. This needs further testing.

Based on the variable importance analysis, the predictors with the highest contributions to the XGBoost model were the daily AOD and albedo (Figure 8c). Compared with study [35], the results are similar, but the feature elevation is at the second place in the ranking, while in this study, the AOD and albedo are the top two features based on the F score, which is defined as the sum number of times a feature have been used in the decision trees of the model. The higher the score is, the more important the feature is. The third to sixth influencing features in this study are wind speed, elevation, NDVI and precipitation. The influence from the NDVI on the estimation in both studies is not high. The first three features have the greatest impact on the estimation, which accounts for up to 76%, but they have similar values.

Next, we use spatially continuous AOD distributions with high resolution that are computed by a BSMP method to train a XGBoost model to estimate the spatiotemporal high-resolution $PM_{2.5}$ distributions of the whole of Shijiazhuang City over the 151 days. Some estimated $PM_{2.5}$ map examples are shown as Figure 9.

From the five subfigures, we can see in the example in March 2019, the $PM_{2.5}$ was distributed with a higher level compared to the other four examples. The average $PM_{2.5}$ value was higher than $115 \mu\text{g}/\text{m}^3$, which belongs to moderate pollution based on the air quality standards of China (<http://english.mee.gov.cn/>), and some of the regions have been covered by heavy pollution ($150\text{--}250 \mu\text{g}/\text{m}^3$) and even serious pollution ($250\text{--}500 \mu\text{g}/\text{m}^3$). For the example on the 1st of January 2019, the $PM_{2.5}$ was still at a higher level and had an average value was about $80 \mu\text{g}/\text{m}^3$. However, in another three cases, the $PM_{2.5}$ was at low levels at minor pollution ($75\text{--}115 \mu\text{g}/\text{m}^3$) or good air quality ($35\text{--}75 \mu\text{g}/\text{m}^3$).

4.3.2. ConvLSTM

For each round of the prediction, the parameters of the model included the kernel size, which we proposed to be 3×3 . We use 40 convolutional filters that can extract important features from the convolution layers, with five units for each. In order to improve the generalization ability and to prevent overfitting (which is the production of an analysis that corresponds too closely or exactly to a particular set of data and may, therefore, fail to fit additional data or predict future observations reliably in machine-learning or deep-learning models), the recurrent weight dropout was set to 0.2 in the model; the number of training times (epochs) was set as 500, whilst the Adam optimizer [86] was used with a learning rate of 0.001 and a decay rate of 0.9. The result is shown as Figure 10.

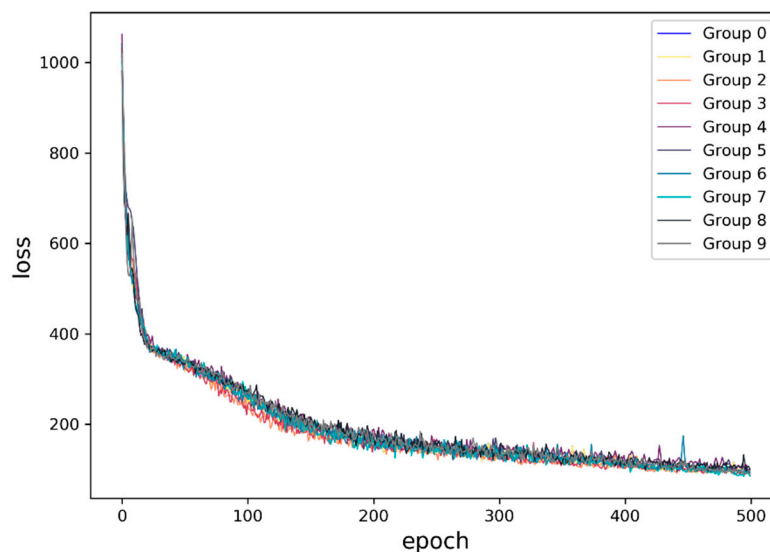


Figure 10. Predicted $PM_{2.5}$ loss versus epoch number. Legend Group 0 to 9 shows the loss changing situations for 1 to 10 groups.

This figure shows the training epoch of the results and the mean absolute error (MAE), which is called a loss function in machine learning or deep learning. It is noted that, when the epoch was less than 30, the loss dropped rapidly, but when the epoch was greater than 30 and less than 200, the average value of the loss of the 10 groups started to fall slowly. After 200 epochs, the average loss decreased very slowly and was relatively stable around 100. In the next section, we analyze the accuracy of the prediction results of the two models.

4.4. Accuracy Analysis

The overall accuracy of the prediction results can be reflected by the RMSE of the tested and predicted values. The total average RMSEs of the two prediction results for the 2500 cells in the 10 validations are reported in Table 3. It is noted that the accuracy of the prediction results of the ConvLSTM model is higher than SARIMA, where the total average of the ConvLSTM model's RMSE is 14.94, while the SARIMA-predicted RMSE is 17.41. Since the structure of the overall process of this study is multilayered both in the spatial and temporal scales, in order to compare the two prediction results clearly, we compute the RMSE results by the method in Section 3.2.3. Then, we assess and discuss the results separately in time and space in the next section.

Table 3. The average root mean square error (RMSE) of the prediction results for all cells in the 10 validation groups.

Model	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	Mean
C	14.78	15.00	15.60	15.27	14.85	15.19	14.88	14.45	15.09	14.30	14.94
S	17.15	22.33	16.14	16.67	21.79	20.53	14.39	15.68	14.10	15.29	17.41

Note: S means the SARIMA model, and C means the convolutional long short-term memory (ConvLSTM) model.

4.4.1. Comparison in Time

The tested ConvLSTM-predicted and SARIMA-predicted $PM_{2.5}$ over the 28 days in the 10 different groups are shown in Figure 11. We can see that the original concentrations change with time in every group, which are the average values of the 2500 cell concentrations. The fitting degree of the other two lines with the original tested line reflects the accuracy of the predictions of the two models.

We notice that, in Figure 11, the $PM_{2.5}$ values are low ($<40 \mu g/m^3$) in all the groups, whereas, in Figure 8, the value ranges have a much higher maximum. The reason why this happens is as follows: Figures 8 and 11 illustrate two totally different processing results, while Figure 8 shows the estimation results and Figure 11 shows the prediction results. Figure 8a shows the relationship of the AOD and $PM_{2.5}$ in the estimation part. There are 1159 sets of the data array that are from 16 station points in 151 days. This means that, in each day's distribution for every factor, there are only 16 real measured $PM_{2.5}$ -value air quality site points. However, Figure 11 shows the tested ConvLSTM-predicted, and the SARIMA-predicted, $PM_{2.5}$ —e.g., (a) illustrates the first group's comparison results. Every point in the line graph ($<40 \mu g/m^3$) is the average value of one distribution that has been divided into a grid format consisting of $50 \times 50 = 2500$ cells.

To summarize, the reason why the values in Figure 11 are so low while the other ones are higher is as follows: First, they are from different processes; one is from the estimation process, and the other is from the prediction process. Second, the lower ones are the average values for a whole region, while the higher ones are from single points. The average ones always reflect the overall situation, while the single ones just show the specific cases. For example, if there is only one value of 11, and the other nine values are 1, the average of the 10 values is two, which is much lower than 11. Third, the lower ones are the estimated results using XGBoost, which have been validated in Section 4.2, while the higher ones are the real measured $PM_{2.5}$ values.

However, it is difficult to see the differences between the prediction results of the two models in Figure 7. Thus, we calculate the RMSE between the 2500 F1s and 2500 F2s to get R1 and, then, calculate the RMSE between the 2500 F1s and 2500 F3s to get the R2. The 28 R1s and 28 R2s over every 28 days in 10 validations is shown in Figure 12.

In Figure 12, we see that the RMSE of the predicted results by ConvLSTM is higher overall than the ones by SARIMA over the 28 days in all 10 validations, which means the accuracy of the prediction of ConvLSTM is higher overall than SARIMA in a temporal scale. Among them, especially G0, G1, G4 and G5 show a better performance for the prediction. For most days, the RMSEs of the ConvLSTM model are lower than SARIMA, keeping to around 10, especially for the 10th to 12th of May 2019. The SARIMA model's prediction accuracy is lower than the average level of itself. However, in some periods, it is higher. For example, on 23 May 2019, the RMSE of the SARIMA was lower than the ConvLSTM one in groups 6, 7 and 8. However, the overall results demonstrate the higher prediction capability of ConvLSTM compared to SARIMA. The less fluctuating curve shows that ConvLSTM is more stable than SARIMA to predict in time.

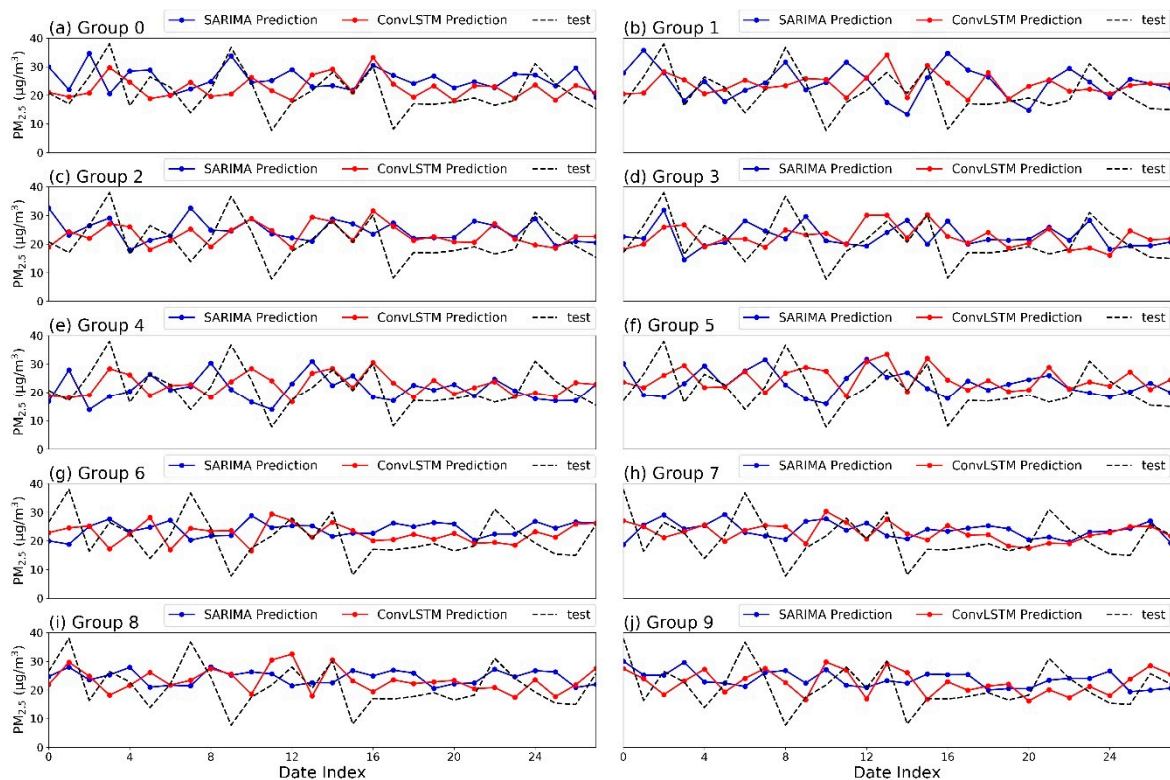


Figure 11. The tested ConvLSTM-prediction, and the SARIMA-prediction, for $PM_{2.5}$. (a–j) illustrate the 10 group comparison results, respectively. “Date index” is defined as the 28 predicted days in each group based on Figure 5, in different subplots; the index represents different ranges of dates, i.e., in Group 0, index 0 to 27 represents the dates 25 April to 22 May, while it represents the dates 26 April to 23 May in Group 1.

4.4.2. Comparison in Space

In the spatial scale, we compute the 168 F1, 168 F2 and 168 F3. Then, we calculate the RMSE between the 28 F1 and 2 F2 to get R3 and calculate the RMSE between 28 F1 and 28 F3 to reach R4. There are 2500 R3 and 2500 R4 in 10 validations. We plot the frequency distribution histogram of R3 and R4 for the 10 groups, as shown in Figure 13, to compare the prediction accuracy of the two models in space. In terms of the SARIMA results, the RMSE ranges from 0 to 60 for all groups, while the ConvLSTM ones range from 0 to 30. For all groups of the SARIMA model, the RMSE occupies the most range, from 0 to 10, and the frequency is about 1150, followed by RMSE ranging from 10 to 20, and the average frequency is about 1100. In terms of the ConvLSTM, the RMSE occupies the most places, from 10 to 20, and the frequency is about 1250, followed by the RMSE ranging from 0 to 10, and the average frequency is about 1150. Although the SARIMA has less RMSEs that are lower than a ConvLSTM, it also has a greater number of higher RMSEs that range from 20 to 60, accounting for 10% $((1100 + 1150) \div 2500 = 0.9 = 90\%)$, while the ConvLSTM has less than 5% of RMSEs distributed in this range $((1150 + 1250) \div 2500 = 0.96 > 95\%)$.

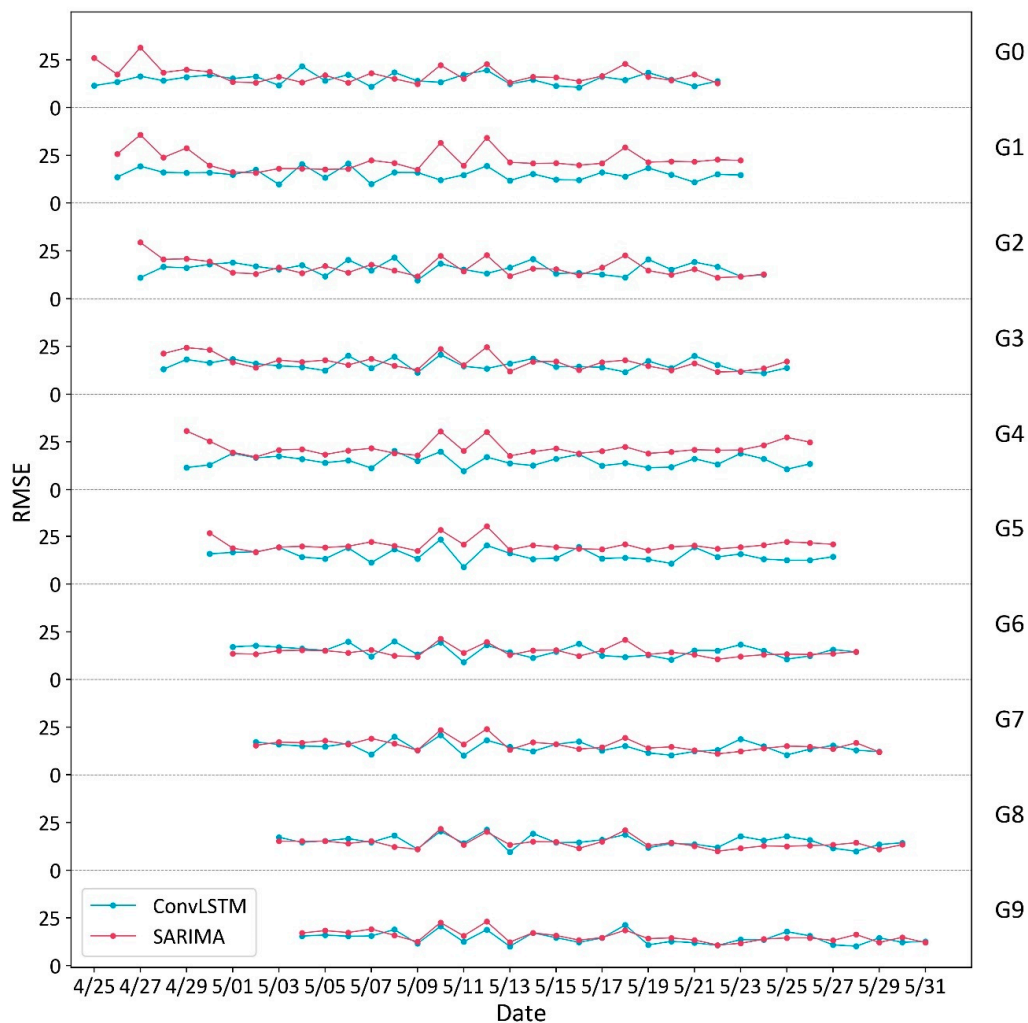


Figure 12. The root mean square errors (RMSEs) of the SARIMA and ConvLSTM models in time. All groups' data share the same x-axis, while, in the y-axis, there are 10 bins, such that every bin spans the RMSE from 0 to 50, where 50 could also be the start of the next bin.

Then, we output the RMSEs as the maps of Shijiazhuang, where the different colors in the cell represent the RMSE values of the prediction. The results are illustrated in Figure 14. The whiter areas in the figure correspond to a higher error (RMSE) of the user density prediction. On the contrary, the darker area indicates a lower error. Consequently, the distribution of errors can indicate the role of the prediction model in space.

Intuitively, in all the prediction RMSE maps of SARIMA, there are much greater numbers of the whiter cells that represent the highest errors of the prediction that are surrounded by the cells with a much blacker color. Intuitively, the overall RMSE distribution results of SARIMA have more whiter cells, which represent a higher error of prediction, while the distributions are uneven, especially in G1, G4 and G5. However, we use ConvLSTM to predict the $PM_{2.5}$ distribution, because the convolution is used when we train the models, which considers the surrounding cells' values, which is different from SARIMA, where the maps of the RMSEs have more blacker cells. The ConvLSTM model considers the spatial autocorrelation that it is more suitable for the prediction of a $PM_{2.5}$ distribution, which proves that the ConvLSTM model is much more accurate than the SARIMA model in space.

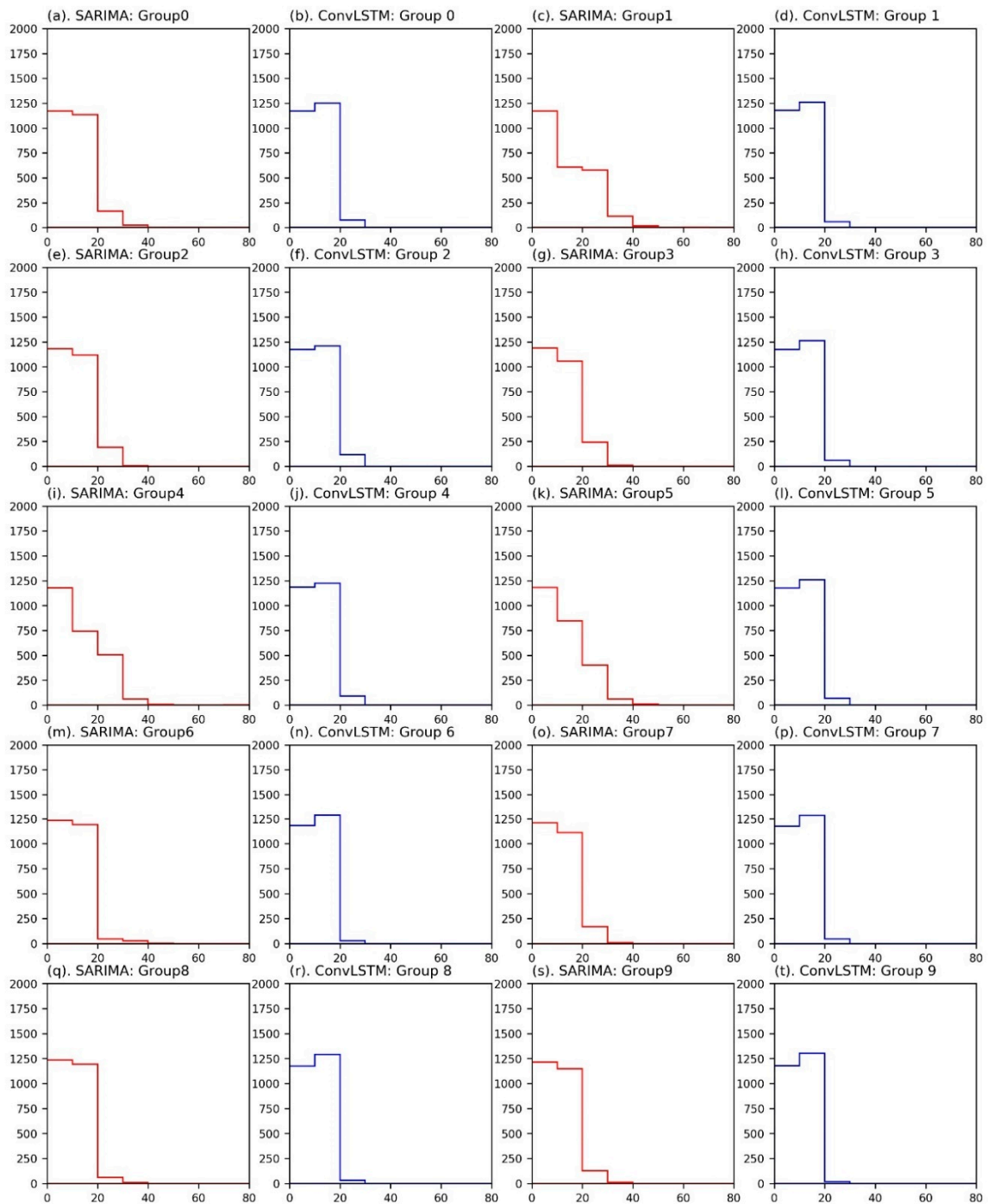


Figure 13. Frequency distribution histogram of the RMSEs of the SARIMA and ConvLSTM models in space. The x-axis of all subplots shows the RMSE values, while the y-axis represents the frequency. (a–t) alternately represent the RMSE frequency distribution of SARIMA (red bars) and ConvLSTM (blue bars).

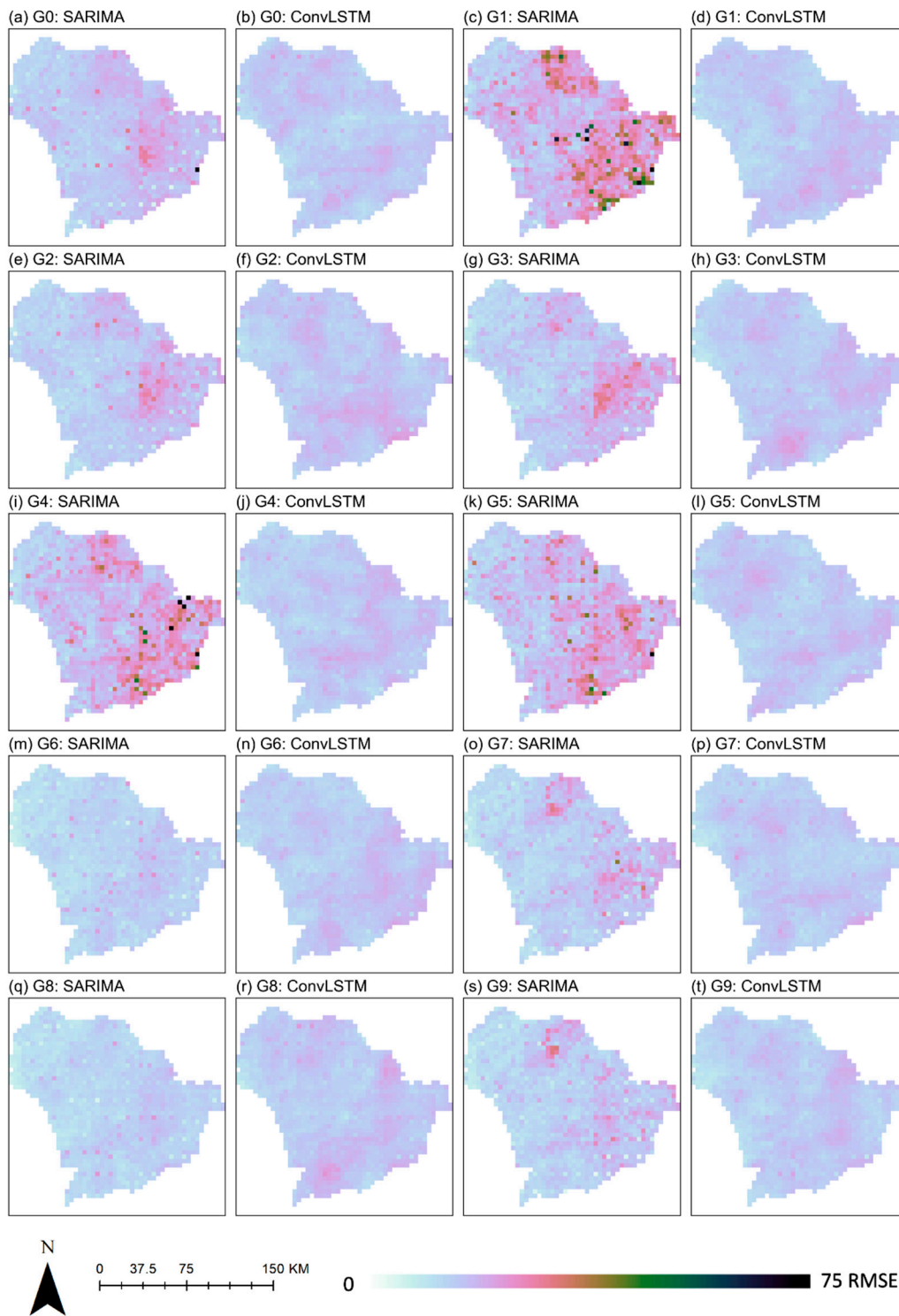


Figure 14. The spatial distribution of the RMSEs of the SARIMA and ConvLSTM models. (a–t) alternately show the RMSE maps of the SARIMA and ConvLSTM in Shijiazhuang. The color bar from left to right are from white, blue to red, green, and final to dark. The more left the color is, the larger the value of RMSE is.

5. Discussion

The above results suggest that our framework is able to be used for high-resolution PM_{2.5} distribution predictions based on the input multi-resource databases. To better understand the ability of our framework, this section examines the methods from two perspectives: first, from the characteristics of the prediction framework, and second, from the contributions of the framework as a value added to existing theories or methods for PM_{2.5} predictions.

5.1. The Characteristics of the Prediction Framework

One of the important characteristics of this framework is that it selects the most suitable prediction method by comparing the estimation or prediction effect of different existing models in a specific scene, so the prediction capability of this framework is relatively objective, with a high precision compared with other related studies that use more limited methods. The second characteristic is that this framework does not need to use any other auxiliary datasets for the time prediction part, because the prediction mechanism of the ConvLSTM model in this study is based upon the internal mechanism of time autocorrelation and space autocorrelation through training the historical spatial-temporal PM_{2.5} distribution. The third characteristic is that the framework starts from a multisource original dataset, and after data fusion, the estimation or automatic selection of parameters of the prediction model and other processes, the final prediction products that can be directly used in other studies are obtained. In this process, the framework considers and solves most of the possible problems: For example, the PM_{2.5} value has only air quality sites' data. Thus, we use AOD as the basis and other meteorological factors as the auxiliary to input the XGBoost model to establish an estimation model to obtain spatial continuous PM_{2.5} distributions to solve. In terms of the AOD spatial missing value problem, we use the BSMP method to make up for or fill in the missing value of the AOD, so that the AOD can cover the whole study area and can then be converted into PM_{2.5} spatial distributions.

5.2. Contributions

This study is of great significance, not only from the perspective of GI Science theory but, also, from the perspective of applying efficient and high-precision prediction PM_{2.5} time-space distributions to practical problems. The main novel contributions of this paper are as follows.

- (1) We present a framework that not only estimates the spatially continuous PM_{2.5} distributions but, also, predicts the distributions in the future at high-resolution resolutions in time and space.
- (2) In the spatial estimation process, we compare some popular regression models and machine-learning methods to select the most accurate model to be used in the framework.
- (3) In the prediction process, we use a ConvLSTM model, as a proven accurate deep-learning model, for the reason that we need to consider the spatial autocorrelation. This is a key process used for prediction related to a continuous region, compared with a traditional time series prediction model and seasonal autoregressive integrated moving average (SARIMA), and is used to test the accuracy and stability of the ConvLSTM.

At the current time, high-precision and high-resolution PM_{2.5} spatiotemporal predictions are becoming more and more important for environmental protection and public health and safety decision-making. At the same time, the availability of multisource heterogeneous PM_{2.5} data and the inconsistent estimation and prediction of PM_{2.5} indicates that a unified framework for use by researchers is lacking. Hence, a complete, unified, efficient and high-precision prediction framework is extremely important. Thus, the framework of this paper makes it possible to predict the concentration of PM_{2.5} based on the original multisource data.

The prediction framework in this paper differs from previous single-prediction research, because ours is based on the raw data from the data-collection machines, such as satellite sensors. After comparing the spatial PM_{2.5} estimation by the XGBoost model, the spatial continuous distribution is obtained,

and finally, the spatial-temporal distribution of $PM_{2.5}$ is predicted using a ConvLSTM model. To the best of our knowledge, there does not yet exist a complete prediction framework at such a high spatial resolution combined with such a high temporal resolution. We applied a ConvLSTM to predict a $PM_{2.5}$ spatial-temporal distribution for the first time; the results also show the utility of our method for this application. Therefore, this study opens up a new perspective for the spatiotemporal predictions of $PM_{2.5}$, which can predict both spatial and temporal distributions of $PM_{2.5}$ at a high resolution.

6. Conclusions

This study presents a framework to predict the daily $PM_{2.5}$ distributions using a case study of Shijiazhuang City, China, based on the daily aerosol optical depth (AOD) datasets and other supplementary raw data. The framework consists of three main parts and nine steps within this. In part 1, the framework obtains high-resolution spatiotemporal AOD distributions, then builds a machine-learning model to estimate the spatial distributions of the $PM_{2.5}$ in part 2. Finally, based on training high-resolution spatiotemporal $PM_{2.5}$ distributions, this framework uses a ConvLSTM model to predict the testing distributions with 3300×3300 m-squared spatial resolution, compared with a SARIMA model, to test the accuracy.

In part 2, we compared some popular regression and machine-learning models, including linear, ridge, LASSO regressions, cubist and XGBoost, to build a relationship among the $PM_{2.5}$ monitoring values: AOD, humidity, precipitation, albedo, NDVI, wind speed and elevation. With the optimal parameters and cross-validation, XGBoost was determined as the estimation model in the framework, with the lowest RMSE $32.86 \mu\text{g}/\text{m}^3$ and highest R^2 0.71 compared to the four other models. As there can be many missing values for the AOD in space, we proposed a Block Statistics and Missing-value Padding (BSMP) method to derive more whole spatially continuous distributions for the AOD. Then, in the third part, the ConvLSTM was selected as a deep-learning method to be used for spatiotemporal $PM_{2.5}$ predictions. After 10 validations and a comparison with SARIMA in both time and space, ConvLSTM clearly produces more accurate predictions, with a total average prediction RMSE of $14.94 \mu\text{g}/\text{m}^3$ compared to SARIMA's $17.41 \mu\text{g}/\text{m}^3$. In terms of the prediction in time, ConvLSTM is more stable, with less fluctuations for prediction, while, on a spatial scale, ConvLSTM can better eliminate the error that is caused by spatial differences than SARIMA; thus, it has a higher prediction accuracy.

In the future, we aim to improve the framework. We can apply the framework to study more areas, such as other cities that suffer severe haze, and apply it on a larger scale (region), such as a province or a country, to test the robustness of the framework. In addition, because we cannot get the missing values of the AOD in our datasets, the BSMP needs to be validated. Plus, in the estimation part, we can analyze other factors that might contribute to the AOD- $PM_{2.5}$ association models, then compare and validate them to get the best key factors to input into the estimation models, which could improve the efficiency of the framework. In the prediction part, we could improve the ConvLSTM model through considering multi-input features—for example, the spatiotemporal features of the pollution source of $PM_{2.5}$ could be considered as a key condition to increase the accuracy of the prediction.

We used 3.3 km as the spatial resolution. This represents about a 10% part of a typical Chinese city district size of 8.7 square kilometers (<http://www.demographia.com/db-beijing-ward.htm>). This spatial resolution is driven by the size of the city we studied (Shijiazhuang, China) and the use of a computationally efficient ConvLSTM model for predictions that divided this city area into 50×50 cells, which meant that our model would typically take 10 h to run on a current laptop/personal computer. An even higher spatial resolution could also be investigated, such as 1.85 km or even 0.925 km, but this would require the use of a much faster computer or a much longer computation time. Note also that the temporal resolution used for evaluating the prediction part is daily, as determined by that of the input AOD dataset, although the predictive capability of our ConvLSTM model is much higher hourly. In the future, we could evaluate the higher hourly temporal resolution predictions of the model through comparisons with the air quality data from air quality (monitoring) stations.

Author Contributions: Conceptualization, G.Z. and X.R.; methodology, G.Z. and S.P.; software, H.L. and J.D.; validation, J.D.; formal analysis, G.Z.; investigation, H.L.; resources, R.L. and X.R.; data curation, X.Z.; writing—original draft preparation, G.Z. and S.P.; writing—review and editing, S.P.; visualization, S.Z.; supervision, S.P. and X.Z.; project administration, X.R. and R.L.; funding acquisition, X.R. and R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the National Key Research and Development Program of China (Grant No. 2017YFB0503605), the National Natural Science Foundation of China (Grant No. 41771478), the Fundamental Research Funds for the Central Universities (Grant No. 2019B02514), the Beijing Natural Science Foundation (Grant No. 8172046), the China Scholarship Council (CSC) and Queen Mary University of London, National Natural Science Foundation of China (Grant No. 41771435).

Acknowledgments: The first author is grateful to the support from the Weipeng Xing and Guangxia Yu.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fan, Y.; Rui, X.; Poslad, S.; Zhang, G.; Yu, T.; Xu, X.; Song, X. A better way to monitor haze through image based upon the adjusted LeNet-5 CNN model. *Signal Image Video Process.* **2019**, *14*, 455–463. [[CrossRef](#)]
2. Tao, M.; Chen, L.; Wang, Z.; Ma, P.; Tao, J.; Jia, S. A study of urban pollution and haze clouds over northern China during the dusty season based on satellite and surface observations. *Atmos. Environ.* **2014**, *82*, 183–192. [[CrossRef](#)]
3. Ziomas, I.C.; Melas, D.; Zerefos, C.S.; Bais, A.F.; Paliatatos, A.G. Forecasting peak pollutant levels from meteorological variables. *Atmos. Environ.* **1995**, *29*, 3703–3711. [[CrossRef](#)]
4. Chaloulakou, A.; Kassomenos, P.; Spyrellis, N.; Demokritou, P.; Koutrakis, P. Measurements of PM₁₀ and PM_{2.5} particle concentrations in Athens, Greece. *Atmos. Environ.* **2003**, *37*, 649–660. [[CrossRef](#)]
5. Hussein, T.; Karppinen, A.; Kukkonen, J.; Härkönen, J.; Aalto, P.P.; Hämeri, K.; Kerminen, V.-M.; Kulmala, M. Meteorological dependence of size-fractionated number concentrations of urban aerosol particles. *Atmos. Environ.* **2006**, *40*, 1427–1440. [[CrossRef](#)]
6. Barlow, P.G.; Brown, D.M.; Donaldson, K.; Maccallum, J.; Stone, V. Reduced alveolar macrophage migration induced by acute ambient particle (PM₁₀) exposure. *Cell Boil. Toxicol.* **2007**, *24*, 243–252. [[CrossRef](#)] [[PubMed](#)]
7. Dockery, D.W. Health Effects of Particulate Air Pollution. *Ann. Epidemiol.* **2009**, *19*, 257–263. [[CrossRef](#)]
8. Zhang, Q.; Quan, J.; Tie, X.; Li, X.; Liu, Q.; Gao, Y.; Zhao, D. Effects of meteorology and secondary particle formation on visibility during heavy haze events in Beijing, China. *Sci. Total Environ.* **2015**, *502*, 578–584. [[CrossRef](#)]
9. Ebenstein, A.; Fan, M.; Greenstone, M.; He, G.; Zhou, M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10384–10389. [[CrossRef](#)]
10. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating Ground-Level PM_{2.5} in China Using Satellite Remote Sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444. [[CrossRef](#)]
11. Hu, J.; Zhang, H.; Chen, S.; Wiedinmyer, C.; Vandenberghe, F.; Ying, Q.; Kleeman, M.J. Predicting Primary PM_{2.5} and PM_{0.1} Trace Composition for Epidemiological Studies in California. *Environ. Sci. Technol.* **2014**, *48*, 4971–4979. [[CrossRef](#)]
12. Chen, F.; Zhang, X.; Zhu, X.; Zhang, H.; Gao, J.; Hopke, P.K. Chemical Characteristics of PM_{2.5} during a 2016 Winter Haze Episode in Shijiazhuang, China. *Aerosol Air Qual. Res.* **2017**, *17*, 368–380. [[CrossRef](#)]
13. Zhang, G.; Rui, X.; Fan, Y. Critical Review of Methods to Estimate PM_{2.5} Concentrations within Specified Research Region. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 368. [[CrossRef](#)]
14. Hwa-Lung, Y.; Chih-Hsin, W. Retrospective prediction of intraurban spatiotemporal distribution of PM_{2.5} in Taipei. *Atmos. Environ.* **2010**, *44*, 3053–3065. [[CrossRef](#)]
15. Zhao, R.; Gu, X.; Xue, B.; Zhang, J.; Ren, W. Short period PM_{2.5} prediction based on multivariate linear regression model. *PLoS ONE* **2018**, *13*, e201011. [[CrossRef](#)]
16. Chen, Y.; Wang, L.; Zhang, L. Research on Application of BP Artificial Neural Network in Prediction of the concentration of PM_{2.5} in Beijing. In Proceedings of the 2015 4th International Conference on Sensors, Measurement and Intelligent Materials, Shenzhen, China, 27–28 December 2015.

17. Ai, H.; Shi, Y. Application of GM (1, 1) model in PM_{2.5} content prediction. In Proceedings of the International Conference on Education, Management and Computing Technology (ICEMCT-16), Hangzhou, China, 9–10 April 2016.
18. Pan, B. Application of XGBoost algorithm in hourly PM_{2.5} concentration prediction. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Harbin, China, 8–10 December 2017; Volume 113, p. 012127.
19. Huang, J.; McQueen, J.; Wilczak, J.; Djalalova, I.; Stajner, I.; Shafran, P.; Allured, D.; Lee, P.; Pan, L.; Tong, D.; et al. Improving NOAA NAQFC PM_{2.5} Predictions with a Bias Correction Approach. *Weather Forecast.* **2017**, *32*, 407–421. [[CrossRef](#)]
20. Song, L.; Pang, S.; Longley, I.; Olivares, G.; Sarrafzadeh, A. Spatio-temporal PM_{2.5} prediction by spatial data aided incremental support vector regression. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 623–630.
21. Zong, R.; Zhang, T.; Chen, Z.; Zhu, Y. Cross-city PM_{2.5} predictions with recurrent neural network. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Seoul, South Korea, 26–29 January 2019.
22. Zhang, G.; Rui, X.; Poslad, S.; Song, X.; Fan, Y.; Wu, B. A Method for the Estimation of Finely-Grained Temporal Spatial Human Population Density Distributions Based on Cell Phone Call Detail Records. *Remote Sens.* **2020**, *12*, 2572. [[CrossRef](#)]
23. Chen, Y. Prediction algorithm of PM_{2.5} mass concentration based on adaptive BP neural network. *Computing* **2018**, *100*, 825–838. [[CrossRef](#)]
24. Sun, W.; Sun, J. Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J. Environ. Manag.* **2017**, *188*, 144–152. [[CrossRef](#)]
25. Bahari, R.A.; Abbaspour, R.A.; Pahlavani, P. Prediction of PM_{2.5} concentrations using temperature inversion effects based on an artificial neural network. In Proceedings of the 1st ISPRS International Conference on Geospatial Information Research, Tehran, Iran, 15–17 November 2014; pp. 73–77.
26. Jiang, P.; Dong, Q.; Li, P. A novel hybrid strategy for PM_{2.5} concentration analysis and prediction. *J. Environ. Manag.* **2017**, *196*, 443–457. [[CrossRef](#)]
27. Lorenz, E.N.; Haman, K. The essence of chaos. *Pure Appl. Geophys.* **1996**, *147*, 598–599.
28. Zheng, Y.; Zhang, Q.; Wang, Z.; Zhu, Y. Application research on PM_{2.5} concentration prediction of multivariate chaotic time series. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Chengdu, China, 7–9 December 2018; Volume 237, p. 022010.
29. Haiming, Z.; Xiaoxiao, S. Study on Prediction of Atmospheric PM_{2.5} Based on RBF Neural Network. In Proceedings of the 2013 Fourth International Conference on Digital Manufacturing & Automation, Qingdao, China, 29–30 June 2013; pp. 1287–1289.
30. Zhang, C.; Wang, X.; Chen, S.; Zou, L.; Tang, C. PM_{2.5} Prediction based on Multifractal Dimension and Artificial Bee Colony Algorithm. In Proceedings of the Journal of Physics: Conference Series, Xi’an, China, 29–31 March 2019; Volume 1237, p. 022085.
31. Liu, Y.; Zheng, H.; Feng, X.; Chen, Z. Short-term traffic flow prediction with Conv-LSTM. In Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 11–13 October 2017; pp. 1–6.
32. Qiao, H.; Wang, T.; Wang, P.; Qiao, S.; Zhang, L. A Time-Distributed Spatiotemporal Feature Learning Method for Machine Health Monitoring with Multi-Sensor Time Series. *Sensors* **2018**, *18*, 2932. [[CrossRef](#)] [[PubMed](#)]
33. Yuan, Z.; Zhou, X.; Yang, T. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London UK, 19–23 August 2018; pp. 984–992.
34. Zhang, G.; Rui, X.; Poslad, S.; Song, X.; Fan, Y.; Ma, Z. Large-Scale, Fine-Grained, Spatial, and Temporal Analysis, and Prediction of Mobile Phone Users’ Distributions Based upon a Convolution Long Short-Term Model. *Sensors* **2019**, *19*, 2156. [[CrossRef](#)] [[PubMed](#)]
35. Xu, Y.; Ho, H.C.; Wong, M.S.; Deng, C.; Shi, Y.; Chan, T.-C.; Knudby, A. Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM_{2.5}. *Environ. Pollut.* **2018**, *242*, 1417–1426. [[CrossRef](#)] [[PubMed](#)]

36. Lasaponara, R. On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecol. Model.* **2006**, *194*, 429–434. [[CrossRef](#)]
37. Jarlan, L.; Mangiarotti, S.; Mougin, E.; Mazzega, P.; Hiernaux, P.; Le Dantec, V. Assimilation of SPOT/VEGETATION NDVI data into a sahelian vegetation dynamics model. *Remote Sens. Environ.* **2008**, *112*, 1381–1394. [[CrossRef](#)]
38. Alizadeh-Choobari, O.; Adibi, P.; Irannejad, P. Impact of the El Niño–Southern Oscillation on the climate of Iran using ERA-Interim data. *Clim. Dyn.* **2017**, *51*, 2897–2911. [[CrossRef](#)]
39. Berrisford, P.; Dee, D.; Fielding, K.; Fuentes, M.; Kallberg, P.; Kobayashi, S.; Uppala, S. *The ERA-Interim Archive*; ERA Report Series; European Centre for Medium-Range Weather Forecasts: Shinfield Park, UK, 2009; pp. 1–16.
40. Rabus, B.; Eineder, M.; Roth, A.; Bamler, R. The shuttle radar topography mission—A new class of digital elevation models acquired by spaceborne radar. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 241–262. [[CrossRef](#)]
41. Farr, T.; Kobrick, M. Shuttle radar topography mission produces a wealth of data. *EOS* **2000**, *81*, 583. [[CrossRef](#)]
42. Zhou, C.H.; Cheng, W.M. Research and compilation of the Geomorphological Atlas of the People’s Republic of China. *Geogr. Res.* **2010**, *29*, 970–979. (In Chinese)
43. Paciorek, C.J.; Liu, Y.; Moreno-Macias, H.; Kondragunta, S. Spatiotemporal Associations between GOES Aerosol Optical Depth Retrievals and Ground-Level PM_{2.5}. *Environ. Sci. Technol.* **2008**, *42*, 5800–5806. [[CrossRef](#)] [[PubMed](#)]
44. Wang, J. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*, 30. [[CrossRef](#)]
45. Koelemeijer, R.; Homan, C.; Matthijsen, J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmos. Environ.* **2006**, *40*, 5304–5315. [[CrossRef](#)]
46. Qu, W.; Wang, J.; Zhang, X.; Sheng, L.; Wang, W. Opposite seasonality of the aerosol optical depth and the surface particulate matter concentration over the north China Plain. *Atmos. Environ.* **2016**, *127*, 90–99. [[CrossRef](#)]
47. Sibson, R. A brief description of natural neighbour interpolation. In *Interpreting Multivariate Data*; John Wiley & Sons: New York, NY, USA, 1981; pp. 21–36.
48. Watson, D. *Contouring: A Guide to the Analysis and Display of Spatial Data*; Pergamon Press: London, UK, 1992; pp. 120–123.
49. Philip, G.M.; Watson, D.F. A precise method for determining contoured surfaces. *APPEA J.* **1982**, *22*, 205–212. [[CrossRef](#)]
50. Watson, D.F.; Philip, G. A refinement of inverse distance weighted interpolation. *Geo-Processing* **1985**, *2*, 315–327.
51. Tian, J.; Chen, N. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM_{2.5}) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens. Environ.* **2010**, *114*, 221–229. [[CrossRef](#)]
52. Chen, Z.-Y.; Zhang, T.; Zhang, R.; Zhu, Z.; Ou, C.-Q.; Guo, Y. Estimating PM_{2.5} concentrations based on non-linear exposure-lag-response associations with aerosol optical depth and meteorological measures. *Atmos. Environ.* **2018**, *173*, 30–37. [[CrossRef](#)]
53. Liu, Y.; Paciorek, C.J.; Koutrakis, P. Estimating Regional Spatial and Temporal Variability of PM_{2.5} Concentrations Using Satellite Data, Meteorology, and Land Use Information. *Environ. Heal. Perspect.* **2009**, *117*, 886–892. [[CrossRef](#)]
54. Sorek-Hamer, M.; Strawa, A.; Chatfield, R.; Esswein, R.; Cohen, A.; Broday, D. Improved retrieval of PM_{2.5} from satellite data products using non-linear methods. *Environ. Pollut.* **2013**, *182*, 417–423. [[CrossRef](#)]
55. Lary, D.J.; Lary, T.; Sattler, B. Using Machine Learning to Estimate Global PM_{2.5} for Environmental Health Studies. *Environ. Health Insights* **2015**, *9* (Suppl. 1), 41–52. [[CrossRef](#)]
56. Song, W.; Jia, H.; Huang, J.; Zhang, Y. A satellite-based geographically weighted regression model for regional PM_{2.5} estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* **2014**, *154*, 1–7. [[CrossRef](#)]
57. Van Donkelaar, A.; Martin, R.V.; Park, R.J. Estimating ground-level PM_{2.5} using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res. Space Phys.* **2006**, *111*, 111. [[CrossRef](#)]

58. Harrison, R.M.; Deacon, A.R.; Jones, M.R.; Appleby, R.S. Sources and processes affecting concentrations of PM₁₀ and PM_{2.5} particulate matter in Birmingham (U.K.). *Atmos. Environ.* **1997**, *31*, 4103–4117. [[CrossRef](#)]
59. Adams, H.; Nieuwenhuijsen, M.J.; Colvile, R. Determinants of fine particle (PM_{2.5}) personal exposure levels in transport microenvironments, London, UK. *Atmos. Environ.* **2001**, *35*, 4557–4566. [[CrossRef](#)]
60. Wang, Y.; Dong, Y.-P.; Feng, J.; Guan, J.-J.; Zhao, W.; Li, H.-J. Characteristics and influencing factors of carbonaceous aerosols in PM_{2.5} in Shanghai, China. *Huan Jing Ke Xue Huanjing Kexue* **2010**, *31*, 1755–1761.
61. Zalakeviciute, R.; López-Villada, J.; Rybarczyk, Y.P. Contrasted Effects of Relative Humidity and Precipitation on Urban PM_{2.5} Pollution in High Elevation Urban Areas. *Sustainability* **2018**, *10*, 2064. [[CrossRef](#)]
62. Seinfeld, J.H.; Pandis, S.N.; Noone, K. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. *Phys. Today* **1998**, *51*, 88. [[CrossRef](#)]
63. Feng, X.; Wang, S. Influence of different weather events on concentrations of particulate matter with different sizes in Lanzhou, China. *J. Environ. Sci.* **2012**, *24*, 665–674. [[CrossRef](#)]
64. Dong, Z.; Yu, X.; Li, X.; Dai, J. Analysis of variation trends and causes of aerosol optical depth in Shaanxi Province using MODIS data. *Chin. Sci. Bull.* **2013**, *58*, 4486–4496. [[CrossRef](#)]
65. Ma, L.; Gao, Y.; Fu, T.; Cheng, L.; Chen, Z.; Li, M. Estimation of Ground PM_{2.5} Concentrations using a DEM-assisted Information Diffusion Algorithm: A Case Study in China. *Sci. Rep.* **2017**, *7*, 15556. [[CrossRef](#)] [[PubMed](#)]
66. McDonald, J.D.; Reed, M.D.; Campen, M.J.; Barrett, E.G.; Seagrave, J.; Mauderly, J.L. Health Effects of Inhaled Gasoline Engine Emissions. *Inhal. Toxicol.* **2007**, *19*, 107–116. [[CrossRef](#)] [[PubMed](#)]
67. Wu, C.-D.; Chen, Y.-C.; Pan, W.-C.; Zeng, Y.-T.; Chen, M.-J.; Guo, Y.-L.L.; Lung, S.-C. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM_{2.5} spatial-temporal variability. *Environ. Pollut.* **2017**, *224*, 148–157. [[CrossRef](#)] [[PubMed](#)]
68. Yao, L.; Lu, N.; Jiang, S. Artificial Neural Network (ANN) for Multi-source PM_{2.5} Estimation Using Surface, MODIS, and Meteorological Data. In Proceedings of the 2012 International Conference on Biomedical Engineering and Biotechnology, Washington, DC, USA, 26–29 May 2012; pp. 1228–1231.
69. Ma, Z.; Xie, Y.; Jiao, J.; Li, L.; Wang, X. The Construction and Application of an Aledo-NDVI Based Desertification Monitoring Model. *Procedia Environ. Sci.* **2011**, *10*, 2029–2035. [[CrossRef](#)]
70. Noi, P.T.; Degener, J.; Kappas, M. Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data. *Remote Sens.* **2017**, *9*, 398. [[CrossRef](#)]
71. Zhou, J.; Li, E.; Wei, H.; Li, C.; Qiao, Q.; Armaghani, D. Random Forests and Cubist Algorithms for Predicting Shear Strengths of Rockfill Materials. *Appl. Sci.* **2019**, *9*, 1621. [[CrossRef](#)]
72. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
73. Quinlan, J.R. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; pp. 343–348.
74. Houborg, R.; McCabe, M.F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 173–188. [[CrossRef](#)]
75. John, R.; Chen, J.; Giannico, V.; Park, H.; Xiao, J.; Shirkey, G.; Yang, Z.; Shao, C.; Laforteza, R.; Qi, J. Grassland canopy cover and aboveground biomass in Mongolia and Inner Mongolia: Spatiotemporal estimates and controlling factors. *Remote Sens. Environ.* **2018**, *213*, 34–48. [[CrossRef](#)]
76. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
77. Chen, B.; Song, Y.; Jiang, T.; Chen, Z.; Huang, B.; Xu, B. Real-Time Estimation of Population Exposure to PM_{2.5} Using Mobile- and Station-Based Big Data. *Int. J. Environ. Res. Public Health* **2018**, *15*, 573. [[CrossRef](#)] [[PubMed](#)]
78. Tobler, W.R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **1970**, *46*, 234. [[CrossRef](#)]
79. Geurts, M.; Box, G.E.P.; Jenkins, G.M. Time Series Analysis: Forecasting and Control. *J. Mark. Res.* **1977**, *14*, 269. [[CrossRef](#)]
80. Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 324–328.

81. Chen, K.; Zhou, Y.; Dai, F. A LSTM-based method for stock returns prediction: A case study of China stock market. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; pp. 2823–2824.
82. Gers, F. Learning to forget: Continual prediction with LSTM. In Proceedings of the 9th International Conference on Artificial Neural Networks: ICANN '99, Edinburgh, UK, 7–10 September 1999.
83. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors* **2017**, *17*, 273. [[CrossRef](#)] [[PubMed](#)]
84. Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Cambridge, MA, USA, 7–12 December 2015; pp. 802–810.
85. Devore, J.L. *Probability and Statistics for Engineering and the Sciences*; Cengage Learning: Boston, MA, USA, 2011.
86. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).