A review of statistical methods used for developing large-scale and long-term PM$_{2.5}$ models from satellite data

Zongwei Ma$^{a,*,}$, Sagnik Dey$^{b,c}$, Sundar Christopher$^d$, Riyang Liu$^a$, Jun Bi$^a$, Palak Balyan$^b$, Yang Liu$^{a,*}$

$^a$ State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing, Jiangsu, China
$^b$ Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India
$^c$ Centre of Excellence for Research on Clean Air, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India
$^d$ Department of Atmospheric and Earth Science, University of Alabama in Huntsville, USA

ARTICLE INFO

Editor: Menghua Wang

Keywords:
Satellite remote sensing
Aerosol optical depth
PM$_{2.5}$ estimates
Statistical methods

ABSTRACT

Research of PM$_{2.5}$ chronic health effects requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks due to their sparse spatial coverage and lack of historical measurements. Estimating PM$_{2.5}$ using satellite-derived aerosol optical depth (AOD) can be used to fill the data gap left by the ground monitors and extend the PM$_{2.5}$ data coverage to suburban and rural areas over long time periods. Two approaches have been applied in large-scale and long-term satellite remote sensing of PM$_{2.5}$, i.e., the scaling and statistical approaches. Compared to the scaling method, the statistical approach has greater prediction accuracy and has been widely used. There is a gap in the current literature and review papers on how the statistical methods work and specific considerations to best utilize them, especially for large-scale and long-term estimates. In this critical review, we summarize the evolution of large-scale and long-term PM$_{2.5}$ statistical models reported in the literature. We describe the framework and guidance of large-scale and long-term satellite-based PM$_{2.5}$ modeling in data preparation, model development, validation, and predictions. Sample computer codes are provided to expedite new model-building efforts. We also include useful considerations and recommendations in covariates selection, addressing the spatiotemporal heterogeneities of PM$_{2.5}$-AOD relationships, and the usage of cross validation, to aid the determination of the final model.

1. Introduction

1.1. Research background

Exposure to particles with aerodynamic diameters less than 2.5 μm (i.e., PM$_{2.5}$) adversely impacts human health (Apte et al., 2018; Pope and Dockery, 2006; Yang and Zhang, 2018). The Global Burden of Disease study has identified air pollution as the fifth largest risk factor for all mortality (Cohen et al., 2017). With rapid economic development and urbanization, widespread PM$_{2.5}$ pollution has attracted great attention worldwide, especially in developing countries.

Studies of the health impacts of PM$_{2.5}$ are the basis for air pollution management. Air pollution exposure data are critical to studies that assess the adverse health impact of ambient PM$_{2.5}$. Traditionally, regional and national ground monitoring networks have been the best means to provide PM$_{2.5}$ estimates. However, establishing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).

Research of chronic health effects often requires knowledge of large-scale and long-term exposure that is not supported by newly established monitoring networks. Traditionally, studies of the health impacts of PM$_{2.5}$ have been based on PM$_{2.5}$ measurements. However, developing and maintaining air pollution monitoring networks are costly, especially at large scales, and may not be a priority for some countries. Spatially, most regions and countries still have sparse or no PM$_{2.5}$ monitors. Approximately 60% of countries have no routine PM$_{2.5}$ monitoring, and only 10% of countries have more than 3 monitors per million inhabitants (Martin et al., 2019). In addition, historical data have been lacking, thereby hampering health studies that require longitudinal data for assessing the role of air quality on health. For example, China established its PM$_{2.5}$ ground monitoring network in late 2012, and data before 2013 have been lacking (Ma et al., 2016a; Ma et al., 2019). In India, ground monitoring of PM$_{2.5}$ started in 2009, but it expanded nationally beyond Delhi only after 2015 (Dey et al., 2020).
monitoring networks due to their sparse spatial coverage and lack of historical measurements. The limitation mentioned above has hindered the studies of the long-term adverse health effects of PM$_{2.5}$ for areas without sufficient PM$_{2.5}$ ground monitors. Few countries have a long-standing network with high spatial coverage, like the US Air Quality System. Even in the US, with good spatiotemporal coverage of ground monitors, it may bias studies of chronic health effects. This is because ground monitors cannot fully cover rural and many suburban areas due to the high operating and maintenance costs, which can lead to spatially mismatched exposure data and populations. To fill the gaps where PM$_{2.5}$ ground monitoring networks are not available, it is critical to estimate the large-scale and long-term PM$_{2.5}$ concentrations to provide enough exposure data that has broader spatiotemporal coverage for chronic health effects studies of PM$_{2.5}$.

With global coverage, satellite remotely sensed aerosol optical depth (AOD) provides an excellent way to estimate surface PM$_{2.5}$ to complement ground monitoring networks both spatially and temporally. Hoff and Christopher (2009) conducted a critical review of the various approaches used to estimate surface PM$_{2.5}$ from satellite data. Since then, hundreds of new studies have been published, attesting to the great importance of this rapidly evolving field of interdisciplinary research.

1.2. Evolution of large-scale and long-term satellite-based PM$_{2.5}$ models

A good starting point to discuss how satellite AOD is related to surface PM$_{2.5}$ is by way of Eq. (1) (Koelemeijer et al., 2000), which shows the dependency of PM$_{2.5}$ and cloud-free AOD relationship on various factors:

$$AOD = PM_{2.5} \times H \times f(RH) \frac{Q_{ext, dry}}{4 \pi \times R_{eff}} = PM_{2.5} \times H \times S$$

(1)

where $H$ is the boundary layer height (BLH), $f(RH)$ is the ratio of ambient and dry extinction coefficients with relative humidity (RH), $\rho$ is the aerosol mass density (g m$^{-3}$), $Q_{ext, dry}$ is the Mie extinction efficiency, and $R_{eff}$ is the particle effective radius. $S$ is the specific extinction efficiency (m$^2$ g$^{-1}$) of the aerosol at ambient RH. This equation assumes aerosols are homogeneously distributed across the BLH.

Based on this theoretical relationship of PM$_{2.5}$ and AOD, two categories of approaches have been developed and applied to estimate large-scale and long-term ground-level PM$_{2.5}$ concentrations, i.e., scaling and statistical approaches. The scaling approach used chemical transport models (CTM) to simulate the association between AOD and PM$_{2.5}$, and then calculate the satellite-derived PM$_{2.5}$ using Eq. (2):

$$\text{Satellite} - \text{derived PM}_{2.5} = \eta \times \text{Satellite AOD}$$

(2)

where $\eta$ is the ratio of the simulated PM$_{2.5}$ to simulated AOD (Liu et al., 2004). An advantage of the scaling model is that it does not require ground measurements to develop the model. Because satellite data and CTM results are generally available globally, air pollutant concentrations can be predicted in any region without any ground measurements (Liu, 2014). It can provide air pollution exposure data over the regions and countries with sparse or no routine air quality monitoring. However, without the calibration of ground measurements, the scaling model tends to have higher prediction errors than those from the statistical models. Furthermore, PM$_{2.5}$-AOD relationships vary spatially and temporally, and aerosol vertical distribution is a major factor (Li et al., 2015a). Numerical model simulations still have challenges in properly characterizing aerosol spatial, temporal, and vertical distributions, thereby leading to uncertainties in the estimation of $\eta$. Due to its independence to ground measurements, the scaling method tends to have higher prediction errors (Liu, 2014).

Compared to the scaling method, another category of methods, i.e., statistical approaches, have greater prediction accuracy. The statistical methods first train the relationship between PM$_{2.5}$ and AOD using various statistical models, including linear regression, advanced statistical models, and machine learning algorithms, using paired ground-measured PM$_{2.5}$ and satellite AOD data. Then the relationship is applied to all satellite AOD data in the study region and period to predict PM$_{2.5}$ concentrations. This extends the PM$_{2.5}$ exposure data from ground monitors to the area and period not covered by ground monitors. Statistical methods require a large amount of ground measured surface PM$_{2.5}$ data to develop a training dataset. Therefore, in regions or countries with sufficient air pollution ground monitors, statistical models have been widely used to estimate with high-accuracy, full-coverage of PM$_{2.5}$ exposures instead of the scaling method.

The simplest form of using AOD to predict PM$_{2.5}$ is to correlate them in a linear regression model. In this method, PM$_{2.5}$ surface concentrations are collocated with satellite AOD in space and time, and a two-variable regression (i.e., AOD is the only independent variable in this model) is formed (Christopher and Gupta, 2002; Wang and Christopher, 2003). Therefore, for a given AOD, using this relationship, a PM$_{2.5}$ value can be derived. This relationship can then be used in satellite grids where ground monitors of PM$_{2.5}$ are not available if the satellite grid is not far from the ground location. Among other reasons, this linear correlation quickly breaks down if aerosols are aloft and require ancillary information to calculate PM$_{2.5}$ (e.g., boundary layer height and relative humidity in Eq. (1)). Therefore, multiple regression techniques tend to include other variables such as meteorological variables and land use information to predict PM$_{2.5}$ (Liu et al., 2009). These variables typically include relative humidity, temperature, boundary layer height, wind speed, wind direction, population density, vegetation cover, road density, etc. It should be noted that the relationship between satellite and ground observations can vary substantially in space and time. This is especially important for large-scale and long-term models. Linear regression models fail to address this issue and often assume a static relationship between PM$_{2.5}$ and AOD, and thus achieve low prediction accuracy. Therefore, few studies have applied linear models for large-scale and long-term estimates. Liu et al. (2005) used a generalized linear model (GLM) to improve the model performance but still failed to address the spatiotemporal heterogeneity of the relationship, with model cross-validation (CV) R$^2$ values below 0.5.

Therefore, the variability of the spatiotemporal relationship is an important factor that the large-scale and long-term statistical models must consider. To account for the spatiotemporal variability of the PM$_{2.5}$-AOD relationship, various advanced statistical models, including generalized additive model (GAM) (Liu et al., 2009), geometrically weighted regression model (GWR) (Hu et al., 2013; Ma et al., 2014; Song et al., 2014), linear mixed-effects model (LME) (Lee et al., 2011; Ma et al., 2016b), geographically and temporally weighted regression (GTWR) (Guo et al., 2017; He and Huang, 2018) were developed. These efforts have greatly improved the prediction accuracy, with the CV R$^2$ generally increased from below 0.5 to the range of 0.6–0.8. Meanwhile, many studies combined two statistical methods to develop two-stage or ensemble models to further improve the model performance (Hu et al., 2014a; Hu et al., 2014b; Liu et al., 2009; Ma et al., 2016a; Ma et al., 2019; Xiao et al., 2018; Xue et al., 2019). Moreover, statistical models have been applied to calibrate the predictions from the scaling approach (Hammer et al., 2020; van Donkelaar et al., 2016) where the residuals of the scaling method derived PM$_{2.5}$ were applied to the GWR model to improve model performance.

In recent years, machine learning algorithms such as neural networks, random forest (RF), and extreme gradient boosting (XGBoost) have been applied to predict PM$_{2.5}$ (Hu et al., 2017; Xiao et al., 2018). Compared to advanced statistical models such as LME, GAM, and GWR, machine learning algorithms have a better ability to address the complex non-linear relationships between PM$_{2.5}$ and AOD and other predictors, thus having greater model accuracy (CV R$^2$ is generally higher than 0.8), but require larger data volume to train the model. To better account for spatiotemporal variability of the PM$_{2.5}$-AOD relationship, some studies developed complex space-time machine learning algorithms, such as geographically-weighted gradient boosting
machine, space-time extremely randomized trees, and random-forest-based spatiotemporal kriging (Shao et al., 2020; Wei et al., 2020a; Zhan et al., 2017; Zhan et al., 2018).

1.3. Gaps in previous reviews

This rapidly growing volume of literature also calls for carefully structured reviews to introduce the latest work to new researchers as well as to reflect on the developments of this field in the past two decades. Three review articles by Xu et al. (2021), Shin et al. (2020), and Chu et al. (2016) summarized the general research progress of satellite-based PM$_{2.5}$ models. These three reviews do not provide adequate details on how statistical methods work for large-scale, long-term PM$_{2.5}$ estimates or short-term pollution episode assessment. A short-term pollution episode assessment mainly focuses on well-defined starting and ending days and times of the exposure period for small scales (city-level or smaller scales), and are usually related to short-term pollution events, e.g., wildfires, dust events and volcano eruptions, and severe pollution episodes (Sorek-Hamer et al., 2020). Short-term pollution episodes are considered to be related to acute health effects. A most recent article has reviewed satellite-based PM$_{2.5}$ modeling strategies for short-term pollution episodes (Sorek-Hamer et al., 2020). Large-scale and long-term PM$_{2.5}$ exposure is related to persistent exposure to air pollution of a population for long periods (one-year or longer period) and large scales (province-level or large scales). However, there are still gaps in the current reviews on how statistical models work and specific considerations to best utilize for large-scale and long-term estimates, especially in the following aspects.

(1) The No Free Lunch (NFL) issue

Although machine learning algorithms seem to outperform advanced statistical models, it should be noted that it is not appropriate to discuss whether a model is good or bad apart from the actual application context. There is not one model which is absolutely better than another. This is NFL theorem (Wolpert and Macready, 1995); that is, if algorithm A outperforms algorithm B in some situations, then there must exist other situations where B outperforms A. For example, the LME model performs well in the New England region of the United States for the year 2003 (Lee et al., 2011). The CV $R^2$ is up to 0.92, which outperforms most studies that use different machine learning approaches in other countries or regions. To compare the performance of different statistical approaches, we need to assess the models in specific application situations, i.e., a specific region, period, and specific dataset. Therefore, comprehensively reviewing the process by which statistical methods work and providing a summary of detailed guidance and recommendations on using these methods are critical for models comparison and selection for specific applications.

(2) Spatiotemporal heterogeneity issue of the PM$_{2.5}$-AOD relationship

Previous studies show that PM$_{2.5}$-AOD relationships have strong spatiotemporal heterogeneities (Guo et al., 2009; Hu et al., 2013). Ignoring the heterogeneities would result in poor model accuracy. Addressing the spatial and temporal nonstationarity is essential for PM$_{2.5}$-AOD statistical modeling. However, the methods and strategies to address this issue have not been summarized in previous reviews.

(3) Model validation issue

To determine a best-performed model, we need to validate different models and compare the validation results. Therefore, validation method and strategy is another key issue for PM$_{2.5}$-AOD statistical modeling in the NFL context. There is a gap in current reviews on model validation methods and their usage for different purposes, e.g., spatial generalization beyond ground monitoring sites and historical hindcasting beyond modeling years, etc.

To promote and enhance the application of the statistical methods of PM$_{2.5}$ satellite remote sensing, this review summarizes the main work flow of statistical methods and provides guidance on how these methods can be used to improve the estimation of surface PM$_{2.5}$, especially for large-scale and long-term estimates. We first describe the overall framework of statistical modeling for satellite-based PM$_{2.5}$. We then review the detailed process and summarize key considerations of PM$_{2.5}$-AOD satellite models, which includes the current progress and problems in the methods of addressing the spatiotemporal heterogeneities of the PM$_{2.5}$-AOD relationship, and model validation. This paper will serve as a reference for the science/application communities of air pollution satellite remote sensing.

2. Framework of statistical modeling

Fig. 1 summarizes the overall framework and workflow of the statistical approach of PM$_{2.5}$ using satellite remote sensing. Overall, it can be divided into four steps: 1) model preparation (Section 3, including variables selection, data collection, and data processing); 2) model development (Section 4, including model selection and model fitting); 3) model validation and prediction in the model years (Section 5, including model cross-validation and model predictions in the model years); and 4) historical data hindcasting (Section 6, including predictions and validations of historical estimates beyond the model years). Each section is highlighted using different color schemes in the diagram (Fig. 1).

3. Model preparation

3.1. Variables selection and data collection

Generally, four types of datasets are required in statistical models, including ground PM$_{2.5}$, satellite AOD, meteorological variables, land use, and other variables. Ground PM$_{2.5}$ is used as the dependent variable, and other variables are used as independent variables (also known as predictors) in statistical models.

(1) Ground PM$_{2.5}$ measurements. Statistical models require adequate ground-measured PM$_{2.5}$ data to develop the models. Various studies indicate that it is important to collect at least one year worth of site-based 24-h ground measurements to develop large-scale and long-term PM$_{2.5}$ prediction models (Liu, 2013). For statistical models at the regional to the national scale, only an extensive routine ground monitoring network, such as ground PM$_{2.5}$ networks of China National Environmental Monitoring Center (CNEMC) and the United States Environmental Protection Agency (USEPA), can provide sufficient, large scale, and long-term ground monitoring data. The CNEMC and USEPA networks have been widely applied in large-scale and long-term satellite-based PM$_{2.5}$ statistical models (Di et al., 2016; Hu et al., 2014a; Liang et al., 2020; Ma et al., 2016a). Data from limited research sites may be used as an independent validation of a large-scale model.

(2) Satellite AOD data. Satellite retrieved AOD is the major dependent variable in PM$_{2.5}$ in statistical modeling. Generally, AOD refers to mid-visible (550 nm) values. Satellite AOD data are usually available at two levels: Level 2 (L2) and Level 3 (L3). L2 data are the retrievals at a finer resolution compared to L3 data. While L2 data are available for each satellite swath, L3 data are gridded at a coarser resolution for the ease of the user community. It is advisable to use L2 data for any quantitative analysis as additional information such as ‘quality flag’, ‘aerosol mixture’, etc., are also provided for data screening. There have been numerous papers that have validated satellite AOD values against ground-based Aerosol Robotic Network (AERONET)
measurements. For example, Gupta et al. (2018) perform global validation of the MODIS 3-km aerosol optical depth (AOD) by comparing it against AERONET and found a high correlation ($R = 0.87$). Sayer et al. (2019) validate MODIS Collection 6.1 and VIIRS Version 1 Deep Blue AOD data over land globally. Such studies provide useful information for the selection, processing, and application of AOD datasets. Commonly used satellite AOD products and relevant details are shown in Table 1. Details about major instruments, algorithms, products, and evaluation results for satellite AOD retrievals can be found in a recent review article (Wei et al., 2020b).

(3) Meteorological variables. Meteorological parameters are important modifiers of the PM$_{2.5}$-AOD relationship and may influence the size, composition, and mixing of particles (Gupta and Christopher, 2009; Liu, 2013) and are therefore incorporated into the satellite-based statistical models as ancillary variables to improve the model accuracy. Commonly used meteorological variables include wind speed (WS), temperature (TEMP), planetary boundary layer height (PBL), surface pressure (SP), relative humidity (RH), and visibility (VIS). For point locations, these parameters are freely available from surface weather stations all over the world. For example, we can download meteorological data of worldwide weather stations from the National Centers for
Environmental Information (NCEI, https://gis.ncdc.noaa.gov/maps/) of the National Oceanic and Atmospheric Administration (NOAA) of the United States. Besides, various numerical model outputs provide gridded assimilated or reanalysis meteorological datasets (Table 2). Examples of PM$_{2.5}$-AOD models using meteorological variables can be found in Table 4.

(4) Land use and other ancillary variables

Land use parameters, such as normalized difference vegetation index (NDVI), urban cover, distance to a major road, forest cover, etc., are directly linked to emission sources and are also valuable ancillary variables in the satellite-based models (Liu, 2013). Besides, other variables, e.g., elevation, population, surface fire counts, etc., are also important ancillary variables since they affect the pollution transport conditions and/or are linked to emissions of PM$_{2.5}$. There have been many available global datasets, such as the MODIS NDVI products (https://earthdata.nasa.gov/eosdis/daacs/landsat) (Huete et al., 1999), European Space Agency (ESA) Global Land Cover data (GlobCover; https://due.esrin.esa.int/page_globcover.php) (Bontemps et al., 2011), the LandScan population dataset (https://landsan.org/landscan-datasets), MODIS active fire spots from Fire Information for Resource Management System (FIRMS, https://earthdata.nasa.gov/data/near-real-time-data/firms), etc. Commonly used datasets for land use and other variables can be found in Table 3. Examples using land use and other variables can be found in Table 4.

### Table 1

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Product Name</th>
<th>Spatial Resolution</th>
<th>Spatial coverage</th>
<th>Temporal coverage</th>
<th>Key References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terra/ Aqua</td>
<td>MODIS</td>
<td>10 km/3</td>
<td>Global</td>
<td>Terra (2000–)</td>
<td>(Gupta et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>km</td>
<td></td>
<td></td>
<td>(Isu et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>MODIS</td>
<td>10 km</td>
<td>Aqua</td>
<td></td>
<td>(Ivy-Justin et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>DB</td>
<td>1 km</td>
<td></td>
<td></td>
<td>(Lin et al., 2010)</td>
</tr>
<tr>
<td></td>
<td>MAIAC</td>
<td></td>
<td></td>
<td></td>
<td>(Garay et al., 2017)</td>
</tr>
<tr>
<td>S-NPP</td>
<td>VIIRS</td>
<td>6 km/750 m</td>
<td>Global</td>
<td>2012-</td>
<td>(Jackson et al., 2013)</td>
</tr>
<tr>
<td>Terra</td>
<td>MISR</td>
<td>17.6 km/4.4 km</td>
<td>Global</td>
<td>2000-</td>
<td>(Kahn et al., 2010)</td>
</tr>
<tr>
<td>Himawari-8</td>
<td>AHI</td>
<td>0.05°</td>
<td>60°S-60°N, 80°E-160°W</td>
<td>2015-</td>
<td>(Yoshida et al., 2013)</td>
</tr>
</tbody>
</table>

Note. The acronyms are as follows, and the full names are in parentheses: MODIS (Moderate Resolution Imaging Spectroradiometer), DT (Dark Target), DB (Deep Blue), MAIAC (multiscale implementation of atmospheric correction), S-NPP (the Suomi National Polar-orbiting Partnership), VIIRS (Visible Infrared Imaging Radiometer), MISR (Multi-angle Imaging Spectroradiometer), AHI (Advanced Himawari Imager).

### Table 2

<table>
<thead>
<tr>
<th>Product</th>
<th>Spatial resolution</th>
<th>Spatial coverage</th>
<th>Temporal resolution</th>
<th>Temporal coverage</th>
<th>Key References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEOS-FP</td>
<td>0.25° x 0.3125°</td>
<td>Global</td>
<td>hourly, 3-hourly</td>
<td>2012-</td>
<td>(Lucchesi, 2018)</td>
</tr>
<tr>
<td>MERRA-2</td>
<td>0.5° x 0.625°</td>
<td>Global</td>
<td>hourly, 3-hourly, 6-hourly, daily</td>
<td>1979-</td>
<td>(Bosilovich et al., 2015)</td>
</tr>
<tr>
<td>ECMWF ERA5</td>
<td>31 km</td>
<td>Global</td>
<td>hourly</td>
<td>1950–1978(Preliminary version), 1979-</td>
<td>(Hersbach et al., 2020)</td>
</tr>
<tr>
<td>NARR</td>
<td>approximately 0.3°</td>
<td>North America</td>
<td>3-hourly</td>
<td>1979-</td>
<td>(Mesinger et al., 2006)</td>
</tr>
<tr>
<td>NLDAS-2</td>
<td>0.125° x 0.125°</td>
<td>CONUS</td>
<td>hourly</td>
<td>1979-</td>
<td>(Cogrover et al., 2005)</td>
</tr>
<tr>
<td>NCEI GHCN</td>
<td>Station-based</td>
<td>Global</td>
<td>daily, monthly</td>
<td>19th century-</td>
<td>(Menne et al., 2012)</td>
</tr>
<tr>
<td>JRA-55</td>
<td>1.25° x 1.25°</td>
<td>Global</td>
<td>3-hourly, 6-hourly, daily</td>
<td>1958-</td>
<td>(Kobayashi et al., 2015)</td>
</tr>
</tbody>
</table>

Note. The acronyms are as follows, and the full names are in parentheses: GEOS-5 (Goddard Earth Observing System, Version 5), GEOS-FP (Goddard Earth Observing System-Forward Processing), MERRA-2 (The Modern-Era Retrospective Analysis for Research and Applications, Version 2), ECMWF (European Center for Medium-Range Weather Forecast), NARR (North American Regional Reanalysis), NLDAS-2 (Phase 2 of the North American Land Data Assimilation System), GHCN (Global Historical Climatology Network), CONUS (continental United States), NCEI (National Centers for Environmental Information), JRA-55 (the Japanese 55-year reanalysis).

### 3.2. Considerations in variables selection

Meteorological and land-use variables are optional; we can use AOD as the only predictor (Gupta and Christopher, 2008; Lee et al., 2011; Yu et al., 2017). However, meteorological and land-use variables are strongly recommended since they can significantly improve model accuracy. For the selection of the ancillary meteorological and land-use variables, the following issues should be considered:

1. While it is tempting to use any variable to see if it improves the model efficiency, it is important to ensure that this is based on the physics of the problem. For example, humidity is an important factor because of particle hygroscopic growth (Titos et al., 2016), wind speed is important because the speed of the air mass governs the spatiotemporal collocations of PM$_{2.5}$ (Chen et al., 2018b), and surface fire counts are directly associated with PM$_{2.5}$ emissions (Hu et al., 2014c), etc.

2. Implementation of multicollinearity analysis. When the predictors in a model are not independent of one another, multicollinearity problems arise and will lead to biased estimation (Yoo et al., 2014). When two or more covariates are highly correlated with each other, only one of them could be used in the model. The method to test multicollinearity is summarized in Section 4.1.

3. Testing significance of variables (for advanced statistical models) or the importance of parameters (for machine learning algorithms), and remove the variables that are not significant/important. For example, Hu et al. (2014a) developed separated models in the southeastern US for each year from 2001 to 2010. For each final annual model, only statistically significant variables were retained.

4. Investigation of dimensionality and assessing model performance by adding/removing variables. Chang et al. (2014) conducted experiments with different sets of predictors, including AOD, meteorology (Met), and land use (LU) variables. They compared the performance of the model with all variables, the AOD only model, AOD + Met model, AOD + LU model, and Met + LU model. In the 10-year study in the southeastern US (Hu et al., 2014a), they compared the model using wind speed and the model U-wind + V-wind, and selected wind speed from five years, while selected U-wind and V-wind in the models of other years.

5. For variables with more than one data source, models can be built separately using different data sources. For example, Hu et al. (2013) extracted meteorological fields from two datasets, i.e., the North American Regional Reanalysis (NARR) and the North American Regional Reanalysis (NARR).
Table 3
Commonly used datasets for land use and other variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Product</th>
<th>Spatial Resolution</th>
<th>Spatial coverage</th>
<th>Temporal resolution</th>
<th>Temporal coverage</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation</td>
<td>MODIS NDVI</td>
<td>250 m</td>
<td>Global</td>
<td>every 16 days</td>
<td>2000-</td>
<td>(Iliuta et al., 1999)</td>
</tr>
<tr>
<td>Land Cover</td>
<td>GlobCover</td>
<td>300 m</td>
<td>Global</td>
<td>multi-year</td>
<td>2004-June 2006 and January-December 2009</td>
<td>(Bontemps et al., 2011)</td>
</tr>
<tr>
<td>Land Cover</td>
<td>CCI-LC</td>
<td>300 m</td>
<td>Global</td>
<td>annual</td>
<td>1992-2015</td>
<td>(Bontemps et al., 2012)</td>
</tr>
<tr>
<td>Fire location</td>
<td>FIRMS</td>
<td>1 km</td>
<td>Global</td>
<td>daily</td>
<td>2000-2012</td>
<td>(Girol et al., 2018)</td>
</tr>
<tr>
<td>Elevation</td>
<td>ASTER GDEM</td>
<td>30 m</td>
<td>83° N-83° S</td>
<td>multi-year</td>
<td>2000-2013</td>
<td>(Tachikawa et al., 2011)</td>
</tr>
<tr>
<td>Emission</td>
<td>SRTM</td>
<td>30 m</td>
<td>Global</td>
<td>/</td>
<td>2000</td>
<td>(Farr et al., 2007)</td>
</tr>
<tr>
<td>Emission</td>
<td>MEIC</td>
<td>0.25° × 0.25°</td>
<td>Global</td>
<td>monthly</td>
<td>1960-2014</td>
<td>(Huang et al., 2017; Huang et al., 2014)</td>
</tr>
</tbody>
</table>

Note. The acronyms are as follows, and the full names are in parentheses: NDVI (Normalized Difference Vegetation Index), GlobCover (Global Land Cover), CCI-LC (Climate Change Initiative - Land Cover), FIRMS (The Fire Information for Resource Management System), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), GDEM (Global Digital Elevation Model Version), SRTM (the Shuttle Radar Topography Mission), PKU (Peking University), MEIC (Multi-resolution Emission Inventory for China).

Table 4
Variables used in some typical studies.

<table>
<thead>
<tr>
<th>Variables</th>
<th>PM&lt;sub&gt;2.5&lt;/sub&gt;, from US Environmental Protection Agency (EPA) compliance network</th>
<th>AOD</th>
<th>Meteorological parameters</th>
<th>Land use and other variables</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM&lt;sub&gt;2.5&lt;/sub&gt;, from US EPA compliance network</td>
<td>Geostationary Operational Satellite (GOES) aerosol/smoke product AOD</td>
<td>/</td>
<td>PBL, RH, TEMP, WS, etc., generated by the rapid update cycle (RUC) model</td>
<td>Road length, elevation, population, land use types, from Environmental Systems Research Institute (ESRI) StreetMap, US Geological Survey’s National Land Cover Database, and US census data</td>
<td>(Lee et al., 2009)</td>
</tr>
<tr>
<td>PM&lt;sub&gt;2.5&lt;/sub&gt;, from US EPA compliance network</td>
<td>Terra and Aqua MODIS DT AOD (Collection 5)</td>
<td>/</td>
<td>/</td>
<td>(Lee et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>PM&lt;sub&gt;2.5&lt;/sub&gt;, from US EPA compliance network</td>
<td>Terra and Aqua MAIAC AOD</td>
<td>WS, from the North American Land Data Assimilation System (NLDAS)</td>
<td>Elevation from US national elevation dataset, road length from ESRI Street Map, forest cover from US National Land Cover Database, point emission from U.S.EPA National Emissions Inventory</td>
<td>(Bu et al., 2014a)</td>
<td></td>
</tr>
<tr>
<td>PM&lt;sub&gt;2.5&lt;/sub&gt;, from China’s air quality monitoring network</td>
<td>Aqua MODIS DT and DB AOD (Collection 6)</td>
<td>WS, PBL, SP, RH, Precipitation, from GEOS 5.2.0 and GEOS 5-FF</td>
<td>Forest and urban covers from GlobCover dataset, fire counts from FIRMS, and grid cell coordinates</td>
<td>(Ma et al., 2016a)</td>
<td></td>
</tr>
</tbody>
</table>

American Land Data Assimilation System (NLDAS). They compared the models using NARR and NLDAS datasets and found the NLDAS model performs slightly better. The data source with better model results can be selected.

Table 4 summarizes the variables used in some references.

3.3. Data processing

Most satellite-based statistical models are developed on a daily (24-h) basis. In this section, we mainly focus on the daily model. It should be noted that some studies used geostationary satellite AOD, and the models were developed on an hourly basis (Chen et al., 2019; Tang et al., 2019). However, the considerations of the data processing of hourly models should be similar.

(1) Data screening and quality assurance

Statistical models are data-driven approaches. Data quality can greatly influence the model performance. For a satellite AOD data product, the quality assurance flag (QAF), which denotes the data quality for each AOD data point, must be carefully used. For example, the quality of MODIS Collection 6 AOD data has been marked as high (QAF = 3), good (QAF = 2), marginal (QAF = 1), and no confidence (QAF = 0) (Levy et al., 2013; Sayer et al., 2014). Over land, retrievals with QAF = 2 and 3 for Deep Blue (DB) AOD and QAF = 3 for Dark Target (DT) AOD are recommended (Hsu et al., 2013; Levy et al., 2013; Sayer et al., 2014). Generally, it is important to use the AOD data with recommended or the highest quality QAF. Critical information such as AOD inversion algorithm and quality assurance can usually be found in user guide documents (e.g., MAIAC AOD’s user guide: https://modis-land.gsfc.nasa.gov/pdf/MCD19_UserGuide_final_Feb-6-2018.pdf) or relevant peer-reviewed papers (e.g., MODIS Collection 6 AOD algorithm paper (Levy et al., 2013)) of AOD products.

Also, special attention should be paid to the data quality of ground PM<sub>2.5</sub> measurements, especially for measurements from a newly established monitoring network. For example, China’s PM<sub>2.5</sub> monitoring network was established in late 2012. Previous studies have observed and addressed the data quality issues of China’s PM<sub>2.5</sub> network, e.g., PM<sub>2.5</sub> measurements exceeding co-located PM<sub>10</sub> levels measured concurrently, same measurements registered for consecutive hours, extremely high or extremely low values for a site compared with its neighboring sites (Liu et al., 2016; Ma et al., 2019; Rohde and Muller, 2015; Xiao et al., 2018). Abnormal values should be removed before further data processing.

(2) Modeling grid definition
The input datasets have various spatial resolutions and are often difficult to match for modeling. Therefore, we need to create a grid and align all data to the predefined grid before developing the algorithm. Additionally, the model grid should cover the whole study region. Since satellite AOD is the main predictor for PM$_{2.5}$ modeling, the resolution of the grid cells should be comparable with the spatial resolution of the AOD products. More considerations about model grid definition can be found in a recent review (Sorek-Hamer et al., 2020).

Typically, satellite-based statistical models are developed on a daily basis. Hourly PM$_{2.5}$ measurements should be averaged to daily mean values or, when available, use the 24-h averages. Meteorological data are usually recorded every hour, 3-h, or 6-h and should be averaged to daily mean data. It should be noted that the time series data are always recorded in UTC. In that case, we should change the UTC to local time before calculating the daily mean values. Land use and population datasets are updated less regularly. Daily values are not required, and the most recent values must be used.

The commonly used AOD data are derived from sensors aboard polar orbit satellites, such as MODIS aboard the Terra satellite, which crosses the equator at approximately 10:30 a.m. local time and has a global coverage of 1–2 days (Hoff and Christopher, 2009). That means for a specific location, such satellite sensors measure the AOD only once in one or more days. For these satellite AOD data, we assume that the instantaneous value can represent the daily average AOD. Sometimes, we can merge two satellite AOD products so that the AOD data can better represent the daily average value. For example, Aqua MODIS crosses the equator at approximately 1:30 p.m. local time. Some studies averaged the Aqua and Terra MODIS AOD (Hu et al., 2013). The geostationary satellites (e.g., GOES and geostationary ocean color imager (GOCI), etc.) (Liu et al., 2009; Ryu et al., 2012) allow AOD retrievals many times in one day and therefore better represent the daily average AOD. However, the limitation of a geostationary satellite is that it only covers a specific region and does not have global coverage.

(4) Aligning data to predefined grid

As we mentioned above, to facilitating data matching, all data should be resampled and aligned to the predefined grid. The resampling and aligning strategy should be determined according to the characteristics of the dataset, as follows.

Daily PM$_{2.5}$ ground measurements should be assigned to the appropriate grid cells. If a grid cell has more than 1 monitor, the measurements from these monitors should be averaged first and then assigned to this grid cell (Ma et al., 2016a; Xue et al., 2019). For point measurements from weather stations and datasets with regular grid cells but coarser resolution than a predefined grid, the data should be interpolated to each grid cell of the predefined grid using spatial interpolation methods such as inverse distance weighted (IDW) or ordinary kriging (OK) methods (Xiao et al., 2018; Xue et al., 2019). For a dataset with regular grid cells but finer resolution than a predefined grid, the data points that fall within the same grid cell of the predefined grid should be averaged and assigned to the corresponding grid cell (Xiao et al., 2018). For some variables such as point emissions and road lengths, values that fall in one grid cell should be summed (Hu et al., 2017).

Some AOD products have irregular grid cells. For example, the sizes and geographical locations of MODIS Level 2 AOD pixels vary in space and time. A predefined 0.1° grid cell may have multiple AOD pixels (e.g., near the center of each satellite swath), or an AOD pixel may cover multiple 0.1° grid cells (e.g., near the edge of each swath). Meanwhile, many grid cells may have missing AOD (up to 40–50% data loss) due to clouds and bright surfaces (Liu, 2014). For data processing of such AOD products, Thiessen polygons for individual MODIS AOD pixels can be created to assign AOD data to the predefined grid (Ma et al., 2016a; Ma et al., 2016b).

(5) Gap filling of AOD data

A large amount of missing AOD values due to clouds and bright surfaces has been a big issue that could induce uncertainties in satellite remote sensing of PM$_{2.5}$ (Liu, 2014). One way to minimize the spatial gap is a fusion of AOD data from multiple AOD products following a statistical approach (Boys et al., 2014; Ma et al., 2016a; Tang et al., 2016). In the most recent years, some studies have developed advanced statistical or machine learning models to fill the gap of the missing AOD, which can improve the AOD coverage to 100% (Bi et al., 2019; Tang et al., 2019; Xiao et al., 2018; Xiao et al., 2017; Zhao et al., 2019).

(6) Data matching

Two matched datasets are required: the modeling dataset and the prediction dataset. The modeling dataset is needed to train the statistical or machine learning model, including spatiotemporally matched ground PM$_{2.5}$, satellite AOD, and ancillary variables. The modeling dataset can be established in three ways: 1) In the first method, all variables, including resampled ground PM$_{2.5}$, satellite AOD, and other ancillary variables, are matched by grid cell ID and day of the year (DOY) (Ma et al., 2016a; Xue et al., 2019); 2) The second method is based on the original ground PM$_{2.5}$ monitoring sites. For each PM$_{2.5}$ site, a square or radius buffer zone can be created. The daily satellite AOD and other ancillary variables within the buffer zone should be averaged and matched to the daily PM$_{2.5}$ concentrations of this site. Alternatively, we can match satellite AOD and ancillary variables to PM$_{2.5}$ ground monitoring sites using the nearest neighbor approach (Hu et al., 2014a); 3) In the third method, we can calculate the AOD and ancillary variables at the locations of ground PM$_{2.5}$ monitors using a spatial interpolation method such as IDW, and allocate them to ground PM$_{2.5}$ measurements. For the modeling dataset, there are no strict rules about which method to select. We can choose the appropriate method according to the actual conditions.

The prediction dataset comprises spatiotemporally matched satellite AOD and other ancillary variables (excluding ground PM$_{2.5}$). The data matching should be based on the resampled data. All variables (including satellite AOD and other ancillary variables, except PM$_{2.5}$) are matched by grid cell ID and DOY.

4. Model development considerations

4.1. Multicollinearity problem

The modeling dataset, i.e., the matched ground PM$_{2.5}$, satellite AOD, and other ancillary variables, is used to develop the statistical or machine learning model. Before model development, we should examine the multicollinearity for independent variables. Multicollinearity becomes a problem when independent variables are highly correlated and therefore are not truly independent, and will lead to biased estimation (Yoo et al., 2014). Variance inflation factor (VIF) analysis can be used to test the severity of the multicollinearity problem. Generally, the multicollinearity level can be accepted if the VIF value is less than 10, and more stringent VIF suggests less than 4 (O’Brien, 2007).

4.2. Single-stage models

As an example, the simplest statistical model is the ordinary least squared (OLS) regression, as follows:

\[
PM_{2.5} = \beta_0 + \beta_1 \times AOD_{st} + \beta_2 \times WS_{st} + \epsilon_{st}
\]  

(3) where PM$_{2.5}$, WS$_{st}$, and $\epsilon_{st}$ are the PM$_{2.5}$ concentration, wind speed, and...
model residual at grid cell $s$ or site $s$ on day $t$. $\beta 0$ is intercept and $\beta 1$, $\beta 2$ are slopes. Here we use AOD and wind speed for demonstration. In practice, more ancillary variables can be added. We use R code for demonstration, which can be also a guidance for the processing using any other tool such as Python. We can use “lm()” function in R for OLS modeling, as the following R code:

\[ \text{OLS}_{\text{model}} \leftarrow \text{lm}(\text{PM} \sim \text{AOD} + \text{WS}, \text{data} = \text{modeling}_{\text{dataset}}) \]

We can use “summary(OLS_model)” to retrieve the regression coefficients $\beta 1$ and $\beta 2$, and their statistical significance. For advanced statistical models, we can test the significance to decide whether to retain a variable. For machine learning models, we can estimate the importance of predictors to guide parameter selection. Table 5 gives sample codes and brief descriptions for typical single-stage models. Detailed descriptions about typical statistical or machine learning models, their advance, and their advantages and disadvantages can be found in three previous reviews (Chu et al., 2016; Shin et al., 2020; Xu et al., 2021).

4.3. Two-stage or hierarchical models

To further improve model performance, many studies developed multi-stage models. The most common type is the two-stage model, which uses the PM$_{2.5}$ residuals from the first-stage model as a dependent variable to develop the second-stage model. Here is an example of a two-stage OLS model:

First stage: same as the single-stage model as shown in Section 4.2.

The second stage (using air temperature as an independent variable for demonstration):

\[ \text{PM}_{\text{resid}} = \beta 0 + \beta 1 \times \text{Temp}_a + \epsilon_{\text{st}} \] (4)

where $\text{PM}_{\text{resid}}$, $\text{Temp}_a$, and $\epsilon_{\text{st}}$ are the PM$_{2.5}$ concentration residual from the first stage model (i.e., $\epsilon_{\text{st}}$ in the Eq. (3)), air temperature, and residual of the second-stage model at grid cell $s$ or site $s$ on day $t$. The R codes are as follows:

\[ \text{PM}_{\text{resid}} \leftarrow \text{OLS}_{\text{model}} \text{residuals} # \text{To obtain residuals from the first stage model and assign to a new variable 'PM}_{\text{resid}}'. \]

\[ \text{modeling}_{\text{dataset}} \leftarrow \text{cbind(modeling}_{\text{dataset}}, \text{PM}_{\text{resid}}) # \text{To combine PM}_{\text{resid}} to the modeling dataset}. \]

\[ \text{OLS}_{\text{model}} \text{2nd} \leftarrow \text{lm}(\text{PM}_{\text{resid}} \sim \text{Temp}, \text{data} = \text{modeling}_{\text{dataset}}) # \text{Second stage model, to model the relationship between PM}_{\text{resid}} and \text{Temp using OLS}. \]

We can use different combinations of single-stage models to construct a two-stage model. For the second stage model, we can select different variables, including AOD or not. For example, Hu et al. (2014b) developed a two-stage model, LME for the first stage and GWR for the second stage. For the second stage model, they used AOD as the only predictor. Ma et al. (2016a) used LME for the first stage model and GAM for the second stage model. They used geographical coordinates, forest cover, and urban cover as independent variables in the second stage model. Although these two-stage models have different combinations, their basic idea is similar to the above-mentioned two-stage OLS model.

4.4. Hybrid models combining scaling and statistical approaches

Statistical models can be applied to calibrate the PM$_{2.5}$ predictions from the scaling approach. First, PM$_{2.5}$ predictions can be derived using the scaling approach. Then the bias between PM$_{2.5}$ predictions from the scaling approach and collocated ground-based PM$_{2.5}$ observations should be calculated. A statistical model can then be developed to describe the relationship between bias and covariates, similar to the second-stage model in Section 4.3. For example, two studies (Hammer et al., 2020; van Donkelaar et al., 2016) used GWR to model the relationship between bias and covariates. They used simulated aerosol composition and land use information as GWR predictors.

Table 5: Typical advanced statistical models or machine learning algorithms for PM$_{2.5}$ satellite remote sensing (using AOD and wind speed as demonstrations).

<table>
<thead>
<tr>
<th>Model *</th>
<th>R package</th>
<th>R code **</th>
<th>Case ***</th>
<th>Major advantage and limitation****</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>stats</td>
<td>GLM_model &lt;- glm (PM ~ AOD + WS, family = gaussian, data = modeling_dataset)</td>
<td>Liu et al., 2005</td>
<td>- Provide interpretable explanations of model decisions - Difficult to capture the complex nonlinear relationship between variables and spatial non-stationary issue - Can display the interaction of sample values often ignored to alleviate spatial non-stationary issue - Can display the proportion of sample values where samples are unevenly distributed or at large spatial scale</td>
</tr>
<tr>
<td>GWR</td>
<td>spgwr</td>
<td>GWR_model &lt;- gwr (PM ~ AOD + WS, coords = modeling_dataset, coords = gwrweight)</td>
<td>Ma et al., 2013; Seng et al., 2014</td>
<td>- Incorporate flexible components of smooth function to capture nonlinearity and interactions - Difficult to incorporate large number of variables because of multimodality problems - Express fixed effects and random interference caused by the relationship between variables - Can display the importance of features - Cannot extrapolate beyond the range of sample values</td>
</tr>
<tr>
<td>GAM</td>
<td>mgcv</td>
<td>GAM_model &lt;- gam (PM ~ s(AOD) + s(WS), family = gaussian, data = modeling_dataset)</td>
<td>Xiao et al., 2018</td>
<td>- Reduce variance and control overfitting problems well - Can display the importance of features</td>
</tr>
<tr>
<td>LME</td>
<td>nlme</td>
<td>LME_model &lt;- lme (fixed = PM ~ AOD + WS, random = list(DOV = ~ 1 + AOD + WS), data = modeling_dataset)</td>
<td>Kloog et al., 2011; Lee et al., 2011</td>
<td>- Incorporate large number of variables because of multimodality problems - Reduce variance and control overfitting problems well - Can display the importance of features</td>
</tr>
<tr>
<td>RF</td>
<td>randomForest</td>
<td>RF_model &lt;- randomForest(PM ~ AOD + WS, data = modeling_dataset, mtry, ntree)</td>
<td>Hu et al., 2017; Xiao et al., 2018</td>
<td>- Incorporate large number of variables because of multimodality problems - Reduce variance and control overfitting problems well - Can display the importance of features</td>
</tr>
<tr>
<td>XGBoost</td>
<td>xgboost</td>
<td>XGBoost_model &lt;- xgb.train(data = xgb.DMatrix(data = as.matrix(modeling_dataset), c('AOD', 'WS')), label = modeling_dataset)</td>
<td>Xiao et al., 2018</td>
<td>- Incorporate large number of variables because of multimodality problems - Reduce both bias and variance of models - Display the importance of features - Require tuning of model parameters</td>
</tr>
</tbody>
</table>

(continued on next page)
4.5. Ensemble models

In ensemble modeling, two or more ensemble members are developed first. For each member, a scaling method, single-stage model, two-stage model, or a hybrid model can be used. Then PM$_{2.5}$ predictions are derived from each member, respectively. Finally, an ensemble model (could be a statistical or machine learning model) is used to fuse the PM$_{2.5}$ predictions derived from all members. For example, Xiao et al. (2018) developed three models, including random forest, generalized additive, and extreme gradient boosting models. Then a generalized additive ensemble model was developed to combine predictions from different algorithms. In another study, Chen et al. (2019) made ensemble PM$_{2.5}$ predictions from three machine learning algorithms using a multiple linear regression model.

4.6. Addressing the spatiotemporal heterogeneities of the PM$_{2.5}$-AOD relationship

Addressing the spatial and temporal nonstationarity for the PM$_{2.5}$-AOD relationship is essential for a large-scale and long-term model. We summarized five strategies to address this issue from previous studies.

(1) Using spatial and temporal models

Statistical models such as GWR, LME, GTWR, and space-time machine learning algorithms have been used to account for the spatial and temporal nonstationarity of PM$_{2.5}$-AOD relationships. GWR is a spatial regression model that can examine spatial variability and nonstationarity by producing local regression results (Brunsdon et al., 1996) and has been widely applied in PM$_{2.5}$-AOD modeling (Hu et al., 2013; Ma et al., 2014; Song et al., 2014). For temporal nonstationarity issues, Lee et al. (2011) developed an LME model with day-specific random slopes and intercepts to establish the day-specific PM$_{2.5}$-AOD relationships. Ma et al. (2016b) proposed a nested LME model with nested month-, week-, and day-specific random effects of PM$_{2.5}$-AOD relationships. He and Huang (2018) and Guo et al. (2017) used a GTWR model, which can simultaneously address the spatial and temporal nonstationarity based on spatiotemporal distances and spatiotemporal weight matrices (Fotheringham et al., 2015) in PM$_{2.5}$-AOD modeling. Zhan et al. (2017) developed a geographically-weighted gradient boosting machine (GW-GBM) by improving GBM through building spatial smoothing kernels to weigh the loss function.

(2) Incorporating spatial and temporal terms

Some studies incorporate spatial and temporal terms into non-spatial and non-temporal models to address the spatiotemporal heterogeneity of PM$_{2.5}$-AOD relationships. For example, a two-dimension smooth term of geographical coordinate has been incorporated into a GAM model in PM$_{2.5}$-AOD modeling studies (Ma et al., 2016a; Xiao et al., 2017). In a machine learning approach, some studies used the haversine approach to calculate the great-circle distance between two points on a sphere specified by their latitudes and longitudes as the spatial term (Wet et al., 2020a) and used DOY as a temporal term (Zhan et al., 2017). To account for spatial and temporal heterogeneity simultaneously, some studies used the convolutional layer as a spatiotemporal term in a machine learning algorithm (Di et al., 2016; Hu et al., 2017).

(3) Developing sub-period models

Developing separate models for each day, each week, or each month, i.e., sub-period models, is also a feasible way to address the temporal nonstationarity issue. For example, separate GWR models were fitted for each day in two previous studies (Hu et al., 2013; Ma et al., 2014).

(4) Developing sub-region models

Some previous studies developed separate models for sub-regions to address the spatial nonstationarity, improving the model accuracy (Ma et al., 2016a; Xiao et al., 2018). For example, Ma et al. (2016a) fitted the first-stage LME model for each province separately and obtained model cross-validation (CV) $R^2$ = 0.78. However, they found that the first-stage CV $R^2$ dropped to 0.63 if a single LME model was fitted for the whole of China.

It should be noted that developing sub-region models or using sub-region as an interaction variable would cause non-smooth (discontinuous) PM$_{2.5}$ predictions in the border areas of each province. To generate a smooth PM$_{2.5}$ concentration surface for the whole domain, a buffer zone should be created for each sub-region and using the matched dataset within the sub-region and its buffer zone to develop the sub-region model, and the overlapping predictions from buffer zones of neighboring regions should be averaged (Ma et al., 2016a; Xiao et al., 2018). Moreover, using buffer zones can ensure sufficient model fitting data in each sub-region model (Ma et al., 2016a; Xiao et al., 2018).

(5) Using a hybrid approach

Many previous studies used a combination of two or more approaches mentioned above to account for the spatiotemporal nonstationarity. For example, Hu et al. (2013) and Ma et al. (2014) used the GWR model to account for spatial nonstationarity and developed separate daily models to account for temporal nonstationarity. Xiao et al. (2017) used a first-stage LME model to address the temporal heterogeneity issue and used a second stage GAM, which included a smooth term of coordinates to account for the spatial heterogeneity of PM$_{2.5}$-AOD relationships. Hu et al. (2014a) and Hu et al. (2014b) developed a two-stage spatiotemporal model (first-stage LME + second-stage GWR).

4.7. Recommendations for model development

According to the NFL theorem, it is difficult to say which model structure or algorithm (including the selection of the model and the method addressing the spatiotemporal heterogeneities of PM$_{2.5}$-AOD relationships) is the best choice for a specific region or country. It depends on the spatial and temporal characteristics and data availability of the region. For examples, 1) advanced statistical models require much fewer ancillary variables than machine learning algorithms, as a result, the former might be better when not many auxiliary variables are available; 2) sub-region models should not be used to address the...
nonstationary issue for a small scale region such as a small province; 3) if
the sample values of PM$_{2.5}$ pollution in a certain place is limited to a
small range, it should be very careful to choose RF for lack of capability
of extrapolation, etc. Overall, such general rules and the advantages and
limitations of different models (such as those presented in Table 5) can
help to determine the direction of model development. However, it is
still hard to determine the detailed model structure or specific algorithm
based on the information mentioned above. One still needs to examine
several approaches and model strategies, compare their validation re-

tults (see Section 5 for validation methods), and then choose a final
approach with the best performance in the very case.

5. Model validations and predictions for the model years

5.1. Model cross-validation

Statistical indicators such as coefficient of determination ($R^2$, the
higher, the better), mean prediction error (MPE, the lower, the better),
root mean squared prediction error (RMSE, the lower, the better),
relative prediction error (RPE, the lower, the better) are usually used for
model evaluation. The formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

$$\text{MPE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (7)$$

$$\text{RPE} = \frac{\text{RMSE}}{\bar{y}} \times 100\% \quad (8)$$

where $y_i$ is ground measured PM$_{2.5}$, $\hat{y}_i$ is model predicted PM$_{2.5}$, $\bar{y}$ is the
average of ground PM$_{2.5}$.

After model fitting, we can calculate the above indicators based on
the whole modeling dataset to evaluate the model performance. That is,
we can apply the fitted model on the same modeling dataset to obtain
the PM$_{2.5}$ predictions and compare them with the ground measurements
to calculate the above indicators (we call them model fitting $R^2$, MPE,
RMSE, and RPE).

For indicators of model fitting, the predictions are made using the
fitted model on the same modeling dataset and thus cannot reflect po-
tential over-fitting, i.e., the model predictions could perform well on the
modeling dataset but badly outside the dataset in the same domain.
Cross validation (CV) is a method to test the potential over-fitting
problem. Typically, 10-fold CV (Rodriguez et al., 2010) is the most
commonly used method. For a standard 10-fold CV (i.e., sample-based
CV), all samples in the modeling dataset are randomly and equally
divided into ten subsets. For the first round, nine subsets are used to fit
the model, which is then applied to the remaining subset to obtain PM$_{2.5}$
predictions. This process is repeated 10 times (see Fig. 2), and then, all
predictions and ground measurements are combined to calculate the
statistical indicators (we call them CV $R^2$, MPE, RMSE, and RPE).

Here is the R code sample to calculate the standard 10-fold CV $R^2$ for
the single-stage OLS model as described in Eq. (1).

```r
> library(pls) # We need the function "cvsegments" in package "pls".
> CV_segments <- cvsegments(nrow(modeling_dataset), 10) # To create 10-fold CV Segments.
> CV_results <- data.frame() # To create an empty data frame to store the CV data.
> for (k in 1:10) {
  > OLS_testing <- modeling_dataset[,names(modeling_dataset) %in% CV_segments[k]$V],]
  > OLS_training <- modeling_dataset[(!(row.names(modeling_dataset) %in% CV_segments[k]$V),]
  > OLS_training_model <- lm(PM ~ AOD + WS, data = OLS_training)
  > OLS_cv_pred <- predict(OLS_training_model, data = OLS_testing)
  > CV_results <- cbind(CV_results, OLS_cv_pred)
} # To repeat 10 rounds.
> summary(lm(CV_results$OLS_cv_pred~CV_results$PM))$r.squared # To calculate the CV $R^2$.
```

To better evaluate model prediction accuracy beyond the grid cells
containing ground monitoring sites, some studies used spatial CV (includ-
ing site-based CV and administrative region-based CV) (Lee et al.,
2011; Li et al., 2015b; Li et al., 2017; Xiao et al., 2018). For site-based
CV, data from one ground monitor or 10% of monitors are used for
validation, and the rest sites are used for model fitting for each round of
validation. The process is repeated k (the number of monitors) or 10
rounds. For administration region-based CV, data from one adminis-
tration region (e.g., a city or a province) is used for validation, and the
rest administration regions are used for model fitting for each round of
validation. The process is repeated k (the number of administration re-

dions) rounds.

Sometimes we need to compare the models developed for different
regions. In this case, we must use relative performance indicators such as
$R^2$ and RPE. Absolute indicators (e.g., MPE and RMSE, etc.), which are
related to the absolute pollution levels, should not be used. For example,
Ma et al. (2016b) developed an LME model in eastern China and found a
CV $R^2$ of 0.72 and MPE of 12.53 $\mu g/m^3$. Hu et al. (2014b) developed a
two-stage model in the southeast US and found a CV $R^2$ of 0.67 and MPE
of 2.54 $\mu g/m^3$. We can compare their CV $R^2$ and conclude that the prior
study is slightly better than the latter. However, we cannot compare their
MPE because the pollution level in the US is much lower than that in
China. Absolute indicators can only be used to compare different
models in the same region and the same year (because for different years,
the pollution levels may change).

5.2. Model adjustment and predictions for model years

There is no unified standard on an acceptable CV $R^2$. It depends on
the quality and availability of the input data. In a region or country with
an extensive network of ground monitors and high-quality ancillary
data, a CV $R^2$ higher than 0.6 is expected. In regions with poor coverage
or a limited quantity of ground observations, a model with a CV $R^2$ of 0.6
or lower could be accepted. If the performance of the initial model is not
satisfactory, we can select different statistical or machine learning al-

orithms, use different combinations for a two-stage model, and adjust
the selection of ancillary variables. We then compare the statistical in-
dicators of model fitting and model CV for different models and select a
model with better CV results.

```
Fig. 2. The diagram of a standard 10-fold cross-validation.
```
Once we preliminarily accept a model, we can apply it to the prediction dataset in the model years to generate PM$_{2.5}$ predictions. For a single-stage OLS model:

$PM\_\text{pred} <\mathbf{-} \text{predict(OLS\_model, data = prediction\_dataset)}$

For a two-stage OLS model, we need to predict the residuals using the second stage model and then sum up the results from the first and second stage models:

$PM\_\text{resid\_pred} <\mathbf{-} \text{predict(OLS\_model\_2nd, data = prediction\_dataset)}$

$PM\_\text{pred\_final} <\mathbf{-} PM\_\text{pred} + PM\_\text{resid\_pred}$

5.3. Recommendations for the usage of CV in the determination of the final model

A standard sample-based CV may not be able to detect the existence of spatial over-fitting, where models perform very well in predicting at locations from training datasets but fail to predict into unknown space (Lv et al., 2016; Meyer et al., 2018). In this case, a spatial CV is recommended rather than a standard sample-based CV to evaluate model generalizability beyond the grid cells containing ground monitoring sites.

Besides, like other environmental or ecological data, the PM$_{2.5}$ concentrations show spatial, temporal, or hierarchical dependence structures, which cannot be captured by CV indicators (Roberts et al., 2017). In practice, the prediction mapping is also a crucial means of checking model performance since model-predicted PM concentration surfaces should be free of unrealistic artifacts. For example, Chen et al. (2018a) developed a satellite-based PM$_{2.5}$ prediction model in China and achieved relatively high CV $R^2$ at monthly and annual levels. However, non-smooth (discontinuous) PM$_{2.5}$ predictions along provincial borders are evident in the prediction maps. As an ambient air pollutant, the spatial distribution of PM$_{2.5}$ cannot possibly conform to administrative boundaries. Using the province as an interaction variable in model fitting is the main reason that causes this abnormality. Sub-region models with overlapping buffer zones can address this issue (see Section 4.6). Another example where the prediction map will have artificial stripes is if we directly incorporate coordinates (e.g., latitude and longitude) to account for spatial nonstationarity in a machine learning model (Zhan et al., 2017), which is also unrealistic. The Haversine approach (Wei et al., 2020a) or convolutional layer (Di et al., 2016; Hu et al., 2017) are potential solutions to address the spatial discontinuity of PM$_{2.5}$ predictions in a large-scale model.

Therefore, while the CV indicators such as $R^2$, RMSE, and MPE are important for model development and evaluation, we should not solely rely on them to determine a final model. In case we find that the prediction map has abnormal spatial patterns that are inconsistent with our common sense, no matter how high the CV $R^2$ is, we should adjust the model and validate the model again.

6. Model hindcasting and validation

For historical years before the ground monitoring network had been established, sufficient ground monitoring data may not be available to develop the statistical model. In this case, we can apply the model developed in the years with sufficient ground PM$_{2.5}$ data on the historical prediction datasets to hindcast historical estimates when satellite and other ancillary data are available. This is similar to the predictions in the model years. If our models are fitted on a daily basis, we must assume that the historical variation of day-to-day relationships are consistent with that of the model training period. Typically, there are two methods to assess historical estimates. First, we can collect historical PM$_{2.5}$ data from published papers or other sources. A limitation of this approach is that we can hardly obtain the daily data and can only evaluate the long-term average estimates such as monthly, seasonal, or annual mean concentrations. The second one is the by-year CV method. This method can evaluate the accuracy of historical estimates at the daily level and monthly or longer average levels. This requires that we have more than one year’s ground monitoring data. The study of Ma et al. (2016a) obtained daily ground PM$_{2.5}$ measurements in China for 2013 and the first half-year of 2014. They developed the PM$_{2.5}$-AOD model using 2013 data and then estimated daily PM$_{2.5}$ concentrations in the first half of 2014 and compared them with the ground measurements to validate the historical PM$_{2.5}$ estimations’ accuracy. Xue et al. (2019) estimated historical PM$_{2.5}$ from 2000 to 2012 using the model developed from data from 2013 to 2016 in China. To evaluate the accuracy of historical estimates, they used three years of data for model training and the remaining one year’s data for testing and then repeated 4 rounds. Since the evaluation is performed in the year that is not used for model fitting, the by-year CV can reflect the accuracy of historical estimates of the years that are not included in model fitting. Xiao et al. (2018) also used the same method in their study. Since the daily variations of the PM$_{2.5}$-AOD relationship are different among years, the historical daily estimates generally have lower accuracy than the daily predictions in the model years. The by-year CV $R^2$ values are usually around 0.5–0.6 at the daily level but are much higher at the monthly level ($>0.7$) (Ma et al., 2016a; Xiao et al., 2018; Xue et al., 2019).

The phenomenon of relatively low accuracy of historical daily estimates can be summarized by the term “concept drift” in the field of data stream mining. The issue of concept drift refers to the inconsistent patterns between known data and unknown data because of hidden variables that are not incorporated in models (Liu et al., 2017a). For example, if we don’t include anthropogenic emissions in PM$_{2.5}$ models, the relationship between PM$_{2.5}$ and other predictors we rely on for prediction may change with changes in the anthropogenic emissions over time, leading to poor performance.

If the purpose of hindcasting is to examine long-term trends, we can also develop a model directly at a coarser temporal resolution (e.g., at a monthly level or longer, etc.) (Huang et al., 2018; Liang et al., 2020). For example, Liang et al. (2020) developed a machine learning model directly at the monthly level in China and then applied it to predict the historical monthly mean PM$_{2.5}$ concentrations from 2000 to 2016.

7. Summary and future directions

Exposure estimates derived from satellite remote sensing data have been successfully applied in many studies, including environmental epidemiology (Crouse et al., 2012; Madrigano et al., 2013; Wang et al., 2018a), health impact assessments (Chowdhury and Dey, 2016; Liu et al., 2017b; Wang et al., 2018b), and social-economic impact studies (Chen and Chen, 2017; Chen and Jin, 2019; Yang and Zhang, 2018). These studies have provided valuable information for regional and national air pollution control and management. Current studies were conducted in a handful of countries that have established a ground monitoring network. There are still many countries without air quality monitoring networks. Once their ground monitoring networks are established, this emerging technology can be promoted to these countries to obtain highly accurate, large-scale, and long-term PM$_{2.5}$ exposure data, which could promote the health effects studies of PM$_{2.5}$ in these regions and countries. The statistical approach can also be applied in satellite remote sensing of other air pollutants, such as PM$_{2.5}$ chemical composition (Geng et al., 2020; Meng et al., 2018), PM$_{1}$ (Chen et al., 2018a; Wang et al., 2019), NO$_2$ (Di et al., 2020; Qin et al., 2020; Qin et al., 2017; Zhan et al., 2018), SO$_2$ (Li et al., 2019; Zhang et al., 2019), and CO (Liu et al., 2019), etc. Satellite remote sensing of air pollution has a broad application prospect.

Overall, this article describes the detailed workflow for statistical models of satellite remote sensing of PM$_{2.5}$. Some major recommendations are also provided, especially for large-scale and long-term models.

(1) Meteorological, land-use, and other covariates are strongly recommended to be incorporated in satellite PM$_{2.5}$ statistical models.
to improve the model accuracy. When choosing a covariate, the following issues should be considered: 1) the variable should be selected based on the physical understanding; 2) multicollinearity analysis should be conducted; 3) significance or importance of variables should be tested; 4) we can assess the model performance by adding/removing variables; and 5) we can assess the performance of models using variables from different datasets.

(2) PM$_{2.5}$–AOD relationships have strong spatiotemporal heterogeneities, which must be addressed for a large-scale and long-term model. The methods include 1) using spatial and temporal models; 2) incorporating spatial and temporal terms; 3) developing sub-region/sub-period models; and 4) using a hybrid approach integrating two or more above approaches.

(3) According to the NFL theorem, it is hard to say which approach (including the model selection and the selection of the method addressing the spatiotemporal heterogeneities of PM$_{2.5}$–AOD relationships) is the best choice for a specific region or country. It depends on the spatial and temporal characteristics and data availability of the region. We should examine several approaches, compare their model validation results, and then choose an approach with the best performance.

(4) To better evaluate the model generalization ability beyond the grid cells containing ground monitoring sites, a spatial CV is recommended rather than a standard sample-based CV. To assess the accuracy of historical hindcasting data beyond the years of developing the models, we could use an external evaluation or a by-year CV.

(5) Although model CV indicators such as $R^2$, RMSE, and MPE are important for model development and evaluation, we should not solely rely on them to determine a final model. The prediction maps can serve as another means of checking model performance. In case we find that the prediction map has abnormal spatial patterns that are inconsistent with our common sense, no matter how high the CV $R^2$ is, we should adjust the model and validate the model again.

Given the rapid evolution of this field, new algorithms and technical considerations frequently emerge and may not have been included at the time of this review. However, the detailed and basic framework and model considerations summarized in this paper will help the researchers to better learn, understand new technologies, compare and select the best applicable model among various algorithms, and will be beneficial for model improvement and optimization in future studies. This paper will provide useful information for remote sensing of air pollution, not only for PM$_{2.5}$ but also for gas pollutants such as NO$_2$, SO$_2$, and CO.

Currently, the application of satellite remote sensing technology in air pollution exposure studies is still in its early stage; more efforts are needed to improve the model accuracy of estimated data in the future.

(1) Due to the NFL theorem, analyzing the sensitivity of satellite-based PM$_{2.5}$ statistical models to various situations is needed. In a recent study, Geng et al. (2018) studied the sensitivity of the two-stage Bayesian ensemble model to various data-input scenarios. Investigating sensitivities of more statistical models under more scenarios for satellite remote sensing of PM$_{2.5}$ is a future direction.

(2) Continued interdisciplinary collaboration is an important way to promote the application of such technology. For example, the strategies of detecting and adapting to the issues of concept drift from the discipline of artificial intelligence, such as the model ensemble for recurring drift, can help develop new model architectures (Liu et al., 2019), which can help to improve the accuracy of historical hindcasting data.

(3) New satellite instruments, such as TEMPO (Tropospheric Emissions: Monitoring Pollution) (Zoogman et al., 2017) and MAIA (Multi-Angle Imager for Aerosols) (Liu et al., 2017c), provide new satellite capabilities for air pollution exposure modeling. Utilizing new satellite data to obtain more accurate air pollution exposure data is also a future direction.

Declaration of Competing Interest

The authors declare that they have no actual or potential competing financial interests.

Acknowledgements

The work of Z. Ma and J. Bi was supported by the National Natural Science Foundation of China (71761147002 and 71921003). The work of S. Dewy was supported by the Central Pollution Control Board, India, DST-FIST grant (SR/FST/ESI-016/2014), DBT grant under Indo-UK program (BT/IN/UK/APH/41/KB/2016-17) and Institute Chair at IIT Delhi. The work of Y. Liu was supported by the NASA Applied Sciences Program (Grant # 8N0SSC19KO119) and the MAIA science team at the JPL, California Institute of Technology (Subcontract #1588347). The content is solely the responsibility of the authors and does not necessarily represent the official views of NASA.

References


Liu, Z., Ma et al. Remote Sensing of Environment 269 (2022) 112827