Article

Optimal Inversion of Conversion Parameters from Satellite AOD to Ground Aerosol Extinction Coefficient Using Automatic Differentiation

Lianfa Li ^{1,2}

¹ State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Datun Road, Beijing 100101, China; lilf@reis.ac.cn

² University of Chinese Academy of Sciences, Beijing 100049, China

Received: 26 December 2019; Accepted: 30 January 2020; Published: 4 February 2020

Abstract: Satellite aerosol optical depth (AOD) plays an important role for high spatiotemporal-resolution estimation of fine particulate matter with diameters $\leq 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$). However, the MODIS sensors aboard the Terra and Aqua satellites mainly measure column (integrated) AOD using the aerosol (extinction) coefficient integrated over all altitudes in the atmosphere, and column AOD is less related to $\text{PM}_{2.5}$ than low-level or ground-based aerosol (extinction) coefficient (GAC). With recent development of automatic differentiation (AD) that has been widely applied in deep learning, a method using AD to find optimal solution of conversion parameters from column AOD to the simulated GAC is presented. Based on the computational graph, AD has considerably improved the efficiency in applying gradient descent to find the optimal solution for complex problems involving multiple parameters and spatiotemporal factors. In a case study of the Jing-Jin-Ji region of China for the estimation of $\text{PM}_{2.5}$ in 2015 using the Multiangle Implementation of Atmospheric Correction AOD, the optimal solution of the conversion parameters was obtained using AD and the loss function of mean square error. This solution fairly modestly improved the Pearson's correlation between simulated GAC and $\text{PM}_{2.5}$ up to 0.58 (test R^2 : 0.33), in comparison with three existing methods. In the downstream validation, the simulated GACs were used to reliably estimate $\text{PM}_{2.5}$, considerably improving test R^2 up to 0.90 and achieving consistent match for GAC and $\text{PM}_{2.5}$ in their spatial distribution and seasonal variations. With the availability of the AD tool, the proposed method can be generalized to the inversion of other similar conversion parameters in remote sensing.

Keywords: parameter inversion; aerosol optical depth; PBLH; ground-based AOD; $\text{PM}_{2.5}$; automatic differentiation

1. Introduction

As one of the criteria air pollutants [1], particular matter (PM) refers to a mixture of solid and liquid particles suspended in the air, including tiny inhalable particles that may penetrate the thoracic region of the respiratory system. Studies [2,3] show PMs with a diameter of less than $10 \mu\text{m}$ (PM_{10}) and with a diameter of less than $2.5 \mu\text{m}$ ($\text{PM}_{2.5}$) are closely associated with short- (e.g., asthma, respiratory) and long-term (e.g., lung cancer, cardiopulmonary mortality) health effects. Recently, several studies [4–6] have also shown a significant association between $\text{PM}_{2.5}$ and diabetes/neurological disorders. Although PM can be emitted directly (primary PM) or formed in the atmosphere (secondary PM), owing to increasing fleets of vehicles and strict emission regulations for industry and coal, traffic emissions have recently contributed more to $\text{PM}_{2.5}$ concentrations, which might also result in high exposure for the population since more persons live along major traffic routes. Given acute and long-term adverse health effects of $\text{PM}_{2.5}$, its monitoring and accurate estimation are important for the studies of its health

effects and control. However, although exhibiting an increasing trend, the number of PM_{2.5} monitoring stations is still limited worldwide, including in China [7].

As one primary component of the aerosol mass, spatiotemporal variability of the PM_{2.5} concentration ($\mu\text{g}/\text{m}^3$) is affected by multiple factors, including various anthropogenic and natural emission sources [8–10], meteorology likely involved in complex atmospheric chemical processes for generation of secondary PM_{2.5} [11,12], local elevation, and terrain [12,13]. Complex atmospheric chemical processes involving these factors and their interactions result in high uncertainty that presents a challenge in the estimation of PM_{2.5}, particularly before the launch of the Moderate Resolution Imaging Spectroradiometer (MODIS) in 1999, when no such strong predictors as satellite aerosol optical depth (AOD) were available. Since the MODIS sensors of the Terra and Aqua satellites started collecting data in 1999 and 2002, respectively, they have been providing measurements of AOD. AOD measures the vertically integrated extinction of the solar beam by dust and haze using sun photometers. For example, an AOD of 0.01 indicates an extremely clean atmosphere, but an AOD of 0.4 suggests very hazy conditions. As one aerosol property, AOD has been widely used as a proxy for the number of particles within a vertical column and as a primary predictor of PM_{2.5} [14]. However, AOD is the vertical integral of aerosol extinction coefficient at all altitudes along the orientation [15], and the aerosol vertical profile presents an uneven and probably exponential distribution [15,16]. Furthermore, the seasonally varying mixing vertical height has an important influence on the association between AOD and ground-level aerosol extinction coefficient or PM_{2.5} [17]. For northern China, column AOD has the maximum value in summer (higher than in winter), but low-level AOD is highest during winter when more aerosols concentrate close to the ground surface, contributing to high ground-level aerosol loading (e.g., high PM_{2.5} concentration) due to lower atmospheric boundary layer in winter than in summer [18,19]. This results in inconsistency in the seasonal variations between column AOD and ground-level aerosol extinction coefficient, which, in turn, affects the use of AOD as a predictor of primary interest for estimation of PM_{2.5} concentration. Since the satellite data gathered by the passive sensors of MODIS have not provided the vertical profile of aerosols, appropriate conversion of satellite AOD to low- or ground-level aerosol extinction coefficient can considerably improve use of satellite AOD for estimation of PM_{2.5} since the low-level aerosol coefficient has more consistent seasonal variations with PM_{2.5} than column integrated AOD.

The satellite AOD-PM_{2.5} association is affected by multiple factors including complex atmospheric processes, emission sources and terrains, etc. and the interactions between them, and varies with different regions [20]. As mentioned above, since AOD characterizes the extinction of the whole atmosphere, aerosol vertical profile is an important factor for conversion of column (satellite) AOD to ground/surface aerosol extinction coefficient. In addition, surface aerosol mass consists of particles of different sizes and relative humidity also affects the conversion to dry particle mass. For the conversion of satellite AOD to ground extinction coefficient, there have been two types of methods [20,21]: empirical relationship between satellite AOD and PM_{2.5}, and using available vertical profile information. For empirical conversion, planetary boundary layer height has been used as the divisor to adjust satellite AOD given AOD obtained through the integration of aerosol extinction coefficient by boundary mixing height in most studies [15,21–25]. Several studies [20,26] employed chemical transport models (CTM) such as the Community Multiscale Air Quality (CMAQ) to simulate the relationship between modeled PM_{2.5} and AOD and then used such a relationship as a linear converter of satellite AOD. However, although most CTMs can provide the simulated vertical distribution of aerosols at varied spatial and temporal scales [27], they just have coarse spatial resolution and may considerably underestimate aerosol mass concentrations [28,29], and may be not available for some study regions and periods. With relative humidity, the hygroscopic growth factor of aerosols was considered in the conversion [15,22,25,30]. Furthermore, Li et al. [23] also conducted particle correction of AOD for PM_{2.5} and Zeng et al. [25] also used visibility for vertical correction. For available vertical profile information in conversion, cloud-Aerosol LIDAR with Orthogonal Polarization (CALIOP) can provide the extinction profile data for conversion [31] with limited spatiotemporal coverage [17] and coarse

spatial resolution of 2° (latitude) $\times$ 5° (longitude) [27]. However, for the purpose of applications, a formula of conversion from satellite AOD of MODIS to ground extinction coefficient is more practical than using LIDAR sensors to obtain the vertical profile that is not always available. Wang et al. (2010) [15] suggested a simplified formula for the conversion from column total AOD to ground aerosol extinction coefficient that substantially improved $PM_{2.5}$ estimation.

In this paper, in comparison with the existing methods, a more flexible conversion method is proposed. Considering multiple complex atmospheric and environmental factors (e.g., emission sources [8,10,32], meteorology [11,12], and elevation [12,13]) involved in the aerosol vertical profile, this method introduces a scaling factor (slope) and a shift factor (intercept) for both planetary boundary layer height (PBLH) and relative humidity (RH) to capture the influence of potential other confounders (e.g., atmospheric chemical processes and other meteorological factors) or random factors based on the simplified conversion formula in [15]. For the optimization of multiple parameters in the proposed conversion formula, automatic differentiation (AD) was used in gradient descent to improve learning efficiency. Automatic differentiation is essential in deep learning, and with the availability of deep learning software such as Tensorflow, PyTorch, or Caffe, it can provide an effective way of finding an optimal solution in many practical applications [33]. Additionally, the linear and polynomial scaling and shift factors were also introduced into the simulated ground aerosol coefficient as proxy to the aerosol extinction coefficient to construct a suitable loss function of mean square error (MSE) for training of the models. With the 2015 measurement data of the Jing-Jin-Ji metropolitan study area, the proposed method was used to reliably convert column (satellite) AOD to ground aerosol extinction coefficient and the results were compared with the empirical conversion methods.

2. Materials and Methods

2.1. Study Region

Covering most of the Jing-Jing-Ji metropolitan region of northern China, the study area (Figure 1) is located between the latitudes of $38^\circ 05'N$ and $41^\circ 04'N$ and the longitudes of $115^\circ 14'E$ and $118^\circ 54'E$. The Jing-Jin-Ji is also known as Beijing-Tianjing-Hebei as the biggest urbanized megalopolis region in northern China. It is the area surrounding the municipalities of Beijing and Tanjing, and 13 smaller cities along the coast of Bohai Sea. The study area has an area of approximately 64,000 km^2 with a total population of approximately 94 million. This area is the core of smog in northern China, one of the most polluted areas in China. It has multiple emission sources including industrial pollutions, bulk coal combustion, and motor vehicle exhausts [34]. This study area has a lower elevation than that of surrounding high Mount Taihang-Yanshan, which can result in heat island effects and hence concentrations of air pollutants with local not-conductive-to-disperse meteorology of high temperature and low air pressure [35,36]. Compared with the other seasons, winter in the study area has much higher pollution with up to several times the emissions of $PM_{2.5}$, organic carbon, black carbon and other pollutants due to heating, and adverse meteorology of lower wind speed and higher atmospheric stability.

2.2. Dataset

2.2.1. Satellite AOD

The Multiangle Implementation of Atmospheric Correction (MAIAC) has been recently used to retrieve satellite AOD from the MODIS sensors aboard the Terra and Aqua satellites at a high spatial (1 km) and temporal (daily) resolution with better atmospheric correction over both dark and moderately bright surfaces than the previous Dark Target and Deep Blue retrieval algorithms. In this study, the 2015 MAIAC AOD images covering the study area were collected from the website (<https://lpdaac.usgs.gov/news/release-of-modis-version-6-maiac-data-products>) of the U.S. Geological Survey. The 1954 Beijing coordinate system was used as the target kilometer projection in this study

and the projection transformation to the target system was conducted for these images. The quality assurance (QA) data provided in the MAIAC dataset were used to filter out invalid pixels due to closeness to cloud or snow or due to high surface reflectance. A non-linear generalized additive model (GAM) was used to fuse Aqua and Terra daily images to increase spatial coverage of available AOD when one AOD was missing and the other was available. Specifically, the fusion was conducted using the averages of both Aqua and Terra (observed/estimated) AODs when both observed AODs were available, or when one observed AOD was available and the other AOD is missing. For the latter, the missing AOD was estimated by the GAM trained using the samples of both observed AOD available. Similar methods were used in [37,38].

Figure 1. Study region of Jing-Jin-Ji metropolitan region.

2.2.2. PM_{2.5} Measurement

The hourly PM_{2.5} measurement (unit: $\mu\text{g}/\text{m}^3$) data covering the study area were from 102 monitoring stations operated by China's governmental agency (<http://www.pm25.in>) (see Figure 1 for spatial distribution of these stations and their 2015 mean PM_{2.5} concentrations). The monitoring data of PM_{2.5} were measured using a tapered element oscillating microbalance (TEOM) with the Thermo Fisher 1405F [39]. Specifically, the ambient air was first pumped in by the TEOM and heated up to 40 °C to remove the influence of water vapor. Then, the particles with a smaller size than 2.5 μm were selected using a special filter and weighted. Finally, the dry mass of PM_{2.5} within a unit volume of ambient air was acquired and recorded. Therefore, relative humidity played an important role in the conversion under moisture conditions to dry particle mass. Using a 75% completeness criterion, the hourly measurements were averaged into daily PM_{2.5} concentration.

2.2.3. PBLH

The 2015 PBLH (unit: meter) data were gathered from the reanalysis data source of MERRA2 (<https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2>). The reanalysis data have a coarse spatial resolution of 0.25° (latitude) $\times$ 0.3125° (longitude) and a high temporal resolution of 3 h. Daily means were

obtained by averaging over 3 h reanalysis data. Re-sampling was conducted to convert the coarse spatial-resolution PBLH data to 1-km spatial resolution data to match the spatial resolution of MAIAC AOD.

2.2.4. Relative Humidity

The measurement data of RH (unit: %) were gathered from the dataset of daily ground observation values in China Mainland (Version 3.0) of the China Meteorological Data Service (<http://data.cma.cn>). In total, the measurement data came from 824 national meteorological stations, and spatial interpolation was conducted using these measurement data with other factors. The method of a residual deep network was used to obtain the daily grid images with a spatial resolution of 1 km (cross-validation R-squared (R^2) = 0.85 for RH) [40]. The study area's grids of RH were extracted from the interpolated dataset.

2.3. Simulation of Ground Aerosol Extinction Coefficient

As a metric of total column integrated aerosol extinction, satellite AOD is determined by the aerosol scale height and the corresponding aerosol extinction coefficient (Figure 2a):

$$\tau_a = \int_0^{z_{TOA}} k_a(\lambda, z) dz \quad (1)$$

where τ_a is column AOD, $k_a(\lambda, z)$ is the aerosol extinction coefficient (k_a) at the altitude of z and the spectral wavelength λ (for satellite AOD, $\lambda = 0.55 \mu\text{m}$), TOA represents the top of the atmosphere [41].

Figure 2. Relationship between satellite aerosol optical depth (AOD), ground aerosol extinction coefficient, and PBLH (a), and vertical exponential distribution (b) of aerosol extinction coefficient.

Although the vertical profile of aerosol extinction coefficient does not directly follow a simple exponential decay with altitude, the scale height simulated by an exponential distribution, although not perfect, is a good approximation for the aerosol vertical extent (Figure 2b) with a limited degree of uncertainty and one feature of aerosol vertical mixing in the low-level troposphere [42]:

$$k_a(\lambda, z) \approx k_{a,0}(\lambda) \cdot \exp(-z/H_A) \quad (2)$$

where $k_{a,0}(\lambda)$ represents the ground-level aerosol extinction coefficient at the wavelength of λ , and H_A is the scale height of aerosol, approximately represented by PBLH [43].

Introducing Equation (2) into Equation (1), the following formula can be obtained:

$$\begin{aligned}
\tau_a &= -k_{a,0}(\lambda) \cdot H_A \cdot \exp\left(-\frac{z}{H_A}\right) \Bigg|_0^{z_{TOA}} \\
&= k_{a,0}(\lambda) \cdot H_A \cdot \left(1 - \exp\left(-\frac{z_{TOA}}{H_A}\right)\right) \\
&= k_{a,0}(\lambda) \cdot (H_A + it)
\end{aligned} \tag{3}$$

where it stands for the intercept, $-H_A \cdot \exp\left(-\frac{z_{TOA}}{H_A}\right)$, which is often negligible due to a high value of z_{TOA} in many studies [15,17].

In addition, studies [44,45] show that the association between k_a and PM concentration varies with the chemical components of particles and the RH of the ambient air. Thus, the factor of RH was introduced into the adjustment of ground aerosol extinction coefficient in Wang et al. (2010) [15]:

$$k_{a,Dry}(\lambda) = k_{a,0}(\lambda) / f(RH) \tag{4}$$

where $k_{a,Dry}(\lambda)$ represents the “dry mass” of the particles, closely associated with $PM_{2.5}$ concentration. $f(RH)$ stands for the factor of relative humidity:

$$f(RH) = (1 - RH/100)^{-g} \tag{5}$$

where g is an empirical fit coefficient.

Introducing Equations (3) and (5) into Equation (4), the following formula can be obtained:

$$\begin{aligned}
k_{a,Dry}(\lambda) &= \frac{\tau_a}{(H_A + it) \cdot f(RH)} \\
&= \frac{\tau_a \cdot (1 - RH/100)^g}{H_A + it}
\end{aligned} \tag{6}$$

To have flexibility in the solution of $k_{a,Dry}(\lambda)$, a shift factor and scaling factor were introduced to PBLH and RH, respectively, in Equation (6) to account for the influence of the uncertainty arising from the measurement error, or other confounders (e.g., the intercept it in Equation (3)). Thus, the following formula is obtained:

$$k_{a,Dry}(\lambda) = \frac{\tau_a \cdot [s_{RH} \cdot (1 - RH/100)^g + i_{RH}]}{s_{H_A} \cdot H_A + i_{H_A}} \tag{7}$$

where s_{RH} and i_{RH} , respectively, represent a scaling factor and a shift factor for RH, and s_{H_A} and i_{H_A} , respectively, represent a scaling factor and a shift factor (similar to it in Equation (6)) for PBLH.

2.4. Solution by Automatic Differentiation

To solve five parameters $\theta(g, s_{RH}, i_{RH}, s_{H_A}, i_{H_A})$ in the flexible model of Equation (7), it is difficult to use a traditional method such as the analytical or least squares solution. Instead, machine learning of gradient descent may be used to find a locally optimal solution. Automatic differentiation [46], also called algorithmic differentiation, provides an efficient method with which to implement gradient descent to find an optimal solution for the complex problem of multiple parameters to be solved. For differentiation used in gradient descent, a traditional hand-crafted method needs non-trivial formulas and laborious work that is impractical for complex problems; numerical differentiation may have the issue of plague of truncation and round-off errors; symbolic differentiation predefines set of operations such as chain-rule for products but it may have a problem of expression swell that refers to a much larger representation of the derivative as opposed to that of the original function [47]. Different from these previous methods, AD is not limited by the nature of function and can go across complicated control flows, and makes gradient descent easy to apply in many practical optimization problems. It consists of a series of operations in a computational graph that is based on dynamic programming. A computational graph is a directed graph of basic operations (e.g., add, multiply, divide, trigonometric function) and the final value can be obtained by propagation through its corresponding path in the graph. Specifically, AD first generates a computational graph of domain variables and

then continues to replace the domain of the variables to incorporate derivatives per the chain rule of differential calculus. In essence, AD applies symbolic differentiation at the elementary level and keeps intermediate numerical results with the evaluation of the main function. By the chain rule, the errors can be backpropagated through gradients to update the parameters until convergence is attained or an optimal solution is obtained.

The efficiency of automatic differentiation and convenience of its use have led to considerable developments of deep learning and are essential for searching for the optimal solution for general problems [33]. In this study, Tensorflow (<https://www.tensorflow.org/>) (Version 1.8) was used as the AD tool to solve the parameters in Equation (7).

For a typical optimization problem using gradient descent, a loss function is first needed to quantify how well the trained model fits the training input X . Then, AD is used in gradient descent to solve the minimum of the following loss function:

$$\operatorname{argmin} L(X, Y, \theta) \quad (8)$$

where X denotes the input data including AOD (τ_a), PBLH (H_A), and relative humidity (RH), Y represents the target variable of ground aerosol extinction coefficient to be converted to $PM_{2.5}$, and θ refers to the parameters to be optimized and estimated in learning.

The final target of Equation (8) is to improve the correlation between ground aerosol extinction coefficient and $PM_{2.5}$. Therefore, the following loss function may be used to make the correlation differentiable:

$$L = 1 - \left(c(k_{a,Dry}(\lambda), PM_{2.5}) \right)^2 \quad (9)$$

where c denotes the Pearson's correlation between $k_{a,Dry}(\lambda)$ in Equation (7) and ground monitoring $PM_{2.5}$.

In order to achieve efficiency in learning, a suitable loss function should be selected to avoid the ill-conditioned issue and have an efficient convergence in optimization. However, the correlation coefficient cannot be used in a loss function due to poor convergence and the ill-conditioned issue; the sensitivity test also demonstrated this. For regression, as the most common loss function, the MSE has good properties of efficient convergence and no outlier predictions with huge errors [33] and can be used instead. Since the ground aerosol extinction coefficient, $k_{a,Dry}(\lambda)$ is finally used to estimate $PM_{2.5}$ in the downstream procedure, we can define a conversion function from it to $PM_{2.5}$ using the following non-linear polynomial formula:

$$\hat{y} = s_{k_{a,Dry}}^2 k_{a,Dry}^2(\lambda) + s_{k_{a,Dry}} k_{a,Dry}(\lambda) + i_{k_{a,Dry}} \quad (10)$$

where $\hat{y}$ denotes $PM_{2.5}$ concentration to be estimated, $k_{a,Dry}^2(\lambda)$ is the square term of $k_{a,Dry}(\lambda)$, $s_{k_{a,Dry}}^2$ is its scaling (slope) factor, and $s_{k_{a,Dry}}$ and $i_{k_{a,Dry}}$ represent the scaling (slope) and shift (intercept) factors, respectively, for $k_{a,Dry}(\lambda)$.

With Equation (10) defined as the estimated $PM_{2.5}$, we can get the MSE loss function:

$$L = \operatorname{MSE}(\hat{\mathbf{y}}, \mathbf{y}) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (11)$$

where $\mathbf{y} = \{y_i\}$ is the vector of the observed $PM_{2.5}$, and $\hat{\mathbf{y}} = \{\hat{y}_i\}$ is the vector of the estimated $PM_{2.5}$ by $k_{a,Dry}(\lambda)$ and the other covariates (PBLH and relative humidity) from Equation (10). If the minimum loss (MSE) is attained, the correlation of ground aerosol extinction coefficient with $PM_{2.5}$ is reasonably assumed to be improved correspondingly.

In addition, gradient descent generally finds a local optimal solution. Owing to the introduction of scaling and shift factors, such a solution cannot make the estimated $k_{a,Dry}(\lambda)$ directly equal to the ground-level aerosol extinction coefficient, but the solution can be regarded as a proxy variable to the

ground true low-level aerosol extinction coefficient. The proxy variable can be used as an alternative to improve the estimation of $PM_{2.5}$. For simplicity, the proxy variable is named the ground aerosol coefficient (GAC) throughout this paper.

The computational graph of AD represents the model to be trained with the nodes used, including input variables, target variable, the parameters to be solved, and interconnections (forward and backpropagation) between them (Figure 3). In the computational graph, a composite computation may consist of multiple basic and/or composite computations. For gradient descent, the target loss function (Equation (11)) was a starting point for decomposition and complex computations were progressively decomposed until simple and basic operations were obtained. Figure 3 shows the composite and basic nodes in the decomposition, and the connections between them with a line arrow indicating the forward or reverse mode of a gradient operation (black line: forward mode; red dot line: reverse mode). Since there were multiple parameters to be solved, reverse mode or backpropagation was the best method for efficiency in computing [46]. Table 1 shows each node's name, definition, related formula, derivatives in backpropagation, and derivatives for the parameters to be solved, if any. For backpropagation, the starting point was the last row of the loss target function (Equation (11); Table 1), Column 4 shows the derivatives of the previous node against the current node, and Column 5 shows the derivatives against a target parameter, if any. In the computational graph, the target loss function (Equation (11)), simulated $PM_{2.5}$ (Equation (10)), and the composite formula for GAC (Equation (7)) were involved (Table 1).

Figure 3. Computational graph of automatic differentiation to solve Equation (11).

As a local optimization method, gradient descent needs to select the optimizer for training. In this study, the adaptive Ada optimizer was used with a mini-batch size of 1024 to obtain an optimal solution. As an advanced optimizer, Ada [48] is an optimization algorithm with adaptive learning rate that is dynamically adjusted based on the squared gradient and momentum.

Table 1. Nodes and their derivatives in the computational graph for conversion from satellite AOD to ground aerosol coefficient (GAC).

Name	Definition	Formula	Derivative for Back Propagation	Derivative of the Parameters to Be Solved
U_1	$1 - RH/100$	$1 - RH/100$	$U'_2(\partial U_2/\partial U_1) = -2s_{RH} \cdot \tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot g \cdot U_1^{g-1}/U_7$	$U'_2(\partial U_2/\partial g) = -2s_{RH} \cdot \tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot \ln(g) \cdot U_1^g/U_7$
U_2	U_1^g	$(1 - RH/100)^g$	$U'_3(\partial U_3/\partial U_2) = -2s_{RH} \cdot \tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7$	$U'_3(\partial U_3/\partial s_{RH}) = -2\tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7 \cdot U_2$
U_3	$S_{RH}U_2$	$S_{RH}(1 - RH/100)^g$	$U'_4(\partial U_4/\partial U_3) = -2\tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7$	$U'_4(\partial U_4/\partial i_{RH}) = -2\tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7$
U_4	$U_3 + i_{RH}$	$S_{RH}(1 - RH/100)^g + i_{RH}$	$U'_5(\partial U_5/\partial U_4) = -2\tau_a \cdot U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7$	
U_5	$\tau_a \cdot U_4$	$\tau_a(S_{RH}(1 - RH/100)^g + i_{RH})$	$U'_8(\partial U_8/\partial U_5) = -2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)/U_7$	$U'_6(\partial U_6/\partial S_{HA}) = 2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot s_{k_{\alpha,Dry}} \cdot (U_5/U_7^2) \cdot H_A$
U_6	$s_{HA} \cdot H_A$	$s_{HA} \cdot H_A$	$U'_7(\partial U_7/\partial U_6) = 2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot s_{k_{\alpha,Dry}} \cdot (U_5/U_7^2)$	$U'_7(\partial U_7/\partial i_{HA}) = 2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot s_{k_{\alpha,Dry}} \cdot (U_5/U_7^2)$
U_7	$U_6 + i_{HA}$	$s_{HA} \cdot H_A + i_{HA}$	$U'_8(\partial U_8/\partial U_7) = 2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2) \cdot s_{k_{\alpha,Dry}} \cdot (U_5/U_7^2)$	
U_8	U_5/U_7	$k_{a,Dry}(\lambda) = \frac{\tau_a \cdot [s_{RH} \cdot (1 - RH/100)^g + i_{RH}]}{s_{HA} \cdot H_A + i_{HA}}$ [Equation (7)]	$U'_{13}(\partial U_{13}/\partial U_8) = U'_{13} \left(\frac{\partial U_{10}/\partial U_8 + \frac{\partial U_{10}}{\partial U_{12}} \cdot \frac{\partial U_{12}}{\partial U_8}}{\frac{\partial U_{10}}{\partial U_{12}} \cdot \frac{\partial U_{12}}{\partial U_{11}} \cdot \frac{\partial U_{11}}{\partial U_8}} \right)$ $= U'_{13} \left(\frac{\partial U_{10}/\partial U_9 \cdot \partial U_9/\partial U_8 + \frac{\partial U_{10}}{\partial U_{12}} \cdot \frac{\partial U_{12}}{\partial U_{11}} \cdot \frac{\partial U_{11}}{\partial U_8}}{\frac{\partial U_{10}}{\partial U_{12}} \cdot \frac{\partial U_{12}}{\partial U_{11}} \cdot \frac{\partial U_{11}}{\partial U_8}} \right)$ $= -2U_{14} \cdot (s_{k_{\alpha,Dry}} + 2U_8 \cdot s_{k_{\alpha,Dry}}^2)$	$U'_9(\partial U_9/\partial s_{k_{a,Dry}}) = -2U_{14} \cdot U_8$
U_9	$s_{k_{a,Dry}} U_8$	$s_{k_{a,Dry}} k_{a,Dry}$	$U'_{10}(\partial U_{10}/\partial U_9) = -2U_{14}$	$U'_{10}(\partial U_{10}/\partial i_{k_{a,Dry}}) = -2U_{14}$
U_{10}	$U_9 + i_{k_{a,Dry}}$	$s_{k_{a,Dry}} k_{a,Dry} + i_{k_{a,Dry}}$	$U'_{13}(\partial U_{13}/\partial U_{10}) = -2U_{14}$	
U_{11}	U_8^2	$k_{a,Dry}^2$	$U'_{12}(\partial U_{12}/\partial U_{11}) = -2U_{14} \cdot s_{k_{\alpha,Dry}}^2$	$U'_{12}(\partial U_{12}/\partial s_{k_{\alpha,Dry}}^2) = -2U_{14} \cdot U_{11}$
U_{12}	$U_{11} \cdot s_{k_{\alpha,Dry}}^2$	$k_{a,Dry}^2 \cdot s_{k_{\alpha,Dry}}^2$	$U'_{13}(\partial U_{13}/\partial U_{12}) = -2U_{14}$	
U_{13}	$U_{10} + U_{12}$	$\hat{y} = s_{k_{\alpha,Dry}}^2 \cdot k_{a,Dry}^2 + s_{k_{a,Dry}} \cdot k_{a,Dry} + i_{k_{a,Dry}}$ [Equation (10)]	$U'_{14}(\partial U_{14}/\partial U_{13}) = -2U_{14}$	
U_{14}	$y - U_{13}$	$y - \hat{y}$	$U'_{15}(\partial U_{15}/\partial U_{14}) = 2U_{14}$	
U_{15}	U_{14}^2	$(y - \hat{y})^2$	$Loss'(\partial Loss/\partial U_{15}) = 1$	
Loss	$\sum U_{15}$	$\sum (y - \hat{y})^2$ [Equation (11)]	$\partial Loss/\partial Loss = 1$	

2.5. Validation and Comparison

A total of 70% of the 53,040 data samples of available AOD and PM_{2.5} measurements for the study area were randomly selected to train the model, and the remaining 30% of the data samples were used to test the trained models. This procedure was repeated five times with different training and testing samples selected, and the final result was obtained by averaging over the results of the five procedures.

The proxy variables for ground aerosol extinction coefficient generated by three methods were generated and compared. These variables include original AOD and GACs simulated just using PBLH, using the empirical formula proposed in [15], and using the AD method proposed in this paper. The test Pearson's correlation of AOD or simulated GAC with measured PM_{2.5}, R², and root mean square error (RMSE) in the linear regressions of PM_{2.5} estimation using these simulated values was reported and compared. The grid surfaces of AOD and simulated GAC were also presented and compared.

3. Results

3.1. Description Statistics

For MAIAC AOD of the data samples, the 2015 mean was 0.79, higher in the South than in the North (0.92 vs. 0.69), and higher in summer than in winter (1.05 vs. 0.62). The Pearson's correlation with PM_{2.5} is 0.39. For PM_{2.5}, the 2015 mean concentration was 78 µg/m³, higher in the South than in the North (89 vs. 65 µg/m³), and more in winter than in summer (119 vs. 69 µg/m³). The sources of industrial and traffic emissions (including carbon emissions from heating in winter) were mainly from the central and southern regions of the study area and the difference in mean concentration between the regions and between the seasons showed such a distribution pattern of PM_{2.5}. For training the model to solve the parameters of conversion, 37,128 data samples were used. Table 2 shows the annual and seasonal statistics of MAIAC AOD, PBLH, RH, and PM_{2.5} for the 2015 dataset.

Table 2. Statistics for the 2015 dataset.

Variable	Units	Range ^a	IQR ^b	Mean			Standard Deviation			Correlation ^c		
				Y ^d	S ^e	W ^f	Y ^d	S ^e	W ^f	Y ^d	S ^e	W ^f
Satellite AOD	-	(0–3.77)	0.92	0.79	1.05	0.62	0.63	0.71	0.54	0.39	0.60	0.59
PBLH	m	(76, 3016)	777	841	1240	356	492	293	205	−0.25	−0.06	−0.0026
RH	%	(18, 96)	27	59	66	54	17	12	16	0.19	0.10	0.50
PM _{2.5}	µg/m ³	(2, 735)	70	78	56	119	69	34	95	1	1	1

^a Value range (minimum, maximum); ^b IQR: Inter-quantile range, i.e., the difference between 75th and 25th percentiles of the data samples; ^c Pearson's correlation with PM_{2.5}; ^d 2015; ^e summer (June, July, and August) of 2015; ^f winter (December, January, and February) of 2015.

3.2. Learning and Validation

For the empirical conversion formula (Equation (6) with $it = 0$), different values of the exponential parameter g were used from a small value (−0.3) to a large value (0.8) to test Pearson's correlation of satellite AOD with PM_{2.5}. The result shows that the optimal correlation (0.54) was attained when $g = 0.208$ (Figure 4).

Figure 4. Optimal solution using the empirical method by varying g .

In machine learning, an epoch is one complete presentation of the data samples to be learned to a model [49]. Usually, a machine learner needs multiple epochs to approach the optimal solution. To obtain the conversion parameters (five in total), 20,000 epochs of training were conducted for this study. The learning curves (Figure 5) show that the loss, RMSE and Pearson's correlation with $PM_{2.5}$ in training and testing, and the exponential coefficient of RH (g in Equation (6)) reached a plateau with increasing number of training epochs. The final optimal solution is $g = 0.51$, $s_{RH} = 164$, $i_{RH} = 65$, $s_{HA} = 0.40$, and $i_{HA} = 33$ with a significant Pearson's correlation of 0.58 with $PM_{2.5}$ for the test dataset. Supplementary Materials Figure S1 shows the trend of the other parameters (scaling and shift factors for RH and PBLH) during learning. As an effective validation for our method, the independent test results show little or very small difference (Pearson's correlation with $PM_{2.5}$: 0.59 vs. 0.58) from the results of the training samples, indicating little over-fitting for our method.

Figure 5. Trends of loss (a), root mean square error (RMSE) (b) and Pearson's correlation with $PM_{2.5}$ (c) in training and testing, and g (d).

3.3. Comparison of the Methods

The test results (Table 3) show that original column (satellite) AOD without any adjustment had the lowest correlation (0.38) with $PM_{2.5}$ and R^2 (0.15) and the highest RMSE ($65 \mu\text{g}/\text{m}^3$) for linear regression; column AOD adjusted by PBLH improved the correlation from 0.38 to 0.53, an increase of 15% (R^2 increased by 0.13 and RMSE decreased by $5 \mu\text{g}/\text{m}^3$). Simulated GAC by the empirical method (Figure 4) further improved the correlation to 0.54 (R^2 increased by 0.01 and RMSE decreased by $1 \mu\text{g}/\text{m}^3$). The optimal simulated GAC by automatic differentiation achieved the highest correlation (0.58) with $PM_{2.5}$ (the highest R^2 being 0.33, and the lowest RMSE being $56 \mu\text{g}/\text{m}^3$).

Table 3. Metrics of AOD and GACs simulated by different methods in the test.

Method	Mean	Range	Standard Deviation	Correlation with PM _{2.5}	R ² in Linear Regression	RMSE in Linear Regression (µg/m ³)
Satellite AOD	0.78	0, 3.77	0.63	0.38	0.15	65
GAC (only by PBLH)	0.14	0, 1.40	0.16	0.53	0.28	60
Empirical GAC	0.11	0, 1.10	0.12	0.54	0.29	59
Optimal GAC by AD	0.33	0, 2.41	0.32	0.58	0.33	56

The histograms (Figure 6) show the value distributions of original AOD and GACs simulated (see Supplementary Materials Figure S2 for the scatter plots between observed vs. predicted PM_{2.5} using original AOD and simulated GAC). The results show a better match of optimal GAC by AD with PM_{2.5} than that by the other methods.

Figure 6. Distributions of satellite AOD, simulated GACs, and PM_{2.5}.

The time series (Figure 7) of total means of PM_{2.5}, AOD, and GAC show a seasonal variation of simulated GAC using AD, quite similar to that of PM_{2.5} (low in summer and high in winter) for the study area, compared with original satellite AOD (high in summer and low in winter).

3.4. Spatial Distributions of Simulated Ground Aerosol Coefficient and Predicted PM_{2.5}

With the optimal simulated GAC using AD, the grid surfaces of the simulated ground aerosol coefficient and predicted PM_{2.5} were made. A residual deep network was used as the regression model [50]. Compared with the models trained just using the original MAIAC AOD with the other covariates, the models trained using the optimal GAC improved the test R² from 0.78 to 0.90 (improvement of 12%). Figure 8 presents the grid surfaces of a typical Summer day (a, b, and c for 07/29/2015) and a typical Winter day (d, e, and f for 12/31/2015): a and c for original satellite AOD, b and d for optimal GAC simulated, and c and e for predicted PM_{2.5} using the optimal GAC.

Figure 7. Time-series plots for PM_{2.5}, satellite AOD, and simulated GAC by automatic differentiation (AD).

Figure 8. Cont.

Figure 8. Grid surfaces of Multiangle Implementation of Atmospheric Correction (MAIAC) AOD (a,d), optimal ground aerosol coefficient by AD (b,e), and PM_{2.5} predicted using optimal GAC (c,f).

As shown in the spatial distributions (Figure 8) of simulated AOD/GAC and estimated PM_{2.5} in summer and winter, the optimal GAC (Figure 8b,e) converted by AD better matched seasonal distribution of PM_{2.5} (higher concentration in winter than in summer) (Figure 7); otherwise, the original AOD (Figure 8a,d) had higher values in summer than in winter, opposite of the seasonal variation of PM_{2.5}. In addition, spatial distribution of AOD/GAC and PM_{2.5} consistently showed higher values in the central and southern regions than the other regions within the study area. As aforementioned, the study area's primary emission sources include industrial pollutions, coal, and vehicle exhausts, and these sources have been aggregated in the populated central and southern regions, thus may be resulting in heavier concentration of PM_{2.5}. In winter for northern China, with the aggravating factors of low elevation, low wind speed, high atmospheric stability, and less rainfall, much more coal combustion due to heating in these populated regions has resulted in much more pollution than in the other seasons and regions. In summer, the difference in PM_{2.5} concentration between the regions was small due to less coal emission and more favorable weather conditions for pollutant dispersion than in winter. For Beijing, China's capital, more heavy industry factories had been relocated and so, on average, it had less pollution than the other cities and regions, shown as Figure 8c,f.

4. Discussion

Although estimation of PM_{2.5} concentration at high spatial (e.g., 1 km) and temporal (e.g., daily) resolution is very helpful for evaluation of its emissions and health effects [51,52], it is very challenging given the limited amount of monitoring data sources and the influence of multiple factors (diverse emission source, meteorology, atmospheric process, and elevation) and their interactions on spatiotemporal variability of PM_{2.5} [53]. Satellite AOD, such as recent high spatiotemporal-resolution MAIAC AOD, has been used as a predictor of primary interest for spatiotemporal estimation of PM_{2.5} due to its global spatial and temporal coverage. It can compensate unavailability or limited availability of the other predictors, e.g., emission sources and inventory at a fine resolution [38]. However, for satellite AOD, the potential difficulty is the significant inconsistency or a difference between its seasonal patterns and ground aerosol mass including PM₁₀ and PM_{2.5} [15,17] given that AOD measures a column integrated aerosol extinction coefficient, different from ground-level aerosol mass. Previously, a simplified conversion [15] from satellite AOD to ground aerosol coefficient was given to improve the representative and prediction power of column AOD for PM_{2.5}. However, this approach of a single-parameter (just an exponent parameter of RH to be solved) formula, similar to a fixed-effect model, did not consider the influence of uncertainty deriving from the measurement errors and/or other confounders and thus might not capture the influence from such uncertainty on the results. The other empirical conversion methods considered more influential factors such as particle size [23] and visibility [25] although such covariates are not available for the case study. Many of these existing methods used a conversion formula of a fixed format and used empirical values to fill in several parameters to finish the conversion. The empirical values of these parameters may not fit in with the specific context and data samples, thus are potentially suboptimal.

Compared with the existing methods [15,20–26,30,31], although just having a modest improvement (4% for the empirical method) for the test correlation between the simulated GAC and PM_{2.5}, this proposed method provides a more flexible framework to fuse the influence of multiple scaling, shift, and other potential factors on the conversion that involves complex atmospheric and chemical processes of aerosols; this method is convenient to train and use in support of the AD tool, and makes the conversion easily adjusted or improved using the data of additional influential factors if available. The polynomial conversion from proxy GAC to PM_{2.5} was conducted to employ the stable RSE loss function to obtain a slightly better correlation than linear conversion (0.58 vs. 0.56). These factors worked similarly as random effect factors that aimed to account for more variance and the influence of uncertainty from the other potential confounders. In addition, the proposed method can conveniently incorporate more polynomial terms of the parameters, and/or the other available covariates to obtain an optimal solution of multiple parameters if these terms or covariates can significantly affect the conversion from satellite

AOD to GAC. A potential extension of the proposed method is to infer the optimal parameters for the probabilistic distribution of the vertical profile of aerosol extinction coefficient if vertical data of multiple layers are available to simulate such a distribution. Automatic differentiation provides a convenient and reliable tool with which to be able to employ gradient descent to find a local optimal combinational solution for multiple parameters. To use automatic differentiation, the scaling and shift factors were also introduced for the proxy variable simulated for the ground aerosol extinction coefficient to use the MSE loss function, which can work better than the loss of Pearson's correlation coefficient in training [33]. The results show that the proposed conversion method achieved the best performance—the highest test correlation (0.58) with $PM_{2.5}$ and highest test R^2 (0.33) and lowest test RMSE ($56 \mu g/m^3$) for regression of $PM_{2.5}$ —compared with the other methods. The ground aerosol coefficient simulated using AD had a distribution and seasonal variation quite similar to those of $PM_{2.5}$. Compared with original column (MAIAC) AOD, the simulated GAC improved Pearson's correlation with a ground true $PM_{2.5}$ from 0.38 to 0.58.

Use of gradient descent in AD made the solution locally optimal, not globally optimal. Thus, the output simulated in Equation (7) could not be completely equal to the ground aerosol extinction coefficient, but rather it would be called a proxy variable for the ground aerosol extinction coefficient. Direct estimation of the ground aerosol extinction coefficient is difficult since just very limited vertical profile data of the aerosol extinction coefficient have been available for training a conversion function [15,17].

For estimation of $PM_{2.5}$, given the simulated GAC available, the other factors, including meteorology (wind speed, relative humidity, air temperature, precipitation, etc.), elevation, and traffic density, should be considered in the models since these factors affect spatiotemporal variability of $PM_{2.5}$ [53,54]. With fusion of these data within the model of a residual deep network [50], the test R^2 of $PM_{2.5}$ estimations reached 0.90. Compared with original column (MAIAC) AOD, the ground aerosol coefficient simulated contributed to a 12% improvement in R^2 . The simulated GAC was quite similar to $PM_{2.5}$ in terms of probabilistic distribution (Figure 6) and spatial distribution (Figure 8) and had a greater contribution than original AOD to spatiotemporal estimation of $PM_{2.5}$. The results show a better match between optimal simulated GAC and $PM_{2.5}$ than original AOD. Although the correlation between $PM_{2.5}$ and simulated GAC was moderately high, such a correlation was just linear. Given multiple different emission sources and complex atmospheric and chemical processes in generation of secondary $PM_{2.5}$, the practical relationship is complex and non-linear. As aforementioned, the spatiotemporal estimation model of $PM_{2.5}$ should consider the influence of multiple factors and their non-linear interactions with GAC. The test showed that without GAC as a predictor, the trained model just had a test R^2 of 0.60 in estimation of $PM_{2.5}$, compared to the test R^2 of 0.90 achieved by the models using GAC. Thus, simulated GAC made a contribution of about 30% in R^2 over the model not using AOD.

The proposed method has three limitations. First, the solution of the presented method is not globally but locally optimal. The flexible model proposed may make a difference in predicted parameters (e.g., scaling and shift factors) between different training cycles. However, results of sensitivity tests show a small standard deviation for the final simulated GAC, illustrating the reliability of the proposed method. Second, the proposed method did not incorporate specific multiple layers in a vertical profile such as those provided by the active LIDAR satellite due to data unavailability. Such vertical information may be helpful for the conversion. However, in practical applications, such multilayer information at a sufficient spatial resolution and spatiotemporal coverage is often unavailable, and many existing studies for $PM_{2.5}$ estimation did not incorporate it in using satellite AOD for $PM_{2.5}$ estimation [15,55]. This paper used an exponential distribution to simplify and simulate such a vertical profile for practical applications. Based on a flexible model and automatic differentiation, although the simulation of the vertical profile is not perfect, the proposed method achieved the state-of-the-art performance in the conversion and downstream application of $PM_{2.5}$ estimation. Third, all the discussed conversion methods work best in the areas dominated by regional aerosols residing

in the boundary layer, but work less well in the areas dominated by long-range aerosol transport at higher altitudes [56,57].

With the availability of deep-learning software, it is convenient to use AD extensively to solve the complex formula (e.g., involving exponential, scaling, and shift parameters) for inversion of conversion parameters for satellite AOD and other environmental surface variables of remote sensing.

5. Conclusions

For inconsistency between satellite column AOD (e.g., MAIAC AOD) and ground aerosol extinction coefficient/PM_{2.5}, in this paper a novel method of automatic differentiation for convenient conversion from satellite AOD to ground aerosol extinction coefficient is proposed. The proposed method introduced scaling and shift factors for PBLH and RH to account for the uncertainty deriving from the measurement and other confounders. Based on the computational graph, AD efficiently automates the differentiation, which has driven extensive applications of gradient descent in machine and deep learning. As the base of the proposed method, AD provides a convenient tool with which to obtain an optimal combinational solution for multiple parameters of the proposed method. In order to use the stable MSE loss function, a polynomial conversion from GAC to PM_{2.5} was suggested to achieve a better correlation of GAC with PM_{2.5} than a linear conversion. In a case study of the Jing-Jin-Ji metropolitan area, the proposed method achieved the highest test correlation (0.58) of the ground aerosol coefficient simulated with PM_{2.5} and the highest test R² (0.33) in linear regression, compared with original AOD and GAC obtained using the single-parameter method. In the extensive downstream validation of PM_{2.5} estimation, our method improved the test R² from 0.78 to 0.90, and achieved a better match of GAC with PM_{2.5} in their spatial distributions and seasonal variation. Similar methods can be also applied to solve the inversion parameters for other surface grid variables in remote sensing.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-4292/12/3/492/s1>, Figure S1: Learning curves of scaling (slope) and shift (intercept) factors for relative humidity and PBLH, Figure S2: Scatter plots of the observed vs. predicted PM_{2.5} in univariate linear regression, illustrating improvement of the optimal ACG by automatic differentiation.

Funding: This work was supported in part by the Strategic Priority Research Program of Chinese Academy of Sciences Grant XDA19040501 and in part by the National Natural Science Foundation of China under Grant 41471376.

Acknowledgments: The author gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan Xp GPUs used for this research.

Conflicts of Interest: The author declares no conflict of interest.

References

1. EPA. Criteria Air Pollutants. Available online: <https://www.epa.gov/criteria-air-pollutants#self> (accessed on 18 August 2019).
2. EPA. Particulate Matter Emissions. Available online: https://cfpub.epa.gov/roe/indicator_pdf.cfm?i=19 (accessed on 5 July 2019).
3. WHO. *Health Effects of Particulate Matter: Policy Implications for Countries in Eastern Europe, Caucasus and Central Asia*; World Health Organization: Copenhagen, Denmark, 2013.
4. Bowe, B.; Xie, Y.; Li, T.; Yan, Y.; Xian, H.; Al-Aly, Z. The 2016 global and national burden of diabetes mellitus attributable to PM_{2.5} air pollution. *Lancet Planet. Health* **2018**, *2*, e301–e312. [CrossRef]
5. Potera, C. Toxicity beyond the Lung: Connecting PM_{2.5}, Inflammation, and Diabetes. *Environ. Health Perspect.* **2014**, *122*, A29. [CrossRef] [PubMed]
6. Zanobetti, A.; Dominici, F.; Wang, Y.; Schwartz, J.D. A national case-crossover analysis of the short-term effect of PM_{2.5} on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environ. Health-Glob.* **2014**, *13*, 38. [CrossRef] [PubMed]
7. Zheng, J. Monitoring Network to be Further Expanded. Available online: http://www.chinadaily.com.cn/china/2017-04/07/content_28827498.htm (accessed on 20 July 2019).

8. BMEPB. A New Round of Beijing PM_{2.5} Source Analysis Officially Released. Available online: <http://www.bjepb.gov.cn/bjhrb/xxgk/jgsz/jjggszjzz/xcyj/xwfb/607219/index.html> (accessed on 1 August 2019).
9. Hong'e, M. Sources of Beijing PM_{2.5} Pollutants Mainly Local. Available online: <https://www.ecns.cn/cns-wire/2018/05-14/302510.shtml> (accessed on 1 August 2019).
10. Zíková, N.; Wang, Y.; Yang, F.; Li, X.; Tian, M.; Hopke, P.K. On the source contribution to Beijing PM_{2.5} concentrations. *Atmos. Environ.* **2016**, *134*, 84–95. [CrossRef]
11. Liu, M.; Lin, J.; Wang, Y.; Sun, Y.; Zheng, B.; Shao, J.; Chen, L.; Zheng, Y.; Chen, J.; Fu, T.-M. Spatiotemporal variability of NO₂ and PM_{2.5} over Eastern China: Observational and model analyses with a novel statistical method. *Atmos. Chem. Phys.* **2018**, *18*, 12933–12952. [CrossRef]
12. Wang, X.; Dickinson, R.; Su, L.; Zhou, C.; Wang, K. PM_{2.5} Pollution in China and How It Has Been Exacerbated by Terrain and Meteorological Conditions. *Bull. Am. Meteorol. Soc.* **2017**, *99*, 105–119. [CrossRef]
13. Xu, M.; Sbihi, H.; Pan, X.; Brauer, M. Local variation of PM_{2.5} and NO₂ concentrations within metropolitan Beijing. *Atmos. Environ.* **2019**, *200*, 254–263. [CrossRef]
14. NASA. Aerosol Optimal Depth. Available online: https://aeronet.gsfc.nasa.gov/new_web/Documents/Aerosol_Optical_Depth.pdf (accessed on 12 July 2017).
15. Wang, Z.F.; Chen, L.F.; Tao, J.H.; Zhang, Y.; Su, L. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sens. Environ.* **2010**, *114*, 50–63. [CrossRef]
16. Wong, M.S.; Nichol, J.; Lee, K.H. Modeling of aerosol vertical profiles using GIS and remote sensing. *Sensors* **2009**, *9*, 4380–4389. [CrossRef]
17. Li, J.; Carlson, E.B.; Laci, A.A. How well do satellite AOD observations represent the spatial and temporal variability of PM_{2.5} concentration for the United States? *Atmos. Environ.* **2015**, *102*, 260–273. [CrossRef]
18. Seidel, D.J.; Ao, C.O.; Li, K. Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. *J. Geophys. Res.-Atmos.* **2010**, *115*. [CrossRef]
19. Yang, Q.Q.; Yuan, Q.Q.; Yue, L.W.; Li, T.W.; Shen, H.F.; Zhang, L.P. The relationships between PM_{2.5} and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. *Environ. Pollut.* **2019**, *248*, 526–535. [CrossRef] [PubMed]
20. Chu, Y.Y.; Liu, Y.S.; Li, X.Y.; Liu, Z.Y.; Lu, H.S.; Lu, Y.A.; Mao, Z.F.; Chen, X.; Li, N.; Ren, M.; et al. A Review on Predicting Ground PM_{2.5} Concentration Using Satellite Aerosol Optical Depth. *Atmosphere* **2016**, *7*, 129. [CrossRef]
21. Hoff, R.M.; Christopher, S.A. Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land? *J. Air Waste Manag. Assoc.* **2009**, *59*, 645–675. [CrossRef] [PubMed]
22. Chew, B.N.; Campbell, J.R.; Hyer, E.J.; Salinas, S.V.; Reid, J.S.; Welton, E.J.; Holben, B.N.; Liew, S.C. Relationship between Aerosol Optical Depth and Particulate Matter over Singapore: Effects of Aerosol Vertical Distributions. *Aerosol Air Qual. Res.* **2016**, *16*, 2818–2830. [CrossRef]
23. Li, Y.; Xue, Y.; Guang, J.; She, L.; Fan, C.; Chen, G. Ground-Level PM_{2.5} Concentration Estimation from Satellite Data in the Beijing Area Using a Specific Particle Swarm Extinction Mass Conversion Algorithm. *Remote Sens.* **2018**, *11*, 10. [CrossRef]
24. Shrestha, B.; Joseph, E. Retrieval of PM_{2.5} profile using the Doppler Lidar across New York State Mesonet. In Proceedings of the 19th Coherent Laser Radar Conference, Okinawa, Japan, 18–21 June 2018.
25. Zeng, Q.L.; Chen, L.F.; Zhu, H.; Wang, Z.F.; Wang, X.H.; Zhang, L.; Gu, T.Y.; Zhu, G.Y.; Zhang, Y. Satellite-Based Estimation of Hourly PM_{2.5} Concentrations Using a Vertical-Humidity Correction Method from Himawari-AOD in Hebei. *Sensors* **2018**, *18*, 3456. [CrossRef]
26. Jin, X.M.; Fiore, A.M.; Curci, G.; Lyapustin, A.; Civerolo, K.; Ku, M.; van Donkelaar, A.; Martin, R.V. Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. *Atmos. Chem. Phys.* **2019**, *19*, 295–313. [CrossRef]
27. Li, S.; Zhang, L.; Cai, K.; Ge, W.; Zhang, X. Comparisons of the vertical distributions of aerosols in the CALIPSO and GEOS-Chem datasets in China. *Atmos. Environ. X* **2019**, *3*, 100036. [CrossRef]
28. Appel, K.W.; Chemel, C.; Roselle, S.J.; Francis, X.V.; Hu, R.M.; Sokhi, R.S.; Rao, S.T.; Galmarini, S. Examination of the Community Multiscale Air Quality (CMAQ) model performance over the North American and European domains. *Atmos. Environ.* **2012**, *53*, 142–155. [CrossRef]

29. Quennehen, B.; Raut, J.C.; Law, K.S.; Daskalakis, N.; Ancellet, G.; Clerbaux, C.; Kim, S.W.; Lund, M.T.; Myhre, G.; Olivie, D.J.L.; et al. Multi-model evaluation of short-lived pollutant distributions over east Asia during summer 2008. *Atmos. Chem. Phys.* **2016**, *16*, 10765–10792. [CrossRef]
30. Sun, T.Z.; Che, H.Z.; Qi, B.; Wang, Y.Q.; Dong, Y.S.; Xia, X.G.; Wang, H.; Gui, K.; Zheng, Y.; Zhao, H.J.; et al. Aerosol optical characteristics and their vertical distributions under enhanced haze pollution events: Effect of the regional transport of different aerosol types over eastern China. *Atmos. Chem. Phys.* **2018**, *18*, 2949–2971. [CrossRef]
31. Toth, T.D.; Zhang, J.L.; Reid, J.S.; Vaughan, M.A. A bulk-mass-modeling-based method for retrieving particulate matter pollution using CALIOP observations. *Atmos. Meas. Tech.* **2019**, *12*, 1739–1754. [CrossRef]
32. BMEPB. Main sources of PM_{2.5} in Beijing: Vehicles, Coal Burning, Industry, Dust and Neighboring Cities. Available online: <https://cleanairasia.org/node12353/> (accessed on 20 August 2019).
33. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
34. Sun, L. Why the Smog in North China Is So Big? Available online: <https://zhuanlan.zhihu.com/p/49894955> (accessed on 12 October 2019).
35. Miao, Y.; Zheng, J.; Wang, S.; Liu, S. Recent advances in, and future prospects of, research on haze formation over Beijing–Tianjin–Hebei, China. *Clim. Environ. Res.* **2015**, *20*, 36–368.
36. Wang, Y.; Zhang, J.; Wang, L.; Hu, B.; Tang, G.; Liu, Z.; Sun, Y.; Ji, D. Researching significance, status and expectation of haze in Brijing–Tianjing–Hebei region. *Adv. Earth Sci.* **2014**, *29*, 387–396.
37. Hu, X.F.; Waller, L.A.; Lyapustin, A.; Wang, Y.J.; Al-Hamdan, M.Z.; Crosson, W.L.; Estes, M.G.; Estes, S.M.; Quattrochi, D.A.; Puttaswamy, S.J.; et al. Estimating ground-level PM_{2.5} concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sens. Environ.* **2014**, *140*, 220–232. [CrossRef]
38. Xiao, Q.; Wang, Y.; Chang, H.H.; Meng, X.; Geng, G.; Lyapustin, A.; Liu, Y. Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* **2017**, *199*, 437–446. [CrossRef]
39. Wang, Z.; Fang, C.; Xu, G.; Pan, Y. Spatial-temporal characteristics of the PM_{2.5} in China in 2014. *Acta Geogr. Sin.* **2015**, *70*, 1720–1734.
40. Fang, Y.; Li, L. Estimation of high-precision high-resolution meteorological factors based on machine learning. *J. Geo-Inf. Sci.* **2019**, in press.
41. Chung, C.E. Aerosol direct radiative forcing: A review. In *Atmospheric Aerosols—Regional Characteristics—Chemistry and Physics*; Abdul-Razzak, H., Ed.; Scitus Academics: Wilmington, NC, USA, 2012; pp. 379–394.
42. Léon, J.-F.; Derimian, Y.; Chiapello, I.; Tanré, D.; Podvin, T.; Chatenet, B.; Diallo, A.; Deroo, C. Aerosol vertical distribution and optical properties over M'Bour (16.96°W; 14.39°N), Senegal from 2006 to 2008. *Atmos. Chem. Phys.* **2009**, *9*, 9249–9261.
43. 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]
44. Lisheng, Z.; Guangyu, S. The impact of relative humidity on the radiative property and radiative forcing of sulfate aerosol. *J. Meteorol. Res.* **2001**, *15*, 465–476.
45. Malm, W.C.; Day, D.E.; Kreidenweis, S.M. Light scattering characteristics of aerosols as a function of relative humidity: Part I—A comparison of measured scattering and aerosol concentrations using the theoretical models. *J. Air Waste Manag. Assoc.* **2000**, *50*, 686–700. [CrossRef] [PubMed]
46. Baydin, G.A.; Pearlmutter, B.; Radul, A.A.; Siskind, J. Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* **2018**, *18*, 1–43.
47. Laue, S. On the equivalence of forward mode automatic differentiation and symbolic differentiation. *arXiv* **2019**, arXiv:1904.02990.
48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
49. Bishop, M.C. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
50. Li, L.; Fang, Y.; Wu, J.; Wang, J. Autoencoder Based Residual Deep Networks for Robust Regression Prediction and Spatiotemporal Estimation. *arXiv* **2018**, arXiv:1812.11262.
51. Giannadaki, D.; Lelieveld, J.; Pozzer, A. Implementing the US air quality standard for PM_{2.5} worldwide can prevent millions of premature deaths per year. *Environ. Health-Glob.* **2016**, *15*, 88. [CrossRef]

52. Shi, L.; Zanobetti, A.; Kloog, I.; Coull, B.A.; Koutrakis, P.; Melly, S.J.; Schwartz, J.D. Low-concentration PM_{2.5} and mortality: Estimating acute and chronic effects in a population-based study. *Environ. Health Perspect.* **2015**, *124*, 46–52. [\[CrossRef\]](#)
53. Li, L.; Zhang, J.; Meng, X.; Fang, Y.; Ge, Y.; Wang, J.; Wang, C.; Wu, J.; Kan, H. Estimation of PM_{2.5} concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with MAIAC aerosol optical depth. *Remote Sens. Environ.* **2018**, *217*, 573–586. [\[CrossRef\]](#)
54. Li, L.; Wu, A.; Cheng, I.; Chen, J.; Wu, J. Spatiotemporal estimation of historical PM_{2.5} concentrations using PM₁₀, meteorological variables, and spatial effect. *Atmos. Environ.* **2017**, *166*, 182–191. [\[CrossRef\]](#)
55. Krishna, R.K.; Ghude, S.D.; Kumar, R.; Beig, G.; Kulkarni, R.; Nivdange, S.; Chate, D. Surface PM_{2.5} Estimate Using Satellite-Derived Aerosol Optical Depth over India. *Aerosol Air Qual. Res.* **2019**, *19*, 25–37. [\[CrossRef\]](#)
56. Guo, P.; Yu, S.C.; Wang, L.Q.; Li, P.F.; Li, Z.; Mehmood, K.; Chen, X.; Liu, W.P.; Zhu, Y.N.; Yu, X.; et al. High-altitude and long-range transport of aerosols causing regional severe haze during extreme dust storms explains why afforestation does not prevent storms. *Environ. Chem. Lett.* **2019**, *17*, 1333–1340. [\[CrossRef\]](#)
57. Zhao, Z.Z.; Cao, J.J.; Shen, Z.X.; Xu, B.Q.; Zhu, C.S.; Chen, L.W.A.; Su, X.L.; Liu, S.X.; Han, Y.M.; Wang, G.H.; et al. Aerosol particles at a high-altitude site on the Southeast Tibetan Plateau, China: Implications for pollution transport from South Asia. *J. Geophys. Res.-Atmos.* **2013**, *118*, 11360–11375. [\[CrossRef\]](#)

© 2020 by the author. 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/>).