LiDAR-Based Remote Sensing of the Vertical Profile of Aerosol Liquid Water Content Using a Machine-Learning Model

Tong Wu, Zhanqing Li, Xiaoai Jin, Wei Wang, Hao Wu, Rongmin Ren, Dongmei Zhang, Lu Chen, Yunfei Su, and Maureen Cribb

***Abstract*— The aerosol liquid water content (ALWC) dictates the hygroscopicity of aerosol particles. To date, measurements of ALWC have been confined primarily to ground-based observations although vertical profiles of ALWC are crucial for understanding its interactions with meteorology. This study proposes a novel method for deriving profiles of ALWC using data acquired by a Light Detection and Ranging (LiDAR), a microwave radiometer, and a suite of aerosol instruments measuring aerosol physical and chemical properties, deployed during a five-month field experiment in Guangzhou, China. The retrieval approach is based on a machine-learning model named the gradient-boosted decision tree model. The inversion accuracy and stability are assessed through comparisons with ALWC data acquired on the ground and at the top of the Guangzhou tower of 532 m above ground. The agreements are encouraging: with the coefficient of determination (R^2) = 0.870 and root-mean-square error (RMSE) = 3.28 $\mu g \cdot m^{-3}$ for all data; R^2 = 0.776 and RMSE = 2.18 $\mu g \cdot m^{-3}$ for tower data; and R^2 = 0.872 and RMSE = 4.1 $\mu g \cdot m^{-3}$ for ground data. From the vertical distribution of the retrieved ALWC in Guangzhou, ALWC is higher in the lower boundary layer, especially when air pollution is severe. The proportion of liquid water in aerosol particles is closely related to the relative humidity in the environment, which will affect the morphology of aerosol particles (with about every 10% increase in liquid water, the depolarization ratio decreases by 0.02). The model may be of general use for studying air pollution and secondary aerosol generation.**

***Index Terms*— Aerosol liquid water content (ALWC), Light Detection and Ranging (LiDAR), machine learning (ML), vertical profile.**

Manuscript received May 13, 2021; revised August 2, 2021 and October 12, 2021; accepted November 10, 2021. Date of publication November 23, 2021; date of current version March 29, 2022. This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 42030606, in part by the National Key Research and Development Program of China under Grant 2017YFC1501702, and in part by the Open Fund of the State Key Laboratory of Remote Sensing Science under Grant 202015. *(Corresponding author: Zhanqing Li.)*

Tong Wu, Xiaoai Jin, Wei Wang, Rongmin Ren, Dongmei Zhang, Lu Chen, and Yunfei Su are with the State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China (e-mail: wutong0207@outlook.com; 201631490012@mail.bnu.edu.cn; wangw_14@mail.bnu.edu.cn; renrongmin@mail.bnu.edu.cn; 201921490030@mail.bnu.edu.cn; 201821490020@mail.bnu.edu.cn; 201731490012@mail.bnu.edu.cn).

Zhanqing Li and Maureen Cribb are with the Department of Atmospheric and Oceanic Science and the Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD 20742 USA (e-mail: zli@atmos.umd.edu; mcribb@umd.edu).

Hao Wu is with the Key Laboratory of China Meteorological Administration Atmospheric Sounding, School of Electrical Engineering, Chengdu University of Information Technology, Chengdu 610225, China (e-mail: wcgse@live.cn).

Digital Object Identifier 10.1109/TGRS.2021.3130204

I. Introduction

AEROSOLS directly and indirectly affect the Earth's energy budget and water cycle by altering cloud properties [1]. Both effects are associated with aerosol hygroscopicity, which denotes the amount of water contained in the aerosol particles, which is referred to as the aerosol liquid water content (ALWC). Changes in ALWC not only change the physical characteristics of aerosol particles but also provide a medium for complex chemical reactions in the atmosphere [2], [3]. Knowledge of the ALWC can help understand and account for the optical properties of aerosols, as well as the aqueous phase chemical reaction that is important for studying the development of secondary aerosols [4]–[6].

The ALWC is mainly determined by the hygroscopic growth characteristics of aerosol particles and the relative humidity (RH) of the environment [7]. There are two effective ways to obtain the ALWC. A common method is to obtain the size distribution of aerosol particles and their hygroscopicity [8]. Stanier *et al.* [9] obtained the particle distribution of aerosols in dry and wet states and calculated the ALWC through the volume difference. Another way to calculate the ALWC is based on aerosol chemical composition and the phase state of aerosol particles that are used as inputs to an aerosol thermodynamic model to simulate the hygroscopicity and ALWC. Examples of such models are ISORROPIA and the aerosol inorganics model (AIM) [10], [11]. Research on ALWC in China has also been carried out in recent years. Liu *et al.* [12] studied the differences in the hygroscopicity of aerosol particles at different scales and with different compositions under a hazy condition in the North China Plain, laying the foundation for studying ALWC and secondary aerosol formation in northern China. Wu *et al.* [13] found that, in the Beijing area, due to the hygroscopic effect of aerosols, aerosol liquid water can act as an effective medium to accelerate the reaction of gaseous pollutants on the surface of aerosol particles, increasing the degree of haze. In this study, the ISORROPIA II model is selected to simulate the ALWC, with aerosol chemical composition as input [14].

The ALWC increases aerosol' scattering characteristics, making aerosol particles more morphologically regular in shape [15]–[17]. Thus, Kuang *et al.* [18] proposed a new method to calculate the ALWC based on the aerosol optical scattering coefficient obtained by dry and wet nephelometer

systems. As mentioned before, almost all methods for calculating the ALWC are based on observational near-surface data. However, the vertical distribution of ALWC has an important impact on the atmospheric environment and cloud–aerosol interactions. The concentration of aerosol particles and the RH of the environment vary significantly with altitude [19], [20]. Examining the variation of ALWC with height helps toward understanding the evolution of pollution and its causes and development regimes, especially in megacities [21]. Compared with more expensive aircraft-based observation experiments, a ground-based polarized aerosol Light Detection and Ranging (LiDAR) provides optical information about the vertical distribution of aerosol continuously over long periods. Studying the vertical variation of ALWC is, thus, feasible. However, to date, few such studies have been conducted. Tan *et al.* [22] used a polarized LiDAR to establish an empirical model based on the relationship between the depolarization ratio (DR), the backscattering coefficient, and ALWC, generating vertical ALWC profiles during a heavy pollution event in Beijing. Note that the ALWC empirical model developed in this study is aimed at heavy air pollution, so its applicability to other atmospheric conditions needs to be verified. There has also been a general lack of high-altitude data for such investigations.

Here, profiles of aerosol optical parameters obtained by a LiDAR together with aerosol hygroscopic properties are used to calculate changes in the backscattering coefficient caused by ALWC. In combination with coincident weather observations, ALWC profiles can be retrieved continuously over a long timeframe. The machine-learning (ML) method called the gradient-boosted decision tree (GBDT) model is used for the first time to infer the ALWC profile. ML algorithms have been increasingly used to develop complex relationships that cannot be handled by conventional regression approaches [23]–[25], often leading to significant improvements in the accuracy and efficiency of calculations. Compared with other regression calculation methods, the biggest advantage of the ML method used here is that it can handle discrete and continuous data flexibly [26]. Due to its simple structure, the calculation is efficient. The retrieved ALWC in the vertical direction is validated using observation data at two different heights. Section II introduces the observation stations, observation instruments, and data. Section III describes the details of the algorithm for calculating the ALWC profile, and Section IV presents validation of the estimates of the vertical distribution of the ALWC. Section V summarizes the study and provides conclusions.

II. Description of the Sites and Datasets

A. Observation Sites

Observations were made at two sites in Guangzhou. Fig. 1 shows the locations of the sites and the locations of the instruments at these sites. One site is the ground observation site (23°00′30″ N, 113°19′03″ E) of the Guangzhou Meteorological Bureau, next to the Starlight Expressway and Changlong Zoo. The other site is the Guangzhou Tower observation site (23°06′31″ N, 113°19′09″ E), where instruments are deployed 532 m above the ground. Both sites are located in a southern part of Guangzhou, about 10 km apart, and have similar topographies. This study uses observations acquired from November 1, 2019, to March 30, 2020, at the ground site and from January 16 to March 30, 2020, at the tower site.

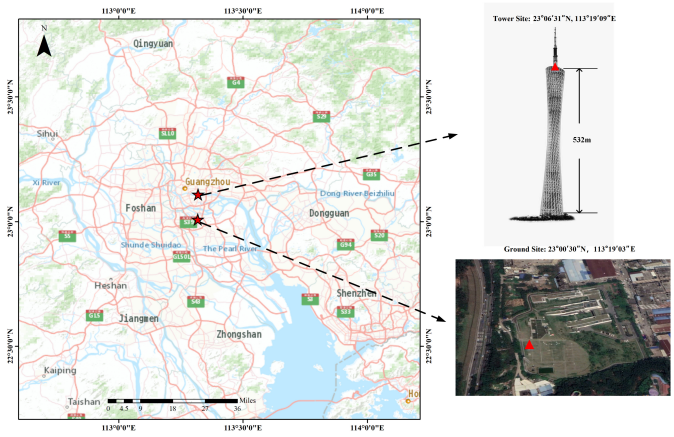

Fig. 1. Map on (Left) shows the locations of the two observation sites in Guangzhou (red stars) and their surroundings. Red triangles in the right-hand side pictures show the locations of the instruments on the Guangzhou Tower (Upper Right) and the Guangzhou Meteorological Bureau site (Lower Right).

B. Datasets

The micropulse LiDAR system (MPL, Sigma Space Corporation) provided measurements of the backscattering coefficient (β_{aer}) of aerosol particles at different elevations (H) at the 532-nm wavelength beyond the blind zone of the lower 260 m. The main aerosol optical parameter used here is the volume DR that can characterize the regularity of the aerosol surface and aerosol concentration [27]. After preprocessing, the volume DR can be obtained by the signal ratio of the two channels [28]. Finally, β_{aer} was calculated using the Fernald method [29]. The aerosol hygroscopic enhancement factor ($f(\text{RH})$) was measured using a dual-nephelometer system (Ecotech) with a high time resolution. Dry (RH $<$ 40%) and wet (a given, elevated RH) aerosol optical scattering characteristics were recorded by the two nephelometers every minute [30]. After 45 min, a complete observation cycle of $f(\text{RH})$ (RH$\sim$30%–90%) was obtained.

To train the fit model and verify the accuracy of the model, ALWC data are needed. As described in Section I, the ALWC can be estimated by the chemical composition of aerosols and aerosol thermodynamic models. An aerosol chemical speciation monitor (ACSM) can obtain the mass concentration of various chemical inorganic species and organic species of aerosols in real time [31]. The ACSM with a 2.5-μm air cutting head was used to collect chemical composition data of fine-mode aerosols during the Guangzhou observation period [32]. Half of the data were observations made near the ground, and the other half were observations made from the Guangzhou tower. The RH and temperature (T) profiles were retrieved by a microwave radiometer (MWR, HATPRO; Radiometer Physics GmbH, Germany), conveying water vapor information in the 22–30-GHz band and temperature information in the 51–59-GHz band [33].

TABLE I
LIST OF INSTRUMENTS AND PARAMETERS

Instruments	Parameters	Type	Location	Time Resolution	Vertical Spatial Resolution
MPL	Backscatter coefficient profile, extinction coefficient profile, volume depolarization ratio profile	Sigma Space Corp., model 4202	Ground	30 s	30 m
ACSM	Aerosol chemical composition	Aerodyne Q-ACSM	Ground (19.11.1-20.1.15); Tower (20.1.16-20.3.30)	15 min	
Dual-Nephelometer	Aerosol hygroscopic enhancement factor	Ecotech	Ground	45 min	
MWR	Temperature and humidity profile	RPG-HATPRO-G4	Ground	3 s	30 m
Doppler Wind Lidar	Wind speed and direction profile	Wind Print S4000	Ground	1 s	15 m

Particulate matter that is 2.5 μm or smaller in size ($PM_{2.5}$) data is from the China National Environmental Monitoring Center measured at a site 2 km from the ground observation site [34]. According to the classification criteria provided by the Environmental Protection Agency, $PM_{2.5}$ concentrations are divided into different pollution levels (PLs): 0–12 $\mu g \cdot m^{-3}$ (good level), 12–35.5 $\mu g \cdot m^{-3}$ (moderate level), 35.5–55.4 $\mu g \cdot m^{-3}$ (sensitive level), 55.4–150.4 $\mu g \cdot m^{-3}$ (unhealthy level), 150.5–250.4 $\mu g \cdot m^{-3}$ (very unhealthy level), and 250.5–500.4 $\mu g \cdot m^{-3}$ (hazardous level) [35].

Based on the Doppler coherent detection principle [36], [37], a wind profiler LiDAR operating at 1550-nm retrieved wind speeds (WSs) and wind directions (WDs). Table I lists the instruments used in this study. In data processing, to ensure that there is enough data for modeling, all data are unified to a 15-min time resolution.

III. METHODOLOGY

A. Potential Features

The main factors affecting ALWC are the hygroscopic properties of aerosols and the RH of the environment [10], [21]. Moisture changes in aerosols affect their optical and microphysical properties [38]. The hygroscopic characteristics of aerosols can, thus, be defined by optical parameters [39]

$$f_\beta(\mathrm{RH}, \lambda) = \frac{\beta(\mathrm{RH}, \lambda)}{\beta(\mathrm{RH_{dry}}, \lambda)} \tag{1}$$

where $\beta(\mathrm{RH}, \lambda)$ is the backscattering coefficient at a particular RH and wavelength λ. Note that $\beta(\mathrm{RH_{dry}}, \lambda)$ is similar to $\beta(\mathrm{RH}, \lambda)$, representing the backscattering coefficient at a dry reference, generally RH = 40% [40]. When the RH of the environment is greater than 40%, hygroscopic growth may start to affect aerosol optical properties [41]. The conditions required to study the hygroscopicity of aerosols by LiDAR alone are limited because it entails continuous changes over time [42], [43]. Many studies have shown that the hygroscopic properties of aerosols at the same time and the same place obtained by LiDAR and nephelometer have a high degree of consistency [44], [45]. Because the conditions required by a LiDAR to obtain the hygroscopic characteristics of aerosols for continuous periods of time are rigid, to ensure continuous data, aerosol hygroscopic enhancement factors at different times were obtained by using wet and dry nephelometers. The effect of aerosol hygroscopicity on the backscattering coefficient, $\Delta\beta$, can roughly represent the effect of aerosol hygroscopicity and environmental RH on the backscattering coefficient (β)

$$\Delta\beta = \begin{cases} 0, & \mathrm{RH} < 40\% \\ \beta - \beta_{\mathrm{dry}}, & \mathrm{RH} \geq 40\%. \end{cases} \tag{2}$$

As mentioned earlier, ambient T and humidity directly affect ALWC. Since RH is a key factor affecting ALWC, RH and T are two potential features to be considered. ALWC affects the optical and microphysical properties of aerosol particles, reflected by the parameters DR and β. They, respectively, represent the sphericity of aerosol particles and the light-blocking ability of aerosol particles. WS and WD affect aerosol transport, connected with aerosol sources. PL refers to the concentration of environmental pollutants, which affects the hygroscopicity of aerosols, so are also considered. The height difference (H) between the ground and the tower is also taken into account. The correlation coefficients between RH and ALWC and between $\Delta\beta$ and ALWC are 0.470 and 0.498, respectively. This suggests that other factors should be considered when estimating ALWC, such as meteorology and the environmental PL. Nine characteristic variables that may affect the ALWC (i.e., RH, T, DR, $\Delta\beta$, β, PL, WS, WD, and H) were selected as potential factors influencing ALWC.

B. Model Selection and Feature Importance

Some of the selected potential features may have nonlinear relationships with ALWC. The GBDT ML model can more effectively distinguish multiple features and combine them when dealing with regression problems [46]. When dealing with regression problems, GBDT has high accuracy and strong robustness. The GBDT model is an iterative decision tree algorithm consisting of multiple trees. The conclusions of all trees are added together to determine the final variable selection [47]. The purpose is to iteratively find the decision tree and make the loss of samples as small as possible. Friedman [48], [49] used the value of the negative gradient of the loss function in the GBDT model as an approximate value of the residual of the boosting tree algorithm in the regression problem to fit a regression tree. The basic flow of this method is given as follows.

Input: Training data (see Table III), $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$.

Output: Regression boosted tree model, $\hat{f}(x)$.

1) Initialize the weak learner, and estimate the constant value that minimizes the loss function, $f_0(x)$

$$f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L_F(y_i, \gamma). \tag{3}$$

2) Calculate the value of the negative gradient of the loss function in the current model as the residual estimate formula (4); estimate the regression leaf node area, and fit the residual approximate value formula (5); and use a linear search to estimate the value of the leaf node area, minimize the loss function, and update the regression tree formula (6)

$$r_{im} = -\left[\frac{\partial L_F(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}} \tag{4}$$

$$r_{jm} = \text{argmin}_\gamma \sum_{x_i \in R_{jm}} L_F(y_i, f_{m-1}(x_i + \gamma)) \tag{5}$$

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}). \tag{6}$$

3) Output the final model

$$\hat{f}(x) = f_M(x). \tag{7}$$

where r_{im} represents the negative gradient of the loss function of the ith sample in the mth round; r_{jm} is the residuals where j is the number of leaf nodes of the regression tree m, R_{jm} is the regression leaf node area, x_i and y_i are the samples for $i = 1, 2, 3 \ldots$, and $L_F(y, \gamma)$ is the least-squares loss function, $L_F(y, \gamma) = (y - \gamma)^2$.

In addition to the GBDT algorithm, there are other commonly used ML algorithms that can deal with similar regression problems. The performance of the GBDT algorithm was, thus, compared with those of several popular ML methods. One ML algorithm is the backpropagation neural network (BPNN); the NN is a fully connected NN and uses three hidden layers and 250, 150, and 50 nodes in each respective layer after optimizing the parameters [50]. The random forest (RF) and the support vector machine (SVM) are two other algorithms chosen for comparison. The RF algorithm is a novel, ensemble ML technique, which can handle continuous and discrete variables and runs efficiently when handling large datasets [51]. The SVM has been employed widely in different classification and regression problems because of its effectiveness in working with nonseparable and high-dimensional datasets [52]. The same nine selected potential features were used as inputs to the models. Statistical metrics, such as the root-mean-square error (RMSE), the mean absolute error (MAE), the root-mean-square-percentage error (RMSPE), and the coefficient of determination (R^2), were also used to evaluate the results calculated at the ground and 532-m tower level. Table II gives details of the validation results of these methods. Among these methods, the GBDT has the best performance in terms of RMSE (3.40), R^2 (0.86), MAE (1.95), and RMSRE (0.61). This means that, when dealing with our regression problem, encountered in this study, GBDT has stronger robustness, higher accuracy, and smaller errors than other models. Therefore, we choose the GBDT algorithm to retrieve the ALWC.

TABLE II
COMPARISON OF THE PERFORMANCES OF DIFFERENT METHODS

Method	RMSE	R^2	MAE	RMSRE
BPNN	4.59	0.74	3.02	1.12
GBDT	3.40	0.86	1.95	0.61
RF	3.89	0.81	2.32	0.81
SVM	3.99	0.80	2.15	1.16

To select the most suitable features for predicting ALWC, the correlation coefficient from linear regression between each of the nine potential feature variables and ALWC, and the feature importance of each parameter to aerosol liquid water were determined [47]. The feature importance score (FIS) J is the average value of the importance of feature J in a single tree [48]

$$\hat{J}_i^2 = \frac{1}{M} \sum_{m=1}^{M} \hat{J}_i^2(T_m) \tag{8}$$

where M is the number of trees. The importance of feature J in a single tree is

$$\hat{J}_i^2(T) = \sum_{t=1}^{L-1} \hat{i}_t^2 1(v_t = j) \tag{9}$$

where L is the number of leaf nodes of the single classification and regression tree (CART), $L - 1$ is the number of nonleaf nodes of the tree, v_t is the feature associated with node t, and $\hat{i}_t^2$ is the reduction value of the square loss after node t is split. In the process of model training, noise information is randomly added. If a certain feature causes a large error in the result after adding the noise information, it means that this feature has a greater impact on the target, and the higher the importance score.

By calculating the importance of features, the FIS was obtained (see Fig. 2). Results show that $\Delta\beta$ is the most important variable, contributing as much as 32.9% to ALWC and positively correlated with it ($R = 0.498$). This means that the contribution of the parameter $\Delta\beta$ in all features to the final prediction result is 32.9%. Although β (R = 0.476; FIS = 13.2%) and PL (R = 0.361; FIS = 10.7%) are positively correlated and have high correlation coefficients, their importance is much lower than $\Delta\beta$. The reason may be that these two parameters have an indirect impact on ALWC, noting that the feature importance index indicates the impact of each feature on the decision tree rather than on the real physical mechanism [24]. Both DR (R = −0.1; FIS = 5.1%) and H (R = −0.244; FIS = 12.39%) are negatively correlated with ALWC. On the one hand, this shows that, as the ALWC increases, aerosol particles gradually become regular (DR decreases). On the other hand, as the altitude increases, the aerosol concentration and water vapor content gradually decrease, and the liquid water content decreases. The lower feature importance of DR may be caused by the relatively regularly shaped aerosol particles in the Guangzhou area [53], the smaller proportion of irregularly shaped particles, and the

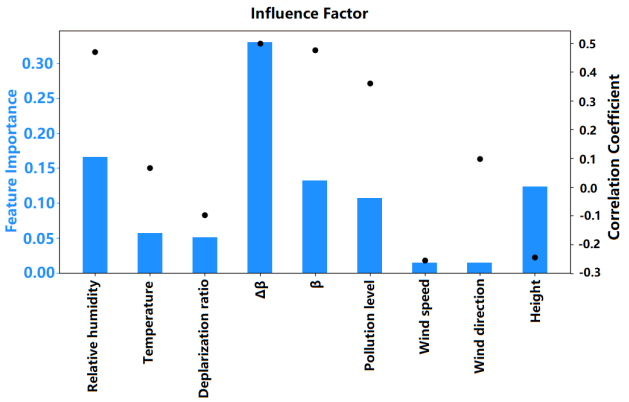

Fig. 2. Selected independent features and FISs in the GBDT model for estimating ALWC.

TABLE III
LIST OF INPUT PARAMETERS

Input parameters	Symbol	Used for training	Used for testing
Relative humidity	RH	80% of all features observed on the ground level	Remaining 20% of all features observed on the ground
Temperature	T		
Depolarization ratio	DR		
Difference in backscattering coefficient between wet and dry	$\Delta\beta$		
Backscattering coefficient	β	80% of all features observed on the tower level	Remaining 20% of all features observed on the tower
Pollution level	PL		
Hight	H		

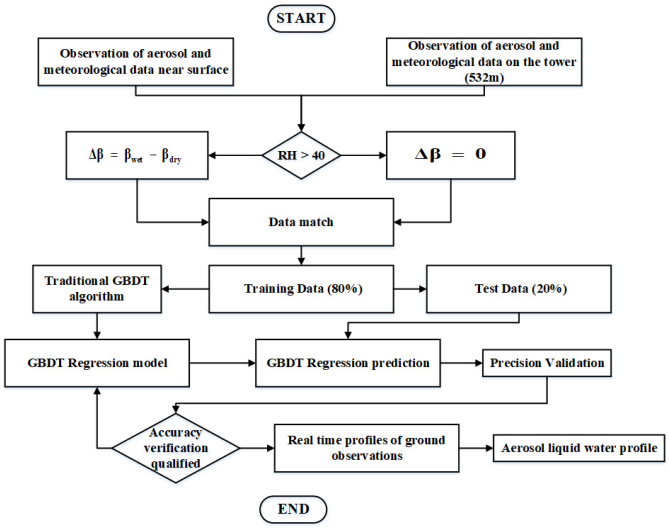

Fig. 3. Flowchart showing the main processes in calculating the aerosol liquid water profile.

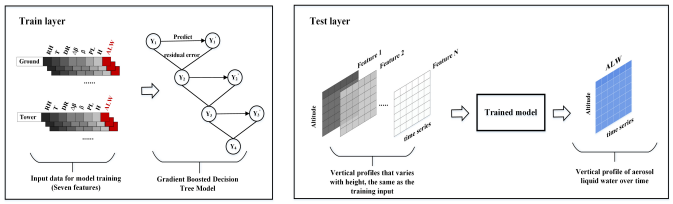

Fig. 4. Main structure of the GBDT model and the processes of training and testing.

smaller DR range. Among the meteorological factors, RH has the greatest influence on ALWC (R = 0.47; FIS = 16.6%), and T has less influence (R = 0.06; FIS = 5.7%). The contribution of WS (1.5%) and WD (1.5%) to the estimation of ALWC is low. Although the tower and ground sites are ∼10 km apart, both sites are located in an urban area where the air is well mixed [54]. This may explain why the wind factors have low contributions. They are, thus, excluded from the model. The remaining seven parameters are used in the GBDT model (see Table III for details).

C. Overview of the ALWC Vertical Profile Estimation Process and the Evaluation Method

Fig. 3 presents the flowchart of estimating the ALWC vertical profile. ALWC data observed at the ground level and on the tower level were first acquired. In the model training and verification stage, since the observed ALWC only has two height values, the data corresponding to these two heights of the input profile parameters match the ALWC data in time. The observed β was distinguished according to the ambient humidity, and changes in β due to the influence of hygroscopic growth were estimated. These data were then spatiotemporally matched. To ensure that the amount of data was sufficient, the time resolution of the training data was unified to 15 min. Note that there is an imbalance between the amount of ground observation data and the amount of observation data from the tower. To ensure that the number of data samples was uniform when training the model, the data matching adopted a downsampling method [55], resulting in a total data volume of 4500. Data from the tower and at the ground level were both randomly divided into two datasets: 80% (3600 data samples) and 20% (900 data samples) of the total data volume were used for training and validating the GBDT model, respectively. To show the specific modeling process more clearly, Fig. 4 describes the training and testing parts of the model in more detail. In the training layer, multiple tree models are built, and each tree model fits the calculated residuals, so as to continuously reduce the error of the entire model and obtain the best results. After validating the model estimates, the profile data corresponding to the input features are input into the trained model. Driven by the vertical variations of the profile data of various input parameters, the derived ALWC also varies with height, as shown in the test layer in Fig. 4. At this point, the inversion of the ALWC profile is completed.

To evaluate the prediction performance of the constructed model, the performance of the trained model with new data as input needed assessment, in particular, with regard to overfitting [56]. After the model training was completed, cross-validation (CV) using the k-fold CV (k-CV) method used to judge the stability of the model was done. This validation method randomly divides the original training dataset into k parts, using one part as the test set and the other $k-1$ parts as the training set [57]. This process loops k times, and the

average accuracy of the k times is used as the performance index of the model. In this study, the value of k was 10.

IV. Results and Discussion

A. Validation of the GBDT Model for Predicting ALWC Based on Multiple Observations

As previously described, the remaining 20% of the observation data (half from the ground and half from the tower) not used to train the model was first used to verify the model's prediction results. To verify the estimation results more comprehensively, three validation exercises were carried out: comparison validations of all validation data, ground validation data only, and tower validation data only. Fig. 5(a) compares observed and estimated ALWCs for all data, showing good agreement, with R^2, RMSE, RMSPE, and MAE being 0.870, 3.28 $\mu g \cdot m^{-3}$, 0.57, and 1.94 $\mu g \cdot m^{-3}$, respectively. The slope of the best-fit line from linear regression is 0.86, suggesting a slight underestimation of the estimated ALWC. Most of the higher ALWC values are associated with higher RH levels. Fig. 5(b) presents the histogram of retrieval errors, which are small and concentrated around zero, with a mean value of 0.1 and a median value of −0.1. Similar comparisons were made for the subsets of data acquired on the ground [see Fig. 5(c) and (d)]. As with the comparison using all data, small biases were also seen here.

The retrieval accuracy based on tower data [see Fig. 5(e) and (f)] was overall reduced, with R^2, RMSE, RMSPE, and MAE values of 0.776, 2.18 $\mu g \cdot m^{-3}$, 0.68, and 1.5 $\mu g \cdot m^{-3}$, respectively. The value of ALWC on the tower (at 532 m) was generally low (<30 $\mu g \cdot m^{-3}$). The performance of the model for different ALWC ranges was assessed, categorized as follows: I (0–10 $\mu g \cdot m^{-3}$), II (0–20 $\mu g \cdot m^{-3}$), III (0–30 $\mu g \cdot m^{-3}$), and IV (0–80 $\mu g \cdot m^{-3}$) (see Fig. 6). The accuracy of ALWC retrievals at both sites gradually increased as the range increased, becoming gradually consistent. The poor performance in the lowest range of ALWC values was mainly due to fewer aerosols with low ALWC [58]. In this case, aerosols played a much lesser role than did RH, incurring more uncertainty in the retrievals as echoed by the dispersion of the validation data points.

Tenfold CV was done to test the degree of stability of the model and the reliability of the results. The average accuracy was 0.82, suggesting that the algorithm is stable.

B. ALWC Vertical Distribution

Using the aforementioned model and profile data of the seven features listed in Table III, ALWCs estimated in the vertical at the ground site were inverted. Fig. 7 shows the time series of the vertical profiles of the four parameters at the Guangzhou Meteorological Bureau ground site in November 2019, with a time resolution of 1 h. Not included are data with high signal-to-noise ratios. Fig. 7(a) shows the inverted vertical ALWC distribution, with the black line showing the height of the planetary boundary layer (PBL) calculated by the gradient method [59], [60]. Aerosol liquid water was mainly concentrated below 1.5 km, which is where the bulk of aerosols resided. The height of the boundary layer in autumn

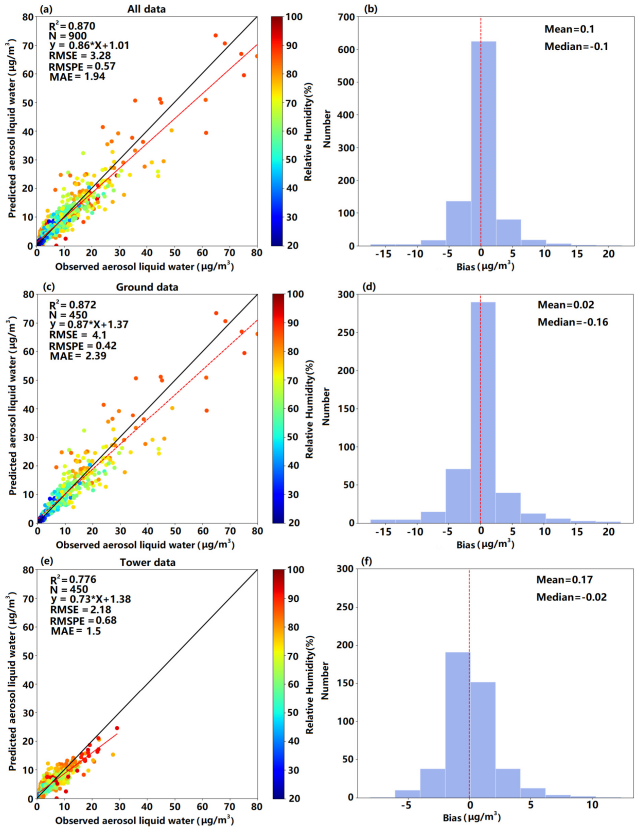

Fig. 5. Estimated ALWC as a function of observed ALWC for (a) all data, (c) ground data only, and (e) tower data only, colored according to RH. The solid black line is the 1:1 line, and the red dashed line is the best-fit line from linear regression. (Right) Histograms of the corresponding biases. R^2: coefficient of determination; N: number of samples; RMSE: root-mean-square error ($\mu g \cdot m^{-3}$); RMSPE: root-mean-square-percentage error ($\mu g \cdot m^{-3}$); and MAE: mean absolute error ($\mu g \cdot m^{-3}$).

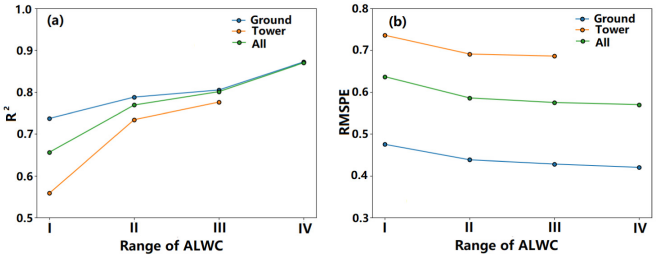

Fig. 6. (a) Coefficient of determination (R^2) and (b) RMSPE of ALWC retrievals for different ranges of ALWC: I (0–10 $\mu g \cdot m^{-3}$), II (0–20 $\mu g \cdot m^{-3}$), III (0–30 $\mu g \cdot m^{-3}$), and IV (0–80 $\mu g \cdot m^{-3}$).

and winter in Guangzhou ranges from 0.5 to 1.1 km [61]. In most cases, higher ALWCs were distributed within the PBL, a part of the atmosphere that is generally more affected by human activities [62]. Fig. 7(b)–(d) shows the time series of the vertical distributions of RH, $\Delta\beta$, and DR, respectively. Some high values of liquid water content are associated with large changes in the backscattering coefficient and relatively low aerosol volume DRs. This suggests, to a certain extent, that the amount of aerosol liquid water affects the morphology of aerosol particles [21]. Overall, missing data aside, continuous vertical profiles of ALWC are achievable.

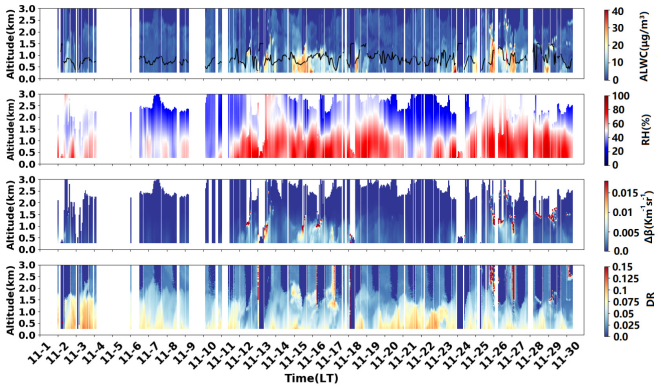

Fig. 7. Time series of vertical (a) ALWC (the input data of the near ground and the corresponding height on the tower to obtain the vertical time series of ALWC are also used as input during model training). (b) RH. (c) $\Delta\beta$ and (d) DR in the observation site of the Guangzhou Meteorological Bureau ground site in November 2019. The black line in (a) shows the height of the boundary layer. Blank sections are missing data or data that are excluded because of their high signal-to-noise ratios.

Fig. 8 shows the distributions of ALWC profiles under four pollution conditions. The PBL height (PBLH) gradually decreased as the PL increased. Under relatively clean atmospheric conditions, the average boundary layer height was about 1 km [see Fig. 8(a)], and when the air pollution was severe, the average boundary layer height was about 0.7 km [see Fig. 8(d)]. When the atmosphere was clean [see Fig. 8(a)], the overall ALWC level was low, and the change with altitude was small. Compared with the other scenarios, there was a weak peak at 1.5 km, with a large standard deviation. This may be because, when the aerosol loading is low, the high RH of the environment provided by low clouds becomes the main factor affecting aerosol liquid water, increasing the uncertainty [63]. For moderate and sensitive PLs [see Fig. 8(b) and (c)], changes in the vertical distribution of the ALWC were roughly the same. The average liquid water content near the ground was $\sim 10\ \mu g \cdot m^{-3}$. When pollution was severe [see Fig. 8(d)], the degree of change in ALWC with height was the greatest, and the average liquid water content near the ground was highest ($\sim 20\ \mu g \cdot m^{-3}$).

C. Influence of the Proportion of ALWC on the Morphology of Aerosol Particles

As shown in Fig. 7, the larger the proportion of the backscattering coefficient caused by aerosol hygroscopicity and the higher the RH of the environment, the more the aerosol particles tended to be spherical. To more clearly understand the influence of the proportion of liquid water in the backscattering coefficient on the morphology of aerosol particles, the fraction of liquid water and the volume DR of aerosols were examined for different levels of pollution within and outside the boundary layer (at 0.5 and 1 km, respectively, selected according to Fig. 8) (see Fig. 9). On the whole, aerosol particles within the boundary layer were more irregularly shaped. The RH of the environment is one of the key factors affecting the proportion of liquid water. A higher proportion of liquid water corresponds to a higher ambient RH. When the air was clean during the observation period [see Fig. 9(a)], there was no direct relationship between the proportion of liquid

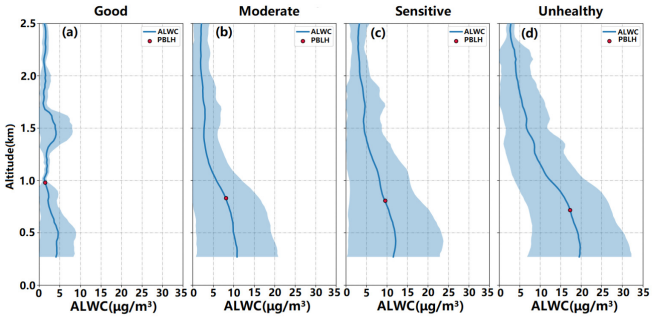

Fig. 8. Vertical distributions of mean ALWC profiles (blue lines) in the observation site at the Guangzhou Meteorological Bureau ground site under different pollution conditions during the observation period. The shaded blue areas represent the standard deviations of the vertical distributions. The red dots represent the mean PBLH. (a)–(d) Four levels of air quality: good, moderate, sensitive, and unhealthy, respectively.

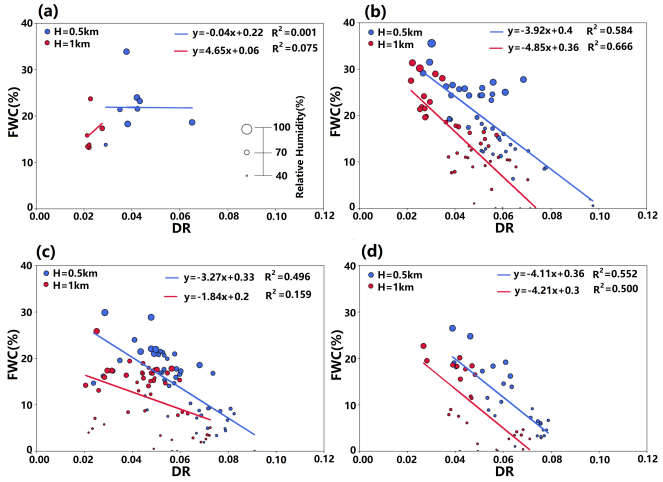

Fig. 9. Relationship between the fraction of backscattering caused by aerosol liquid water in the total backscattering coefficient (FWC) and the volume DR. (a)–(d) Four levels of air quality: good, moderate, sensitive, and unhealthy, respectively. The colors correspond to two different heights, i.e., 0.5 (blue) and 1 km (red). The lines are the best-fit lines from linear regression, and the different sizes of the dots reflect the magnitudes of the corresponding environmental RH levels. R^2: coefficient of determination.

water and the shape of aerosol particles, likely due to the low aerosol loading and smaller sample size. When the aerosol loading was higher [see Fig. 9(b)–(d)], the fraction of ALWC was negatively correlated with the volume DR of aerosols [64]. As the pollution increased, the range of variation of DR of aerosol particles gradually becomes smaller. The reason for this phenomenon may be that, when the pollution was severe [see Fig. 9(b)–(d)], the concentration of aerosol particles per unit volume was higher, and the shape of aerosol particles was more uniform [65]. This phenomenon is manifested both inside and outside the boundary layer, suggesting that human activities will also have a certain impact on the characteristics of aerosols higher in the atmosphere [66]. Results show that the proportion of backscattering coefficient caused by aerosol liquid water during winter pollution events in the Guangzhou area increased by about 10%, and the DR decreased by 0.02.

V. Conclusion

This study developed a new approach for estimating ALWC vertical profiles. The ALWC is associated with aerosol hygroscopicity and loading, RH, and other meteorological and

environmental factors, with complex nonlinear relationships. To date, the vast majority of ALWC-related studies focus on the surface *in situ* measurements. This new remote sensing method estimates vertical profiles of ALWC based on aerosol information obtained by a Mie LiDAR, combined with meteorological profile data and hygroscopicity information, and incorporated into the GBDT model.

The development and validation of the model benefited greatly from a field experiment conducted at ground and tower (532 m) sites in Guangzhou, China, in spring 2020. The sites were equipped with a large number of state-of-the-art instruments acquiring all pertinent measurements. The model was proven to have high accuracy and high stability. According to the derived vertical distribution of ALWC, ALWC was mainly distributed within the boundary layer. Overall, ALWC gradually decreased with height. As air pollution becomes more severe, the rate at which ALWC decreased with height gradually increased. Increasing ALWC also made aerosol particles more spherical.

The retrieved vertical profiles of ALWC do not only provide richer information about the atmospheric environment but also help improve our understanding of the influence of ALWC on aerosol optical parameters in the vertical direction, atmospheric chemistry, secondary aerosol formation, and global environmental changes.

REFERENCES

[1] *The Physical Science Basis, Contribution Working Group I to Fifth Assessment Report Intergovernmental Panel on Climate Change*, IPCC: Climate Change, Geneva, Switzerland, 2013, doi: 10.1038/446727a.

[2] B. J. Krueger, "The transformation of solid atmospheric particles into liquid droplets through heterogeneous chemistry: Laboratory insights into the processing of calcium containing mineral dust aerosol in the troposphere," *Geophys. Res. Lett.*, vol. 30, no. 3, p. 1148, 2003, doi: 10.1029/2002GL016563.

[3] T. K. V. Nguyen, Q. Zhang, J. L. Jimenez, M. Pike, and A. G. Carlton, "Liquid water: Ubiquitous contributor to aerosol mass," *Environ. Sci. Technol. Lett.*, vol. 3, no. 7, pp. 257–263, Jul. 2016, doi: 10.1021/acs.estlett.6b00167.

[4] X. Pan *et al.*, "Real-time observational evidence of changing Asian dust morphology with the mixing of heavy anthropogenic pollution," *Sci. Rep.*, vol. 7, no. 1, p. 335, Dec. 2017, doi: 10.1038/s41598-017-00444-w.

[5] C. J. Hennigan, M. H. Bergin, J. E. Dibb, and R. J. Weber, "Enhanced secondary organic aerosol formation due to water uptake by fine particles," *Geophys. Res. Lett.*, vol. 35, no. 18, 2008, Art. no. L18801, doi: 10.1029/2008GL035046.

[6] X. Jin *et al.*, "Significant contribution of organics to aerosol liquid water content in winter in Beijing, China," *Atmos. Chem. Phys.*, vol. 20, no. 2, pp. 901–914, Jan. 2020, doi: 10.5194/acp-20-901-2020.

[7] Y. Kitamori, M. Mochida, and K. Kawamura, "Assessment of the aerosol water content in urban atmospheric particles by the hygroscopic growth measurements in Sapporo, Japan," *Atmos. Environ.*, vol. 43, no. 21, pp. 3416–3423, Jul. 2009, doi: 10.1016/j.atmosenv.2009.03.037.

[8] Y. F. Cheng *et al.*, "Relative humidity dependence of aerosol optical properties and direct radiative forcing in the surface boundary layer at Xinken in Pearl River Delta of China: An observation based numerical study," *Atmos. Environ.*, vol. 42, no. 25, pp. 6373–6397, Aug. 2008, doi: 10.1016/j.atmosenv.2008.04.009.

[9] C. O. Stanier, A. Y. Khlystov, W. R. Chan, M. Mandiro, and S. N. Pandis, "A method for the *in situ* measurement of fine aerosol water content of ambient aerosols: The dry-ambient aerosol size spectrometer (DAASS) special issue of Aerosol science and technology on findings from the fine particulate matter supersites program," *Aerosol Sci. Technol.*, vol. 38, no. 1, pp. 215–228, Dec. 2004, doi: 10.1080/02786820390229525.

[10] A. Nenes, S. N. Pandis, and C. Pilinis, "ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols," *Aquatic Geochem.*, vol. 4, pp. 123–152, Mar. 1998.

[11] A. S. Wexler and J. H. Seinfeld, "Second-generation inorganic aerosol model," *Atmos. Environ. A, Gen. Topics*, vol. 25, no. 12, pp. 2731–2748, Jan. 1991.

[12] H. J. Liu *et al.*, "Aerosol hygroscopicity derived from size-segregated chemical composition and its parameterization in the north China plain," *Atmos. Chem. Phys.*, vol. 14, no. 5, pp. 2525–2539, Mar. 2014, doi: 10.5194/acp-14-2525-2014.

[13] Z. Wu *et al.*, "Aerosol liquid water driven by anthropogenic inorganic salts: Implying its key role in haze formation over the north China plain," *Environ. Sci. Technol. Lett.*, vol. 5, no. 3, pp. 160–166, Mar. 2018, doi: 10.1021/acs.estlett.8b00021.

[14] C. Fountoukis and A. Nenes, "ISORROPIA II: A computationally efficient thermodynamic equilibrium model for K^+-Ca^{2+}-Mg^{2+}-NH_4^+–Na^+-SO_4^{2-}–NO_3–Cl–H_2O aerosols," *Atmos. Chem. Phys.*, vol. 7, pp. 4639–4659, Jan. 2007.

[15] J. Chen *et al.*, "Aerosol hygroscopic growth, contributing factors, and impact on haze events in a severely polluted region in northern China," *Atmos. Chem. Phys.*, vol. 19, no. 2, pp. 1327–1342, Jan. 2019, doi: 10.5194/acp-19-1327-2019.

[16] M. J. Granados-Muñoz *et al.*, "Hygroscopic growth of atmospheric aerosol particles based on active remote sensing and radiosounding measurements: Selected cases in southeastern Spain," *Atmos. Meas. Techn.*, vol. 8, no. 2, pp. 705–718, Feb. 2015, doi: 10.5194/amt-8-705-2015.

[17] M. Lv *et al.*, "Hygroscopic growth of atmospheric aerosol particles based on lidar, radiosonde, and *in situ* measurements: Case studies from the Xinzhou field campaign," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 188, pp. 60–70, Feb. 2017, doi: 10.1016/j.jqsrt.2015.12.029.

[18] Y. Kuang *et al.*, "A novel method for calculating ambient aerosol liquid water content based on measurements of a humidified nephelometer system," *Atmos. Meas. Techn.*, vol. 11, no. 5, pp. 2967–2982, May 2018.

[19] I. Veselovskii, D. N. Whiteman, A. Kolgotin, E. Andrews, and M. Korenskii, "Demonstration of aerosol property profiling by multi-wavelength lidar under varying relative humidity conditions," *J. Atmos. Ocean. Technol.*, vol. 26, no. 8, pp. 1543–1557, Aug. 2009, doi: 10.1175/2009JTECHA1254.1.

[20] T. Wu *et al.*, "Aerosol optical properties over China from RAMS-CMAQ model compared with CALIOP observations," *Atmosphere*, vol. 8, no. 12, p. 201, Oct. 2017, doi: 10.3390/atmos8100201.

[21] H. Tan *et al.*, "An analysis of aerosol liquid water content and related impact factors in Pearl River Delta," *Sci. Total Environ.*, vol. 579, pp. 1822–1830, Feb. 2017, doi: 10.1016/j.scitotenv.2016.11.167.

[22] W. Tan *et al.*, "Profiling aerosol liquid water content using a polarization lidar," *Environ. Sci. Technol.*, vol. 54, no. 6, pp. 3129–3137, Mar. 2020, doi: 10.1021/acs.est.9b07502.

[23] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.2013.

[24] J. Wei *et al.*, "Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111221, doi: 10.1016/j.rse.2019.111221.

[25] J. Wei *et al.*, "Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space–time extremely randomized trees," *Atmos. Chem. Phys.*, vol. 20, no. 6, pp. 3273–3289, Mar. 2020, doi: 10.5194/acp-20-3273-2020.

[26] X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, "Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2303–2310, Sep. 2017, doi: 10.1109/TITS.2016.2635719.2017.

[27] T. Wu *et al.*, "Hygroscopicity of different types of aerosol particles: Case studies using multi-instrument data in megacity Beijing, China," *Remote Sens.*, vol. 12, no. 5, p. 785, Mar. 2020, doi: 10.3390/rs12050785.

[28] M. Tesche *et al.*, "Vertically resolved separation of dust and smoke over Cape Verde using multiwavelength Raman and polarization lidars during Saharan mineral dust experiment 2008," *J. Geophys. Res.*, vol. 114, no. D13, 2009, Art. no. D13202, doi: 10.1029/2009JD011862.

[29] F. G. Fernald, "Analysis of atmospheric lidar observations: Some comments," *Appl. Opt.*, vol. 23, no. 5, p. 652, Mar. 1984, doi: 10.1364/AO.23.000652.

[30] P. Yan, X. Pan, J. Tang, X. Zhou, R. Zhang, and L. Zeng, "Hygroscopic growth of aerosol scattering coefficient: A comparative analysis between urban and suburban sites at winter in Beijing," *Particuology*, vol. 7, no. 1, pp. 52–60, Feb. 2009, doi: 10.1016/j.partic.2008.11.009.

[31] Y. Wang *et al.*, "Enhanced hydrophobicity and volatility of submicron aerosols under severe emission control conditions in Beijing," *Atmos. Chem. Phys.*, vol. 17, no. 8, pp. 5239–5251, Apr. 2017, doi: 10.5194/acp-17-5239-2017.

[32] S. Li *et al.*, "Characterizing the ratio of nitrate to sulfate in ambient fine particles of urban Beijing during 2018–2019," *Atmos. Environ.*, vol. 237, Sep. 2020, Art. no. 117662, doi: 10.1016/j.atmosenv.2020.117662.

[33] G. Massaro, I. Stiperski, B. Pospichal, and M. W. Rotach, "Accuracy of retrieving temperature and humidity profiles by ground-based microwave radiometry in truly complex terrain," *Atmos. Meas. Techn.*, vol. 8, no. 8, pp. 3355–3367, Aug. 2015, doi: 10.5194/amt-8-3355-2015.

[34] J.-P. Guo *et al.*, "Correlation between PM concentrations and aerosol optical depth in eastern China," *Atmos. Environ.*, vol. 43, no. 37, pp. 5876–5886, Dec. 2009, doi: 10.1016/j.atmosenv.2009.08.026.

[35] Z. Li, X. Yang, C. Zhao, and T. Fan, "Ratio of $PM_{2.5}$ to PM_{10} mass concentrations in Beijing and relationships with pollution from the north China plain," *Asia–Pacific J. Atmos. Sci.*, vol. 57, no. 3, pp. 421–434, Aug. 2021, doi: 10.1007/s13143-020-00203-4.

[36] M. L. Chanin, A. Garnier, A. Hauchecorne, and J. Porteneuve, "A Doppler lidar for measuring winds in the middle atmosphere," *Geophys. Res. Lett.*, vol. 16, no. 11, pp. 1273–1276, Nov. 1989, doi: 10.1029/GL016i011p01273.

[37] C. O. Hines, "Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 1: Basic formulation," *J. Atmos. Solar-Terr. Phys.*, vol. 59, no. 4, pp. 371–386, doi: 10.1016/S1364-6826(96)00079-X.1997.

[38] Y. Kuang, C. S. Zhao, J. C. Tao, Y. X. Bian, and N. Ma, "Impact of aerosol hygroscopic growth on the direct aerosol radiative effect in summer on north China plain," *Atmos. Environ.*, vol. 147, pp. 224–233, Dec. 2016, doi: 10.1016/j.atmosenv.2016.10.013.

[39] G. Zhao *et al.*, "Impact of aerosol hygroscopic growth on retrieving aerosol extinction coefficient profiles from elastic-backscatter lidar signals," *Atmos. Chem. Phys.*, vol. 17, no. 19, pp. 12133–12143, Oct. 2017, doi: 10.5194/acp-17-12133-2017.

[40] R. Zhang, A. F. Khalizov, J. Pagels, D. Zhang, H. Xue, and P. H. Mcmurry, "Variability in morphology, hygroscopicity, and optical properties of soot aerosols during atmospheric processing," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 30, pp. 10291–10296, Jul. 2008, doi: 10.1073/pnas.0804860105.

[41] P. F. Liu *et al.*, "Hygroscopic properties of aerosol particles at high relative humidity and their diurnal variations in the north China plain," *Atmos. Chem. Phys.*, vol. 11, no. 7, pp. 3479–3494, Apr. 2011, doi: 10.5194/acp-11-3479-2011.

[42] A. E. Bedoya-Velásquez *et al.*, "Hygroscopic growth study in the framework of EARLINET during the SLOPE I campaign: Synergy of remote sensing and *in situ* instrumentation," *Atmos. Chem. Phys.*, vol. 18, no. 10, pp. 7001–7017, May 2018, doi: 10.5194/acp-18-7001-2018.

[43] H. Eichler *et al.*, "Hygroscopic properties and extinction of aerosol particles at ambient relative humidity in South-Eastern China," *Atmos. Environ.*, vol. 42, no. 25, pp. 6321–6334, Aug. 2008, doi: 10.1016/j.atmosenv.2008.05.007.

[44] R. M. Hoff *et al.*, "Lidar, nephelometer, and *in situ* aerosol experiments in southern Ontario," *J. Geophys. Research: Atmos.*, vol. 101, no. D14, pp. 19199–19209, Aug. 1996.

[45] M. Pahlow *et al.*, "Comparison between lidar and nephelometer measurements of aerosol hygroscopicity at the southern great plains atmospheric radiation measurement site," *J. Geophys. Res.*, vol. 111, no. D5, 2006, Art. no. D05S15.

[46] G. Forkuor, O. K. L. Hounkpatin, G. Welp, and M. Thiel, "High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0170478, doi: 10.1371/journal.pone.0170478.

[47] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Statist.*, vol. 28, no. 2, pp. 337–407, 2000.

[48] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[49] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002, doi: 10.1016/S0167-9473(01)00065-2.2002.

[50] S. A. Bogdanov and O. S. Sidelnikov, "Use of complex fully connected neural networks to compensate for nonlinear effects in fibre-optic communication lines," *Quantum Electron.*, vol. 51, no. 5, pp. 459–462, May 2021, doi: 10.1070/QEL17518.

[51] I. A. Ibrahim and T. Khatib, "A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm," *Energy Convers. Manage.*, vol. 138, pp. 413–425, Apr. 2017, doi: 10.1016/j.enconman.2017.02.006.

[52] B. Kalantar, B. Pradhan, S. A. Naghibi, A. Motevalli, and S. Mansor, "Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)," *Geomatics, Natural Hazards Risk*, vol. 9, no. 1, pp. 49–69, 2018, doi: 10.1080/19475705.2017.1407368.

[53] T. Nishizawa *et al.*, "Vertical distribution of water-soluble, sea salt, and dust aerosols in the planetary boundary layer estimated from two-wavelength backscatter and one-wavelength polarization lidar measurements in Guangzhou and Beijing, China," *Atmos. Res.*, vol. 96, no. 4, pp. 602–611, Jun. 2010, doi: 10.1016/j.atmosres.2010.02.002.

[54] J. Duan, X. Bi, J. Tan, G. Sheng, and J. Fu, "Seasonal variation on size distribution and concentration of PAHs in Guangzhou city, China," *Chemosphere*, vol. 67, no. 3, pp. 614–622, Mar. 2007, doi: 10.1016/j.chemosphere.2006.08.030.

[55] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 539–550, Apr. 2009, doi: 10.1109/TSMCB.2008.2007853.

[56] X. Yan, C. Liang, Y. Jiang, N. Luo, Z. Zang, and Z. Li, "A deep learning approach to improve the retrieval of temperature and humidity profiles from a ground-based microwave radiometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8427–8437, Dec. 2020, doi: 10.1109/TGRS.2020.2987896.2020.

[57] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010, doi: 10.1109/TPAMI.2009.187.2010.

[58] E. O. Edney *et al.*, "Impact of aerosol liquid water on secondary organic aerosol yields of irradiated toluene/propylene/NO_x/$(NH_4)_2SO_4$/air mixtures," *Atmos. Environ.*, vol. 34, no. 23, pp. 3907–3919, Jan. 2000.

[59] L. Menut, C. Flamant, J. Pelon, and P. H. Flamant, "Urban boundary-layer height determination from lidar measurements over the Paris area," *Appl. Opt.*, vol. 38, no. 6, p. 945, Feb. 1999, doi: 10.1364/AO.38.000945.

[60] D. J. Seidel, C. O. Ao, and K. Li, "Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis," *J. Geophys. Res.*, vol. 115, no. D16, 2010, Art. no. D16113, doi: 10.1029/2009JD013680.

[61] D. Wu *et al.*, "An extremely low visibility event over the Guangzhou region: A case study," *Atmos. Environ.*, vol. 39, no. 35, pp. 6568–6577, Nov. 2005, doi: 10.1016/j.atmosenv.2005.07.061.

[62] N. Bellouin *et al.*, "Bounding global aerosol radiative forcing of climate change," *Rev. Geophys.*, vol. 58, no. 1, Mar. 2020, Art. no. e2019RG000660, doi: 10.1029/2019RG000660.

[63] A. Khain, D. Rosenfeld, and A. Pokrovsky, "Aerosol impact on the dynamics and microphysics of deep convective clouds," *Quart. J. Roy. Meteorolog. Soc.*, vol. 131, no. 611, pp. 2639–2663, Oct. 2005, doi: 10.1256/qj.04.62.

[64] C. M. Carrico, "Mixtures of pollution, dust, sea salt, and volcanic aerosol during ACE-Asia: Radiative properties as a function of relative humidity," *J. Geophys. Res. Atmos.*, vol. 108, no. D23, p. 8650, 2003, doi: 10.1029/2003JD003405.2003.

[65] H. Volten *et al.*, "Scattering matrices of mineral aerosol particles at 441.6 nm and 632.8 nm," *J. Geophys. Res., Atmos.*, vol. 106, no. D15, pp. 17375–17401, Aug. 2001, doi: 10.1029/2001JD900068.

[66] J. G. Canadell *et al.*, "Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 47, pp. 18866–18870, Nov. 2007, doi: 10.1073/pnas.0702737104.

Tong Wu received the B.Sc. and M.Sc. degrees from the Shandong University of Science and Technology, Qingdao, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree in global environmental change with Beijing Normal University, Beijing, China.

His research interests include ground-based aerosol LiDAR remote sensing and aerosol vertical hygroscopic growth characteristics.

Zhanqing Li received the B.Sc. and M.Sc. degrees from the Nanjing University of Information Science and Technology, Nanjing, China, in 1983 and 1989, respectively, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 1991.

He is currently a Professor with the University of Maryland, College Park, MD, USA. He has authored over 360 articles. His research interests include remote sensing, atmospheric physics, and climate and environment focusing on aerosol, cloud, radiation budget, and precipitation.

Dr. Li is also a fellow of the American Association for Aerosol Research (AAAR), the American Geophysical Union (AGU), and the American Meteorological Society (AMS). He was a recipient of numerous awards in the U.S., Canada, and Germany.

Xiaoai Jin received the bachelor's degree in atmospheric science from Lanzhou University, Lanzhou, China, in 2016. She is currently pursuing the Ph.D. degree in global environmental change with Beijing Normal University, Beijing, China.

Her research interests include atmospheric aerosol hygroscopicity and its impact on atmospheric pollution and aerosol optical properties.

Wei Wang received the B.Sc. degree from Lanzhou University, Lanzhou, China, in 2014. He is currently pursuing the Ph.D. degree in global environmental change with Beijing Normal University, Beijing, China.

His research interest includes multi-instrument joint inversion of vertical temperature and humidity profile.

Hao Wu received the bachelor's degree in environmental science from the Nanjing University of Information Science and Technology, Nanjing, China, in 2009, the M.S. degree from the Chengdu University of Information Technology, Chengdu, China, in 2016, and the Ph.D. degree from Beijing Normal University, Beijing, China, in 2020.

He is currently an Assistant Professor with the Chengdu University of Information Technology. His research interests include atmospheric physics and environmental science, focusing on aerosol, turbulence, greenhouse gas, and atmospheric boundary layer research.

Rongmin Ren received the B.Sc. degree from Zhejiang Agriculture and Forestry University, Hangzhou, China, in 2018, and the M.Sc. degree from Beijing Normal University, Beijing, China, in 2021.

Her research interest is in atmospheric aerosol hygroscopicity.

Dongmei Zhang received the bachelor's degree from Henan University, Kaifeng, China, in 2019. She is currently pursuing the master's degree in global environmental change with Beijing Normal University, Beijing, China.

Her research interests include ambient ultrafine particles, particle size distribution, and source apportionment of particle number concentration.

Lu Chen received the B.Sc. degree from the Nanjing University of Information Science and Technology, Nanjing, China, in 2018, and the M.Sc. degree from Beijing Normal University, Beijing, China, in 2021, where she is currently pursuing the Ph.D. degree in global environmental change.

Her research interests include regional climate models and atmospheric aerosol research.

Yunfei Su received the bachelor's degree in physics science from Beijing Normal University, Beijing, China, in 2017, where she is currently pursuing the Ph.D. degree in global environmental change.

Her research interest includes cloud remote sensing by ground-based and satellite and cloud precipitation research.

Maureen Cribb received the M.Sc. degree in atmospheric science from Dalhousie University, Halifax, NS, Canada, in 1999.

She is currently a Senior Faculty Specialist with the Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA, providing various research supports to Dr. Li's Team.