A novel method for calculating ambient aerosol liquid water content based on measurements of a humidified nephelometer system

Ye Kuang[1], ChunSheng Zhao[2], Gang Zhao[2], JiangChuan Tao[1], Wanyun Xu[3], Nan Ma[1], YuXuan Bian[3]

[1]{Institute for Environmental and Climate Research, Jinan University, Guangzhou 511443, China}

[2]{Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China}

[3]{State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences}

Correspondence to: C. S. Zhao (zcs@pku.edu.cn)

Abstract

Water condensed on ambient aerosol particles plays significant roles in atmospheric environment, atmospheric chemistry and climate. Before now, no instruments were available for real-time monitoring of ambient aerosol liquid water contents (ALWC). In this paper, a novel method is proposed to calculate ambient ALWC based on measurements of a three-wavelength humidified nephelometer system, which measures aerosol light scattering coefficients and backscattering coefficients at three wavelengths under dry state and different relative humidity (RH) conditions, providing measurements of light scattering enhancement factor $f(\mathrm{RH})$. The proposed ALWC calculation method includes two steps. The first step is the estimation of the dry state total volume concentration of ambient aerosol particles, $V_a(\mathrm{dry})$, with a machine learning method called random forest model based on measurements of the "dry" nephelometer. The estimated $V_a(\mathrm{dry})$ agrees well with the measured one. The second step is the estimation of the volume growth factor $\mathrm{Vg(RH)}$ of ambient aerosol particles due to water uptake, using $f(\mathrm{RH})$ and Ångstr$\ddot{o}$m exponent. The ALWC is calculated from the estimated $V_a(\mathrm{dry})$ and $\mathrm{Vg(RH)}$. To validate the new method, the ambient ALWC calculated from measurements of the humidified nephelometer system during the Gucheng campaign was compared with ambient ALWC calculated from ISORROPIA thermodynamic model using aerosol chemistry data. A good agreement was achieved, with a slope and intercept of 1.14 and -8.6 $\mu m^3/cm^3$ (r^2=0.92), respectively. The advantage of this new method is that the ambient ALWC can be obtained solely based on measurements of a three-wavelength humidified nephelometer system, facilitating the real-time monitoring of the ambient ALWC and promoting the study of aerosol liquid water and its role in atmospheric chemistry, secondary aerosol formation and climate change.

1. Introduction

Atmospheric aerosol particles play significant roles in atmospheric environment, climate, human health and the hydrological cycle, and have received much attention in recent decades. One of the most important constituents of ambient atmospheric aerosol is liquid water. The content of condensed water on ambient aerosol particles depends mostly on the aerosol hygroscopicity and the ambient relative humidity (RH). Results of previous studies demonstrate that liquid water contributes greatly to the total mass of ambient aerosol particles when the ambient RH is higher than 60% (Bian et al., 2014). Aerosol liquid water also has large impacts on aerosol optical properties and aerosol radiative effects (Tao et al., 2014;Kuang et al., 2016). Liquid water condensed on aerosol particles can also serves as a site for multiphase reactions which perturb local chemistry and further influence the aging processes of aerosol particles (Martin, 2000). Recent studies have shown that aerosol liquid water serves as a reactor, which can efficiently transform sulphur dioxide to sulphate during haze events, aggravating atmospheric environment in the North China Plain (NCP) (Wang et al., 2016;Cheng et al., 2016). Hence, to gain more insight into the role of aerosol liquid water in atmospheric chemistry, aerosol aging processes and aerosol optical properties, the real-time monitoring of ambient aerosol liquid water content (ALWC) is of crucial importance.

Few techniques are currently available for measuring the ALWC. The humidified tandem differential mobility analyser systems (HTDMAs) are useful tools and widely used to measure hygroscopic growth factors of ambient aerosol particles (Rader and McMurry, 1986;Wu et al., 2016;Meier et al., 2009). Hygroscopicity parameters retrieved from measurements of HTDMAs can be used to calculate the volume of liquid water. Nevertheless, HTDMAs cannot be used to measure the total aerosol water volume, because they are not capable of measuring the hygroscopic properties of the entire aerosol population. With size distributions of aerosol particles in their ambient state and dry state, the aerosol water volume can be estimated. Engelhart et al. (2011) deployed a Dry-Ambient Aerosol Size Spectrometer to measure the aerosol liquid water content and volume growth factor of fine particulate matter. This system provides only aerosol water content of aerosol particles within certain size range (particle diameter less than 500 nm for the setup of Engelhart et al. (2011)). In addition, in conjunction with aerosol thermodynamic equilibrium models, ALWC can also be estimated with detailed aerosol chemical information. However, simulations of aerosol hygroscopicity and phase state by using thermodynamic equilibrium models are still very complicated even under the thermodynamic equilibrium hypothesis and these models may cause large bias when used for estimating ALWC (Bian et al., 2014).

The idea of using the humidified nephelometer system for the study of aerosol hygroscopicity has already been proposed very early on(Covert et al., 1972). The instrument measures aerosol light scattering coefficient (σ_{sp}) under dry state and different RH conditions, providing information on aerosol light scattering enhancement factor $f(\mathrm{RH})$. One advantage of this method is that it has a fast response time and continuous measurements can be made, facilitating the monitoring of changes in ambient conditions. Another advantage of this method is that it provides information on the overall aerosol hygroscopicity of the entire aerosol population (Kuang et al., 2017a). Both measured σ_{sp} of aerosol particles in dry state and $f(\mathrm{RH})$ vary strongly with parameters of particle number size distribution (PNSD), making it difficult to directly link them with the dry state aerosol particle volume ($V_a(\mathrm{dry})$) and the volume growth factor $\mathrm{Vg(RH)}$ of the entire aerosol population. So far, the ALWC could not be directly estimated based solely on measurements of the humidified nephelometer system. Several studies have shown that given the PNSDs at dry state, an iterative algorithm together with the Mie theory can be used to calculate an overall aerosol hygroscopic growth factor $\mathrm{g(RH)}$ based on measurements of $f(\mathrm{RH})$ (Zieger et al., 2010;Fierz-Schmidhauser et al., 2010). In such an iterative algorithm, the $\mathrm{g(RH)}$ is assumed to be independent of the aerosol diameter. Then ALWC at different RH levels can be calculated based on derived g(RH) and the measured PNSD. This method not only requires additional measurements of PNSD, but also may result in significant deviations of the estimated ALWC, because g(RH) should be a function of aerosol diameter rather than a constant value. Another method, which directly connects f(RH) to Vg(RH) (Vg(RH) = $f(\mathrm{RH})^{1.5}$), is also used for predicting ALWC based on measurements of the humidified nephelometer system and mass concentrations of dry aerosol particles (Guo et al., 2015). This method assumes that the average scattering efficiency of aerosol particles at dry state and different RH conditions are the same, and requires additional measurements of PNSD or mass concentrations of dry aerosol particles (Guo et al., 2015). However, the scattering efficiency of aerosol particles vary with particle diameters, which will change under ambient conditions due to aerosol hygroscopic growth.

In this paper, we propose a novel method to calculate the ALWC based only on measurements of a humidified nephelometer system. The proposed method includes two steps. The first step is calculating V_a(dry) based on measurements of the "dry" nephelometer using a machine learning method called random forest model. With measurements of PNSD and BC, the six parameters measured by the nephelometer can be simulated using the Mie theory, and the V_a(dry) can also be calculated based on PNSD. Therefore, the random forest model can be trained with only regional historical datasets of PNSD and BC. In this study, datasets of PNSD and BC measured from multiple sites are used in the machine learning model to characterize a regional aerosol and these datasets have covered a wide range of aerosol loadings. The second step is calculating Vg(RH) based on the Ångström exponent and f(RH) measured by the humidified nephelometer system. In this step, both the influences of the variations in PNSD and aerosol hygroscopicity are both considered to derive Vg(RH) from measured f(RH). Finally, based on calculated $V_a(\text{dry})$ and Vg(RH), ALWCs at different RH points can be estimated. The used datasets are introduced in Sect.2. Calculation method of $V_a(\text{dry})$ based only on measurements of the nephelometer, which measures optical properties of aerosols in dry state, is described in Sect.3.2. The way of deriving Vg(RH) based on measurements of the humidified nephelometer system is introduced and discussed in Sect.3.3. The final formula of calculating ambient ALWC is described in Sect.3.4. The verification of the $V_a(\text{dry})$ predicted by using the machine learning method is described in Sect.4.1. The validation of ambient ALWC calculated from measurements of the humidified nephelometer system is presented in Sect.4.2. The contribution of ambient ALWC to the total ambient aerosol volume is discussed in Sect.4.3.

2. Instruments and datasets

Datasets from six field campaigns were used in this paper. The six campaigns were conducted at four different measurement sites (Wangdu, Gucheng and Xianghe in Hebei province and Wuqing in Tianjin) of the North China Plain (NCP), the locations of these field campaign sites are displayed in Fig.S1. Time periods and datasets used from these field campaigns are listed in Table 1. During these field campaigns, aerosol particles with aerodynamic diameters less than 10 µm were sampled (by passing through an impactor). The PNSDs in dry state, which range from 3nm to 10µm, were jointly measured by a Twin Differential Mobility Particle Sizer (TDMPS, Leibniz-Institute for Tropospheric Research , Germany; Birmili et al. (1999)) or a scanning mobility particle size spectrometer (SMPS) and an Aerodynamic Particle Sizer (APS, TSI Inc., Model 3321) with a temporal resolution of 10 minutes. The mass concentrations of black carbon (BC) were measured using a Multi-Angle Absorption Photometer (MAAP Model 5012, Thermo, Inc., Waltham, MA USA) with a temporal resolution of 1 minute during field campaigns of F1 to F5, and using an aethalometer (AE33) (Drinovec et al., 2015) during field campaign F6. The aerosol light scattering coefficients (σ_{sp}) at three wavelengths (450 nm, 550 nm, and 700 nm) were measured using a TSI 3563 nephelometer (Anderson and Ogren, 1998) during field campaigns of F1 to F5, and using an Aurora 3000 nephelometer (Müller et al., 2011) during field campaign F6.

Datasets of PNSD, BC and σ_{sp} from campaigns F2, F4 and F5 are referred to as D1. Measurements of PNSD and measurements from the humidified nephelometer system during campaign F6 (Gucheng campaign) are used to verify the proposed method of calculating the ambient ALWC. Details about the humidified nephelometer system during Wangdu and Gucheng campaigns are introduced in detail in (Kuang et al., 2017a). During the Gucheng campaign, an In situ Gas and Aerosol Compositions Monitor (IGAC, Fortelice International Co.,Taiwan) was used for monitoring water-soluble ions (Na^+, K^+, Ca^{2+}, Mg^{2+}, NH^{4+}, $SO4^{2-}$, NO^{3-}, Cl^-) of $PM_{2.5}$ and their precursor gases: NH_3, HCl, and HNO_3. The time resolution of IGAC measurements is one hour. Ambient air was drawn into the IGAC system through a stainless-steel pipe wrapped with thermal insulation at a flow rate of 16.7 L/min. The ambient RH and temperature were observed using an automatic weather station with a time resolution of one minute.

3. Methodology

3.1 Closure calculations

To ensure the datasets of σ_{sp} and PNSD used are of high quality, a closure study between measured σ_{sp} and that calculated based on measured PNSD and BC with Mie theory (Bohren and Huffman, 2008) is first performed. Measured σ_{sp} bears uncertainties introduced by angular truncation errors and nonideal light source. To achieve consistency between measured and modelled σ_{sp}, modelled σ_{sp} are calculated according to practical angular situations of the nephelometer (Anderson et al., 1996). During the σ_{sp} modelling process, BC was considered to be half externally and half coreshell mixed with other aerosol components. The mass size distribution of BC used in Ma et al. (2012), which was also observed in the NCP, was used in this research to account for the mass distributions of BC at different particle sizes. The applied refractive index and density of BC were $1.80 - 0.54\mathrm{i}$ and $1.5\mathrm{g}\, cm^{-3}$ (Kuang et al., 2015). The refractive index of non light-absorbing aerosol components (other than BC) was set to $1.53 - 10^{-7}i$ (Wex et al., 2002). For the Mie theory calculation details please refer to Kuang et al. (2015).

The closure results between modelled σ_{sp} and σ_{sp} measured by TSI 3563 or Aurora 3000 using datasets observed during six field campaigns (Table 2) are depicted in Fig.1. In general, for all six field campaigns, modelled σ_{sp} values correlate very well with measured σ_{sp} values. Considering the measured PNSD has an uncertainty of larger than 10% (Wiedensohler et al., 2012), and the measured σ_{sp} has an uncertainty of about 9% (Sherman et al., 2015), modelled σ_{sp} values agree well with measured σ_{sp} values in campaigns F1, F4, F5 and F6, with all points lying nearby the 1:1 line, and most points falling within the 20% relative difference lines. For the closure results of field campaign F2, the modelled σ_{sp} values are systematically lower than measured σ_{sp} values. For the closure results of field campaign F3, most points also lie nearby 1:1 line, but points are relatively more dispersed.

3.2 Calculation of $V_a(\mathrm{dry})$ based on measurements of the "dry" nephelometer

3.2.1 Theoretical relationship between $V_a(\mathrm{dry})$ and σ_{sp}

Previous studies demonstrated that the σ_{sp} of aerosol particles is roughly proportional to $V_a(\text{dry})$ (Pinnick et al., 1980). Here, the quantitative relationship between $V_a(\text{dry})$ and σ_{sp} is analyzed.

The σ_{sp} and $V_a(\text{dry})$ can be expressed as the following:

$$\sigma_{sp} = \int \pi r^2 Q_{sca}(m,r)\text{n(r)dr} \quad (1)$$

$$V_a(\text{dry}) = \int \frac{4}{3}\pi r^3 \text{n(r)dr} \quad (2)$$

where $Q_{sca}(m,r)$ is scattering efficiency for a particle with refractive index m and particle radius r, while n(r) is the aerosol size distribution. As presented in equation (1) and (2), relating $V_a(\text{dry})$ with σ_{sp} involves the complex relation between $Q_{sca}(m,r)$ and particle diameter, which can be simulated using the Mie theory. According to the aerosol refractive index at visible spectral range, aerosol chemical components can be classified into two categories: the light absorbing component and the almost light non-absorbing components (inorganic salts and acids, and most of the organic compounds). Near the visible spectral range, the light absorbing component can be referred to as BC. BC particles are either externally or internally mixed with other aerosol components. In view of this, Q_{sca} at 550 nm, as a function of particle diameter for four types of aerosol particles, is simulated using Mie theory: almost non-absorbing aerosol particle, BC particle, BC particle core-shell mixed with non-absorbing components with the radius of the inner BC core being 50 nm and 70 nm, respectively. Same with those introduced in Sect.2.2, the refractive indices of BC and light non-absorbing components used here are $1.80 - 0.54\text{i}$ and $1.53 - 10^{-7}\text{i}$, respectively.

The simulated results are shown in Fig.2a. Near the visible spectral range, most of the ambient aerosol components are almost non-absorbing, and their Q_{sca} varies more like the blue line shown in Fig.2a. In that case, aerosol particles have diameters less than about 800 nm and Q_{sca} increases almost monotonously with particle diameter and can be approximately estimated as a linear function of diameter. Fig.2b shows the simulated size-resolved accumulative contribution to the scattering coefficient at 550 nm for all PNSDs measured during the Wangdu campaign. The results indicate that, for continental aerosol particles without influences of dust, in most cases, all particles with diameter less than about 800 nm contribute more than 80% to the total σ_{sp}. Therefore, for equation (1), if we express $Q_{sca}(m,r)$ as $Q_{sca}(m,r) = \mathrm{k} \cdot \mathrm{r}$, then equation (1) can be expressed as the following:

$$\sigma_{sp} = \mathrm{k} \cdot \int \pi r^3 \mathrm{n(r)dr} \qquad (3)$$

This explains why $\sigma_{sp}(550\ nm)$ is roughly proportional to $V_a(\mathrm{dry})$. However, the value k varies greatly with particle diameter. The ratio $\sigma_{sp}(550\ nm)/V_a(\mathrm{dry})$ (hereinafter referred to as R_{Vsp}) is mostly affected by the PNSD, which determines the weight of influence different particle diameters have on R_{Vsp}. The discrepancy between the blue line and black line shown in Fig.2a indicates that the fraction of externally mixed BC particles and their sizes has large impact on R_{Vsp}. The difference between the black line and the red line as well as the difference between the solid red line and the dashed red line shown in Fig.2a indicate that the way and the amount of BC mixed with other components also exert significant influences on R_{Vsp}. In summary, the variation of R_{Vsp} is mainly determined by variations in PNSD, mass size distribution and the mixing state of BC. It is difficult to find a simple function describing the relationship between measured σ_{sp} and $V_a(\mathrm{dry})$.

Based on PNSD and BC datasets of field campaigns F1 to F6, the relationship between σ_{sp} at 550 nm and $V_a(\mathrm{dry})$ of PM_{10} or $PM_{2.5}$ are simulated using the Mie theory. The results are shown in Fig.3. The results demonstrate that the σ_{sp} at 550 nm is highly correlated with the V_a(dry) of PM_{10} and $PM_{2.5}$. The square of the correlation coefficient (r^2) between σ_{sp} at 550 nm and V_a(dry) of PM_{10} or $PM_{2.5}$ are 0.94 and 0.99, respectively. A roughly proportional relationship exists between V_a(dry) and $\sigma_{sp}(550\ nm)$, especially for V_a(dry) of $PM_{2.5}$. However, both R_{Vsp} of PM_{10} and $PM_{2.5}$ vary significantly. R_{Vsp} of PM_{10} mainly ranges from 2 to 6 $cm^3/(\mu m^3 \cdot Mm)$, with an average of 4.2 $cm^3/(\mu m^3 \cdot Mm)$. R_{Vsp} of $PM_{2.5}$ mainly ranges from 3 to 6.5 $cm^3/(\mu m^3 \cdot Mm)$, with an average of 5.1 $cm^3/(\mu m^3 \cdot Mm)$. Simulated size-resolved accumulative contributions to σ_{sp} at 550 nm for all PNSDs measured during campaigns F1 to F6 and corresponding size-resolved accumulative contributions to V_a(dry) of PM_{10} are shown in Fig.S2. The results indicate that particles with diameter larger than 2.5 μm usually contribute negligibly to σ_{sp} at 550 nm but contribute about 20% of the total PM_{10} volume. Hence σ_{sp} at 550 nm is insensitive to changes in particles mass of diameters between 2.5 to 10 μm. This may partially explain why V_a(dry) of $PM_{2.5}$ correlates better with σ_{sp} at 550 nm than V_a(dry) of PM_{10}.

3.2.2 Machine learning

Based on analyses in Sect.3.2.1, R_{Vsp} varies a lot with PNSD being the most dominant influencing factor. The “dry” nephelometer provides not only one single σ_{sp} at 550 nm, it measures six parameters including σ_{sp} and back scattering coefficients (σ_{bsp}) at three wavelengths (for TSI 3563: 450 nm, 550 nm, 700 nm). The Ångström exponent calculated from spectral dependence of σ_{sp} provides information on the mean predominant aerosol size and is associated mostly with PNSD. The variation of the hemispheric backscattering fraction (HBF), which is the ratio between σ_{bsp} and σ_{sp}, is also essentially related to the PNSD. HBFs at three wavelengths

(450 nm, 550 nm, 700 nm) and the Ångstr$\ddot{o}$m exponents calculated from σ_{sp} at different wavelengths (450-550 nm, 550-700 nm, 450-700 nm) for typical non-absorbing aerosol particles with their diameters ranging from 100 nm to 3 μm are simulated using the Mie theory. The results are shown in Fig.4a and Fig.4b. HBF values at three different wavelengths and their differences are more sensitive to changes in PNSD of particle diameters less than about 400 nm. Ångstr$\ddot{o}$m exponents calculated from σ_{sp} at different wavelengths almost decrease monotonously with particle diameter when particle diameter is less than about 1 μm, however, they differ distinctly when particle diameter is larger than 300 nm. These results indicate that HBFs at three wavelengths and Ångstr$\ddot{o}$m exponents calculated from σ_{sp} at different wavelengths are sensitive to different diameter ranges of PNSD.

Thus, all six parameters measured by the “dry” nephelometer together can provide valuable information about variations in R_{Vsp}. However, no explicit formula exists between these six parameters and $V_a(\mathrm{dry})$. How to use these six optical parameters is a problem. Machine learning methods which can handle many input parameters are capable of learning from historical datasets and then make predictions, and strict relationships among variables are not required. Machine learning methods are powerful tools for tackling highly nonlinear problems and are widely used in different areas. In the light of this, predicting $V_a(\mathrm{dry})$ based on six optical parameters measured by the “dry” nephelometer might be accomplished by using a machine learning method. In this study, random forest is chosen for this purpose.

Random forest is a machine learning technique that is widely used for classification and non-linear regression problems (Breiman, 2001). For non-linear regression cases, random forest model consists of an ensemble of binary regression decision tress. Each tree has a randomized training scheme, and an average over the whole ensemble of regression tree predictions is used for final prediction. In this study, the function RandomForestRegressor from the Python Scikit–Learn machine learning library (http://scikit-learn.org/stable/index.html) is used. This model has several strengths. First, by averaging over an ensemble of decision trees, there is a significantly lower risk of overfitting. Second, it involves fewer assumptions about the dependence between inputs and outputs when compared with traditional parametric regression models. The random forest model has two parameters: the number of input variables (N_{in}) and the number of trees grown (N_{tree}). In this study, N_{in} and N_{tree} are six and eight, respectively. The six input parameters the three scattering coefficients, three backscattering coefficients.

The quality of input datasets is critical to the prediction accuracy of the machine learning method. As discussed in Sect.3.1, modeled σ_{sp} during some field campaigns are not completely consistent with measured σ_{sp}, large bias might exist between them due to the measurement uncertainties of PNSD and σ_{sp}. To avoid that the measurements uncertainties are involved in the training processes of the random forest model. In this study, both the required datasets of six optical parameters which corresponding to measurements of TSI 3563 and V_a(dry) for training the random forest model are calculated or simulated based on measurements of PNSD and BC from field campaigns F1 to F4 and F6. Datasets of PNSD and six optical parameters measured by the nephelometer during campaign F5 are used to verify the prediction ability of the trained random forest model. The performance of this random forest model on predicting both V_a(dry) of PM_{10} and $PM_{2.5}$ are investigated. A schematic diagram of this method is shown in Fig.5.

3.3 Connecting $f(\mathrm{RH})$ to Vg(RH)

3.3.1 κ-Köhler theory

κ-K$\ddot{o}$hler theory is used to describe the hygroscopic growth of aerosol particles with different sizes, and the formula expression of κ-K$\ddot{o}$hler theory can be written as follows (Petters and Kreidenweis, 2007):

$$\mathrm{RH} = \frac{D^3 - {D_d}^3}{D^3 - {D_d}^3(1-\kappa)} \cdot \exp\left(\frac{4\sigma_{s/a} \cdot M_{water}}{R \cdot T \cdot D_p \cdot g \cdot \rho_w}\right) \qquad (4)$$

where D is the diameter of the droplet, D_d is the dry diameter, $\sigma_{s/a}$ is the surface tension of solution/air interface, T is the temperature, M_{water} is the molecular weight of water, R is the universal gas constant, ρ_w is the density of water, and κ is the hygroscopicity parameter. By combining the Mie theory and the κ-K$\ddot{o}$hler theory, both $f(\mathrm{RH})$ and $\mathrm{Vg(RH)}$ can be simulated. In the processes of calculations for modelling $f(\mathrm{RH})$ and $\mathrm{Vg(RH)}$, the treatment of BC is same with those introduced in Sect.2.2. As aerosol particle grow due to aerosol water uptake, the refractive index will change. In the Mie calculation, impacts of aerosol liquid water on the refractive index are considered based on volume mixing rule. The used refractive index of liquid water is $1.33 - 10^{-7}i$ (Seinfeld and Pandis, 2006).

3.3.2 Parameterization schemes for $f(\mathrm{RH})$ and Vg(RH)

The $f(\mathrm{RH})$ is defined as $f(\mathrm{RH}) = \sigma_{sp}(RH, 550\ nm)/\sigma_{sp}(dry, 550\ nm)$ where $\sigma_{sp}(RH, 550\ nm)$ and $\sigma_{sp}(dry, 550\ nm)$ represents σ_{sp} at wavelength 550 nm under certain RH and dry conditions. Additionally, $\mathrm{Vg(RH)}$ is defined as $\mathrm{Vg(RH)} = V_a(\mathrm{RH})/V_a(\mathrm{dry})$, where $V_a(\mathrm{RH})$ represents total volume of aerosol particles under certain RH conditions.

A physically based single-parameter representation is proposed by Brock et al. (2016) to describe $f(\mathrm{RH})$. The parameterization scheme is written as:

$$f(\mathrm{RH}) = 1 + \kappa_{sca}\frac{RH}{100-RH} \quad (5)$$

where κ_{sca} is the parameter which fits $f(\mathrm{RH})$ best. Here, a brief introduction is given about the physical understanding of this parameterization scheme. For aerosol particles whose diameters larger than 100 nm, regardless of the Kelvin effect, the hygroscopic growth factor for a aerosol particle can be approximately expressed as $\mathrm{g(RH)} \cong (1 + \kappa\frac{RH}{100-RH})^{1/3}$ (Brock et al., 2016). Enhancement factor in volume can be expressed as the cube of $\mathrm{g(RH)}$. Aerosol particles larger than 100 nm contribute the most to σ_{sp} and $V_a(\mathrm{dry})$ (as shown in Fig.S2).If a constant κ which represents the overall aerosol hygroscopicity of ambient aerosol particles, is used as the κ of different particle sizes, then $\mathrm{Vg(RH)}$ can be approximately expressed as $\mathrm{Vg(RH)} = 1 + \kappa\frac{RH}{100-RH}$. In addition, σ_{sp} is usually proportional to $V_a(\mathrm{dry})$ which indicates that the relative change in σ_{sp} due to aerosol water uptake is roughly proportional to relative change in aerosol volume. Therefore, $f(\mathrm{RH})$ might also be well described by using the formula form of equation (5). Previous studies have shown that this parameterization scheme can describe $f(\mathrm{RH})$ well (Brock et al., 2016;Kuang et al., 2017b).

During processes of measuring $f(\mathrm{RH})$, the sample RH in the “dry” nephelometer (RH_0) is not zero. According to equation (5), the measured $f(\mathrm{RH})_{measure} = \frac{f(\mathrm{RH})}{f(RH_0)}$ should be fitted using the following formula:

$$f(\mathrm{RH})_{measure} = (1 + \kappa_{sca}\frac{RH}{100-RH})/(1 + \kappa_{sca}\frac{RH_0}{100-RH_0}) \quad (6)$$

Based on this equation, κ_{sca} can be calculated from measured $f(\mathrm{RH})$ directly. The typical value of RH_0 measured in the “dry” nephelometer during Wangdu campaign is about 20%. The importance of the RH_0 correction changes under different aerosol hygroscopicity and RH_0 conditions. The parameter κ_{sca} is fitted with and without consideration of RH_0 for f(RH) measurements during Wangdu campaign, and the results are shown in Fig.S3. The results demonstrate that, overall, the κ_{sca} will be underestimated if the influence of RH_0 is not considered, and the larger the κ_{sca}, the more that the κ_{sca} will be underestimated.

In addition, based on discussions about the physical understanding of equation (5), the Vg(RH) should be well described by the following equation:

$$\mathrm{Vg(RH)} = 1 + \kappa_{Vf}\frac{RH}{100-RH} \quad (7)$$

where κ_{Vf} is the parameter which fits Vg(RH) best. To validate this conclusion, a simulative experiment is conducted. In the simulative experiment, average PNSD in dry state and mass concentration of BC during the Haze in China (HaChi) campaign (Kuang et al., 2015) are used. During HaChi campaign, size-resolved κ distributions are derived from measured size-segregated chemical compositions (Liu et al., 2014) and their average is used in this experiment to account the size dependence of aerosol hygroscopicity. Modelled results of f(RH) and Vg(RH) are shown in Fig.7. Results demonstrate that modelled f(RH) and Vg(RH) can be well parameterized using the formula form of equation (5) and (7). Fitted values of κ_{sca} and κ_{Vf} are 0.227 and 0.285, respectively. This result indicates that if linkage between κ_{sca} and κ_{Vf} is established, measurements of f(RH) can be directly related to Vg(RH).

3.3.3 Bridge the gap between f(RH) and Vg(RH)

Many factors have significant influences on the relationships between f(RH) and Vg(RH), such as PNSD, BC mixing state and the size-resolved aerosol hygroscopicity. To gain insights into the relationships between κ_{sca} and κ_{Vf}, a simulative experiment using Mie theory and κ-K$\ddot{o}$hler theory is designed. In this experiment, all PNSDs at dry state along with mass concentrations of BC from D1 are used, characteristics of these PNSDs can be found in Kuang et al. (2017b). As to size-resolved aerosol hygroscopicity, a number of size-resolved κ distributions were derived from measured size-segregated chemical compositions during HaChi campaign (Liu et al., 2014). Results from other researches also show similar size dependence of aerosol hygroscopicity (Meng et al., 2014). In view of this, the shape of the average size-resolved κ distribution during HaChi campaign (black line shown in Fig.S5) is used in the designed experiment. Other than the shape of size-resolved κ distribution, the overall aerosol hygroscopicity which determines the magnitude of $f(\mathrm{RH})$ also have large impacts on the relationship between κ_{sca} and κ_{Vf}. In view of this, ratios range from 0.05 to 2 with an interval of 0.05 are multiplied with the average size-resolved κ distribution (the black line shown in Fig.S5) to produce a number of size-resolved κ distributions which represent aerosol particles from nearly hydrophobic to highly hygroscopic. During simulating processes, each PNSD is modelled with all produced size-resolved κ distributions. In the following, the ratio κ_{Vf}/κ_{sca} termed as R_{Vf} is used to indicate the relationship between κ_{sca} and κ_{Vf}.

In consideration of that values of Ångstr$\ddot{o}$m exponent contain information about PNSD (Kuang et al., 2017b) and values of κ_{sca} represent overall hygroscopicity of ambient aerosol particles, and both the two parameters can be directly calculated from measurements of a three-wavelength humidified nephelometer system (Kuang et al., 2017b). Simulated R_{Vf} values are spread into a two-dimensional gridded plot. The first dimension is Ångstr$\ddot{o}$m exponent with an interval of 0.02 and the second dimension is κ_{sca} with an interval of 0.01. Average R_{Vf} value within each grid is represented by color and shown in Fig.6a. Values of Ångström exponent corresponding to used PNSDs are calculated from simultaneously measured σ_{sp} values at 450 nm and 550 nm from TSI 3563 nephelometer. Results shown in Fig.6a exhibit that both PNSD and overall aerosol hygroscopicity have significant influences on R_{Vf}. Simulated values of R_{Vf} range from 0.8 to 1.7 with an average of 1.2. Overall, R_{Vf} value is lower when value of Ångström exponent is larger. The percentile value of standard deviation of R_{Vf} values within each grid divided by its average is shown in Fig.6b. In most cases, these percentile values are less than 10% (about 90%) which demonstrates that R_{Vf} varies little within each grid shown in Fig.6a. Figure 6 shows the influence of aerosol size and chemistry on R_{Vf}. For Ångström exponent less than ~1.1, R_{Vf} varies strongly with κ_{sca}. However, for Ångström exponent values greater than ~1.1, the R_{Vf} relative standard deviation exhibits a higher variability with the Ångström exponent. Thus, showing the sensitivity of R_{Vf} to changes in aerosol size for small particles. In general, results shown in Fig.6 imply that results of Fig.6a can serve as a look up table to estimate R_{Vf} and thereby κ_{Vf}, such that these values can be directly predicted from measurements of a three-wavelength humidified nephelometer system.

For the look up table shown in Fig.6a, a fixed size-resolved κ distribution is used, which might not be able to capture variations of R_{Vf} induced by different types of size-resolved κ distributions under different PNSD conditions. A simulative experiment is conducted to investigate the performance of this look up table. In this experiment, the following datasets are used: PNSDs and mass concentrations of BC from D1 (the number of used PNSD is 11996), and size-resolved κ distributions from HaChi campaign (Liu et al., 2014) which are presented in Fig.7a (the number is 23). Results shown in Fig.7a imply that the shape of size-resolved κ distribution is highly variable yet has no apparent correlation with aerosol loading. During the simulating processes, for each PNSD, it is used to simulate R_{Vf} values corresponding to all used size-resolved κ distributions, therefore, 275908 R_{Vf} values are modelled. Also, modelled values of κ_{sca} and corresponding values of modelled Ångstr$\ddot{o}$m exponent are together used to estimate R_{Vf} values using the look up table shown in Fig.7a. Results of relative differences between estimated and modelled R_{Vf} values under different pollution conditions are shown in Fig.7b. Overall, 88% of points have absolute relative differences less than 15%, and 68% of points have absolute relative differences less than 10%. This look up table performs better when the air is relatively polluted.

3.4 Calculation of ambient ALWC

According to the equation $\mathrm{Vg(RH)} = 1 + \kappa_{Vf}\frac{RH}{100-RH}$, volume concentrations of aerosol liquid water (ALWC) at different RH points can be expressed as:

$$ALWC = V_a(\mathrm{dry}) \times (\mathrm{Vg(RH)} - 1) = V_a(\mathrm{dry}) \cdot \kappa_{sca} \cdot R_{Vf} \cdot \frac{RH}{100-RH}. \quad (7)$$

According to discussions of Sect.3.2, $V_a(\mathrm{dry})$ can be predicted based only on measurements from the “dry” nephelometer by using a random forest model. The training of the random forest model requires only regional historical datasets of simultaneously measured PNSD and BC. The κ_{sca} is directly fitted from $f(\mathrm{RH})$ measurements. The R_{Vf} can be estimated using the look up table introduced in Sect.3.3. Thus, based only on measurements from a three-wavelength humidified nephelometer system, ALWCs of ambient aerosol particles at different RH points can be estimated. If both measurements from the humidified nephelometer system and ambient RH are available, ambient ALWC can be calculated. The flowchart of calculating ambient ALWC based on measurements of the humidified nephelometer system is shown in Fig.8. The used nephelometer corresponding to this flowchart should be TSI 3563. If nephelometer of the used humidified nephelometer system is Aurora 3000, wavelengths in this flowchart will change but other steps are totally the same.

4. Results and discussions

4.1 Validation of the random forest model for predicting $V_a(\mathrm{dry})$ based on measurements of the "dry" nephelometer

The machine learning method, random forest model, is proposed to predict $V_a(\mathrm{dry})$ based only on σ_{sp} and σ_{bsp} at three wavelengths measured by the "dry" nephelometer. Datasets of PNSD and BC from field campaigns F1 to F4 and F6 are used to train the random forest model. Datasets of PNSD and optical parameters measured by the "dry" nephelometer from field campaign F5 are used to verify the trained random forest model. The schematic diagram of this method is shown in Fig.5. The comparison results between calculated and predicted $V_a(\mathrm{dry})$ of PM_{10} and $PM_{2.5}$ are shown in Fig.9. The square of correlation coefficient between predicted and calculated $V_a(\mathrm{dry})$ of PM_{10} is 0.96. And almost all points lie between or near 20% relative difference lines. The square of correlation coefficient between predicted and calculated $V_a(\mathrm{dry})$ of $PM_{2.5}$ is 0.997. And almost all points lie between or near 10% relative difference lines. The standard deviations of relative differences between predicted and calculated $V_a(\mathrm{dry})$ of PM_{10} and $PM_{2.5}$ are 10% and 4% , respectively. These results indicate that $V_a(\mathrm{dry})$ of $PM_{2.5}$ can be well predicted by using the machine learning method. While $V_a(\mathrm{dry})$ of PM_{10} predicted by using the machine learning method has a relatively larger bias.

Machine learning methods do not explicitly express relationships between many variables, however, they learn and implicitly construct complex relationships among variables from historical datasets. Many different and comprehensive machine learning methods are developed for diverse applications and can be directly used as a tool for solving a lot of nonlinear problems which may not be mathematically well understood. We suggest that using machine learning method for estimating $V_a(\text{dry})$ based on measurements of the "dry" nephelometer. The way of estimating $V_a(\text{dry})$ with machine learning method might be applicable for different regions around the world if used estimators are trained with corresponding regional historical datasets.

4.2 Comparison between ambient ALWC calculated from ISORROPIA and measurements of the humidified nephelometer system.

So far, widely used tools for prediction of ambient ALWC are thermodynamic models. ISORROPIA-II thermodynamic model (http://nenes.eas.gatech.edu/ISORROPIA/index_old.html) is a famous one, and is widely used in researches for predicting pH and ALWC of ambient aerosol particles (Guo et al., 2015;Cheng et al., 2016;Liu et al., 2017;Fountoukis and Nenes, 2007). Water soluble ions and gaseous precursors are required as inputs of thermodynamic model. During Gucheng campaign, measurements from both the humidified nephelometer system and IGAC are available. Thus, the ambient ALWC can be calculated through two independent methods: thermodynamic model based on IGAC measurements and the method proposed in Sect.3.4 which is based on measurements of the humidified nephelometer system. In this study, the forward mode in ISORROPIA-II is used, and water-soluble ions in PM2.5 and gaseous precursors (NH_3, HNO_3, HCl) measured by the IGAC instrument along with simultaneously measured RH and T are used as inputs. The aerosol water associated with organic matter are not considered in the method of ISORROPIA model, due to the lack of measurements of organic aerosol mass. However, results from previous studies indicate that organic matter induced particle water only account for about 5% of total ALWC (Liu et al., 2017). For the ALWC calculated from the humidified nephelometer system. The needed V_a(dry) of $PM_{2.5}$ in equation (7) is calculated from simultaneously measured PNSD.

The comparison results between ambient ALWC calculated from these two independent methods are shown in Fig.10a. The square of correlation coefficient between them is 0.92, most of the points lie within or nearby 30% relative difference lines. The slope is 1.14, and the intercept is -8.6 $\mu m^3/cm^3$. When ambient RH is higher than 80%, the ambient ALWCs calculated from measurements of the humidified nephelometer system are relatively higher than those calculated based on ISORROPIA-Ⅱ. When ambient RH is lower than 60%, the ambient ALWCs calculated from measurements of the humidified nephelometer system are relatively lower than those calculated based on ISORROPIA-Ⅱ. Overall, a good agreement is achieved between ambient ALWC calculated from measurements of the humidified nephelometer system and ISORROPIA thermodynamic model.

Guo et al. (2015) conducted the comparison between ambient ALWC calculated from ISORROPIA model and ambient ALWC calculated from measurements of the humidified nephelometer system by assuming Vg(RH)= $f(\mathrm{RH})^{1.5}$. Thus, the comparison results between ambient ALWC calculated based on ISORROPIA and ambient ALWC calculated by assuming Vg(RH)= $f(\mathrm{RH})^{1.5}$ are also shown in Fig.10b. The square of correlation coefficient between them is also 0.92. However, the slope and intercept are 1.7 and -21 $\mu m^3/cm^3$, respectively. When the ambient RH is higher than about 80%, calculated ambient ALWC will be significantly overestimated if assumes that Vg(RH)= $f(\mathrm{RH})^{1.5}$. This method assumes that average scattering efficiency of aerosol particles at dry state and different RH conditions are the same. When ambient RH is high, the particle diameters changes a lot. As the results shown in Fig.S6, for non-absorbing particle, when diameter of aerosol particle in dry state is less than 500 nm, the aerosol scattering efficiency increase almost monotonously with increasing RH especially when RH is higher than 80%. Therefore, it is not suitable to assume that average scattering efficiency of aerosol particles at dry state and different RH conditions are the same.

4.3 Volume fractions of ALWC in total ambient aerosol volume

During Wangdu campaign, κ_{sca} ranges from 0.05 to 0.3 with an average of 0.19. Estimated values of R_{Vf} ranges from 0.86 to 1.47, with an average of 1.15. Estimated values of κ_{Vf} ranges from 0.05 to 0.35, with an average of 0.22. The calculated volume fractions of water in total volume of ambient aerosols during Wangdu campaign are shown in Fig.11a. The results indicate that during Wangdu campaign, when ambient RH is higher than 70%, the κ_{Vf} values are relatively higher. The volume fractions of water is always higher than 50% when ambient RH is higher than 80%.

During Gucheng campaign, κ_{sca} ranges from 0.008 to 0.22 with an average of 0.1, κ_{Vf} ranges from 0.01 to 0.21 with an average of 0.12. The aerosol hygroscopicity during Gucheng campaign is much lower than aerosol hygroscopicity during Wangdu campaign. The calculated volume fractions of water in total volume of ambient aerosols during Gucheng campaign are shown in Fig.11b. During Gucheng campaign, the maximum volume fraction of water in ambient aerosol is 42% when ambient RH is at 80%.On average, when ambient RH is higher than 90%, the volume fraction of water in ambient aerosols reaches higher than 50%.

4.4 Discussions about the applicability of the proposed method

The method proposed in this research is based on datasets of PNSD, σ_{sp} and size-resolved κ distribution which are measured on the NCP without influences of dust events and sea salt. Cautions should be exercised if using the proposed method to estimate the ALWC when the air mass is significantly influenced by sea salt or dust. The way of estimating $V_a(\mathrm{dry})$ with machine learning method might be applicable for different regions around the world. However, the used predictor from machine learning should be trained with corresponding regional historical datasets of PNSD and BC. The way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg}(\mathrm{RH})$ might also be applicable for other continental regions. Still, we suggest that the used look up table is simulated from regional historical datasets.

Note that the humidified nephelometer usually operates with RH less than 95%. Aerosol water, however, increase dramatically with increasing RH when RH is greater than 95%. Such high RH conditions can occur during the haze events. This may limit the usage of the proposed method when ambient RH is extremely high. As discussed in Sect.3.3, the proposed way of connecting $f(\mathrm{RH})$ and $\mathrm{Vg}(\mathrm{RH})$ is based on the κ-K$\ddot{o}$hler theory. If κ does not change with RH, the proposed method should be applicable when RH is higher than 95%, even the measurements of humidified nephelometer system are conducted when RH is less than 95%. Many studies have done researches about the change of κ with the changing RH (Rastak et al., 2017;Renbaum-Wolff et al., 2016), their results demonstrate that the κ changes with increasing RH. However, few studies have investigated the variation of κ of ambient aerosol particles with changing RH when RH is less than 100%. Liu et al. (2011) have measured κ of ambient aerosol particles at different RHs (90%, 95%, 98.5%) on the NCP. Their results demonstrated that κ at different RHs differ little for ambient aerosol particles with different diameters. Results of Kuang et al. (2017a) indicated that κ values retrieved from $f(\mathrm{RH})$ measurements agree well with κ values at RH of 98%

of aerosol particles with diameter of 250 nm. In this respect, the proposed method might be applicable even when ambient RH is extremely high for ambient aerosol particles on the NCP. Moreover, for calculating the ambient ALWC, the measured ambient RH is required. If the ambient RH is higher than 95%, the measured ambient RH with current techniques is highly uncertain. Given this, cautions should be exercised if the ambient ALWC is calculated when the ambient RH is higher than 95%.

5. Conclusions

In this paper, a novel method is proposed to calculate ALWC based on measurements of a three-wavelength humidified nephelometer system. Two critical relationships are required in this method. One is the relationship between $V_a(\mathrm{dry})$ and measurements of the “dry” nephelometer. Another one is the relationship between Vg(RH) and $f(\mathrm{RH})$. The ALWC can be calculated from the estimated $V_a(\mathrm{dry})$ and Vg(RH).

Previous studies have shown that an approximate proportional relationship exists between $V_a(\mathrm{dry})$ and corresponding σ_{sp}, especially for fine particles (particle diameter less than 1 μm). However, PNSD and other factors still have significant influences on this proportional relationship. It is difficult to directly estimate $V_a(\mathrm{dry})$ from measured σ_{sp}. In this paper, a random forest predictor from machine learning procedure is used to estimate $V_a(\mathrm{dry})$ based on measurements of a three-wavelength nephelometer. This random forest predictor is trained based on historical datasets of PNSD and BC from several field campaigns conducted on the NCP. This method is then validated using measurements from Wangdu campaign. The square of correlation coefficient between measured and estimated $V_a(\mathrm{dry})$ of PM_{10} and $PM_{2.5}$ are 0.96 and 0.997, respectively.

The relationship between $\mathrm{Vg(RH)}$ and $f\mathrm{(RH)}$ is investigated in Sect.3 by conducting a simulative experiment. It is found that the complicated relationship between $\mathrm{Vg(RH)}$ and $f\mathrm{(RH)}$ can be disentangled by using a look up table, and parameters required in the look up table can be directly calculated from measurements of a three-wavelength humidified nephelometer system. Given that the $V_a\mathrm{(dry)}$ can be estimated from a three-wavelength "dry" nephelometer, the ambient ALWC can be estimated from measurements of a three-wavelength humidified nephelometer system in conjunction with measured ambient RH. We have conducted the comparison between ambient ALWC calculated from ISORROPIA and ambient ALWC calculated from measurements of the humidified nephelometer system. The square of correlation coefficient between them is 0.92, and most of the points lie within or nearby 30% relative difference lines. The slope and intercept are 1.14 and -8.6 $\mu m^3/cm^3$, respectively. Overall, a good agreement is achieved between ambient ALWC calculated from measurements of the humidified nephelometer system and ISORROPIA thermodynamic model.

Results introduced in this research have bridged the gap between $f\mathrm{(RH)}$ and $\mathrm{Vg(RH)}$. The advantage of using measurements of a humidified nephelometer system to estimate ALWC is that this technique has a fast response time and can provide continuous measurements of the changing ambient conditions. The new method proposed in this research will facilitate the real-time monitoring of the ambient ALWC and further our understanding of roles of ALWC in atmospheric chemistry, secondary aerosol formation and climate change.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (41590872 and 41505107), the National Key R&D Program of China (2016YFC020000: Task 5) and the National research program for key issues in air pollution control (DQGG0103).

Data availability. The data used in this study are available from the corresponding author upon request (zcs@pku.edu.cn)

References

Anderson, T., Covert, D., Marshall, S., Laucks, M., Charlson, R., Waggoner, A., Ogren, J., Caldow, R., Holm, R., and Quant, F.: Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer, Journal of Atmospheric and Oceanic Technology, 13, 967-986, 1996.

Anderson, T. L., and Ogren, J. A.: Determining aerosol radiative properties using the TSI 3563 integrating nephelometer, Aerosol Science and Technology, 29, 57-69, 10.1080/02786829808965551, 1998.

Bian, Y. X., Zhao, C. S., Ma, N., Chen, J., and Xu, W. Y.: A study of aerosol liquid water content based on hygroscopicity measurements at high relative humidity in the North China Plain, Atmos. Chem. Phys., 14, 6417-6426, 10.5194/acp-14-6417-2014, 2014.

Birmili, W., Stratmann, F., and Wiedensohler, A.: Design of a DMA-based size spectrometer for a large particle size range and stable operation, Journal of Aerosol Science, 30, 549-553, 10.1016/s0021-8502(98)00047-0, 1999.

Bohren, C. F., and Huffman, D. R.: Absorption and scattering of light by small particles, Wiley, New York, USA, 2008.

Breiman, L.: Random forests, Machine Learning, 45, 5-32, 10.1023/a:1010933404324, 2001.

Brock, C. A., Wagner, N. L., Anderson, B. E., Attwood, A. R., Beyersdorf, A., Campuzano-Jost, P., Carlton, A. G., Day, D. A., Diskin, G. S., Gordon, T. D., Jimenez, J. L., Lack, D. A., Liao, J., Markovic, M. Z., Middlebrook, A. M., Ng, N. L., Perring, A. E., Richardson, M. S., Schwarz, J. P., Washenfelder, R. A., Welti, A., Xu, L., Ziemba, L. D., and Murphy, D. M.: Aerosol optical properties in the southeastern United States in summer – Part 1: Hygroscopic growth, Atmos. Chem. Phys., 16, 4987-5007, 10.5194/acp-16-4987-2016, 2016.

Cheng, Y., Zheng, G., Wei, C., Mu, Q., Zheng, B., Wang, Z., Gao, M., Zhang, Q., He, K., Carmichael, G., Pöschl, U., and Su, H.: Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China, Science Advances, 2, 10.1126/sciadv.1601530, 2016.

Covert, D. S., Charlson, R., and Ahlquist, N.: A study of the relationship of chemical composition and humidity to light scattering by aerosols, Journal of applied meteorology, 11, 968-976, 1972.

Drinovec, L., Močnik, G., Zotter, P., Prévôt, A. S. H., Ruckstuhl, C., Coz, E., Rupakheti, M., Sciare, J., Müller, T., Wiedensohler, A., and Hansen, A. D. A.: The "dual-spot" Aethalometer: an improved measurement of aerosol black carbon with real-time loading compensation, Atmospheric Measurement Techniques, 8, 1965-1979, 10.5194/amt-8-1965-2015, 2015.

Engelhart, G. J., Hildebrandt, L., Kostenidou, E., Mihalopoulos, N., Donahue, N. M., and Pandis,

S. N.: Water content of aged aerosol, Atmos. Chem. Phys., 11, 911-920, 10.5194/acp-11-911-2011, 2011.

Fierz-Schmidhauser, R., Zieger, P., Vaishya, A., Monahan, C., Bialek, J., O'Dowd, C. D., Jennings, S. G., Baltensperger, U., and Weingartner, E.: Light scattering enhancement factors in the marine boundary layer (Mace Head, Ireland), Journal of Geophysical Research: Atmospheres, 115, D20204, 10.1029/2009JD013755, 2010.

Fountoukis, C., and Nenes, A.: ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K⁺–Ca²⁺–Mg²⁺–NH₄⁺–Na⁺–SO₄^{2−}–NO₃^{−}–Cl^{−}–H₂O aerosols, Atmos. Chem. Phys., 7, 4639-4659, 10.5194/acp-7-4639-2007, 2007.

Guo, H., Xu, L., Bougiatioti, A., Cerully, K. M., Capps, S. L., Hite Jr, J. R., Carlton, A. G., Lee, S. H., Bergin, M. H., Ng, N. L., Nenes, A., and Weber, R. J.: Fine-particle water and pH in the southeastern United States, Atmos. Chem. Phys., 15, 5211-5228, 10.5194/acp-15-5211-2015, 2015.

Kuang, Y., Zhao, C. S., Tao, J. C., and Ma, N.: Diurnal variations of aerosol optical properties in the North China Plain and their influences on the estimates of direct aerosol radiative effect, Atmos. Chem. Phys., 15, 5761-5772, 10.5194/acp-15-5761-2015, 2015.

Kuang, Y., Zhao, C. S., Tao, J. C., Bian, Y. X., and Ma, N.: Impact of aerosol hygroscopic growth on the direct aerosol radiative effect in summer on North China Plain, Atmospheric Environment, 147, 224-233, http://dx.doi.org/10.1016/j.atmosenv.2016.10.013, 2016.

Kuang, Y., Zhao, C., Tao, J., Bian, Y., Ma, N., and Zhao, G.: A novel method for deriving the aerosol hygroscopicity parameter based only on measurements from a humidified nephelometer system, Atmos. Chem. Phys., 17, 6651-6662, 10.5194/acp-17-6651-2017, 2017a.

Kuang, Y., Zhao, C., Tao, J., Bian, Y., Ma, N., and Zhao, G.: A novel method to derive the aerosol hygroscopicity parameter based only on measurements from a humidified nephelometer system, Atmos. Chem. Phys. Discuss., 2017, 1-25, 10.5194/acp-2016-1066, 2017b.

Liu, H. J., Zhao, C. S., Nekat, B., Ma, N., Wiedensohler, A., van Pinxteren, D., Spindler, G., Müller, K., and Herrmann, H.: Aerosol hygroscopicity derived from size-segregated chemical composition and its parameterization in the North China Plain, Atmos. Chem. Phys., 14, 2525-2539, 10.5194/acp-14-2525-2014, 2014.

Liu, M., Song, Y., Zhou, T., Xu, Z., Yan, C., Zheng, M., Wu, Z., Hu, M., Wu, Y., and Zhu, T.: Fine particle pH during severe haze episodes in northern China, Geophys. Res. Lett., 44, 5213-5221, 10.1002/2017GL073210, 2017.

Liu, P. F., Zhao, C. S., Göbel, T., Hallbauer, E., Nowak, A., Ran, L., Xu, W. Y., Deng, Z. Z., Ma, N., Mildenberger, K., Henning, S., Stratmann, F., and Wiedensohler, A.: Hygroscopic properties of aerosol particles at high relative humidity and their diurnal variations in the North China Plain, Atmos. Chem. Phys., 11, 3479-3494, 10.5194/acp-11-3479-2011, 2011.

Ma, N., Zhao, C. S., Müller, T., Cheng, Y. F., Liu, P. F., Deng, Z. Z., Xu, W. Y., Ran, L., Nekat, B., van Pinxteren, D., Gnauk, T., Müller, K., Herrmann, H., Yan, P., Zhou, X. J., and Wiedensohler, A.: A new method to determine the mixing state of light absorbing carbonaceous using the measured aerosol optical properties and number size distributions, Atmos. Chem. Phys., 12, 2381-2397, 10.5194/acp-12-2381-2012, 2012.

Martin, S. T.: Phase Transitions of Aqueous Atmospheric Particles, Chem. Rev., 100, 3403-3454, 10.1021/cr990034t, 2000.

Meier, J., Wehner, B., Massling, A., Birmili, W., Nowak, A., Gnauk, T., Brüggemann, E., Herrmann, H., Min, H., and Wiedensohler, A.: Hygroscopic growth of urban aerosol particles in Beijing (China) during wintertime: a comparison of three experimental methods, Atmos. Chem. Phys., 9, 6865-6880, 10.5194/acp-9-6865-2009, 2009.

Meng, J. W., Yeung, M. C., Li, Y. J., Lee, B. Y. L., and Chan, C. K.: Size-resolved cloud condensation nuclei (CCN) activity and closure analysis at the HKUST Supersite in Hong Kong, Atmos. Chem. Phys., 14, 10267-10282, 10.5194/acp-14-10267-2014, 2014.

Petters, M. D., and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmospheric Chemistry and Physics, 7, 1961-1971, 2007.

Pinnick, R. G., Jennings, S. G., and Chýlek, P.: Relationships between extinction, absorption, backscattering, and mass content of sulfuric acid aerosols, Journal of Geophysical Research: Oceans, 85, 4059-4066, 10.1029/JC085iC07p04059, 1980.

Rader, D. J., and McMurry, P. H.: Application of the tandem differential mobility analyzer to studies of droplet growth or evaporation, Journal of Aerosol Science, 17, 771-787, http://dx.doi.org/10.1016/0021-8502(86)90031-5, 1986.

Rastak, N., Pajunoja, A., Acosta Navarro, J. C., Ma, J., Song, M., Partridge, D. G., Kirkevåg, A., Leong, Y., Hu, W. W., Taylor, N. F., Lambe, A., Cerully, K., Bougiatioti, A., Liu, P., Krejci, R., Petäjä, T., Percival, C., Davidovits, P., Worsnop, D. R., Ekman, A. M. L., Nenes, A., Martin, S., Jimenez, J. L., Collins, D. R., Topping, D. O., Bertram, A. K., Zuend, A., Virtanen, A., and Riipinen, I.: Microphysical explanation of the RH-dependent water affinity of biogenic organic aerosol and its importance for climate, Geophys. Res. Lett., 44, 5167-5177, 10.1002/2017GL073056, 2017.

Renbaum-Wolff, L., Song, M., Marcolli, C., Zhang, Y., Liu, P. F., Grayson, J. W., Geiger, F. M., Martin, S. T., and Bertram, A. K.: Observations and implications of liquid–liquid phase separation at high relative humidities in secondary organic material produced by α-pinene ozonolysis without inorganic salts, Atmos. Chem. Phys., 16, 7969-7979, 10.5194/acp-16-7969-2016, 2016.

Seinfeld, J. H., and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, 2006.

Sherman, J. P., Sheridan, P. J., Ogren, J. A., Andrews, E., Hageman, D., Schmeisser, L., Jefferson, A., and Sharma, S.: A multi-year study of lower tropospheric aerosol variability and systematic relationships from four North American regions, Atmos. Chem. Phys., 15, 12487-12517, 10.5194/acp-15-12487-2015, 2015.

Tao, J. C., Zhao, C. S., Ma, N., and Liu, P. F.: The impact of aerosol hygroscopic growth on the single-scattering albedo and its application on the NO2 photolysis rate coefficient, Atmos. Chem. Phys., 14, 12055-12067, 10.5194/acp-14-12055-2014, 2014.

Wang, G., Zhang, R., Gomez, M. E., Yang, L., Levy Zamora, M., Hu, M., Lin, Y., Peng, J., Guo, S., Meng, J., Li, J., Cheng, C., Hu, T., Ren, Y., Wang, Y., Gao, J., Cao, J., An, Z., Zhou, W., Li, G., Wang, J., Tian, P., Marrero-Ortiz, W., Secrest, J., Du, Z., Zheng, J., Shang, D., Zeng, L., Shao, M., Wang, W., Huang, Y., Wang, Y., Zhu, Y., Li, Y., Hu, J., Pan, B., Cai, L., Cheng, Y., Ji, Y., Zhang, F., Rosenfeld, D., Liss, P. S., Duce, R. A., Kolb, C. E., and Molina, M. J.: Persistent sulfate formation from London Fog to Chinese haze, Proc Natl Acad Sci U S A, 10.1073/pnas.1616540113, 2016.

Wex, H., Neususs, C., Wendisch, M., Stratmann, F., Koziar, C., Keil, A., Wiedensohler, A., and Ebert, M.: Particle scattering, backscattering, and absorption coefficients: An in situ closure and sensitivity study, Journal of Geophysical Research-Atmospheres, 107, 18, 10.1029/2000jd000234, 2002.

Wu, Z. J., Zheng, J., Shang, D. J., Du, Z. F., Wu, Y. S., Zeng, L. M., Wiedensohler, A., and Hu, M.: Particle hygroscopicity and its link to chemical composition in the urban atmosphere of Beijing, China, during summertime, Atmos. Chem. Phys., 16, 1123-1138, 10.5194/acp-16-1123-2016, 2016.

Zieger, P., Fierz-Schmidhauser, R., Gysel, M., Ström, J., Henne, S., Yttri, K. E., Baltensperger, U., and Weingartner, E.: Effects of relative humidity on aerosol light scattering in the Arctic, Atmos. Chem. Phys., 10, 3875-3890, 10.5194/acp-10-3875-2010, 2010.

Table 1 Abbreviations

RH	relative humidity
$PM_{2.5}$	particulate matter with aerodynamic diameter of less than 2.5 μm
PM_{10}	particulate matter with aerodynamic diameter of less than 10 μm
$f(\mathrm{RH})$	aerosol light scattering enhancement factor at 550 nm
ALWC	aerosol liquid water content: volume concentrations of water in ambient aerosols
$V_a(\mathrm{dry})$	total volume of ambient aerosol particles in dry state
Vg(RH)	aerosol volume enhancement factor due to water uptake
NCP	North China Plain
HTDMA	humidified tandem differential mobility analyser system
PNSD	particle number size distribution
BC	black carbon
g(RH)	hygroscopic growth factor
APS	Aerodynamic Particle Sizer
SMPS	scanning mobility particle size spectrometer
σ_{sp}	aerosol light scattering coefficient
σ_{bsp}	aerosol back scattering coefficient
σ_{ext}	aerosol extinction coefficient
R_{Vsp}	$\sigma_{sp}(550\ nm)/V_a(\mathrm{dry})$
F1 to F6	referred as to five field campaigns listed in Table 2
D1	PNSD, BC and nephelometer measurements from F2, F4 and F5

Table 2. Locations, time periods and used datasets of six field campaigns

Location	Wuqing	Wuqing	Xianghe	Xianghe	Wangdu	Gucheng
Time period	7 march to 4 April, 2009	12 July to 14 August, 2009	22 July to 30 August, 2012	9 July to 8 August, 2013	4 June to 14 July, 2014	15 October to 25 November, 2016
PNSD	TSMPS+APS	TSMPS+APS	SMPS+APS	TSMPS+APS	TSMPS+APS	SMPS+APS
BC	MAAP	MAAP	MAAP	MAAP	MAAP	AE33
σ_{sp}	TSI 3563	TSI 3563	TSI 3563	TSI 3563	TSI 3563	Aurora 3000
$f(\mathrm{RH})$					Humidified nephelometer system	Humidified nephelometer system
Water soluble Ions						IGAC
Campaign Name	F1	F2	F3	F4	F5	F6

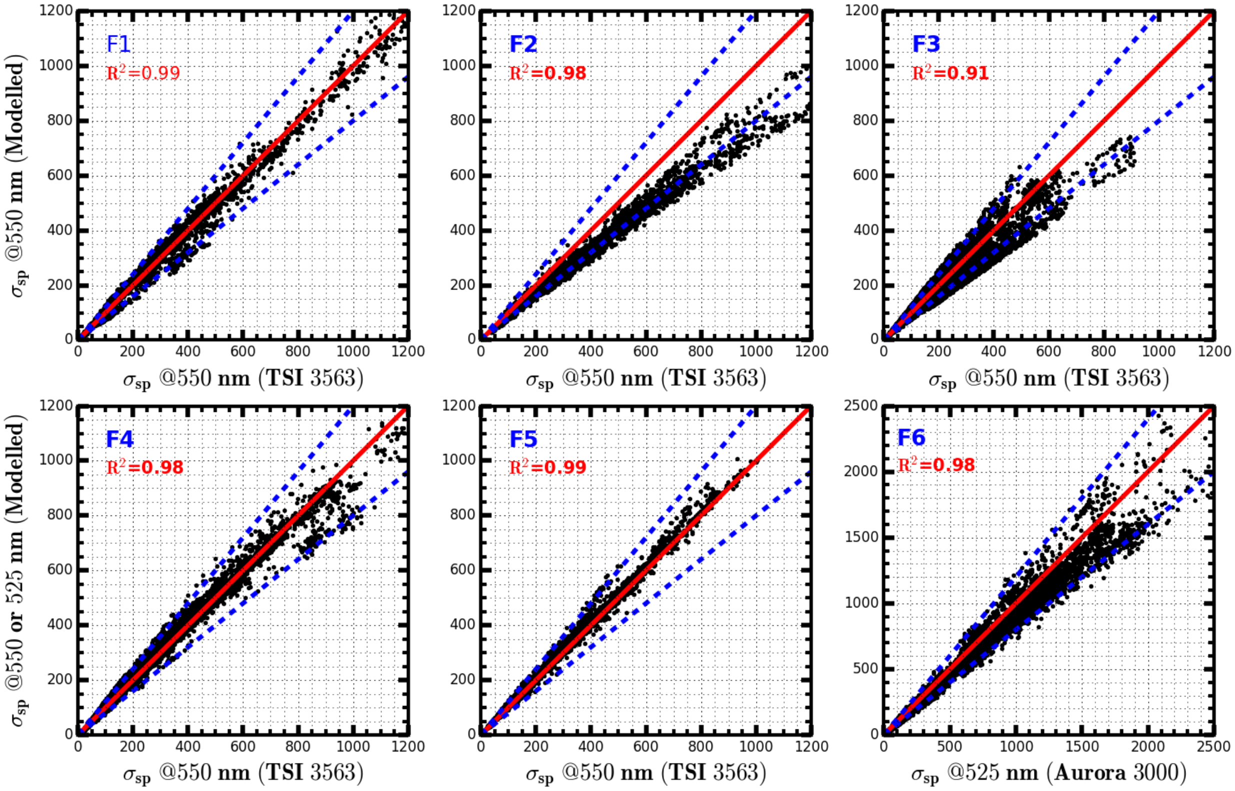

Figure 1. Comparisons between measured and calculated σ_{sp} (Mm^{-1}), solid red lines are 1:1 references lines. Dashed blue lines are 20% relative difference lines. R^2 is square of correlation coefficient between measured and modelled σ_{sp}. Blue texts at the upper left corners are corresponding field campaigns as listed in Table2.

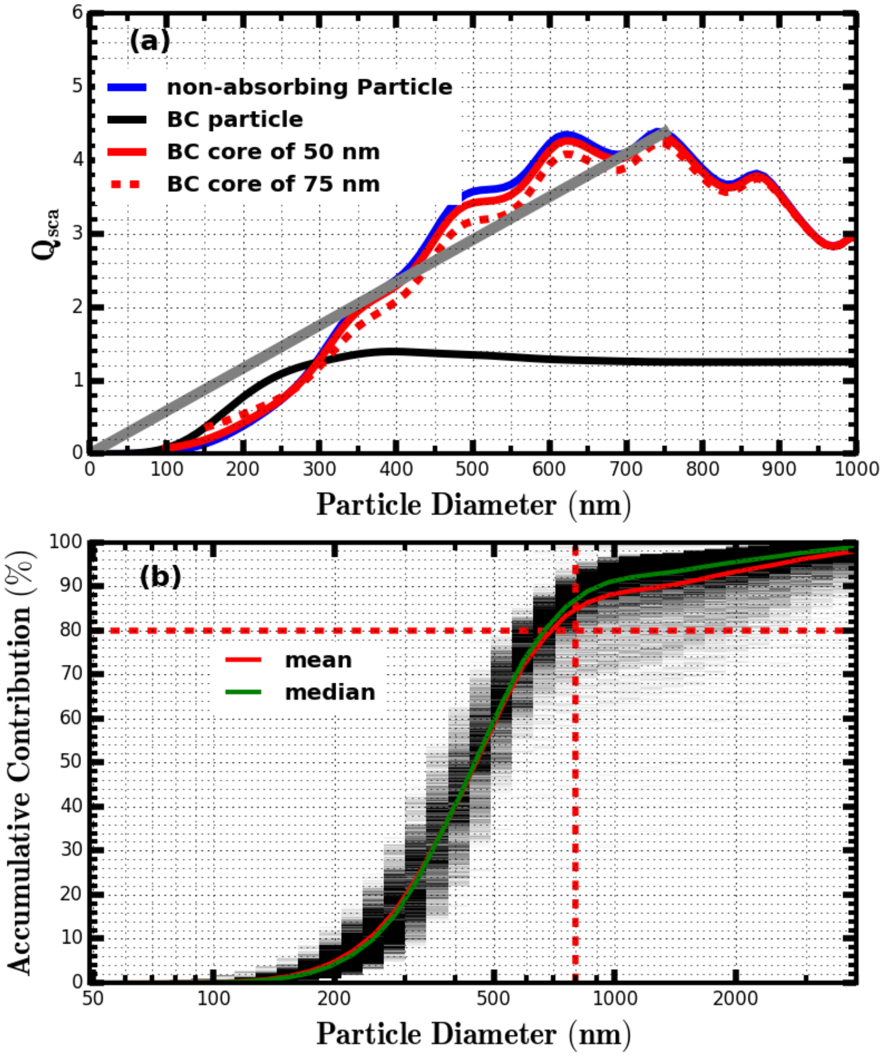

Figure 2. (a) Q_{sca} at 550 nm as a function of particle diameter for four types of aerosol particles: almost non-absorbing aerosol particle, BC particle, BC particle core-shell mixed with non-absorbing components and the radius of inner BC core are 50 nm and 70 nm. The gray line corresponds to the fitted linear line for the case of non-absorbing particle when particle diameter is less than 750 nm. (b) Simulated size-resolved accumulative contribution to σ_{sp} at 550 nm for all PNSDs measured during Wangdu campaign, the color scales (from light gray to black) represent occurrences. The dashed dotted lines in (b) represents the position of 800 nm and 80% contribution, respectively.

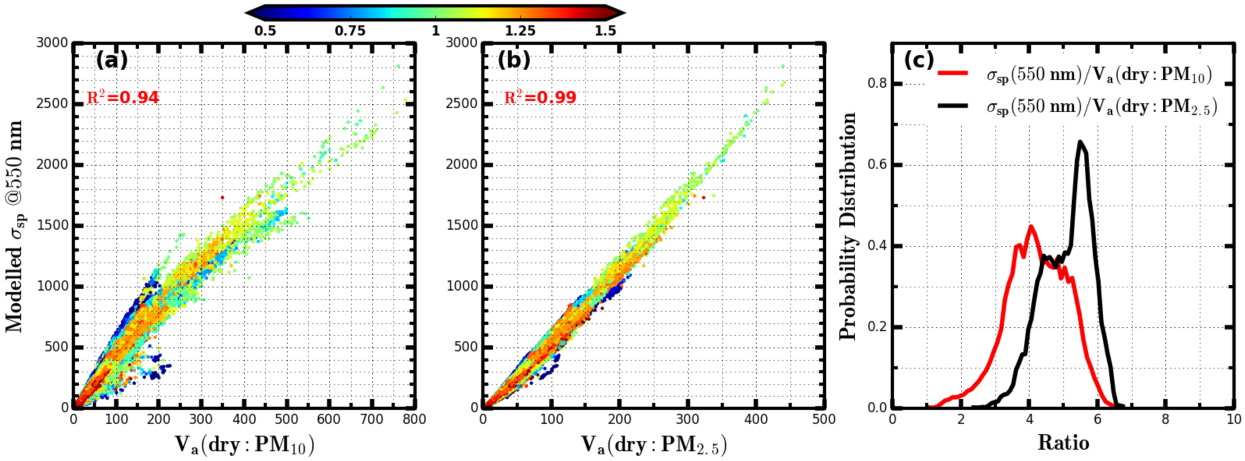

Figure 3. (a) and (b): Modelled σ_{sp} at 550 nm based on PNSD and BC versus V_a(dry) of PM$_{10}$ or PM$_{2.5}$ calculated from measured PNSD. PNSD and BC datasets from six field campaigns listed in Table 2 are used. The unit of V_a(dry) is $\mu m^3/cm^3$, the unit of σ_{sp} is Mm^{-1}. Colors of scattered points in (a) and (b) represent corresponding values of Ångström exponent. R^2 is the square of correlation coefficient. (c) The probability distribution of the modelled ratio between σ_{sp} at 550 nm and V_a(dry) of PM$_{10}$ or PM$_{2.5}$.

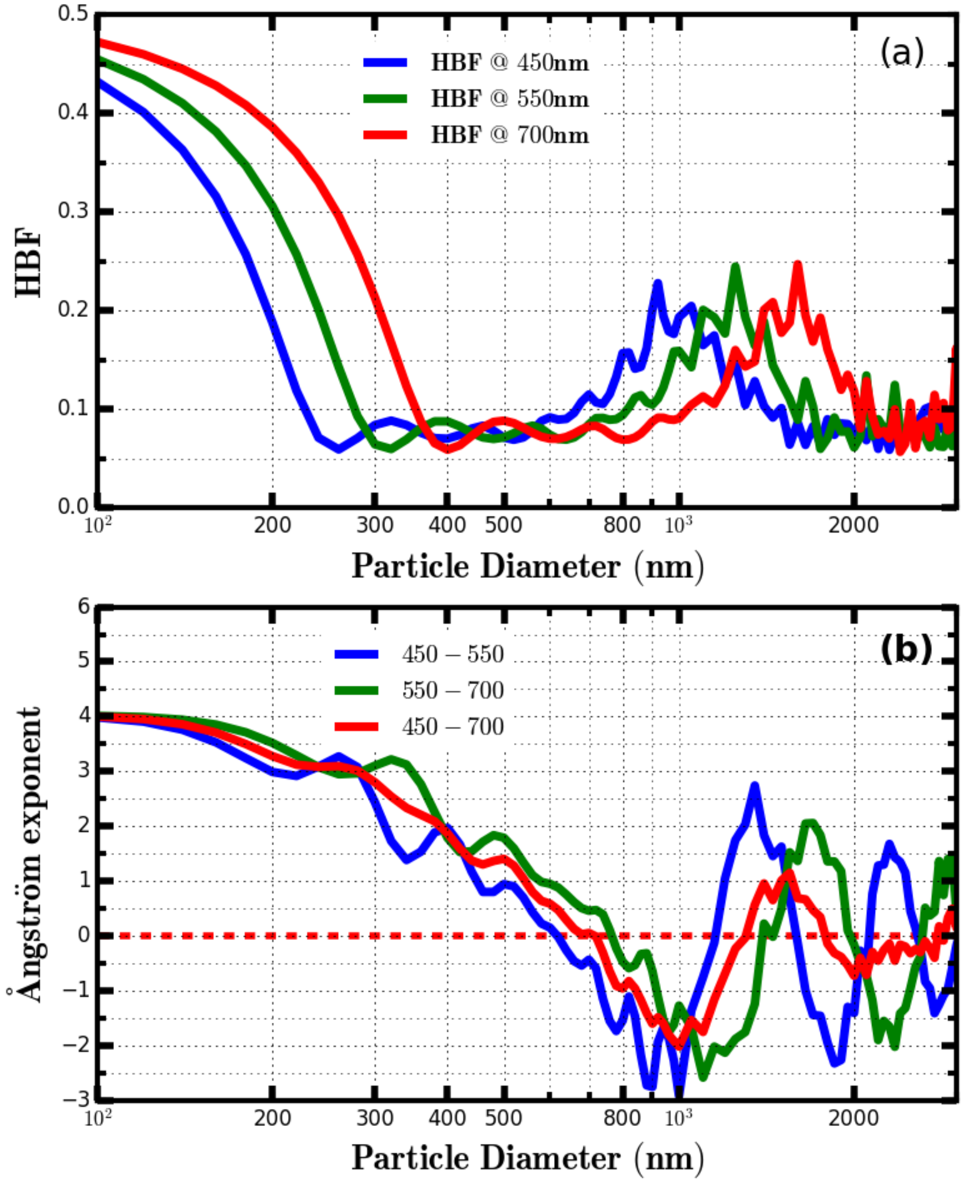

Figure 4. (a) Simulated HBF at three wavelengths as a function particle diameter. (b) Simulated Ångstr*ö*m exponent values as a function a particle diameter.

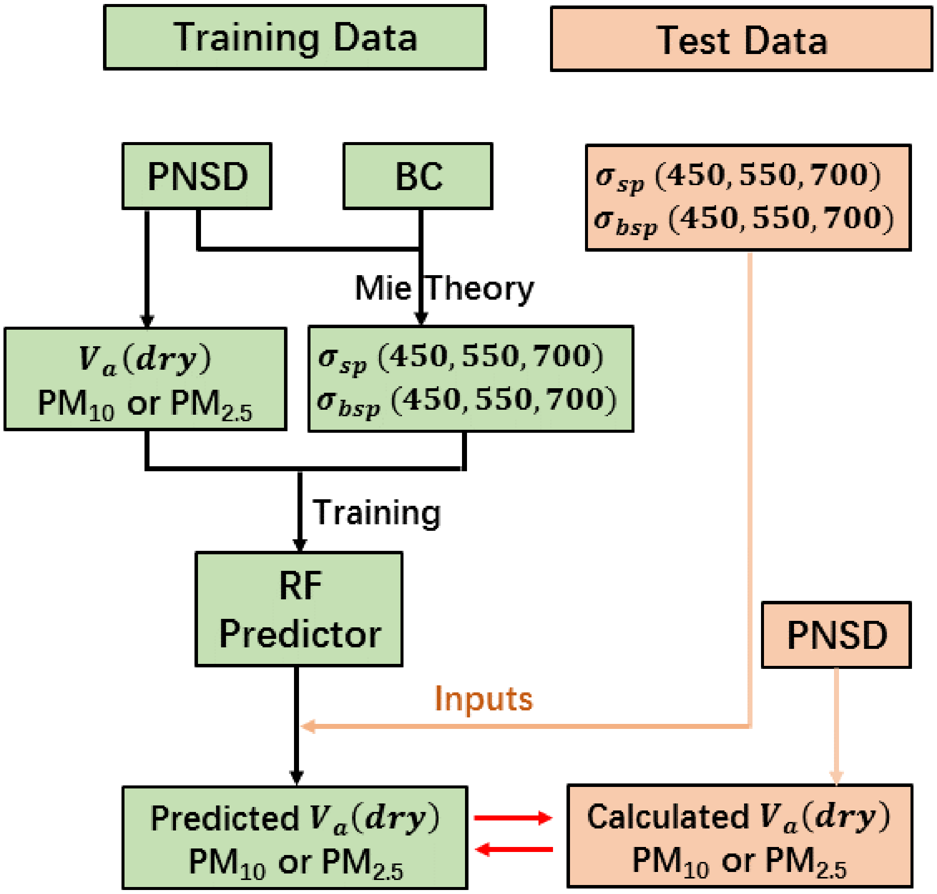

Figure 5. Schematic diagram of training the random forest (RF) model and verifying the performance of trained RF predictor. The trained datasets of PNSD and BC are from field campaigns F1 to F4 and F6, the test datasets of PNSD and optical parameters are from campaign F5, σ_{bsp} is the back scattering coefficient.

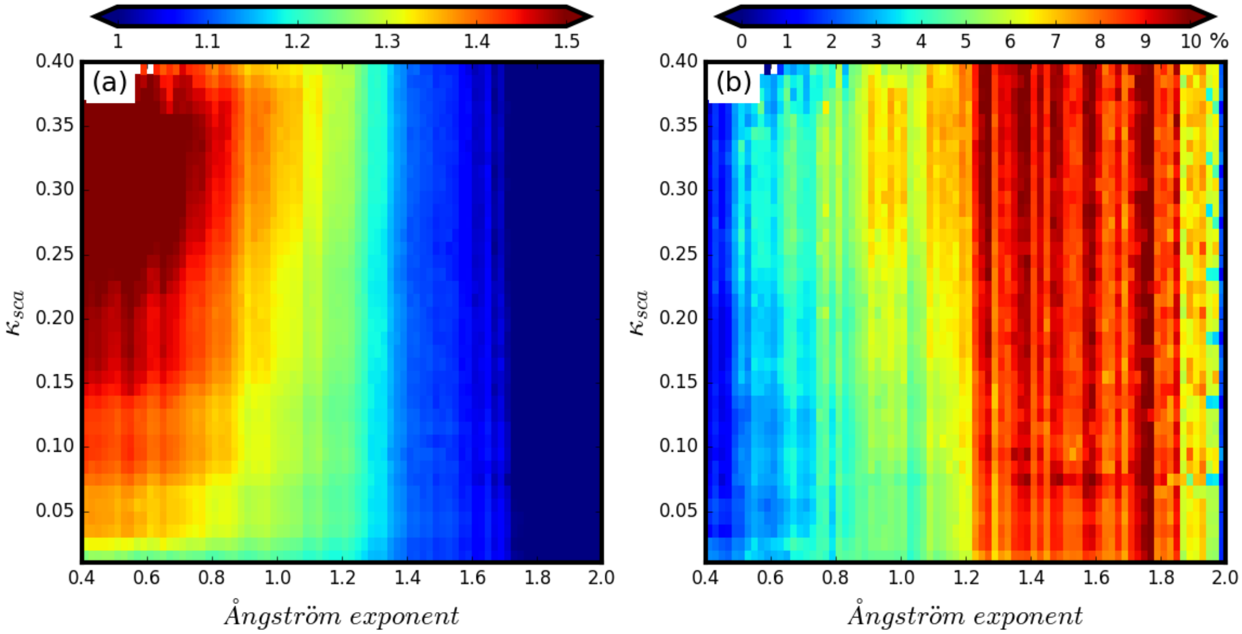

Figure 6. (a) Colors represent R_{Vf} values and the colorbar is shown on the top of this figure, x-axis represents Ångström exponent and y-axis represents κ_{sca}. (b) Meanings of x-axis and y-axis are same with them in (a), however, color represents the percentile value of the standard deviation of R_{Vf} values within each grid divided by their average.

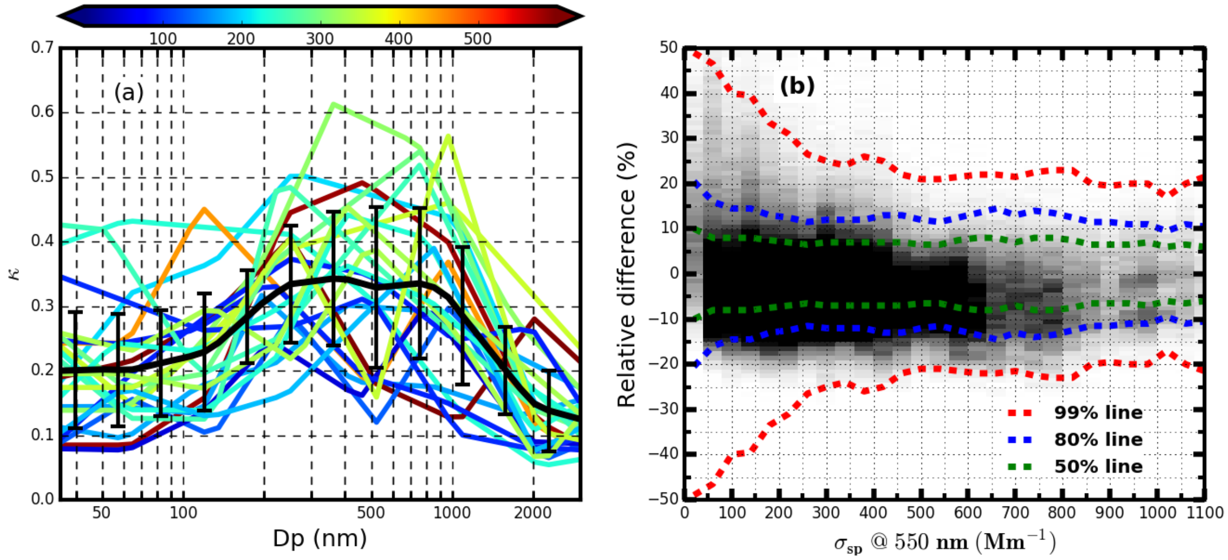

Figure 7. (a) All size-resolved κ distributions which are derived from measured size-segregated chemical compositions during HaChi campaign, colors represent corresponding values of average σ_{sp} at 550 nm (Mm^{-1}), black solid line is the average size-resolved κ distribution and error bars are standard deviations ; (b) The gray

colors represent the distribution of relative differences between modelled and estimated R_{Vf} values, darker grids have higher frequency, dashed lines with the same color mean that corresponding percentile of points locate between the two lines.

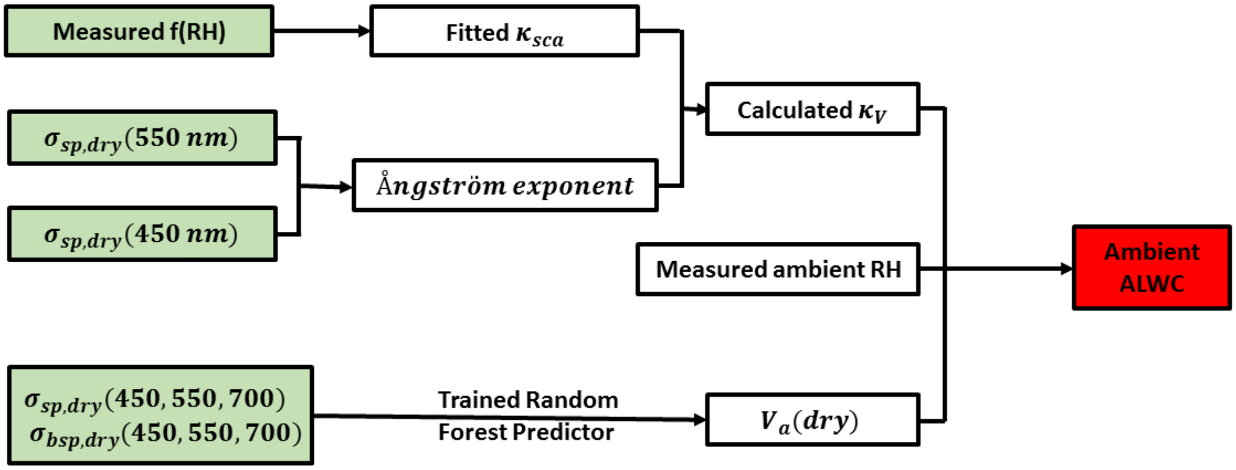

Figure 8. The flowchart of calculating ambient aerosol liquid water contents based on measurements of a three-wavelength humidified nephelometer system.

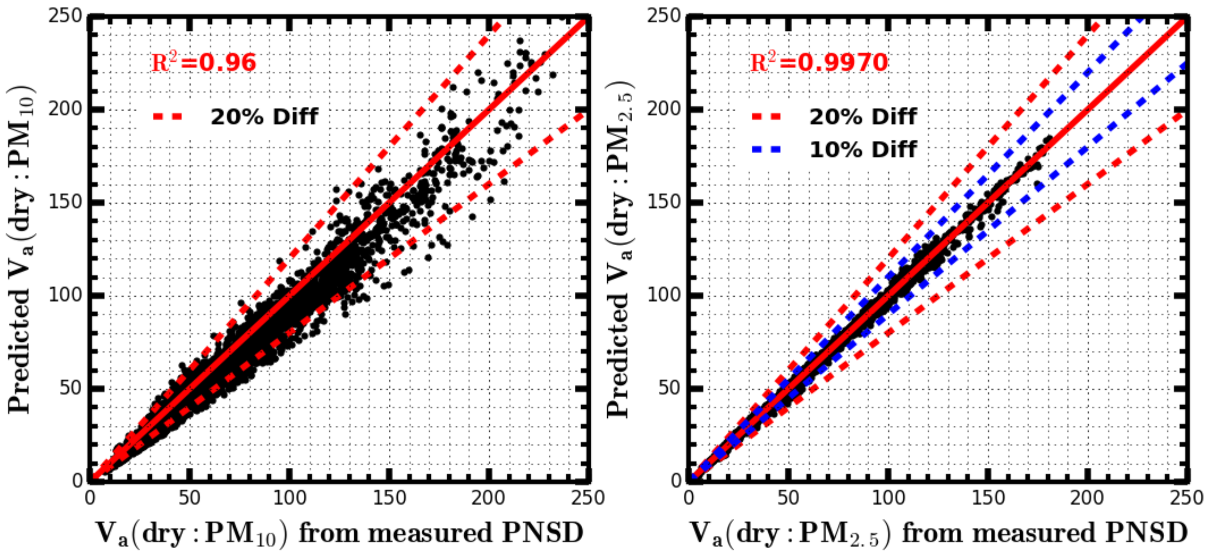

Figure 9. The comparison between V_a(dry) ($\mu m^3/cm^3$) of PM10 or PM2.5 calculated from measured PNSD and V_a(dry) of PM10 or PM2.5 which are predicted based on six optical parameters measured by the "dry" nephelometer by using the random forest model.. R^2 is the square of correlation coefficient? Solid red line is the 1:1 line, dashed red lines and dashed blue lines represent 20% and 10% relative difference lines.

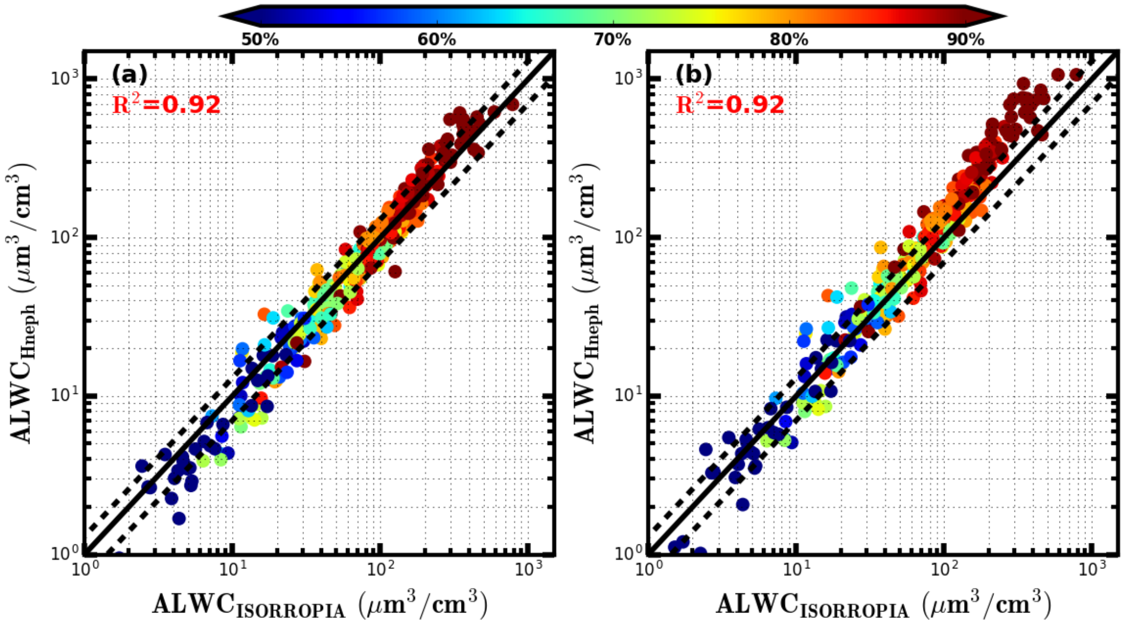

Figure 10. The comparison between ALWC calculated from ISORROPIA thermodynamic model ($ALWC_{ISORROPIA}$) and ALWC calculated from measurements of the humidified nephelometer system ($ALWC_{Hneph}$). The black solid line is the 1:1 line, the two dashed black lines are 30% relative difference lines. R^2 is the square of correlation coefficient? Colors of scatter points represent ambient RH. (a) $ALWC_{Hneph}$ is calculated using the method proposed in this research. (b) $ALWC_{Hneph}$ is calculated by assuming Vg(RH)= f(RH)$^{1.5}$ (Guo et al., 2015).

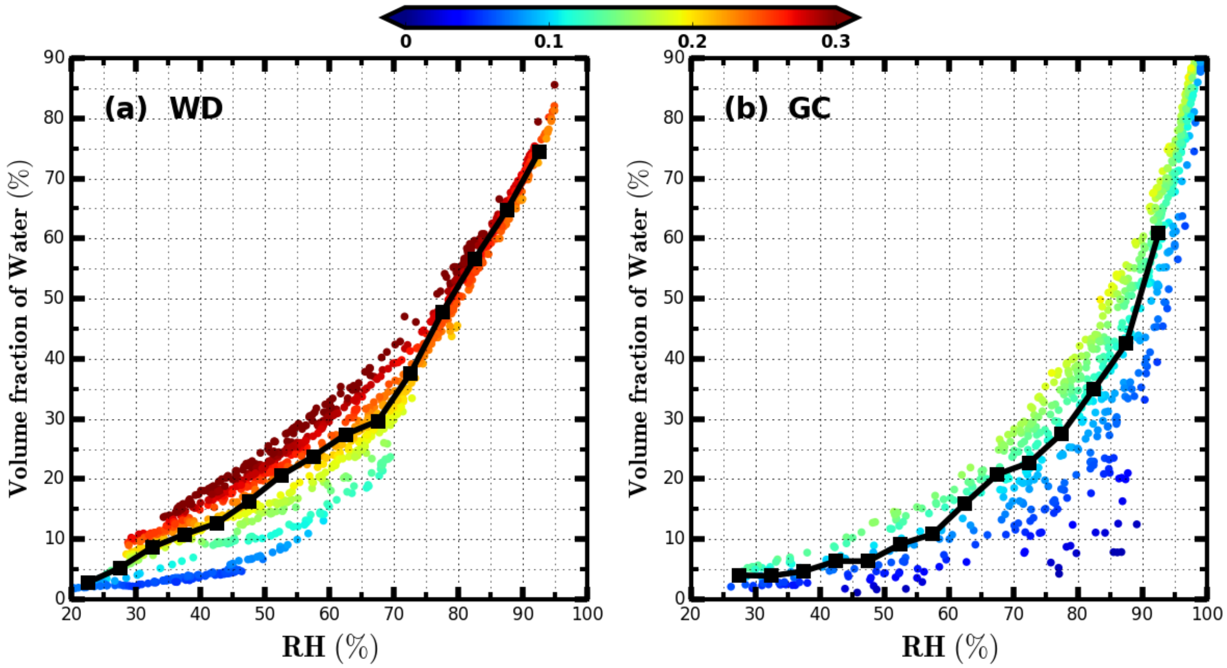

Figure 11. Volume fractions of water in total volume of ambient aerosols during Wangdu (WD) and Gucheng (GC) campaigns. X-axis represents measured ambient RH. Y-axis represents volume fractions of water. Colors of scatter points represent corresponding κ_{Vf}. Black solid lines in (a) and (b) show the average volume fractions of water under different ambient RH conditions.