Response to anonymous referee #2

General comments:

The paper, "A novel method for calculating ambient aerosol liquid water contents based on measurements of a humidified nephelometer system" describes a new technique to measure aerosol liquid water content (LWC). The method described has advantages over traditional methods (ie HTDMA) as it measures LWC of aerosols in real time. In addition, this three wavelength nephelometer system measures the entire size distribution at once without assuming a constant growth factor for the entire distribution, as is the case with previous nephelometer measurements. The LWC of aerosols has important implications for climate forcing and atmospheric chemistry and there is a need for a more accurate measurement to reduce uncertainties.

Response: Thanks for your comments.

Comment: In general, there are large sections of this paper that should be omitted. The content does not support the conclusions of the paper. More importantly, these sections are sometimes confusing and will only distract the reader. This includes the paragraphs describing the relationship between scattering and aerosol volume, which the authors show to be correlated but not easily parameterized. The next section, which describes using the angstrom exponent and HBF to constrain the ratio between scattering and aerosol volume can also be summarized or omitted. The author's conclusion that large bias' will occur if using the "look up table" (fig 4) isn't necessary for the sake of the papers conclusions. More emphasis needs to be placed on what was done that produced usable results.

Responses: Thanks for your comments. We have revised the manuscript substantially. Now the paper has four sections. Section 1 is the introduction. Section 2 is instruments and datasets. Section 3 is methodology. Section 4 is results and discussions. The proposed method has 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. The second step is calculating $\mathrm{Vg(RH)}$ based on Ångstr$\ddot{o}$m exponent and $f(\mathrm{RH})$ measured by the humidified nephelometer system. Therefore, Section 3.1 is the closure calculations between measured and calculated scattering coefficients, because the selection of datasets for training is crucial for machine learning method. In Section 3.2, the calculation of V_a(dry) using the random forest model is discussed. In Section 3.3, the method of calculating $\mathrm{Vg(RH)}$ based on Ångstr$\ddot{o}$m exponent and $f(\mathrm{RH})$ are presented and discussed. In Section 3.4, the formula of calculating ambient ALWC is described. In Section 4.1, the method of predicting V_a(dry) using the trained random forest model is validated by using measurements of Wangdu campaign. In Section 4.2, the calculated ambient ALWC using the proposed method of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ is verified with ambient ALWC calculated from ISORROPIA thermodynamic model. In Section 4.3, the volume fractions of water in ambient aerosols are described and discussed. In Section 4.4, the applicability of the proposed method for calculating ALWC is discussed.

Comment: Finally, the authors describe the machine learning method, which improves the ability to predict the volume of aerosol in the dry state. During the Wangdu campaign (Fig6), this method is able to reproduce the dry volume very well. My main concern is how applicable this method would be in a different type of environment. How difficult would it be to train the estimator to respond to different data sets and measurement conditions?

Responses: Thanks for your comment. Due to the lack of PNSD and nephelometer measurements under different environment conditions, the proposed machine learning method is not validated using measurements from different environment conditions. We cannot be sure if this trained estimator can be applicable under different environment conditions. Thus, we have recommended that the estimator should be trained with regional historical datasets. As what's shown in Fig.5, the training of this random forest model requires only datasets of simultaneously measured PNSD and BC which are already being measured for years in some regions.

Comment: The next section describes parametrizing the relationship between f(RH) and humidified aerosol volume using the "look up table" shown in fig 8. I would recommend referring to both fig 8 and fig 4 as something other than a lookup table, which is not an appropriate description for this plot. This approach, once again seems limited by the specific environment. It would be nice if the authors showed results from a different less-polluted region.

Response: Thanks for your comment. Based on your comment, results of Fig.4 are omitted. As to results of Fig.8 (Fig.6 in the revised manuscript), we have validated the way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ by using results in Fig.6 as a look up table with ambient ALWC calculated from ISORROPIA model. And we also have compared results of this method with the results of using the traditional way of calculating $\mathrm{Vg(RH)}$ based on $f(\mathrm{RH})$ (Guo et al., 2015). The results indicate that the proposed method can improve the calculation of $\mathrm{Vg(RH)}$ based on measured $f(\mathrm{RH})$.

Comment: The paper has multiple typographical and grammatical errors. Line 365 is a good example of the grammatical/typographical errors that are found throughout. Line 405 references a figure 20a, which doesn't exist. Prior to publication I would recommend careful editing for these errors as well as re-formatting to streamline the paper for only the most pertinent information.

Response: Thanks for your comment. We have edited the English with a copy editor and checked typographical and grammatical errors.

References

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.

Response to anonymous referee #3

General Comments:

The authors present both empirical and machine-learning methods for determining an aggregate or effective volume and the aerosol water content from the nephelometer Angstrom and backscattering coefficients. The empirical method uses cross correlations between the aerosol scattering coefficient, measured volume, backscatter fraction, and Angstrom exponent to estimate the aerosol volume. The machine learning method uses backscatter and Angstrom exponents to mimic the aerosol scattering to volume ratio. The machine-learning method offers a valuable tool that could be applied to many aspects of atmospheric aerosol and chemical predictions. The paper needs further development. I think it's important to refine the methodology, improve the paper organization, and clarify some of text to make this a stronger paper. The section linking scattering hygroscopic growth to volume hygroscopic growth doesn't follow a valid analysis method. The fRH and gRH data come from different measurement sites. I suggest leaving out the sections on volume hygroscopic growth as well as the discussion on kappa-Kohler. I don't recommend publication until the paper is restructured.

I suggest resubmitting the paper after removal of the sections on hygroscopic growth and water content.

Responses: Thanks for your comments and insightful suggestions. Based on your comments, we have refined the methodology and reorganized the paper. As to the section linking scattering hygroscopic growth to volume hygroscopic growth, this section is an important part of our methodology, and is moved to the methodology section. For the size-resolved κ distributions used for simulating the relationship between scattering hygroscopic growth to volume hygroscopic growth, only the average shape of the size-resolved κ from HaChi is used because that 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.7a) to produce a number of size-resolved κ distributions which represent aerosol particles from nearly hydrophobic to highly hygroscopic. Results from other studies have shown similar size dependence of aerosol hygroscopicity (Meng et al., 2014). We also have done the comparison between ambient ALWC (aerosol liquid water content) calculated from measurements of the humidified nephelometer system by using the proposed method and ambient ALWC calculated from ISORROPIA thermodynamic model. A good agreement is achieved between them. But if use the traditional way of connecting $f(\mathrm{RH})$ (scattering enhancement factor) to $\mathrm{Vg(RH)}$ (volume growth factor), the ambient ALWC tends to be significantly overestimated, especially when RH is higher than 80%. Thus, we think this part provides a new way for connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ which is useful for estimating ambient ALWC based on measurements of the humidified nephelometer system.

Comment: The paper needs better organization and clear, step-wise presentation of the methods and results. The methodology is scattered throughout the paper. Description of the results is vague and tends to gloss over important features.

Response: Thanks for your comment. We have revised the methodology part substantially. The method of estimating $V_a(\mathrm{dry})$ using the machine learning method and the way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ are moved to the methodology section.

Comment: Please edit the English grammar and word order. Avoid run-on sentences. You have attendency to state conclusions without providing supporting evidence. State the methods used and then the data results. The methods and results are interspersed in the paper, which adds to confusion.

Response: Thanks for your comment. We have edited the English by another copy editor. In the revised manuscript, the methods and results are separated. The used datasets are introduced in Sect.2. Calculation method of $V_a(\mathrm{dry})$ based only on measurements of the nephelometer is described in Sect.3.2. The way of deriving $\mathrm{Vg(RH)}$ 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 $V_a(\mathrm{dry})$ 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. And the contribution of ambient ALWC to total ambient aerosol volume is discussed in Sect.4.3.

Comment: Introduction 1. Although the methodology is different from other inversion techniques, such as those of Ziegar et al., that calculate an "effective" gRH from fRH, the methodology is similar to that used in Aeronet retrievals of the aerosol effective radius from the AOD, Angstrom exponent and asymmetry parameter. There was a paper that attempted to calculate fRH or aerosol water using Aeronet data, but it suffered from low signal and spatial resolution. Can the authors describe how their method is similar to or different from the Aeronet retrievals and also speculate if this method could be used with remote sensing AOD measurements? Below is a link that has links to their "spectral deconvolution algorithm".

https://aeronet.gsfc.nasa.gov/new_web/Documents/Inversion_products_V2.pdf or Atmos. Meas. Tech., 10, 695-708, 2017, https://doi.org/10.5194/amt-10-695-2017.

Response: Thanks for your comment. We think our method is different with Aeronet retrievals. Our method is dealing with optical properties of aerosols at one location. Aeronet retrievals are dealing with integral optical properties of aerosols that distributed from the surface to the top of the atmosphere and assumes that aerosols are homogeneously distributed across the vertical layer. However, in real world, microphysical properties of aerosol particles (aerosol size distributions, aerosol hygroscopicity, aerosol mixing state, et al) at different altitudes are different, and relative humidity of the air at different altitudes also differs greatly (Kuang et al., 2016). Our method is based on machine learning which learn from historical datasets, and six parameters are used to constrain variations in PNSD in the machine learning method. The way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg}(\mathrm{RH})$ is based on simulative experiment which only assumes an average shape of size dependence of aerosol hygroscopicity and the variation of bulk aerosol hygroscopicity is considered. The Aeronet retrievals retrieves the particle number size distribution, complex refractive index and partition of spherical/non- spherical particles which fits the observed data best. We think the method proposed in this research can not be used with remote sensing AOD measurements, because it is difficult to use several parameters to constrain PNSD and RH variations at different altitudes, too much about aerosol properties and aerosol vertical distribution are unknown.

Comment: 2. The introduction needs to state more about the methodology other than saying it's "a novel method". Add a paragraph describing the two techniques. Describe the empirical model use of size-dependent parameters (backscatter and Angstrom) to predict the ratio of scattering/volume. Describe how machine-learning methods, using large data sets, over a long-time period as input, mimic the system behavior via feed-back loops to predict an output.

Response: Thanks for your comment. The following paragraph is added to the introduction: "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 $V_a(\mathrm{dry})$ 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 $V_a(\mathrm{dry})$ can also be calculated based on PNSD. Therefore, the random forest model can be trained with only regional historical datasets of PNSD and BC. The second step is calculating $\mathrm{Vg(RH)}$ based on the $\text{Ångstr}ö\text{m}$ exponent and $f(\mathrm{RH})$ measured by the humidified nephelometer system. In this step, both the influences of the variations in PNSD and aerosol hygroscopicity are both taken into account to derive $\mathrm{Vg(RH)}$ from measured $f(\mathrm{RH})$. Finally, based on calculated $V_a(\mathrm{dry})$ and $\mathrm{Vg(RH)}$, ALWCs at different RH points can be estimated. The used datasets are introduced in Sect.2. Calculation method of $V_a(\mathrm{dry})$ 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 $\mathrm{Vg(RH)}$ 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 $V_a(\mathrm{dry})$ 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. ". The random forest model is introduced in the methodology section.

Comment: 3. In Section 2.1 Transfer Table S1 to Section 2.1. You can leave the detailed measurement description in the supplement. Give general information on the breakdown of the aerosol composition between organics/sulfate/nitrate/dust at this time of year. Sections 2.2 to 3.2 need to be reorganized. I suggest segregating Section 2 into Section 2.2 is closure, 2.3 is Mie theory, 2.4 Machine learning method.

I suggest renaming section 2.2 as "Closure Calculations". Show the scattering closure between the measured and calculated scattering coefficient. The integrity of the volume and fRH closure depends on the scattering closure. Figure 2 would validate the measurements better if 2a showed the scattering closure and 2b showed the scattering vs volume.

Response: Thanks for your comment. Section 2 is about the instruments and datasets. Table S1 is transferred to this section, more details are listed in this Table. Section 2 and Section 3 are reorganized. The closure results between the measured and calculated scattering coefficient during different campaigns are introduced in Section 2.1. The discussions about theoretical relationships between scattering coefficient and aerosol volume are introduced and discussed in Section 2.2.1. The machine learning method is introduced in Section 2.2.2. The proposed method of connecting $f(\mathrm{RH})$ to $\mathrm{Vg}(\mathrm{RH})$ is introduced in Section 3.

Comment: Figure 2 shows 2 branches or subsets of the data; one above and a 2nd below the fit line. Is this behavior present in the scattering closure? Do these two branches represent 2 different aerosol types or multiple size modes?

Response: Thanks for your comment. These two branches exist for datasets of campaign F1 (please refer to Table 2 of the revised manuscript for detailed campaign information). As shown in Fig.1 of the revised manuscript. This behavior does not present in the scattering closure. The relationship between $V_a(dry)$ and σ_{sp} for measurements of

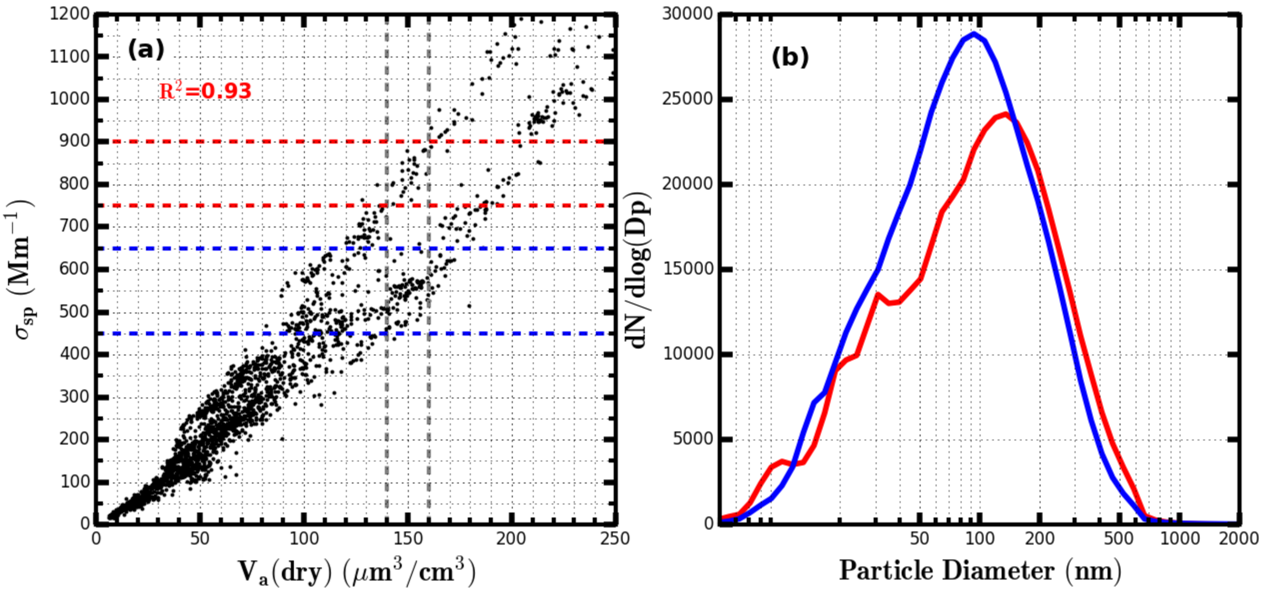


Figure 1. (a) The relationship between $V_a(dry)$ and σ_{sp} at 550 nm for measurements of campaign F1. (b) The blue one corresponds to average PNSD of data points of lower branch which locate in the range of dashed blue lines in (a), red one corresponds to average PNSD of data points of upper branch which locate in the range of dashed red lines in (a).

campaign F1 is shown in Figure 1a. And the average PNSDs of chosen data points of lower and upper branches are shown in Fig.1b. These results indicate that two branches corresponding two different PNSD shapes, but without multiple size modes.

Comment: The application of an average Rvsp to estimate the volume doesn't add to the paper and distracts from the other methods. I suggest removing it from the analysis.

Response: Thanks for your comment. This part is removed from the manuscript.

Comment: Section 2.3 Mie Move equations 5 and 6 to the start of section 2.3 and explain how you relate scattering to volume and the assumptions in this approach. Move the discussion of Mie theory from section 3.1 to section 2.3. State your adaptation of Mie Theory in a clear, stepwise, logical fashion. Show that going from equations 5 > 7 >6 assumes that Q is roughly linear with r such that Q=k*Q(m).

Describe Mie model using simulated data with 4 aerosol types and results in Figure 3.

Describe limitations of assumption that Q is linearly proportional to r.

Describe what the dotted red lines represent in the Figure 3 caption.

Response: Thanks for your comment. We have moved equations 5 and 6 to the section 3.2. The name of this section is "calculation of $V_a(\mathrm{dry})$ based on measurements of the "dry" nephelometer". In Sect.3.2.1, we have described the theoretical relationship between $V_a(\mathrm{dry})$ and σ_{sp}. In Sect.3.2.2, we have described the machine learning method. The discussion of the Mie theory is also moved to Sect.3.2. Limitations of the assumption that Q is linearly proportional to r is discussed in Sect.3.2.1. The dotted red lines in Figure 3 (Figure 2 of the revised manuscript) are described.

Comment: Section 2.5 Machine learning

Describing an alternate method of machine learning using size-dependent scattering parameters; Angstrom and backscatter to mimic the measured scattering/volume ratio. Give some background on "machine learning" and the algorithm name. Can you add a simplified algorithm decision tree with basic logic steps or diagram that would help explain the process?

Response: Thanks for your comment. The background on machine learning and the algorithm name is described in Sect.3.2.2. A schematic diagram of training the machine learning method is also shown in Figure 5.

Comment: Results and Discussion: Section 3.1 Empirical method Refer back to Figure

2 and need for estimating a variable kscat or R. Introduce the empirical method of determining R from HBF and Angstrom exponent. Explain figure 4.

In your description of the results, use and simple and direct language. Leave out extra information on the impact of BC and mixing state on the HBF and Angstrom exponents until the discussion of the data fit lines as these are secondary contributions.

Explain Figure 5a and variance about fit line in relation to HBF. This variance likely stems from the Angstrom exponent and HBF describing a fraction of the PNSD. Show a plot of HBF(450, 550, 700nm) and Angstrom exponents(450/550, 450/700 and 550/700) versus r for a lognormal size distribution. The plot will show the sensitivities of these parameters to aerosol size.

Response: Thanks for your comment. Comments from another reviewer suggest that results of Figure 4 and Figure 5a should be omitted. We agree with reviewer and focus on the machine learning method. HBF(450, 550, 700nm) and Angstrom exponents (450/550, 450/700 and 550/700) as a function of particle diameters are shown in Fig.4 of the revised manuscript. The results shown in Fig.4 indicate that HBFs at three wavelengths and Ångström exponents calculated from σ_{sp} at different wavelengths are sensitive to different diameter ranges of PNSD.

Comment: Section 3.2 Machine learning Explain how machine method uses 6 parameters to describe the aerosol volume relative to the scattering. Explain figures 5b and Figure 6 and how machine learning is an improvement over the empirical method.

Response: Thanks for your comment. A schematic diagram of training the machine learning method is shown in Figure 5. Why machine learning is an improvement is explained in Sect.3.2.2.

Comment: 3.3 fRH and Vrh

The method of simulating the Kd size distribution from variations of the average and then applying this to the measured size distribution to obtain a 4 modeled volume growth values isn't valid. The PNSD shape will change with aerosol type as will the Kd size distribution. Aerosol size-dependent growth varies with size such that multiplying an entire Kd distribution by a constant won't reproduce the Kd distribution of a different aerosol types. In addition the Kd and frh data come from two different measurement sites.

Response: Thanks for your comment. Section 3.3 is an important part of our methodology, and is moved to the methodology section of the revised manuscript. For the size-resolved κ distributions used for simulating the relationship between scattering hygroscopic growth to volume hygroscopic growth, only the average shape of the size-resolved κ from HaChi is used because that 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.7a) to produce a number of size-resolved κ distributions which represent aerosol particles from nearly hydrophobic to highly hygroscopic. We agree with the reviewer that PNSD shape as well as size-resolved κ distribution will change. However, $f(\mathrm{RH})$ and $\mathrm{Vg(RH)}$ are integral variables which are sensitive to integral variables which can represents variations in PNSD and overall aerosol hygroscopicity. This is why we establish a look up table which can take the variations of bulk aerosol hygroscopicity and the Ångström exponent into account. We also have examined how much the variations in shape of size-resolved κ distribution and PNSD will impact on the prediction ability of the established look up table based on measure size-resolved κ distributions. The results are shown in Figure 7b. Moreover, results from other studies have also shown similar size dependence of aerosol hygroscopicity (Meng et al., 2014). We also have done the comparison between ambient ALWC (aerosol liquid water content) calculated from measurements of the humidified nephelometer system by using the proposed method and ambient ALWC calculated from ISORROPIA thermodynamic model. A good agreement is achieved between them. A traditional way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ (Guo et al., 2015) is also described and discussed in Sect.4.3. If we use the traditional way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$, the ambient ALWC tends to be overestimated significantly, especially when RH is higher than 80%. Thus, we think this part provides a new way of connecting $f(\mathrm{RH})$ to $\mathrm{Vg(RH)}$ which is useful for estimating ambient ALWC based on measurements of the humidified nephelometer system.

References

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., 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.

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.

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

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

~~[1~~ [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}

~~[2]{Institute for Environmental and Climate Research, Jinan University, Guangzhou 511443, China}~~

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

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

Abstract

Formatted: Justified

Formatted: Not Superscript/ Subscript

Formatted: Not Superscript/ Subscript

Water condensed on ambient aerosol particles plays significant roles in atmospheric environment, atmospheric chemistry and climate. So far, no instruments ~~are~~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~~. A humidified nephelometer system~~, which measures aerosol light scattering coefficients and backscattering coefficients at three wavelengths under dry state and different relative humidity (RH) conditions, ~~and therefore provides~~providing measurements of light scattering enhancement factor $f(\mathrm{RH})$. The proposed ALWC calculation method ~~of calculating ALWC~~ includes two steps. The first step is ~~estimating~~ the estimation of the dry state total volume concentration of ambient aerosol particles~~ in dry state (V_a,~~ V_a(dry)~~)~~, with a machine learning method called random forest model based on measurements of the “dry” nephelometer. The estimated V_a(dry) agrees well with the measured ~~V_a(dry)~~ ~~.~~one. The second step is ~~estimating~~the estimation of the volume growth factor Vg(RH) of ambient aerosol particles due to water uptake, using $f(\mathrm{RH})$ and Ångström exponent. The ALWC is calculated from the estimated V_a(dry) and Vg(RH). ~~Uncertainty analysis~~ To validate the new method, the ambient ALWC calculated from measurements of the ~~estimated V_a(dry)~~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 ~~Vg(RH) is conducted. This research have bridged the gap between $f(\mathrm{RH})$~~intercept of 1.14 and ~~Vg(RH).~~ -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 ~~using only~~solely based on measurements ~~from~~of a three-wavelength humidified nephelometer system~~. This method will facilitate~~, facilitating the real-time monitoring of the ambient ALWC and ~~help for studying roles~~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 ~~both~~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). ~~And aerosol~~Aerosol liquid water also has large impacts on aerosol optical properties and aerosol radiative effects (Tao et al., 2014;Kuang et al., 2016). ~~Condensed liquid~~Liquid water condensed on aerosol particles can also ~~serve~~serves as a site for multiphase reactions which perturb local chemistry and ~~also~~ further influence the aging processes of aerosol particles (Martin, 2000). Recent studies have shown that aerosol liquid water serves as a reactor, which ~~help for~~can efficiently ~~transforming sulfur~~transform sulphur dioxide to ~~sulfate~~sulphate during haze events ~~and plays crucial roles in worsening~~, aggravating atmospheric environment ~~on~~in the North China Plain (NCP) (Wang et al., 2016;Cheng et al., 2016). Hence, ~~the real-time monitoring of ambient aerosol liquid water content (ALWC) is of crucial importance~~ to gain more ~~insights~~insight into the ~~roles~~role of aerosol liquid water in atmospheric chemistry, aerosol aging processes and aerosol optical

Formatted: Justified

properties~~.~~, the real-time monitoring of ambient aerosol liquid water content (ALWC) is of crucial importance..

Few techniques are curruently 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 ~~volumes~~the volume of liquid water. Nevertheless, HTDMAs ~~can not~~cannot be used to measure the total aerosol water volume, because they are not capable of measuring the hygroscopic properties of the entire aerosol ~~size distribution~~population. With size distributions of aerosol particles in their ambient state and dry state, the aerosol water volume can ~~de~~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).

~~A~~The idea of using the humidified nephelometer system, ~~which measures aerosol light scattering coefficient (σ_{sp}) under dry and different RH conditions is a relatively early method proposed~~ for ~~studying~~the study of aerosol hygroscopicity has already been proposed very early on(Covert et al., 1972). ~~It provides information about~~The instrument measures aerosol light

Formatted: English (U.S.)

scattering coefficient (σ_{sp}) under dry state and different RH conditions, providing information on aerosol light scattering enhancement factor f(RH) . One advantage of this method is that it has a fast response time and ~~thus measurements can be made continuously which facilitates the monitoring of changing ambient conditions. Another advantage of this method is that it provides information about the overall aerosol hygroscopicity of the entire aerosol size distribution.~~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(RH) vary strongly with parameters of particle number size distribution (PNSD), making it ~~is~~ difficult to directly link them with ~~volume of aerosol particles in~~the dry state aerosol particle volume (V_a(dry)) and the volume growth factor Vg(RH) of the entire aerosol population. So far, the ALWC ~~can~~could not be directly estimated ~~with~~ based solely on measurements ~~from only a~~of the humidified nephelometer system. Several studies have shown that ~~if~~given the PNSDs ~~in~~at dry state ~~are measured, then~~, an iterative algorithm together with the Mie theory can be used to calculate an overall aerosol hygroscopic growth factor g(RH)- based on measurements of f(RH) (Zieger et al., 2010;Fierz-Schmidhauser et al., 2010). In ~~this~~such an iterative algorithm, the g(RH) is assumed to be independent of the aerosol diameter. Then ALWC at different RH ~~points~~levels can be calculated based on derived g(RH) and the measured PNSD. This method not only requires additional measurements ~~about PNSD, but also may result in significant deviations of the estimated ALWC because that g(RH) should be a function of aerosol diameter rather than a constant value. In this paper, we proposed~~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(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. 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 taken into account to derive Vg(RH) from measured f(RH). Finally, based on calculated V_a(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(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(dry) predicted by using the machine learning

Formatted: Justified

Formatted: Font: Times New Roman, English (U.S.)

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

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. ~~Materials~~ Instruments and ~~methods~~datasets

~~2.1 Datasets~~

Datasets from six field campaigns ~~are~~were used in this paper. The six campaigns ~~are~~were conducted at four different measurement sites (Wangdu, Gucheng and Xianghe in Hebei province and Wuqing in Tianjin) of the North China Plain (NCP), ~~site~~the locations ~~-~~of these field ~~campaigns~~campaign sites are ~~shown~~displayed in Fig.S1. Time periods and ~~used~~ datasets used from these ~~filed~~field campaigns are listed in Table 1. During these field campaigns, ~~sampled~~ aerosol particles ~~have~~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 ~~called~~ (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.

Formatted: Justified

~~2.2~~ Datasets ~~about PNSD, BC and σ_{sp} from campaigns F1 to F5 are referred to as D1. Datasets about~~of PNSD, BC and σ_{sp} from campaigns F2, F4 and F5 are referred to as ~~D2~~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 ~~detailedly~~ introduced in ~~(Kuang et al., 2017b).**Mie theory**~~

~~The first goal of this research is estimating V_a(dry) from σ_{sp}~~ 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.

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

3. Methodology

3.1 Closure calculations

To ensure the V_a(dry) ~~can be integrated from measured PNSD. Thus,~~ datasets of σ_{sp} ~~and PNSD are needed to investigate relationships between σ_{sp}~~ and ~~V_a(dry). To make sure the data quality of~~PNSD used ~~datasets~~are of high quality, a closure study between measured σ_{sp} and ~~PNSD, the closure between measured σ_{sp} and σ_{sp}~~that calculated based on measured PNSD and BC with Mie theory (Bohren and Huffman, 2008) is first ~~investigated~~performed. Measured σ_{sp}

Formatted: Justified

~~has problems regarding~~bears uncertainties introduced by angular truncation errors and ~~nonideality of~~nonideal light source. ~~In order to make sure the~~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). ~~Moreover, during processes of~~During the σ_{sp} modelling ~~σ_{sp},~~process, BC ~~is~~was considered to be half externally and half coreshell mixed with other aerosol components~~, and the~~. The mass size distribution of BC used in Ma et al. (2012), which was also observed ~~on~~in the NCP~~ is~~, was used in this research to account for the mass distributions of BC at different particle sizes. The ~~used~~applied refractive index and density of BC ~~are~~were $1.80 - 0.54\mathrm{i}$ and 1.5g cm^{-3} (Kuang et al., 2015). ~~Used~~The refractive index of non light-absorbing aerosol components (other than BC) ~~is1~~was set to 1.53 $- 10^{-7}i$ (Wex et al., 2002). ~~Calculation details based on~~For the Mie theory calculation details please refer to Kuang et al. (2015).~~ Datasets about PNSD and σ_{sp} from D1 are used to perform the closure investigation. Finally, during processes of investigating relationships between σ_{sp} and V_a(dry), data points in D1 with relative differences between measured σ_{sp} at 550 nm and modelled σ_{sp} at 550 nm greater than 10% are excluded. 10% is chosen because of that measured PNSD has uncertainty of larger than 10% (Wiedensohler et al., 2012), and measured σ_{sp} has uncertainty of about 9%, this threshold can make sure that most used data points are measured when instruments operated well. .~~

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 $V_a(\mathrm{dry})$ (Pinnick et al., 1980). Here, the quantitative relationship between $V_a(\mathrm{dry})$ and σ_{sp} is analyzed.

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

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

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

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 $V_a(\mathrm{dry})$ 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

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: Font: Times

Formatted: Justified

component can be referred to as BC. BC particles are neither externally nor 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 $1.80-0.54\mathrm{i}$ and $1.53-10^{-7}\mathrm{i}$, 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 diameteer. 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 (5) can be expressed as the following:

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

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

Formatted: Justified, Indent: First line: 2 ch 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(dry).

Based on PNSD and BC datasets of field campaigns F1 to F6, the relationship between σ_{sp} at 550 nm and V_a(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 Ångström 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. Ångström 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 Ångström 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(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(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(RH) to Vg(RH)

~~2.3~~3.3.1 κ-K$\ddot{o}$hler theory~~(Wiedensohler et al., 2012)~~

~~To simulate the relationships between f(RH) and Vg(RH),~~ 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):

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

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

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: English (U.S.)

Formatted

Formatted: Justified

combining the Mie theory and the κ-Köhler theory, both $f(\mathrm{RH})$ and Vg(RH) can be simulated. In the processes of calculations for modelling $f(\mathrm{RH})$ and 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 on the basis of volume mixing rule. The used refractive index of liquid water is $1.33 - 10^{-7}i$ (Seinfeld and Pandis, 2006).

~~2.4~~3.3.2 Parameterization ~~scheme~~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. ~~Additionaly~~Additionally, 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:

$$f(\mathrm{RH}) = 1 + \kappa_{sca}\frac{RH}{100-RH}$$ (~~2~~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 ~~the following~~ $\mathrm{g(RH)} \cong (1+\kappa\frac{RH}{100-RH})^{1/3}$ (Brock et al., 2016)~~: $\mathrm{g(RH)} \cong (1+\kappa\frac{RH}{100-RH})^{1/3}$~~. Enhancement factor in volume can be expressed as the cube of g(RH). Of particular note is that aerosol particles larger than 100 nm contribute the most

Formatted: Justified

Formatted: English (U.S.)

Formatted: Justified

to σ_{sp} and $V_a(\text{dry})$, which means that if κ values of aerosol particles of different sizes are the same, then Vg(RH) can be approximately expressed as $\text{Vg(RH)} = 1 + \kappa\frac{RH}{100-RH}$. In addition, σ_{sp} is usually proportional to $V_a(\text{dry})$ which indicates that the relative change in σ_{sp} due to aerosol water uptake is roughly proportional to relative change in aerosol volume. Therefore, $f(\text{RH})$ might also be well described by using the formula form of equation (~~2~~5). Previous studies have shown that this parameterization scheme can describe $f(\text{RH})$ well (Brock et al., 2016;Kuang et al., ~~2017a~~2017b).

During processes of measuring $f(\text{RH})$, the sample RH in the "dry" nephelometer (RH_0) is not zero. According to equation (~~2~~5), the measured $f(\text{RH})_{measure} = \frac{f(\text{RH})}{f(RH_0)}$ should be fitted using the following formula:

$$f(\text{RH})_{measure} = (1 + \kappa_{sca}\frac{RH}{100-RH})/(1 + \kappa_{sca}\frac{RH_0}{100-RH_0}) \quad (\cancel{3}6)$$

Based on this equation, κ_{sca} can be calculated from measured $f(\text{RH})$ 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 $f(\text{RH})$ measurements during Wangdu campaign, and the results are shown in Fig.~~1~~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.

Formatted: Justified

In addition, based on discussions about the physical understanding of equation (~~2~~5), the Vg(RH) should be well described by the following equation:

$$\mathrm{Vg(RH)} = 1 + \kappa_{Vf}\frac{RH}{100-RH} \quad (47)$$

where κ_{Vf} is the parameter which fits Vg(RH) best.

~~3. Results and discussions~~

~~3.1 Estimation of V_a(dry) from measurements of the "dry" nephelometer~~

~~The first step of the proposed method is estimating V_a(dry) from measurements of the "dry" nephelometer. The investigation about the relationship between~~ V_a(dry) ~~and parameters measured by the "dry" nephelometer is required. Results of previous studies demonstrated that σ_{sp} of aerosol particles is roughly proportional to V_a(dry) (Pinnick et al., 1980). To confirm this conclusion, datasets of concurrently measured σ_{sp} (not corrected for angular truncation error) and PNSD of aerosol particles in dry state from D1 are used to investigate the relationships between measured σ_{sp} and V_a(dry). The measured V_a(dry) is integrated from simultaneously measured PNSD. To gain a first glimpse about the roughly proportional relationship between σ_{sp} and V_a(dry). All valid data points of measured σ_{sp} at 550 nm and~~ V_a(dry) ~~are plotted against each other and presented in Fig.2a. The results demonstrate that the σ_{sp} is highly correlated with V_a(dry), and the square of correlation coefficient between them is 0.92. The roughly proportional relationship exists between V_a~~(dry) ~~and $\sigma_{sp}(550\ nm)$. However, the ratio $\sigma_{sp}(550\ nm)/V_a$(dry) (hereinafter referred to as R_{Vsp}) varies significantly. The R_{Vsp} for points in Fig.2a range 1.54 to 6.9 $cm^3/(\mu m^3 \cdot Mm)$, and the average R_{Vsp} is 4.35 $cm^3/(\mu m^3 \cdot Mm)$. If this average R_{Vsp} is used for estimations of V_a(dry) based on measured $\sigma_{sp}(550\ nm)$, large bias may occur. Datasets of PNSD and σ_{sp} measured by the "dry" nephelometer during Wangdu campaign are used for investigating the performance of using the average R_{Vsp} in Fig.2a for~~

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: Font: Times New Roman, English (U.S.)

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

~~estimating V_a(dry), and the results are shown in Fig.2b. The x-axis represents measured V_a(dry) which is integrated from measured PNSD. The y-axis represents estimated~~ V_a(dry) ~~with an average R_{Vsp}. The results demonstrate that although a good correlation exists between estimated~~ V_a(dry) ~~and measured V_a(dry) (square of correlation coefficient between them is 0.83), large errors might occur, about 30% of data points have relative differences larger than 30%. More sophisticated method which can partially account for the variation of~~ R_{Vsp} ~~is needed for estimating V_a(dry) based on measurements of the "dry" nephelometer.~~

~~For developing a method which can partially consider the variation of R_{Vsp}, factors which determine the variation in R_{Vsp} should be aware of. Here, the quantitative relationship between V_a(dry) and σ_{sp} is analyzed. The σ_{sp} and V_a(dry) can be expressed as the following:~~

$$\sigma_{sp} = \int \pi r^2 Q_{sca}(m,r)\mathrm{n(r)dr} \quad \cancel{(5)}$$

$$V_a(\mathrm{dry}) = \int \frac{4}{3}\pi r^3 \mathrm{n(r)dr} \quad \cancel{(6)}$$

~~where $Q_{sca}(m,r)$ is scattering efficiency for a particle with refractive index m and particle radius r, n(r) is the aerosol size distribution. As presented in equation (5) and (6), relating V_a(dry) with σ_{sp} involves complex relation between $Q_{sca}(m,r)$ and particle diameter, and this relationship can be simulated using Mie theory. In consideration of 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~~

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

Formatted: Font: Times

Formatted: Justified

Formatted: Justified, Indent: First line: 2 ch

~~particle, BC particle, BC particle core-shell mixed with non-absorbing components and the radius of inner BC core are 25 nm and 100 nm, respectively. Same with those introduced in Sect.2.2, used refractive indices of BC and light non-absorbing components are 1.80 − 0.54i and 1.53 − 10^−7^i, respectively. The simulated results are shown in Fig.3a. Near the visible spectral range, most of ambient aerosol particles are almost non-absorbing, and their~~ $\cancel{Q_{sca}}$ ~~varies more like the blue line shown in Fig.3a. In the case of the blue line, aerosol particles with diameter less than about 800 nm, their~~ $\cancel{Q_{sca}}$ ~~increases almost monotonously with the particle diameter and can be approximately as a linear function to some extent. Fig.3b shows the simulated size-resolved accumulative contribution to scattering coefficient at 550 nm for all PNSDs measured during 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 total~~ $\cancel{\sigma_{sp}}$. ~~Therefore, for equation (5), If we express~~ $\cancel{Q_{sca}}(m,r)$ ~~as~~ $\cancel{Q_{sca}(m,r) = k \cdot r}$~~, then equation (5) can be expressed as the following:~~

$$\cancel{\sigma_{sp} = k \cdot \int \pi r^3 n(r) dr} \quad \cancel{(7)}$$

~~This explains why~~ $\cancel{\sigma_{sp}(550\ nm)}$ ~~is roughly proportional to~~ $\cancel{V_a}$~~(dry). However, the value k varies a lot for different particle diameters, which lead to the~~ $\cancel{R_{Vsp}}$ ~~affected greatly by the PNSD which determines weights of influences of aerosol particles with different diameters on~~ $\cancel{R_{Vsp}}$~~. The difference between the blue line and black line shown in Fig.3a indicates that fraction of externally mixed BC particles in all particles and their sizes will impact on~~ $\cancel{R_{Vsp}}$ ~~largely. 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.3a indicate that how BC mixed with and how much BC core-shell mixed with other components also exert significant influences on~~ $\cancel{R_{Vsp}}$~~. In summary,~~

~~the variation of R_{VSP} is mainly determined by variations in PNSD, mass size distribution and mixing state of BC. It is difficult to find a simple functional relationship between measured σ_{sp} and V_a(dry).~~

~~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. The Ångström exponent calculated from spectral dependence of σ_{sp} provide information on mean predominant aerosol size and is associated mostly with PNSD. However, the mass size distribution and mixing state of BC also impact on Ångström exponent. The variation of the hemispheric backscattering fraction (HBF) which is the ratio between σ_{bsp} and σ_{sp}, is essentially related with mass size distribution and mixing state of BC if the PNSD is fixed (Ma et al., 2012). If the PNSD and mass size distribution of BC are fixed, higher HBF at 550 nm means that BC particles are more internally (core-shell) mixed with other aerosol components (Ma et al., 2012). Hence, variations in both Ångström exponent and HBF are associated with PNSD, mass size distribution and mixing state of BC. As a result, the Ångström exponent and HBF together might constrain the variation of R_{VSP} better. In keeping with this philosophy, R_{VSP} shown in Fig.2a are spread into a two-dimensional gridded plot as shown in Fig.4a. Ångström exponent values are calculated based on concurrently measured σ_{sp} at 450 nm and 550 nm from TSI 3563 nephelometer. In Fig.4a, two regions are distinctly differed. In general, when HBF at 550 nm is larger than 0.14 and Ångström exponent is larger than 1, the R_{VSP} tends to be much smaller. This can be qualitatively understood. For the case of the blue line shown in Fig.3a, if particle diameter is smaller than about 750 nm, overall, the k value is larger if the particle diameter is larger. Smaller Ångström exponent and HBF at 550 nm together correspond to relatively larger~~

~~particle diameter and therefore larger R_{Vsp}. However, more details about the average variation pattern of R_{Vsp} with changes of HBF at 550 nm and Ångström exponent are difficult to be disentangled, due to that influences of PNSD, mass size distribution and mixing state of BC on R_{Vsp} are highly nonlinear. The percentile value of standard deviation of R_{Vsp} values within each grid of Fig.4a divided by their average is shown in Fig.4b. If HBF at 550 nm is less than 0.13, in most cases, percentile values shown in Fig.4b are less than 7%, which means that in this region R_{Vsp} varies little within each grid. However, if HBF at 550 nm is larger than 0.14, in most cases, percentile values shown in Fig.4b are near or even larger than 20%, which means that in this region even HBF and Ångström exponent are fixed, R_{Vsp} still varies a lot. These results imply that if using results shown in Fig.4a as a look up table for estimating R_{Vsp}, large bias may occur when HBF at 550 nm is larger than 0.14.~~

~~Datasets of σ_{sp} and σ_{bsp} measured by the "dry" nephelometer and concurrently measured PNSD during Wangdu campaign are used for verifying the performance of using results shown in Fig.4a as a look up for estimating R_{Vsp} and further estimating V_a(dry), and results are shown in Fig.5a. Compared with the results shown in Fig.2b, the look up table method has improved the estimation of V_a(dry) markedly (square of correlation coefficient between measured and estimated V_a(dry) increased from 0.83 to 0.9). It is noticeable that for points with HBF at 550 nm larger than about 0.13, V_a(dry) are systematically underestimated. This result is consistent with the previous analysis that if using results shown in Fig.4a as a look up table for estimating R_{Vsp}, large bias may occur when HBF at 550 nm is larger than 0.14.~~

Formatted: Font: Times New Roman

Formatted: Font: Times New Roman

~~Six parameters are measured by the "dry" nephelometer, however, only three parameters including σ_{sp} at 450 nm and 550 nm, and σ_{bsp} at 550 nm are used if using the look up table~~

~~shown in Fig.4a for estimating V_a(dry). It can be seen from the results shown in Fig.4b, when the HBF at 550 nm is larger than 0.14, variations in R_{Vsp} are poorly constrained. Based on the improvement achieved by using a look up table, we speculate that if all six parameters measured by the "dry" nephelometer are used together, then HBF at three wavelengths and Ångström exponent calculated both from σ_{sp} at 450 nm and 550 nm and σ_{sp} at 550 nm and 700 nm together can constrain variation in R_{Vsp} better. Machine learning methods which can handle many input parameters are capable of learning from historical datasets and then make predictions are powerful tools for tackling highly nonlinear problems. In the light of this, the idea came out that predicting V_a(dry) based on six optical parameters measured by the "dry" nephelometer might be accomplished by using a machine learning method. In this paper, we choose the machine learning function RidgeCV (ridge regression) from the linear model of module Scikit-learn of computer language Python (Pedregosa et al., 2011) for training the historical datasets of concurrently measured V_a(dry) and six raw parameters measured by the "dry" nephelometer from several field campaigns (Corresponding to data points shown in Fig.2a). Measurements during Wangdu campaign again are used for evaluating this machine learning method and the results are shown in Fig.5b. Compared with results shown in Fig.5a, the estimation of V_a(dry) is further improved, not only reflected in the increase of square of correlation coefficient, but also reflected in the change of the slope. And almost all points with HBF at 550 nm larger than 0.13 distributed within or near the 20% relative difference line. For the machine learning method, the square of correlation coefficient between measured and estimated V_a(dry) is 0.93, with 75% and 43% points have absolute relative differences less than 20% and 10%, respectively. And the standard deviations of absolute and relative differences between measured and estimated V_a(dry) are 8.4 $\mu m^3/cm^3$ and 10%, respectively.~~

~~Measured PNSDs and values of σ_{sp} at 550 nm during Wangdu campaign are shown in Fig.6a and Fig.6b, respectively. The results show that new particle formation phenomena are frequently observed during Wangdu campaign. In addition, both time series of estimated values of V_a(dry) using the machine learning method and time series of V_a(dry) which are integrated from measured PNSDs are shown in Fig.6c. The results demonstrate that overall, under different pollution levels and during periods with and without new particle formation phenomena, predicted V_a(dry) agrees well with measured V_a(dry). If a reasonable aerosol density is assumed, measurements from a three-wavelength nephelometer can also be used to estimate total mass concentrations of ambient aerosol particles in dry state.~~

~~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(dry) based on measurements of the "dry" nephelometer. The way of estimating V_a(dry) with machine learning method might be applicable for different regions around the world if used estimators are trained with corresponding regional historical datasets.~~

Formatted: Justified

~~3.2~~1.1.1 ~~Bridge the gap between f(RH) and Vg(RH)~~

Formatted: Justified

~~The approximate proportional relationship between σ_{sp} and V_a(dry) introduced in Sect.3.1 is only applicable for aerosol particles of constant refractive index, which is not the case for aerosol particles growing by addition of water with increasing RH (Hegg et al., 1993). As aerosol particles grow under conditions of increasing RH, the aerosol scattering efficiency change~~

Formatted: Justified, Indent: First line: 0 ch

~~nonlinearly and can even decrease. It is difficult to use the same method as introduced in Sect.3.1 to estimate the total aerosol volume of aerosol particles in ambient RH conditions. If Vg(RH) can be directly estimated from measured f(RH), then the ALWC can be estimated. Relating f(RH) to Vg(RH) involves complicated variations of aerosol scattering efficiency as a function of growing particle diameter due to aerosol water uptake, and this relationship can be described using Mie theory and κ-Köhler theory. As discussed in Sect.2.4, f(RH) and Vg(RH) can be described by the formula form of equation (2) and (4).~~ To consolidate 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 (~~2~~5) and (~~4~~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-Köhler theory is designed. In this experiment, all PNSDs at dry state along with mass concentrations of BC from ~~D2~~D1 are used, characteristics of these PNSDs can be found in

Formatted: Justified

Formatted: Justified

~~Kuang et al. (2017a).~~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). ~~Their results demonstrate that overall, size-resolved κ distributions have three modes: highly hygroscopic mode with diameters of aerosol particles ranging from 150 nm to 1 μm, less hygroscopic mode with diameters of aerosol particles less than 150 nm and nearly hydrophobic mode with diameters of aerosol particles larger than 1 μm. The~~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.~~9a~~7a) 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(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 ~~aforementioned~~ average size-resolved κ distribution (the black line shown in Fig.~~9a~~7a) 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 Ångström exponent contain information about PNSD ~~(Kuang et al., 2017a)~~(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., 2017a).~~(Kuang et al., 2017b). Simulated R_{Vf} values are spread into a two-dimensional gridded plot. The first dimension is Ångström 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.~~8a~~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.~~8a~~6a exhibit that both PNSD and overall aerosol ~~hygroscopicty~~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. With respect to influences of κ_{sca} on R_{Vf}, if Ångström exponent is larger than about 1.1, κ_{sca} have small influences on R_{Vf} while its influence is remarkable when Ångström exponent is lower than 1.1. In addition, the percentile value of standard deviation of R_{Vf} values within each grid divided by its average is shown in Fig.~~8b~~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.~~8a~~6a. This implies that results of Fig.~~8a~~6a can serve as a look up table to estimate R_{Vf} and thereby κ_{Vf} values can be directly predicted from measurements of a three-wavelength humidified nephelometer system.

For the look up table shown in Fig.~~8a~~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 ~~D2~~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.~~9a~~7a (the number is 23). Results shown in Fig.~~9a~~7a imply that the shape of size-resolved κ distribution has no apparent correlation with pollution degrees and varies a lot. 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 Ångström exponent are together used to estimate R_{Vf} values using the look up table shown in Fig.~~8a~~7a. Results of relative differences between estimated and modelled R_{Vf} values under different pollution conditions are shown in Fig.~~9b~~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.

~~3.3~~3.4 ~~Estimation~~Calculation of ~~the~~ ambient ALWC

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

According to the equation $\mathrm{Vg(RH)} = 1 + \kappa_{Vf}\frac{RH}{100-RH}$, ~~During the Wangdu campaign, there are ten days~~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(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(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~~ are available. Values of κ_{sca} are first fitted from observed f(RH) curves and then linearly interpolated to times~~, 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.

4. Results and discussions

4.1 Validation of the random forest model for predicting V_a(dry) based on measurements of the "dry" nephelometer

The machine learning method, random forest model, is proposed to predict V_a(dry) based only on σ_{sp} and σ_{bsp} ~~RH points (one f(RH) curve lasts about 45 minutes, the time resolution of~~ at three wavelengths measured by the "dry" nephelometer. Datasets of PNSD and BC from field campaigns F1 to F4 and F6 are used ~~ambient RH is five minutes), and the~~ 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(dry) of PM_{10} and $PM_{2.5}$ are shown in Fig.8. The square of correlation coefficient between predicted and calculated V_a(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(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(dry) of PM_{10} and $PM_{2.5}$ are 10% and 4% , respectively. These results indicate that V_a(dry) of $PM_{2.5}$ can be well predicted by using the machine learning method. While V_a(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

Formatted: Font: Times New Roman
Formatted: Font: Times New Roman
Formatted: Font: Times New Roman
Formatted: Font: Times New Roman
Formatted: Font: Times New Roman
Formatted: Justified

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 $V_a(\mathrm{dry})$ based on measurements of the “dry” nephelometer. The way of estimating $V_a(\mathrm{dry})$ 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 ~~20a. The RH range of one~~ 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://isorropia.eas.gatech.edu) 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). 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 $V_a(\mathrm{dry})$ 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.9a. 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 $\mathrm{Vg(RH)} = f(\mathrm{RH})^{1.5}$. Thus, the comparison results between ambient ALWC calculated based on ISORROPIA and ambient ALWC calculated by assuming $\mathrm{Vg(RH)} = f(\mathrm{RH})$~~ cycle is about 50% to 90%. The estimated values of κ_{vf} using results shown in Fig.20a as a look up table is~~ $^{1.5}$ are also shown in Fig.9b. 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. ~~20a.~~ When the ambient RH is higher than about 80%, calculated ambient ALWC will be significantly overestimated if assumes that $\mathrm{Vg(RH)} = f(\mathrm{RH})^{1.5}$. This method assumes that average scattering efficiency of aerosol particles at dry state and different RH conditions are the

Formatted: Font: Times, 12 pt

same. When ambient RH is high, the particle diameters changes a lot. As the results shown in Fig.S5, 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 ~~this observation period~~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. ~~Time series of ambient RH is shown in Fig.20b, and RH points with RH larger than 95% are excluded because the measurements of ambient RH at this range is highly uncertain. With estimated values of~~ $\cancel{\kappa_{Vf}}$ ~~and measured ambient RH, Vg(RH) of aerosol particles in ambient RH states can be estimated. Then, with measured~~ $\cancel{V_a}$~~(dry) (shown in Fig.20c) which is integrated from measured PNSD, water volumes of ambient aerosol particles are estimated and shown in Fig.20c. During this observation period, estimated water volume of ambient aerosol particles mainly range from 1 to 300~~ $\cancel{\mu m^3/cm^3}$~~, with an average of 42~~ $\cancel{\mu m^3/cm^3}$~~.~~The calculated volume fractions of water in total volume of ambient aerosols during Wangdu campaign are shown in Fig.10a. 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%.

~~3.4 Uncertainty analysis~~

Formatted: Justified

During Gucheng campaign, κ_{sca} ranges from 0.008 to 0.22with 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. During Gucheng campaign, the maximum volume fraction of water in ambient aerosol is 42% when ambient RH is at 80%. Averagely speaking, 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

~~According to the equation Vg(RH) = 1 +~~ $\cancel{\kappa_{Vf}\frac{RH}{100-RH}}$, ~~the estimated volume of aerosol liquid water (~~$\cancel{V_{water}}$~~) can be expressed as:~~ $\cancel{V_{water}} = \cancel{V_a}(\text{dry}) \cdot \kappa_{sca} \cdot R_{Vf} \cdot \frac{RH}{100-RH}$. ~~Neglecting measurement uncertainty of ambient RH, uncertainties contribute to~~ $\cancel{V_{water}}$ ~~include uncertainty of~~ $\cancel{V_a}$~~(dry), uncertainties of~~ $\cancel{\kappa_{sca}}$ ~~and~~ $\cancel{R_{Vf}}$.

~~Results introduced in Sect.3.1 suggest that using the machine learning method to predict~~ $\cancel{V_a}$~~(dry) from measurements of a three-wavelength nephelometer is feasible but non-negligible bias still exists between measured and estimated~~ $\cancel{V_a}$~~(dry). The standard deviation of relative differences between measured and estimated~~ $\cancel{V_a}$~~(dry) is 15%. If using triple the standard deviation (99% of points locate within this range) as the uncertainty of this method, the uncertainty is 45%. Here, sources of this large bias is discussed. The~~ $\cancel{V_a}$~~(dry) is determined from PNSD which is high-dimensional. Six parameters provided by the "dry" nephelometer cannot accurately constrain~~ $\cancel{R_{Vsp}}$ ~~in the machine learning method. This should be the largest uncertainty source. In addition, used datasets for training the estimator carried some uncertainties which should also influence the performance of the estimator. Using a Monte Carlo method based on~~

~~uncertainties of measured PNSD (see Table 3 of Ma et al. (2014) for the uncertainty parameters of PNSD), V_a(dry) integrated from measured PNSD have uncertainty of about 5%. The TSI 3563 nephelometer also carry some uncertainties, it is about 9% (Heintzenberg et al., 2006;Sherman et al., 2015). Their uncertainties will propagate in the processes of training and verifying the estimator. If the datasets for training the estimator are more comprehensive (like a year of observation in several sites), the uncertainty of this machine learning method might be smaller.~~

~~The κ_{sca} is directly fitted from f(RH) measurements.~~ ~~Results of Titos et al. (2016) demonstrate that, for moderately hygroscopic aerosols (e.g., $f(\mathrm{RH}=80\%)$ less than 2.2), f(RH) errors are around 15%. Since most values of $f(\mathrm{RH}=80\%)$ observed on continental regions are less than 2.2 (Zhang et al., 2015;Titos et al., 2016), 15% is used as the uncertainty of f(RH) as well as κ_{sca}.~~

~~As to uncertainty of estimated R_{Vf}. Many factors exert influences on R_{Vf}, such as PNSD, mixing state of BC and size-resolved κ distribution. If using the 99% line (triple the standard deviation) shown in Fig.9b as uncertainties of R_{Vf} from influences of size-resolved κ distribution and PNSD, then this aspect of uncertainties of R_{Vf} under different pollution conditions range from 17% to 49%. Additionally, the mixing state of BC can also impact on R_{Vf}. In this study, the BC is assumed to be half externally and half coreshell mixed with other aerosol components. A simple simulative test is performed to investigate the influence of BC mixing state on R_{Vf}. In this test, we simulated R_{Vf} values for three kinds of BC mixing states: external; half external and half coreshell; core-shell, and the average PNSD and mass concentration of BC during Wangdu campaign are used. Simulated R_{Vf} values for these three mixing state are 1.13, 1.18 and 1.25,~~

~~respectively. Thus, we consider 6% as the uncertainty of R_{VF} caused by the variation of BC mixing state. The synthesized uncertainties of estimated R_{VF} under different pollution levels are presented in Fig.21, which have considered the variations of BC mixing state and size-resolved κ distribution and PNSD. Uncertainties of estimated R_{VF} by using the look up table shown in Fig.8a range from 18% to 49.4%.~~

~~With estimated uncertainties of V_a(dry), κ_{sca} and R_{VF}, the uncertainties of estimated V_{water} under different pollution levels can be estimated. In the processes of estimating V_{water}, two methods can be used to estimate V_a(dry). The first mehtod is estimating V_a(dry) from measurements of the three-wavelength "dry" nephelometer (Method 1). However, if PNSD is available, V_a(dry) can be directly integrated from measured PNSD (Method 2). The calculated uncertainties of V_{water} under different pollution levels with V_a(dry) estimated from these two methods are presented in Fig.21. For Method 1, uncertainties of estimated V_{water} range from 24% to 52%, with an average of 31%. For Method 2, uncertainties of estimated V_{water} range from 51% to 68%%, with an average of 55%. Compared to clean conditions, the uncertainty of estimated V_{water} is smaller when the air is highly polluted. We recommend that if measured PNSD is available, V_a(dry) should be calculated from measured PNSD, otherwise, V_a(dry) can be estimated from measurements of the "dry" nephelometer.~~

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 and sea salt. Cautions should be exercised if using the proposed method to estimate the ALWC when the air mass is influenced by sea salt or dust. The way of estimating V_a(dry) with machine learning method might be applicable for different regions around the world. However, the used ~~estimator~~predictor

Formatted: Justified

from machine learning should be trained with corresponding regional historical datasets and PNSD and BC. The way of connecting f(RH) to Vg(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 ~~limits~~limit the usage of the proposed method when ambient RH is extremely high. As discussed in Sect.~~2.4~~3.3, the proposed way of connecting f(RH) and Vg(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. (2017b)~~Kuang et al. (2017a) indicated that κ values retrieved from f(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. ~~However, if~~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%.

~~4. conclusions~~

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 $\mathrm{Vg(RH)}$ and $f(\mathrm{RH})$. The ALWC can be calculated from the estimated $V_a(\mathrm{dry})$ and $\mathrm{Vg(RH)}$.

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, ~~an estimator~~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 ~~estimator~~random forest predictor is trained ~~with~~based on historical datasets of PNSD and ~~σ_{sp}~~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 $V_a(\mathrm{dry})$ ~~is 0.93~~of PM_{10} and $PM_{2.5}$ are 0.96 and 0.997, respectively.

The relationship between Vg(RH) and f(RH) is ~~then~~ investigated in Sect.3 by conducting a simulative experiment. It is found that the complicated relationship between Vg(RH) and f(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(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. ~~During Wangdu campaign, calculated water volumes of ambient aerosol particles range from 1 to 300~~ $\cancel{\mu m^3/cm^3}$~~, with an average of 42~~ $\cancel{\mu m^3/cm^3}$ ~~.~~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(RH) and 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.

Formatted: Justified

Acknowledgments

This work is supported by the National Natural Science Foundation of China (41590872~~, 41375134). The data used are listed in the references and a repository at http://pan.baidu.com/s/1c2Nzc5a.~~ and 41505107) and the National Key R&D Program of China (2016YFC020000: Task 5).

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.

~~Hegg, D., Larson, T., and Yuen, P.-F.: A theoretical study of the effect of relative humidity on light scattering by tropospheric aerosols, Journal of Geophysical Research: Atmospheres, 98, 18435-18439, 10.1029/93JD01928, 1993.~~

~~Heintzenberg, J., Wiedensohler, A., Tuch, T. M., Covert, D. S., Sheridan, P., Ogren, J. A., Gras, J., Nessler, R., Kleefeld, C., Kalivitis, N., Aaltonen, V., Wilhelm, R. T., and Havlicek, M.: Intercomparisons and Aerosol Calibrations of 12 Commercial Integrating Nephelometers of Three Manufacturers, Journal of Atmospheric and Oceanic Technology, 23, 902-914, 10.1175/JTECH1892.1, 2006.~~

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, ~~2017a~~2017b.

~~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, 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.

Formatted: Hyperlink

Phys., 12, 2381-2397, 10.5194/acp-12-2381-2012, 2012.
~~Ma, N., Birmili, W., Müller, T., Tuch, T., Cheng, Y. F., Xu, W. Y., Zhao, C. S., and Wiedensohler, A.: Tropospheric aerosol scattering and absorption over central Europe: a closure study for the dry particle state, Atmos. Chem. Phys., 14, 6241-6259, 10.5194/acp-14-6241-2014, 2014.~~
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.
~~Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830, 2011.~~
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.

Formatted: Hyperlink

Phys., 14, 12055-12067, 10.5194/acp-14-12055-2014, 2014.

~~Titos, G., Cazorla, A., Zieger, P., Andrews, E., Lyamani, H., Granados-Muñoz, M. J., Olmo, F. J., and Alados-Arboledas, L.: Effect of hygroscopic growth on the aerosol light-scattering coefficient: A review of measurements, techniques and error sources, Atmospheric Environment, 141, 494-507, http://dx.doi.org/10.1016/j.atmosenv.2016.07.021, 2016.~~

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.

~~Wiedensohler, A., Birmili, W., Nowak, A., Sonntag, A., Weinhold, K., Merkel, M., Wehner, B., Tuch, T., Pfeifer, S., Fiebig, M., Fjäraa, A. M., Asmi, E., Sellegri, K., Depuy, R., Venzac, H., Villani, P., Laj, P., Aalto, P., Ogren, J. A., Swietlicki, E., Williams, P., Roldin, P., Quincey, P., Hüglin, C., Fierz-Schmidhauser, R., Gysel, M., Weingartner, E., Riccobono, F., Santos, S., Grüning, C., Faloon, K., Beddows, D., Harrison, R., Monahan, C., Jennings, S. G., O'Dowd, C. D., Marinoni, A., Horn, H. G., Keck, L., Jiang, J., Scheckman, J., McMurry, P. H., Deng, Z., Zhao, C. S., Moerman, M., Henzing, B., de Leeuw, G., Löschau, G., and Bastian, S.: Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions, Atmos. Meas. Tech., 5, 657-685, 10.5194/amt-5-657-2012, 2012.~~

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.

~~Zhang, L., Sun, J. Y., Shen, X. J., Zhang, Y. M., Che, H., Ma, Q. L., Zhang, Y. W., Zhang, X. Y., and Ogren, J. A.: Observations of relative humidity effects on aerosol light scattering in the Yangtze River Delta of China, Atmos. Chem. Phys., 15, 8439-8454, 10.5194/acp-15-8439-2015, 2015.~~

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

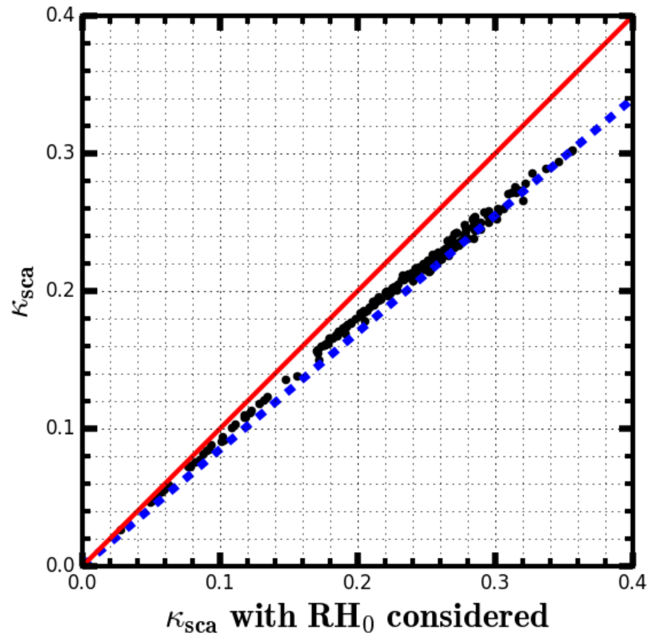
RH	relative humidity
$PM_{2.5}$	particulate matter with aerodynamic diameter of less than 2.5 µm
PM_{10}	particulate matter with aerodynamic diameter of less than 10 µm
f(RH)	aerosol light scattering enhancement factor at 550 nm
ALWC	aerosol liquid water content: volume concentrations of water in ambient aerosols
V_a(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}	$\sigma_{sp}(550\ nm)/V_a$(dry)
F1 to ~~F5~~F6	referred as to five field campaigns listed in Table ~~S1~~2
	~~PNSD, BC and nephelometer measurements from field campaigns F1 to F4~~
D1	PNSD, BC and nephelometer measurements from F2, F4 and F5
~~D2~~	

Formatted: English (U.S.)

Table 2. Locations, time periods and used datasets of ~~five~~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(RH)					Humidified nephelometer system	Humidified nephelometer system
Water soluble Ions						IGAC
Campaign Name	F1	F2	F3	F4	F5	F6


0.0
0.1
0.2
0.3
0.4
κ_{sca}
κ_{sca} with RH_0 considered

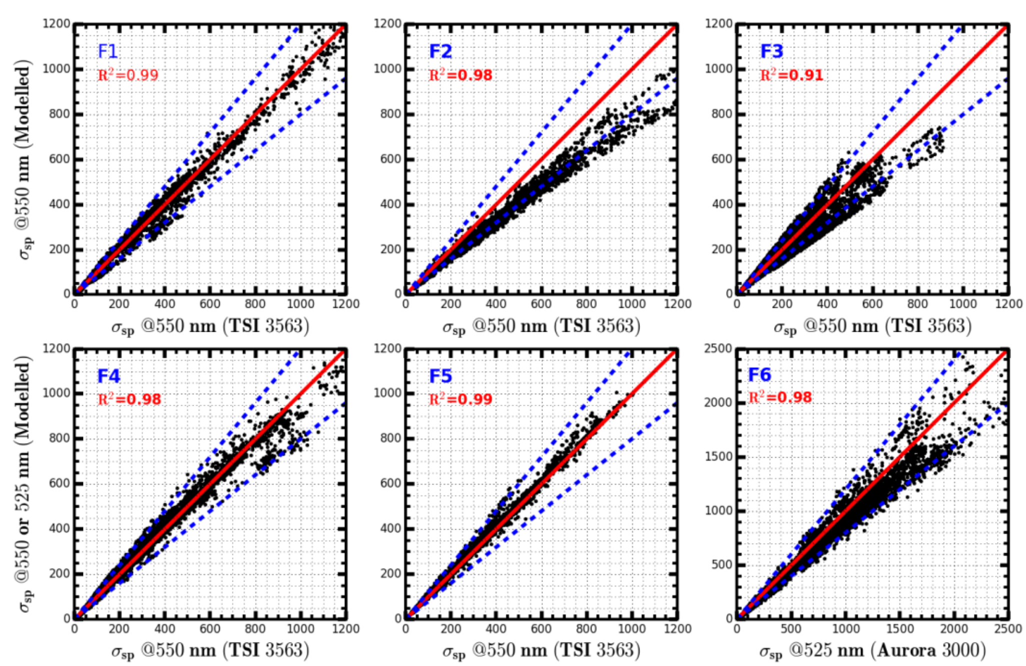


Figure 1. ~~X axis~~ Comparisons between measured and ~~y axis represent κ_{sca} are fitted with and without consideration of RH_0 in the "dry" nephelometer,~~calculated σ_{sp}, solid ~~respectively. The~~ red ~~line is~~lines are 1:1 ~~line, the~~ references lines. Dashed blue ~~dashed line is the 15~~lines are 20% relative difference ~~line.~~

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

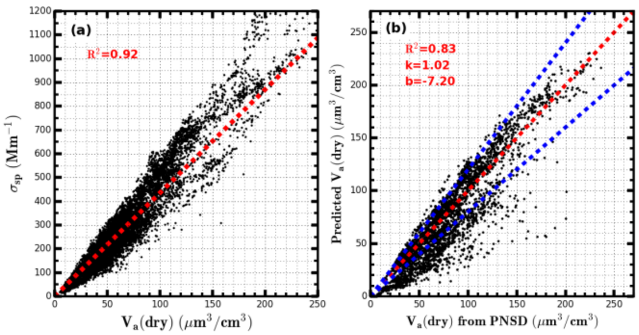
Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt



~~**Figure 2**. (a) Scatter plot of all valid data points of~~ ~~V_a(dry) and σ_{sp} at 550 nm from D1, dashed red line is the line whose slope is equal to the average ratio $\sigma_{sp}(550\ nm)/V_a$(dry)(R_{Vsp}). (b) The comparison between V_a(dry) estimated from a fixed average R_{Vsp} of (a) and measured V_a(dry). In figure (a), the dashed red line is the line whose slope is the average ratio R_{Vsp}. In figure (b), red line is the 1:1 line, two dashed blue line are lines with relative difference of 20%. R^2 is the suqare~~lines. R^2 is square of correlation coefficient~~, k is the slope, b is the intercept~~ between measured and modelled σ_{sp}. Blue texts at the upper left corners are corresponding field campaigns as listed in Table2.

Formatted: Caption, Line spacing: single

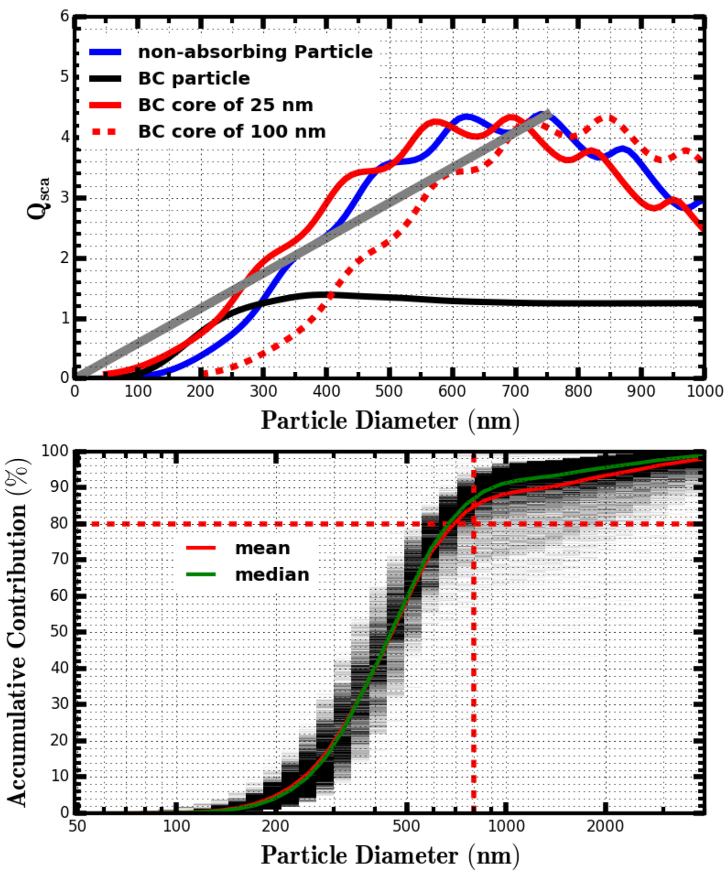
Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt, English (U.K.)

Formatted: English (U.K.)

Formatted: TA_Main_Text, Space After: 12 pt


non-absorbing Particle
BC particle
BC core of 25 nm
BC core of 100 nm
Q_{sca}
Particle Diameter (nm)
mean
median
Accumulative Contribution (%)
Particle Diameter (nm)

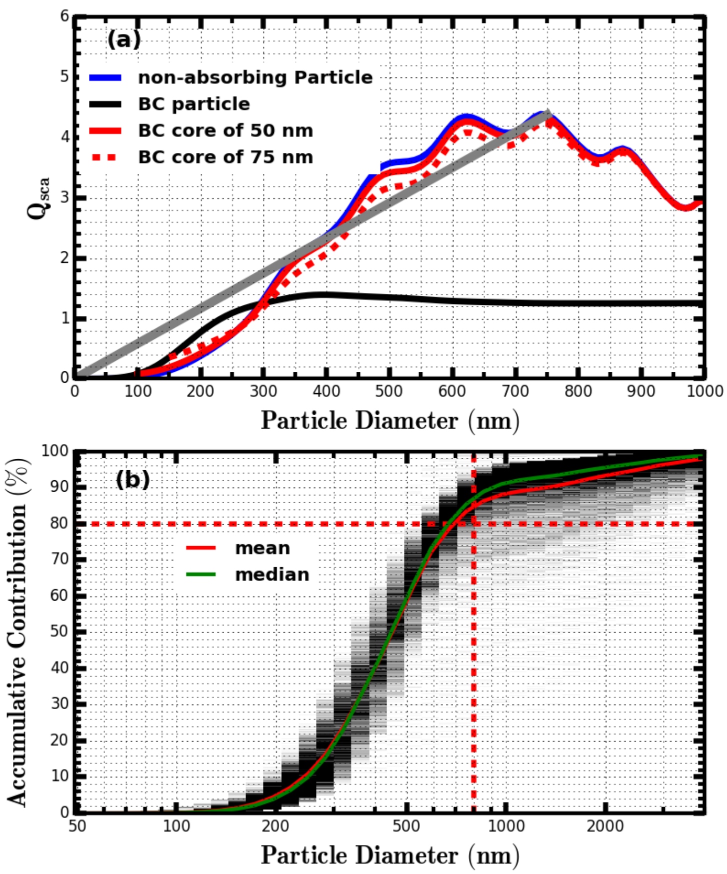
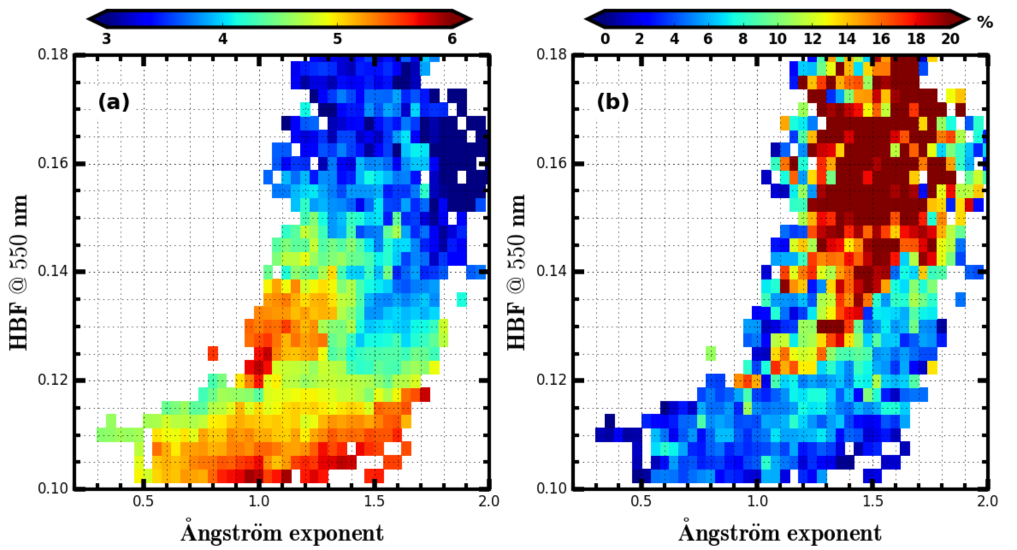


Figure ~~3~~2. (a) Q_{sca} at 550 nm as a function of particle diameter for four types of aerosol particles~~, the~~: almost non-absorbing aerosol particle, BC particle, BC particle core-shell mixed with non-absorbing components and the radius of inner BC core are 50 nm and 70 nm. The gray line corresponds 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 ~~scattering coefficient~~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.



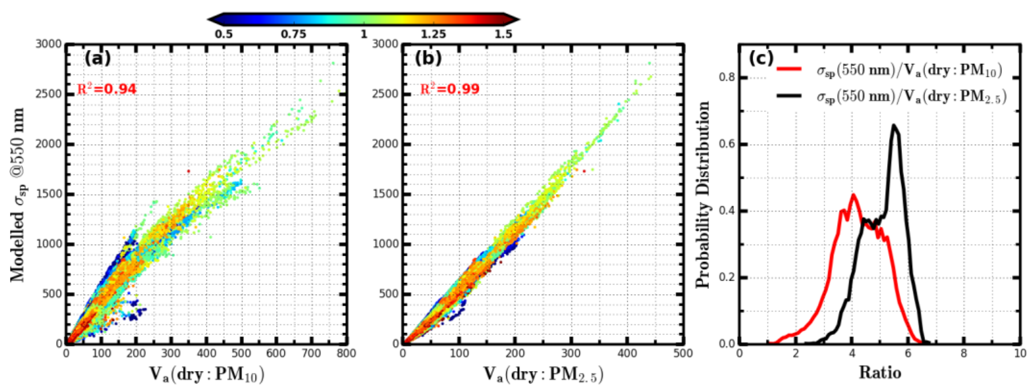


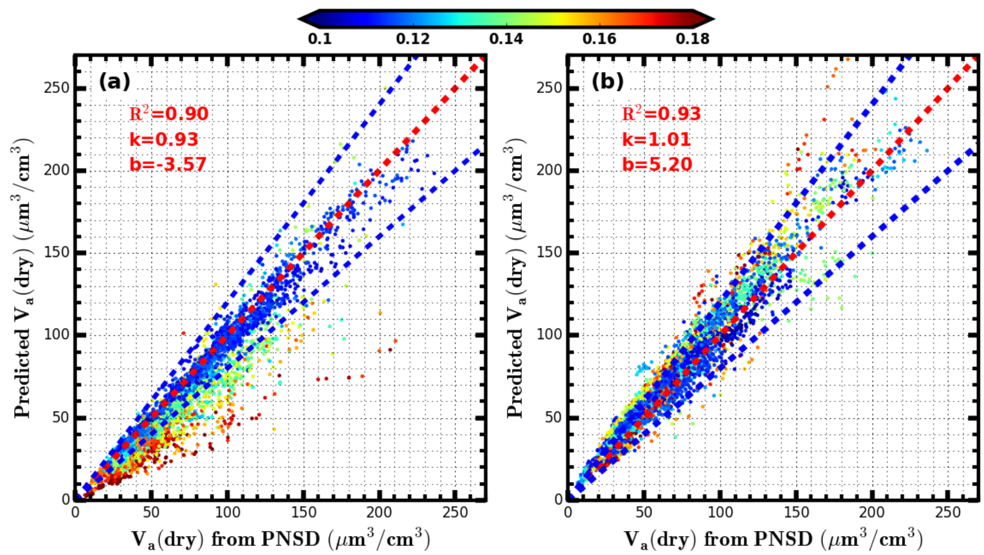
Figure ~~4~~3. (a) ~~Colors represent R_{vsp} values~~ and ~~the color bar is shown on the top of this figure, x-axis represents Ångström exponent and y-axis represents HBF at 550 nm.~~ (b~~) Meanings of x-axis and y-axis are same with them in (a), however, color represents the percentile value of the standard deviation of R_{vsp} values within each grid divided by their average.~~

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt

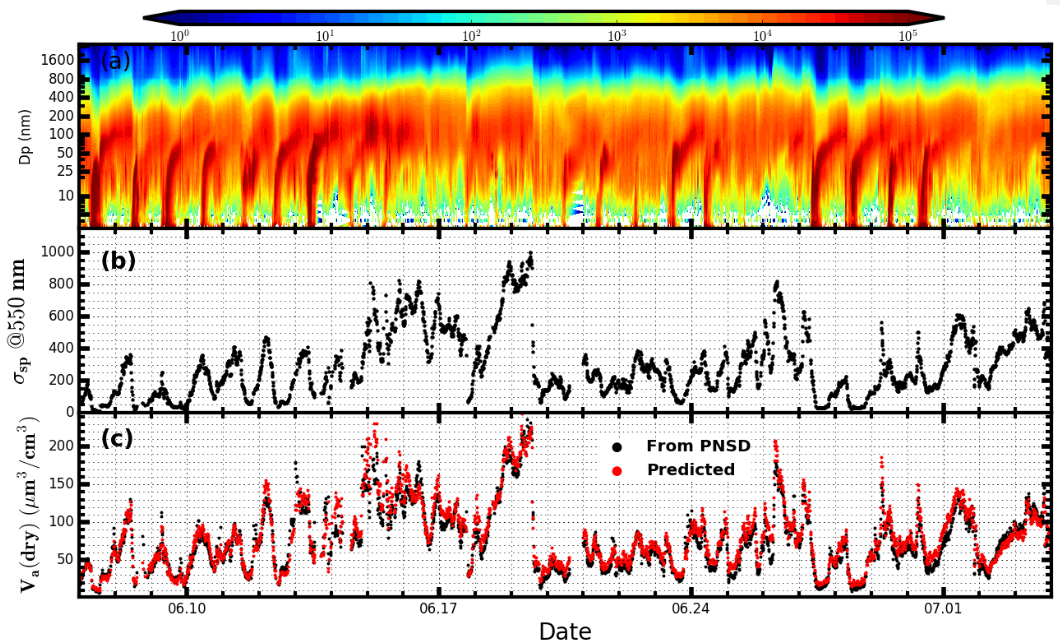


~~Figure 5~~~~. Comparisons between predicted and measured~~): Modelled σ_{sp} at 550 nm based on PNSD and BC versus V_a(dry)~~, the red dashed line is the 1:1 line, two blue dashed line shown in (a) and (b) are lines with relative difference of 20%.~~ of PM_{10} or $PM_{2.5}$ calculated from measured 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 ~~HBF~~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 the color bar is shown on the top. (a) V_a(dry) in the y-axis is predicted by using results shown in Fig.4a as a look up table. (b)~~ and V_a(dry)~~ in the y-axis is predicted by using the machine learning method.~~

Formatted: Font: +Headings (Cambria)

Formatted: Font: +Headings (Cambria)

Formatted: Font: +Headings (Cambria)



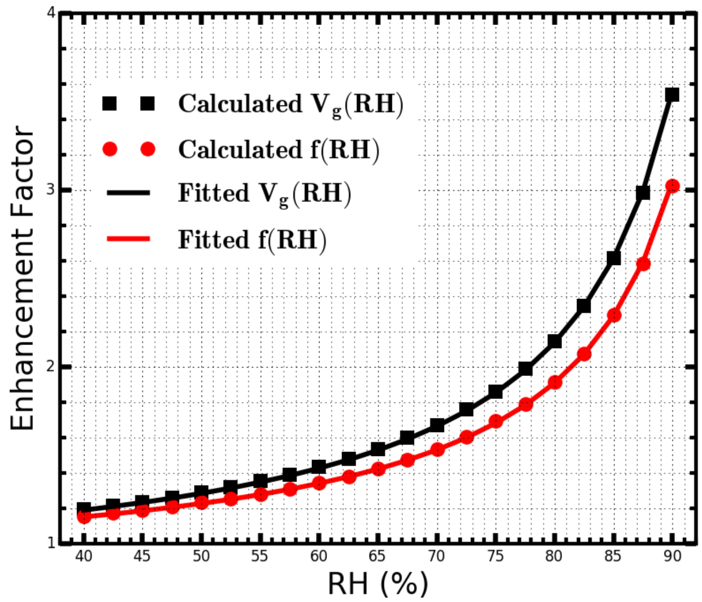
~~**Figure 6**. Measurements~~ of ~~PNSD and σ_{sp} during Wangdu campaign; (a) Time series of PNSD in dry state, colors represent $dN/dlog(Dp)(cm^{-3})$; (b) Time series of measured σ_{sp} at 550 nm; (c) V_a(dry) integrated from measured PNSD and V_a(dry) predicted by using the machine learning method~~ PM_{10} or $PM_{2.5}$.

Formatted: Line spacing: single

Formatted: Font: +Headings (Cambria)

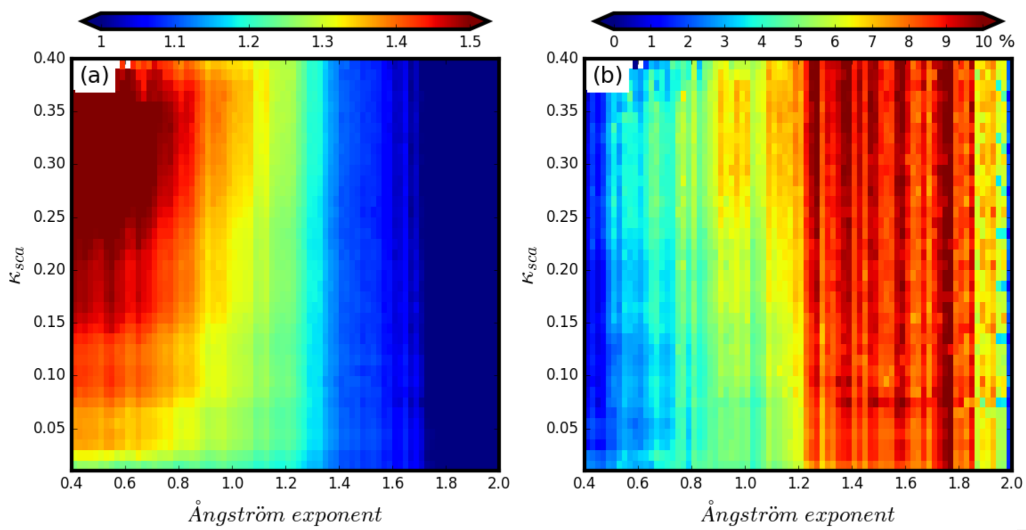
Formatted: Font: +Headings (Cambria)

Formatted: Font: +Headings (Cambria), 10 pt, English (U.K.)

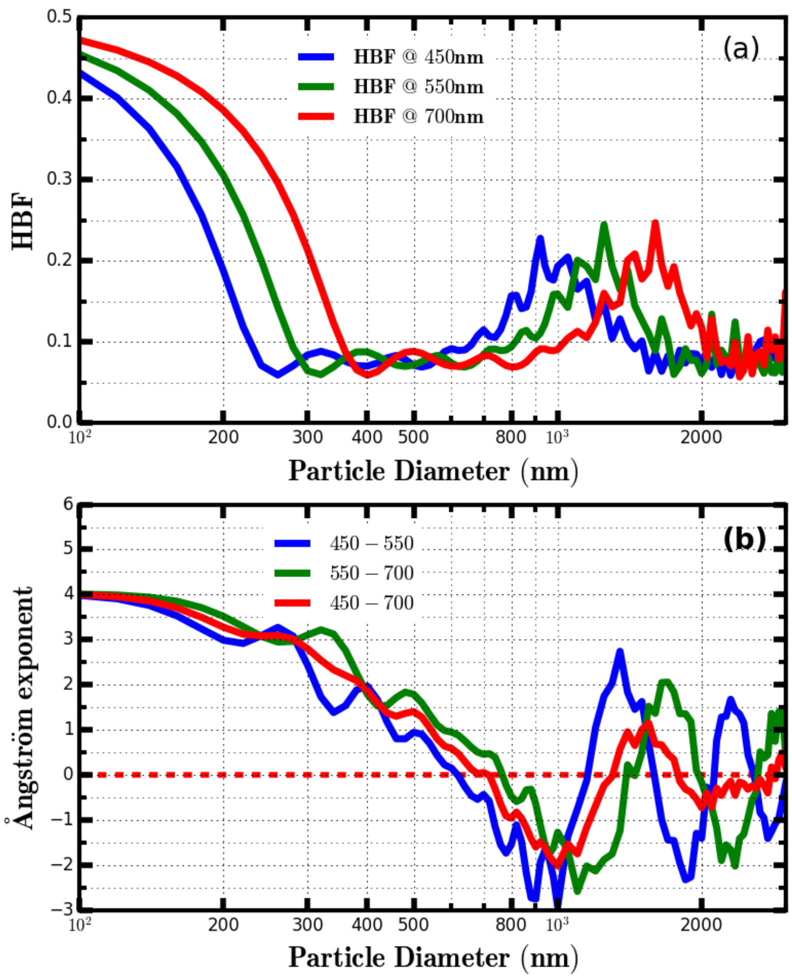


~~**Figure 7**. Modelled *f*(RH) and Vg(RH) (scatter points) and fitted *f*(RH) and Vg(RH) (solid lines) using formula form of equation (2).~~

1071



(a)
(b)
1
1.1
1.2
1.3
1.4
1.5
0 1 2 3 4 5 6 7 8 9 10 %
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
κ_{sca}
0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0
Ångström exponent



~~**Figure 8**~~

Figure 4. (a) Simulated HBF at three wavelengths as a function particle diameter. (b) Simulated Ångström exponent values as a function a particle diameter.

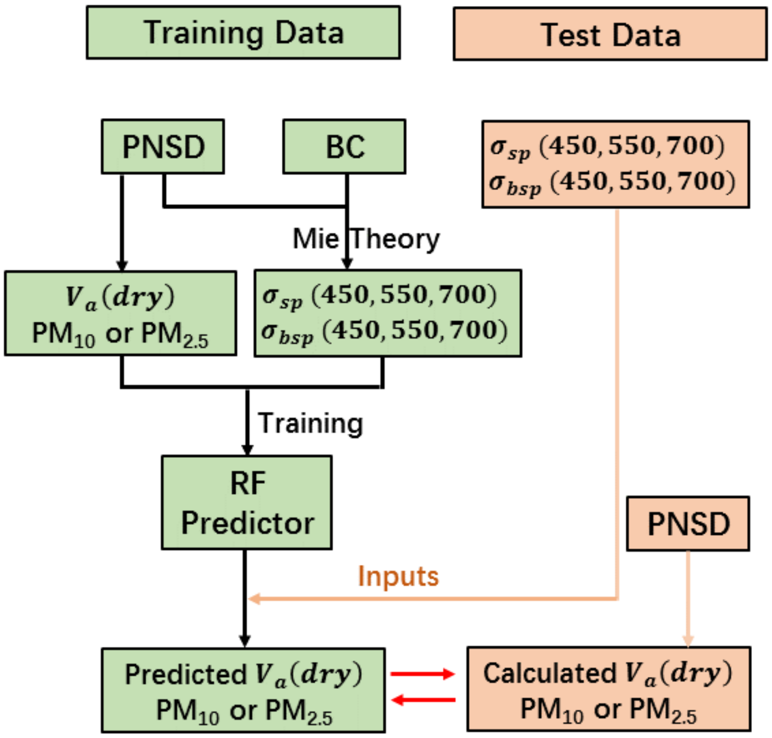


Figure 5. Schematic diagram of training the random forest (RF) model and verifying the performance of trained RF predictor. The trained datasets of PNSD and BC are from field campaigns F1 to F4 and F6, the test datasets of PNSD and optical parameters are from campaign F5. σ_{bsp} is the back scattering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

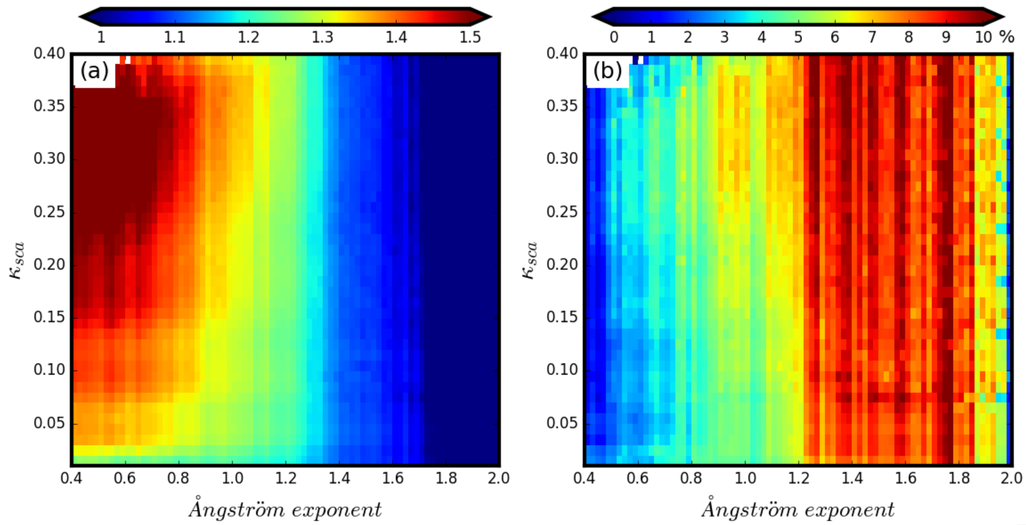


Figure 6. (a) Colors represent R_{Vf} values and the colorbar is shown on the top of this figure, x-axis represents Ångström exponent and y-axis represents κ_{sca}. (b) Meanings of x-axis and y-axis are same with them in (a), however, color represents the percentile value of the standard deviation of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

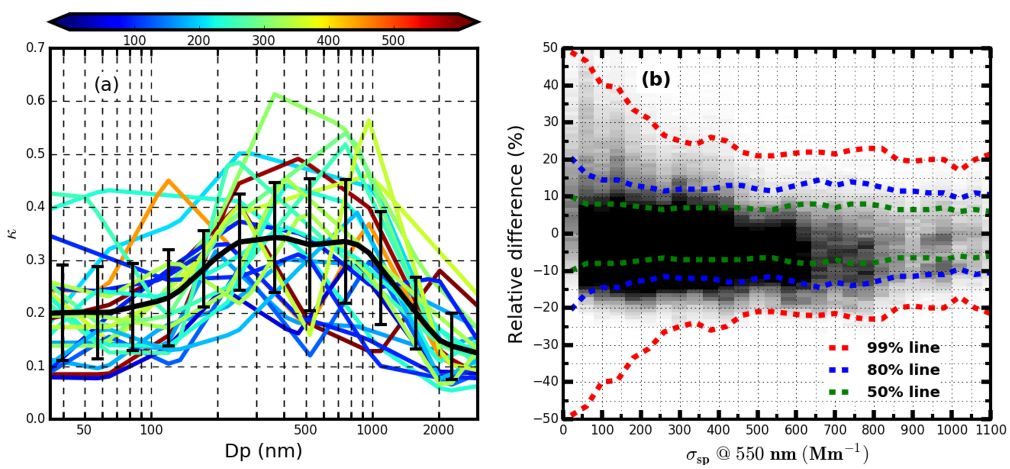
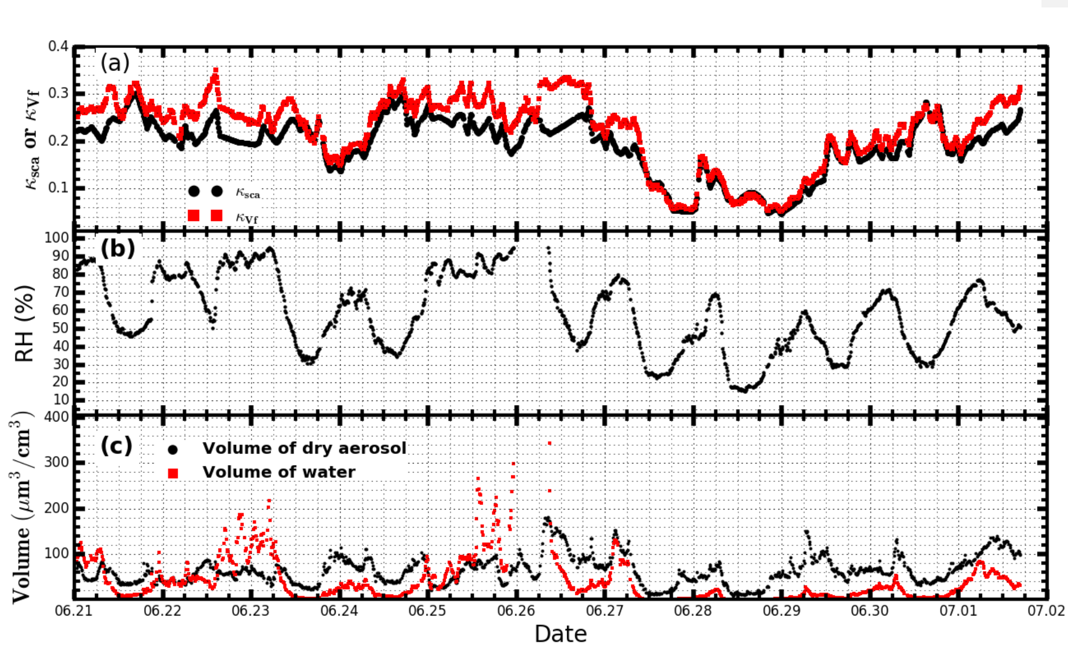
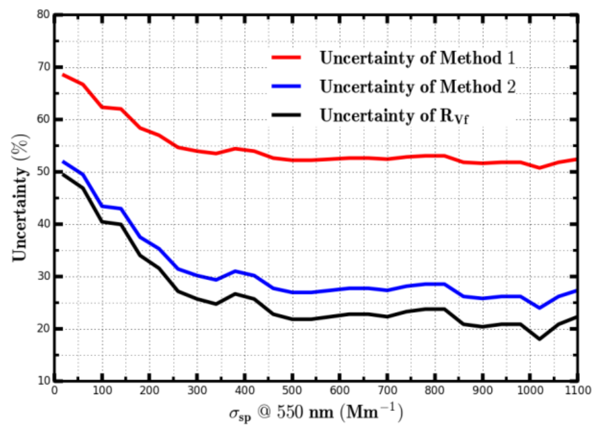


Figure 97. (a) All size-resolved κ distributions which are derived from measured size-segregated chemical compositions during HaChi campaign, colors represent corresponding values of average σ_{sp} at 550 nm (Mm^{-1}), black solid line is the average size-resolved κ distribution and error bars are standard 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 10.** (a) Time series of values of κ_{sca} fitted from observed f(RH) curves and predicted values of κ_{Vf} by using results shown in Fig.8a as a look up table; (b) Measured amibient RH; (c) Time series of V_a(dry) ($\mu m^3/cm^3$) which is integrated from measured PNSD and volume of aerosol liquid water estimated from combination of κ_{Vf} and ambient RH.~~



~~**Figure 11**. Black line corresponds to uncertainty of predicted R_{Vf} by using results shown in Fig.5a as the look up table. Blue and red lines represent uncertainties of volumes of aerosol liquid water which are estimated from the following two methods: Method 1 corresponds to V_a(dry) is estimated from the machine learning method, Method 2 corresponds to V_a(dry) is integrated from the concurrently measured PNSD.~~

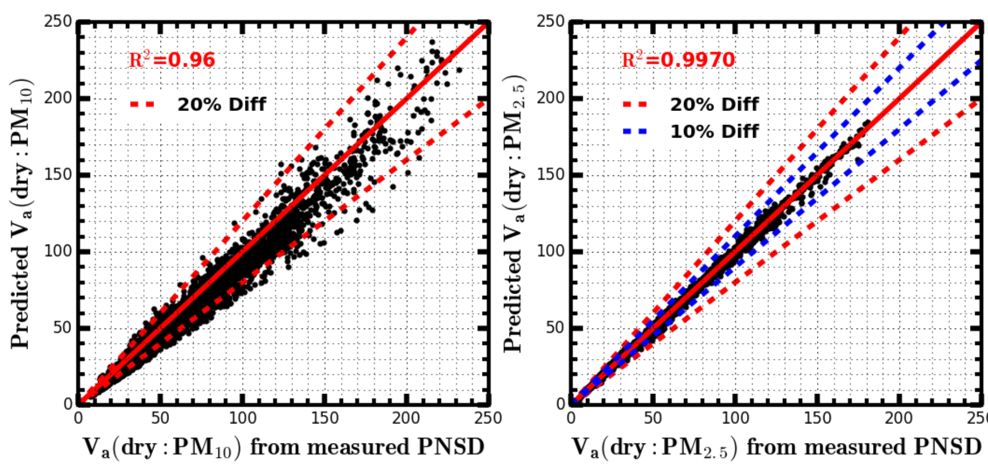


Figure 8. The comparison between V_a(dry) of PM_{10} or $PM_{2.5}$ calculated from measured PNSD and V_a(dry) of PM_{10} or $PM_{2.5}$ which are predicted based on six optical parameters measured by the "dry" nephelometer by using the random forest model. The unit of V_a(dry) is $\mu m^3/cm^3$. R^2 is 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.

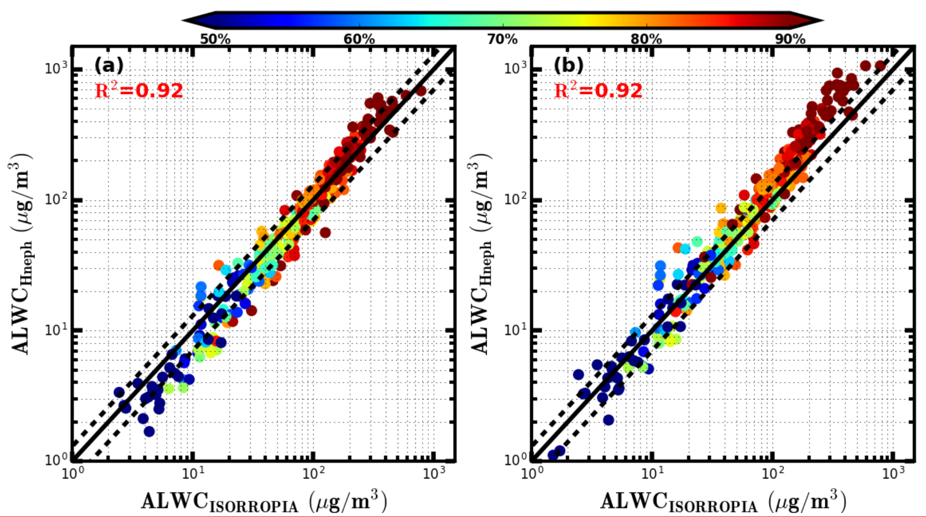


Figure 9. The comparison between ALWC calculated from ISORROPIA thermodynamic model ($ALWC_{ISORROPIA}$) and ALWC calculated from measurements of the humidified nephelometer system ($ALWC_{Hneph}$). The black solid line is the 1:1 line, the two dashed black lines are 30% relative difference lines. R^2 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).

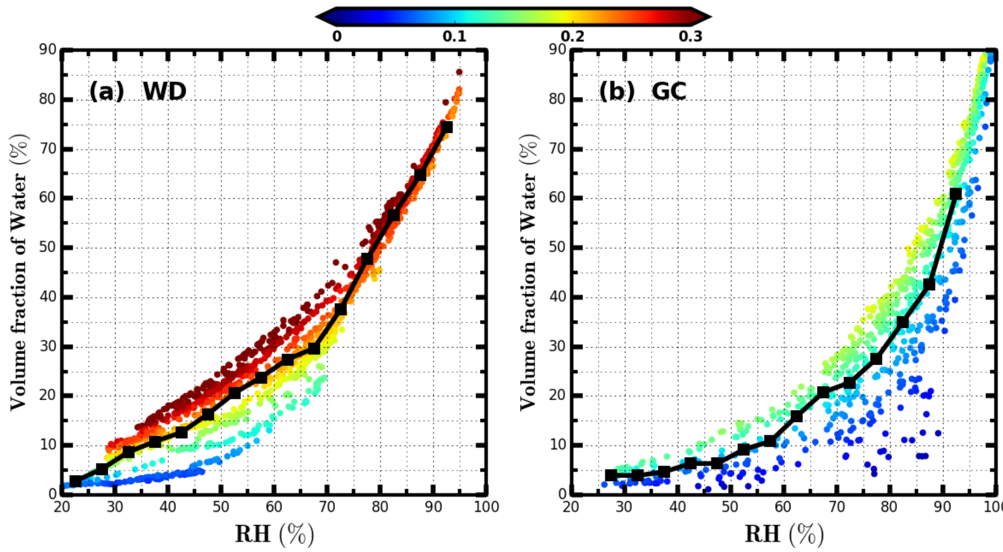


Figure 10. Volume fractions of water in total volume of ambient aerosols during Wangdu (WD) and Gucheng (GC) campaigns. X-axis represents measured ambient RH. Y-axis represents volume fractions of water. Colors of scatter points represent corresponding κ_{Vf}. Black solid lines in (a) and (b) show the average volume fractions of water under different ambient RH conditions.