Method to calculate the aerosol asymmetry factor based on measurements from the humidified nephelometer system

Gang Zhao[1], Chunsheng Zhao[1], Ye Kuang[2], Yuxuan Bian[3], Jiangchuan Tao[2], Chuanyang Shen[1], Yingli Yu[1]

[1]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, Beijing, 100081, China

Corresponding author: Chunsheng Zhao (zcs@pku.edu.cn)

Abstract

The aerosol asymmetry factor (g) is one of the most important factors for assessing direct aerosol radiative forcing. ~~So far, few studies have focused on~~However, little attention has been paid to the measurements and parameterization of g. In this study, t~~T~~he characteristics of g are studied based on field measurements over the North China Plain by using the Mie scattering theory. The results show that calculated g values for the dry aerosol can vary over a wide range (between 0.54 and 0.67). When ambient relative humidity (RH) reaches 90%, g is significantly enhanced by a factor of 1.2 due to aerosol hygroscopic growth ~~of the continental aerosol~~. For the first time, a novel method to calculate g based on measurements from the humidified nephelometer system is proposed. This method can constrain the uncertainty of g within 2.56% for dry aerosol populations and 4.02% for ambient aerosols, where the ~~taking into account~~ aerosol hygroscopic growth has been taken into account. Sensitivity studies show that aerosol hygroscopicity ~~is the most important factor that influences the accuracy of predicting~~ g plays a vital role in the accuracy of predicting g.

1 Introduction

In addition to aerosol optical depth and aerosol single-scattering albedo, the aerosol phase function is the most important factor for assessing direct aerosol radiative forcing (DARF) (Andrews et al., 2006;Russell et al., 1997). The Henyey-Greenstein phase function (PF_{HG}) is a widely used method to parameterize the phase function (Toublanc, 1996;Boucher, 1998;Pandey and Chakrabarty, 2016) because it uses the aerosol asymmetry factor (g) as the only free parameter. The PF_{HG} is expressed as

$$PF_{HG}(\theta) = \frac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}}, \quad (1)$$

where θ is the angle between the incident light direction and the scattered light direction. In this respect, the free parameter g can reflect the angular aerosol scattering energy distribution.

g is defined as:

$$g = \frac{1}{2}\int_0^{\pi} \cos\theta P(\theta) \sin(\theta)\, d\theta, \quad (2)$$

where P(θ) is the normalized scattering phase function. As a result, g can be a computationally efficient ~~factor that replaces~~parameter to replace the phase function ~~to~~ in the study of aerosol radiative transfer properties (Toublanc, 1996;Hansen, 1969;Boucher, 1998). This replacement proves to be useful and has been widely accepted in previous researches ~~Some researchers have widely accepted the use of g as a replacement of the phase function~~ (Hansen, 1969;Wiscombe and Grams, 1976;Sagan and Pollack, 1967;Andrews et al., 2006) but significant bias may arise in~~. However, the~~ g-related PF_{HG} ~~may cause significant bias~~ when estimating photo-dissociation rates (Toublanc, 1996) and aerosol radiative forcing effects (Boucher, 1998). Up to now, there have been few studies that have assessed the ~~bias~~ deviation when replacing the ambient phase function with the g-related PF_{HG} (Pandey and Chakrabarty, 2016;Boucher, 1998;Wiscombe and Grams, 1976) and there is no study that uses field measurements of aerosol optical properties to estimate the bias. Moreover, variations in g can influence the evolution of the atmospheric vertical structure through its effects on the atmospheric radiative distribution. Kudo et al. (2016) also found that the vertical profile of the asymmetry factor plays an important role in altering vertical variations in the solar heating rate. Marshall et al. (1995) reported that a 10% overestimation of g can systematically reduce aerosol climatic forcing by 12% or more. Andrews et al. (2006) found that a 10% reduction in g would result in a 19% overestimation of atmosphere radiative forcing at the top of atmosphere (TOA). An accurate estimation of g can greatly help improve the assessment of the aerosol radiative effect.

There are ~~many~~ several methods available to derive the g of aerosol ~~g~~ particles under ~~for~~ the dry and ambient condition respectively. Horvath et al. (2016) measured the phase function of aerosols, calculated the g of aerosols, and found that the g-related PF_{HG} can be used as a good approximation of the measured phase function. Many works used the Mie model (Bohren and Huffman, 2007) to calculate the phase function and proved its reliability (Andrews et al., 2006;Marshall et al., 1995;Shettle and Fenn, 1979;Bian et al., 2017). Comprehensive attempts have been made to relate g

带格式的：缩进：首行缩进： 0 厘米

带格式的：缩进：首行缩进： 0 厘米

with the hemispheric backscatter fraction (b). The value of b is the ratio of light scattered into the backward hemisphere compared to total light scattered in all directions (Wiscombe and Grams, 1976;Andrews et al., 2006;Horvath et al., 2016), with the definition of

$$b = \frac{\int_{\frac{\pi}{2}}^{\pi} P(\theta)\cdot \sin\theta \cdot d\theta}{\int_{0}^{\pi} P(\theta)\cdot \sin\theta \cdot d\theta}. \quad (3)$$

The main advantage of the backscatter ratio is that it can be measured with an integrating nephelometer equipped with a backscatter shutter (Charlson et al., 1974).

The free parameter g varies significantly for different aerosol types and different seasons. In ~~the~~ previous study, the g values are studied mainly by using the Mie scattering theory and the measured aerosol particle numbers size distribution (PNSD). D'Almeida et al. (1991) suggested that g at a wavelength of 500 nm ranges from 0.64 to 0.83 depending on the aerosol type and season. A mean value of 0.67 at an ambient relative humidity (RH) was also recommended (D'Almeida et al., 1991). Hartley and Hobbs (2001) reported a median g value of 0.7 for aerosols along the east coast of the United States. Formenti et al. (2000) measured Saharan dust aerosol and found that the aerosol g values ranged from 0.72-0.73. Biomass burning aerosols in Brazil had a low g value of 0.54 (Ross et al., 1998).

Some works have studied the impacts of aerosol hygroscopic growth on the parameter g (Hartley and Hobbs, 2001;Kuang et al., 2015;Andrews et al., 2006) and found that variations in g with RH can have significant influences on aerosol radiative effects (Kuang et al., 2015;Kuang et al., 2016;Andrews et al., 2006). A parameterization scheme of g, that takes RH and aerosol hygroscopic growth into account, is necessary.

When exposed to the ambient atmosphere, aerosols can grow by taking up water, which causes their corresponding optical properties to ~~considerable~~ change considerably. The κ-Köhler theory (Petters and Kreidenweis, 2007) is widely used to describe the hygroscopic growth of aerosol particles by using a single aerosol hygroscopic growth parameter (κ) and the κ-Köhler equation, which is shown as

$$\frac{RH}{100} = \frac{gf^3 - 1}{gf^3 - (1-\kappa)} \cdot \exp\left(\frac{4\sigma_{s/a} M_{water}}{R\cdot T\cdot D_d \cdot gf \cdot \rho_w}\right), \quad (4)$$

where D_d is the dry particle diameter; gf(RH) is the aerosol growth factor, ~~which is~~ defined as the ratio of the aerosol diameter at a given RH ~~and~~ to the dry aerosol diameter (D_{RH}/D_d); T is the

temperature; $\sigma_{s/a}$ is the surface tension of the solution; M_{water} is the molecular weight of water; R is the universal gas constant and ρ_w is the density of water. The aerosol hygroscopic growth parameter κ can be further used to investigate the influence of aerosol hygroscopic growth on aerosol optical properties (Tao et al., 2014;Kuang et al., 2015;Zhao et al., 2017) and aerosol liquids water contents (Bian et al., 2014).

According to the Mie theory, g is associated with aerosol particle number size distribution, the particle complex refractive index, the aerosol mixing state and ambient RH. At the same time, the aerosol morphology has significant influence on g. Datasets from the humidified nephelometer system can partially account for all of these factors. The humidified nephelometer system consists of two parallel nephelometers, ~~:~~ one ~~nephelometer~~ of which measures dry aerosol scattering properties and the other measures aerosol scattering properties under well-controlled RH conditions. This ~~results in~~system can give the light scattering enhancement factor (f_{RH}), which is defined as $f_{RH}(\lambda)= \sigma_{sca(\lambda)}/\sigma_{sca(\lambda)}$, or the ratio of the aerosol scattering coefficient under given RH conditions to that ~~of~~ under dry conditions. Each nephelometer can provide a scattering coefficient (σ_{sca}) and back-scattering coefficient (β_{sca}) at three wavelengths (450, 525 and 635nm). σ_{sca} can be used to calculate the aerosol scattering Ångstrom index, which reflects the aerosol PNSD to some extent. In general, a larger value for the Ångstrom index always corresponds to a smaller predominant aerosol size. Variations in β_{sca} and σ_{sca} can be used to deduce the aerosol BC mixing state (Ma et al., 2012). At the same time, datasets from the humidified nephelometer system can also be used alone to measure the aerosol hygroscopicity and provide an overall hygroscopic parameter κ (Kuang et al., 2017). ~~All in all~~In conclusion, measurements from the humidified nephelometer system might be used for estimating g under the given RH conditions. However, there is no clear relationship between the measured datasets from the humidified nephelometer and g. The non-linear influence of the above listed factors on g makes it difficult to parameterize the g.

Random forest machine learning model is a powerful technique that can be used for classification and non-linear regression (Huttunen et al., 2016;Breiman, 2001;Hu et al., 2017). This model is a widely used nonparametric machine learning algorithm that has several strengths. First, it involves fewer assumptions regarding the dependence between observations and outcomes when compared with traditional parametric regression models. Second, strict relationships among variables are not needed before implementing the random forest model. Third, this learning model requires much less

computing resource than ~~that of the~~ deep learning. Finally, this model has very low risk of over fitting by averaging over an ensemble of decision trees. Thus, the random forest machine learning model is used in this work to study the calculation of g based on the datasets of the humidified nephelometer system.

In this study, the Mie scattering theory and field measurements over the North China Plain (NCP) are used to study the characteristics of g. Section 2 describes the related datasets used in this study. Details of the study on the characteristics of g and impacts of aerosol hygroscopic growth on g are shown in section 3.1. A new method, which is based on a random forest machine learning model, is introduced to calculate g in section 3.2. We also discuss the impacts of g variations on the uncertainties of DARF in section 3.3, and the corresponding results are presented in section 4.3. Section 4.1 gives the calculated characteristics of g and section 4.2 proves the feasibility of using the machine learning model to calculate g. At the same time, this method is validated by the ambient aerosol phase function measured with a charge-coupled device -laser aerosol detective system (CCD-LADS). Conclusions are in section 5.

2. Instruments and datasets

Datasets used in this study come from three field campaigns, which were conducted at three different sites in the NCP. These three field measurements ~~are~~ were conducted at Gucheng in Hebei Province (Gucheng, 39°09' N, 115°44' E) from 15 October to 25 November in 2016, the AERONET BEIJING_PKU station in Beijing (PKU, 39°59' N, 116°18' E) from 21 March to 10 April in 2017, and the Yanqi Campus of the University of Chinese Academy of Sciences (UCAS, 40°24' N, 116°40' E) in the Huairou district, Beijing from 3 January to 27 January in 2016. Details of these locations are shown in Fig. S1. The PKU station is located at the northwest of Beijing, between the 4^{th} and 5^{th} ring road. It is 11km from the center of the Megacity Beijing, which is adjacent to Hebei Province and the megacity Tianjin. In the above three cities, the industrial manufacturing has led to heavy air pollution. Datasets for this location are representative of urban aerosols in the NCP. Gucheng is located between two megacities (120 km from Beijing and 190 km from Shijiazhuang) of NCP and the pollution conditions of Gucheng can be a good representation of the continental background in the NCP. Details for the Gucheng station can be found at Kuang et al. (2017). The UCAS station is 60 km away from the center of Beijing and is at the edge of the NCP, which makes it suitable for measuring the regional

pollution properties of the NCP (Ma et al., 2016). More details of the measurement sites ~~can refer to~~ are available in section 1 of the supplementary materials.

Table 1 lists the information for the field campaigns and the datasets used in this study. During the campaigns, sampled aerosols that had an aerodynamic diameter of less than 10 μm are selected by an impactor (Mesa Labs, Model SSI2.5) at the inlet. These aerosols are then dried to below 30% RH with a Nafion drying tube and then lead to each instrument. Aerosol PNSDs ranging from 3 nm to 10 μm are measured by using the scanning mobility particle size spectrometer (SMPS, TSI Inc., model 3936) and an aerodynamic particle sizer (APS, TSI Inc., model 3321) with a temporal resolution of 5 min. Black carbon (BC) mass concentrations are measured by a multi-angle absorption photometer (MAAP model 5012, Thermo, Inc., Waltham, MA USA) at UCAS and by an Aethalometer 33 (Hansen et al., 1984;Drinovec et al., 2015) at PKU and Gucheng. The aerosol σ_{sca} at wavelengths of 450 nm, 525 nm and 635 nm is measured by an Aurora 3000 nephelometer and the corresponding values are recorded every minute (Müller et al., 2011).

The f_{RH} is measured by a self-constructed humidified nephelometer system. ~~The detail information of the humidified nephelometer is described elsewhere (Kuang et al., 2017). Some brief descriptions about the humidified nephelometer are introduced here. A~~ In this system, a humidifier is used to control the RH of the sample aerosol and σ_{sca} is measured for each of the controlled RH. The sample aerosol is humidified through a Gore-Tex tube, which is surrounded by a circulating water layer in a stainless steel tube. The RH is changed by changing the temperature of the circulating water, which is controlled by the water bath and software. For each cycle, the RH points are set to range from about 50% to about 90% over 45 minutes. For most of the cases, the aerosol PNSDs are consistent over the cycle. These cycles of f_{RH} values are abandoned when the measured maximum and the minimum σ_{sca} value are beyond the range of 1.4 and 0.6 times of the mean measured scattering coefficient of each cycle. The detail information of the humidified nephelometer is described by Kuang et al. (2017).

Ambient aerosol phase function with a time resolution of 5 minutes is measured at UCAS by using a CCD-LADS. This system consists of a continuous laser, two charge-coupled device cameras and the corresponding fish eye lenses. The wavelength of the laser is 532nm and a quarter-wave plate was mounted in front of the laser emitter to change the polarization state of the laser from linear to circular. The CCD-LADS can measure the ambient aerosol phase function at a wide angular range of 10-170°

with a high resolution of 0.1°. More details of the measurement system can be found at Bian et al. (2017).

3. Methodology

3.1 Calculating characteristics of g based on the Mie scattering theory (g_{Mie})

The Mie model (Bohren and Huffman, 2007) is ~~employed~~applied to calculate the characteristics of g_{Mie}. When running the Mie model, aerosol PNSD, aerosol complex refractive index, BC mixing state and BC mass concentration are essential. Its results include aerosol phase function, and g_{Mie} can be calculated by the definition shown in formula 2.

Mixing states of the BC come from the measurements of the field measurements. From the work of Ma et al. (2012), the mixing states of BC in the NCP were presented as both core-shell mixed and externally mixed. Ma et al. (2012) provides the ratio of BC mass concentrations under an externally mixed state, M_{ext_BC}, to total BC mass concentration, M_{BC}, as follows:

$$r_{ext\_BC} = \frac{M_{ext\_BC}}{M_{BC}}. \quad (5)$$

The mean value of r_{ext_BC}=0.51 (Ma et al., 2012) is used ~~as a representation of the mixing state~~ in this study. The size-resolved distribution of BC mass concentration is the same as that used by Ma et al. (2012). The κ-Köhler theory and the Mie scattering model are employed to calculate g_{Mie} under different RH conditions. When the aerosol grows by taking up water~~gets hygroscopic growth~~, the BC is treated as a non-hygroscopic and insoluble core~~the water is assumed to mix only with the shell~~. The real time value κ, which is derived from the measurement of f_{RH}, is used to account for aerosol hygroscopic growth. For each RH value, the growth factor can be calculated based on formula 4. The corresponding ambient aerosol PNSD at a given RH can be determined too by applying the κ and formula 4. The refractive index ($\widetilde{m}$), which accounts for water content in the particle, is derived as a volume mixture between the dry aerosol and water (Wex et al., 2002a):

$$\widetilde{m} = f_{V,dry}\,\widetilde{m}_{aero,dry} + (1 - f_{V,dry})\,\widetilde{m}_{water}, \quad (6)$$

where $f_{v,dry}$ is the ratio of the dry aerosol volume to the total aerosol volume under a given RH condition; $\widetilde{m}_{aero,dry}$ is the refractive index for dry ambient aerosols and $\widetilde{m}_{water}$ is the refractive index of water.

The refractive indices of BC, non-light-absorbing aerosols and water, which are used in this study, are 1.8+0.54i (Kuang et al., 2015), $1.53+10^{-7}$i (Wex et al., 2002b) and $1.33+10^{-7}$i, respectively. Then,

the corresponding g values under the given RH and PNSD can also be calculated. More details on using the Mie model to calculate the aerosol phase function for different RH conditions can be found in Zhao et al. (2017).

3.2 Calculating g by using the random forest machine learning model (g_{ML})

In this study, the random forest machine learning model from the Scikit-Learn machine learning library (Hu et al., 2017;Pedregosa, 2011) was used to calculate g. The random forest model has two parameters: the number of input variables (n_{pre}) and the number of trees grown (n_{tree}). In this study, the n_{pre} and n_{tree} are determined by minimize the relative difference of the g_{ML} and g_{Mie}. Details of choosing the values of n_{pre} and n_{tree} are shown in section 2 of the supplementary. The n_{pre} and n_{tree} are set as eight and thirty-two in this study, respectively. The eight input parameters include the three dry scattering coefficients, three dry backscattering coefficients, RH and κ.

The measured datasets are divided into two parts: one ~~for~~ as the training data ~~of~~ for the random forest model, and the other ~~for~~ as the testing data. All training datasets come from field measurements at Gucheng station, whereas the datasets from PKU are employed to test the accuracy of the model. With split datasets from different sites, the feasibility of the random forest model in the NCP can be guaranteed. Before calculating g_{Mie}, we compare the measured σ_{sca} from the dry nephelometer and calculate σ_{sca} from the Mie scattering model. These data, where the relative difference between the measured and calculated σ_{sca} is within 30%, are used for the following analyses. With this, the inaccuracy form the measurement of the instruments can be avoided to some extent. More details regarding the used data are shown in section 3 of the supplementary material.

To further avoid the uncertainties of the measurements when training the random forest machine learning model, both the required input parameters and the predictors, g values, come from the calculation of the Mie scattering model using the measurement of the aerosol PNSD and BC from the field campaign of Gucheng. For each measured PSND and BC, the corresponding σ_{sca} and β_{sca} under dry condition at the wavelength of 450nm, 525nm and 635nm are modeled based on the Mie theory. With the concurrently measured κ values from the humidified nephelometer, the g_{Mie} values under different RH can be determined too. Then the modeled σ_{sca}, β_{sca} under dry condition, the κ values and the RH are used as the input data for the model and the corresponding g_{Mie} values are used as the prediction data.

3.3 Aerosol DARF estimations

The earth-atmosphere systems can be significantly influenced by aerosols through the scattering and absorption of the energy.~~, which scatter and absorb the energy.~~ In this study, the Santa Barbara DISORT (discrete ordinates radiative transfer) Atmospheric Radiative Transfer (SBDART) model (Ricchiazzi et al., 1998) is employed to estimate the DARF. The characteristics of DARF with the variations in g are studied.

The instantaneous DARF is calculated at the TOA for cloud-free conditions. DARF is defined as the difference between radiative flux at the TOA under present aerosol conditions and aerosol-free conditions:

$$\mathrm{DARF} = \left(f_a \downarrow - f_a \uparrow\right) - \left(f_m \downarrow - f_m \uparrow\right), \quad (7)$$

where $\left(f_a \downarrow - f_a \uparrow\right)$ is the downward radiative irradiance flux with given aerosol distributions and $\left(f_m \downarrow - f_m \uparrow\right)$ is the radiative irradiance flux under aerosol free conditions. The DARF at 50km is calculated because almost all of the aerosols are distributed within the height of 50 km~~located at the range of 0 and 50km~~ in the parameterization scheme ~~of aerosol vertical distribution~~ (Liu et al., 2009). The wavelengths in the range from 0.25 to 4 μm are calculated for irradiance in this study.

Input data for the SBDART are listed below. Vertical profiles of the aerosol optical properties, which include the aerosol extinction coefficient (σ_{ext}), aerosol single scattering albedo (SSA) and g. They all have a vertical ~~with a height~~ resolution of 50 m and~~,~~ come from the results of the Mie scattering and the parameterized aerosol vertical distributions. Methods for parameterization and calculation of the aerosol optical profiles can be found in section 4 of supplementary material or ~~relate~~ refer to Kuang et al. (2016) and Zhao et al. (2017). Atmospheric meteorological parameter profiles come from the results of the intensive radiosonde observations at the Meteorological Bureau of Beijing (39°48' N, 116°28' E) at the local time of 13:30 from July to September in 2008. Kuang et al. (2016) studied these measured profiles and found that the vertical distributions of these parameters, which include profiles for water vapor, pressure and temperature, can be used as a good representation of the meteorological parameter profiles in the NCP during the summer. The corresponding measured mean results during field measurement are used in this study and the details of these profiles are shown in section 4 of the supplementary material. Surface albedo values are obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) V005 Climate Modeling Grid (CMG) Albedo Product (MCD43C3). The mean results of the surface albedo of Beijing from July to September in

2008 are used. The remaining input data for the SBDART are set to their default values (Ricchiazzi et al., 1998).

4 Results and Discussion

4.1 Characteristics of g_{Mie}

4.1.1 Characteristics of g_{Mie} at different sites

Fig. 1 gives the statistical results for the calculated g properties at Gucheng, PKU and UCAS. The RH at the three sites shows almost the same diurnal variation pattern in Fig. 1 (a) (b) and (c). The RH reaches a peak in the morning at approximately 6:00 am, and then reaches its lowest value at approximately 16:00 in the afternoon. However, the mean values of RH are 77.7%±20.9% at Gucheng, 47.8%±20.8% at PKU and 33.49±15.22% at UCAS. The g_{Mie} values under dry conditions that are calculated by the measured PNSD have almost no diurnal patterns. The g_{Mie} values at PKU (0.614±0.025) are slightly lower than those at Gucheng (0.601±0.021) and UCAS (0.595±0.023) as shown in Fig. 1 (d), (e) and (f). The difference in g_{Mie}values results from different aerosol properties at these sites. From fig. S6, the peak diameter of the mean and median PNSD at Gucheng locates around 150nm. However, the peak diameter of the mean and median PNSD at PKU locates at around 100nm. The peak values of the mean and median diameter of the aerosol PNSD at UCAS locates at around 60nm. At the same time, there are large partitions of small particles that are lower than 60nm at PKU and UCAS. However, these particles, which are lower than 100nm, contribute little to the total aerosol scattering. The aerosol PNSD at PKU is more dispersed than that of the Gucheng and UCAS, which corresponds to a larger variation in the g values. From fig. S6 (h), (i) and (j), the size distribution of the aerosol scatter coefficient at around 500nm contributes less to the scatter coefficient at PKU than at that of the Gucheng and UCAS. Thus these particles with the diameter larger than 500nm contribute more to the aerosol scattering coefficient. As g_{Mie} increase with the aerosol diameter, the aerosol g_{Mie} under dry conditions at PKU tends to be larger than that at Gucheng and UCAS.

However, ambient g_{Mie} values have different patterns at different sites, as shown in Fig. 1 (h), (i) and (j). The g_{Mie} values have an RH-related diurnal pattern at Gucheng, with a mean value of 0.668±0.073, but show no diurnal variation at PKU and UCAS, where the mean values of g_{Mie} are 0.639±0.049 and 0.618±0.033, respectively. The variations of ambient g_{Mie} values are mainly resulted from the variation of the aerosol hygroscopic growth under the ambient condition, which is highly

related to the ambient RH. The g_{Mie} value is significantly influenced by RH when the RH is higher than 80%, which will be detailed in section 4.1.2. Ambient g_{Mie} values at Gucheng, PKU and UCAS can vary from 0.57 to 0.8, 0.55 to 0.76 and 0.56 to 0.72 respectively, comparable to those of Andrews et al. (2006), which range from 0.59 to 0.72.

4.1.2 Influence of RH on g

To assess the influence of RH on g, the g_{Mie} values are calculated under different RH conditions for each aerosol PNSD. The statistical results of g_{Mie} versus RH are shown in Fig. 2. The g_{Mie} value has a wide variation range between 0.54 and 0.67 with the mean value located at 0.61 under dry conditions~~a mean of 0.61 at dry conditions and can vary widely (0.54 to 0.67), which corresponds to approximately 10% of the variation~~. However, the mean g_{Mie} value can ~~vary~~ change from 0.65 to 0.8 when the RH reaches 90%. The g_{Mie} enhancement factor, which is defined as the ratio of g_{Mie} at a given RH and g_{Mie} under dry conditions, can reach a mean value of 1.2 at an RH of 90%, which means that the g_{Mie} value under wet conditions is approximately 20% higher than that under the dry conditions. This finding is consistent with that of Hartley and Hobbs (2001), who found that g is highly related to the RH.

~~However~~Contrary to RH, the aerosol complex refractive index has little influence on g and the uncertainties for g are less than 0.004 based on the Monte Carlo simulation of the g at different complex refractive index values. More details of discussing the influence of aerosol complex refractive index on g can be referred to in~~relate to~~ section 6 of the supplementary materials.

4.2 Calculating g_{ML} by using the machine learning model

4.2.1 Feasibility of using the random forest model

We establish two independent random forest machine learning models to predict g_{ML} values under dry conditions and under ambient RH conditions separately.

When ~~running~~ the random forest machine learning model are run for g values under dry conditions, σ_{sca} and β_{sca} at three different wavelengths, are used as the input for independent variables. The other two input parameters, RH and κ, are set equal to zero. The predictor g values come from the results of the Mie scattering model. Fig. 3(a) shows the calculated g_{Mie} values and predicted g_{ML} values by the random forest machine learning model under dry conditions at the site of PKU. The results show that the g_{Mie} values and g_{ML} values have good consistency with an R^2 value of 0.98. There are 95%

of the cases that the relative difference between g_{Mie} and g_{ML} are within the relative differences of 2.56%.

Fig. 3(b) shows the comparison of the predicted g_{ML} values under different RH conditions and g_{Mie} values calculated by the Mie scattering model. The correlation coefficient between g_{Mie} and g_{ML} reaches 0.93 and 95% of the relative differences are within 4.02%. The random forest model can be a good method to predict g values under different RH conditions with high accuracy and the uncertainties of predicting g values using the random forest machine learning model is estimated to be 4.02%.

The filled colors of the dots in Fig. 3 represent the concurrently measured σ_{sca}. It is shown that with an increase in σ_{sca}, g values tend to be larger, which is in accordance with the particle scattering properties. When a particle has larger diameters, the σ_{sca} of the particle is higher, and there tends to be a larger partition of forward scattering light.

Wiscombe and Grams (1976) studied the relationship between b and g and gave the expression between them as follows:

$$g = -7.143889 \cdot b^3 + 7.464439 \cdot b^2 - 3.96356 \cdot b + 0.9893 \quad (8).$$

This equation is widely used to calculate g from b (Andrews et al., 2006;Horvath et al., 2016;Kassianov et al., 2007). We use the field measurement results to test its reliability. The comparison results between calculated g values from the Mie scattering model and parameterized g values from equation 6 are shown in Fig.S9. From fig.S9, we can see that the parameterized g values are prevalently larger than the calculated g values by approximately 10%. When the σ_{sca} is smaller, the deviations become larger. Some other empirical relationships between b and g (Moosmüller and Ogren, 2017) are also tested. These parameterization scheme has almost the same result as Wiscombe and Grams (1976). This result means that the previously established parameterization scheme is not applicable in the NCP

4.2.2 Sensitivity of the random forest model

Sensitivity studies are carried out to assess the influence of each input variable on g_{ML}. Based on the works of Müller et al. (2011), the uncertainties in total scattering are 4% (450nm), 2% (525nm), 5% (635nm) for experiments with ambient air and laboratory generated white particles. For backscattering, the differences are higher and amount 7% (450nm), 3% (525nm) and 11% (635nm). The uncertainties of the measured RH by the RH sensors is 1.7% for RH ranges from 0 to 90%(Kuang et al., 2017) and

the uncertainties of the derived κ values is 6% (Kuang et al., 2017). The Monte Carlo simulations are conducted to study the sensitivities of the g_{ML} to the input parameters in three steps. First, the mean results of the measured dry σ_{sca}, dry β_{sca}, RH and κ values are used to predict the g value. Second, the dry σ_{sca} at 450 nm are randomly changed with a mean value of 0 and standard deviation of 4% and the other input are kept unchanged as the input. The corresponding standard deviation of the predicted g value is used as the sensitivities of the g_{ML} to the σ_{sca} at 450nm. At last, the sensitivities are carried out accordingly for each of the input parameter. The uncertainties of g_{ML} values to the input parameters are estimated. The total uncertainties of predicting g RH are derived when all of the input parameters are randomly changed with their corresponding uncertainties. For each test, the Monte Carlo simulations are carried out for 20000 times.

Table 2 gives the two time of the standard deviation of the g_{ML} values corresponding to the uncertainties of the input parameters. Form table 2, it is shown that the uncertainties of measured σ_{sca} has little influence of the g_{ML} with 0.487%, 0.492% and 0.486% for wavelength of 450nm, 525nm and 635nm respectively. However, the measurement of the three β_{sca} have larger uncertainties and lead to greater influence on predicting g_{ML} with 0.651%, 0.486% and 0.710%. The uncertainty of the RH has little influence on predicting g_{ML} with 0.487%. However, the uncertainty of derived κ values (6%) influence the g values most with 1.92%. The total uncertainties of predicting g due to the uncertainties of the measurment is 1.95%. All in all, the total uncerntaities of predicting the g_{ML} is estimated to be 4.47% considering the 4.02% uncertainties of the random forest machine learning model from section 4.2.1.

4.2.3 Validation of the random forest machine learning model

Datasets of the UCAS campaign are also used to validate the random forest machine learning model. On one hand, the g_{ML} values are calculated by using the random forest machine learning model with the measurements of the humidified nephelometer. On the other hand, ambient g values are calculated by using the measured phase function from the CCD-LADS g_{CCD} according to the definition shown in formula 2. Then the g values calculated with the two methods are compared.

Comparison results of these two kinds of g values are shown in fig. 4. Form fig.4, the values of g_{ML} and g_{CCD} show good consistence. There are 95% of the conditions that the relative differences between the g_{ML} and g_{CCD} are in the range of 6.5% which is a little higher than the relative difference of the g values (4.02%) between machine learning method and the Mie scattering method.

During the period, the σ_{sca} range from 30 to 260 Mm^{-1} which lead to cleaner conditions in UCAS than in Gucheng and PKU. Correspondingly, most of the g_{Mie} values are small and locate at the range of 0.54 to 0.62 which are obviously lower than those in other campaigns. At the same time, the surrounding condition at UCAS during the winter is relative dry, which results to small g values. These conditions may partially explain the relatively higher difference between the g_{ML} and g_{CCD}. With this validation, we conclude that the random forest machine learning model can give a reasonable g value based on the measurements of the humidified nephelometer system.

4.3 Estimating the impacts of g on DARF

4.3.1 Uncertainties of replacing the calculated phase function with the PF_{HG}

When the PF_{HG} is used to parameterize the calculated phase function by using the Mie theory (PF_{Mie}), there are some deviations and the influence of these deviations should be estimated. The relative difference between the DARF from the PF_{Mie} and from the PF_{HG} is used to estimate uncertainties when using the PF_{HG}. First, the PF_{Mie} profiles are used as inputs to estimate DARFs. The PF_{Mie} is then replaced with the g-related PF_{HG} which is parameterized by g_{Mie} from the PF_{Mie}, and the DARFs are calculated again. These relative differences between the DARFs from the above two steps are recorded and compared. The relative differences at different zenith angle conditions are calculated to comprehensively estimate the influence of the PF_{HG}.

Fig.5 shows the estimated DARFs at different zenith angles. In Fig. 5(a), DARF at the TOA can vary from -2.55 to -4.8 w/m^2. When the PF_{Mie} is replaced by the PF_{HG}, the calculated DARF ranges from -2.6 to -5.1 w/m^2. The relative difference of the DARFs between the two methods ranges from 1.3% to 7.1%, as shown in Fig. 5(b). It is concluded that using the g-related PF_{HG} to replace the PF_{Mie} to estimate aerosol radiative effects is applicable in the NCP, with a deviation of less than 7%.

4.3.2 Impacts of g variations on DARF estimation

Variations in g can lead to significant ~~variations~~ changes in the estimated DARF (Kuang et al., 2016;Andrews et al., 2006;Mccomiskey et al., 2008). In this study, the uncertainties of the g values from the input parameter is estimated to be 1.95% ~~when predicting g~~ and the total variation in running the random forest machine learning model is estimated to be 4.47%. At the same time, the g can varies about 10% for different aerosol PNSD and can be enhanced by 20% with the increment of RH from 30% to 90%. It is very important to know the extent of the variation in DARF corresponding to the uncertainties from g.

The variation in DARF from the uncertainties of g is calculated by increasing or decreasing g by 1.95%, 4.47%, and 10% to the original g values, and then comparing the corresponding DARFs with the original ~~DARFs~~ones. To study the influence of RH on g and DARF, the DARF with the g values calculated from the dry parameterized aerosol population profile, is estimated.

Fig. 6 shows the estimated DARFs with different variation in g and the corresponding variations in the estimated DARF. The results show that when g varies by 1.95%, the DARF can vary 4%. However, variations of 4.47% and 10% in g values can lead to variations in the estimated DARF with 9.4% and 21%, respectively.

The estimated DARF using the parameterized aerosol profile, which considers the aerosol hygroscopic growth, is smaller than the DARF using the g profiles from the dry aerosol population. The g values under dry condition are smaller than that of the wet ambient. Thus, there is larger partition of energy that is scattered forward which leads to less outgoing backscattering energy and a larger value of the estimated DARF.

When the DARF are estimated ignoring the impacts of aerosol hygroscopic growth on g, the relative difference can be as high as 20% for all of the zenith angles. Thus, i~~I~~t is necessary to consider~~ing~~ the aerosol hygroscopic growth when calculating the g values.

5 Conclusions

The characteristics of g in the NCP are studied based on the Mie scattering theory and field measurements from sites of Gucheng and PKU. The results show that g_{Mie} values are 0.604±0.025 at Gucheng and 0.615±0.021 at PKU. The ambient g_{Mie} values at Gucheng show obvious diurnal variations due to variations in RH. When the ambient RH reaches 90%, g_{Mie} can be enhanced by 20% and the g values ~~under~~ from different aerosol population can vary 10%. Comparison of the calculated g_{Mie} values from the Mie scattering model and the parameterized g values from the Wiscombe and Grams (1976) method shows that the parameterized g is overestimated by approximately 10% and that the deviations are ~~even greater~~become larger when the measured σ_{sca} is below 200 M m^{-1}.

The random forest machine learning model and datasets from the humidified nephelometer are employed to calculate g_{ML} values. The input data of the random forest model contain measured σ_{sca} and β_{sca} at three wavelengths, RH and the hygroscopic parameter κ. Except for RH, all input data came from measurements from the humidified nephelometer system (Kuang et al., 2017). The random forest model can significantly improve the accuracy of ~~predicting~~ g_{ML} prediction. The uncertainties of the

predicted g_{ML} values are constrained ~~to be~~ within 2.56% under dry conditions and 4.02% under ambient conditions and the uncertainties from the measurement of the humidified nephelometer can lead to a variation of 1.95% in g, which is mainly resulted from the inaccuracy of the derived κ. The total uncertainty of ~~calculating~~ g calculation using the random forest machine learning model is 4.47%. This is the first time that machine learning model and datasets from the humidified nephelometer system ~~and machine learning~~ are combined to study g. At the same time, this method can accounting for the influence of aerosol hygroscopic growth on g.

This ~~e~~ new method for calculating g is validated by comparing the g_{ML} values from the random forest machine learning model and the g_{CCD} values from the measured phase function by using the CCD-LADS. The g values with this two methods show good consistence with 95% of the data within the relative difference of 6.5%.

SBDART model is used to study the impacts of g on DARF. We first studied the relative differences between the estimated DARFs by using the PF_{HG} and the calculated phase function by using the Mie theory, the measured mean aerosol PNSD and BC mass concentration at the site of Gucheng and PKU. The results show that the relative differences in DARF can be contained within 7.1% when replacing the PF_{Mie} with g-related PF_{HG}. The PF_{HG} can be a feasible parameterization scheme to study DARF in the NCP.

The sensitivity study shows that the maximum uncertainties of DARF are 4%, 9.4% and 21%, which correspond to the uncertainties of the g from the instrument measurement, the machine learning model and the variation of aerosol PNSD. However, when the DARF are estimated ignoring the effects of aerosol hygroscopic growth on g, the relative differences of the DARF is as large as 20% for all of the zenith angles. It is necessary to parameterize the g ~~with~~ accounting for the effect of aerosol hygroscopic growth.

This work can further our understanding of the role of g in the aerosol radiative effects ~~of aerosols~~ and can help reduce uncertainties in estimating DARF.

Data availability. The measurement data involved in this study are available upon request to the authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (41590872) and the National Key R&D Program of China (2016YFC020000:Task 5).

References

Andrews, E., Sheridan, P. J., Fiebig, M., McComiskey, A., Ogren, J. A., Arnott, P., Covert, D., Elleman, R., Gasparini, R., Collins, D., Jonsson, H., Schmid, B., and Wang, J.: Comparison of methods for deriving aerosol asymmetry parameter, Journal of Geophysical Research, 111, 10.1029/2004jd005734, 2006.

Bian, Y., Zhao, C., Xu, W., Zhao, G., Tao, J., and Kuang, Y.: Development and validation of a CCD-laser aerosol detective system for measuring the ambient aerosol phase function, Atmos. Meas. Tech., 10, 2313-2322, 10.5194/amt-10-2313-2017, 2017.

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, Atmospheric Chemistry and Physics, 14, 6417-6426, 10.5194/acp-14-6417-2014, 2014.

Bohren, C. F., and Huffman, D. R.: Absorption and Scattering by a Sphere, in: Absorption and Scattering of Light by Small Particles, Wiley-VCH Verlag GmbH, 82-129, 2007.

Boucher, O.: On Aerosol Direct Shortwave Forcing and the Henyey–Greenstein Phase Function, J Atmos Sci, 55, 128-134, 10.1175/1520-0469(1998)055<0128:OADSFA>2.0.CO;2, 1998.

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

Charlson, R. J., Porch, W. M., Waggoner, A. P., and Ahlquist, N. C.: Background aerosol light scattering characteristics: nephelometric observations at Mauna Loa Observatory compared with results at other remote locations, Tellus, 26, 345-360, 1974.

D'Almeida, G. A., Koepke, P., and Shettle, E. P.: Atmospheric Aerosols: Global Climatology and Radiative Characteristics, Journal of Medical Microbiology, 54, 55-61, 1991.

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.

Formenti, P., Andreae, M. O., and Lelieveld, J.: Measurements of aerosol optical depth above 3570 m asl in the North Atlantic free troposphere: results from ACE - 2, Tellus B, 52, 678-693, 10.1034/j.1600-0889.2000.00006.x, 2000.

Hansen, A. D. A., Rosen, H., and Novakov, T.: The aethalometer — An instrument for the real-time measurement of optical absorption by aerosol particles, Science of The Total Environment, 36, 191-196, http://dx.doi.org/10.1016/0048-9697(84)90265-1, 1984.

Hansen, J. E.: EXACT AND APPROXIMATE SOLUTIONS FOR MULTIPLE SCATTERING BY CLOUDY AND HAZY PLANETARY ATMOSPHERES, J Atmos Sci, 26, 478-&, 10.1175/1520-0469(1969)026<0478:eaasfm>2.0.co;2, 1969.

Hartley, W. S., and Hobbs, P. V.: An aerosol model and aerosol-induced changes in the clear-sky albedo off the east coast of the United States, Journal of Geophysical Research: Atmospheres, 106, 9733-9748, 10.1029/2001jd900025, 2001.

Horvath, H., Kasahara, M., Tohno, S., Olmo, F. J., Lyamani, H., Alados-Arboledas, L., Quirantes, A., and Cachorro, V.: Relationship between fraction of backscattered light and asymmetry parameter, Journal of Aerosol Science, 91, 43-53, 10.1016/j.jaerosci.2015.09.003, 2016.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., and Liu, Y.: Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach, Environ Sci Technol, 51, 6936-6944,

513 10.1021/acs.est.7b01210, 2017.
514 Huttunen, J., Kokkola, H., Mielonen, T., Mononen, M. E. J., Lipponen, A., Reunanen, J., Lindfors, A. V., Mikkonen, S.,
515 Lehtinen, K. E. J., Kouremeti, N., Bais, A., Niska, H., and Arola, A.: Retrieval of aerosol optical depth from surface solar
516 radiation measurements using machine learning algorithms, non-linear regression and a radiative transfer-based look-up
517 table, Atmospheric Chemistry and Physics, 16, 8181-8191, 10.5194/acp-16-8181-2016, 2016.
518 Kassianov, E. I., Flynn, C. J., Ackerman, T. P., and Barnard, J. C.: Aerosol single-scattering albedo and asymmetry parameter
519 from MFRSR observations during the ARM Aerosol IOP 2003, Atmos. Chem. Phys., 7, 3341-3351,
520 10.5194/acp-7-3341-2007, 2007.
521 Kuang, Y., Zhao, C. S., Tao, J. C., and Ma, N.: Diurnal variations of aerosol optical properties in the North China Plain and
522 their influences on the estimates of direct aerosol radiative effect, Atmos. Chem. Phys., 15, 5761-5772,
523 10.5194/acp-15-5761-2015, 2015.
524 Kuang, Y., Zhao, C. S., Tao, J. C., Bian, Y. X., and Ma, N.: Impact of aerosol hygroscopic growth on the direct aerosol radiative
525 radiative effect in summer on North China Plain, Atmospheric Environment, 147, 224-233, 2016.
526 Kuang, Y., Zhao, C., Tao, J., Bian, Y., Ma, N., and Zhao, G.: A novel method for deriving the aerosol hygroscopicity parameter
527 based only on measurements from a humidified nephelometer system, Atmos. Chem. Phys., 17, 6651-6662,
528 10.5194/acp-17-6651-2017, 2017.
529 Kudo, R., Nishizawa, T., and Aoyagi, T.: Vertical profiles of aerosol optical properties and the solar heating rate estimated
530 by combining sky radiometer and lidar measurements, Atmospheric Measurement Techniques, 9, 3223-3243,
531 10.5194/amt-9-3223-2016, 2016.
532 Liu, P., Zhao, C., Zhang, Q., Deng, Z., Huang, M., Xincheng, M. A., and Tie, X.: Aircraft study of aerosol vertical distributions
533 over Beijing and their optical properties, Tellus Series B-Chemical & Physical Meteorology, 61, 756–767, 2009.
534 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.,
535 Müller, K., Herrmann, H., Yan, P., Zhou, X. J., and Wiedensohler, A.: A new method to determine the mixing state of light
536 absorbing carbonaceous using the measured aerosol optical properties and number size distributions, Atmos. Chem. Phys.,
537 12, 2381-2397, 10.5194/acp-12-2381-2012, 2012.
538 Ma, N., Zhao, C., Tao, J., Wu, Z., Kecorius, S., Wang, Z., Größ, J., Liu, H., Bian, Y., Kuang, Y., Teich, M., Spindler, G., Müller, K.,
539 van Pinxteren, D., Herrmann, H., Hu, M., and Wiedensohler, A.: Variation of CCN activity during new particle formation
540 events in the North China Plain, Atmospheric Chemistry and Physics, 16, 8593-8607, 10.5194/acp-16-8593-2016, 2016.
541 Marshall, S. F., Covert, D. S., and Charlson, R. J.: Relationship between asymmetry parameter and hemispheric backscatter
542 ratio: implications for climate forcing by aerosols, Applied Optics, 34, 6306-6311, 10.1364/AO.34.006306, 1995.
543 Mccomiskey, A., Schwartz, S. E., Schmid, B., Guan, H., Lewis, E. R., Ricchiazzi, P., and Ogren, J. A.: Direct aerosol forcing:
544 Calculation from observables and sensitivities to inputs, Journal of Geophysical Research Atmospheres, 113, -, 2008.
545 Moosmüller, H., and Ogren, J. A.: Parameterization of the Aerosol Upscatter Fraction as Function of the Backscatter
546 Fraction and Their Relationships to the Asymmetry Parameter for Radiative Transfer Calculations, Atmosphere, 8, 133,
547 10.3390/atmos8080133, 2017.
548 Müller, T., Laborde, M., Kassell, G., and Wiedensohler, A.: Design and performance of a three-wavelength LED-based total
549 scatter and backscatter integrating nephelometer, Atmos. Meas. Tech., 4, 1291-1303, 10.5194/amt-4-1291-2011, 2011.
550 Pandey, A., and Chakrabarty, R. K.: Scattering directionality parameters of fractal black carbon aerosols and comparison
551 with the Henyey-Greenstein approximation, Opt Lett, 41, 3351-3354, 10.1364/OL.41.003351, 2016.
552 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion,B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R.,
553 Dubourg,V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot,M., and Duchesnay, E.:: Scikit-learn: Machine
554 Learning in Python, J. Mach. Learn. Res., 12, 2825-2830, 2011.
555 Petters, M. D., and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation
556 nucleus activity, Atmos. Chem. Phys., 7, 1961-1971, 10.5194/acp-7-1961-2007, 2007.

Ricchiazzi, P., Yang, S., Gautier, C., and Sowle, D.: SBDART: A Research and Teaching Software Tool for Plane-Parallel Radiative Transfer in the Earth's Atmosphere, Bulletin of the American Meteorological Society, 79, 2101-2114, 10.1175/1520-0477(1998)079<2101:sarats>2.0.co;2, 1998.

Ross, J. L., Hobbs, P. V., and Holben, B.: Radiative characteristics of regional hazes dominated by smoke from biomass burning in Brazil: Closure tests and direct radiative forcing, Journal of Geophysical Research Atmospheres, 103, 31925–31941, 1998.

Russell, P. B., Kinne, S. A., and Bergstrom, R. W.: Aerosol climate effects: Local radiative forcing and column closure experiments, Journal of Geophysical Research Atmospheres, 102, 9397-9407, 1997.

Sagan, C., and Pollack, J. B.: Anisotropic nonconservative scattering and the clouds of Venus, Journal of Geophysical Research, 72, 469-477, doi:10.1029/JZ072i002p00469, 1967.

Shettle, E. P., and Fenn, R. W.: Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties, Lancet, 48, 504, 1979.

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

Toublanc, D.: Henyey-Greenstein and Mie phase functions in Monte Carlo radiative transfer computations, Applied Optics, 35, 3270-3274, 10.1364/ao.35.003270, 1996.

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, J Geophys Res-Atmos, 107, 10.1029/2000jd000234, 2002a.

Wex, H., Neusüß, 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, LAC 4-1-LAC 4-18, 10.1029/2000jd000234, 2002b.

Wiscombe, W. J., and Grams, G. W.: The Backscattered Fraction in two-stream Approximations, J Atmos Sci, 33, 2440-2451, 10.1175/1520-0469(1976)033-2440, 1976.

Zhao, G., Zhao, C., Kuang, Y., Tao, J., Tan, W., Bian, Y., Li, J., and Li, C.: Impact of aerosol hygroscopic growth on retrieving aerosol extinction coefficient profiles from elastic-backscatter lidar signals, Atmos. Chem. Phys. Discuss., 2017, 1-24, 10.5194/acp-2017-240, 2017.

588

589 **Table 1.** Field information, dataset information and instruments that are used in this study.

Field information		Datasets and instruments				
Location	Time period	PSND	BC	σ_{sc}	f_{RH}	Phase function
Gucheng, Hebei (39°09' N, 115°44' E)	15 Oct to 25 Nov, 2016	SMP, APS	AE33	Aurora 3000	Humidified Nephelometer	None
PKU, Beijing (39°59' N, 116°18' E)	21 Mar to 10 Apr, 2017	SMPS, APS	AE33	Aurora 3000	Humidified Nephelometer	None
UCAS, Beijing (40°24' N, 116°40' E)	3 Jan to 27 Jan, 2016	SMPS, APS	MAAP	Aurora 3000	Humidified Nephelometer	CCD-LADS

590

591

592 **Table 2.** The sensitivities of g to the input parameters.

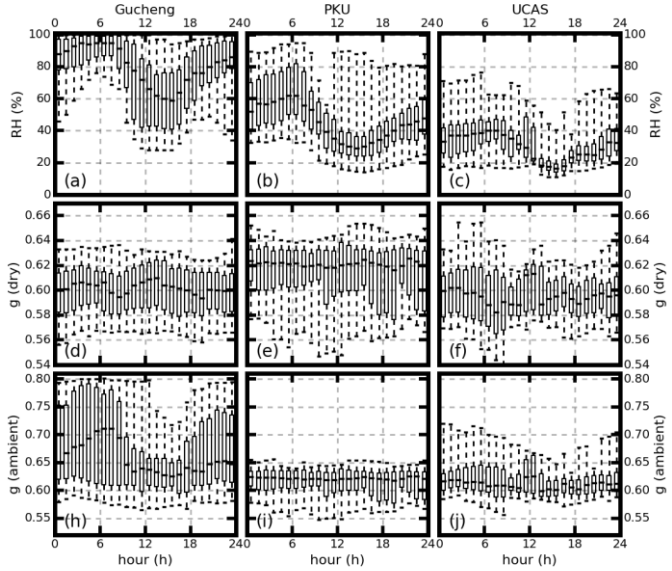
Parameter	$\sigma_{sca,450}$	$\sigma_{sca,525}$	$\sigma_{sca,635}$	$\beta_{sca,450}$	$\beta_{sca,525}$	$\beta_{sca,635}$	RH	κ	total
Parameter(%)[*1]	4	2	5	7	3	11	6	6	
g(%)[*2]	0.487	0.492	0.486	0.651	0.487	0.710	0.486	1.920	1.950

593 *1. The uncertainties of the measured parameters.

594 *2. The uncertainties of g values due to the uncertainties of the measurement parameters.

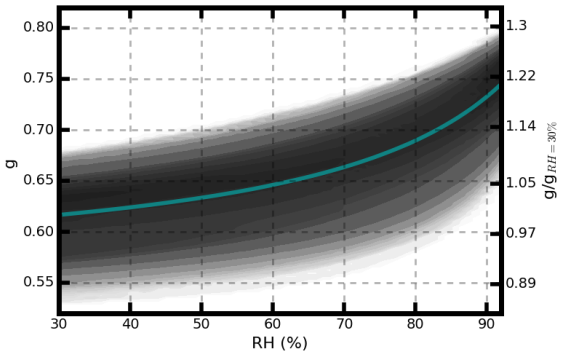
595

596



597 **Figure 1.** (a)(b)(c) Average diurnal pattern of RH, (d)(e)(f) g values calculated from dry aerosols, and
598 (h)(i)(g) g values from ambient aerosols. The panels (a), (d) and (h) are the results from Gucheng.
599 Panels (b), (e) and (i) are the results from PKU. Panels (c),(f) and (g) are the results of UCAS. The box
600 and whisker plots represent the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles.
601

602



603 **Figure 2.** Probability distributions of g under different RH conditions. The ticks on the left show g
604 values at different RH values, and the ticks on the right show the g enhancement factor, which is
605 defined as the ratio of g at a given RH to the g value at dry conditions (RH=30%). The solid line (cyan)
606 shows the mean result of g values and the enhancement factor at different RH values.
607

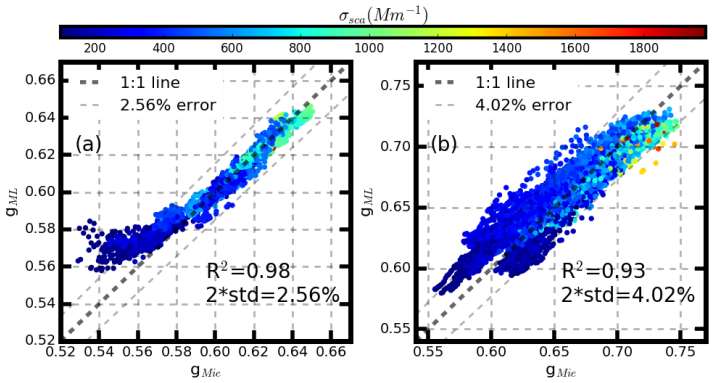
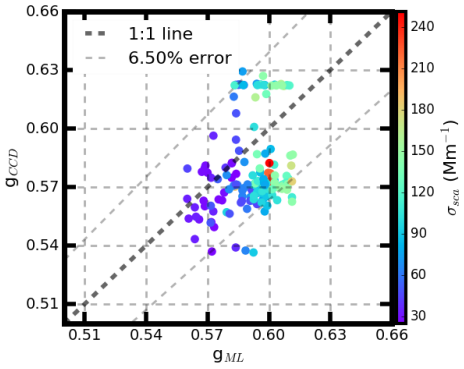
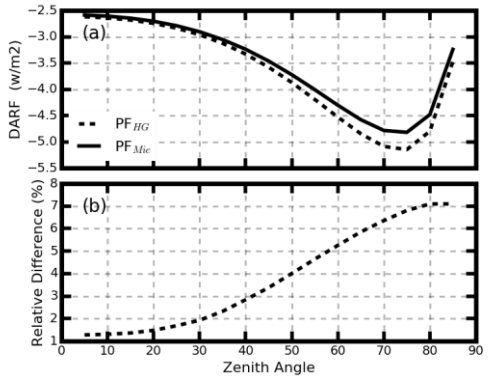


Figure 3. Comparison of calculated g values (g_{Mie}) from the Mie model and predicted g values (g_{ML}) from the random forest model under (a) dry conditions and (b) ambient conditions at the site of PKU. Colored dots represent the concurrently measured σ_{sca} corresponding to the time of g.



613 **Figure 4.** Comparison of the calculated g values (g_{CCD}) from the CCD-LADS measured phase function
614 and the calculated g values (g_{ML}) by using the random forest machine learning model.
615

616
617

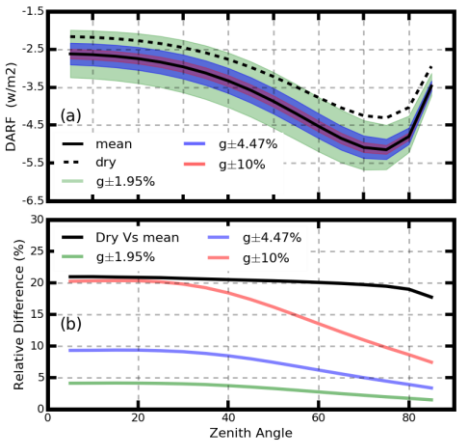


618

619 **Figure 5. (a)** Estimated DARFs at different zenith angles when using the g-related PF_{HG} (dotted line)
620 and the phase function calculated by using the Mie scattering theory (solid line). (b) The relative
621 difference between the DARFs in (a).

622

623



624 **Figure 6.** The variation in DARF when g varies by a range of 1.95% (the filled dark color), 10% (grey
625 color), and 20% (light grey color). Different line styles represent the corresponding mean relative
626 differences in DARF compared to the original value.