New correction method of scattering coefficient measurements of a three-wavelength nephelometer

Jie Qiu[1], Wangshu Tan[1,2], Gang Zhao[1,3], Yingli Yu[4,1], Chunsheng Zhao[1]

[1]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China
[2]School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China
[3]State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing 100871, China
[4]Economics & Technology Research Institute, China National Petroleum Corporation, Beijing 100724, China

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

Abstract. The aerosol scattering coefficient is an essential parameter for estimating aerosol direct radiative forcing, which can be measured by nephelometers. Currently, nephelometers have the problem of non-ideal Lambertian light source and angle truncation. Hence, the observed raw scattering coefficient data need to be corrected. In this study, based on the random forest machine learning model and taking Aurora 3000 as an example, we have proposed a new method to correct the scattering coefficient measurements of a three-wavelength nephelometer under different relative humidity conditions. The result shows that the empirical corrected values match Mie-calculation values very well at all the three wavelengths and under all the measured relative humidity conditions, with more than 85 % of the corrected values in error by less than 2 %. The correction method is valid to obtain scattering coefficient with high accuracy and there is no need for additional observation data.

1 Introduction

Atmospheric aerosol particles directly impact the earth radiative balance by scattering or absorbing the solar radiation. However, the uncertainty of aerosol direct radiative forcing varies greatly, ranging between -0.77 and 0.23 W/m^2 (IPCC, 2013), which poses a great challenge to the accurate quantification of its effects on the earth climate system. Aerosol scattering and absorbing coefficients are the two most important parameters for estimating aerosol direct radiative forcing, and part of the estimation uncertainty comes from the inaccuracy in their measurements. Therefore, more precise measurements are needed. In recent years, two commercial integrating nephelometers (Aurora 3000 and TSI 3563) have been developed to measure aerosol scattering coefficients and hemispheric backscattering coefficients at three different wavelengths (450 nm, 525 nm, 635 nm for Aurora 3000 and 450 nm, 550 nm, 700 nm for TSI 3563). The three-wavelength integrating nephelometer is widely employed in field measurements and laboratory studies due to its high accuracy in measuring aerosol scattering coefficients (Anderson et al., 1996). However, it has two primary drawbacks, namely the angle truncation and nonideal Lambertian light source, contributing to a certain systematic error (Bond et al. 2009). The angle truncation indicates the lack of illumination

near 0° and 180° and the nonideal Lambertian light source means that the measured scattered signal is non-sinusoidal. The two drawbacks render the nephelometer measurement less precise.

In order to correct the measurement errors of nephelometer, Anderson and Ogren (1998) used a single parameter as the scattering correction factor (hereinafter CF) to quantify the nonideal effects. CF is defined as the ratio of Mie-calculated scattering coefficient to that measured by the nephelometer and is closely related to the aerosol size and chemical composition. Müller et al. (2011) demonstrated that several methods have been proposed to derive CF. Initially, researchers simulated the nephelometer measurements based on the Mie model. In more detail, they replaced the ideal sinusoidal function with nephelometer's actual scattering angle sensitivity function to derive the scattering coefficient under the condition of nephelometer light source. The scattering coefficient under the condition of ideal Lambertian light is also obtained by the Mie model, thereby calculating the CF. However, this method needs the additional information of particle number size distribution (PNSD), particle shape and refractive index (Quirantes et al., 2008). It is not convenient to obtain simultaneous PNSD data, because the measurement instrument is expensive and not easy to maintain.

An alternative popular correction mechanism is to constrain CF simply by the wavelength dependence of scattering (scattering Ångström exponent (SAE)). Considering that SAE and CF both rely on particle size, Anderson and Ogren (1998) established the linear relationship between them for each TSI nephelometer's wavelength. This ingenious method is convenient, because the scattering properties at different wavelengths, or SAE, can be directly measured by nephelometer itself. However, Bond et al. (2009) found that SAE is also affected by the particle refractive index, while CF is scarcely impacted by it. This difference renders the regression method less accurate. Furthermore, the absorption properties of sampled particles can alter the wavelength dependence of scattering, contributing to errors of this correction method for absorbing aerosols (Bond et al., 2009). Therefore, it is not an accurate correction method to establish the simple linear relationship between a single parameter SAE and CF.

In this study, the measurement limitations of angle truncation and nonideal Lambertian light source are both considered. In view of the disadvantages of these methods mentioned above, we put forward a new correction method of scattering coefficient measurements of a three-wavelength nephelometer, with the use of machine learning model and taking Aurora 3000 correction as an example. A description of data and methodology under dry and other relative humidity conditions is given in Sect. 2. The verifications of linear regression method and our new method are presented in Sect. 3. At last, the conclusions are presented in Sect. 4.

2 Data and Method

2.1 Site description

Eight field observations (Table 1) were conducted at different time periods in China, including two observations in Wuqing (39°38' N, 117°04' E), two in Xianghe (39°76' N, 117°01' E), each observation in Wangdu (38°40' N, 115°08' E), Zhangqiu (36°71' N, 117°54' E), Beijing(39°59' N, 116°18' E) and Gucheng (38°9' N, 115°44' E). Five sites (Wuqing, Xianghe,

Wangdu, Zhangqiu, Gucheng) are located in suburban areas, representing the characteristics of regional anthropogenic aerosols in the Northern China Plain. Measurement in Beijing was conducted at Peking University (downtown Beijing), surrounded by two heavy traffic roads, and hence it can well represent the typical case of urban pollution. The number size distribution measurements of eight campaigns cover a wide range of 10-1000 nm (Fig.1), and the single scattering albedo (SSA) varies between 0.235 and 0.997 (Fig.2).

Table 1. The summary of eight field observations used in this paper.

Site	(1)Wuqing	(2)Wuqing	(3)Xianghe	(4)Xianghe	(5)Wangdu	(6)Zhangqiu	(7)Beijing	(8)Gucheng
Date	07 March-04 April	12 July-14 August	22 July-30 August	09 July-08 August	04 June-14 July	23 July-24 August	25 March-09 April	15 October-25 November
Year	2009	2009	2012	2013	2014	2017	2017	2016
PNSD	TDMPS +APS	TDMPS +APS	SMPS +APS	TDMPS +APS	TDMPS +APS	SMPS +APS	SMPS +APS	SMPS +APS
BC	MAAP	MAAP	MAAP	MAAP	MAAP	AE33	AE33	AE33
f(RH)	/	/	/	/	TSI 3563	Aurora 3000	Aurora 3000	Aurora 3000

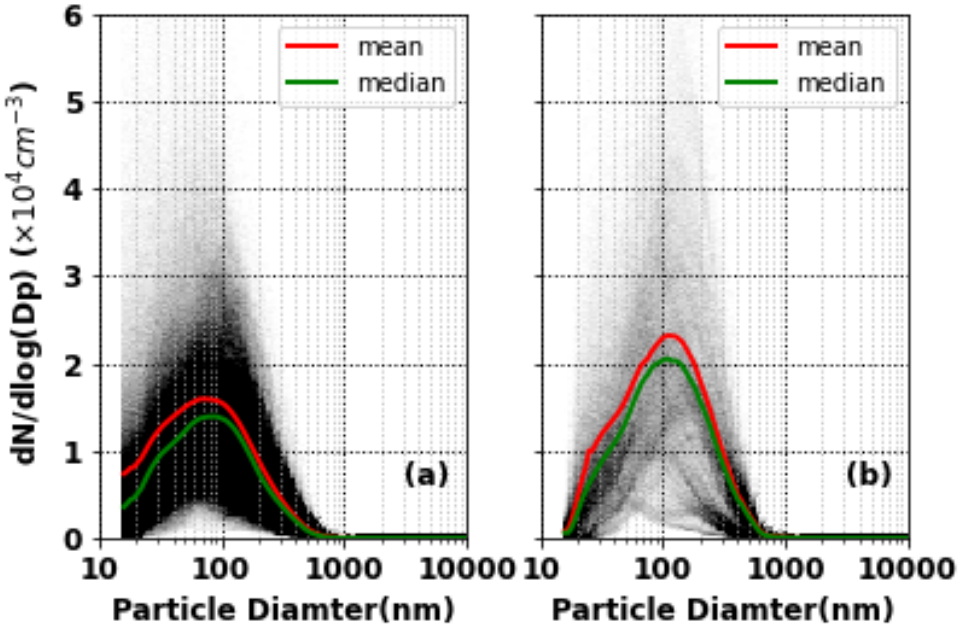


Figure 1. The number size distribution of the measured aerosol in the (a) field observation (1)-(7) and (b) field observation (8).

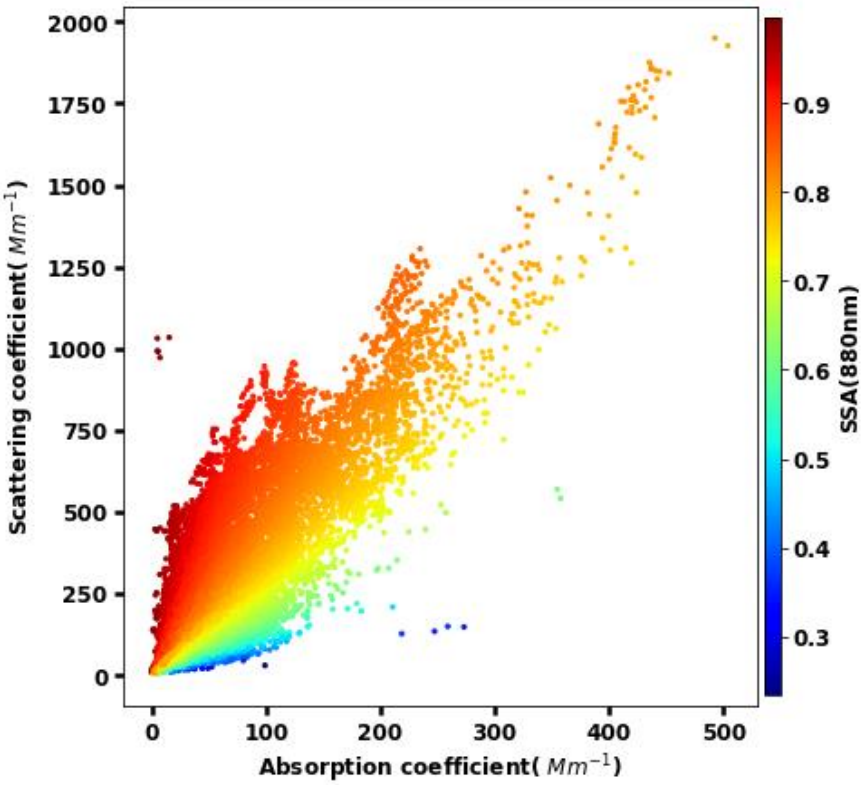


Figure 2. The SSA of field observation (1)-(8).

2.2 Method

This paper expects to put forward a simple and precise method of deriving CF. Inspired by establishing the linear relationship between SAE and CF (Anderson and Ogren, 1998; Müller et al., 2011), this paper first elucidates more parameters that exert impacts on CF and can be directly obtained by nephelometer measurements. Considering of complex relationships among parameters and the requirements of ordinary regression method (e.g. independent variables), it is not an appropriate means to use regression analysis to derive the relationship between CF and some variables at each wavelength. Therefore, a random forest (RF) machine learning model from the scikit-learn machine learning library (Prettenhofer et al., 2011), an effective method that can be used for classification and nonlinear regression (Breiman, 2001), is adopted. The RF model has several advantages (Zhao et al., 2018) as follows: First, it involves fewer assumptions of dependency between observations and results than traditional regression models. Second, there is no need for a strict relationship among variables before implementing model simulation. Third, this model requires much fewer computing resources than deep learning. Finally, it has a lower over-fitting risk. Based on this machine learning model, our new method splits the above datasets into seven training datasets and one test dataset, then uses Mie model and the training datasets to calculate CF. The CFs of training datasets, combined with parameters that impact CF and can be directly obtained from nephelometer, are used to train the machine learning model. The

derived RF models are verified by the test dataset. If the verification results are credible, the RF models can be directly used in field measurements to predict in-situ CF and finally obtain the corrected scattering coefficient.

2.2.1 Correction under dry conditions

An important feature of Mie scattering is that the larger the particle, the more forward scattering, meaning that the ratio of backscattering coefficient to total scattering coefficient, or hemispheric backscattering fraction (HBF), would become smaller. Therefore, HBF can to some extent stand for aerosol size and this paper aims to find out whether HBF can be used as one parameter to predict CF or not. Considering that both SAE and CF relate to particle size, this paper uses the datasets of field observation (1)-(7) to explore the relationship between CF and calculated SAE and HBF at different wavelengths (Fig. 3).

Following the method of Anderson and Ogren (1998) and Müller et al.(2011), we established the linear regression equation between CF and SAE (black dashed lines). It is found that the change of CF could be constrained by the change of SAE to a certain extent, but the data points are dispersed from the regression equation. The larger the HBF, the greater the slope of CF changing with SAE. Therefore, besides SAE, HBF can be utilized to provide extra information on particle size and thus predict CF.

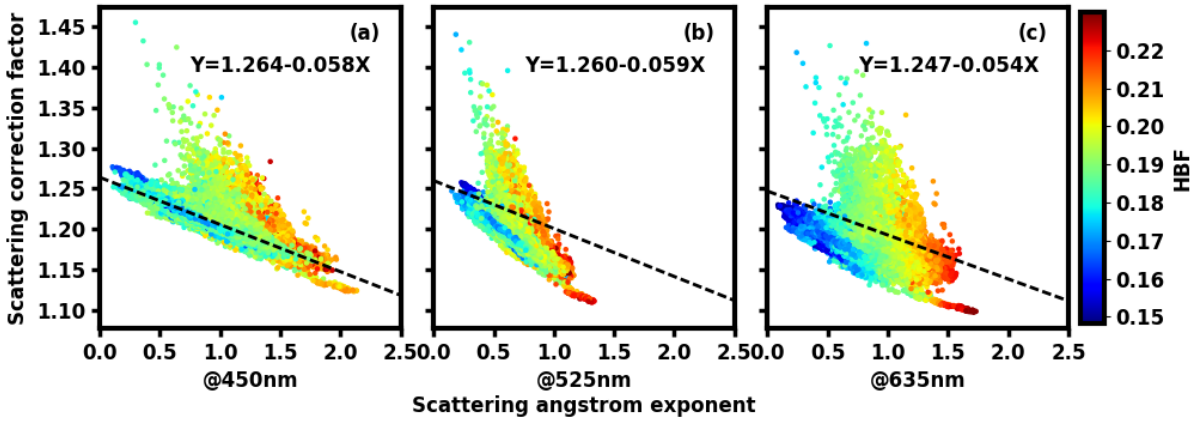
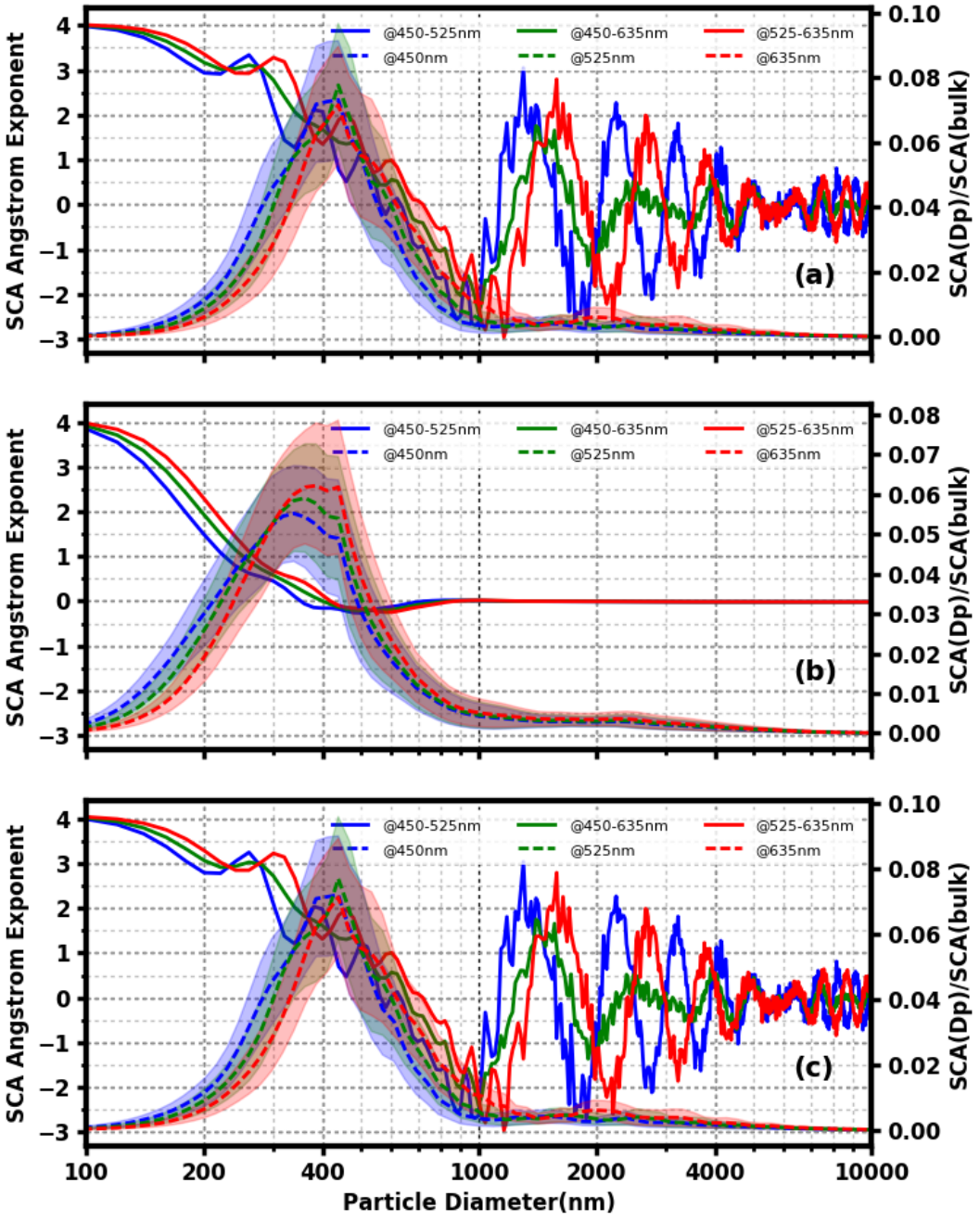


Figure 3. Scattering correction factors versus the scattering Ångström index. (a), (b), (c) respectively represent the results at the wavelengths of 450 nm, 525 nm, and 635 nm. The black dashed line is a statistical linear relationship, and the color of points represents the hemispheric backscattering fraction HBF.

Before establishing the relationship between CF and calculated SAE and HBF, it is necessary to obtain the size range of which particles contribute much to the variations of SAE and HBF. The paper makes the assumption of three independent types of particle composition: scattering particles, absorbing particles, and core-shell mixed particles with the core radius of 35 nm. The refractive index is 1.80-0.54i (Ma et al., 2012) for the absorbing materials and $1.53\text{-}10^{-7}$i (Wex et al., 2002) for the scattering materials. Based on this assumption and all measured size distributions mentioned above, the variation of SAE

at the three wavelength combinations (450+525 nm, 450+635 nm, 525+635 nm) and HBF at the three wavelengths (450 nm, 525 nm and 635 nm) in the particle size range (100 nm-10 μm) is calculated by the Mie model, under the condition of Aurora 3000 light source (The light angle is limited from 10° to 171° and for the detail of the angular sensitivity function, please refer to Müller et al. (2011)). Additionally, to distinguish the particle size range where the change of SAE and HBF can be obviously manifested in the overall optical properties of aerosols, the paper also calculates the ratio of size-resolved scattering and hemispheric backscattering to total scattering for three types of assumed aerosols.

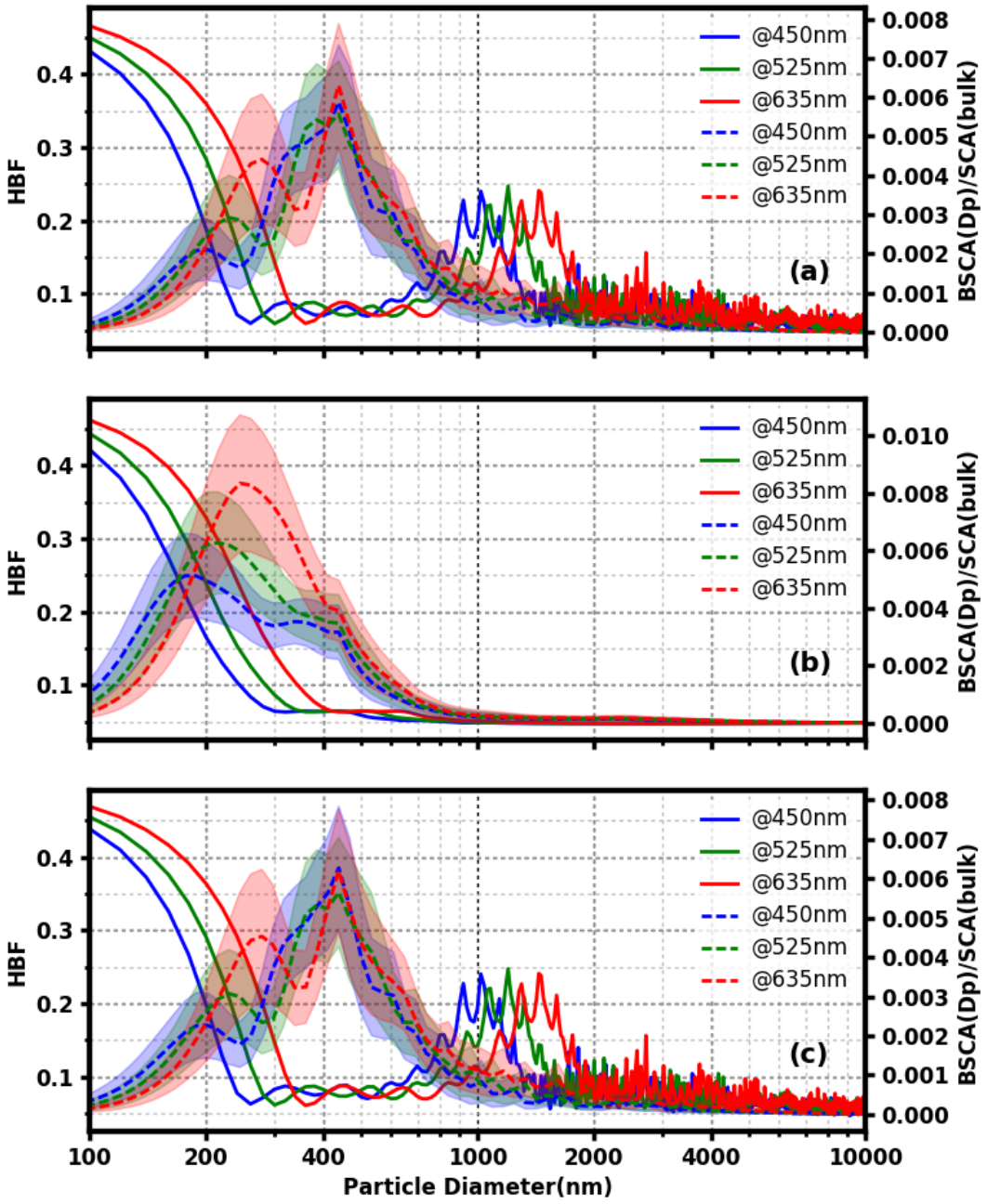
As shown in Fig. 4, for all the three types of aerosols, scattering is mainly concentrated in the size range of 100-1000 nm; particles larger than 1000 nm contribute little to the total scattering, and hence there is no follow-up discussion of SAE change of these large particles. When particles are smaller than 1000 nm, the overall trend of SAE is decreasing with the increase of particle size and SAE calculated at different wavelengths is obviously different. Especially when the particle is greater than 300 nm, the SAE variation with particle diameter is large; while particles in the size range of 100-300 nm contribute little to SAE variations. Therefore, the SAE variability is mostly sensitive to the concentration of particles in the 300-1000 nm size range.



125

Figure 4. The SAE change of scattering particles (a), absorbing particles (b), and core-shell mixing particles of core radius 35 nm (c) with the change in particle diameter (solid line). The dashed lines represent the ratio of scattering at a certain diameter relative to the total scattering.

From Fig. 5, for environmental aerosol particles, the backscattering of particles in the 100-1000 nm range also contributes a lot to the total scattering, and the HBF characteristics of particles greater than 1000 nm are no longer discussed below. For particles with a size less than 300 nm, all three types of aerosol particles show a noticeable HBF variation with the change of particle size. However, particles larger than 300 nm contribute little to HBF variations. In other words, HBF variability is mostly sensitive to the concentration of particles in the 100-300 nm size range.



135 **Figure 5.** The HBF change of scattering particles (a), absorbing particles (b), and core-shell mixing particles of core radius 35 nm (c) with the change in particle diameter (solid line). The dashed lines represent the ratio of hemispheric backscattering at a certain diameter relative to the total scattering.

Based on the above analysis, it is known that SAE and HBF can represent different size information of aerosol particles (300-1000 nm for SAE and 100-300 nm for HBF), finally deriving the particle size information of 100-1000 nm. Therefore, SAE and HBF are two parameters that can be used for machine learning process.

In order to calculate accurate SAE and HBF, scattering and backscattering information should be accurate. Considering that it is also affected by the mass concentration of BC and aerosol mixing states, not only PNSD but also black carbon (BC) data are needed to run the Mie model. According to Ma et al. (2012), when calculating the amount of externally mixed BC and core-shell mixed BC, R_{ext} is used to represent the ratio of the mass concentration of the externally mixed BC (M_{ext-BC}) to that of the total BC (M_{BC}):

$$R_{ext} = M_{ext-BC}/M_{BC}. \quad (1)$$

Ma et al. (2012) pointed out that HBF is sensitive to R_{ext}. Therefore, on the basis of Mie model, we use PNSD, M_{BC} and the assumed R_{ext} value to calculate HBF. Next, the calculation HBF is compared with the observation result of nephelometers. If their difference is minimal, the assumed R_{ext} value is considered true. Deriving mass concentration of BC and PNSD data, assuming that the true R_{ext} is consistent at each size and there is no difference in the radius of core-shell mixed particles with the same size, we can calculate the number size distribution of core-shell mixed BC and externally mixed BC. Furthermore, the refractive index can also be obtained, making it possible to derive more precise information of scattering, backscattering and then SAE and HBF. Details about this method of retrieving PNSD and refractive indices can be found in Ma et al. (2012).

In summary, our correction method of nephelometers under the dry condition encompasses the following procedures (Fig. 6):

(1) Obtain information on particle number size distribution (PNSD), black carbon (BC), and mixing state (R_{ext}) of field observation (1)-(7).

(2) Calculate the scattering and backscattering by Mie model under the conditions of the nephelometer light source at the wavelengths of 450 nm, 525 nm and 635 nm.

(3) Calculate the hemispheric backscattering fraction HBF at the three wavelengths.

(4) Calculate the scattering Ångström index SAE of the three wavelength combinations (450+525 nm, 450+635 nm, 525+635 nm).

(5) Calculate the scattering and backscattering by Mie model under the conditions of the ideal light source at the wavelengths of 450 nm, 525 nm and 635 nm.

(6) Based on the results of the second and fifth steps, calculate theoretical CF at the three wavelengths.

(7) Use six parameters, including three HBF and three SAE, and theoretical CF of each wavelength to train the machine learning model, deriving RF predictor.

(8) Verify the predictive validity of the trained model with the dataset of Gucheng.

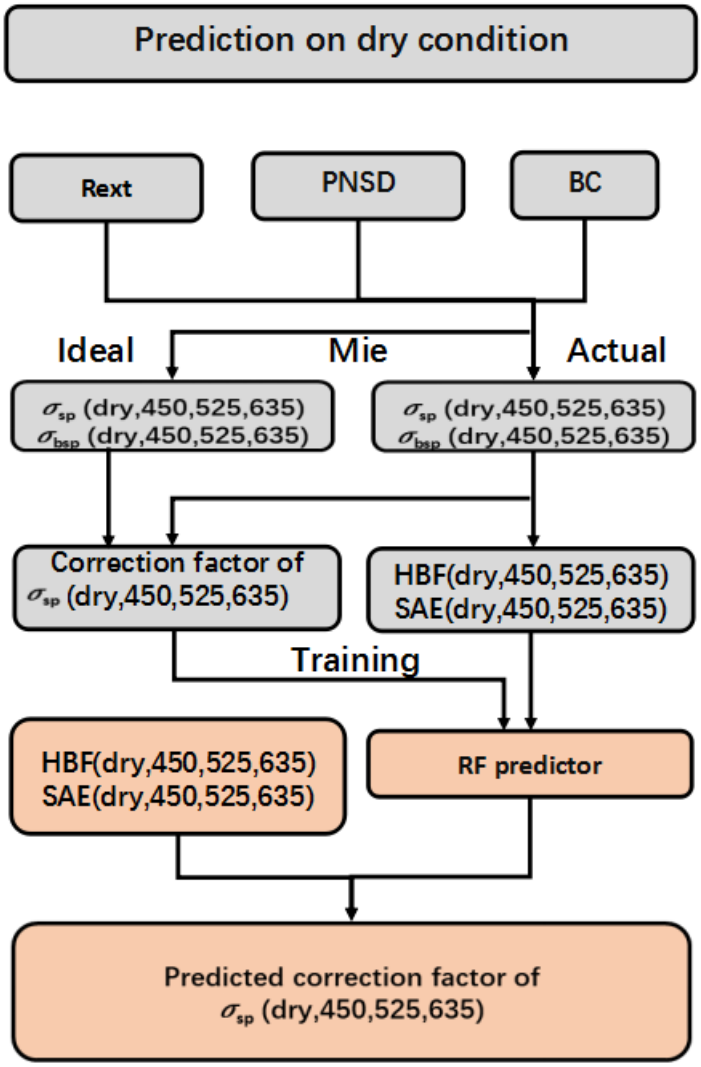


Figure 6. Flow chart of estimating CF under dry conditions by machine learning.

2.2.2 Correction under different RH conditions

Under elevated relative humidity conditions, a correction method taking the hygroscopicity into account is needed, because, with the increment of relative humidity, the non-absorbing component in the aerosol particle can take up water due to its hygroscopicity and then grow up. Accordingly, the water content and particle size may change, resulting in a certain change of CF for the same group of aerosol particles. Therefore, besides SAE and HBF, more parameters relating to hygroscopicity should be considered when deriving CF under elevated relative humidity conditions.

The hygroscopicity or aerosol hygroscopic growth could be indicated by the scattering hygroscopic growth curve f(RH) and the backscattering hygroscopic growth curve $f_b(RH)$: At low relative humidity, the growth due to aerosol taking up water is weak and thus the change of f(RH) and $f_b(RH)$ is small; as relative humidity goes up, the aerosol hygroscopic growth is obvious. Correspondingly, the change of f(RH) and $f_b(RH)$ is large. Referring to researches of Kuang et al. (2017) and Brock et al. (2016), the following formulas are used to describe f(RH) and $f_b(RH)$:

$$\mathrm{f(RH)} = 1 + \kappa_{\mathrm{sca}} \frac{\mathrm{RH}}{100 - \mathrm{RH}} \tag{2}$$

$$f_b(RH) = 1 + \kappa_{\mathrm{bsca}} \frac{\mathrm{RH}}{100 - \mathrm{RH}} \tag{3}$$

where κ_{sca} and κ_{bsca} are fitting parameters representing the hygroscopic growth rate in aerosol scattering and backscattering.

When it comes to the aerosol overall hygroscopicity, according to 24 size distributions of κ obtained from Hachi field observation (Liu et al., 2014), the paper takes their average size distribution (the total volume-weighted κ is 0.281) as the basis; next, in order to obtain a sequence of size distributions of κ, the basis κ is multiplied from 0.05 to 2, with 0.01 as the interval. According to the PNSD of outfield observation (1)-(7) and these assumed size distributions of κ, the theoretical Mie-calculation values are presented as scatter points in Fig. 7. On the basis of the above formulas, the lines represent fitted curves under the condition of nephelometer light source. As can be seen, for the three wavelengths, Eq. (2) and Eq. (3) basically describe the trend of f(RH) and $f_b(RH)$ values. In other words, aerosol scattering and hemispheric backscattering hygroscopic growth can be represented by parameters of κ_{sca} and κ_{bsca}. As a result, we wonder whether or not the hygroscopic growth of CF (hereinafter C(RH)), could be fitted similarly as above formulas with parameter κ_c. The black scatter points in the figure do not lie close to the black dashed lines, and accordingly, the fit formula cannot accurately describe C(RH).

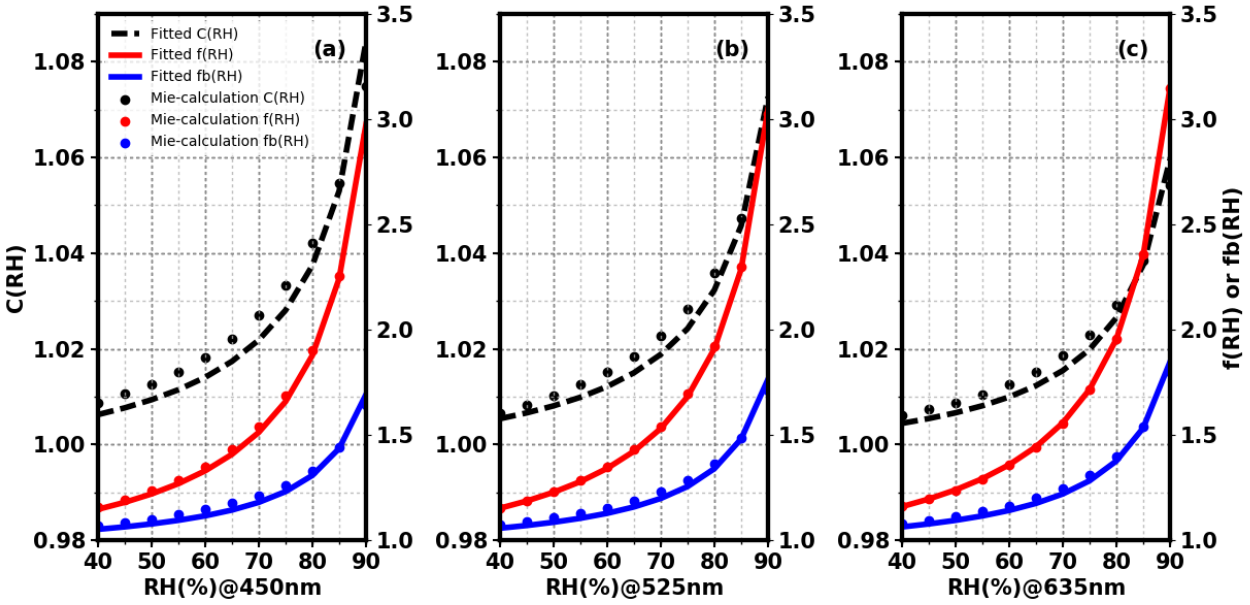


Figure 7. The comparison between κ fitted and theoretical Mie-calculation f(RH), $f_b(RH)$ and C(RH) at the wavelengths of 450 nm (a), 525 nm (b), and 635 nm (c) under the condition of nephelometer light source. The scatter points represent each theoretical Mie-calculation value. The red solid line is the f(RH) fitted curve and the blue solid line is the $f_b(RH)$ fitted curve, both corresponding to the right ordinate value. The black dashed line is the C(RH) fitted curve, which corresponds to the left ordinate value.

Therefore, this paper attempts to derive CF under different RH conditions in a similar machine learning way as described for the dry state. First of all, we need to find out parameters impacting CF under different RH conditions. Aerosol size accounts for CF, referring to Sect. 2.2.1, and thereby SAE and HBF in the dry state at three wavelengths are needed. Besides, hygroscopicity matters to a large extent; κ-Köhler theory (Petters and Kreidenweis, 2007) is thus applied, which uses hygroscopicity parameter κ to describe the hygroscopic growth of aerosol particles under different relative humidity conditions:

$$S = \frac{D^3 - D_d^3}{D^3 - D_d^3(1-\kappa)} \cdot \exp\left(\frac{4\sigma_{s/\alpha} \cdot M_{water}}{R \cdot T \cdot D_d \cdot g \cdot \rho_w}\right). \quad (4)$$

Where S is saturation ratio; D is the diameter of the aerosol particle after hygroscopic growth; D_d is the diameter of the aerosol particle in the dry state; $\sigma_{s/\alpha}$ is the surface tension at the interface between the solution and air; T represents absolute temperature; M_{water} is the molar mass of water; R is the universal gas constant and ρ_w is the density of water.

With the information of PNSD, refractive index of dry aerosol, mixing state, size distribution of κ, and water refractive index of $1.33 - 10^{-7}$i (Seinfeld and Pandis, 2006), on the basis of κ-Köhler theory (Eq. (4)), this paper can calculate the aerosol optical parameters at different RH, deriving f(RH) and $f_b(RH)$. Next, Eq. (2) and Eq. (3) are used to fit the curve of f(RH) and $f_b(RH)$ at each wavelength, deriving fitting parameters κ_{sca} and κ_{bsca} which can imply the size-resolved hygroscopicity. Combined with relative humidity, the estimated change of CF with the relative humidity involves up to 13 physical quantities.

To summarize, our correction method of nephelometers under different relative humidity conditions encompasses the following procedures (Fig. 8):

(1) Obtain information on particle number size distribution (PNSD), black carbon (BC), mixing state (R_{ext}), aerosol hygroscopicity parameter (κ), and relative humidity RH of field observation (1)-(7).

(2) Calculate the scattering and backscattering by Mie model under the conditions of the nephelometer light source at the wavelengths of 450 nm, 525 nm and 635 nm in the dry state.

(3) Calculate the hemispheric backscattering fraction HBF at the three wavelengths under dry conditions.

(4) Calculate the scattering Ångström index SAE of the three wavelength combinations (450+525 nm, 450+635 nm, 525+635 nm) under dry conditions.

(5) Under different relative humidity conditions and assumptions of aerosol hygroscopicity, according to the κ-Köhler theory, aerosol scattering and hemispheric backscattering after the hygroscopic growth are calculated on the basis of nephelometer light source at three wavelengths.

(6) Calculate f(RH) and $f_b(RH)$ curves of the three wavelengths based on the scattering and hemispheric backscattering under dry and different relative humidity conditions.

(7) Calculate the fitting parameters of κ_{sca} and κ_{bsca} from f(RH) and $f_b(RH)$.

(8) Calculate the scattering and hemispheric backscattering after the hygroscopic growth under the conditions of the ideal light source at three wavelengths.

(9) Based on the results of the fifth and eighth steps, calculate theoretical CF at the three wavelengths.

(10) Use thirteen parameters, including three HBF and three SAE, relative humidity RH, three κ_{sca} and three κ_{bsca} at this RH, and theoretical CF of each wavelength to train the machine learning model, deriving the RF predictor.

(11) Verify the predictive validity of the trained model with the dataset of Gucheng.

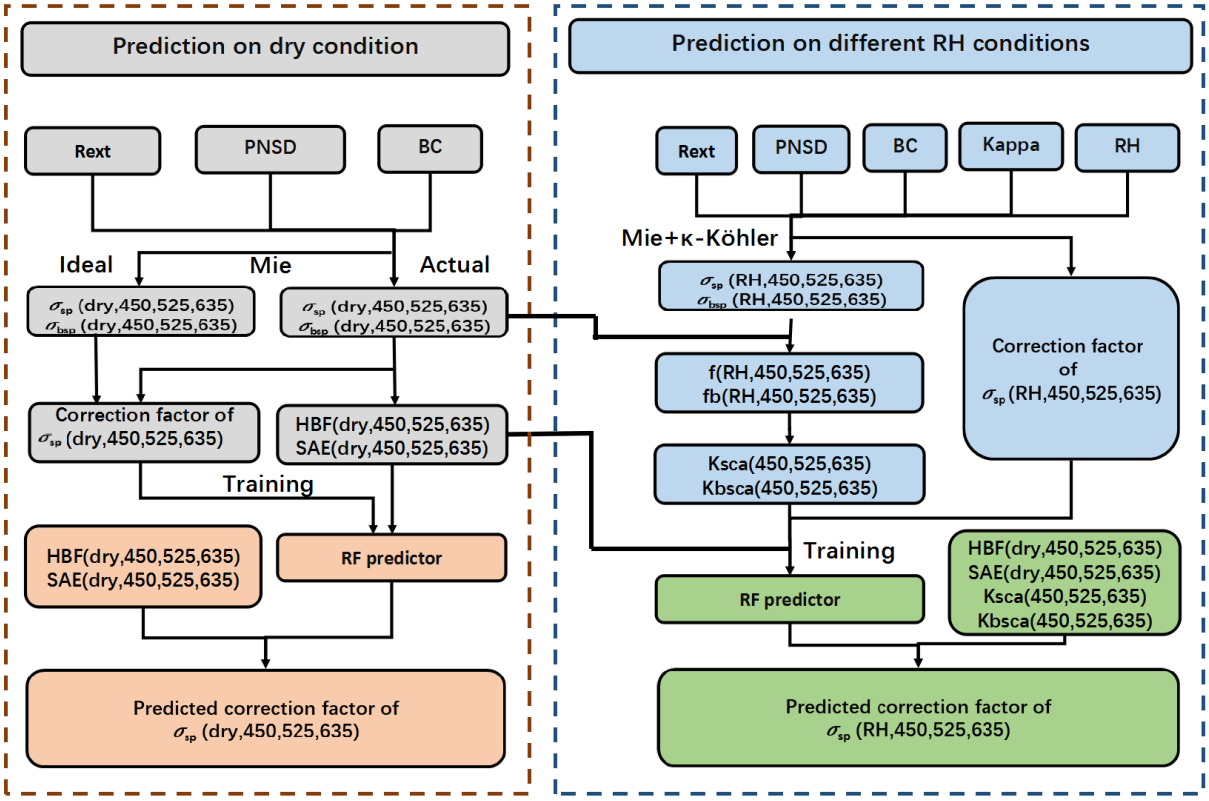


Figure 8. Flow chart of estimating CF under different relative humidity conditions by machine learning.

3 Results and discussions

In order to verify the methods introduced above, on the basis of Gucheng data and the derived RF predictor, we have predicted CF and compared it with the theoretical Mie-calculated CF. First of all, for comparison, Gucheng data are used to verify the simple linear parameterization shown in Anderson and Ogren (1998) and Müller et al. (2011).

3.1 Verification of linear regression method

The PNSD and BC data of Gucheng are used to establish linear fit relationships between CF and the corresponding SAE at three different wavelengths (450 nm, 525 nm and 635 nm), which are respectively represented as:

$$\mathrm{CF} = 1.264 - 0.058\mathrm{SAE}, \quad (5)$$

$$\mathrm{CF} = 1.260 - 0.059\mathrm{SAE}, \quad (6)$$

$$\mathrm{CF} = 1.247 - 0.054\mathrm{SAE}. \quad (7)$$

As shown in Fig. 9, CF ranges between 1.1 and 1.35. There is a relatively large gap between the predicted results derived from linear relationships and the theoretical simulation result, especially at the wavelengths of 525 nm and 635 nm.

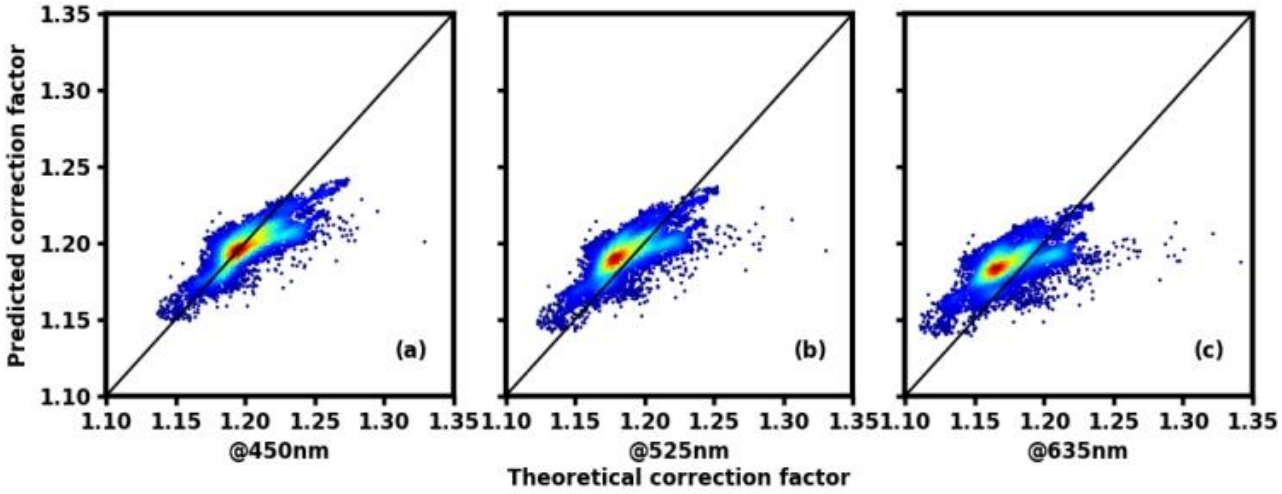


Figure 9. The comparison of theoretical correction factor and predicted correction factor calculated by the Ångström index at different wavelengths; figure (a), (b), (c) are the comparison results for 450 nm, 525 nm and 635 nm, respectively. The black solid line represents that theoretical correction factor equals to the predicted one. The color of data points represents the data density; the warmer the hue, the denser the data points.

When aerosols take up water and then grow up, with Gucheng data, this paper establishes different linear statistical relationships under different relative humidity conditions in order to estimate CF. The data points gradually become dispersed from the 1:1 line as the relative humidity increases (Fig.10). The reason is that, under the condition of high humidity, the hygroscopic growth and thus particle size can vary greatly due to differences of aerosol hygroscopicity. Moreover, refractive indices also show a great distinction owing to the change of water content.

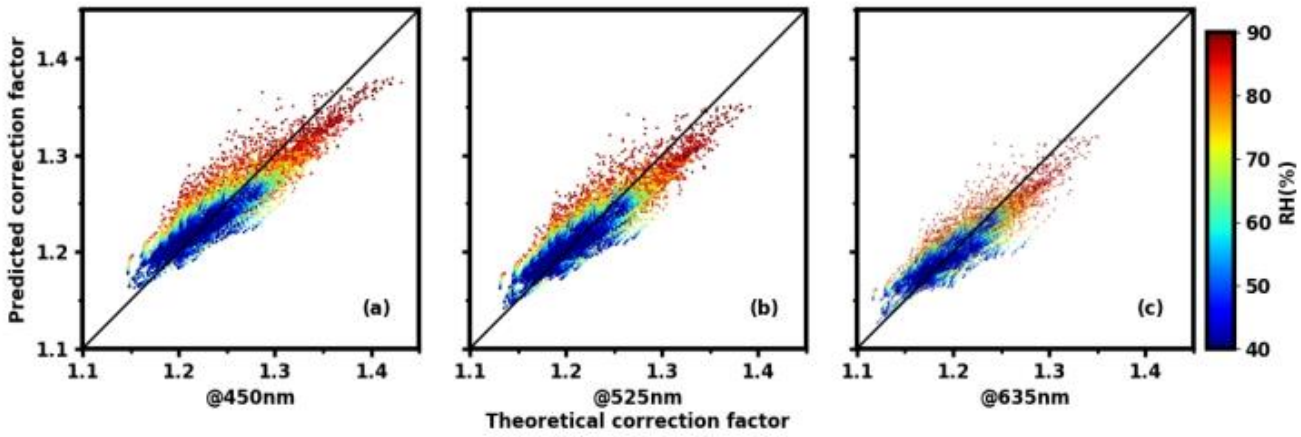


Figure 10. The comparison of theoretical correction factor and predicted correction factor calculated by the Ångström index at different wavelengths; the figure (a), (b), (c) are the comparison results for 450 nm, 525 nm and 635 nm, respectively. The black solid line represents that theoretical correction factor equals to the predicted one. The color of data points represents different relative humidity conditions.

Therefore, the ordinary linear regression method of establishing relationship between CF and a single parameter SAE (Anderson and Ogren, 1998; Müller et al., 2011) cannot be applied to most cases, especially under the condition of high relative humidity.

3.2 Under dry conditions

When it comes to the results of our new method, as shown in Fig. 11, for 450 nm and 525 nm, the prediction performance is relatively good, and the correlation coefficient between prediction value and the theoretical Mie-calculation is 0.88 and 0.84, respectively; more than 90 % of the points fall within the error range of 2 %, and most of them are basically concentrated near the 1:1 line. For 635 nm, the result is slightly worse, with the correlation coefficient at 0.76 and 85.88 % of points in error by less than 2 %. In general, compared with the traditional correction method, our method does not need to consider whether or not the aerosol has strong or wavelength dependent absorption, which improves the accuracy of the CF estimation in the dry state; in addition, the input parameters can be obtained by the nephelometer's observation.

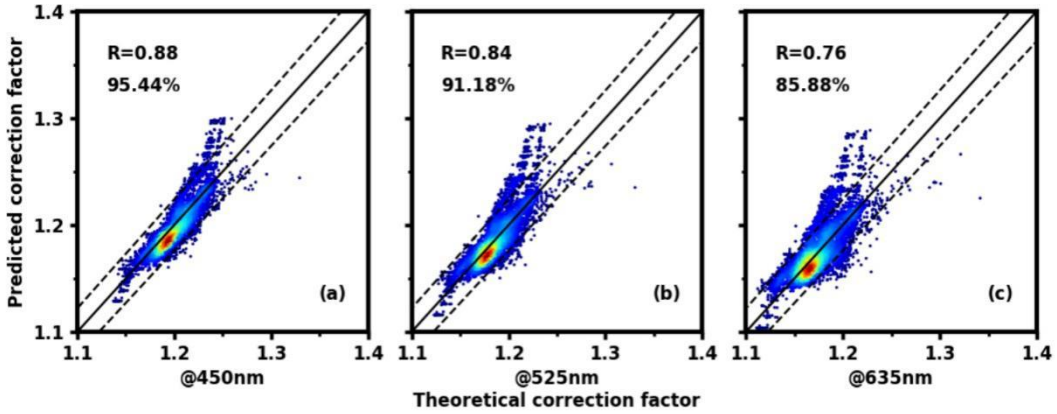


Figure 11. On dry conditions, the comparison of the correction factors calculated by our method and theoretical Mie-calculation values at the wavelengths of 450 nm (a), 525 nm (b), and 635 nm (c), respectively, with a black solid 1:1 line and two dashed lines representing a deviation of 2 %. The color of the data point represents data density; the warmer the hue, the denser the data point. R is the correlation coefficient, and the percentage indicates the percentile of points falling within the error range of 2 %.

3.3 Under different RH conditions

The paper uses each PNSD of the field observation (1)-(7) and averages them to plot Fig. 12 which represents the variation characteristics of CF with the change of relative humidity and aerosol population hygroscopicity, at three wavelengths of 450 nm, 525 nm and 635 nm, respectively.

Under all the different relative humidity conditions, CF of the 450 nm is the largest, with that of 525 nm coming second, and that of 635 nm is the smallest (Fig. 12). All CFs at the three wavelengths increase with the increment of relative humidity. Furthermore, if the relative humidity remains constant, CF also increases with aerosol hygroscopicity increasing. It is reasonable since the environment relative humidity and the hygroscopicity of aerosols have positive impacts on particle sizes and thus CF.

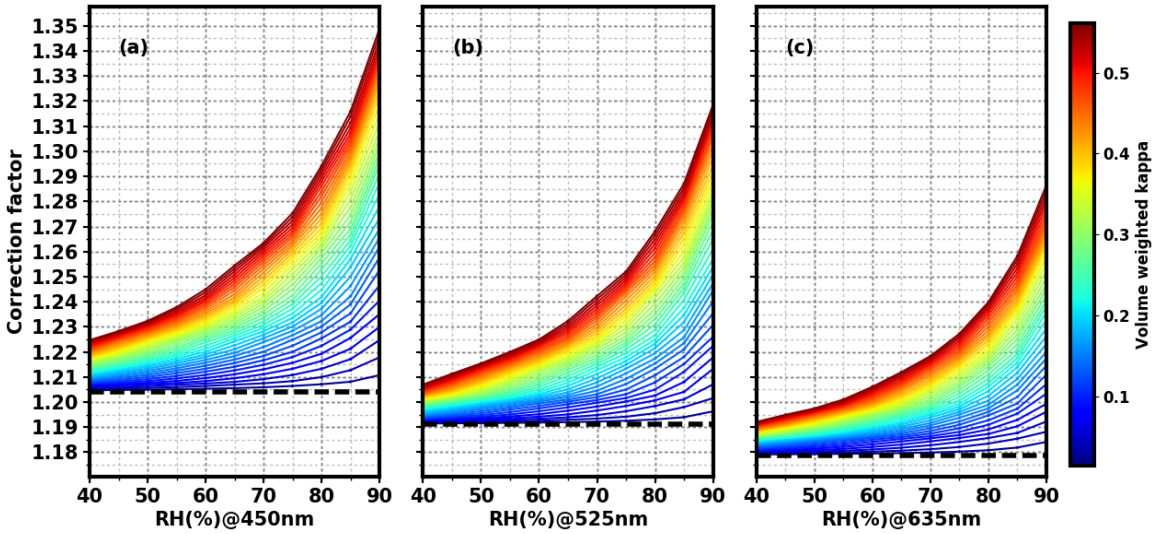


Figure 12. The theoretical calculation of the scattering correction factors (CF) versus relative humidity (RH) and hygroscopicity κ at the wavelengths of 450 nm (a), 525 nm (b), 635 nm(c). The dashed line represents scattering correction factor in the dry state, and the color represents the overall hygroscopicity of aerosols (κ). The color bar is derived from multiplying the total volume-weighted κ of 0.281 by 0.05 to 2, with 0.01 as the interval.

Our correction method under different RH conditions takes the humidity and hygroscopicity into account. As depicted in Fig. 13, the new method predicts CF very well at all the three wavelengths, and nearly all scatter points at the three wavelengths are centered near the 1:1 line. For the 450 nm wavelength, the correlation coefficient between prediction value and theoretical Mie-calculation reaches 0.99, with 99.54 % of the points falling within the error range of 2 %; for the 525 nm wavelength, the correlation coefficient is 0.98, with 98.99 % of the points falling within the error range of 2 %; for the 635 nm wavelength, the correlation coefficient is 0.95, with 96.37 % of the points in error by less than 2 %. From graphs of (d), (e) and (f), the new method's estimation of CF is basically consistent in accuracy at each relative humidity. Another advantage of our new method is that all these input parameters can be obtained by the nephelometer's observation, achieving the goal of self-correction.

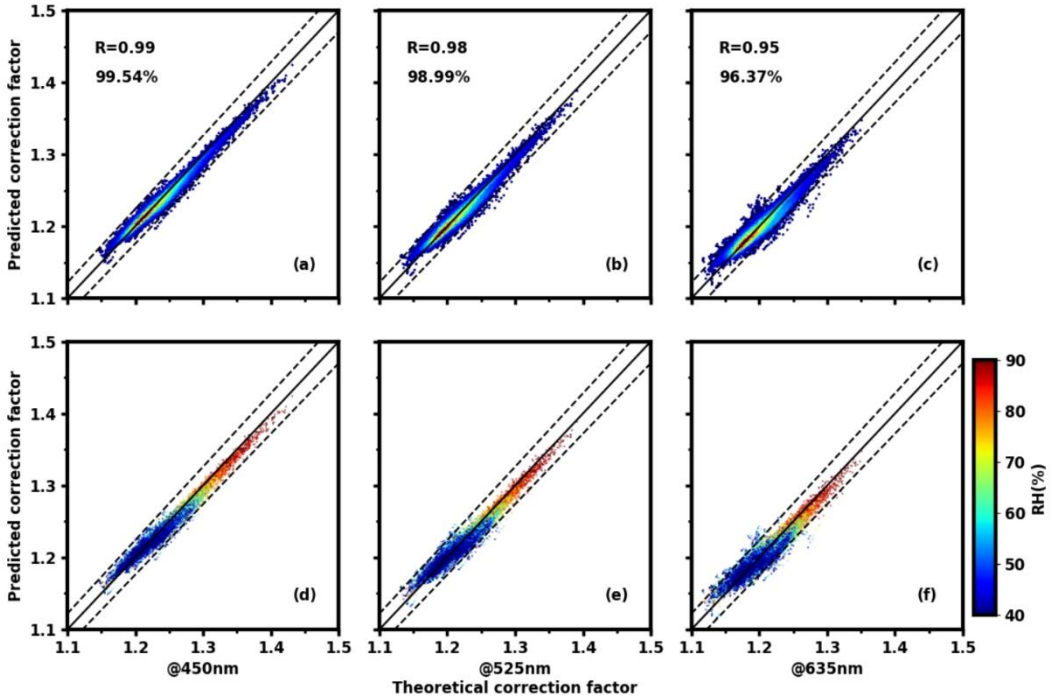


Figure 13. On different relative humidity conditions, the comparison of the correction factors calculated by our method and theoretical Mie-calculation values at the wavelengths of 450 nm (a), 525 nm (b), and 635 nm (c), respectively, with a black solid 1:1 line and two dashed lines representing a deviation of 2 %. The color of the data point represents data density; the warmer the hue, the denser the data point. R is the correlation coefficient, and the percentage indicates the percentile of points falling within the error range of 2 %. The data in (d), (e), (f) are the same as those in (a), (b), (c), but the color here stands for different relative humidity conditions rather than density.

4 Conclusions

The aerosol scattering coefficient is an essential parameter for estimating aerosol direct radiative forcing, which can be measured by nephelometers. However, nephelometers have the problem of non-ideal Lambertian light source and angle truncation, and hence the observed scattering coefficient data need to be corrected. The scattering correction factor (CF) is thus put forward and it depends on the aerosol size and chemical composition. The most direct calibration method is to combine the particle number size distribution, black carbon data and Mie scattering model to correct the nephelometer. However, this method requires auxiliary measurement data. Later, scattering Ångström index (SAE) measured by nephelometer itself is utilized to establish the linear relationship with CF. After verification, it is found that the method lacks precision and accuracy. Therefore, our paper has proposed a new method of nephelometer self-correction.

Under dry conditions, after analysis, SAE and HBF can represent different ranges of aerosol particle size information (300-1000 nm for SAE and 100-300 nm for HBF). With the use of the existing observation results of PNSD, black carbon and

R_{ext} to obtain SAE and HBF, the paper applies random forest (RF) machine learning model to establish the relationship between CF and calculated SAE and HBF, deriving the trained RF model. With the dataset of Gucheng, the verification results show that this method is relatively accurate. The commonly used integrating nephelometer can derive in-situ scattering and backscattering coefficients at three wavelengths to calculate three SAE and three HBF. Therefore, with the use of derived RF model and nephelometer calculation of SAE and HBF, CF could be predicted by the nephelometer itself.

Under other relative humidity conditions, the humidified nephelometer system is utilized. In addition to the dry aerosol particle size information, we should also consider the change in water content and particle size brought by the growth of aerosol taking up water. This paper finds that CF increases with the increment of relative humidity and aerosol hygroscopicity. Therefore, on the basis of κ-Köhler theory, the existing observation results of PNSD, black carbon, R_{ext}, aerosol hygroscopicity parameter κ, and relative humidity are used to run the Mie model, obtaining the theoretical CF and 13 quantities relating to the change of CF under different RH conditions. Similarly, the machine learning model is trained to obtain the relationship between CF and the 13 quantities. With the dataset of Gucheng, the verification results show that the accuracy of CF obtained by this method is very high. The humidified nephelometer system can observe scattering and hemispheric backscattering coefficients at three wavelengths under both dry and elevated RH conditions, obtaining corresponding f(RH) and $f_b(RH)$ under the nephelometer light source condition. As a result, all the 13 quantities, including six physical quantities of SAE and HBF representing dry aerosol size at each wavelength, six fitting parameters κ_{sca} and κ_{bsca} representing particle size-resolved hygroscopicity at each wavelength, and the relative humidity, can be directly obtained from nephelometers. Therefore, with the use of derived RF model and the above 13 quantities, CF could be predicted in situ by humidified nephelometer system.

The strengths of our new method are summed up as follows: Under either dry or any other relative humidity conditions, the prediction performance of CF at three wavelengths is excellent. Furthermore, at each relative humidity, the accuracy of CF estimation is almost the same. All inputs can be obtained through the nephelometer's observation, achieving self-correction; that is, on the basis of ensuring accuracy of correction, there is no need for other aerosol microphysical observations.

When it comes to the weaknesses, due to limitations of the Mie theory, our method cannot be applied to analyse datasets which include desert and marine aerosols and hence further studies are needed. In this study, the new method is put forward only based on datasets in the Northern China Plain. There might be errors in applying our RF models to predict CF all over the world. Therefore, more field observation datasets are needed to verify and perfect this method, hopefully establishing a database of RF model in the future.

Data availability. The data used in this study is available when requesting the authors

Author contributions. Jie Qiu, Wangshu Tan, Gang Zhao, Yingli Yu and Chunsheng Zhao discussed the results; Wangshu Tan offered his help in the coding; Jie Qiu wrote the manuscript.

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

Acknowledgments. This work is supported by the National Natural Science Foundation of China (41590872).

References

Anderson, T. L., Covert, D. S., Marshall, S. F., Laucks, M. L., Charlson, R. J., Waggoner, A. P., Ogren, J. A., Caldow, R., Holm, R. L., Quant, F. R., Sem, G. J., Wiedensohler, A., Ahlquist, N. A., and Bates, T. S.: Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer, J. Atmos. Ocean. Tech., 13, 967-986, doi:10.1175/1520-0426(1996)013<0967:Pcoahs>2.0.Co;2, 1996.

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

Bond, T. C., Covert, D. S., and Müller, T.: Truncation and angular-scattering corrections for absorbing aerosol in the TSI 3563 nephelometer, Aerosol Sci. Tech., 43, 866-871, doi:10.1080/02786820902998373, 2009.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, doi:10.1023/a:1010933404324, 2001.

Brock, C. A., Wagner, N. L., Anderson, B. E., Attwood, A. R., and Murphy, D. M.: Aerosol optical properties in the southeastern United States in summer- part 1: hygroscopic growth, Atmos. Chem. Phys., 16, 4987-5007, doi:10.5194/acp-16-4987-2016, 2016.

IPCC: Climate Change 2013 – The Physical Science Basis: Contribution of the Working Group I to the Fifth Assessment Report of the IPCC, Cambridge University Press, New York, NY, 2013.

Kuang, Y., Zhao, C. S., Tao, J. C., Bian, Y. X., 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, doi:10.5194/acp-17-6651-2017, 2017.

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, doi:10.5194/acp-14-2525-2014, 2014.

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

Müller, T., Laborde, M., Kassell, G., and Wiedensohler, A.: Design and performance of a three-wavelength LED-based total scatter and backscatter integrating nephelometer, Atmos. Meas. Tech., 4, 1291-1303, doi:10.5194/amt-4-1291-2011, 2011.

Petters, M. D., and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961-1971, doi: 10.5194/acp-8-6273-2008, 2007.

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, J. Mach. Learn. Res., 12, 2825–2830, doi:10.1524/auto.2011.0951, 2011.

Quirantes, A., Olmo, F. J., Lyamani, H., and Alados-Arboledas, L.: Correction factors for a total scatter/backscatter nephelometer, J. Quant. Spectrosc. Radiat. Transfer, 109, 1496-1503, doi:10.1016/j.jqsrt.2007.12.014, 2008.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, New York, USA, 701-1118, 2006.

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, 8122, doi:10.1029/2000jd000234, 2002.

Zhao, G., Zhao, C. S., Kuang, Y., Bian, Y. X., Tao, J. C., Shen, C. Y., and Yu, Y. L.: Calculating the aerosol asymmetry factor based on measurements from the humidified nephelometer system, Atmos. Chem. Phys., 18, 9049–9060, doi:10.5194/acp-18-9049-2018, 2018.